



N° d'ordre : 2525

Lois Gamma multivariées pour le traitement d'images radar

THÈSE

présentée et soutenue publiquement le 25 octobre 2007

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Toulouse

(spécialité traitement de l'image et mathématiques appliquées)

par

Florent Chatelain

Composition du jury

<i>Président :</i>	Gérard Letac	Professeur émérite à l'U.P.S., Toulouse
<i>Rapporteurs :</i>	Patrice Abry Philippe Réfrégier Cédric Richard	Directeur de Recherche à l'E.N.S. de Lyon Professeur à l'Institut Fresnel de Marseille Professeur à l'U.T.T., Troyes
<i>Examineurs :</i>	André Ferrari Jordi Inglada Sophie Lambert-Lacroix Jean-Yves Tourneret	Professeur à l'université de Nice Sophia Antipolis Ingénieur de Recherche, C.N.E.S, Toulouse Maître de Conférence à l'U.J.F., Grenoble Professeur à l'E.N.S.E.E.I.H.T., Toulouse

Remerciements

Rédiger les remerciements me permet de me remémorer, avec un grand plaisir, les moments importants qui sont venus ponctuer ces trois années de doctorat, ainsi que les personnes qui m'ont accompagné et encouragé. Ces années à Toulouse ont, en effet, été l'occasion de rencontrer de nombreuses personnes qui m'ont beaucoup apporté du point de vue scientifique, mais également de partager des amitiés fortes et durables.

Ce travail n'aurait pas pu voir le jour sans mon directeur de thèse, Jean-Yves TOURNERET. Que mettre en avant parmi ses qualités : son enthousiasme communicatif, son énergie débordante, ses qualités humaines ou son sens aigu des jeux de cartes? Merci pour la qualité de son encadrement tant d'un point de vue scientifique qu'humain, mais aussi pour la confiance et la liberté accordées durant ma thèse. Il m'a également offert le luxe de mener un grand nombre de collaborations durant mes recherches et de participer à de nombreuses conférences ou écoles d'été, de Toulouse à Hawaï! Je souhaite à tout doctorant la chance d'avoir un tel directeur de thèse, dont l'encadrement a déjà suscité de nombreuses vocations...

Merci aux membres du jury de thèse, à commencer par les rapporteurs Patrice ABRY, Philippe RÉFRÉGIER et Cédric RICHARD, que je remercie vivement d'avoir accepté cette difficile tâche de relecture. Je remercie chaudement André FERRARI, qui m'a guidé et accompagné dès le début de ma thèse en partageant notamment ses connaissances et ses idées. Sophie LAMBERT-LACROIX qui m'a accueilli à Grenoble, et avec laquelle j'ai eu beaucoup de plaisir à travailler. Jordi INGLADA pour toutes les applications qui ont motivé une bonne partie de ma thèse et les collaborations fructueuses qu'il m'a proposées. Je tiens également à remercier tout particulièrement Gérard LETAC, pour sa gentillesse et l'intérêt qu'il a toujours montré pour mes problèmes et mes applications. Son aide, si précieuse, et ses arguments, toujours lumineux, ont guidé ma thèse. Qu'il soit en outre remercié pour m'avoir fait l'honneur de présider mon jury. Merci également aux membres du jury venus de loin, qui ont fait le déplacement en pleine grève des avions au risque de se retrouver piégés sur Toulouse!

Merci à Sylvie pour l'aide précieuse qu'elle apporte à tous ses doctorants. Merci également à tous les autres membres du labo pour leurs conseils, encouragements, et pour la sympathie qu'il m'ont témoigné : Nathalie (merci aussi pour tous les chocolats), Corinne, Marie, Martial (toujours prêt à monter une mission prospective), Jérôme (toujours prêt à partager un mail ou une vidéo), André-Luc (qui a toujours pris régulièrement des nouvelles de Cyrielle), Benoît, Marie-Laure, Jean-Luc, Manu, Daniel, Julien, Xavier et Marc que j'ai eu la chance de retrouver au labo...

Un grand merci pour mes parents, leur soutien et leur aide, tout particulièrement durant les derniers mois de la thèse. C'est donc avec le plus grand plaisir que je remercie ma très chère maman, pourvoyeuse en petits plats et relectrice appliquée, Raphaël et papa. Je n'oublie pas Michèle, Jean-Luc, Alexis et Nathalie et plus généralement toute ma famille pour son soutien. Merci également à Monique et Robert, qui m'ont toujours si bien accueilli dans le calme des sous-bois tarnais. Merci aussi pour leur générosité, grâce à laquelle cette thèse a pu être arrosée comme il se doit.

C'est maintenant l'occasion de remercier toutes les personnes dont j'ai pu croiser la route durant ces années et qui ont toutes, à leur manière, embelli ces années de thèse. A commencer par Audrey – dont la trop grande spontanéité devant mes changements capillaires estivaux n'occulte cependant pas la profonde gentillesse – ainsi que sa complice Virginie avec lesquelles j'ai eu la

chance de partager le bureau. Alexandra et Rahim, piliers du grand bureau des thésards. Jean-Rémi, Henri et Frédéric, nouveaux venus dans le labo mais toujours prêts à rendre service, qui ont eu le mérite de supporter mes longues et quotidiennes interprétations musicales lorsque nous partagions le même bureau. Je n'oublie pas ceux du TéSA, que ce soit Sarah et Marie-Jo (toujours au top pour gérer nos missions), ou Anchalee, Mariana et Xavier (toujours au top pour participer aux dites missions). Jérôme, doctorant du bureau à côté. Nicolas, compagnon de galère mais précurseur en tout, véritable bouée de sauvetage pour tous les problèmes techniques et administratifs auxquels je fus confronté et dont la gentillesse ne gâche rien. Farid, merci à toi l'homme puma pour m'avoir initié aux trésors trop méconnus du 7ème art. PJ celui qui voulait qu'on l'appelle JP (toujours prêt à lancer une bonne boule de neige quand il le faut). Vincent H., plus grand blagueur du labo, qui outre sa bonne humeur, a toujours des anecdotes incroyables et/ou des descriptions de maladies terrifiantes, à partager. Mathieu, une idole pour beaucoup et avec qui je garde de super souvenirs de randos dans les pyrénées. Vincent M., organisateur lors de nos inoubliables escapades, à connotation scientifique, à Florence et Hawaï ("c'est génial!"). Toute la communauté des doctorants lui sera également redevable de la tradition de "l'anglaise". A commencer bien sûr par Wilou, icône du PL ("c'est grâce à Wilou qu'on est là!") dont l'amitié m'est si précieuse et qui m'a tant marqué et influencé durant ces années toulousaines que ce soit au PL, sur les berges de la Garonne, ou ailleurs... C'est d'ailleurs l'occasion de remercier tous nos partenaires du PL : Aniela et Denis, Val, Hélène, Héloïse et Juju, Antoine, Olivier que j'ai également croisé dans l'ancien TéSA, Cyril, Fred, Zarah, Amela... sans oublier Yannick et Alex. Mes derniers remerciements vont à Cyrielle, qui m'accompagne avec tant de bonheur depuis de nombreuses années. Merci pour ta patience et ton aide à nulle autre pareille. Ses années à Toulouse furent, à tes côtés, un ravissement et une joie que nous continuons à partager.

Enfin, je dédie cette thèse à mes grand-parents qui, malgré les vicissitudes de la vie, m'ont toujours témoigné un soutien et un amour inestimable. A Josette donc, ma très chère grand-mère si rudement mise à l'épreuve durant ces derniers mois, et Henri qui restera pour moi, mais aussi pour beaucoup d'autres, un modèle, et dont je suis si fier d'être le petit fils.

A Josette et Henri

Résumé

Dans de nombreux systèmes d'imagerie, et notamment dans des systèmes d'imagerie active tels que le radar, l'amplitude du front de l'onde étudiée est classiquement modélisée par une loi gaussienne complexe. Les intensités mesurées correspondent alors au module au carré de lois gaussiennes complexes. L'estimation des paramètres, appliquée à des problèmes de détection, nécessite de connaître alors précisément les statistiques des intensités reçues. L'objet de cette thèse consiste à étudier des extensions multivariées des lois paramétriques, telles que les lois Gamma, rencontrées dans de tels systèmes d'imagerie. Cette modélisation multivariée permet, en effet, de tenir compte des dépendances statistiques entre des images d'une même scène acquises à des dates différentes et/ou par des capteurs différents. C'est pourquoi les familles de lois multivariées étudiées s'avèrent particulièrement utiles pour résoudre de nombreux problèmes d'estimation et de détection rencontrés en traitement de l'image, telles que la détection de changements temporels ou l'analyse polarimétrique d'une scène.

Abstract

The wavefront amplitude of many optical systems can be modeled as a sum of complex components distributed according to Gaussian distributions. In particular, this is the case for active imaging systems such as radar systems. The resulting intensity measurements are the sum of the squared modulus of these complex Gaussian components. Parameter estimation and detection problems require to determine accurately the statistical properties of the collected intensities. The subject of this thesis consists of studying families of multivariate gamma based distributions useful to solve estimation and detection problems in different image processing applications. Multivariate extensions of the uni-dimensional gamma distribution make possible to model the correlation between images of a same scene acquired at different times and/or by different sensors. Therefore, these distributions are interesting for parameter estimation and detection in many image processing applications, such as in change detection problems or in polarimetric analysis.

Table des matières

Glossaire	xiii
------------------	-------------

Introduction	1
---------------------	----------

Chapitre 1	
Lois Gamma Multivariées	5

1.1 Généralités	6
1.1.1 Domaines d'application historiques des lois Gamma Multivariées	6
1.1.2 La loi Gamma Univariée	6
1.1.3 Modèles statistiques multivariés et copules	10
1.2 Lois Gamma Bivariées	11
1.2.1 Une zoologie de familles de lois Gamma Bivariées	11
1.2.2 Le modèle des lois Gamma Bivariées	16
1.3 Lois Multivariées	26
1.3.1 Une zoologie de familles de lois Gamma Multivariées	27
1.3.2 Le modèle des Lois Gamma Multivariées (LGM)	29
1.3.3 Lois multinomiales négatives (LMN)	38
1.3.4 Lois de Poisson Mélangées Multivariées (LPMM)	42

Chapitre 2	
Recalage et détection de changements pour des images RSO mono-capteur	47

2.1 Modélisation statistique des images RSO	48
2.1.1 Phénomène de speckle et loi Gamma	49
2.1.2 Mesures de similarité entre images RSO	52

2.2	Images RSO mono-capteur	56
2.2.1	Estimation par maximum de vraisemblance	57
2.2.2	Estimation par la méthode des moments	61
2.2.3	Estimation sur données synthétiques	63
2.2.4	Recalage pour des images réelles	64
2.2.5	Détection de changements	69
2.3	Conclusion	75

<p>Chapitre 3</p> <p>Détection de changements pour des images RSO multi-capteurs</p>	77
----------------------------------------------------------------------------------------------------	-----------

3.1	Lois Gamma multivariées multi-capteurs	78
3.1.1	Définition	78
3.1.2	La loi Gamma bivariée multi-capteurs	79
3.1.3	Moments d'une LGBMu	80
3.1.4	Densité de Probabilité d'une LGBMu	81
3.1.5	Mesures de similarités pour une LGBMu	81
3.2	Estimation des paramètres des LGBMu	83
3.2.1	Méthode du Maximum de Vraisemblance (MV)	83
3.2.2	Fonctions d'Inférence par les Marges (FIM)	85
3.2.3	Méthode de Moments	88
3.3	Données synthétiques	89
3.3.1	Méthodes du maximum de vraisemblance et des moments	89
3.3.2	Fonctions d'inférence par les marges (FIM) et méthode du maximum de vraisemblance (MV)	91
3.4	Détection de Changements	92
3.4.1	Images synthétiques	92
3.4.2	Images Réelles	94
3.5	Comparaison des détecteurs mono et multi-capteurs	97
3.5.1	Estimation à partir de vecteurs synthétiques	97
3.5.2	Détection de changements à l'aide d'images synthétiques et réelles	98

3.6	Estimation des nombres de vues équivalents	100
3.7	Conclusion	104

Chapitre 4

Lois Gamma multivariées pour l'imagerie polarimétrique 105

4.1	Imagerie polarimétrique	107
4.1.1	Principe d'un système imageur en imagerie polarimétrique cohérente . . .	107
4.1.2	Lumière totalement polarisée	109
4.1.3	Lumière partiellement polarisée	109
4.2	Imagerie polarimétrique "fort-flux"	112
4.2.1	Distribution du vecteur des intensités	113
4.2.2	Estimation de la matrice de polarisation et du DdP	114
4.2.3	Estimation sur données synthétiques	122
4.2.4	Estimation à partir d'images synthétiques	126
4.3	Imagerie polarimétrique "faible-flux"	129
4.3.1	Modèle statistique	129
4.3.2	Estimation de la matrice de polarisation et du DdP	131
4.3.3	Estimation sur données synthétiques	136
4.3.4	Estimation à partir d'images synthétiques	139
4.4	Conclusion	141

Conclusions 143

Annexes

Annexe A

Transformées d'éléments aléatoires

A.1	Fonction génératrice	147
A.2	Transformée de Laplace	150

Annexe B

Familles Exponentielles Naturelles

B.1	Définition	157
B.2	Moments d'une FEN	160
B.3	Autres paramétrisations d'une FEN	161
B.4	Fonction Variance d'une FEN	163
B.5	Transformation affine et convolution de FEN	164
B.6	Estimateurs du maximum de vraisemblance	166

Annexe C

Lois Gamma Multivariées

C.1	Loi Gamma Bivariée	169
C.1.1	Log-convexité de la transformée de Laplace bivariée	169
C.1.2	Moments joints d'une loi Gamma bivariée	170
C.1.3	Développement de la loi Gamma bivariée en série de polynômes de Laguerre	172
C.1.4	Approximation de l'information mutuelle d'une loi Gamma bivariée lorsque $r \rightarrow 1$	172
C.2	Loi Gamma multivariée	174
C.2.1	LGM éuicorrélée	174
C.2.2	LGM à structure de corrélation markovienne	175
C.2.3	Développement en série de polynômes de Laguerre de la densité d'une LGM	177
C.3	Loi négative multinomiale bivariée	178
C.3.1	Masses de probabilité d'une loi négative multinomiale bivariée	178
C.3.2	Valeurs admissibles des paramètres d'une loi multinomiale négative bivariée	179

Annexe D

Estimation des paramètres des lois Gamma multivariées

D.1	Estimateur du maximum de vraisemblance d'une loi Gamma bivariée	181
D.1.1	Dérivée de la log-vraisemblance	181
D.1.2	Log-concavité de la fonction f_q	182
D.2	Loi Gamma bivariée multi-capteurs	182
D.2.1	Densité	183
D.2.2	Calcul numérique de la fonction de Horn Φ_3	184

D.2.3	Dérivées de la fonction de Horn Φ_3	185
D.2.4	Estimateurs du maximum de vraisemblance	185
D.3	Lois Gamma multivariées dont le polynôme affine est quadratique	186
D.3.1	Inverse d'une matrice appartenant à l'espace vectoriel H	186
D.3.2	Estimateurs du maximum de vraisemblance pour le modèle des trois images d'intensité polarimétriques	187
D.4	Propriétés des estimateurs du maximum de vraisemblance composite des lois mul- tinomiales négatives	189

Liste des publications

Bibliographie	197
----------------------	------------

Glossaire

Acronymes

BCR	Borne de Cramer-Rao
DdP	Degré de Polarisation
EMV	Estimateur du Maximum de Vraisemblance
EMVC	Estimateur du Maximum de Vraisemblance Composite
EQM	Erreur Quadratique Moyenne
FEN	Famille Exponentielle Naturelle
FIM	Fonctions d'Inférences par les Marges
LGB	Loi Gamma Multivariée
LGBMu	Loi Gamma Multivariée Multi-capteurs
LGM	Loi Gamma Bivariée
LGMMu	Loi Gamma Bivariée Multi-capteurs
LMN	Loi Multinomiale Négative
LMNB	Loi Multinomiale Négative Bivariée
LPMM	Loi de Poisson Mélangée Multivariée
MdM	Méthode des Moments
MCNL	Moindres Carrés Non Linéaires
RSO	Radar à Synthèse d'Ouverture
v.a.	variable ou vecteur aléatoire
i.i.d.	variables ou vecteurs aléatoires indépendants et de lois identiques

Notations

x	scalaire
X	variable aléatoire
\mathbf{x}	vecteur ou matrice, selon le contexte
\mathbf{X}	vecteur ou matrice aléatoire, selon le contexte
$\mathbf{0}$	vecteur nul $(0, \dots, 0)^T$
$\mathbf{1}$	vecteur $(1, \dots, 1)^T$
$ \mathbf{a} $	déterminant de la matrice \mathbf{a}
$\text{tr}(\mathbf{a})$	trace de la matrice \mathbf{a}
$\text{etr}(\mathbf{a})$	exponentielle de la trace de la matrice \mathbf{a}
\mathbf{I}_d	matrice identité de taille $d \times d$
μ	mesure
$\mathcal{M}(E)$	ensemble des mesures de Radon positives sur l'espace euclidien E et non concentrées sur un hyperplan affine
$\mathbb{E}[\mathbf{X}]$	espérance du vecteur aléatoire \mathbf{X}
$\text{var}(\mathbf{X})$ ou $\text{cov}(\mathbf{X})$	matrice de covariance du vecteur aléatoire \mathbf{X}
$\text{cov}(X, Y)$	covariance du couple (X, Y)
r	coefficient de corrélation
$r(X, Y)$	coefficient de corrélation du couple (X, Y)
$\Pr(A)$	probabilité associée à l'évènement A
$\mathbb{I}_A(\mathbf{x})$	fonction indicatrice sur l'ensemble A : $\mathbb{I}_A(\mathbf{x}) = 1$ si $\mathbf{x} \in A$, 0 sinon
$\delta_{\mathbf{a}}(d\mathbf{x})$	mesure de Dirac au point \mathbf{a}
$\Gamma(x)$	fonction Gamma au point $x > 0$
$(a)_n$	symbole de Pochhammer si $a > 0$ alors $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} \forall n \in \mathbb{N}$. Plus généralement si $a \in \mathbb{R}$, alors $(a)_0 = 1$ et $(a)_{n+1} = (a+n)(a)_n$ pour tout entier naturel n .

Distributions de probabilité

$\mathcal{G}(q, p)$	loi Gamma de paramètres de forme q et d'échelle p
$\mathcal{G}(q, P)$	loi Gamma Multivariée de paramètre le polynôme affine P et de paramètre de forme le scalaire q
$\mathcal{G}(q, P)$	loi Gamma Multivariée Multi-capteurs de paramètre le polynôme affine P et de paramètre de forme le vecteur q
$\mathcal{BN}(q, p)$	loi Binomiale Négative de paramètres (q, p)
$\mathcal{MN}(q, P)$	loi Multinomiale Négative de paramètre le polynôme affine P et le scalaire q
$\mathcal{P}(\lambda)$	loi de Poisson de paramètre λ
$\mathcal{N}(\mathbf{m}, \Sigma)$	loi Normale de moyenne \mathbf{m} et de matrice de covariance Σ

Table des figures

1.1	Densité de la loi $\mathcal{G}(q, 1)$ pour différentes valeurs du paramètre de forme q	7
1.2	Densité de la loi Gamma bivariée de moyennes $m_1 = qp_1 = m_2 = qp_2 = 1$ pour différentes valeurs du paramètre de forme q et du coefficient de corrélation entre les marges r	23
2.1	Image d'un objet uniforme éclairé par un laser [Goo06]	49
2.2	Image RSO de la région d'Izmit, Turquie (CNES)	50
2.3	Information mutuelle et ses approximations lorsque $r \rightarrow 0$ et $r \rightarrow 1$ ($q = 1$).	56
2.4	Courbes typiques de l'opposée de la log-vraisemblance $f(r) = -l(\mathbf{x}; \widehat{m}_{1ML}, \widehat{m}_{2ML}, r)$ en fonction de r ($q = 2, n = 500$).	60
2.5	Courbes typiques de l'opposée de la log-vraisemblance $-l(\mathbf{x}; \widehat{m}_{1ML}, \widehat{m}_{2ML}, r)$ en fonction de r ($q = 2, n = 500$).	61
2.6	Erreurs quadratiques moyennes en fonction de $\log(n)$ pour $r = 0.2$	64
2.7	Erreurs quadratiques moyennes en fonction de $\log(n)$ pour $r = 0.8$	64
2.8	Images Radarsat du volcan Nyiragongo.	65
2.9	Moyennes des coefficients de corrélation estimés en fonction de τ et barres d'erreurs pour les pixels non affectés par l'éruption (pixels noirs de la figure 2.8(c)) (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments) pour différentes tailles de fenêtres.	67
2.10	Histogramme des décalages estimés τ pour les pixels non affectés par l'éruption (pixels noirs du masque 2.8(c)) des images 2.8(a) et 2.8(b) (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).	68
2.11	Moyennes des coefficients de corrélation estimés en fonction de τ pour les pixels blancs appartenant à la région rectangulaire (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).	68
2.12	Histogramme des décalages estimés τ lorsque les images 2.8(a) et 2.8(b) sont corrompues par un bruit additif gaussien (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).	69
2.13	Courbes COR pour données synthétiques pour différentes tailles de fenêtres.	71
2.14	Images Radarsat de Gloucester avant et après inondations.	72
2.15	Courbes COR pour les images de Gloucester pour différentes tailles de fenêtres.	73

2.16	Courbes COR pour différentes valeurs du paramètre q ($n = 15 \times 15$).	74
2.17	Cartes de corrélations pour différentes valeurs du paramètre q (de gauche à droite : $q = 1, q = 2, q = 3$ et le masque des pixels affectés en noir).	74
2.18	Probabilité d'erreur moyenne $Pe = \frac{1}{2}(P_{FA} + P_{ND})$ en fonction du seuil λ pour $n = 9 \times 9$	75
3.1	Courbes typiques de la log-vraisemblance négative en fonction de r' ($q_1 = 2, q_2 = 4, n = 500$) pour différents échantillons.	86
3.2	Courbes typiques de la log-vraisemblance négative en fonction de r' ($q_1 = 2, q_2 = 4, n = 500$) pour différents échantillons.	87
3.3	log EQM en fonction de $\log n$ pour le paramètre r ($r' = 0.2, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	89
3.4	log EQM en fonction de $\log n$ pour le paramètre r ($r' = 0.5, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	90
3.5	log EQM en fonction de $\log n$ pour le paramètre r ($r' = 0.8, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	90
3.6	log EQM en fonction de $\log n$ pour le paramètre m_1 ($r' = 0.8, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	91
3.7	log EQM en fonction de $\log n$ pour le paramètre m_1 ($r' = 0.9, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	91
3.8	Rapport d'efficacité asymptotique (REA) pour les estimateurs de m_1 et r ($q_1 = 1, q_2 = 5, m_1 = 1$ et $m_2 = 1$) en fonction de r'	92
3.9	log EQM en fonction de $\log n$ pour le paramètre r ($r' = 0.9, q_1 = 1, q_2 = 2, m_1 = 100$ et $m_2 = 100$).	92
3.10	Courbes COR obtenues à partir des données synthétiques pour différents couples de paramètres de formes (q_1, q_2) et pour différentes tailles de fenêtre ($n = 9 \times 9$ ou $n = 21 \times 21$).	93
3.11	P_D en fonction des paramètres de forme q_1 et q_2 , ($P_{FA} = 0.3, n = 1$).	94
3.12	clique de 5 pixels	95
3.13	Images Radarsat de Gloucester avant et après inondation.	95
3.14	Courbes COR pour les images de Gloucester ($q_1 = 1, q_2 = 5$).	96
3.15	log EQM en fonction de $\log(n)$ pour le paramètre r' ($q_1 = 1, q_2 = 2$).	97
3.16	log EQM en fonction de $\log(n)$ pour le paramètre r' ($q_1 = 1, q_2 = 5$).	98
3.17	Courbes COR synthétiques pour une taille de fenêtre d'estimation $n = 15 \times 15$	99
3.18	Courbes COR synthétiques pour une taille de fenêtre d'estimation $n = 5 \times 5$	99
3.19	Courbes COR pour les images de Gloucester ($q_1 = 1, q_2 = 5$).	100
3.20	NVE estimés pour les images de Gloucester ($n = 31 \times 31$)	101
3.21	Courbes COR pour les images de Gloucester	102
3.22	Cartes de détection pour les différentes mesures de similarités ($n = 15 \times 15$)	103

4.1	Images d'une scène composée d'éléments en bois, métal et papier illuminée par un laser[BC00]	106
4.2	Principe d'un système imageur pour l'imagerie polarimétrique cohérente [BC00, Fad06].	108
4.3	Ellipse de polarisation du champ électrique E (diagramme de Lissajous)	110
4.4	log EQM des estimateurs à 2 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7 .	124
4.5	log EQM des estimateurs à 3 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7 .	125
4.6	log EQM des estimateurs du DdP à partir de 2, 3 et 4 images en fonction de P^2 ($n = 15 \times 15$).	126
4.7	Composition de la scène polarimétrique.	127
4.8	Images des Intensités mesurées et du DdP.	128
4.9	Estimées du DdP P^2 sur la scène simulée (les nombres insérés représentent les moyennes des estimées de P^2 pour chaque objet) ($n = 15 \times 15$).	130
4.10	log EQM des estimateurs à 2 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7 .	137
4.11	log EQM des estimateurs à 4 images en fonction du logarithme de la taille des échantillons pour la matrice Γ_2 .	138
4.12	log EQM des estimateurs à 4 images en fonction du logarithme de la taille des échantillons pour la matrice Γ_2 .	138
4.13	log EQM des estimateurs du DdP à partir de 2 et 4 images en fonction de P^2 ($n = 15 \times 15$).	139
4.14	Intensités photonisées mesurées et DdP théorique.	140
4.15	Estimées du DdP P^2 sur la scène simulée (les nombres insérés représentent les moyennes des estimées de P^2 pour chaque objet) ($n = 15 \times 15$).	142

Introduction

Contexte

Dans de nombreux systèmes d'imagerie, la description de la lumière, ou plus généralement du champ électrique de l'onde étudiée, ne peut être menée en dehors d'un cadre statistique. C'est en particulier le cas des systèmes d'imagerie cohérents, tels que le radar. L'intérêt de ces systèmes réside dans le caractère actif du capteur. Grâce à quoi l'on peut acquérir rapidement les images, et ceci indépendamment des conditions extérieures, contrairement aux images optiques, soumises par exemple aux aléas météorologiques. La cohérence de l'onde émise est cependant la cause de phénomènes d'interférences multiples qui entachent l'image d'un bruit statistique, appelé *bruit de speckle*. Sous des conditions qui seront précisées par la suite, les intensités mesurées en chaque pixel de l'image suivent alors une loi Gamma.

Le présent travail consiste alors à étudier des extensions multivariées des lois paramétriques rencontrées dans de tels systèmes d'imagerie. Des modèles de lois Gamma multivariées ont en effet reçu une attention soutenue d'un point de vue théorique dans les communautés des probabilités et des statistiques, mais ces familles ont cependant été peu étudiées dans le domaine du traitement d'images. Nous cherchons donc ici à évaluer l'apport de ces modèles multivariés pour des problèmes d'estimation et de détection lorsque l'on dispose de données qui sont corrélées.

En imagerie radar, lorsqu'une série temporelle d'images d'une même scène est disponible, il peut s'avérer intéressant d'estimer les corrélations entre les différentes acquisitions. Puisque chaque image est distribuée suivant une loi Gamma, les lois Gamma multivariées offrent un cadre approprié afin de modéliser les dépendances entre les diverses images. L'estimation des paramètres de telles lois s'applique aux problèmes de recalage d'images ou de détection de changements temporels entre les images. La question est alors de savoir si cette modélisation apporte des améliorations par rapport à des approches non paramétriques, ou des approches qui ne prennent en compte que les seules statistiques unidimensionnelles de chaque image.

L'imagerie polarimétrique consiste à étudier l'état de polarisation de la lumière rétrodiffusée par une scène. Cet état de polarisation reflète les propriétés des matériaux imagés et s'avère pertinent afin de les caractériser. Dans le but d'estimer cet état de polarisation, il est nécessaire d'effectuer plusieurs mesures d'intensités de l'onde rétrodiffusée, obtenues pour différentes polarisations de cette dernière. Sous les mêmes conditions que pour l'imagerie radar, chaque intensité est distribuée suivant une loi Gamma. Afin de déterminer l'état de polarisation de la lumière, et donc les propriétés polarimétriques des matériaux imagés, il convient d'estimer les corrélations entre ces intensités. On peut donc se poser la question de la pertinence du modèle des lois Gamma multivariées par rapport à des stratégies d'estimation classiques, comme des méthodes des moments, qui ne sont pas forcément optimales.

Organisation du manuscrit

Le manuscrit s'organise en 4 chapitres.

Le chapitre 1 est consacré au modèle des lois Gamma multivariées qui sera appliqué à l'imagerie radar ou plus généralement à des problèmes d'optique statistique dans les chapitres suivants. Nous étudions dans un premier temps les propriétés théoriques des lois Gamma bivariées et, dans un second temps, celles des lois Gamma multivariées, qui sont une extension du modèle bivarié en dimension quelconque. Cette étude nous amène à considérer le modèle des lois multinomiales négatives et des lois de Poisson mélangées par une loi Gamma multivariée, utilisé en optique statistique pour modéliser les problèmes de "faible-flux".

Le chapitre 2 traite des problèmes du recalage d'images et de la détection de changements lorsque sont disponibles plusieurs images d'une même scène acquises à des dates différentes. La méthode proposée consiste à modéliser les couples de pixels appartenant à chacune des deux images par une loi Gamma bivariée.

Le chapitre 3 prolonge le chapitre précédent lorsque les capteurs radars ont des caractéristiques différentes. Une nouvelle famille de lois Gamma multivariées est introduite afin de prendre en compte les différences entre les distributions statistiques de chaque image.

Le chapitre 4 est dédié à l'imagerie polarimétrique cohérente. L'étude effectuée aboutit à différentes méthodes d'estimation des propriétés polarimétriques de la lumière en fonction du nombre d'images d'intensité disponibles. Une étude similaire est menée dans le cas dit "faible-flux", où les intensités mesurées sont alors distribuées selon des lois de Poisson mélangées par des lois Gamma multivariées.

Ces chapitres utilisent les notions de transformée de Laplace, ou de fonction génératrice dans le cas discret. Ces outils permettent de caractériser et d'étudier les lois multivariées considérées. Outils, dont les principales propriétés, bien connues dans le cas univarié, sont exposées à l'annexe A. De plus, les conditions selon lesquelles les lois étudiées appartiennent à des familles exponentielles naturelles sont étudiées dans le premier chapitre, et les suivants utilisent les résultats liés aux propriétés de ces familles pour l'estimation des paramètres. Pour ces raisons, l'annexe B présente une introduction à la théorie des familles exponentielles naturelles. Les résultats énoncés dans ces deux annexes, A et B, sont illustrés par les exemples des lois Gamma univariées et des lois binomiales négatives. Ces distributions sont en effet les lois marginales des lois multivariées auxquelles nous nous intéressons dans cette thèse.

Principales contributions

Les principales contributions de cette thèse sont listées ci-dessous.

Chapitre 2 : L'originalité du travail tient à la modélisation des intensités de couples d'images radar par des lois Gamma bivariées. Des méthodes de maximum de vraisemblance sont alors proposées afin d'estimer des mesures de similarité entre ces images.

Chapitre 3 : Nous proposons une nouvelle famille de lois multivariées afin de modéliser des séries d'images radar d'une même scène acquises par des capteurs différents. Des stratégies d'estimation adaptées à ces familles de lois sont alors développées.

Chapitre 4 : Nous montrons que sous des hypothèses classiques en optique statistique, les intensités des images polarimétriques suivent une loi Gamma multivariée. Ceci permet de prouver

que la méthode classique d'estimation de la matrice de polarisation est efficace, et conduit à de nouvelles méthodes d'estimation des propriétés polarimétriques de la lumière lorsque l'on ne dispose que d'un nombre réduit d'images polarimétriques. La même étude est menée en faible-flux où un bruit supplémentaire, le bruit de Poisson, est pris en compte. Des méthodes originales d'estimation sont alors développées.

Conseils de lecture

Le chapitre 1 présente les propriétés théoriques des familles multivariées étudiées dans les chapitres suivants. Une lecture intégrale de ce chapitre n'est cependant pas nécessaire pour la compréhension du reste du travail. Quelques indications sont données sur ce chapitre afin de permettre à chacun de le parcourir selon son bon vouloir.

pp. 16-18 et 21 Présentation des lois Gamma bivariées, principales propriétés et densité associée.

pp. 28-30 Présentation des lois Gamma multivariées.

p. 38 et 40 Présentation des lois multinomiales négatives et expression des masses de probabilité dans le cas bivarié.

pp. 42-44 Présentation des lois de Poisson mélangées multivariées.

pp. 45-46 Lois de Poisson mélangées par une loi Gamma multivariée.

Lois Gamma Multivariées

Sommaire

1.1 Généralités	6
1.1.1 Domaines d'application historiques des lois Gamma Multivariées	6
1.1.2 La loi Gamma Univariée	6
1.1.3 Modèles statistiques multivariés et copules	10
1.2 Lois Gamma Bivariées	11
1.2.1 Une zoologie de familles de lois Gamma Bivariées	11
1.2.2 Le modèle des lois Gamma Bivariées	16
1.3 Lois Multivariées	26
1.3.1 Une zoologie de familles de lois Gamma Multivariées	27
1.3.2 Le modèle des Lois Gamma Multivariées (LGM)	29
1.3.3 Lois multinomiales négatives (LMN)	38
1.3.4 Lois de Poisson Mélangées Multivariées (LPMM)	42

Ce chapitre présente une étude théorique des familles de lois multivariées qui seront appliquées à l'imagerie Radar ou plus généralement à des problèmes d'optique statistique dans les chapitres suivants. Le problème de l'estimation des paramètres de ces lois n'est pas abordé dans ce chapitre et sera traité lors des chapitres suivants en fonction des applications et du modèle statistique qui en découle. Ce chapitre se décompose en trois parties.

La première partie situe le contexte historique dans lequel les modèles de lois Gamma multivariées ont été étudiés. Les principales propriétés de la loi Gamma sur \mathbb{R} sont également rappelées. Enfin la notion de distribution statistique multivariée est précisée.

Les familles de lois Gamma bivariées sont définies et étudiées dans la deuxième partie. Un panorama de familles de lois Gamma bivariées est tout d'abord présenté. Le modèle de loi Gamma étudié par la suite, appelée dans la littérature *loi Gamma bivariée de Kibble*, est ensuite étudié de manière plus approfondie. Les propriétés les plus importantes de ces lois sont obtenues. En particulier, plusieurs représentations possibles de la densité de ces familles de lois sont données, ces diverses formes ayant été étudiées par différents auteurs.

Dans une dernière partie, les familles de lois multivariées qui seront utilisées par la suite sont définies en dimension $d \in \mathbb{N}$ quelconque. Outre les lois Gamma multivariées, les lois multinomiales négatives sont elles aussi caractérisées, le cas bivarié étant étudié de manière plus approfondie. Enfin les lois de Poisson mélangées multivariées qui sont utilisées en optique sta-

tistique pour modéliser les problèmes de “faible-flux” sont définies et étudiées lorsque la loi de mélange est une loi Gamma multivariée.

1.1 Généralités

1.1.1 Domaines d’application historiques des lois Gamma Multivariées

Les modèles de lois Gamma multivariées ont reçu une attention particulière en hydrologie [YOB01]. En effet les données issues de phénomènes hydrologiques correspondent aux réalisations de variables aléatoires positives modélisant par exemple l’intensité des précipitations, la hauteur des crues ou le débit des cours d’eau, le niveau des bassins hydrologiques, ou même la durée des précipitations ou la fréquence d’apparition d’inondations. Le modèle statistique des lois Gamma s’avère particulièrement bien adapté à la modélisation de ces phénomènes. L’approche multivariée permet en outre de prendre en compte les diverses corrélations, que ce soit entre les différentes grandeurs physiques rencontrées en hydrologie [Cla80, Yue01], entre les niveaux de plusieurs bassins ou réservoirs hydrologiques [Pha76] ou entre les évolutions au cours d’une saison du niveau d’un réservoir [War92].

Les modèles de lois Gamma multivariées ont bien sûr été étudiés dans d’autres domaines d’application que l’hydrologie. Par exemple, le couple formé par l’amplitude et la durée des rafales de vent subies par des fusées ou des missiles lors de leur ascension a été modélisé dans [SA81, SAT82] par une famille de loi Gamma bivariée. Ces modèles multivariés sont aussi fréquemment utilisés dans le cadre de la théorie de fiabilité et de l’analyse de survie [BR81, CD05, CT07].

1.1.2 La loi Gamma Univariée

La loi Gamma est une distribution continue dont le support est l’ensemble des réels strictement positifs. Avec la paramétrisation classique, cette distribution notée $\mathcal{G}(q, p)$ admet pour mesure :

$$\mathcal{G}(q, p)(dx) = \frac{e^{-\frac{x}{p}}}{\Gamma(q)} \left(\frac{x}{p}\right)^{q-1} \mathbb{I}_{]0, +\infty[}(x) \frac{dx}{p} \quad \forall (q, p) \in]0, +\infty[^2, \quad (1.1)$$

Notons que cette loi est parfois paramétrée non pas par le paramètre p , mais en fonction de son inverse $\lambda = 1/p$. Dans ce manuscrit, la paramétrisation classique correspondra toujours à la paramétrisation par le couple (q, p) introduite à l’équation (1.1) ci-dessus. Comme cela est montré aux exemples A.2.1 et B.1.1, lorsque le paramètre $q \in]0, +\infty[$ est fixé, la famille $F_q = \{\mathcal{G}(q, p) \mid p \in]0, +\infty[\}$ est une famille exponentielle naturelle (FEN) dont les moments sont

$$\mathbb{E}[X^n] = \frac{\Gamma(q+n)}{\Gamma(q)} p^n \quad \forall n \in \mathbb{N}. \quad (1.2)$$

On obtient en particulier l’expression suivante de la moyenne et de la variance :

$$\begin{aligned} \mathbb{E}[X] &= qp, \\ \text{var}(X) &= qp^2. \end{aligned} \quad (1.3)$$

De plus cette FEN peut être caractérisée par la transformée de Laplace de ses éléments. En effet si $X \sim \mathcal{G}(q, p)$ alors

$$\mathcal{L}_X(z) = \mathbb{E} \left[e^{-zX} \right] = (1 + pz)^{-q}, \quad \forall z \in]-\frac{1}{p}, \infty[, \quad (1.4)$$

et donc F_q est l'ensemble des mesures de probabilité dont les transformées de Laplace correspondent à l'expression (1.4) pour $p \in]0, +\infty[$. Cette distribution est infiniment divisible car quel que soit $p > 0$, la loi Gamma de transformée de Laplace (1.4) existe pour tout $q > 0$. De plus, soient q_1, \dots, q_n n paramètres réels strictement positifs et $p \in]0, \infty[$. Si X_1, \dots, X_n sont n v.a. indépendantes de loi respectives $\mathcal{G}(p, q_i)$ pour $1 \leq i \leq n$, alors la loi de la somme

$$X = X_1 + \dots + X_n,$$

est distribuée suivant la loi Gamma $\mathcal{G}(q, p)$ où $q = q_1 + \dots + q_n$.

Les paramètres q et p issus de la paramétrisation classique (1.1) de la loi Gamma ont reçu une dénomination standard :

- Le paramètre p est appelé le paramètre d'échelle de la loi Gamma pour la raison suivante : si X est une v.a. de loi $\mathcal{G}(q, p)$, alors la v.a. $\frac{X}{p}$ est distribuée suivant la loi $\mathcal{G}(q, 1)$. Les lois Gamma de la FEN F_q où le paramètre q est fixé se déduisent donc toutes les unes des autres par homothétie et les densités sont identiques au facteur d'échelle près. C'est pourquoi la loi Gamma $\mathcal{G}(q, 1)$ est appelée *loi Gamma standard* de paramètre de forme q .
- Le paramètre q est appelé le paramètre de forme. Contrairement au paramètre d'échelle, sa valeur influe en effet sur la forme de la distribution. Ceci est illustré sur la figure 1.1 où la densité $\mathcal{G}(q, 1)$ est représentée pour trois valeurs du paramètre q : $q = \frac{1}{2} < 1$ à la figure 1.1(a), $q = 1$ à la figure 1.1(b) et $q = 2 > 1$ à la figure 1.1(c).

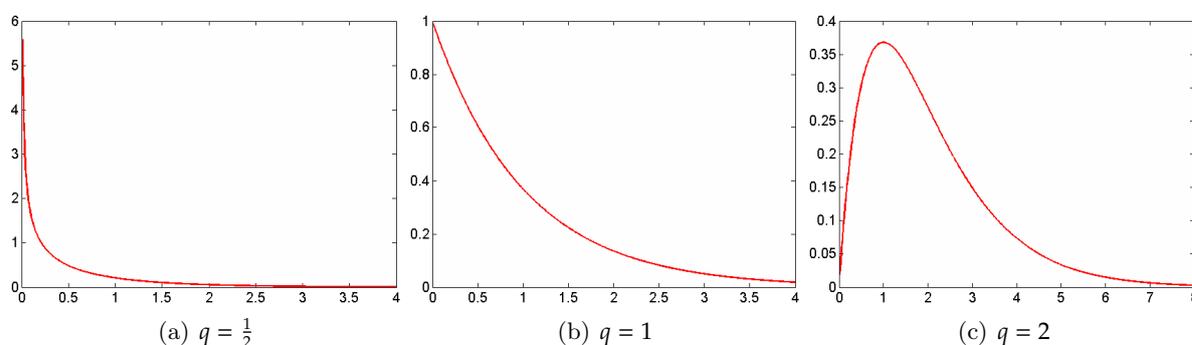


FIG. 1.1 – Densité de la loi $\mathcal{G}(q, 1)$ pour différentes valeurs du paramètre de forme q .

Le cas particulier où le paramètre de forme est $q = 1$ correspond à une loi exponentielle. Cette loi intervient naturellement dans de nombreuses applications et est donc très importante en statistique. En particulier la loi exponentielle est sans mémoire et est utilisée pour modéliser la durée de vie de systèmes sans usure.

La loi dite du χ^2 est un autre cas particulier de la loi Gamma. En effet pour tout entier naturel non nul n , la loi du χ^2 à n degrés de liberté, notée χ_n^2 , correspond à la loi $\mathcal{G}(\frac{n}{2}, 2)$. La loi du χ_n^2 étant la loi de la somme de n v.a. normales centrées réduites au carré, la loi Gamma $\mathcal{G}(q, p)$ lorsque le paramètre de forme $q = \frac{n}{2}$ est un demi entier est obtenue de la façon suivante :

$$\frac{p}{2} (X_1^2 + \dots + X_n^2) \sim \mathcal{G}\left(\frac{n}{2}, p\right) \quad \forall p \in]0, +\infty[, \forall n \in \mathbb{N}^*,$$

où les variables X_i sont des v.a. identiques et indépendantes (i.i.d.) de loi normale centrée réduite.

L'interprétation probabiliste de la loi Gamma lorsque le paramètre de forme q n'est pas un demi-entier est plus complexe. Une des plus simples est la suivante [Yor03] :

$$\left(\int_0^\infty e^{-\frac{t}{q} + B(t)} dt \right)^{-1} \sim \mathcal{G}(q, 1) \quad \forall q \in]0, +\infty[.$$

où $B(t)$ est un mouvement brownien standard sur \mathbb{R} . Ces processus jouent un rôle important dans la formule de Black-Scholes utilisée en mathématiques financières pour modéliser la valeur d'une option. La formule donnée ci-dessus demeure cependant trop complexe pour être utilisée afin de simuler efficacement une loi Gamma. Par conséquent lorsque le paramètre q n'est pas un demi-entier, la loi Gamma est simulée par des méthodes de type acceptation-rejet [MT00], l'expression analytique de la densité (1.1) étant disponible. Même si les interprétations probabilistes de cette loi sont rarissimes, ces cas peuvent être intéressants en imagerie radar où le paramètre de forme correspond à un nombre de vues équivalent (NVE) qui n'est pas forcément un demi-entier comme cela sera expliqué au chapitre suivant.

Loi binomiale négative et la loi Gamma

La loi Gamma univariée est par ailleurs étroitement liée à la loi Binomiale Négative présentée notamment dans les exemples A.1.1, B.1.2 et B.3.2 et dont les propriétés sont rappelées ici.

La loi Binomiale Négative est une distribution discrète dont le support est l'ensemble des entiers naturels \mathbb{N} . Avec la paramétrisation classique, cette distribution notée $\mathcal{BN}(q, p)$ admet pour mesure :

$$\mathcal{BN}(q, p)(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} (1-p)^k p^q \delta_k(dx) \quad \forall (q, p) \in]0, +\infty[\times]0, 1[. \quad (1.5)$$

Comme cela est montré aux exemples A.1.1 et B.1.2, lorsque le paramètre $q \in]0, +\infty[$ est fixé, la famille $F_q = \{\mathcal{BN}(q, p) \mid p \in]0, 1[\}$ est une FEN dont les moments factoriels sont

$$\mathbf{E}[N(N-1)\dots(N-r+1)] = \frac{\Gamma(q+r)}{\Gamma(q)} \left(\frac{1-p}{p} \right)^r \quad \forall r \in \mathbb{N}. \quad (1.6)$$

On obtient en particulier l'expression suivante de la moyenne et de la variance :

$$\begin{aligned} \mathbf{E}[X] &= q \frac{1-p}{p}, \\ \text{var}(X) &= q \frac{1-p}{p^2} > \mathbf{E}[X]. \end{aligned} \quad (1.7)$$

La propriété $\text{var}(X) > \mathbf{E}[X]$ indique la surdispersion des lois binomiales négatives. De plus cette FEN peut être caractérisée par la fonction génératrice de ses éléments. En effet si $N \sim \mathcal{BN}(q, p)$ alors

$$G_N(z) = \mathbf{E}[z^N] = \left(\frac{p}{1-(1-p)z} \right)^q = \left(\frac{1-(1-p)z}{p} \right)^{-q}, \quad \forall |z| < \frac{1}{1-p}, \quad (1.8)$$

et donc F_q est l'ensemble des mesures de probabilité dont la fonction génératrice correspond à l'expression (1.8) pour un certain $p \in]0, 1[$. Cette expression de la fonction génératrice explique le nom de cette loi. En effet si $-q$ était positif on obtiendrait la fonction génératrice d'une loi binomiale, et donc la variable aléatoire N serait distribuée selon une loi binomiale. Ici l'exposant $-q$ est négatif, d'où le nom de la loi.

Comme la loi Gamma, cette distribution est infiniment divisible car quel que soit $p \in]0, 1[$, la loi binomiale négative de fonction génératrice (1.8) existe pour tout $q > 0$. De plus soient q_1, \dots, q_n n paramètres réels strictement positifs et $p \in]0, 1[$, si X_1, \dots, X_n sont n v.a. indépendantes de loi respectives $\mathcal{BN}(p, q_i)$ pour $1 \leq i \leq n$, alors la loi de la somme

$$X = X_1 + \dots + X_n,$$

est distribuée suivant la loi binomiale négative $\mathcal{BN}(q, p)$ où $q = q_1 + \dots + q_n$. Le cas particulier $q = 1$ correspond, à une translation de 1 près, à la loi géométrique $\mathcal{BN}(1, p)$ et modélise le nombre d'échecs avant le premier succès lors de la répétition d'épreuves identiques et indépendantes de probabilité de succès p^1 . Plus généralement, lorsque q est un entier naturel non nul, la loi $\mathcal{BN}(q, p)$ modélise alors le nombre d'échecs avant le q -ième succès (voir la note de bas de page à l'exemple A.1.1).

La loi négative binomiale et la loi Gamma sont liées par des opérations de composition et de mélange.

Proposition 1.1.1. *La loi binomiale négative composée par une loi Gamma est une loi Gamma : si $N \sim \mathcal{BN}(q, p)$ alors la variable obtenue par la composition suivante² $X \sim \mathcal{G}(q + N, 1)$ est distribuée selon la loi Gamma $\mathcal{G}(q, 1/p)$.*

Preuve. La variable X ainsi construite admet pour transformée de Laplace

$$\begin{aligned} \mathcal{L}_X(z) &= \mathbb{E} \left[e^{-zX} \right] = \mathbb{E} \left[\mathbb{E} \left[e^{-zX} \mid N \right] \right], \\ &= \mathbb{E} \left[\left(\frac{1}{1+z} \right)^{q+N} \right] = \left(\frac{1}{1+z} \right)^q G_N \left(\frac{1}{1+z} \right), \\ &= \left(\frac{1}{1+z} \right)^q \left(\frac{p}{1 - \frac{1-p}{1+z}} \right)^q = \left(1 + \frac{z}{p} \right)^q. \end{aligned} \tag{1.9}$$

La deuxième ligne est obtenue grâce à l'expression de la transformée (1.4) de Laplace d'une loi Gamma, et la troisième ligne grâce à l'expression de la fonction génératrice d'une loi binomiale négative (1.8). \square

Un autre résultat classique [Gra97, p. 17] sur ces lois concerne les lois de Poisson mélangées. Les lois de Poisson mélangées sont les familles de lois obtenues lorsque le paramètre X de la loi de Poisson $\mathcal{P}(X)$ est une variable aléatoire. La loi suivie par la v.a. X est alors appelée la loi de mélange et la loi de la v.a. N telle que $N|X \sim \mathcal{P}(X)$ est appelée la loi de Poisson mélangée par la loi de X .

¹Dans la littérature, la loi géométrique de paramètre p correspond au rang du premier succès lors de la répétition d'une épreuve de Bernoulli de probabilité de succès p . Cette loi correspond donc à la loi $\mathcal{BN}(1, p)$ traduite de 1.

²On parle de loi binomiale négative composée par une loi Gamma car la v.a. $X \sim \mathcal{G}(q + N, 1)$ peut être générée de la façon suivante : $X = \sum_{i=0}^N X_i$ où $N \sim \mathcal{BN}(q, p)$ et les v.a. $(X_i)_{i \geq 0}$ sont i.i.d. de loi $\mathcal{G}(q, 1/p)$.

Proposition 1.1.2. *La loi de Poisson mélangée par une loi gamma est une loi négative binomiale : si $X \sim \mathcal{G}(q, \frac{1-p}{p})$ avec $(q, p) \in]0, +\infty[\times]0, 1[$, alors la variable $N \sim \mathcal{P}(X)$ est distribuée suivant la loi binomiale négative $\mathcal{BN}(q, p)$.*

Preuve. La variable N ainsi obtenue admet pour fonction génératrice :

$$\begin{aligned} G_N(z) &= \mathbb{E}[z^N] = \mathbb{E}\left[\mathbb{E}[z^N | X]\right], \\ &= \mathbb{E}\left[e^{X(z-1)}\right] = \mathcal{L}_X(1-z), \\ &= \frac{1}{\left(1 + \frac{1-p}{p}(1-z)\right)^q} = \left(\frac{p}{1 - (1-p)z}\right)^q, \end{aligned} \tag{1.10}$$

la deuxième ligne étant obtenue grâce à l'expression de la fonction génératrice d'une loi de Poisson. La variable N suit donc la loi négative binomiale $\mathcal{BN}(q, p)$. \square

Une conséquence de la proposition précédente est que l'ensemble des lois binomiales négatives $\{\mathcal{BN}(q, p) \mid (q, p) \in]0, +\infty[\times]0, 1[\}$ peut être obtenu par mélange d'une loi de Poisson par une loi Gamma. Bien que le paramètre p de la loi binomiale négative $\mathcal{BN}(q, p)$ ne soit pas un paramètre d'échelle et que sa valeur caractérise aussi la forme de la distribution, par analogie avec la loi Gamma le paramètre q , sera appelé ici le *paramètre de forme*.

1.1.3 Modèles statistiques multivariés et copules

Si les distributions de probabilité sur \mathbb{R} sont toutes définies sans ambiguïté, il n'en n'est pas de même pour les distributions multivariées. En effet, mis à part le cas des lois normales multivariées et des lois de Wishart, lesquelles peuvent être vues comme une généralisation des lois du χ^2 , qui sont clairement codifiées, il n'existe pas de caractérisation unique des autres lois multivariées. En général, les auteurs s'accordent sur le point suivant : étant donné une famille de loi "X" définie sur \mathbb{R} , une loi multivariée est qualifiée de loi multivariée de type X si ses lois marginales (univariées) sont toutes du type X. Le problème est alors qu'il existe un très grand nombre de modèles statistiques vérifiant cette condition. Ces différents modèles peuvent cependant tous être obtenus à partir des loi marginales et de copules.

La notion de copule a été introduite par Abe Sklar en 1959 avant d'être étudiée de façon plus intensive à partir du milieu des années 1980 suite à l'article de Genest et MacKay [GM86]. Parmi les ouvrages de référence sur le sujet, on peut notamment citer les livres de Nelsen et de Joe [Nel99, Joe97].

Une copule est une fonction de répartition multivariée, notée traditionnellement C, définie sur $[0, 1]^d$, d étant la dimension de la loi multivariée, dont les marges sont celles de lois uniformes sur $[0, 1]$. La caractérisation d'une copule C est alors la suivante :

1. C est une fonction définie sur le domaine $[0, 1]^d$
2. $C(u_1, \dots, u_d) = 0$ si une des composantes u_i est nulle, pour tout $1 \leq i \leq d$,
3. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ pour tout $1 \leq i \leq d$,
4. la fonction C est d -croissante sur $[0, 1]^d$.

Les conditions ci-dessus assurent que la copule C est une fonction de répartition dont les lois marginales sont toutes des lois uniformes sur l'intervalle $[0, 1]$. L'interprétation de la notion

de d -croissance se fait en notant qu'une mesure de probabilité est nécessairement positive. Par exemple en dimension $d = 2$, si le couple de v.a. (U, V) admet pour fonction de répartition la copule C , alors

$$\Pr(u_1 < U < u_2, v_1 < V < v_2) = C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2).$$

Par définition, la propriété de 2-croissance sur $[0, 1]^2$ assure alors que $C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0$ pour tout $(u_1, v_1) \in [0, 1]^2$ et $(u_2, v_2) \in [0, 1]^2$.

Le théorème de Sklar affirme que si C est une copule de dimension d , et si F_1, \dots, F_d sont des fonctions de répartition (univariées), alors $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$ est une fonction de répartition de dimension d , dont les marges sont précisément F_1, \dots, F_d . Réciproquement, si F est une fonction de répartition en dimension d , il existe alors une copule C telle que $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, où les F_i sont les lois marginales de F .

Si les fonctions de répartition des lois marginales sont en plus toutes continues, la copule C est alors unique, et donnée par la relation $C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$ (où F_i^{-1} désigne l'inverse généralisée de la fonction croissante F_i). Dans ce cas, on peut alors parler de "la" copule associée à un vecteur aléatoire (X_1, \dots, X_d) .

Afin de construire des distributions multivariées de lois marginales données, de nombreuses familles de copules, correspondant à différentes méthodes de construction, ont été étudiées. Les copules dites archimédiennes [Joe97, p. 86] sont une des familles les plus célèbres.

Pour être exhaustive, l'étude des lois multivariées d'un type donné devrait donc porter sur l'étude de toutes les fonctions copules associées aux lois marginales du type de loi étudié. Cette étude relève alors du cas non paramétrique. Cette approche n'est cependant pas adoptée ici, le travail proposé se limitant à l'étude d'un modèle paramétrique de loi Gamma multivariée précis.

1.2 Lois Gamma Bivariées

Avant d'étudier des extensions multivariées des lois Gamma en dimension d quelconque, il est judicieux d'étudier au préalable le cas bivarié ($d = 2$). Ce cas est intéressant, car il existe en général une expression analytique explicite de la densité des lois bivariées, et certaines constructions de modèles multivariés de dimension supérieure se déduisent des modèles bivariés.

1.2.1 Une zoologie de familles de lois Gamma Bivariées

Un nombre important de modèles de lois Gamma Bivariées a été recensé, notamment dans [KBJ00, chap. 48]. En particulier, l'application de cinq modèles différents de lois Gamma Bivariées à des données hydrologiques a été étudiée dans [YOB01]. Les propriétés de la somme $X + Y$ et le rapport $X/(X + Y)$ de couples de v.a. (X, Y) distribués suivant cinq modèles classiques de lois Gamma bivariées ont aussi été passés en revue dans [NK06]. Quelques-uns de ces modèles, qui portent le nom de leurs auteurs, sont présentés dans cette partie. Pour une revue plus complète, le lecteur peut se référer à [KBJ00, chap. 48]. Dans cette section, les transformées de Laplace sont toutes définies avec la convention $\varepsilon = -1$ selon la définition A.2.1 et sont donc toutes au moins définies sur $[0, \infty]^2$.

La loi Gamma bivariée de MacKay

Cette distribution est issue d'un des premiers modèles de loi Gamma bivariée, et a été introduite par MacKay dans [McK34] pour caractériser la loi jointe du couple formé par la variance empirique d'un échantillon de variables gaussiennes centrées i.i.d. et par la variance empirique associée à un sous-échantillon de l'échantillon précédent. Cette distribution est définie par la densité du couple (X_1, X_2) suivante :

$$p_{X_1, X_2}(x_1, x_2) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)} x_1^{a-1} (x_2 - x_1)^{b-1} e^{-cx_2} \mathbb{I}_{x_2 > x_1 > 0}(x_1, x_2), \quad (1.11)$$

avec $a, b, c > 0$. Les lois marginales d'une telle distribution sont alors des lois Gamma de paramètres de forme respectifs a et $a + b$ et de paramètre d'échelle c : $X_1 \sim \mathcal{G}(a, c)$ et $X_2 \sim \mathcal{G}(a + b, c)$. Le cas où les paramètres d'échelles des lois marginales sont des réels positifs quelconques (pas forcément égaux) s'obtient facilement en appliquant une homothétie sur une des deux marges.

Cette distribution est un des types de loi (IVa) définie par le système de Pearson bivariée [KBJ00, chap. 44]. Elle peut également être obtenue de la manière suivante : soient U, V deux v.a. indépendantes de lois respectives $\mathcal{G}(a, c)$ et $\mathcal{G}(b, c)$. Alors le couple $(X_1, X_2) = (U, U + V)$ admet la densité (1.11) exprimée ci dessus. Cette méthode de construction peut être étendue pour construire des lois Gamma multivariées de dimension quelconque. À partir de cette construction, on obtient directement grâce à l'expression des moments d'une loi Gamma univariée (1.3) que le coefficient de corrélation du couple (X_1, X_2) est :

$$r(X_1, X_2) = \sqrt{\frac{a}{a+b}} \geq 0.$$

Cette structure de corrélation n'est donc pas très souple, car si les paramètres de forme sont fixés alors la valeur de la corrélation entre les marges est imposée selon la formule précédente. De plus ce modèle bivarié est asymétrique et impose la contrainte supplémentaire $X_2 > X_1$.

Grâce à l'expression de la transformée de Laplace d'une loi Gamma univariée (1.4), on obtient également l'expression de la transformée de Laplace du couple (X_1, X_2) :

$$\begin{aligned} \mathcal{L}_{(X_1, X_2)}(z_1, z_2) &= \mathbb{E} \left[e^{-z_1 X_1 - z_2 X_2} \right] = \mathbb{E} \left[e^{-(z_1 + z_2)U} \right] \mathbb{E} \left[e^{-z_2 V} \right] = \mathcal{L}_U(z_1 + z_2) \mathcal{L}_V(z_2), \\ &= (1 + c(z_1 + z_2))^{-a} (1 + cz_2)^{-b}. \end{aligned} \quad (1.12)$$

Ces lois bivariées ont été utilisées pour modéliser le couple formé par le débit d'écoulement de cours d'eau (X_1) et le niveau des précipitations (X_2) [Cla80], où pour modéliser la perméabilité des sols [ADFS03]. Dans ces domaines d'application, les contraintes du modèle évoquées ci-dessus ne posent en effet pas problème.

La loi Gamma bivariée de Cheriyan et Ramabhadran

Ces lois attribuées à Cheriyan et Ramabhadran [Che41, Ram51] ont aussi été étudiées un peu plus tard par Prékopa et Szántzai [PS78] en dimension quelconque (voir section 1.3.1). Ce modèle est obtenu d'une façon assez semblable à celui de MacKay, mais présente l'avantage d'être symétrique. Si Y_0, Y_1 et Y_2 sont des lois Gamma standard (i.e. de paramètre d'échelle

égal à 1) indépendantes, de paramètres de forme respectifs θ_0 , θ_1 et θ_2 , cette loi bivariée est celle du couple (X_1, X_2) tel que :

$$\begin{aligned} X_1 &= Y_0 + Y_1, \\ X_2 &= Y_0 + Y_2. \end{aligned} \tag{1.13}$$

Par construction, les marges X_1 et X_2 du couple ainsi défini sont des lois Gamma standard de paramètres de formes respectifs $\theta_0 + \theta_1$ et $\theta_0 + \theta_2$. Les lois correspondantes de paramètres d'échelle réels positifs quelconques s'obtiennent par homothétie. La densité de cette loi est alors de la forme :

$$p_{X_1, X_2}(x_1, x_2) = \frac{e^{-(x_1+x_2)}}{\Gamma(\theta_0)\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^{\tilde{x}} y_0^{\theta_0-1} (x_1 - y_0)^{\theta_1-1} (x_2 - y_0)^{\theta_2-1} e^{y_0} dy_0,$$

où $\tilde{x} = \min(x_1, x_2)$. Des développements de cette densité en série entière de la variable \tilde{x} ont été étudiés dans [DF61] pour certaines valeurs du triplet $(\theta_0, \theta_1, \theta_2)$. D'autres développements plus généraux en terme de séries de produits de polynômes de Laguerre généralisés ont été obtenus dans [Szá86].

Par ailleurs, à partir de la construction donnée à l'équation 1.13, on obtient directement l'expression du coefficient de corrélation entre les marges :

$$r(X_1, X_2) = \frac{\theta_0}{\sqrt{(\theta_0 + \theta_1)(\theta_0 + \theta_2)}} \geq 0.$$

Par conséquent, lorsque la valeur des paramètres de forme des lois marginales est imposée, par exemple $\theta_0 + \theta_1 = q_1$ et $\theta_0 + \theta_2 = q_2$, on obtient alors que $r(X_1, X_2) = \frac{\theta_0}{\sqrt{q_1 q_2}}$ et, lorsque le paramètre θ_0 parcourt l'intervalle $]0, \min(q_1, q_2)[$, le coefficient de corrélation entre les marges parcourt l'intervalle suivant :

$$0 < r(X_1, X_2) < \sqrt{\frac{\min(q_1, q_2)}{\max(q_1, q_2)}}.$$

Ce modèle est donc à dépendance positive ($r(X_1, X_2) > 0$), et peut être reparamétré en fonction des paramètres de forme q_1 et q_2 des marges et du coefficient de corrélation $r = r(X_1, X_2)$. En particulier lorsque les marges ont le même paramètre de forme $q = q_1 = q_2$, alors les valeurs possibles du paramètre r sont celles de l'intervalle $]0, 1[$.

À partir de la définition du modèle donnée à l'équation (1.13) et de la transformée de Laplace d'une loi Gamma univariée (1.4), l'expression suivante de la transformée de Laplace du couple (X_1, X_2) est finalement obtenue :

$$\begin{aligned} \mathcal{L}_{(X_1, X_2)}(z_1, z_2) &= \mathbf{E} \left[e^{-z_1 X_1 - z_2 X_2} \right] = \mathbf{E} \left[e^{-(z_1+z_2)Y_0} \right] \mathbf{E} \left[e^{-z_1 Y_1} \right] \mathbf{E} \left[e^{-z_2 Y_2} \right], \\ &= \mathcal{L}_{Y_0}(z_1 + z_2) \mathcal{L}_{Y_1}(z_1) \mathcal{L}_{Y_2}(z_2), \\ &= (1 + z_1 + z_2)^{-\theta_0} (1 + z_1)^{-\theta_1} (1 + z_2)^{-\theta_2}. \end{aligned} \tag{1.14}$$

La loi Gamma bivariée de type Farlie-Gumbel-Morgenstern

Les lois bivariées de type Farlie-Gumbel-Morgenstern (FGM) sont des familles dont les lois marginales suivent des lois données obtenues grâce à la célèbre copule proposée par Morgenstern

[Mor56], définie pour $\theta \in [-1, 1]$ comme :

$$C_{\theta}^{\text{FGM}}(u, v) = uv + \theta u(1 - u)v(1 - v).$$

Cette copule a été initialement introduite afin d'obtenir des familles de lois qui incluent le cas où les deux marges sont indépendantes ($\theta = 0$). Un de ses inconvénients est que la corrélation entre les marges d'un tel modèle appartient à un intervalle assez restreint. Il a en effet été montré [SPB78] que pour ces familles de lois bivariées, le coefficient de corrélation entre les marges est toujours compris dans l'intervalle $[-\frac{1}{3}, \frac{1}{3}]$. Ces distributions ont été généralisées par Sarmanov [Sar66] permettant ainsi d'obtenir des structures de corrélation entre les marges plus larges. Ces dernières familles de distributions bivariées ont été récemment étudiées en détail dans [SL04].

Dans le cas où les marges X_1 et X_2 sont des lois Gamma standard de paramètres de forme respectifs q_1 et q_2 , de densités notées p_{X_1} et p_{X_2} et de fonctions de répartition notées F_{X_1} et F_{X_2} respectivement, on obtient l'expression suivante de la densité du couple (X_1, X_2) :

$$p_{(X_1, X_2)}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) [1 + \theta (2F_{X_1}(x_1) - 1)(2F_{X_2}(x_2) - 1)].$$

La transformée de Laplace de cette distribution, calculée dans [GW84], est

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = \prod_{i=1}^2 (1 + z_i)^{-q_i} \left[1 + 4\theta \prod_{i=1}^2 \frac{I(q_i; (1 + z_i)^{-1})}{I(q_i; 1)} \right], \quad (1.15)$$

où

$$I(a; x) = \int_0^x \frac{z^{a-1}}{(1+z)^{2a}} dz.$$

Toujours selon [GW84], le coefficient de corrélation $r = r(X_1, X_2)$ de ce modèle de loi Gamma bivariée est toujours compris (i.e. pour toutes les valeurs de q_1 et q_2) dans l'intervalle $]-\rho, \rho[$ avec $\rho = 1/\pi \simeq 0.3183$, les bornes de l'intervalle étant les valeurs limites du coefficient de corrélation r lorsque $\theta = \pm 1$ et lorsque les paramètres de forme q_1 et q_2 tendent vers l'infini.

La loi Gamma bivariée de Becker et Roux

Cette distribution proposée par Becker et Roux [BR81] a été reparamétrée par la suite [SR89] par Steel et Le Roux. Cette extension des lois Gamma au cas bivarié est issue de modèles physiques pour la théorie de fiabilité et l'analyse de survie. La distribution du couple (X_1, X_2) modélise en effet l'espérance de vie respective de deux composants d'un même système soumis à des perturbations (modèle de chocs) dont les occurrences sont modélisées par quatre processus de Poisson homogènes indépendants. Ainsi les nombres de perturbations S_1 et S_2 subies par chaque composant suivent des processus de Poisson homogènes indépendants de paramètres respectifs $1/\alpha_1$ et $1/\alpha_2$. Cependant si le composant numéro $i \in \{1, 2\}$ tombe en panne, les perturbations reçues par le composant numéro $j = \{1, 2\} \setminus i$ suivent alors un processus de Poisson homogène de paramètre λ_j/α_j .

Les différentes situations physiques modélisées se déduisent des valeurs des paramètres réels positifs λ_1 et λ_2 . Si $\lambda_1, \lambda_2 < 1$, on est alors dans le cas d'un modèle de compétition où les composants concourent pour une même ressource dont les réserves sont limitées. Si au contraire $\lambda_1, \lambda_2 > 1$, on est alors dans le cas d'un modèle de partage de charge et la panne d'un composant

provoque un accroissement des perturbations subies par le second. Si le modèle est maintenant asymétrique, par exemple $\lambda_1 < 1$ et $\lambda_2 > 1$, le composant 1 est alors considéré comme un parasite du composant 2, considéré comme l'hôte : si l'hôte tombe en panne le premier, alors l'espérance de vie du parasite diminue. Inversement si le parasite tombe en panne le premier, l'espérance de vie de l'hôte s'accroît. Enfin le cas $\lambda_1 = \lambda_2 = 1$ correspond à la situation où les deux composants agissent de manière indépendante l'un de l'autre. Les paramètres de la loi du couple (X_1, X_2) ainsi obtenu sont alors les 4 paramètres $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ des processus de Poisson, ainsi que le nombre de perturbations, notés respectivement $q_1 \in \mathbb{N}$ et $q_2 \in \mathbb{N}$, que peut supporter chacun des 2 composants avant sa panne.

Ces lois sont qualifiées d'extension bivariée des lois Gamma car si $\lambda_1 = \lambda_2 = 1$, alors les variables X_1 et X_2 sont, bien sûr, indépendantes et un résultat classique sur les processus de Poisson assure que ces variables suivent des lois Gamma de paramètres d'échelle α_1 et α_2 et de forme q_1 et q_2 respectivement. Cependant dans le cas général $\lambda_1, \lambda_2 > 0$ les marges ne sont pas des lois Gamma. Une expression de la densité de ces lois est donnée dans [SR89] ou dans [KBJ00, p. 447]. Par ailleurs, aucune expression simple de la transformée de Laplace de ces distributions n'est disponible.

La loi Gamma bivariée de Kibble

Ce sont ces familles de lois Gamma bivariées qui vont nous intéresser par la suite. Les propriétés de ces lois seront étudiées de manière plus approfondie dans les sections à venir, ce paragraphe se limitant à une caractérisation de ces lois et à leur historique.

Ces lois ont été introduites en 1933 par Wicksell [Wic33] et Kibble [Kib41] en 1941 qui considèrent les distributions bivariées des couples (X_1, X_2) pour lesquels il existe 4 paramètres p_1, p_2, p_{12}, q , avec $q > 0$, tels que la transformée de Laplace de ces couples soit de la forme :

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = \mathbb{E} \left[e^{-z_1 X_1 - z_2 X_2} \right] = \frac{1}{(1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^q}. \quad (1.16)$$

Ce modèle est qualifié de loi Gamma bivariée car les lois marginales de X_1 et X_2 s'obtiennent en notant que

$$\begin{aligned} \mathcal{L}_{X_1}(z_1) &= \mathcal{L}_{(X_1, X_2)}(z_1, 0) = \frac{1}{(1 + p_1 z_1)^q}, \\ \mathcal{L}_{X_2}(z_2) &= \mathcal{L}_{(X_1, X_2)}(0, z_2) = \frac{1}{(1 + p_2 z_2)^q}. \end{aligned}$$

Les lois de X_1 et X_2 sont donc des lois Gamma, de même paramètre de forme q et de paramètres d'échelles respectifs p_1 et p_2 , notées $\mathcal{G}(q, p_1)$ et $\mathcal{G}(q, p_2)$.

Ces lois ont été étudiées par la suite sous diverses formes par D'jachenko [D'j62], Sarmanov [Sar66] ainsi que par Moran [Mor69] (Jonhson et Kotz ont d'ailleurs baptisé ces distributions *lois Gamma bivariées de Kibble et Moran* [KBJ00, chap. 48]).

Enfin des généralisations de ces lois ont été proposées par Griffiths [Gri69] et Dussauchoy et Berland [DB72] qui ont considéré les couples (Y_1, Y_2) dont la transformée de Laplace est de la forme :

$$\mathcal{L}_{(Y_1, Y_2)}(z_1, z_2) = (1 + p_1 z_1)^{-q_1} (1 + p_2 z_2)^{-q_2} \left(1 - \frac{r p_1 p_2 z_1 z_2}{(1 + p_1 z_1)(1 + p_2 z_2)} \right)^{-q_3}, \quad (1.17)$$

avec $q_1, q_2, q_3 > 0$ et $0 \leq r \leq 1$. À partir de cette transformée de Laplace bivariée, on déduit par un raisonnement identique à celui effectué pour les lois de Kibble que les marges sont distribuées selon les lois gamma $\mathcal{G}(q_1, p_1)$ et $\mathcal{G}(q_2, p_2)$. Lorsque $q_1 = q_2 = q_3 = q$, on retrouve alors la transformée de Laplace de la loi de Kibble. Ces lois sont donc plus générales que les lois de Kibble. Cependant il n'existe pas d'expression analytique simple de la densité de ces lois. Une condition nécessaire [Gri69] pour que la loi associée à cette transformée de Laplace existe est $r q_3 \leq \min(q_1, q_2)$, et si de plus $q_3 < \min(q_1, q_2)$, alors cette transformée de Laplace s'écrit

$$\mathcal{L}_{(Y_1, Y_2)}(z_1, z_2) = (1 + p_1 z_1)^{-(q_1 - q_3)} (1 + p_2 z_2)^{-(q_2 - q_3)} (1 + p_1 z_1 + p_2 z_2 + p_1 p_2 (1 - r) z_1 z_2)^{-q_3},$$

et d'après la proposition A.2.3, le couple (Y_1, Y_2) peut être obtenu comme la somme du couple (X_1, X_2) distribué suivant la loi de Kibble (1.16) de paramètres $p_1, p_2, p_1 p_2 (1 - r)$ et q_3 et du couple (U_1, U_2) dont les marges sont indépendantes et de lois respectives $\mathcal{G}(q_1 - q_3, p_1)$ et $\mathcal{G}(q_2 - q_3, p_2)$. D'autres généralisations sur lesquelles nous reviendrons ont été proposées et étudiées dans [Sar66] et [SA81, SAT82].

1.2.2 Le modèle des lois Gamma Bivariées

Les lois Gamma bivariées étudiées dans ce manuscrit correspondent aux lois de Kibble caractérisées par leur transformée de Laplace donnée à l'équation (1.16). À partir de maintenant, l'expression *loi Gamma bivariée* ne sera employée que pour désigner ces lois.

Une première constatation, directement obtenue à partir de l'expression de la transformée de Laplace (1.16), est que les paramètres de cette loi vérifient les conditions : $q > 0, p_1 > 0, p_2 > 0$. En effet si le couple (X_1, X_2) admet l'expression (1.16) pour transformée de Laplace, l'hypothèse $q > 0$ est importante, sinon les marges X_i ne seraient pas distribuées suivant des lois Gamma. De plus on peut supposer sans perte de généralité que $p_i > 0$, car si $p_i < 0$ alors la variable $-X_i$ suit la loi $\mathcal{G}(q, |p_i|)$, pour tout $i \in \{1, 2\}$. Quitte à effectuer un changement de signe, on peut donc supposer que X_1 et X_2 sont des variables aléatoires positives, i.e. $p_1 > 0$ et $p_2 > 0$ (le cas $p_i = 0$ correspond au cas dégénéré $X_i = 0$ qui est sans intérêt). La positivité des v.a. X_1 et X_2 entraîne que la transformée de Laplace est au moins définie sur le convexe $[0, \infty]^2$. La transformée de Laplace d'une mesure de probabilité (ou plus généralement d'une mesure de $\mathcal{M}(\mathbb{R}^2)$) étant une fonction log-convexe et analytique réelle sur son domaine de définition selon la proposition A.2.5, son hessien doit donc être semi-défini positif sur $]0, \infty[^2$. Le calcul montre (voir annexe C.1.1) que le paramètre p_{12} doit donc vérifier la condition nécessaire :

$$0 \leq p_{12} \leq 2p_1 p_2. \tag{1.18}$$

De plus si p_{12} est nul, la transformée de Laplace de la loi se réduit à

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2)^{-q},$$

qui est la transformée de Laplace du couple $(p_1 Y, p_2 Y)$ lorsque Y suit la loi univariée Gamma standard de paramètre de forme q notée $\mathcal{G}(q, 1)$. Dans ce cas la loi est dégénérée puisqu'elle est concentrée sur l'hyperplan de \mathbb{R}^2 (i.e. une droite) orthogonale au vecteur $(-p_2, p_1)^T$. Les deux marges du couple $X_1 = p_1 Y$ et $X_2 = p_2 Y$ sont alors totalement corrélées : $r(X_1, X_2) = 1$ car $p_1, p_2 > 0$.

Il est possible de déterminer exactement les conditions nécessaires et suffisantes sur les paramètres pour lesquelles la loi Gamma bivariée existe. D'après les remarques précédentes, il a déjà

été prouvé que des conditions nécessaires sont $p_1, p_2 > 0$ et $0 \leq \frac{p_{12}}{p_1 p_2} \leq 2$, la loi étant dégénérée si $p_{12} = 0$. La démonstration du théorème suivant sera conclue suite au calcul de la densité associée à une loi Gamma bivariée, p. 21.

Théorème 1.2.1. *Les conditions nécessaires et suffisantes pour que la loi Gamma bivariée du couple (X_1, X_2) , dont la transformée de Laplace est*

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = \mathbb{E} \left[e^{-z_1 X_1 - z_2 X_2} \right] = \frac{1}{(1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^q},$$

existe et ne soit pas dégénérée sont

1. $q > 0$,
2. $p_1 > 0, p_2 > 0$,
3. $0 < p_{12} \leq p_1 p_2$.

Si $p_{12} = 0$ la loi est dégénérée, les marges étant totalement corrélées : $r(X_1, X_2) = 1$.

Une conséquence immédiate de ce théorème est que les lois Gamma bivariées sont infiniment divisibles, puisque les conditions d'existence portant sur les paramètres p_1, p_2, p_{12} ne dépendent pas de q . Pour un triplet (p_1, p_2, p_{12}) vérifiant ces conditions, la loi de transformée de Laplace (1.16) existe quel que soit le paramètre $q > 0$, et est une loi Gamma bivariée.

Moments joints

Les moments joints de la loi Gamma bivariée peuvent être obtenus à partir de sa transformée de Laplace selon la proposition A.2.2. En développant en série entière la transformée de Laplace, on obtient alors par unicité du développement en série entière (voir annexe C.1.2) la formule suivante :

$$\mathbb{E} \left[X_1^{n_1} X_2^{n_2} \right] = p_1^{n_1} p_2^{n_2} (q)_{n_1} (q)_{n_2} \sum_{k=0}^{\min(n_1, n_2)} \frac{(-n_1)_k (-n_2)_k}{(q)_k} \frac{r^k}{k!}, \quad \forall (n_1, n_2) \in \mathbb{N}^2, \quad (1.19)$$

où

$$r = \frac{p_1 p_2 - p_{12}}{p_1 p_2}, \quad (1.20)$$

et où $(a)_n$ est le symbole de Pochhammer [AS64, p. 256] défini si a est un réel strictement positif comme :

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} \quad \forall n \in \mathbb{N}, \quad (1.21)$$

ou plus généralement pour tout réel a comme :

$$(a)_0 = 1, \quad (a)_{n+1} = (a+n)(a)_n \quad \forall n \in \mathbb{N}. \quad (1.22)$$

On retrouve notamment les expressions (1.2) des moments d'une loi Gamma univariée :

$$\mathbb{E} \left[X_1^{n_1} \right] = \frac{\Gamma(q+n_1)}{\Gamma(q)} p_1^{n_1}, \quad \mathbb{E} \left[X_2^{n_2} \right] = \frac{\Gamma(q+n_2)}{\Gamma(q)} p_2^{n_2},$$

On obtient également les expressions suivantes du moment joint d'ordre $(1, 1)$, de la covariance et du coefficient de corrélation d'une loi Gamma bivariée :

$$\begin{aligned} \mathbf{E}[X_1 X_2] &= q^2 p_1 p_2 \left(1 + \frac{r}{q}\right) = q((q+1)p_1 p_2 - p_{12}), \\ \text{cov}(X_1, X_2) &= q(p_1 p_2 - p_{12}), \\ r(X_1, X_2) &\equiv \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1) \text{var}(X_2)}} = \frac{p_1 p_2 - p_{12}}{p_1 p_2} \equiv r. \end{aligned} \tag{1.23}$$

Le paramètre r introduit à l'équation (1.20) et intervenant dans l'expression des moments joints (1.23) est donc le coefficient de corrélation entre les marges. Ce paramètre est lié au paramètre p_{12} et l'on obtient

$$p_{12} = p_1 p_2 (1 - r). \tag{1.24}$$

Par conséquent, les contraintes sur les paramètres énoncées au théorème 1.2.1 pour que la loi Gamma bivariée existe imposent :

$$0 \leq r < 1, \tag{1.25}$$

Les lois Gamma bivariées sont donc à dépendance positive puisque $r \geq 0$ et toutes les valeurs du coefficient de corrélation dans l'intervalle $[0, 1]$ sont possibles. Concernant les valeurs limites du coefficient de corrélation, on obtient que :

- la limite $r = 1$ est atteinte pour le cas dégénéré où $p_{12} = 0$, la loi bivariée étant alors concentrée sur la droite $X_2 = \frac{p_2}{p_1} X_1$ de \mathbb{R}^2 comme cela a été expliqué un peu plus tôt dans cette section.
- le cas où la corrélation est nulle correspond à $p_{12} = p_1 p_2$. Dans ce cas la transformée de Laplace se factorise :

$$\begin{aligned} \mathcal{L}_{(X_1, X_2)}(z_1, z_2) &= \left(\frac{1}{1 + p_1 z_1 + p_2 z_2 + p_1 p_2 z_1 z_2} \right)^q, \\ &= \left(\frac{1}{1 + p_1 z_1} \right)^q \left(\frac{1}{1 + p_2 z_2} \right)^q = \mathcal{L}_{X_1}(z_1) \mathcal{L}_{X_2}(z_2). \end{aligned} \tag{1.26}$$

La transformée de Laplace du couple se décompose comme le produit des transformées de Laplace des marges, donc les marges sont indépendantes entre elles. Ceci montre que comme dans le cas d'un vecteur gaussien, *l'absence de corrélation est équivalente à l'indépendance pour les lois Gamma bivariées.*

On remarque que les conditions nécessaires (1.18) obtenues en considérant la log-convexité de la transformée de Laplace reviennent à dire que $-1 \leq r \leq 1$. Comme cela a déjà été signalé, ces conditions ne sont cependant pas suffisantes.

Les paramètres p_1 et p_2 sont des paramètres d'échelle de la loi Gamma bivariée : si (X_1, X_2) suit une loi Gamma bivariée telle que $X_1 \sim \mathcal{G}(q, p_1)$ et $X_2 \sim \mathcal{G}(q, p_2)$ et telle que $r(X_1, X_2) = r$, alors $(Y_1, Y_2) = \left(\frac{X_1}{p_1}, \frac{X_2}{p_2}\right)$ suit la loi Gamma bivariée telle que $Y_1 \sim \mathcal{G}(q, 1)$, $Y_2 \sim \mathcal{G}(q, 1)$ et $r(Y_1, Y_2) = r$. Bien que le paramètre r , et donc le paramètre p_{12} , ne soit pas un paramètre d'échelle et que sa valeur influe aussi sur la forme de la distribution, le paramètre q est appelé le *paramètre de forme* de la distribution par analogie avec le cas univarié. En définitive, la définition donné à l'équation (1.16) de la loi Gamma bivariée et les relations entre les paramètres (p_1, p_2, p_{12}) et

les moments montrent que cette loi est caractérisée par 4 types de paramètres : Le paramètre de forme q , les paramètres d'échelles (p_1, p_2) ou bien les moyennes $(m_1 = qp_1, m_2 = qp_2)$, et le paramètre de corrélation p_{12} ou bien r .

Une propriété importante de ces lois bivariées est qu'elles appartiennent à des familles exponentielles naturelles.

La FEN $F(\mu_{c,q})$ des lois Gamma Bivariées

Soient $c \geq 0$ et $q > 0$ des paramètres réels positifs fixés. On considère alors la mesure $\mu_{c,q}$ de $\mathcal{M}(\mathbb{R}^2)$ dont la transformée de Laplace a pour expression :

$$\mathcal{L}^{\mu_{c,q}}(\theta_1, \theta_2) = (\theta_1\theta_2 - c)^{-q}, \quad (1.27)$$

définie en particulier sur l'ouvert

$$\Theta_{\mu_{c,q}} = \{(\theta_1, \theta_2) \mid \theta_1 > 0, \theta_2 > 0, \theta_1\theta_2 - c > 0\}.$$

L'ensemble $\Theta_{\mu_{c,q}}$ correspond à l'intérieur d'une branche d'hyperbole sur le premier quadrant. Sur cet ensemble l'expression (1.27) de la transformée de Laplace est bien définie, positive (car la transformée de Laplace d'une mesure positive est positive) et log-convexe (voir la proposition A.2.5).

Selon la proposition B.1.2, la FEN $F(\mu_{c,q})$ engendrée par la mesure $\mu_{c,q}$ est l'ensemble des mesures de probabilité $\nu_{c,q}$ dont la transformée de Laplace, pour tout (θ_1, θ_2) dans l'ouvert $\Theta_{\mu_{c,q}}$ et pour tout $(z_1 + \theta_1, z_2 + \theta_2) \in \Theta_{\mu_{c,q}}$, est de la forme :

$$\begin{aligned} \mathcal{L}^{\nu_{c,q}}(z_1, z_2) &= \frac{\mathcal{L}^{\mu_{c,q}}(\theta_1 + z_1, \theta_2 + z_2)}{\mathcal{L}^{\mu_{c,q}}(\theta_1, \theta_2)}, \\ &= \left(\frac{\theta_1\theta_2 - c}{(\theta_1 + z_1)(\theta_2 + z_2) - c} \right)^q, \\ &= \left(1 + \frac{\theta_2}{\theta_1\theta_2 - c}z_1 + \frac{\theta_1}{\theta_1\theta_2 - c}z_2 + \frac{1}{\theta_1\theta_2 - c}z_1z_2 \right)^{-q}. \end{aligned} \quad (1.28)$$

En identifiant les paramètres de la transformée de Laplace obtenue avec ceux de la transformée de Laplace de la loi Gamma bivariée exprimée à l'équation (1.16), on obtient le système suivant pour tout $(\theta_1, \theta_2) \in \Theta_{\mu_{c,q}}$:

$$\begin{cases} p_1 &= \frac{\theta_2}{\theta_1\theta_2 - c}, \\ p_2 &= \frac{\theta_1}{\theta_1\theta_2 - c}, \\ p_{12} &= \frac{1}{\theta_1\theta_2 - c}. \end{cases} \quad (1.29)$$

Par conséquent, toutes les lois Gamma bivariées vérifiant les conditions énoncées au théorème 1.2.1 sont obtenues. En inversant maintenant le système (1.29) précédent, on obtient pour tout $p_1 > 0, p_2 > 0$ et $0 < p_{12} \leq p_1p_2$:

$$\begin{cases} \theta_1 &= \frac{p_2}{p_{12}}, \\ \theta_2 &= \frac{p_1}{p_{12}}, \\ c &= \frac{p_1p_2 - p_{12}}{p_{12}^2}, \end{cases} \quad (1.30)$$

Donc toute loi Gamma bivariée s'exprime sous la forme d'une mesure de probabilité appartenant à une famille exponentielle naturelle. Cette propriété sera notamment utilisée dans les chapitres suivants pour l'estimation des paramètres de ces lois.

Densité de la loi Gamma Bivariée

Toutes les lois Gamma bivariées non dégénérées appartiennent à une FEN $F(\mu_{c,q})$ donnée selon la section précédente. Il suffit donc d'obtenir une expression de la mesure $\mu_{c,q}$ pour tout $q, c > 0$ pour obtenir la densité d'une loi Gamma bivariée (voir section B.1)³.

La mesure $\mu_{c,q}$ est caractérisée par sa transformée de Laplace donnée à l'équation (1.27). Pour tout $(\theta_1, \theta_2) \in \Theta_{\mu_{c,q}}$, on obtient donc

$$\mathcal{L}^{\mu_{c,q}}(\theta_1, \theta_2) = \left(\frac{1}{\theta_1 \theta_2} \right)^q \left(\frac{1}{1 - \frac{c}{\theta_1 \theta_2}} \right)^q,$$

avec $0 < \frac{c}{\theta_1 \theta_2} < 1$. En utilisant le développement en série entière suivant,

$$\left(\frac{1}{1-z} \right)^p = \sum_{k=0}^{\infty} \frac{p(p+1)\dots(p+k-1)}{k!} z^k = \sum_{k=0}^{\infty} \frac{(p)_k}{k!} z^k \text{ pour tout } |z| < 1 \quad (1.31)$$

il vient

$$\mathcal{L}^{\mu_{c,q}}(\theta_1, \theta_2) = \sum_{k=0}^{\infty} \frac{(q)_k}{k!} \frac{c^k}{\theta_1^{q+k} \theta_2^{q+k}},$$

À l'aide de l'expression de la mesure de probabilité d'une loi Gamma univariée (1.1) et du fait que l'image de \mathbb{R} par cette mesure vaut 1 :

$$\frac{1}{\theta_i^p} = \int_0^{\infty} e^{-\theta_i x_i} \frac{x_i^{p-1}}{\Gamma(p)} dx_i,$$

on obtient

$$\begin{aligned} \mathcal{L}^{\mu_{c,q}}(\theta_1, \theta_2) &= \sum_{k=0}^{\infty} \frac{(q)_k c^k}{k!} \int_0^{\infty} e^{-\theta_1 x_1} \frac{x_1^{q+k-1}}{\Gamma(q+k)} dx_1 \int_0^{\infty} e^{-\theta_2 x_2} \frac{x_2^{q+k-1}}{\Gamma(q+k)} dx_2, \\ &= \int_0^{\infty} \int_0^{\infty} e^{-\theta_1 x_1} e^{-\theta_2 x_2} \sum_{k=0}^{\infty} \frac{(q)_k c^k}{k!} \frac{x_1^{q+k-1}}{\Gamma(q+k)} \frac{x_2^{q+k-1}}{\Gamma(q+k)} dx_1 dx_2, \end{aligned}$$

pour tout (θ_1, θ_2) dans l'ensemble ouvert non vide $\Theta_{\mu_{c,q}}$. On reconnaît l'expression d'une transformée de Laplace sur \mathbb{R}^2 et la proposition A.2.1 assure, par unicité de la transformée de Laplace, que

$$\begin{aligned} \mu_{c,q}(dx_1, dx_2) &= \sum_{k=0}^{\infty} \frac{(q)_k c^k}{k!} \frac{x_1^{q+k-1}}{\Gamma(q+k)} \frac{x_2^{q+k-1}}{\Gamma(q+k)} \mathbb{I}_{[0,\infty]^2}(x_1, x_2) dx_1 dx_2, \\ &= \frac{x_1^{q-1} x_2^{q-1}}{\Gamma(q)} \sum_{k=0}^{\infty} \frac{(cx_1 x_2)^k}{\Gamma(q+k) k!} \mathbb{I}_{[0,\infty]^2}(x_1, x_2) dx_1 dx_2, \\ &= \frac{x_1^{q-1} x_2^{q-1}}{\Gamma(q)} f_q(cx_1 x_2) \mathbb{I}_{[0,\infty]^2}(x_1, x_2) dx_1 dx_2, \end{aligned}$$

³On pourrait bien sûr inverser directement la transformée de Laplace d'une loi Gamma bivariée (1.16), mais l'expression (1.27) de la transformée de Laplace de la mesure $\mu_{c,q}$ est plus simple et les calculs sont plus aisés dans ce cas.

avec

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(q+k)k!} \quad \forall z \in \mathbb{C}. \quad (1.32)$$

La fonction f_q est une fonction hypergéométrique confluyente limite et est liée à fonction de Bessel modifiée I_ν [AS64, p. 374] de la manière suivante :

$$\left(\frac{z}{2}\right)^{q-1} f_q\left(\frac{z}{2}\right) = I_{q-1}(z) \quad \forall z \in \mathbb{C}. \quad (1.33)$$

Finalement on obtient les expressions suivantes de la densité de la loi Gamma bivariée (en faisant attention au fait que la convention prise pour la transformée de Laplace est ici $\varepsilon = -1$) pour les différentes paramétrisations équivalentes et pour un paramètre de forme $q > 0$ donné :

- La paramétrisation canonique de la FEN $F(\mu_{c,q})$

$$p(x_1, x_2) = (\theta_1 \theta_2 - c)^q e^{-\theta_1 x_1 - \theta_2 x_2} \frac{x_1^{q-1} x_2^{q-1}}{\Gamma(q)} f_q(cx_1 x_2) \mathbb{I}_{[0, \infty]^2}(x_1, x_2), \quad (1.34)$$

définie pour tout $c \geq 0$ et $(\theta_1, \theta_2) \in \Theta_{\mu_{c,q}}$.

- La paramétrisation par les paramètres de la transformée de Laplace p_1 , p_2 et p_{12} , obtenue grâce aux relations (1.30)

$$p(x_1, x_2) = \frac{1}{p_{12}^q} \exp\left[-\left(\frac{p_2}{p_{12}}x_1 + \frac{p_1}{p_{12}}x_2\right)\right] \frac{x_1^{q-1} x_2^{q-1}}{\Gamma(q)} f_q(cx_1 x_2) \mathbb{I}_{[0, \infty]^2}(x_1, x_2), \quad (1.35)$$

où

$$c = \frac{p_1 p_2 - p_{12}}{p_{12}^2}, \quad (1.36)$$

définie pour tout $p_1 > 0$, $p_2 > 0$ et $0 < p_{12} \leq p_1 p_2$.

- La paramétrisation par les moyennes $m_1 = qp_1$ et $m_2 = qp_2$ et par le coefficient de corrélation r exprimé à l'équation (1.23)

$$p(x_1, x_2) = \left(\frac{q^2}{m_1 m_2 (1-r)}\right)^q e^{-\left(\frac{q}{m_1(1-r)}x_1 + \frac{q}{m_2(1-r)}x_2\right)} \frac{x_1^{q-1} x_2^{q-1}}{\Gamma(q)} f_q(cx_1 x_2) \mathbb{I}_{[0, \infty]^2}(x_1, x_2), \quad (1.37)$$

où

$$c = \frac{r}{(1-r)^2} \frac{q^2}{m_1 m_2}, \quad (1.38)$$

définie pour tout $m_1 > 0$, $m_2 > 0$ et $0 \leq r < 1$.

L'expression (1.35) de la densité d'une loi Gamma bivariée explique pourquoi le paramètre p_{12} doit respecter la contrainte $p_{12} \leq p_1 p_2$. En effet l'expression (1.35) obtenue en inversant la transformée de Laplace reste valide pour tout $p_{12} \in]0, 2p_1 p_2[$, mais si $p_{12} > p_1 p_2$ il découle de

la formule (1.36) que le paramètre c est strictement négatif. Or la fonction de Bessel modifiée admet tous ses zéros (une infinité) sur l'axe imaginaire. Selon la relation (1.33) entre la fonction f_q et la fonction de Bessel modifiée, la fonction f_q qui est continue sur \mathbb{R} admet une infinité de zéros sur $] -\infty, 0[$ et n'est donc pas positive sur cet ensemble. Par conséquent si c est négatif, la fonction définie à l'équation (1.35) n'est pas une densité car elle n'est pas positive.

Réciproquement si les conditions du théorème 1.2.1 sont vérifiées alors le paramètre c est positif et par conséquent le terme $f_q(cxy)$ est positif en tant que somme de termes positifs. La fonction $(x_1, x_2) \in \mathbb{R}^2 \mapsto p(x_1, x_2)$ définie en (1.35) est alors positive en tant que produit de fonctions positives. Par construction, cette fonction admet l'expression (1.16) comme transformée de Laplace, laquelle vaut 1 au point $(0, 0)$. L'intégrale sur \mathbb{R}^2 de la fonction est donc égale à 1, aussi cette fonction est elle bien une densité et la loi existe. Ceci achève la démonstration du théorème 1.2.1.

Les graphes de la densité de la loi Gamma bivariée sont représenté sur la figure 1.2 pour différentes valeurs du paramètre de forme $q \in \{\frac{1}{2}, 1, 2\}$ et différentes valeurs du coefficient de corrélation $r \in \{0.2, 0.5, 0.8\}$, les moyennes des marges étant $m_1 = qp_1 = m_2 = qp_2 = 1$.

Une autre expression de cette densité est donnée à la section suivante.

Lois Gamma bivariées, polynômes de Laguerre et probabilité de Lancaster

Les probabilités de Lancaster introduites dans [Lan58, Lan63] sont des classes de probabilité sur \mathbb{R}^2 construites à partir des lois marginales et de suites de polynômes orthonormaux pour ces lois.

Considérons deux mesures de probabilité sur \mathbb{R} , $\alpha(dx)$ et $\beta(dy)$, non concentrées sur un nombre fini de points et pour lesquelles il existe $a > 0$ tel que $\int e^{-ax} \alpha(dx) < \infty$ et $\int e^{-ay} \beta(dy) < \infty$. On note alors $(P_n)_{n \geq 0}$ et $(Q_n)_{n \geq 0}$ les suites de polynômes à degrés étagés, n étant le degré des polynômes P_n et Q_n , orthonormales pour les produits scalaires respectifs :

$$\begin{aligned} \langle P, Q \rangle_\alpha &= \int_{\mathbb{R}} P(x)Q(x)\alpha(dx), \\ \langle P, Q \rangle_\beta &= \int_{\mathbb{R}} P(x)Q(x)\beta(dx). \end{aligned}$$

Une condition suffisante pour qu'une loi $\mu(dx, dy)$ soit de Lancaster est qu'il existe une suite de réels $(\rho_n)_{n \geq 0}$ telle que la loi $\mu(dx, dy)$ est absolument continue par rapport à $\alpha(dx)\beta(dy)$ et égale à

$$\mu(dx, dy) = \left[\sum_{n=0}^{\infty} \rho_n P_n(x)Q_n(y) \right] \alpha(dx)\beta(dy). \quad (1.39)$$

De nombreuses propriétés [Kou98] découlent de cette représentation. En particulier si $\mu(dx, dy)$ et $\alpha(dx)\beta(dy)$ sont absolument continues par rapport à la mesure de Lebesgue, la divergence du χ^2 entre $\mu(dx, dy)$ et $\alpha(dx)\beta(dy)$ est définie comme :

$$\begin{aligned} I_{\chi^2}(\mu, \alpha\beta) &= \int_{\mathbb{R}^2} \frac{(\mu(x, y) - \alpha(x)\beta(y))^2}{\alpha(x)\beta(y)} dx dy, \\ &= \int_{\mathbb{R}^2} \frac{\mu(x, y)^2}{\alpha(x)\beta(y)} dx dy - 1. \end{aligned} \quad (1.40)$$

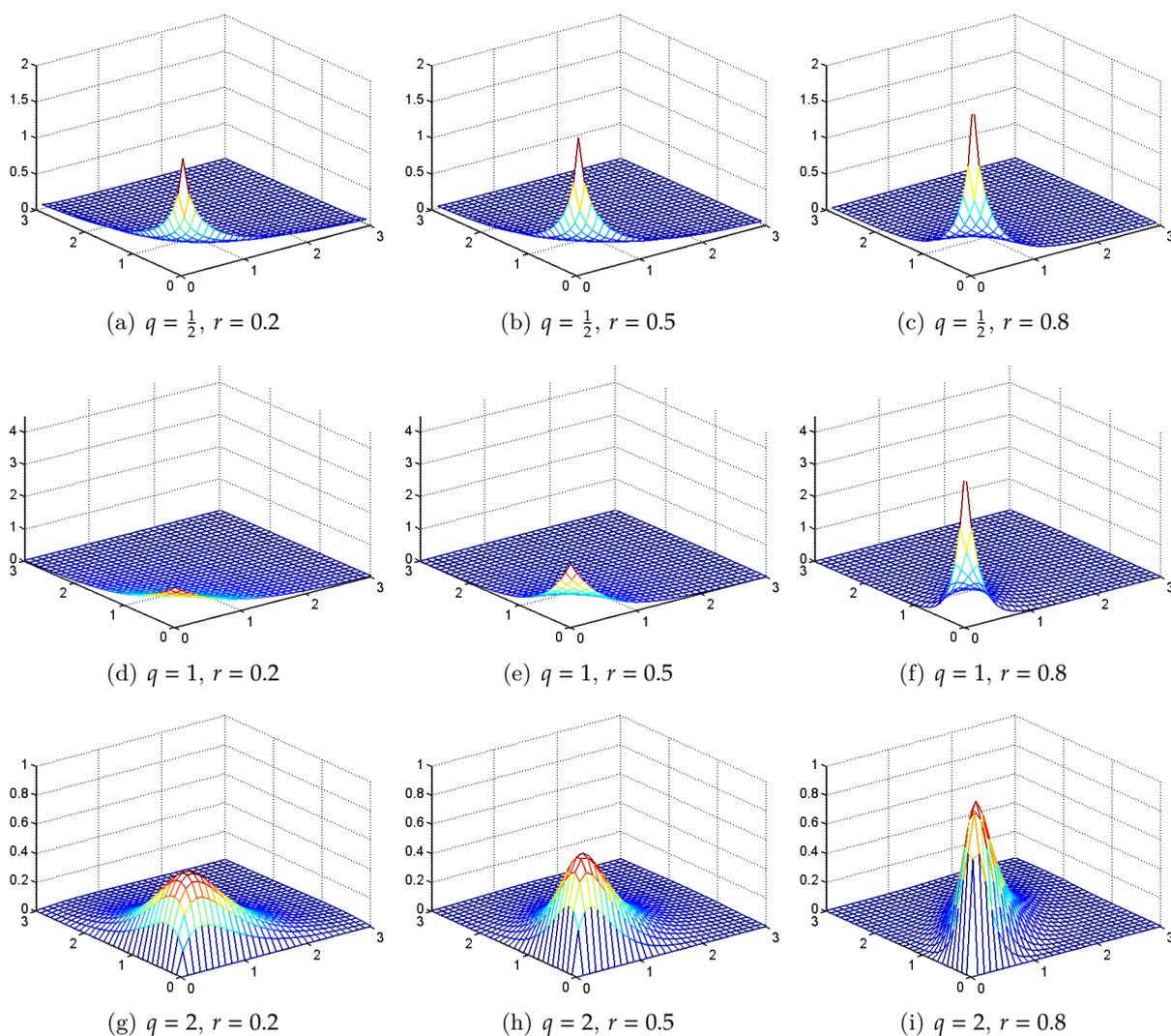


FIG. 1.2 – Densité de la loi Gamma bivariée de moyennes $m_1 = qp_1 = m_2 = qp_2 = 1$ pour différentes valeurs du paramètre de forme q et du coefficient de corrélation entre les marges r .

Cette divergence entre une mesure bivariée et le produit des mesures marginales est encore appelée *la mesure d'indépendance du χ^2* ⁴ car c'est une grandeur positive qui s'annule si et seulement si les marges sont indépendantes.

Si $\mu(dx, dy)$ est une probabilité de Lancaster, alors un calcul direct montre que :

$$I_{\chi^2}(\mu, \alpha\beta) = -1 + \sum_{n \geq 0} \rho_n^2. \quad (1.41)$$

Les polynômes généralisés de Laguerre [AS64, p. 771–802] sont des familles de polynômes

⁴Cette divergence est qualifiée de divergence du χ^2 car dans le cas discret on retrouve la statistique de test d'adéquation à une loi proposée par Pearson. Sous certaines hypothèses (la loi théorique est une loi multinomiale) cette statistique suit une loi χ^2 .

$(L_n^{(\nu)})_{n \geq 0}$, n étant le degré du polynôme, orthogonaux sur \mathbb{R}^+ pour le produit scalaire

$$\langle P, Q \rangle_{w_\nu} = \int_{\mathbb{R}^+} w_\nu(x) P(x) Q(x) dx, \quad (1.42)$$

où w_ν est la fonction poids strictement positive sur $]0, \infty[$

$$w_\nu(x) = e^{-x} x^\nu \mathbb{I}_{]0, \infty[}(x) \quad \forall \nu > -1,$$

et $\nu > -1$.

Selon cette caractérisation, les polynômes $(L_n^{(\nu)})_{n \geq 0}$ sont tous définis à une constante multiplicative près. L'expression standard de ces polynômes est

$$\begin{aligned} L_n^{(\nu)}(x) &= \frac{e^x x^{-\nu}}{n!} \frac{d^n}{dx^n} (e^{-x} x^{\nu+n}), \\ &= \sum_{i=0}^n \binom{\nu+n}{n-i} \frac{(-x)^i}{i!}. \end{aligned} \quad (1.43)$$

Les deux premiers polynômes sont alors :

$$\begin{aligned} L_0^{(\nu)}(x) &= 1, \\ L_1^{(\nu)}(x) &= -x + \nu + 1, \end{aligned}$$

et les suivants peuvent se calculer par la relation de récurrence :

$$L_{n+1}^{(\nu)}(x) = \frac{1}{n+1} \left((2n+1+\nu-x)L_n^{(\nu)}(x) - (n+\nu)L_{n-1}^{(\nu)}(x) \right).$$

De plus, pour tout $\nu > 0$

$$\left\langle L_n^{(\nu-1)}, L_m^{(\nu-1)} \right\rangle_{w_{\nu-1}} = \frac{\Gamma(\nu+n)}{n!} \delta_{m,n} \quad \forall (m, n) \in \mathbb{N}^2, \quad (1.44)$$

où $\delta_{m,n}$ est le symbole de Kronecker qui vaut 1 si $m = n$, 0 sinon.

Afin de déterminer la densité d'un couple (X_1, X_2) distribué suivant une loi Gamma bivariée, on peut supposer sans perte de généralité que les paramètres d'échelle de cette loi sont $p_1 = p_2 = 1$. En effet la densité dans le cas général $p_1, p_2 > 0$ s'obtient à partir de la précédente par le changement de variable $(Y_1, Y_2) = (p_1 X_1, p_2 X_2)$. On suppose donc que le couple (X_1, X_2) est distribué suivant une loi Gamma bivariée dont les marges suivent les lois $\mathcal{G}(q, 1)$ et admettent pour coefficient de corrélation le paramètre r . Le calcul présenté à l'annexe C.1.3 montre alors que la densité de loi Gamma bivariée associée au couple (X_1, X_2) a pour expression :

$$p_{(X_1, X_2)}(x_1, x_2) = \mathcal{G}_{q,1}(x_1) \mathcal{G}_{q,1}(x_2) \sum_{k=0}^{\infty} r^k P_k^{(q)}(x_1) P_k^{(q)}(x_2), \quad (1.45)$$

où la fonction $x \mapsto \mathcal{G}_{q,1}(x)$ correspond à la densité de la loi Gamma univariée $\mathcal{G}(q, 1)$ et où

$$P_k^{(q)}(x) = \left(\frac{k! \Gamma(q)}{\Gamma(q+k)} \right)^{1/2} L_k^{(q-1)}(x) = \left(\frac{k!}{(q)_k} \right)^{1/2} L_k^{(q-1)}(x) \quad (1.46)$$

La propriété d'orthogonalité (1.42) des polynômes de Laguerre implique alors que la famille de polynômes $(P_n^{(q)})_{n \geq 0}$ est orthonormale pour le produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{G}_{q,1}}$. Par conséquent la loi Gamma bivariée de transformée de Laplace (1.16) est de la forme définie à l'équation (1.39). Cette famille de lois est donc une famille de probabilité de Lancaster pour les lois marginales \mathcal{G}_{q,p_1} , la famille de polynôme orthogonaux $(P_n^{(q)})_{n \geq 0}$ et la suite $\rho_n = r^n, \forall n \in \mathbb{N}$. Ce résultat est connu et a notamment été montré par D'jachenko [D'j62]. Une conséquence immédiate est que la mesure d'indépendance du χ^2 pour le couple (X_1, X_2) , i.e. la divergence du χ^2 entre les lois bivariées $p_{(X_1, X_2)}$ et $p_{X_1} p_{X_2}$, a pour expression :

$$I_{\chi^2}(X_1, X_2) = -1 + \sum_{k \geq 0} (r^k)^2 = \frac{r^2}{1 - r^2} \quad \forall 0 \leq r < 1.$$

Cette mesure de similarité ne dépend pas par construction des paramètres d'échelle puisque $I_{\chi^2}(X_1, X_2) = I_{\chi^2}(X_1/p_1, X_2/p_2)$ comme cela sera montré à la section 2.1.2. De manière plus surprenante, cette divergence ne dépend pas non plus du paramètre de forme q . Elle est cependant une fonction croissante du coefficient de corrélation r . L'expression de cette divergence illustre également le fait qu'absence de corrélation et indépendance sont des propriétés équivalentes pour la loi Gamma bivariée puisque si $r = 0$, alors $I_{\chi^2} = 0$.

Bien que l'expression (1.45) soit absolument convergente pour tout $0 \leq r < 1$, il est difficile de calculer numériquement cette série lorsque r est proche de 1 car les termes ne sont pas de signes constants. Il est possible à partir de l'équation (1.45) de retrouver l'expression (1.30) de la densité grâce à la formule de Hille-Hardy [AET81, p. 189]. Les familles de lois de Lancaster sont par ailleurs des cas particuliers des familles dites de Sarmanov qui sont construites par extension du modèle de Farlie-Gumbel-Morgenstern (voir [Sar66, SL04] pour une caractérisation plus précise de ces familles) présenté à la section 1.2.1. La loi Gamma bivariée a ainsi été proposée sous la forme (1.45) par Sarmanov.

Simulation de la loi Gamma bivariée

Toutes les lois Gamma bivariées de transformée de Laplace (1.16) et dont les paramètres vérifient les conditions énoncées au théorème 1.2.1 peuvent être simulées. Une première remarque, est que la densité de ces lois est connue. Par conséquent des méthodes de type acceptation-rejet peuvent être employées pour simuler la loi bivariée. Cependant ces méthodes ne sont pas toujours très efficaces notamment si la proportion de rejet est importante. Il est donc en général préférable d'utiliser une méthode directe de simulation. La méthode de simulation ici étudiée a été proposée dans [Gav70] et repose sur la composition de lois négatives binomiales par des lois Gamma. Ces compositions ont déjà été étudiées dans le cas univarié à la section 1.1.2 à travers la proposition (1.1.1).

On peut ici supposer sans perte de généralité que la transformée de Laplace du couple (X_1, X_2) à simuler est de la forme :

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = \left(\frac{1}{1 + \frac{z_1}{1-r} + \frac{z_2}{1-r} + \frac{z_1 z_2}{1-r}} \right)^q, \quad (1.47)$$

avec $0 \leq r < 1$. En effet si (X_1, X_2) admet la transformée de Laplace ci-dessus alors celle du vecteur $(p_1(1-r)X_1, p_2(1-r)X_2)$ correspond bien à l'expression générale (1.16). Ici le couple (X_1, X_2) à simuler est tel que ses marges suivent la loi Gamma $\mathcal{G}(q, \frac{1}{1-r})$ et ont pour coefficient

de corrélation le paramètre r d'après la formule (1.23). La loi du couple (X_1, X_2) est qualifiée de *loi Gamma bivariée standard* de paramètre de corrélation r pour des raisons liés à l'estimation des paramètres et qui seront expliquées dans le chapitre suivant.

Proposition 1.2.2. *Soit N une v.a. de loi $\mathcal{BN}(q, 1 - r)$ avec $0 \leq r < 1$. Soient X_1 et X_2 deux v.a. obtenues par composition de la loi négative binomiale par les lois Gamma suivante $X_1 \sim \mathcal{G}(q + N, 1)$ et $X_2 \sim \mathcal{G}(q + N, 1)$, les v.a. $X_1|N$ et $X_2|N$ étant indépendantes. Alors le couple (X_1, X_2) est distribué suivant une loi Gamma de paramètre de forme q et dont la transformée de Laplace correspond à celle définie à l'équation (1.47).*

Preuve. La démonstration est similaire à celle faite pour le cas univarié. En effet pour tout $z_1, z_2 \geq 0$, la transformée de Laplace du couple (X_1, X_2) est obtenue en conditionnant l'espérance par rapport à la v.a. N :

$$\begin{aligned} \mathcal{L}_{(X_1, X_2)} &= \mathbb{E} \left[e^{-z_1 X_1 - z_2 X_2} \right] = \mathbb{E} \left[\mathbb{E} \left[e^{-z_1 X_1 - z_2 X_2} \mid N \right] \right] = \mathbb{E} \left[\mathbb{E} \left[e^{-z_1 X_1} \mid N \right] \mathbb{E} \left[e^{-z_2 X_2} \mid N \right] \right], \\ &= \mathbb{E} \left[\left(\frac{1}{1 + z_1} \right)^{q+N} \left(\frac{1}{1 + z_2} \right)^{q+N} \right] = \left(\frac{1}{(1 + z_1)(1 + z_2)} \right)^q G_N \left(\frac{1}{(1 + z_1)(1 + z_2)} \right), \\ &= \left(\frac{1}{(1 + z_1)(1 + z_2)} \right)^q \left(\frac{1 - r}{1 - r \frac{1}{(1 + z_1)(1 + z_2)}} \right)^q = \left(\frac{1}{1 + \frac{z_1}{1-r} + \frac{z_2}{1-r} + \frac{z_1 z_2}{1-r}} \right)^q. \end{aligned}$$

□

Toutes les lois Gamma bivariées non dégénérées peuvent être obtenues selon la proposition précédente si l'on sait simuler des v.a. Gamma et binomiales négatives de paramètres quelconques. Le cas dégénéré $p_{12} = 0 \iff r = 1$ est par ailleurs trivial, car dans ce cas le couple $(X_1 = p_1 Y, X_2 = p_2 Y)$ où $Y \sim \mathcal{G}(q, 1)$ a bien l'expression (1.16) pour transformée de Laplace. La simulation des lois Gamma univariées ne pose pas problème, comme cela est expliqué à la section précédente consacrée à la loi Gamma univariée. La simulation de lois binomiales négatives dont le paramètre de forme n'est pas entier est par contre moins standard⁵. La loi négative binomiale peut être obtenue dans ce cas par mélange d'une loi de Poisson par une loi Gamma d'après la proposition 1.1.2. Cette méthode amène à l'algorithme 1 de simulation d'une loi Gamma bivariée de transformée de Laplace (1.16).

D'autres méthodes peuvent être utilisées pour simuler directement des lois Gamma bivariées. Dans la section 1.3.2 consacrées aux lois Gamma multivariées une méthode de simulation obtenue à partir de vecteurs Gaussiens est présentée. Cependant cette méthode ne permet de simuler que les lois Gamma multivariées (ou bivariées) dont le paramètre de forme est un demi-entier : $q = \frac{n}{2}$ avec $n \in \mathbb{N}/\{0\}$.

1.3 Lois Multivariées

Dans cette partie, les différentes lois multivariées utilisées par la suite sont introduites. La définition de ces lois ainsi que leurs principales propriétés sont donc exposées au cours de cette section. Les lois qui nous intéressent sont, bien sûr, les lois Gamma multivariées, mais aussi les

⁵si le paramètre de forme q est entier, on peut utiliser directement la définition de la loi négative binomiale (distribution du nombre d'échecs avant le q -ième succès lors de la répétition d'épreuves i.i.d. de probabilité de succès p) pour la simuler, où l'obtenir comme la somme de q lois géométriques i.i.d. de paramètre p .

Algorithm 1 gambivrnd(q, p_1, p_2, r)

Require: procédure `gamrnd(q, p)` pour simuler la loi $\mathcal{G}(q, p)$
procédure `negbinrnd(n, p)` pour simuler la loi $\mathcal{BN}(n, p)$ avec $n \in \mathbb{N}$
procédure `poissrnd(λ)` pour simuler la loi de Poisson $\mathcal{P}(\lambda)$
 $q > 0, p_1 > 0, p_2 > 0, 0 \leq r < 1$
/ simulation de la loi $\mathcal{BN}(q, (1-r))$ */*
1: **if** q est entier **then**
2: $N = \text{negbinrnd}(q, p)$
3: **else**
4: $U = \text{gamrnd}(q, \frac{r}{1-r})$
5: $N = \text{poissrnd}(U)$
6: **end if**
/ simulation de la loi Gamma bivariée */*
7: $X_1 = \text{gamrnd}(q+N, p_1(1-r))$
8: $X_2 = \text{gamrnd}(q+N, p_2(1-r))$
9: **return** X_1, X_2

lois multinomiales négatives et les lois de Poisson mélangées multivariées qui sont obtenues par “poissonisation” d’une loi multivariée. Lorsque la loi de mélange est une loi Gamma multivariée, il est montré que les familles de lois de Poisson mélangées sont un cas particulier de lois multinomiales négatives. Bien que toutes ces lois multivariées soient définies et caractérisées de manière relativement simple, il n’existe pas en général de forme analytique explicite de leurs densités ou de leurs masses.

Cette section se décompose en quatre sous-parties. Quelques modèles célèbres de lois Gamma multivariées sont tout d’abord présentés. Le modèle que nous allons étudier est ensuite défini et étudié. Les familles de lois appelées lois multinomiales négatives sont à leur tour étudiées, le cas bivarié recevant une attention plus particulière. La quatrième sous-partie est finalement consacrée à l’étude des lois de Poisson mélangées, notamment lorsque la loi de mélange est la loi Gamma bivariée.

1.3.1 Une zoologie de familles de lois Gamma Multivariées

La plupart des modèles de lois Gamma multivariées sont directement obtenus en généralisant en dimension d quelconque les modèles existants de lois Gamma bivariées. Quelques uns de ces modèles sont présentés dans les paragraphes suivants.

Lois Gamma multivariées de Cheriyan et Ramabhadran

Le modèle bivarié a été présenté à la section 1.2.1. L’extension au cas multivarié en dimension $d \in \mathbb{N}$ est obtenue de la façon suivante : soient $Y_0, Y_1, Y_2, \dots, Y_d$ $d+1$ v.a. indépendantes distribuées suivant des lois Gamma (i.e. de paramètre d’échelle égal à 1) de paramètres de forme respectifs $\theta_0, \theta_1, \theta_2, \dots, \theta_d$. La loi Gamma multivariée est construite comme la loi du vecteur

$\mathbf{X} = (X_1, \dots, X_d)^T$ pour lequel :

$$\begin{aligned} X_1 &= Y_0 + Y_1, \\ &\vdots \\ X_d &= Y_0 + Y_d. \end{aligned} \tag{1.48}$$

Par construction, les marges X_i du vecteur ainsi défini sont des lois Gamma standard de paramètres de formes $\theta_0 + \theta_i$ pour tout $1 \leq i \leq d$. Les lois correspondantes de paramètres d'échelle réels positifs quelconques s'obtiennent par homothétie. De plus la loi de chaque couple (X_i, X_j) correspond à la loi bivariée de Cheriyan et Ramabhadran de paramètres $(\theta_0, \theta_i, \theta_j)$ pour tout $1 \leq i < j \leq d$. Par conséquent le coefficient de corrélation entre les marges vaut :

$$r(X_i, X_j) = \frac{\theta_0}{\sqrt{(\theta_0 + \theta_i)(\theta_0 + \theta_j)}} \geq 0 \quad \forall 1 \leq i < j \leq d,$$

selon la section 1.2.1. Lorsque la valeur des paramètres de forme des lois marginales est imposée, par exemple $\theta_0 + \theta_i = q_i$ pour tout $1 \leq i \leq d$ et que la corrélation entre les marges de chaque couple est elle aussi imposée, alors tous les paramètres $\theta_0, \theta_1, \dots, \theta_d$ sont fixés et la loi n'a plus de degré de liberté. La valeur des coefficients de corrélation pour les autres marges est alors donnée par la formule précédente.

À partir de la définition du modèle multivarié, donnée à l'équation (1.48), et de la transformée de Laplace d'une loi Gamma univariée (1.4), l'expression suivante de la transformée de Laplace du vecteur $\mathbf{X} = (X_1, \dots, X_d)^T$ est finalement obtenue :

$$\begin{aligned} \mathcal{L}_{\mathbf{X}}(\mathbf{z}) &= \mathbb{E} \left[e^{-\mathbf{z}^T \mathbf{X}} \right] = \mathbb{E} \left[e^{-(z_1 + \dots + z_d) Y_0} \right] \prod_{i=1}^d \mathbb{E} \left[e^{-z_i Y_i} \right] = \mathcal{L}_{Y_0}(z_1 + \dots + z_d) \prod_{i=1}^d \mathcal{L}_{Y_i}(z_i), \\ &= (1 + z_1 + \dots + z_d)^{-\theta_0} \prod_{i=1}^d (1 + z_i)^{-\theta_i}. \end{aligned}$$

Comme il n'existe pas de forme analytique simple de la densité, les paramètres de ces lois sont souvent estimés par la méthode des moments. Pour pallier ce problème, une méthode d'inférence bayésienne a été proposée par Tsonas [Tsi04] lorsque les *a priori* sont des lois Gamma. Ces modèles multivariés ont aussi été étudiés par Mathai et Moschopoulos [MM91] dans un cadre un peu plus général, où les vecteurs Y_i suivent une loi Gamma étendue à 3 paramètres, le troisième paramètre étant un paramètre de localisation.

Lois Gamma multivariées de Prékopa et Szántai

Ces lois sont une généralisation du modèle multivarié de Cheriyan et Ramabhadran présenté à la section 1.3.1. On s'intéresse dans ce modèle [PS78] à la loi du vecteur de \mathbb{R}^d \mathbf{X} obtenu de la manière suivante :

$$\mathbf{X} = \mathbf{A}\mathbf{W},$$

où \mathbf{A} est une matrice déterministe à d lignes dont les éléments sont 0 ou 1 (il existe donc $2^d - 1$ colonnes différentes possibles pour une telle matrice si l'on exclut la colonne composée uniquement de 0) et \mathbf{W} est une matrice aléatoire dont les éléments W_i sont indépendants et suivent une loi Gamma standard de paramètre θ_i . Plusieurs résultats théoriques ont alors été montrés pour ces distributions dans [PS78].

1.3.2 Le modèle des Lois Gamma Multivariées (LGM)

Le modèle des lois Gamma multivariées qui nous intéresse a été identifié par Bar-Lev *et al* [BLBE⁺94] au cours de leur classification des familles exponentielles naturelles sur \mathbb{R}^n . Par la suite on désignera par lois Gamma multivariées (LGM) les familles de lois définies dans cette section. Ces familles multivariées sont caractérisées par leur transformée de Laplace qui est une extension en dimension d quelconque de la transformée de Laplace de la loi Gamma univariée et de la loi Gamma bivariée de Kibble.

Afin de caractériser ces LGM, on définit préalablement la notion de polynôme affine.

Définition 1.3.1. *Un polynôme $P(\mathbf{z})$ est dit affine par rapport au vecteur $\mathbf{z} = (z_1, \dots, z_d)$ si la fonction polynomiale d'une seule variable $z_j \mapsto P(\mathbf{z})$ peut s'exprimer comme $A^{(-j)}z_j + B^{(-j)}$ (pour tout $j = 1, \dots, d$), où $A^{(-j)}$ et $B^{(-j)}$ ne dépendent pas de z_j et sont des polynômes par rapport aux variables z_i avec $i \neq j$.*

Il faut noter que, pour tout polynôme affine P ,

$$P_1(z_1, \dots, z_d) = P(a_1z_1 + b_1, \dots, a_dz_d + b_d),$$

est aussi un polynôme affine, quels que soient les coefficients réels a_i et b_i .

On obtient maintenant la caractérisation d'une LGM.

Définition 1.3.2. *Soit $q > 0$ et $P(\mathbf{z})$ un polynôme affine en dimension d . La loi Gamma multivariée concentrée sur \mathbb{R}_+^d de paramètres q et $P(\mathbf{z})$, notée $\mathcal{G}(q, P)$, est définie par sa transformée de Laplace :*

$$\mathcal{L}_{\mathcal{G}(q, P)}(\mathbf{z}) = [P(\mathbf{z})]^{-q} \quad \forall \mathbf{z} \in \Theta_{q, P}, \quad (1.49)$$

où $\Theta_{q, P}$ est l'intérieur du domaine d'existence, le polynôme affine satisfaisant de plus la contrainte évidente $P(\mathbf{0}) = 1$.

Remarques :

- le polynôme P est appelé paramètre, mais les paramètres sont en fait les coefficients de ce polynôme.
- par analogie avec le cas univarié on appellera q le paramètre de forme même si cette appellation est abusive car le polynôme P ne correspond pas en général à un paramètre d'échelle.

Exemple 1.3.1 : *LGM en dimension $d \in \{1, 2, 3\}$.*

- En dimension $d = 1$, le polynôme affine est de la forme :

$$P(z) = 1 + pz,$$

et l'expression (1.49) correspond donc à la transformée de Laplace de loi Gamma sur \mathbb{R} $\mathcal{G}(q, p)$ pour $p > 0$.

- En dimension $d = 2$, le polynôme affine est de la forme :

$$P(z_1, z_2) = 1 + p_1z_1 + p_2z_2 + p_{12}z_1z_2,$$

et la définition 1.3.2 de la LGM dans le cas bivarié correspond bien à la loi de Kibble définie à la section 1.2.1.

– En dimension $d = 3$, l'expression du polynôme affine devient :

$$P(z_1, z_2, z_3) = 1 + p_1 z_1 + p_2 z_2 + p_3 z_3 + p_{12} z_1 z_2 + p_{13} z_1 z_3 + p_{23} z_2 z_3 + p_{123} z_1 z_2 z_3.$$

Contrairement au cas bivarié où il est possible de déterminer facilement les conditions nécessaires et suffisantes sur les paramètres pour que la loi bivariée existe, en revanche déterminer les conditions nécessaires et suffisantes sur la paire de paramètres (q, P) pour que la distribution multivariée $\mathcal{G}(q, P)$ existe est un problème difficile. Une condition nécessaire est bien sûr $P(\mathbf{0}) = 1$ puisque la transformée de Laplace d'une mesure de probabilité vaut 1 au point $\mathbf{0}$. Le cas où ces distributions sont infiniment divisibles, i.e. le cas où la loi existe pour un polynôme affine P donné quel que soit le paramètre de forme $q > 0$, a par ailleurs été étudié par Bernardoff dans [Ber06].

Proposition 1.3.1. *Les lois marginales d'un vecteur $\mathbf{X} = (X_1, \dots, X_d)^T$ de loi $\mathcal{G}(q, P)$ sont des lois Gamma multivariées qui ont toutes le même paramètre de forme q .*

Cette propriété importante des lois Gamma multivariées traduit le fait que la loi de $\tilde{\mathbf{X}} = ((X_i)_{i \in I})^T$, où I est un sous-ensemble de $\{1, \dots, d\}$ est une loi gamma multivariée (univariée si $\text{card } I = 1$).

Démonstration. Cette propriété découle directement de la définition 1.3.2 d'une loi Gamma multivariée à partir de sa transformée de Laplace, car en notant $\tilde{\mathbf{z}} = (z_i)_{i \in I}$, on a :

$$\mathcal{L}_{\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}) = \mathbb{E} \left[e^{\sum_{i \in I} -z_i X_i} \right] = \mathbb{E} \left[e^{\sum_{i=1}^d -\mathbb{1}_I(i) z_i X_i} \right] = \mathcal{L}_X(z_1 \mathbb{1}_I(1), \dots, z_d \mathbb{1}_I(d)) = [\tilde{P}(\tilde{\mathbf{z}})]^{-q},$$

où $\mathbb{1}_I(k)$ est la fonction indicatrice sur l'ensemble I , le polynôme $\tilde{P}(\tilde{\mathbf{z}}) = P(z_1 \mathbb{1}_I(1), \dots, z_d \mathbb{1}_I(d))$ étant par construction un polynôme affine par rapport à $\tilde{\mathbf{z}}$. \square

Par conséquent les coefficients des monômes z_i, z_j et $z_i z_j$ pour $1 \leq i < j \leq d$ du polynôme affine $P(z_1, \dots, z_d)$ (i.e. les triplets notés (p_i, p_j, p_{ij}) dans les exemples 1.3.1) doivent vérifier les conditions énoncées au théorème 1.2.1. Ces conditions sont nécessaires mais ne sont pas suffisantes si $d > 2$.

Les moments joints d'une LGM peuvent se calculer à partir de sa transformée de Laplace (1.49) grâce à la proposition A.2.2.

Les FEN générés par les LGM

Le résultat suivant, qui montre que les FEN engendrées par une LGM sont des familles de LGM, est une généralisation en dimension quelconque du résultat démontré dans le cas bivarié à la section 1.2.2. On considère la LGM $\mathcal{G}(q, P)$ sur \mathbb{R}^d caractérisée par sa transformée de Laplace (1.49) et l'on note $\Theta_{q, P}$ l'intérieur du domaine de définition de la transformée de Laplace. On obtient alors la proposition suivante :

Proposition 1.3.2. *Soit $q > 0$ et P un polynôme affine tel que $P(\mathbf{0}) = 1$ tels que la loi $\mathcal{G}(q, P)$ existe et ne soit pas concentrée sur un hyperplan affine. La FEN engendrée par la loi $\mathcal{G}(q, P)$, notée $F_{\mathcal{G}(q, P)}$ est la famille des lois Gamma Multivariées de paramètre de forme q et dont les polynômes affines notés P_θ sont de la forme :*

$$P_\theta(\mathbf{z}) = \frac{P(\theta_1 + z_1, \dots, \theta_d + z_d)}{P(\theta_1, \dots, \theta_d)}, \quad (1.50)$$

pour tout $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T \in \Theta_{q,P}$.

Démonstration. Selon la proposition B.1.2, la FEN engendrée par la loi $\mathcal{G}(q,P)$ est l'ensemble des mesures de probabilité μ pour lesquelles il existe $\boldsymbol{\theta} \in \Theta_{q,P}$ telle que la transformée de Laplace de μ soit :

$$\begin{aligned} \mathcal{L}\mu(\mathbf{z}) &= \frac{\mathcal{L}_{\mathcal{G}(q,P)}(\boldsymbol{\theta} + \mathbf{z})}{\mathcal{L}_{\mathcal{G}(q,P)}(\boldsymbol{\theta})}, \\ &= \left(\frac{P(\theta_1 + z_1, \dots, \theta_d + z_d)}{P(\theta_1, \dots, \theta_d)} \right)^{-q}. \end{aligned}$$

La FEN $F_{\mathcal{G}(q,P)}$ fait donc partie des familles de LGM car le polynôme $P_{\boldsymbol{\theta}}(\mathbf{z}) = P(\theta_1 + z_1, \dots, \theta_d + z_d)/P(\theta_1, \dots, \theta_d)$ est un polynôme affine. \square

La FEN $F_{\mathcal{G}(q,P)}$ est donc, selon la proposition précédente, une sous-famille des LGM de paramètre de forme q fixé. Bien qu'en dimension $d = 1$, cette FEN corresponde exactement à la famille des LGM non dégénérées de paramètre de forme q fixé selon les résultats présentés à la section 1.1.2, ce n'est plus le cas lorsque $d \geq 2$ (voir le cas $d = 2$ à la section 1.2.2).

LGM et Lois de Wishart

Dans cette section, la relation entre les lois de Wishart et les LGM est mise en évidence. Une méthode de simulation de familles de LGM à partir de vecteurs gaussiens est alors proposée.

Les lois de Wishart [Mui82, p. 85–108] peuvent en effet être construites à partir de vecteurs gaussiens.

Définition 1.3.3. Soient L vecteurs aléatoires i.i.d. $\mathbf{Z}_1, \dots, \mathbf{Z}_L$ de \mathbb{R}^d distribuées suivant la loi normale multivariée centrée de matrice de covariance $\boldsymbol{\Sigma}$ notée $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Alors la matrice aléatoire symétrique de dimension $d \times d$:

$$\mathbf{A} = \frac{1}{2} \sum_{i=1}^L \mathbf{Z}_i \mathbf{Z}_i^T,$$

est distribuée selon une loi de Wishart centrée de paramètres L et $\boldsymbol{\Sigma}$ notée $\mathcal{W}_d(L, \boldsymbol{\Sigma})$. L'entier L est appelé par analogie avec la loi du χ^2 sur \mathbb{R} , le nombre de degrés de liberté de la loi de Wishart centrée.

À partir de la définition, on voit qu'en dimension $d = 1$ une loi de Wishart à L degrés de liberté correspond, à un facteur multiplicatif près, à une loi du χ^2 à L degrés de liberté. Par ailleurs le facteur $\frac{1}{2}$ dans la définition ci-dessus de la matrice \mathbf{A} n'est pas standard, mais il permet de simplifier les expressions et notamment celle de la transformée de Laplace d'une loi de Wishart. L'expression de la transformée de Laplace d'une loi de Wishart est en effet un résultat classique : si $\mathbf{A} \sim \mathcal{W}_d(L, \boldsymbol{\Sigma})$, alors la transformée de Laplace de \mathbf{A} qui existe pour toute matrice symétrique \mathbf{S} , telle que $\mathbf{I}_d + \mathbf{S}\boldsymbol{\Sigma}$ soit définie positive, est :

$$\mathcal{L}_{\mathbf{A}}(\mathbf{S}) = \mathbf{E}[\text{etr}(-\mathbf{S}\mathbf{A})] = \det(\mathbf{I}_d + \mathbf{S}\boldsymbol{\Sigma})^{-L/2}, \quad (1.51)$$

où \mathbf{I}_d est la matrice identité d'ordre d et $\text{etr}(\mathbf{M})$ est l'exponentielle de la trace de la matrice \mathbf{M} . La transformée de Laplace de la diagonale notée \mathbf{X} d'une matrice de Wishart $\mathbf{A} \sim \mathcal{W}_d(L, \boldsymbol{\Sigma})$ est

par ailleurs obtenue lorsque la matrice \mathbf{S} pour laquelle on évalue la transformée de Laplace est diagonale. En effet :

$$\text{tr}(\text{Diag}(\mathbf{z})\mathbf{A}) = \sum_{i=1}^d z_i A_{ii} = \mathbf{z}^T \mathbf{X},$$

si $\mathbf{A} = (A_{ij})_{1 \leq i, j \leq d}$, $\text{Diag}(\mathbf{z})$ est la matrice diagonale dont les éléments diagonaux sont $\mathbf{z} = (z_1, \dots, z_d)^T$ et si $\text{tr}(\mathbf{M})$ est la trace de la matrice \mathbf{M} . On obtient le résultat suivant :

Proposition 1.3.3. *Pour toute matrice définie positive Σ et tout entier non nul L , la diagonale notée \mathbf{X} de la matrice de Wishart $\mathbf{A} \sim \mathcal{W}_d(L, \Sigma)$ est une LGM de paramètre de forme $q = \frac{L}{2}$ et dont la transformée de Laplace est :*

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \mathbb{E} \left[e^{-\mathbf{z}^T \mathbf{X}} \right] = \mathbb{E} [\text{etr}(-\text{Diag}(\mathbf{z})\mathbf{A})] = \det(\mathbf{I}_d + \text{Diag}(\mathbf{z})\Sigma)^{-L/2}. \quad (1.52)$$

Démonstration. D'après la définition 1.3.2 d'une LGM, il suffit de montrer que l'expression $\det(\mathbf{I}_d + \text{Diag}(\mathbf{z})\Sigma)$ est un polynôme affine pour le vecteur \mathbf{z} . C'est chose faite en remarquant que pour tout $1 \leq i \leq d$, les éléments de la i -ème ligne, notée \mathbf{l}_i , de la matrice $\mathbf{I}_d + \text{Diag}(\mathbf{z})\Sigma$ sont des fonctions affines de la variable z_i et ne dépendent pas des variables z_j pour $1 \leq j \neq i \leq d$. Par conséquent il existe deux vecteurs lignes \mathbf{a} et \mathbf{b} de \mathbb{R}^d qui ne dépendent pas de \mathbf{z} tels que $\mathbf{l}_i = \mathbf{a}z_i + \mathbf{b}$. En utilisant la propriété de multilinéarité du déterminant par rapport à ses lignes, il vient que pour tout $1 \leq i \leq d$ le déterminant s'exprime comme suit :

$$\begin{aligned} \det(\mathbf{l}_1, \dots, \mathbf{l}_i, \dots, \mathbf{l}_d) &= \det(\mathbf{l}_1, \dots, \mathbf{a}z_i + \mathbf{b}, \dots, \mathbf{l}_d), \\ &= \det(\mathbf{l}_1, \dots, \mathbf{a}, \dots, \mathbf{l}_d) z_i + \det(\mathbf{l}_1, \dots, \mathbf{b}, \dots, \mathbf{l}_d), \\ &= A^{(-i)} z_i + B^{(-i)}, \end{aligned}$$

où $A^{(-i)}$ et $B^{(-i)}$ ne dépendent pas de z_i et sont des polynômes des variables z_j pour $1 \leq j \neq i \leq d$. On obtient bien un polynôme affine d'après la définition 1.3.1. \square

Les familles de lois ainsi définies, qui sont donc des sous-familles de LGM, ont notamment été étudiées par Krishnamoorthy, Parthasarathy [KP51] et par Griffiths [Gri84]. Ce dernier a identifié les conditions sur la matrice Σ pour que la loi soit infiniment divisible, i.e. pour que la loi dont la transformée de Laplace est (1.52) existe pour tout $L \in \mathbb{R}$. Il est d'ailleurs important de remarquer que si les diagonales de lois de Wishart sont des LGM, la réciproque n'est pas vraie. Bien que les lois de Wishart peuvent être généralisées à partir de l'expression de leur transformée de Laplace (1.51) au cas où le paramètre L n'est pas entier, ces lois sur $\mathbb{R}^{d \times d}$ n'existent que pour $L \in \{1, 2, \dots, d-2\} \cup [d-1, \infty[$. Au contraire, la famille de lois identifiée dans cette section, et plus généralement les LGM, peuvent être définies pour tout $q = \frac{L}{2} > 0$.

Il est facile d'obtenir les moyennes et les coefficients de corrélation entre les marges de la LGM obtenue comme la diagonale \mathbf{X} d'une loi de Wishart $\mathcal{W}_d(L, \Sigma)$. En effet pour tout $1 \leq i < j \leq d$, la transformée de Laplace du couple (X_i, X_j) a pour expression :

$$\mathcal{L}_{(X_i, X_j)}(z_i, z_j) = \mathcal{L}_{\mathbf{X}}(\tilde{\mathbf{z}}_{ij}) = \det(\mathbf{I}_d + \text{Diag}(\tilde{\mathbf{z}}_{ij})\Sigma)^{-L/2},$$

où $\tilde{\mathbf{z}}_{ij}$ est le vecteur de \mathbb{R}^d tel que ses i -ème et j -ème composantes sont respectivement z_i et z_j et toutes ses autres composantes sont nulles. Dans ce cas, on trouve en développant le déterminant obtenu par rapport à chaque ligne m tel que $1 \leq m \neq i, j \leq d$ l'expression suivante :

$$\begin{aligned} \mathcal{L}_{(X_i, X_j)}(z_i, z_j) &= \left[\det \begin{pmatrix} 1 + \sigma_{i,i}z_i & \sigma_{i,j}z_i \\ \sigma_{i,j}z_j & 1 + \sigma_{j,j}z_j \end{pmatrix} \right]^{-\frac{L}{2}}, \\ &= \left(1 + \sigma_{i,i}z_i + \sigma_{j,j}z_j + (\sigma_{i,i}\sigma_{j,j} - \sigma_{i,j}^2)z_i z_j \right)^{-\frac{L}{2}}, \end{aligned}$$

avec $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq d}$. D'après la formule (1.23), les moyennes et les coefficients de corrélation du vecteur \mathbf{X} sont :

$$\begin{aligned} \mathbf{E}[X_i] &= \frac{L}{2} \sigma_{i,i}, \quad \forall 1 \leq i \leq d, \\ r(X_i, X_j) &= \frac{L}{2} \frac{\sigma_{i,j}^2}{\sigma_{i,i}\sigma_{j,j}} \equiv r_{ij}, \quad \forall 1 \leq i < j \leq d. \end{aligned}$$

On en déduit alors l'algorithme 2 de simulation de LGM lorsque le paramètre de forme $q = \frac{L}{2}$ est un demi-entier fixé, pour toutes les moyennes $m_i > 0$ (ou de manière équivalente les paramètres d'échelle $p_i = \frac{m_i}{q} > 0$) et pour tous les coefficients de corrélation entre les marges $r_{ij} \in [0, 1[$.

Algorithm 2 Simulation de LGM à partir de vecteurs Gaussiens

1. simulation de $2q$ vecteurs Gaussiens de \mathbb{R}^d $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(2q)}$ i.i.d distribués suivant la loi $\mathcal{N}(\mathbf{0}, R_{1/2})$ où $R_{1/2} = (a_{ij})_{1 \leq i, j \leq d}$ avec :

$$\begin{aligned} a_{i,i} &= 1 \quad \forall 1 \leq i \leq d, \\ a_{i,j} &= a_{j,i} = \sqrt{r_{i,j}} \quad \forall 1 \leq i < j \leq d, \end{aligned}$$

2. sommation des carrés des composantes de chaque vecteur gaussien :

$$\begin{aligned} X_1 &= \frac{p_1}{2} \left[(Z_1^{(1)})^2 + \dots + (Z_1^{(2q)})^2 \right], \\ &\vdots \\ X_d &= \frac{p_d}{2} \left[(Z_d^{(1)})^2 + \dots + (Z_d^{(2q)})^2 \right]. \end{aligned}$$

Retour : le vecteur $(X_1, \dots, X_d)^T$

Il faut remarquer que même lorsque le paramètre de forme est un demi-entier fixé, on ne génère de cette façon qu'une sous-famille de LGM lorsque la dimension est $d > 2$. En effet les seuls degrés de liberté d'une telle loi sont les coefficients des monômes z_i pour $1 \leq i \leq d$ et $z_i z_j$ pour $1 \leq i < j \leq d$ du polynôme affine P de la LGM. Ces paramètres correspondent aux moyennes des marges et aux coefficients de corrélation entre chaque marge d'après l'équation (1.23). Les autres coefficients du polynôme affine P se déduisent des coefficients précédents et ne sont donc pas des paramètres pour ces familles de lois. Cependant comme cela a déjà été mentionné, dans

de nombreuses applications, la loi Gamma mutivariée provient de ce modèle à partir de vecteurs gaussiens.

Il n'existe en général pas de forme analytique exploitable de la densité de ces lois. Il est cependant possible d'exhiber quelques cas particuliers pour lesquels la densité admet une forme simple. Dans les deux sections ci-après, des cas particuliers de LGM pour lesquels une forme analytique simple de la densité est obtenue sont en effet présentés, tandis que dans la dernière section une forme générale, mais inexploitable numériquement, de la densité de ces lois est étudiée.

LGM éuicorrélées

On définit la famille de LGM *éuicorrélées* de la manière suivante : pour tout paramètre de forme $q > 0$, les familles de LGM éuicorrélées sur \mathbb{R}^d sont obtenues lorsque le polynôme affine noté P_{eq} est de la forme :

$$P_{\text{eq}}(\mathbf{z}) = \alpha \prod_{i=1}^n (1 + \beta_i z_i) + 1 - \alpha, \quad (1.53)$$

avec $\mathbf{z} = (z_1, \dots, z_d)$, $\beta_i > 0$ pour tout $1 \leq i \leq d$ et $\alpha = \frac{1}{1-r}$ pour tout $r \in [0, 1[$. Selon la proposition 1.3.1 et l'équation (1.23), si $\mathbf{X} = (X_1, \dots, X_d) \sim \mathcal{G}(q, P)$, le coefficient de corrélation entre les marges est alors :

$$r(X_i, X_j) = \frac{\alpha^2 \beta_i \beta_j - \alpha \beta_i \beta_j}{\alpha^2 \beta_i \beta_j} = \frac{\alpha - 1}{\alpha} = r \quad \forall 1 \leq i \leq d.$$

Le coefficient de corrélation est donc le même pour chaque couple de marges, d'où le nom de LGM *éuicorrélée*. Ces familles de lois ont été proposées par Gaver [Gav70]. Comme cela est expliqué dans le cas bivarié à la section 1.2.2, on peut supposer sans perte de généralité que les paramètres d'échelle β_i sont tous égaux à 1 pour tout $1 \leq i \leq d$. La transformée de Laplace associée à cette distribution devient alors :

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \left[\frac{-r}{1-r} + \frac{1}{1-r} \prod_{i=1}^n (1 + z_i) \right]^{-q}. \quad (1.54)$$

Le calcul en annexe C.2.1 démontre que si le vecteur \mathbf{X} a pour transformée de Laplace l'expression (1.54), alors la densité associée à la loi de ce vecteur est :

$$p_{\mathbf{X}}(\mathbf{x}) = (1-r)^q \left(\prod_{i=1}^d e^{-x_i} \frac{x_i^{q-1}}{\Gamma(q)} \right) \times {}_0F_{d-1}(q, \dots, q; r x_1 \dots x_d) \mathbb{I}_{]0, \infty[^d}(\mathbf{x}) \quad \forall q > 0, r \in [0, 1[, \quad (1.55)$$

où $\mathbf{x} = (x_1, \dots, x_d)^T$ et ${}_0F_{d-1}$ est la fonction hypergéométrique telle que :

$${}_0F_{d-1}(a_1, \dots, a_{d-1}; \mathbf{x}) = \sum_{k \geq 0} \frac{1}{(a_1)_k \dots (a_{d-1})_k} \frac{x^k}{k!} \quad \forall \mathbf{x} \in \mathbb{R},$$

où $(a)_k$ est le symbole de Pochhammer défini aux équations (1.21) et (1.22). Cette fonction hypergéométrique est absolument convergente sur \mathbb{R} . Son évaluation numérique se fait sans difficulté en large dimension, le terme principal de la série décroissant très rapidement.

Proposition 1.3.4. Soit N une v.a. de loi $\mathcal{BN}(q, 1 - r)$ avec $0 \leq r < 1$. Soient X_1, X_2, \dots, X_d d v.a. obtenues par composition de la loi négative binomiale par des lois Gamma $X_1, \dots, X_d \sim \mathcal{G}(q + N, 1)$, les v.a. $X_1|N, \dots, X_d|N$ étant mutuellement indépendantes. Alors le vecteur $\mathbf{X} = (X_1, \dots, X_d)$ est distribué suivant une LGM de paramètre de forme q et dont la transformée de Laplace correspond à celle définie à l'équation (1.54).

Preuve. La démonstration est en tout point similaire à celle effectuée dans le cas bivarié pour la proposition 1.2.2. \square

L'algorithme 1 de simulation présenté à la section 1.2.2 pour une loi Gamma bivariée s'étend alors sans difficulté aux cas des LGM éuicorrélées. Il faut noter que ces familles de lois sont en général différentes des familles présentées à la section 1.3.2, qui sont obtenues à partir de vecteurs gaussiens, et ceci, même lorsque le paramètre de forme est un demi-entier et que ces dernières vérifient la propriété d'éuicorrélation (i.e. que les corrélations entre tous les couples de marges sont identiques). Selon la méthode de construction proposée à l'algorithme 2, ces lois peuvent être obtenues lorsque la matrice de covariance du vecteur gaussien est la matrice $R_{1/2}$ dont les termes diagonaux sont 1 et dont tous les autres termes sont \sqrt{r} , r étant le coefficient de corrélation entre chaque marge. D'après la proposition 1.3.3, ces familles ont pour polynôme affine :

$$P(\mathbf{z}) = \det(\mathbf{I}_d + \text{Diag}(\mathbf{z})R_{1/2}),$$

et lorsque la dimension est $d > 2$, les coefficients des monômes de plus de deux variables de ce polynôme affine sont distincts de ceux du polynôme affine d'une loi Gamma éuicorrélée (1.53), même si les lois sont à la même échelle.

LGM à structure de corrélation markovienne

On appelle LGM à structure de corrélation markovienne la distribution du vecteur $\mathbf{X} = (X_1, \dots, X_d)^T$ dont la transformée de Laplace est de la forme (1.49), le paramètre de forme $q > 0$ étant fixé, et qui vérifie la propriété de Markov :

$$\Pr(X_{i+1} \in B | X_i = x_i, \dots, X_1 = x_1) = \Pr(X_{i+1} \in B | X_i = x_i), \quad (1.56)$$

pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$ et pour tout $i \in \{1, \dots, d-1\}$. Selon la formule de Bayes, on a la relation suivante entre la densité jointe du vecteur \mathbf{X} au point $\mathbf{x} = (x_1, \dots, x_d)^T$, notée $f_{\mathbf{X}}(\mathbf{x})$, et les densités conditionnelles :

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)f_{X_2|X_1=x_1}(x_2)f_{X_3|X_1=x_1, X_2=x_2}(x_3) \dots f_{X_d|X_1=x_1, X_2=x_2, \dots, X_{d-1}=x_{d-1}}(x_d),$$

où $f_{X_{i+1}|X_1=x_1, X_2=x_2, \dots, X_i=x_i}(x_{i+1})$ désigne la densité conditionnelle de la v.a. X_{i+1} sachant le vecteur $(X_1, \dots, X_i)^T$ pour tout $i \in \{1, \dots, d-1\}$. Une conséquence directe de la propriété de Markov 1.56 est :

$$f_{X_{i+1}|X_1=x_1, X_2=x_2, \dots, X_i=x_i}(x_{i+1}) = f_{X_{i+1}|X_i=x_i}(x_{i+1}) \quad \forall (x_1, \dots, x_{i+1})^T \in \mathbb{R}^{i+1}.$$

La densité jointe d'une LGM à structure de corrélation markovienne s'exprime donc :

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{X_1}(x_1)f_{X_2|X_1=x_1}(x_2)f_{X_3|X_2=x_2}(x_3) \dots f_{X_d|X_{d-1}=x_{d-1}}(x_d), \\ &= f_{(X_1, X_2)}(x_1, x_2) \prod_{i=2}^{d-1} \frac{f_{(X_i, X_{i+1})}(x_i, x_{i+1})}{f_{X_i}(x_i)}. \end{aligned}$$

Grâce à l'expression (1.30) de la densité d'une loi Gamma bivariée de transformée de Laplace :

$$\mathcal{L}_{(X_i, X_j)}(z_i, z_j) = \left(1 + p_i z_i + p_j z_j + p_i p_j (1 - r_{i,j}) z_i z_j\right)^{-q},$$

où $r_{i,j} = r(X_i, X_j)$, et en supposant que les paramètres d'échelle sont tous égaux à 1 (i.e. $p_1 = \dots = p_d = 1$), on obtient alors :

$$f_{\mathbf{X}}(\mathbf{x}) = e^{-\frac{x_1}{1-r_{1,2}} - \frac{x_d}{1-r_{d-1,d}} - \sum_{i=2}^{n-1} \frac{1-r_{i-1,i} r_{i,i+1}}{(1-r_{i-1,i})(1-r_{i,i+1})} x_i} \frac{(x_1 x_d)^{q-1}}{\Gamma(q)} \prod_{i=1}^{d-1} \frac{f_q\left(\frac{r_{i,i+1} x_i x_{i+1}}{(1-r_{i,i+1})^2}\right)}{(1-r_{i,i+1})^q}. \quad (1.57)$$

Il est montré en annexe C.2.2 que pour la structure markovienne décrite à l'équation (1.56), les coefficients de corrélation entre les marges de la LGM vérifient la propriété :

$$r_{i,j} = r_{i,k} r_{k,j}, \quad \forall 1 \leq i < k < j \leq d. \quad (1.58)$$

Une conséquence de cette propriété (1.58) est le fait que si R est la matrice de corrélation de la loi Gamma, alors R^{-1} est une matrice symétrique tridiagonale.

Par ailleurs il est également montré à l'annexe C.2.2 que la transformée de Laplace associée à la densité (1.57) a pour expression :

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \det\left(\mathbf{I}_n + \text{Diag}(\mathbf{z}) R_{1/2}\right)^{-q}, \quad (1.59)$$

où \mathbf{I}_d est la matrice identité d'ordre d , $\text{Diag}(\mathbf{z})$ la matrice diagonale dont les éléments diagonaux sont les composantes du vecteur \mathbf{z} , et $R_{1/2} = (a_{i,j})_{1 \leq i, j \leq d}$ la matrice de corrélation (donc symétrique) telle que :

$$\begin{aligned} a_{i,i} &= 1 & \forall 1 \leq i \leq d, \\ a_{i,i+1} &= \sqrt{r_{i,i+1}} & \forall 1 \leq i \leq d-1, \\ a_{i,j} &= \sqrt{r_{i,j}} = \sqrt{r_{i,k}} \sqrt{r_{k,j}}, & \forall 1 \leq i < k < j \leq d. \end{aligned} \quad (1.60)$$

Lorsque le paramètre de forme q est un demi-entier, ces lois peuvent être obtenues à partir de vecteurs gaussiens comme cela est expliqué à la section 1.3.2. Dans le cas général (i.e. q réel), ces familles de lois peuvent être obtenues comme la distribution stationnaire de processus autorégressifs d'ordre 1 [LS79, GL80], ce qui montre également que ces familles de lois sont infiniment divisibles.

Développement en série de polynômes de Laguerre de la densité d'une LGM

Il est possible d'obtenir en dimension d quelconque une expression générale de la densité d'une LGM sous la forme d'une série de polynôme de Laguerre. Le terme principal de cette série est cependant exprimé sous la forme de sommes multiples imbriquées (et le nombre de ces sommes augmente avec la dimension d) si bien que les expressions analytiques trouvées sont inexploitablement numériquement. Cette représentation peut être vue comme une généralisation en dimension quelconque de la représentation de Lancaster d'une loi Gamma bivariée 1.2.2 et peut être utile afin d'obtenir des propriétés théoriques concernant les LGM. Un exemple a déjà été donné en dimension $d = 2$ à la section 1.2.2. Dans cette section, le cas où la dimension est $d = 3$ est traité, afin d'éviter de manipuler des expressions trop lourdes, mais le calcul dans le cas

général $d \in \mathbb{N}$ s'effectue de manière analogue. Dans le cas trivarié, la transformée de Laplace d'un vecteur $\mathbf{X} = (X_1, X_2, X_3)^T$ distribué suivant une LGM est de la forme :

$$\mathcal{L}_{\mathbf{X}}(z_1, z_2, z_3) = (1 + p_1 z_1 + p_2 z_2 + p_3 z_3 + p_{12} z_1 z_2 + p_{13} z_1 z_3 + p_{23} z_2 z_3 + p_{123} z_1 z_2 z_3)^{-q}.$$

Comme $p_{ij} = p_i p_j (1 - r_{ij})$ où $r_{ij} = r(X_i, X_j)$ selon (1.23), en notant $\rho = 1 - \frac{p_{123}}{p_1 p_2 p_3}$ on se ramène, par changement d'échelle $\mathbf{Y} = (X_1/p_1, X_2/p_2, X_3/p_3)^T$ au cas équivalent où la transformée de Laplace a pour expression :

$$\mathcal{L}_{\mathbf{Y}}(z_1, z_2, z_3) = ((1 + z_1)(1 + z_2)(1 + z_3) - r_{12} z_1 z_2 - r_{13} z_1 z_3 - r_{23} z_2 z_3 - \rho z_1 z_2 z_3)^{-q}.$$

Le calcul effectué à l'annexe C.2.3 montre alors que la densité du vecteur \mathbf{Y} s'écrit :

$$p_{\mathbf{Y}}(y_1, y_2, y_3) = \left(\prod_{i=1}^3 e^{-y_i} \frac{y_i^{q-1}}{\Gamma(q)} \right) \sum_{n \geq 0} (q)_k \sum_{(n_i)} \frac{r_{12}^{n_1} r_{13}^{n_2} r_{23}^{n_3} \rho^{n_4}}{n_1! n_2! n_3! n_4!} \prod_{i=1}^3 \frac{N_i!}{(q)_{N_i}} L_{N_i}^{(q-1)}(y_i) \mathbb{I}_{]0, \infty[^3}(y_1, y_2, y_3), \quad (1.61)$$

où le symbole $\sum_{(n_i)}^n$ désigne la sommation pour tout les entiers naturels (n_1, n_2, n_3, n_4) tels que :

$$\sum_{i=1}^4 n_i = n_1 + n_2 + n_3 + n_4 = n,$$

les entiers N_i étant définis par :

$$N_1 = n - n_3, \quad N_2 = n - n_2, \quad N_3 = n - n_1,$$

et $L_n^{(\alpha)}$ désignant le polynôme de Laguerre généralisé, défini à la section 1.2.2. Les coefficients de corrélation étant tous dans l'intervalle $[0, 1]$, la série obtenue est alors absolument convergente si $|\rho| < 1$. Le cas d'une loi Gamma générée à partir de vecteurs gaussiens a ainsi été étudié sous cette forme de séries de polynômes de Laguerre dans [KP51, NKN04]. Dans ce cas, la transformée de Laplace de la loi Gamma obtenue a pour expression :

$$\mathcal{L}_{\mathbf{Y}}(z_1, z_2, z_3) = \left[\det \left(\mathbf{I}_3 + \text{Diag} \left((z_1, z_2, z_3)^T \right) R_{1/2} \right) \right]^{-q},$$

où q est un demi-entier et $R_{1/2} = \left(\sqrt{r_{ij}} \right)_{1 \leq i, j \leq 3}$, avec la convention $r_{ii} = 1$ pour $1 \leq i \leq d$, selon les résultats présentés à la section 1.3.2. Il vient alors en développant le déterminant que $\rho = -2 \sqrt{r_{12} r_{13} r_{23}} + r_{12} + r_{13} + r_{23}$, et l'on peut montrer que pour tout $r_{ij} \in [0, 1]$:

$$| -2 \sqrt{r_{12} r_{13} r_{23}} + r_{12} + r_{13} + r_{23} | < 1.$$

Une conséquence est que la série (1.61) est dans ce cas absolument convergente.

La représentation de la densité d'une LGM sous forme d'une série de polynômes de Laguerre a donc un intérêt purement théorique. Ces représentations sont en effet inexploitablement numériquement lorsque $d \geq 3$ du fait de leur complexité calculatoire due aux sommes multiples imbriquées. De plus le calcul de la série, dont les termes ne sont pas de signe constant, est instable numériquement lorsque les valeurs du coefficient de corrélation sont proches de 1.

1.3.3 Lois multinomiales négatives (LMN)

Tout comme la famille des LGM, le modèle des lois multinomiales négatives qui nous intéressent a été identifié par Bar-Lev *et al* [BLBE⁺94] au cours de leur classification des familles exponentielles naturelles sur \mathbb{R}^d . Ces familles multivariées sont concentrées sur \mathbb{N}^d et sont caractérisées par leur fonction génératrice qui est une extension en dimension d quelconque de la fonction génératrice d'une loi binomiale négative sur \mathbb{R}^d .

Au cours de cette section, ces lois sont dans un premier temps définies. Les familles exponentielles engendrées par ces lois sont ensuite caractérisées. Enfin le cas bivarié, où une expression analytique simple des masses de probabilité est disponible, est étudié plus en détail.

Les LMN sont caractérisées de la manière suivante.

Définition 1.3.4. Soit $q > 0$ et soit $P(\mathbf{z})$ un polynôme affine en dimension d (voir la définition 1.3.1). Une loi multinomiale négative concentrée sur \mathbb{N}^d de paramètres q et $P(\mathbf{z})$, notée $\mathcal{MN}(q, P)$, est définie par sa fonction génératrice :

$$G_{\mathcal{MN}(q,P)}(\mathbf{z}) = [P(\mathbf{z})]^{-q}, \quad (1.62)$$

le polynôme affine satisfaisant la contrainte évidente $P(\mathbf{1}) = 1$.

Remarques :

- comme pour les LGM, le polynôme P est appelé paramètre, mais les paramètres sont en fait les coefficients de ce polynôme.
- par analogie avec les LGM, le paramètre q est appelé le paramètre de forme même si cette appellation est abusive car le polynôme P ne correspond pas à un paramètre d'échelle.
- ces lois sont nommées lois *multinomiales négatives* car si la puissance $-q$ de la fonction génératrice (1.62) était positive, on obtiendrait une loi de type multinomiale. Ici la puissance $-q$ est négative d'où le nom de cette loi.

Des exemples de polynôme affine en dimension $d = 1, 2, 3$ ont été présentés à l'exemple 1.3.1 lorsque les polynômes vérifient la contrainte $P(\mathbf{0}) = 1$, alors que dans le cas des LMN la condition nécessaire est $P(\mathbf{1}) = 1$. Les polynômes affines satisfaisant cette dernière contrainte se déduisent des précédents comme suit $Q(\mathbf{z}) = P(\mathbf{z}) - P(\mathbf{1}) + 1$.

Comme pour les LGM, il n'est en général pas possible de déterminer facilement les conditions nécessaires et suffisantes sur le couple de paramètres (q, P) pour que la loi multivariée existe. Ces conditions peuvent être cependant déterminées dans le cas bivarié, comme cela sera montré à la section 1.3.3. De plus, le cas où ces distributions sont infiniment divisibles, i.e. le cas où la loi existe pour un polynôme affine P donné, quel que soit le paramètre de forme $q > 0$, a été étudié par Bernardoff dans [Ber03].

Proposition 1.3.5. Les lois marginales d'un vecteur $\mathbf{N} = (N_1, \dots, N_d)^T$ de loi $\mathcal{MN}(q, P)$ sont des lois multinomiales négatives (binomiales négatives dans le cas univarié) qui ont toutes le même paramètre de forme q .

Cette propriété importante des LMN traduit le fait que la loi de $\tilde{\mathbf{N}} = ((N_i)_{i \in I})^T$, où I est un sous-ensemble de $\{1, \dots, d\}$, est une loi multinomiale négative (binomiale négative si $\text{card } I = 1$).

Démonstration. Cette propriété découle directement de la définition 1.3.4 d'une loi Gamma multivariée à partir de sa transformée de Laplace, car, en notant $\tilde{\mathbf{z}} = (z_i)_{i \in I}$ et $\widehat{\mathbf{z}}$ le vecteur de \mathbb{R}^d dont la i -ème composante est z_i si $i \in I$, 1 sinon, on a :

$$G_{\tilde{N}}(\tilde{\mathbf{z}}) = \mathbb{E} \left[\prod_{i \in I} z_i^{N_i} \right] = \mathbb{E} \left[\prod_{i=1}^d \widehat{z}_i^{N_i} \right] = G_N(\widehat{\mathbf{z}}) = [P(\widehat{\mathbf{z}})]^{-q} = [\tilde{P}(\tilde{\mathbf{z}})]^{-q},$$

où le polynôme $\tilde{P}(\tilde{\mathbf{z}}) = P(\widehat{\mathbf{z}})$ est par construction un polynôme affine par rapport à $\tilde{\mathbf{z}}$. \square

Les moments joints d'une LGM peuvent se calculer à partir de sa fonction génératrice (1.49) grâce à la proposition A.1.2.

Les FEN générées par les LMN

On montre dans cette section que les FEN engendrées par une LMN sont des familles de LMN. On considère la loi $\mathcal{MN}(q, P)$ sur \mathbb{R}^d caractérisée par sa fonction génératrice (1.62). Selon la proposition A.2.6, on obtient pour la transformée de Laplace associée à cette loi l'expression (avec ici la convention $\varepsilon = +1$) :

$$\mathcal{L}_{\mathcal{MN}(q, P)}(\mathbf{z}) = [P(\exp(\mathbf{z}))]^{-q}, \quad (1.63)$$

et l'on note $\Theta_{q, P}$ l'intérieur du domaine de définition de la transformée de Laplace. On obtient alors la proposition suivante :

Proposition 1.3.6. *Soit $q > 0$ et P un polynôme affine tel que $P(\mathbf{1}) = 1$. La FEN engendrée par la loi $\mathcal{MN}(q, P)$, notée $F_{\mathcal{MN}(q, P)}$ est la famille de lois multinomiales négatives de paramètre de forme q et dont les polynômes affines notés $P_{\boldsymbol{\theta}}$ sont de la forme :*

$$P_{\boldsymbol{\theta}}(\mathbf{z}) = \frac{P(e^{\theta_1+z_1}, \dots, e^{\theta_d+z_d})}{P(e^{\theta_1}, \dots, e^{\theta_d})}, \quad (1.64)$$

pour tout $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T \in \Theta_{q, P}$.

Démonstration. Selon la proposition B.1.2, la FEN engendrée par la loi $\mathcal{MN}(q, P)$ est l'ensemble des mesures de probabilité μ pour lesquelles il existe $\boldsymbol{\theta} \in \Theta_{q, P}$ telle que la transformée de Laplace de μ soit :

$$\mathcal{L}_{\mu}(\mathbf{z}) = \frac{\mathcal{L}_{\mathcal{MN}(q, P)}(\boldsymbol{\theta} + \mathbf{z})}{\mathcal{L}_{\mathcal{MN}(q, P)}(\boldsymbol{\theta})} = \left(\frac{P(e^{\theta_1+z_1}, \dots, e^{\theta_d+z_d})}{P(e^{\theta_1}, \dots, e^{\theta_d})} \right)^{-q}.$$

La fonction génératrice des éléments de cette FEN est donc selon la proposition B.1.2 :

$$G_{\mu}(\mathbf{z}) = \left(\frac{P(e^{\theta_1 z_1}, \dots, e^{\theta_d z_d})}{P(e^{\theta_1}, \dots, e^{\theta_d})} \right)^{-q},$$

pour tout $\mathbf{z} \in \Theta_{q, P}$. La FEN $F_{\mathcal{MN}(q, P)}$ fait donc partie des familles de LMN car le polynôme $P_{\boldsymbol{\theta}}(\mathbf{z}) = \frac{P(e^{\theta_1 z_1}, \dots, e^{\theta_d z_d})}{P(e^{\theta_1}, \dots, e^{\theta_d})}$ est un polynôme affine. \square

La FEN $F_{\mathcal{MN}(q, P)}$ est donc selon la proposition précédente une sous-famille des LMN de paramètre de forme q fixé. Bien qu'en dimension $d = 1$, cette FEN corresponde exactement à la famille des lois binomiales négatives non dégénérées de paramètre de forme q fixé selon les résultats présentés à la section 1.1.2, ce n'est plus le cas lorsque $d \geq 2$.

La loi multinomiale négative bivariée

Le cas bivarié est intéressant car il est alors possible d'obtenir une expression analytique explicite de la mesure de probabilité de ces lois, ainsi qu'une expression explicite des FEN engendrées par ces lois. Dans cette section, une paramétrisation adéquate de ces familles de lois bivariées est introduite dans un premier temps. Quelques propriétés de ces lois sont ensuite données, ainsi que les conditions pour lesquelles ces lois existent. Enfin des expressions explicites des masses de probabilité et des FEN engendrées par ces lois sont proposées.

Afin de déterminer les masses de probabilité d'une LMN bivariée, il est utile de remarquer que selon la définition 1.3.4 d'une LMN, le polynôme affine de cette loi vérifie les conditions nécessaires $P(\mathbf{1}) = 1$ car, par définition, la fonction génératrice d'une loi de probabilité vaut 1 au point $\mathbf{1}$, et $P(\mathbf{0}) \neq 0$, sinon la fonction génératrice de la LMN n'est pas définie au point $\mathbf{0}$. Or selon la définition de la fonction génératrice d'une loi de probabilité, la fonction génératrice est toujours définie en $\mathbf{0}$ (et est même égale à en ce point à la masse $\Pr(\mathbf{N} = \mathbf{0})$).

Finalement il vient qu'en dimension $d = 2$, tout polynôme affine vérifiant les conditions $P(\mathbf{1}) = 1$ et $P(\mathbf{0}) \neq 0$ peut être paramétré par le triplet $(a, b, c) \in \mathbb{R}^3$ tel que $(1 - a)(1 - b) - c \neq 0$ de la manière suivante :

$$P_{a,b,c}(z_1, z_2) = \frac{1 - az_1 - bz_2 + (ab - c)z_1z_2}{(1 - a)(1 - b) - c}. \quad (1.65)$$

En effet un polynôme affine vérifiant la condition $P(\mathbf{1}) = 1$ a pour expression générale $1 - p_1 - p_2 - p_{12} + p_1z_1 + p_2z_2 + p_{12}z_1z_2$ avec $p_1, p_2, p_{12} \in \mathbb{R}$. Si $1 - p_1 - p_2 - p_{12} \neq 0$, i.e. si le polynôme ne s'annule pas au point $(0, 0)$, alors on peut exprimer facilement les paramètres a, b, c de l'équation (1.65) en fonction des paramètres p_1, p_2, p_{12} . On note par la suite $P_{a,b,c}$ le polynôme affine de \mathbb{R}^2 paramétré par (a, b, c) selon l'équation (1.65).

Avec cette paramétrisation, la fonction génératrice du couple $(N_1, N_2) \sim \mathcal{MN}(q, P_{a,b,c})$ a pour expression

$$G_{(N_1, N_2)}(z_1, z_2) = \mathbb{E} \left[z_1^{N_1} z_2^{N_2} \right] = \left[\frac{(1 - a)(1 - b) - c}{1 - az_1 - bz_2 + (ab - c)z_1z_2} \right]^q, \quad (1.66)$$

À partir de l'expression ci-dessus de la fonction génératrice du couple (N_1, N_2) , en posant respectivement $z_1 = 1$ puis $z_2 = 1$ on obtient $N_1 \sim \mathcal{NB}(q, \frac{(1-a)(1-b)+c}{1-b})$ et $N_2 \sim \mathcal{NB}(q, \frac{(1-a)(1-b)+c}{1-a})$ selon l'expression (1.8) de la fonction génératrice de la loi binomiale négative. De plus des calculs directs à partir de l'expression de la fonction génératrice (voir la proposition A.1.2) permettent d'obtenir l'expression du moment joint $\mathbb{E} [N_1 N_2]$. Enfin, on obtient les expressions :

$$\begin{aligned} \mathbb{E} [N_1] &= q \frac{a(1 - b) + c}{(1 - a)(1 - b) - c}, & \mathbb{E} [N_2] &= q \frac{b(1 - a) + c}{(1 - a)(1 - b) - c}, \\ \text{var} (N_1) &= q \frac{(a(1 - b) + c)(1 - b)}{((1 - a)(1 - b) - c)^2}, & \text{var} (N_2) &= q \frac{(b(1 - a) + c)(1 - a)}{((1 - a)(1 - b) - c)^2}, \\ \text{cov} (N_1, N_2) &= q \frac{c}{((1 - a)(1 - b) - c)^2}. \end{aligned} \quad (1.67)$$

Le couple (N_1, N_2) est donc à dépendance positive, car $q > 0$ et $c \geq 0$ si $(a, b, c) \in \Delta$ et donc $\text{cov} (N_1, N_2) \geq 0$, et n'est pas corrélé lorsque $c = 0$. Lorsque $c = 0$, la fonction génératrice du

couple devient :

$$G_{(N_1, N_2)}(z_1, z_2) = \left[\frac{(1-a)(1-b)}{(1-az_1)(1-bz_2)} \right]^q = \left[\frac{1-a}{(1-az_1)} \right]^q \left[\frac{1-b}{(1-bz_2)} \right]^q = G_{N_1}(z_1)G_{N_1}(z_2),$$

et donc les variables N_1 et N_2 sont indépendantes. Comme pour les LGM, absence de corrélation et indépendance entre les marges sont des propriétés équivalentes pour les LMN. On trouve par ailleurs grâce aux expressions de la covariance et des variances que si $a = b$, alors le coefficient de corrélation entre les marges prend toutes les valeurs de l'intervalle $[0, 1[$ lorsque le paramètre c parcourt l'ensemble tel que $(a, a, c) \in \Delta$. L'absence de corrélation correspond au cas où $c = 0$ et la corrélation totale au cas où c tend par valeur inférieure vers $(1-a)^2$.

Il est par ailleurs possible de déterminer les conditions nécessaires et suffisantes sur le couple $(q, P_{a,b,c})$ pour que la loi $\mathcal{MN}(q, P_{a,b,c})$ existe.

Proposition 1.3.7. *La loi du couple $(N_1, N_2) \sim \mathcal{MN}(q, P_{a,b,c})$ caractérisée par la fonction génératrice (1.66) existe pour tout $q > 0$ si les paramètres a, b, c vérifient les conditions suivantes :*

i $a, b, c \leq 0,$

iii $(1-a)(1-b) - c > 0.$

Ces conditions sont également nécessaires, et ce quel que soit la valeur du paramètre de forme $q > 0$.

La preuve de cette démonstration est donnée à l'annexe C.3.2. Une conséquence de la proposition est que ces lois multinomiales négatives bivariées sont infiniment divisibles puisque les conditions énoncées à la proposition 1.3.7 sur le polynôme affine de cette loi ne dépendent pas du paramètre de forme $q > 0$.

En excluant les cas dégénérés $a = 0$ et $b = 0$ (qui correspondent respectivement au cas $N_1 = 0$ et $N_2 = 0$), l'ensemble des paramètres qui satisfont les conditions énoncées à la proposition 1.3.7 peut alors s'écrire :

$$\Delta = \{(a, b, c) \in]0, 1[\times]0, 1[\mid (1-a)(1-b) - c > 0\}. \quad (1.68)$$

En définitive, on obtient également l'expression explicite des masses de probabilité de cette loi bivariée.

Théorème 1.3.8. *Soit $q > 0$ et $P_{a,b,c}$ un polynôme affine tel que $(a, b, c) \in \Delta$. Alors une expression explicite des masses de probabilité du couple $(N_1, N_2) \sim \mathcal{MN}(q, P_{a,b,c})$ est :*

$$\Pr(N_1 = m, N_2 = n) = a^m b^n [(1-a)(1-b) - c]^q \sum_{k=0}^{\min(m,n)} \frac{(q)_k}{k!} \frac{(q+k)_{m-k}}{(m-k)!} \frac{(q+k)_{n-k}}{(n-k)!} \left(\frac{c}{ab}\right)^k, \quad (1.69)$$

pour tout $(m, n) \in \mathbb{N}^2$, $(a)_n$ étant le symbole de Pochhammer défini à l'équation (1.21).

Une démonstration de ce théorème est donnée à l'annexe C.3.1.

En raison du résultat énoncé à la proposition 1.3.6, on montre que les FEN des lois multinomiales négatives bivariées sont les familles obtenues lorsque $q > 0$ est fixé et lorsque le paramètre $\rho = \frac{c}{ab}$ est lui aussi fixé. Une base des FEN obtenues pour chaque couple (q, ρ) est la mesure suivante :

$$\mu_{q,\rho}(dx, dy) = \sum_{m,n \geq 0} \sum_{k=0}^{\min(m,n)} \frac{(q)_k}{k!} \frac{(q+k)_{m-k}}{(m-k)!} \frac{(q+k)_{n-k}}{(n-k)!} \rho^k \delta_{m,n}(dx, dy). \quad (1.70)$$

1.3.4 Lois de Poisson Mélangées Multivariées (LPMM)

Les lois ou les processus de Poisson mélangés univariés ont été étudiés de façon intensive afin de modéliser les occurrences d'événements rares (voir par exemple le livre de Grandell [Gra97] et les références citées dans le livre). Une très grande quantité d'articles sur le sujet appliqués à des domaines divers a ainsi été publiée depuis un siècle. En comparaison, les lois de Poisson mélangées multivariées (LPMM) ont reçu une attention bien moindre. Malgré les travaux recensés depuis les années 1950 dans des domaines aussi variés que la modélisation des accidents aériens [BN52], les assurances de voitures [Pic76], la victimologie [Nel84], l'optique statistique [AS04, RRG06b, FLT04] ou des travaux plus théoriques sur ces processus [Zoc05], beaucoup de résultats théoriques concernant les LPMM sont encore à étudier.

Un exemple de loi de Poisson mélangée univariée a déjà brièvement été présenté à la proposition 1.1.2. Une loi de Poisson mélangée multivariée en est une extension lorsque la loi de mélange est multivariée.

Définition 1.3.5. Soit $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)^T$ un vecteur aléatoire distribué sur \mathbb{R}_+^d . Un vecteur aléatoire $\mathbf{N} = (N_1, \dots, N_d)^T$ est distribué selon une loi de Poisson mélangée multivariée (LPMM) si :

- les v.a. $(N_i|\lambda_i)_{i=1, \dots, d}$ sont mutuellement indépendantes.
- la v.a. $(N_i|\lambda_i)$ suit la loi de Poisson $\mathcal{P}(\lambda_i)$ pour tout $1 \leq i \leq d$.

La loi du vecteur $\boldsymbol{\lambda}$ est qualifiée de loi de mélange de la LPMM, ou encore de loi des intensités du mélange.

On adopte la notation suivante :

$$\mathbf{N} \sim \text{LPMM}(\boldsymbol{\lambda}),$$

afin de désigner la LPMM suivie par le vecteur \mathbf{N} , la loi de mélange étant celle du vecteur aléatoire $\boldsymbol{\lambda}$.

Selon la définition 1.3.5, une LPMM est entièrement caractérisée par sa loi de mélange. Il est d'ailleurs possible d'exprimer la loi obtenue en fonction de la loi de mélange.

Proposition 1.3.9. Lorsque la loi du vecteur intensité $\boldsymbol{\lambda}$ est absolument continue sur son domaine de définition (\mathbb{R}_+^d) , on obtient l'expression suivante des masses du vecteur discret $\mathbf{N} \sim \text{LPMM}(\boldsymbol{\lambda})$:

$$\Pr(\mathbf{N} = \mathbf{k}) = \int \dots \int_{\mathbb{R}_+^d} \prod_{\ell=1}^d \frac{(\lambda_\ell)^{k_\ell}}{k_\ell!} \exp(-\lambda_\ell) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \quad (1.71)$$

où $f(\cdot)$ est la densité du vecteur $\boldsymbol{\lambda}$.

Cette formule est une conséquence directe de la définition d'une LPMM. L'expression analytique obtenue est cependant en général trop complexe pour être utilisée afin d'obtenir directement les masses du vecteur \mathbf{N} .

Moments d'une LPMM

Les moments d'une LPMM se déduisent de ceux de la loi de mélange grâce à une relation simple. Avant de déterminer l'expression des moments joints standard d'une LPMM, il est judicieux d'étudier les moments joints factoriels qui ont une expression plus simple. En notant

$N^{[r]} = N(N-1)\dots(N-r+1)$, on obtient en effet :

$$\mathbb{E} \left[\prod_{k=1}^d N_k^{[r_k]} \right] = \mathbb{E} \left[\mathbb{E} \left[\prod_{k=1}^d N_k^{[r_k]} \mid \boldsymbol{\lambda} \right] \right] = \mathbb{E} \left[\prod_{k=1}^d \mathbb{E} \left[N_k^{[r_k]} \mid \lambda_k \right] \right] = \mathbb{E} \left[\prod_{k=1}^d \lambda_k^{r_k} \right].$$

La dernière égalité provenant de l'expression des moments factoriels d'une loi de Poisson. À partir de l'expression obtenue des moments factoriels, on obtient le résultat énoncé à la proposition ci-dessus.

Proposition 1.3.10. *Les moments joints d'une LPPM s'obtiennent en fonction de ceux de la loi de mélange grâce à la relation :*

$$\mathbb{E} \left[\prod_{k=1}^d N_k^{r_k} \right] = \sum_{j_1=0}^{r_1} \dots \sum_{j_d=0}^{r_d} \prod_{k=1}^d S(r_k, j_k) \mathbb{E} \left[\prod_{k=1}^d \lambda_k^{j_k} \right], \quad (1.72)$$

où $S(j, k)$ est le nombre de Stirling de seconde espèce [JKK92] (le nombre de Stirling $S(r_k, j_k)$ est le nombre de partitions possibles d'un ensemble de r_k éléments en exactement j_k sous-ensembles non vides).

L'expression (1.72) des moments joints classiques est en effet obtenue à partir des moments factoriels en décomposant chaque polynôme $N_k^{r_k}$ de la variable N_k dans la base des polynômes à degré étagés $(N_k^{[r]})_{r \leq r_k}$. Les propriétés suivantes découlent alors de la proposition précédente.

Corollaire 1.3.11. *La moyenne et la covariance du vecteur \mathbf{N} s'expriment respectivement comme :*

$$\begin{aligned} \mathbb{E}[\mathbf{N}] &= \mathbb{E}[\boldsymbol{\lambda}], \\ \text{cov}(\mathbf{N}) &= \text{cov}(\boldsymbol{\lambda}) + \text{Diag}(\mathbb{E}[\boldsymbol{\lambda}]), \end{aligned} \quad (1.73)$$

où $\text{Diag}(\mathbf{x})$ est la matrice diagonale dont les éléments diagonaux sont les composantes du vecteur $\mathbf{x} = (x_1, \dots, x_d)^T$.

La notion de surdispersion n'est pas définie de manière standard dans le cas multivarié, mais, en prenant comme définition le fait qu'un vecteur \mathbf{X} suit une loi surdispersée si la matrice $\text{cov}(\mathbf{X}) - \text{Diag}(\mathbb{E}[\mathbf{X}])$ est définie positive, on obtient alors selon l'équation (1.73) que les LPPM sont des familles de lois surdispersées lorsque la loi de mélange n'est pas dégénérée.

Fonction génératrice

Avec les notations vectorielles, la fonction génératrice du vecteur \mathbf{N} qui suit une LPMM s'exprime simplement en fonction de la transformée de Laplace de loi de mélange du vecteur $\boldsymbol{\lambda}$.

Théorème 1.3.12. *Soit $G_{\mathbf{N}}$ la fonction génératrice du vecteur \mathbf{N} (voir la définition A.1.1), et $\mathcal{L}_{\boldsymbol{\lambda}}$ la transformée de Laplace du vecteur $\boldsymbol{\lambda}$ (avec la convention $\varepsilon = -1$, voir la définition A.2.1). Si $\mathbf{N} \sim \text{LPMM}(\boldsymbol{\lambda})$, pour tout vecteur \mathbf{z} tel que $G_{\mathbf{N}}$ soit fini, on a :*

$$G_{\mathbf{N}}(\mathbf{z}) = \mathcal{L}_{\boldsymbol{\lambda}}(\mathbf{1} - \mathbf{z}). \quad (1.74)$$

Preuve. La démonstration s'effectue en calculant l'espérance conditionnelle par rapport au vecteur du mélange.

$$\begin{aligned} G_N(\mathbf{z}) &= \mathbb{E} \left[\mathbf{z}^N \right] = \mathbb{E} \left[\mathbb{E} \left[\mathbf{z}^N | \boldsymbol{\lambda} \right] \right] = \mathbb{E} \left[\prod_{k=1}^d \mathbb{E} \left[z_k^{N_k} | \lambda_k \right] \right], \\ &= \mathbb{E} \left[\prod_{k=1}^d e^{-(1-z_k)\lambda_k} \right] = \mathcal{L}_{\boldsymbol{\lambda}}(\mathbf{1} - \mathbf{z}). \end{aligned}$$

La deuxième ligne provient de l'expression de la fonction génératrice d'une loi de Poisson. \square

LPMM et familles exponentielles naturelles

Dans ce paragraphe, on parlera de famille exponentielle naturelle $F = F(\mu)$ même pour les cas dégénérés où la famille est concentrée sur un hyperplan affine noté H de \mathbb{R}^d (l'hypothèse 1 de la définition d'une FEN donnée à la section B.1 est donc invalidée). Par conséquent, la log-transformée de Laplace de la mesure génératrice μ est seulement convexe sur l'intérieur Θ_μ de son domaine de définition, même si elle reste strictement convexe sur le convexe $\Theta_\mu \cap H$. De plus, l'estimateur du maximum de vraisemblance des paramètres de la FEN n'est pas défini si la moyenne empirique des observations n'appartient pas à l'hyperplan H en question.

On considère une FEN F dont le support est dans $]0, \infty[^d$ et caractérisée par son ensemble de moyennes M_F et sa fonction variance V_F . On s'intéresse alors à la famille de LPMM obtenue lorsque les lois de mélange parcourent la FEN F :

$$\text{LPMM}(F) = \{\text{LPMM}(\mu) \mid \mu \in F\},$$

où $\text{LPMM}(\mu)$ désigne la LPMM obtenue pour la loi de mélange μ . Bien que l'on puisse penser que l'on obtient une nouvelle FEN G telle que $M_G \supset M_F$ et telle que $V_G(m) = m + V_F(m)$ pour tout $m \in M_F$, selon l'équation (1.73), il est montré dans [FLT04, FLT07] que ce n'est en général pas le cas. Les auteurs exhibent en effet la condition nécessaire et suffisante sur la FEN F , énoncée au théorème ci-dessous, pour que la famille $\text{LPMM}(F)$ soit encore une FEN.

Théorème 1.3.13. *Si la famille $\text{LPMM}(F)$, où F est une FEN concentrée sur $]0, \infty[^d$, est encore une FEN, alors il existe une partition $\{T_0, T_1, \dots, T_N\}$ de l'ensemble $\{1, \dots, d\}$ (avec éventuellement $T_0 = \emptyset$), des nombres réels $a_1, \dots, a_d \geq 0$ et des nombres réels $q_1, \dots, q_N > 0$ tels que la mesure ν , dont la transformée de Laplace (avec la convention $\varepsilon = -1$) est :*

$$\mathcal{L}_\nu(\boldsymbol{\theta}) = e^{-\sum_{k \in T_0} a_k \theta_k} \prod_{m=1}^N \left(1 + \sum_{k \in T_m} a_k \theta_k \right)^{-q_m}, \quad (1.75)$$

soit une base de la FEN F , i.e. $F = F(\nu)$.

Une démonstration de ce théorème est donnée dans [FLT07]. Une première constatation est qu'en dimension $d = 1$, les seules FEN qui vérifient les conditions sont celle engendrées par une masse de dirac (cas dégénéré), ou les FEN des lois Gamma de paramètres de forme $q > 0$ fixé. Une deuxième constatation est que si la loi de mélange est une FEN engendrée par une loi Gamma multivariée (en dimension $d > 1$), la famille de LPMM obtenue est une FEN dans les deux cas suivants :

1. les marges de la loi Gamma sont toutes deux à deux indépendantes. En effet une base ν de la FEN de ces lois Gamma multivariées a pour transformée de Laplace :

$$\mathcal{L}_\nu(\mathbf{z}) = \prod_{m=1}^d (1 + p_m z_m)^{-q}.$$

2. toutes les marges sont totalement corrélées. En effet une base ν de la FEN de ces lois Gamma multivariées a pour transformée de Laplace :

$$\mathcal{L}_\nu(\mathbf{z}) = \left(1 + \sum_{k=1}^d p_k z_k \right)^{-q}.$$

Dans ce cas la mesure de probabilité ν correspond à la loi du vecteur $\mathbf{X} = X(p_1, p_2, \dots, p_d)^T$ où X est une variable aléatoire de loi Gamma standard $\mathcal{G}(q, 1)$. La loi ν est donc concentrée sur la demi-droite de \mathbb{R}^d paramétrée par $t(p_1, p_2, \dots, p_d)^T$ avec $t \in]0, \infty[$.

Ceci nous amène à étudier plus en détail les LPMM dont la loi de mélange est une loi Gamma multivariée.

LPMM générées par une loi Gamma multivariée

On suppose ici que la loi de mélange est la loi Gamma multivariée $\mathcal{G}(q, P)$ dont la transformée de Laplace est donnée à la définition 1.3.2. Les paramètres de la LGM sont donc le polynôme affine P et le paramètre de forme $q > 0$, et l'ensemble de définition de la LGM noté $\Theta_{q,P}$. Soit \mathbf{N} un vecteur distribué suivant la LPMM ainsi obtenue. Alors, selon le théorème 1.3.12, on obtient l'expression suivante de la fonction génératrice du vecteur \mathbf{N} :

$$G_{\mathbf{N}}(\mathbf{z}) = [P(\mathbf{1} - \mathbf{z})]^{-q} \quad \forall \mathbf{z} \in \Theta_{q,P}. \quad (1.76)$$

Le polynôme $P(\mathbf{1} - \mathbf{z})$ étant encore un polynôme affine, on obtient la fonction génératrice d'une loi multinomiale négative selon la définition 1.3.4.

Proposition 1.3.14. *Si la loi de mélange est la loi Gamma multivariée $\mathcal{G}(q, P)$, alors la LPMM obtenue est la loi multinomiale négative $\mathcal{NM}(q, P_1)$ où P_1 est le polynôme affine suivant :*

$$P_1(\mathbf{z}) = P(\mathbf{1} - \mathbf{z}). \quad (1.77)$$

Exemple 1.3.2 : *La LPMM générée par la famille des lois Gamma bivariée ($d = 2$).*

Dans ce cas le polynôme affine P est de la forme :

$$P(z_1, z_2) = 1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2,$$

avec $p_1, p_2 > 0$ et $0 \leq p_{12} \leq p_1 p_2$ selon le théorème 1.2.1. Soit $P_{a,b,c}$ la paramétrisation présentée à l'équation (1.65) du polynôme affine d'une loi multinomiale négative bivariée. Par identification avec le polynôme $P_1(\mathbf{z}) = P(\mathbf{1} - \mathbf{z})$, on obtient les relations suivantes :

$$\begin{aligned} a &= \frac{p_1 + p_{12}}{1 + p_1 + p_2 + p_{12}}, \\ b &= \frac{p_2 + p_{12}}{1 + p_1 + p_2 + p_{12}}, \\ c &= \frac{p_1 p_2 - p_{12}}{(1 + p_1 + p_2 + p_{12})^2}. \end{aligned} \quad (1.78)$$

Les paramètres (a, b, c) appartiennent donc à l'ensemble :

$$\Delta' = \{(a, b, c) \in]0, 1[^2 \times]0, 1[\mid (1-a)(1-b) - c > 0 \text{ et } ab \geq c\},$$

qui est strictement inclus dans l'ensemble Δ introduit à l'équation (1.68). Les LPMM générées par les lois Gamma bivariées sont donc une sous-famille des lois négatives multinomiales bivariées.

De plus, si le paramètre de forme q est fixé, la famille des lois Gamma bivariées est une FEN, éventuellement dégénérée si $p_{12} = 0$. D'après l'équation (1.70), les FEN engendrées par les lois multinomiales négatives bivariées sont les sous-familles obtenues lorsque les paramètres q et $\rho = \frac{c}{ab}$ sont fixés. Selon le système (1.78), cette situation correspond aux deux cas particuliers suivants :

1. $p_{12} = p_1 p_2 \iff \frac{c}{ab} = 0$, et les marges de la loi Gamma bivariée sont indépendantes,
2. $p_{12} = 0 \iff \frac{c}{ab} = 1$, et la loi Gamma bivariée est dégénérée et concentrée sur une droite.

Dans les deux cas, on retrouve bien la condition énoncée au théorème 1.3.13.

En inversant le système, on trouve l'expression suivante des paramètres de la LPMM en fonction des paramètres de la loi Gamma bivariée pour tout $(a, b, c) \in \Delta'_{a,b,c}$:

$$\begin{aligned} p_1 &= \frac{a(1-b) + c}{(1-a)(1-b) - c}, \\ p_2 &= \frac{b(1-a) + c}{(1-a)(1-b) - c}, \\ p_{12} &= \frac{ab - c}{(1-a)(1-b) - c}. \end{aligned} \tag{1.79}$$

Ceci montre que les paramétrisations selon les vecteurs $(a, b, c)^T$ et $(p_1, p_2, p_{12})^T$ sont équivalentes, puisqu'elles se déduisent l'une de l'autre par une transformation bijective.

Recalage et détection de changements pour des images RSO mono-capteur

Sommaire

2.1	Modélisation statistique des images RSO	48
2.1.1	Phénomène de speckle et loi Gamma	49
2.1.2	Mesures de similarité entre images RSO	52
2.2	Images RSO mono-capteur	56
2.2.1	Estimation par maximum de vraisemblance	57
2.2.2	Estimation par la méthode des moments	61
2.2.3	Estimation sur données synthétiques	63
2.2.4	Recalage pour des images réelles	64
2.2.5	Détection de changements	69
2.3	Conclusion	75

Cette étude a pour but de déterminer l'intérêt potentiel des lois gamma bivariées pour le recalage et la détection de changements dans les images obtenues à partir d'un radar à synthèse d'ouverture (radar RSO). L'intérêt de l'imagerie RSO réside dans le caractère actif du capteur. Les images peuvent ainsi être obtenues rapidement de nuit ou par temps nuageux par exemple contrairement aux images optiques. Cependant la cohérence de l'onde électromagnétique confère aux images RSO un aspect visuel très bruité. C'est pourquoi une modélisation statistique est nécessaire pour traiter ces images. On considère deux images d'une même scène notées I pour l'image de référence et J pour l'image secondaire.

Le recalage d'images peut être défini comme la recherche d'une transformation géométrique T qui maximise une certaine mesure de similarité entre l'image I et le résultat de la transformation $T \circ J$. L'élément clé du recalage d'images est alors l'estimation de la mesure de similarité entre les deux images. Cette estimation est généralement effectuée localement à l'aide d'une fenêtre entourant le pixel d'intérêt. Cette fenêtre doit être la plus petite possible pour prendre en compte les non stationnarités de l'image. Cependant, choisir une petite fenêtre pose un problème puisque l'estimation de la mesure de similarité ainsi obtenue n'est en général pas satisfaisante. Afin d'obtenir des estimateurs performants avec des tailles de fenêtres "petites", il est classique d'utiliser des informations a priori sur la statistique de l'image.

La *détection de changements temporels* correspond à la mise en évidence et à la localisation des changements de l'état d'une scène observée à partir de plusieurs images acquises à des dates différentes (séparées de plusieurs jours, mois ou années), avec des modalités et dans des conditions d'acquisition éventuellement différentes. Dans ce chapitre les images acquises correspondront cependant toujours à des images RSO issues de capteurs radars identiques. Le cas de capteurs différents sera étudié dans le prochain chapitre. Les changements à détecter peuvent alors être lents et de faible intensité comme l'évolution des assolements en agriculture, ou beaucoup plus rapides et spectaculaires lorsqu'il s'agit de catastrophes naturelles telle qu'une inondation ou un tremblement de terre. Considérons deux images I and J enregistrées à des instants différents t_I et t_J . La détection de changements consiste à déterminer une carte indiquant les changements qu'a subit l'image entre les dates t_I et t_J . Plus précisément, on veut construire une carte binaire correspondant aux deux classes "changement" et "pas de changement". Le problème peut en général se décomposer en deux parties :

1. génération d'une carte de changements,
2. seuillage de cette carte de changements pour produire une carte binaire indiquant les deux classes précédentes.

La qualité d'une détection de changements sera liée à la qualité de la carte de changements mais aussi à la qualité du seuillage. Dans cette étude, nous nous intéressons à la première étape consistant à déterminer une carte de changements. Cette carte peut être obtenue en estimant localement les mesures de similarités entre les deux images. Comme pour le recalage d'images, on utilise souvent une fenêtre glissante permettant d'estimer cette mesure de similarité avec une bonne résolution. Afin d'avoir une bonne estimation de cette mesure à partir de fenêtres de petites dimensions, il convient à nouveau d'utiliser des informations a priori sur l'image.

Ce chapitre se décompose en deux parties. La première partie présente tout d'abord les propriétés statistiques des images RSO d'intensité. Les mesures de similarité utilisées pour les applications de recalage d'image et de détection de changement sont ensuite introduites, et calculées dans le cas du modèle des lois Gamma bivariées.

La seconde partie traite des problèmes du recalage et de détection de changement pour des couples d'images RSO issues d'un même capteur radar, ou de capteurs radar ayant les mêmes caractéristiques. Sous cette hypothèse, le nombre de vues associée à chaque image est identique et la distribution statistique des couples composés d'un pixel appartenant à l'image de référence et du pixel correspondant dans l'image secondaire est une loi Gamma bivariée. Les estimateurs des paramètres des lois Gamma sont alors étudiés. Le problème du recalage d'images radars basé sur les lois gamma bivariées est ensuite traité. Il est finalement montré que les lois gamma bivariées permettent de mettre en oeuvre des stratégies performantes de détection de changements pour des images obtenues à partir d'un radar à synthèse d'ouverture.

2.1 Modélisation statistique des images RSO

Ce chapitre présente dans un premier temps les propriétés statistiques des images RSO. Sous certaines conditions qui seront supposées vérifiées par la suite, les intensités mesurées sur une zone homogène suivent une loi Gamma. Des mesures de similarités permettant de mesurer la "distance" entre deux images RSO sont ensuite introduites. Les mesures étudiées permettent de caractériser la dépendance entre un couple d'images RSO. Ces mesures sont étudiées lorsque la

loi jointe des intensités de l'image de référence et de l'image secondaire est modélisée par une loi Gamma bivariée.

2.1.1 Phénomène de speckle et loi Gamma

Le *speckle* encore appelé chatoiement, fouillis ou tavelure est un phénomène rencontré fréquemment en optique lors de l'éclairage d'une scène par une lumière cohérente. Ce phénomène intervient également en imagerie RSO et explique la forme de la distribution statistique des intensités mesurées dans ce type d'imagerie.

Contexte

Comme cela est expliqué par Goodman [Goo06], ce phénomène de speckle a reçu une attention toute particulière avec l'apparition des premiers lasers commerciaux au début des années 1960. En éclairant en effet un papier ou un mur avec de la lumière émise par un laser, i.e. une lumière cohérente et quasi-monochromatique, on voit apparaître des zones d'intensités fortement contrastées et des motifs d'aspect granuleux bien que l'éclairage de la scène soit relativement uniforme. Ce phénomène est illustré sur la figure 2.1 qui montre l'image d'un objet uniforme éclairé par un laser.

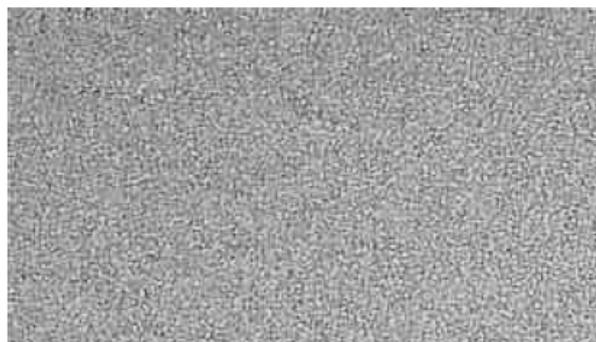


FIG. 2.1 – Image d'un objet uniforme éclairé par un laser [Goo06]

La cause physique de ce phénomène dû au caractère cohérent de l'éclairage laser a été rapidement identifiée : la quasi-totalité des matériaux, à l'exception des miroirs pour lesquels on parle de réflexion spéculaire, sont rugueux à l'échelle des longueurs d'onde optiques visibles. Par conséquent, ces matériaux se présentent sous l'aspect d'une multitude de facettes microscopiques orientées de façon spatialement aléatoire. Chaque facette déphase alors aléatoirement la lumière, et le champ lumineux résultant est la somme de ces contributions qui interfèrent chacune de manière constructive ou destructive avec une autre. Lorsque ces contributions sont constructives, l'intensité observée augmente et, inversement si les ondes réfléchies interfèrent de manière destructrice, alors l'intensité observée est faible. Ces interférences multiples entre les ondes réfléchies par le matériau rugueux à l'échelle de la longueur d'onde expliquent donc les zones fortement inhomogènes observées.

Pour les mêmes raisons que celles évoquées ci-dessus, le phénomène de speckle intervient dans tout autre domaine que l'optique où des champs électromagnétiques sont réfléchies par

des objets rugueux à l'échelle de la longueur d'onde de ce champ. C'est en particulier le cas de l'imagerie RSO où les ondes électromagnétiques émises sont des micro-ondes, ce qui correspond à des longueurs d'onde de l'ordre du centimètre ou du décimètre. La grande majorité des scènes observées sont donc rugueuses à cette échelle, mis à part quelques cas particuliers comme les autoroutes ou les lacs sans vague pour lesquels la réflexion est spéculaire. Le phénomène de speckle en imagerie RSO est illustré sur la figure 2.2

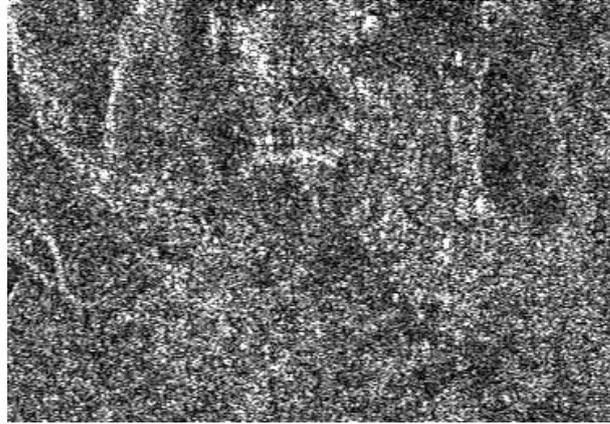


FIG. 2.2 – Image RSO de la région d'Izmit, Turquie (CNES)

Modélisation statistique du phénomène de speckle

Bien que le phénomène de speckle soit déterministe puisqu'il dépend uniquement de la configuration locale du matériau observé, il est impossible de le modéliser en dehors d'un cadre statistique. Afin de le modéliser, on considère une onde monochromatique rétrodiffusée par un objet dont la réflectivité est supposée uniforme. On suppose ici que cette onde est totalement polarisée. En omettant le facteur relatif à la fréquence ν de l'onde incidente, l'onde rétrodiffusée en un point et un instant donné s'écrit alors sous la forme suivante [Goo06] :

$$A = \sum_{n=1}^N a_n e^{j\phi_n}, \quad (2.1)$$

où N est le nombre de diffuseurs élémentaires pour la cellule de résolution considérée, j est nombre complexe tel que $j^2 = -1$ et a_n et ϕ_n sont les termes correspondants aux amplitudes et aux phases qui caractérisent chaque diffuseur. La contribution de chaque diffuseur élémentaire est modélisée comme un phaseur aléatoire et le champ électrique obtenu en est la somme. L'étude statistique du speckle est classiquement basée sur l'hypothèse [Goo06] du *speckle pleinement développé* qui correspond aux propriétés suivantes :

1. l'amplitude a_n et la phase ϕ_n d'un diffuseur donné sont indépendantes de celle des autres diffuseurs dans la cellule de résolution.
2. l'amplitude a_n et la phase ϕ_n d'un diffuseur sont indépendantes l'une de l'autre, pour tout $1 \leq n \leq N$.
3. les amplitudes a_n sont identiquement distribuées et admettent un moment d'ordre 1 et 2 finis (notés $\mathbf{E}[a]$ et $\mathbf{E}[a^2]$).

4. les phases ϕ_n sont uniformément distribuées sur $]-\pi, \pi[$, pour tout $1 \leq n \leq N$.
5. le nombre N de diffuseurs élémentaires dans une cellule de résolution est grand.

Selon ces hypothèses, l'expression (2.1) du champ électrique est la somme de vecteurs dans le plan complexe identiquement distribués et indépendants dont les longueurs et les directions sont aléatoires. Ce type de processus correspond à une *marche aléatoire*. En considérant les parties réelles et imaginaires du champ électrique (2.1) notées respectivement $A_{\mathfrak{R}}$ et $A_{\mathfrak{I}}$, on peut écrire :

$$A_{\mathfrak{R}} = \sum_{i=1}^N a_n \cos(\phi_n), \quad A_{\mathfrak{I}} = \sum_{i=1}^N a_n \sin(\phi_n).$$

L'hypothèse 4) implique les résultats suivants pour tout $1 \leq m, n \leq N$:

$$\begin{aligned} \mathbb{E}[\cos(\phi_n)] &= \mathbb{E}[\sin(\phi_n)] = 0, \\ \mathbb{E}[\cos(\phi_m) \sin(\phi_n)] &= 0, \\ \mathbb{E}[\cos(\phi_m) \cos(\phi_n)] &= \mathbb{E}[\sin(\phi_m) \sin(\phi_n)] = \frac{1}{2} \delta_{m,n}, \end{aligned}$$

où $\delta_{m,n}$ est le symbole de Kronecker qui vaut 1 si $m = n$ et 0 sinon. Par conséquent sous les hypothèses énoncées ci-dessus, les parties réelles et imaginaires vérifient :

$$\begin{aligned} \mathbb{E}[A_{\mathfrak{R}}] &= \mathbb{E}[A_{\mathfrak{I}}] = 0, \\ \text{var}(A_{\mathfrak{R}}) &= \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}[a_m a_n] \mathbb{E}[\cos(\phi_m) \cos(\phi_n)] = \frac{1}{2} \sum_{n=1}^N \mathbb{E}[a_n^2] = \frac{N}{2} \mathbb{E}[a^2], \\ \text{var}(A_{\mathfrak{I}}) &= \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}[a_m a_n] \mathbb{E}[\sin(\phi_m) \sin(\phi_n)] = \frac{1}{2} \sum_{n=1}^N \mathbb{E}[a_n^2] = \frac{N}{2} \mathbb{E}[a^2], \end{aligned}$$

et les parties réelle et imaginaire sont décorréliées :

$$\mathbb{E}[A_{\mathfrak{R}} A_{\mathfrak{I}}] = \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}[a_m a_n] \mathbb{E}[\cos(\phi_m) \sin(\phi_n)] = 0 = \mathbb{E}[A_{\mathfrak{R}}] \mathbb{E}[A_{\mathfrak{I}}].$$

De plus, les v.a. $A_{\mathfrak{R}}$ et $A_{\mathfrak{I}}$ sont obtenues comme la somme de N v.a. identiques et indépendantes. Le nombre de diffuseurs N étant grand, les conditions d'application du théorème centrale limite sont satisfaites. La loi du champ électrique $A = A_{\mathfrak{R}} + jA_{\mathfrak{I}}$ peut donc être approchée par une loi gaussienne complexe circulaire (i.e. que les parties réelles et imaginaires de cette loi sont indépendantes et suivent des lois normales centrées identiques) de variance $\sigma^2 = N\mathbb{E}[a^2]$ selon les propriétés démontrées précédemment, notée $\mathcal{N}_c(0, \sigma^2)$. On considère par la suite que la loi du champ électrique A est la loi gaussienne complexe circulaire $\mathcal{N}_c(0, \sigma^2)$.

Finalement, l'intensité du champ électrique, qui est définie comme $|A|^2 = A_{\mathfrak{R}}^2 + A_{\mathfrak{I}}^2$, est par construction distribuée suivant une la loi Gamma de paramètre de forme $q = 1$ et de paramètre d'échelle $R = \sigma^2 = N\mathbb{E}[a^2]$, notée $\mathcal{G}(1, R)$. Le paramètre R est appelé la réflectivité. Par ailleurs, afin d'améliorer l'interprétation visuelle ou automatique des images obtenues, une méthode classique pour réduire la variance est de moyenner pixel à pixel L vues de l'image qui sont supposées incohérentes entre elles et donc indépendantes. On obtient donc le résultat suivant :

Proposition 2.1.1. *En imagerie active cohérente et sous les hypothèses du speckle pleinement développé, les intensités rétrodiffusées suivent la loi Gamma $\mathcal{G}(L, R/L)$ où le paramètre de forme $L \in \mathbb{N}/\{0\}$ correspond au nombre de vues et le paramètre R , qui est la moyenne de la loi, correspond à la réflectivité de l'objet ou de la scène observé.*

En raison de ce résultat, l'intensité I mesurée est souvent représentée sous la forme

$$I = R S_I,$$

où R est la réflectivité (déterministe) de la scène observée en un point donné et S_I est un bruit multiplicatif de loi $\mathcal{G}(L, 1/L)$, et donc de moyenne 1, dû au phénomène de speckle. Dans cette représentation, la réflectivité représente l'image idéale (i.e. non bruitée) qui est corrompue par le bruit multiplicatif de speckle.

2.1.2 Mesures de similarité entre images RSO

La notion de mesure de similarité est une notion clé pour les problèmes qui nous intéressent. En effet, l'estimation de cette mesure entre les couples d'images RSO étudiées est l'étape préalable à partir de laquelle le recalage ou la détection de changement peuvent être effectués.

Mesures de similarité construites à partir de divergences

Beaucoup de mesures de similarité classiques peuvent être obtenues à partir du modèle des ϕ -divergences. C'est en particulier le cas de la mesure d'indépendance du χ^2 et de l'information mutuelle qui sont étudiées dans cette section.

La notion de ϕ -divergence introduite par Csiszár [Csi67] permet de mesurer une distance⁶ entre deux mesures de probabilité, notées $\alpha(dx)$ et $\beta(dx)$. Si la mesure $\alpha(dx)$ est absolument continue par rapport à $\beta(dx)$, on note alors $\frac{\alpha(x)}{\beta(y)}$ la dérivée de Radon-Nikodym obtenue et les ϕ -divergences sont définies de la manière suivante :

$$C_\phi(\alpha, \beta) = \int \phi\left(\frac{\alpha(x)}{\beta(x)}\right) \beta(dx),$$

où ϕ est une fonction strictement convexe sur \mathbb{R}^+ , deux fois dérivable sur un voisinage de 1 et telle que $\phi(1) = \phi'(1) = 0$ et $\phi''(1) > 0$. Une conséquence de ces conditions est que la fonction ϕ admet un minimum unique sur \mathbb{R}^+ atteint au point 1 et qui vaut 0. Par conséquent, la divergence $C_\phi(\alpha, \beta)$ est une grandeur positive qui est nulle si et seulement si la dérivée de Radon-Nikodym $\frac{\alpha(x)}{\beta(y)}$ est égal à 1 presque sûrement pour la mesure $\beta(dx)$, i.e. si et seulement si les lois sont identiques. Avec cette définition, les distances de Hellinger, de Toussaints, la distance L^1 , les divergences du χ^2 et de Kullback-Leibler peuvent notamment être obtenues. On suppose par la suite que les lois $\alpha(dx)$ et $\beta(dx)$ sont absolument continues pour la mesure de Lebesgue, de densités respectives $\alpha(x)$ et $\beta(x)$.

Les deux divergences qui nous intéressent sont les suivantes :

⁶les divergences obtenues ne sont cependant pas des distances au sens mathématique du terme : elles ne sont pas symétriques et l'inégalité triangulaire n'est pas nécessairement vérifiée

– La divergence dite du χ^2 qui est obtenue lorsque $\phi(x) = (x - 1)^2$. Il vient alors :

$$\begin{aligned}\chi^2(\alpha, \beta) &= \int \left(\frac{\alpha(x)}{\beta(x)} - 1 \right)^2 \beta(x) dx = \int \frac{(\alpha(x) - \beta(x))^2}{\beta(x)} dx, \\ &= \int \frac{\alpha(x)^2}{\beta(x)} dx - 1.\end{aligned}$$

– La divergence de Kullback-Leibler qui est obtenue lorsque $\phi(x) = x \log(x) - x + 1$. On obtient alors :

$$\begin{aligned}\text{KL}(\alpha, \beta) &= \int \left[\log \left(\frac{\alpha(x)}{\beta(x)} \right) \frac{\alpha(x)}{\beta(x)} - \frac{\alpha(x)}{\beta(x)} + 1 \right] \beta(x) dx, \\ &= \int \log \left(\frac{\alpha(x)}{\beta(x)} \right) \alpha(x) dx - \int \alpha(x) dx + \int \beta(x) dx, \\ &= \int \log \left(\frac{\alpha(x)}{\beta(x)} \right) \alpha(x) dx.\end{aligned}$$

On étudie maintenant le cas d'un couple de v.a. (X_1, X_2) sur \mathbb{R}^2 dont la loi admet une densité notée $f_{(X_1, X_2)}$. Soit f_{X_1} et f_{X_2} les densités marginales de ce vecteur. Alors pour une fonction ϕ donnée, une mesure de similarité pour le couple (X_1, X_2) peut-être obtenue comme :

$$I_\phi(X_1, X_2) = C_\phi \left(f_{(X_1, X_2)}, f_{X_1} f_{X_2} \right) = \int_{\mathbb{R}^2} \phi \left(\frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{X_1}(x_1) f_{X_2}(x_2)} \right) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 \quad (2.2)$$

Les mesures de similarité ainsi construites mesurent la dépendance entre les variables X_1 et X_2 . D'après les propriétés des ϕ -divergences, ces mesures sont positives, et sont nulles si et seulement si les variables X_1 et X_2 sont indépendantes. Une autre propriété importante de ces mesures, due à l'invariance par transformation bijective des phi-divergences, est donnée à la proposition suivante⁷ :

Proposition 2.1.2. *Les mesures de similarité obtenues à partir du modèle des ϕ -divergences sont par construction invariantes aux changements d'échelle : si $(Y_1, Y_2) = (X_1/\lambda_1, X_2/\lambda_2)$ où λ_1 et λ_2 sont des réels non nuls, alors*

$$I_\phi(Y_1, Y_2) = I_\phi(X_1, X_2)$$

pour toute fonction ϕ vérifiant les conditions de définition des ϕ -divergences.

Démonstration. Les densités respectives de Y_1, Y_2 et (Y_1, Y_2) se déduisent de celles de X_1, X_2 et (X_1, X_2) :

$$f_{Y_1}(y_1) = \lambda_1 f_{X_1}(\lambda_1 y_1), \quad f_{Y_2}(y_2) = \lambda_2 f_{X_2}(\lambda_2 y_2), \quad f_{(Y_1, Y_2)}(y_1, y_2) = \lambda_1 \lambda_2 f_{(X_1, X_2)}(\lambda_1 y_1, \lambda_2 y_2).$$

⁷On peut montrer un résultat plus général : si $(Y_1, Y_2) = (h_1(X_1), h_2(X_2))$ et si h_1 et h_2 sont des C^1 -difféomorphismes, alors la mesure de similarité associée à (Y_1, Y_2) est la même que celle associée à (X_1, X_2) (cette propriété n'est en général pas vraie pour une transformation du type $(Y_1, Y_2) = h(X_1, X_2)$, même si h est un C^1 -difféomorphisme). Cependant, seule l'invariance aux changements d'échelle nous intéresse par la suite.

Par conséquent

$$I_\phi(Y_1, Y_2) = \int_{\mathbb{R}^2} \phi\left(\frac{f_{(X_1, X_2)}(\lambda_1 y_1, \lambda_2 y_2)}{f_{X_1}(\lambda_1 y_1) f_{X_2}(\lambda_2 y_2)}\right) \lambda_1 f_{X_1}(\lambda_1 x_1) \lambda_2 f_{X_2}(\lambda_2 x_2) dy_1 dy_2.$$

En effectuant le changement de variable $(y_1, y_2) = (\lambda_1 x_1, \lambda_2 x_2)$ dans l'intégrale, on retrouve l'expression (2.2) \square

Les familles de mesures de similarité obtenues à partir du modèle des ϕ -divergences ont été étudiées de manière approfondie pour les problèmes de recalage d'images dans le domaine de l'imagerie médicale [SM99] mais aussi dans le domaine de l'imagerie RSO [IG04]. Ces familles de mesures de similarité sont a priori plus informatives que le coefficient de corrélation, qui est également une mesure de similarité, où seuls les moments d'ordre 2 sont pris en considération.

La mesure de similarité obtenue pour la divergence du χ^2 , appelée *mesure d'indépendance du χ^2* et notée I_{χ^2} , et la mesure de similarité obtenue pour la divergence de Kullback-Leibler, qui correspond à la notion bien connue d'*information mutuelle* (notée IM), ont finalement pour expressions respectives :

$$\begin{aligned} I_{\chi^2}(X_1, X_2) &= \int \frac{f_{(X_1, X_2)}^2(x_1, x_2)}{f_{X_1}(x_1) f_{X_2}(x_2)} dx_1 dx_2 - 1, \\ \text{IM}(X_1, X_2) &= \int \log\left(\frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{X_1}(x_1) f_{X_2}(x_2)}\right) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2. \end{aligned} \tag{2.3}$$

Les deux mesures de similarité ci-dessus sont retenues et étudiées pour les raisons suivantes :

- L'information mutuelle est une mesure très populaire en imagerie satellitaire, qui a une interprétation probabiliste claire à partir de l'entropie associée aux images étudiées et qui a démontré de bonnes performances pour les problèmes de recalage d'image ou de détection de changement [Ing02, CVA03, IG04].
- La mesure d'indépendance du χ^2 est une mesure classique, dont le comportement et les performances sont proches de ceux de l'information mutuelle [SM99], et qui présente l'avantage d'avoir une expression analytique très simple pour le modèle des lois Gamma bivariées.

Application au modèle de loi Gamma bivariée

Les lois Gamma bivariées présentées au chapitre précédent peuvent être paramétrées par leur paramètre de forme $q > 0$, les moyennes des marges $m_1 > 0$ et $m_2 > 0$ et le coefficient de corrélation entre les marges $r \in [0, 1[$. Une conséquence immédiate de la proposition 2.1.2 est que les mesures de similarité obtenues pour les ϕ -divergences et appliquées au modèle des lois Gamma bivariées ne dépendent pas des paramètres d'échelle m_1 et m_2 . La mesure d'indépendance du χ^2 et l'information mutuelle ne dépendent donc que des paramètres q et r . Les développements suivants montrent par ailleurs que les deux mesures de similarités obtenues sont des fonctions bijectives du paramètre r , du moins lorsque q est fixé. Par conséquent, lorsque le nombre de vues associé aux images RSO est connu ce qui sera le cas par la suite, i.e. lorsque le paramètre de forme est fixé selon la proposition 2.1.1, la mesure de similarité définie par le coefficient de corrélation est équivalente à l'information mutuelle ou à la mesure d'indépendance du χ^2 . Certains avantages de l'information mutuelle et de la mesure d'indépendance du χ^2 seront cependant discutés dans les sections suivantes de ce chapitre et du chapitre suivant.

Mesure d'indépendance du χ^2 L'expression de la mesure d'indépendance du χ^2 est directement obtenue à partir de la représentation sous la forme de probabilité de Lancaster de la loi Gamma bivariée. Selon le résultat donné à la section 1.2.2, on obtient en effet :

$$I_{\chi^2}(r) = \frac{r^2}{1-r^2}, \quad \forall r \in [0, 1[, \quad q > 0. \quad (2.4)$$

Cette mesure ne dépend pas du paramètre de forme q et est une fonction croissante de r .

Information Mutuelle Afin de calculer l'information mutuelle, on peut supposer que les paramètres d'échelles m_1 et m_2 sont égaux au paramètre de forme q , puisque l'information mutuelle ne dépend pas de m_1 et m_2 . On obtient finalement à partir de l'expression (1.37) de la densité que

$$\begin{aligned} \text{IM}_q(r) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) \left[-q \log(1-r) - \frac{r}{1-r} (x_1 + x_2) + \log[\Gamma(q) f_q(cx_1 x_2)] \right] dx_1 dx_2, \\ &= -q \log(1-r) - \frac{r}{1-r} (\mathbb{E}[X_1] + \mathbb{E}[X_2]) + \mathbb{E} \left[\log(\Gamma(q) f_q(cX_1 X_2)) \right], \\ &= -q \log(1-r) - 2 \frac{qr}{1-r} + \mathbb{E} \left[\log(\Gamma(q) f_q(cX_1 X_2)) \right], \end{aligned} \quad (2.5)$$

où $c = r/(1-r)^2$. Comme l'application $r \mapsto f_q(cx_1 x_2)$ est strictement croissante par composition d'applications croissantes, on obtient finalement que l'information mutuelle est une fonction de r strictement croissante. Malheureusement, il n'est pas facile de déterminer une expression analytique simple de

$$A = \mathbb{E} \left[\log(\Gamma(q) f_q(cx_1 x_2)) \right].$$

On devra donc utiliser une procédure d'intégration numérique pour calculer ce dernier terme. La valeur de l'information mutuelle peut alors être tabulée pour toutes les valeurs possibles de q et de $r \in [0, 1]$. Des approximations de l'information mutuelle obtenues pour $r \rightarrow 0$ et $r \rightarrow 1$ peuvent également être déterminées et permettent d'éviter la procédure d'intégration numérique requise pour calculer l'information mutuelle au voisinage de $r = 0$ et $r = 1$.

– $r \rightarrow 0$: Le développement de Taylor de $f_q(z)$ au second ordre autour de $z = 0$ s'écrit

$$f_q(z) = 1 + \frac{z}{q\Gamma(q)} + \frac{z^2}{2q(q+1)\Gamma(q)} + o(z^2),$$

où $o(z^2)/z^2$ tend vers 0, quand $z \rightarrow 0$, d'où l'approximation :

$$A = r(1+2r) \frac{\mathbb{E}[X_1 X_2]}{q} - \frac{r^2 \mathbb{E}[(X_1 X_2)^2]}{2q} \left(\frac{1}{q(q+1)} \right) + o(r^2).$$

En utilisant l'expression (1.19) des moments joints de la loi Gamma bivariée, l'information mutuelle IM_q peut finalement être approchée grâce à l'expression suivante :

$$\text{IM}_q \approx \frac{r^2}{2}. \quad (2.6)$$

– $r \rightarrow 1$: Le développement asymptotique de $f_q(z)$ en $+\infty$ s'écrit

$$f_q(z) = \frac{\exp(2\sqrt{z})}{\sqrt{4\pi z} q^{-1/2}} (1 + o(1)), \quad (2.7)$$

où $o(1)$ tend vers 0, lorsque $z \rightarrow \infty$, d'où le résultat suivant :

$$A = \mathbb{E} \left[\log \left[\frac{\exp(2\sqrt{cX_1X_2})}{\sqrt{4\pi(cX_1X_2)^{q-1/2}}} \right] \right],$$

$$= 2\sqrt{c} \mathbb{E} \left[\sqrt{X_1X_2} \right] - \left(\frac{q}{2} - 1 \right) (\log c + \mathbb{E} [\log X_1 + \log X_2]) + \log \left[\frac{\Gamma(q)}{2\sqrt{\pi}} \right] + o(1).$$

Après avoir remplacé les moyennes de $\sqrt{X_1X_2}$, $\log(X_1)$ et de $\log(X_2)$ calculées à l'annexe C.1.4, on obtient les résultats suivants

$$\text{IM}_q \approx -\frac{1}{2} \log(1-r) + \left(q - \frac{1}{2} \right) + \log \left[\frac{\Gamma(q)}{2\sqrt{\pi}} \right] - \left(q - \frac{1}{2} \right) \psi(q). \quad (2.8)$$

La figure 2.3 montre que l'information mutuelle M_q peut être approchée avec précision par (2.6) et (2.8) dès que $r < 0.5$ et $r > 0.9$. Cette figure a été obtenue pour le paramètre de forme $q = 1$.

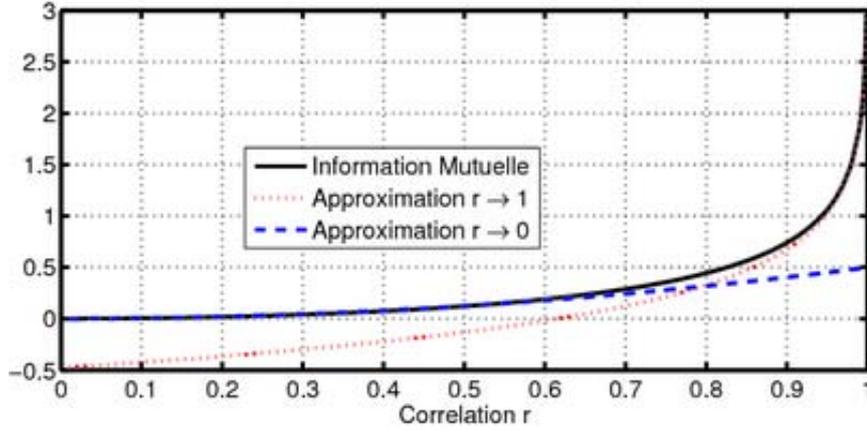


FIG. 2.3 – Information mutuelle et ses approximations lorsque $r \rightarrow 0$ et $r \rightarrow 1$ ($q = 1$).

2.2 Images RSO mono-capteur

Il a été vu à la section 2.1.1 que sous les hypothèses de speckle pleinement développé, qui sont supposées vérifiées dans notre étude, la distribution des intensités d'une image RSO suit une loi Gamma. Il est alors naturel de modéliser la distribution du vecteur formé par les intensités d'un pixel acquises à différentes dates par une loi Gamma multivariée. Ce modèle permet ainsi de prendre en compte les dépendances temporelles entre les différentes images. Bien que l'étude soit généralisable en dimension quelconque, i.e. pour un nombre quelconques d'images de la même scène, on se limite dans cette partie à l'étude d'un couple d'images RSO formé par une image de référence et une image secondaire. Dans cette section, les nombre de vues associées à ces deux images sont supposés être identiques, ce qui est par exemple le cas si les images ont été acquises par le même capteur. Pour cette raison les images RSO étudiées sont qualifiées de *mono-capteur*. Ce nombre de vue étant par ailleurs une caractéristique du capteur, on le suppose connu dans les applications.

L'intérêt de l'approche proposée est de prendre en compte des contraintes strictes sur la forme des distributions des lois marginales. Elle se distingue en ceci de méthodes non paramétriques où les densités jointes sont estimées par exemple grâce aux histogrammes empiriques [IG04], ou à partir des cumulants empiriques [IM07]. Ces dernières méthodes sont plus souples car elles peuvent être appliquées à d'autres types d'images (par exemple un couple d'images RSO et optique), ou lorsque les hypothèses de speckle pleinement développé ne sont plus vérifiées. Cependant, le modèle proposé permet d'obtenir une idée des performances maximales que l'on peut espérer lorsque les hypothèses de speckle pleinement développé sont valides et, peut ainsi servir de référence pour évaluer les performances des méthodes précédentes.

2.2.1 Estimation par maximum de vraisemblance

Dans cette section et la section suivante, des méthodes d'estimation des paramètres des lois gamma bivariées sont étudiées. Plus précisément, on cherche à estimer le vecteur des paramètres inconnus les moyennes et le coefficient de corrélation entre les marges, i.e. le vecteur $\boldsymbol{\theta} = (m_1, m_2, r)$, à partir de n vecteurs indépendants $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^n)$, où $\mathbf{x}^i = (x_1^i, x_2^i)$ est distribué suivant une loi gamma bivariée de vecteur paramètre $\boldsymbol{\theta}$. Notons que le paramètre q est supposé connu dans cette étude même si les travaux présentés permettraient d'estimer ce paramètre.

Principe

La méthode du maximum de vraisemblance peut être appliquée aux lois gamma bivariées car une expression analytique de la densité est disponible⁸ (1.37). Il a été montré à la section 1.2.2 que lorsque les paramètres q et $c = \frac{rq^2}{(1-r)^2 m_1 m_2}$ sont fixés, alors la famille de loi obtenue est une famille exponentielle naturelle. Dans ce cas, les estimateurs du maximum de vraisemblance des moyennes sont les moyennes empiriques selon le théorème B.6.3 :

$$\widehat{m}_{1\text{ML}} = \bar{x}_1, \quad \widehat{m}_{2\text{ML}} = \bar{x}_2. \quad (2.9)$$

Si maintenant le paramètre $c \geq 0$ est inconnu, la log-vraisemblance exprimée en fonction du vecteur (m_1, m_2, c) admet nécessairement son maximum sur la demi-droite de \mathbb{R}^3 d'équation paramétrique $(\widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, c)$ avec $c \geq 0$ selon le résultat précédent. Ceci démontre que même lorsque le paramètre $c \geq 0$ est inconnu les estimateurs du maximum de vraisemblance des moyennes sont définis et sont ceux présentées à l'équation (2.9) ci-dessus. Les estimateurs du maximum de vraisemblance du vecteur $\boldsymbol{\theta} = (m_1, m_2, r)$ se déduisent finalement de ceux du vecteur (m_1, m_2, c) d'après le principe d'invariance fonctionnelle car il existe une relation bijective entre ces vecteurs quel que soit $\boldsymbol{\theta} \in]0, \infty[^2 \times]0, 1[$. En conséquence, les estimateurs du maximum de vraisemblance des moyennes sont les moyennes empiriques même lorsque le paramètre $r \in]0, 1[$ est inconnu⁹.

Le calcul de l'estimateur du maximum de vraisemblance du paramètre r , noté \widehat{r}_{ML} , nécessite alors de maximiser la fonction $r \in]0, 1[\mapsto l(\mathbf{x}; (\widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, r))$, où la fonction $l(\mathbf{x}; \boldsymbol{\theta})$ est la log-vraisemblance. Après avoir éliminé les termes qui ne dépendent pas de $\boldsymbol{\theta}$, le logarithme de la

⁸Le problème est plus compliqué dans le cas général $d > 2$ puisqu'il n'y a pas en général d'expression simple de la densité d'une loi gamma multivariée. Dans ce cas, les moyennes des marges et les coefficients de corrélation entre les marges peuvent néanmoins être estimés efficacement à l'aide d'une méthode de maximum de vraisemblance composite (voir la section 4.3.2 pour plus de détails sur cette méthode).

⁹On peut retrouver ce résultat en cherchant les conditions pour lesquelles le gradient de la log-vraisemblance par rapport au vecteur $(m_1, m_2, r)^T$ est nul.

fonction de vraisemblance peut s'écrire comme suit :

$$l(\mathbf{x}; \boldsymbol{\theta}) = -nq \log(m_1 m_2) - \sum_{j=1}^2 \frac{nq \bar{x}_j}{m_j(1-r)} - nq \log(1-r) + \sum_{i=1}^n \log f_q(c x_1^i x_2^i), \quad (2.10)$$

où $c = \frac{rq^2}{m_1 m_2 (1-r)^2}$, et $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_j^i$ est la moyenne arithmétique de x_j pour $j = 1, 2$. L'estimateur du maximum de vraisemblance dénoté \widehat{r}_{ML} est alors obtenu en minimisant la fonction $f(r) = -l(\mathbf{x}; (\widehat{m}_{1ML}, \widehat{m}_{2ML}, r))$ définie sur $[0, 1[$. Après avoir éliminé les termes qui ne dépendent pas de r , on obtient :

$$2 \frac{nq}{(1-r)} - nq \log(1-r) + \sum_{i=1}^n \log f_q(\widehat{c} x_1^i x_2^i), \quad (2.11)$$

où

$$\widehat{c} = \frac{r}{(1-r)^2} \frac{q^2}{\bar{x}_1 \bar{x}_2}.$$

En dérivant maintenant la fonction $r \in [0, 1[\mapsto f(r)$, on montre (voir le calcul effectué à l'annexe [D.1.1](#)) que si $\widehat{r}_{ML} \in]0, 1[$ alors \widehat{r}_{ML} est une racine de la fonction suivante :

$$g(r) = r - 1 + \frac{q}{n \bar{x}_1 \bar{x}_2} \left(\sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(\widehat{c} x_1^i x_2^i)}{f_q(\widehat{c} x_1^i x_2^i)} \right). \quad (2.12)$$

Malheureusement, il n'est pas possible de trouver une expression analytique explicite de \widehat{r}_{ML} .

Existence

L'estimateur du maximum de vraisemblance existe si la log-vraisemblance admet un maximum unique sur l'ensemble de définition des paramètres. D'après la section précédente, la difficulté est de montrer que l'estimateur de \widehat{r}_{ML} existe quel que soit les paramètres inconnus $\boldsymbol{\theta} \in]0, \infty[^2 \times]0, 1[$. Considérons le cas d'une loi *Gamma bivariée standard*, i.e. dont la transformée de Laplace est de la forme :

$$\mathcal{L}(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^{-q},$$

avec $p_1 = p_2 = p_{12} > 0$. Selon l'expression des moments d'une loi Gamma bivariée donnée à l'équation (1.23), on trouve dans ce cas que $p_1 = p_2 = p_{12} = \frac{1}{1-r}$ où $r \in [0, 1[$ est le coefficient de corrélation entre les marges. Les moyennes de la loi Gamma bivariée sont alors $m_1 = m_2 = \frac{q}{1-r}$. Ce modèle est donc paramétré par le coefficient de corrélation $r \in [0, 1[$.

Théorème 2.2.1. *L'estimateur du maximum de vraisemblance du paramètre r , noté \widehat{r}_{ML} , d'une loi Gamma bivariée standard est bien définie pour tout $r \in [0, 1[$. En outre $\widehat{r}_{ML} = 0$ si et seulement si $\frac{1}{n} \sum_{i \geq 1}^n x_1^i x_2^i \leq q^2$. Sinon \widehat{r}_{ML} est l'unique racine sur $]0, 1[$ de la fonction score suivante :*

$$g(r) = \frac{\partial l(\mathbf{x}; r)}{\partial r} = \frac{-q}{1-r} - \frac{1}{n} \sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(r x_1^i x_2^i)}{f_q(r x_1^i x_2^i)}.$$

Démonstration. La preuve repose sur la concavité de la log-vraisemblance d'une loi Gamma bivariée standard :

$$l(\mathbf{x}; r) = nq \log(1 - r) + \sum_{i=1}^n \log f_q(rx_1^i x_2^i).$$

Il est démontré dans l'annexe D.1.2 que la fonction $z \in \mathbb{R}^+ \mapsto \log f_q(z)$ est strictement concave. Comme la fonction $h : x \mapsto nq \log(1 - r)$ est aussi strictement concave sur $[0, 1[$ (on montre en effet que $h''(x) = -nq/(1 - r)^2 < 0$), la fonction $r \in [0, 1[\mapsto l(\mathbf{x}; r)$ est, en tant que somme de fonctions strictement concave, également strictement concave. Ceci démontre que la fonction score g est une fonction strictement décroissante sur $[0, 1[$ telle que

$$g(0) = -q + \frac{1}{nq} \sum_{i=1}^n x_1^i x_2^i.$$

Suivant la valeur de $g(0)$, deux cas se présentent :

- $g(0) \leq 0$: dans ce cas la log-vraisemblance est strictement décroissante sur $[0, 1[$ et le maximum est atteint au point $\widehat{r}_{\text{ML}} = 0$.
- $g(0) > 0$: la fonction $g(r) \rightarrow -\infty$ lorsque $r \rightarrow 1$. Cette fonction, décroissante et continue, est donc bijective de $[0, 1[$ vers $[g(0), -\infty[$. Par conséquent il existe une unique racine de g sur $[0, 1[$ qui est l'estimateur du maximum de vraisemblance de r .

□

Malheureusement, la propriété de concavité de la log-vraisemblance par rapport à r n'est plus vérifiée dans le cas général. On peut en effet montrer que la fonction $r \mapsto l(\mathbf{x}; (\widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, r))$, dont une expression est donnée (au signe près) à l'équation (2.11), n'est pas forcément concave sur $[0, 1[$. Plusieurs réalisations de l'opposée de cette log-vraisemblance à minimiser en fonction de r sont représentées aux figures 2.4 et 2.5. Ces réalisations sont obtenues pour des échantillons de taille $n = 500$ distribués suivant une loi Gamma bivariée, de paramètre de forme $q = 2$, de moyennes $m_1 = m_2 = 1$ et pour différentes valeurs du coefficient de corrélation $r \in [0, 1[$. Pour de grandes valeurs du coefficient de corrélation telle que $r = 0.98$, on constate que le critère à minimiser n'est pas concave. Cependant au vu de la forme de ce critère, il existe un minimum unique situé autour de la valeur théorique et il est clair que tout algorithme de gradient convergera vers cet unique minimum.

Performance

Les propriétés des estimateurs du maximum de vraisemblance de $\widehat{m}_{1\text{ML}}$ et $\widehat{m}_{2\text{ML}}$ peuvent s'obtenir facilement. Ces estimateurs sont en effet de manière évidente non biaisés et convergents. D'après le théorème B.6.3 ces estimateurs sont efficaces lorsque le paramètre $c = \frac{rq^2}{(1-r)^2 m_1 m_2} \geq 0$ est fixé car dans ce cas les lois obtenues appartiennent à une FEN. Mais ces estimateurs ne dépendent pas du paramètre c , et sont par conséquent efficaces lorsque le paramètre $c \geq 0$ est inconnu, ou de manière équivalente lorsque le paramètre $r \in [0, 1[$ est inconnu¹⁰.

L'étude des performances de l'estimateur \widehat{r}_{ML} est plus difficile à mener. Bien sûr, on sait que l'estimateur du maximum de vraisemblance est asymptotiquement non biaisé, convergent

¹⁰On peut retrouver ce résultat grâce aux propriétés d'efficacité asymptotique des EMV. En effet la variance de chacun des estimateurs $\widehat{m}_{1\text{ML}}$ et $\widehat{m}_{2\text{ML}}$ est de la forme K_i/n avec $K_i = m_i^2/q$ pour $i \in \{1, 2\}$. Par conséquent cette variance est égale à la borne de Cramer-Rao de chacun des paramètres.

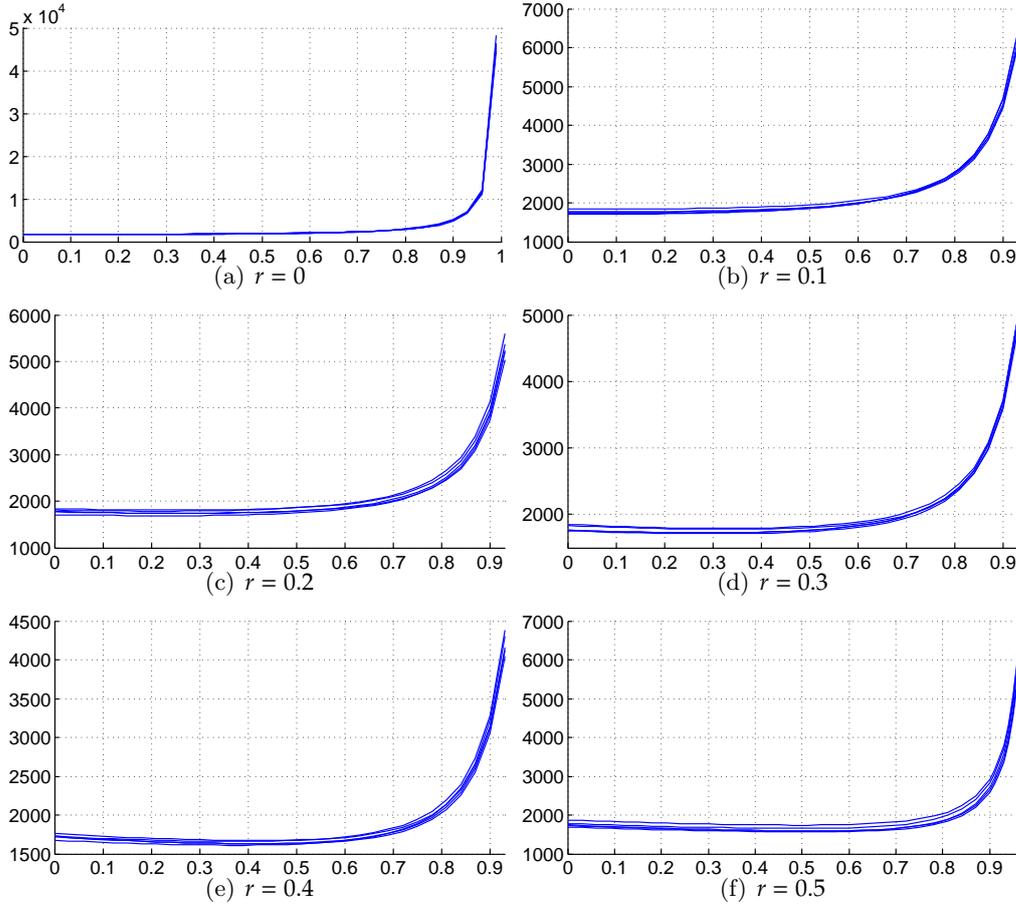


FIG. 2.4 – Courbes typiques de l’opposée de la log-vraisemblance $f(r) = -l(\mathbf{x}; \widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, r)$ en fonction de r ($q = 2$, $n = 500$).

et asymptotiquement efficace moyennant certaines hypothèses de régularité qui sont vérifiées dans ce contexte. En conséquence, l’erreur quadratique moyenne de l’estimateur \widehat{r}_{ML} peut être approchée pour N grand par la borne de Cramer-Rao associée à ce paramètre.

Pour des estimateurs non biaisés, les bornes de Cramer-Rao des paramètres m_1 , m_2 et r s’obtiennent en inversant la matrice d’information de Fisher dont les éléments sont :

$$[\mathcal{I}(\boldsymbol{\theta})]_{ij} = -\mathbb{E} \left[\frac{\partial^2 \ln f_{2D}(\mathbf{X})}{\partial \theta_i \partial \theta_j} \right], \quad (2.13)$$

où $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T = (m_1, m_2, r)^T$ et f_{2D} est la densité de la loi Gamma bivariée de paramètres q (fixé) et $(m_1, m_2, r)^T$. Le calcul de ces espérances est difficile analytiquement à cause du terme $\log f_q$ qui apparaît dans la log-vraisemblance. Dans de telles situations, il est habituel d’approcher ces espérances en moyennant plusieurs estimées obtenus à l’aide de simulations de Monte-Carlo. Plus précisément, cette approche consiste à approcher l’information de Fisher $\mathcal{I}(\boldsymbol{\theta})$ comme suit

$$[\mathcal{I}(\boldsymbol{\theta})]_{ij} \approx -\frac{1}{N} \sum_{k=1}^N \frac{\partial^2 \ln f_{2D}(\mathbf{x}^k)}{\partial \theta_i \partial \theta_j}, \quad (2.14)$$

où \mathbf{x}^k est la réalisation d’un vecteur aléatoire distribué suivant une loi gamma bivariée de densité

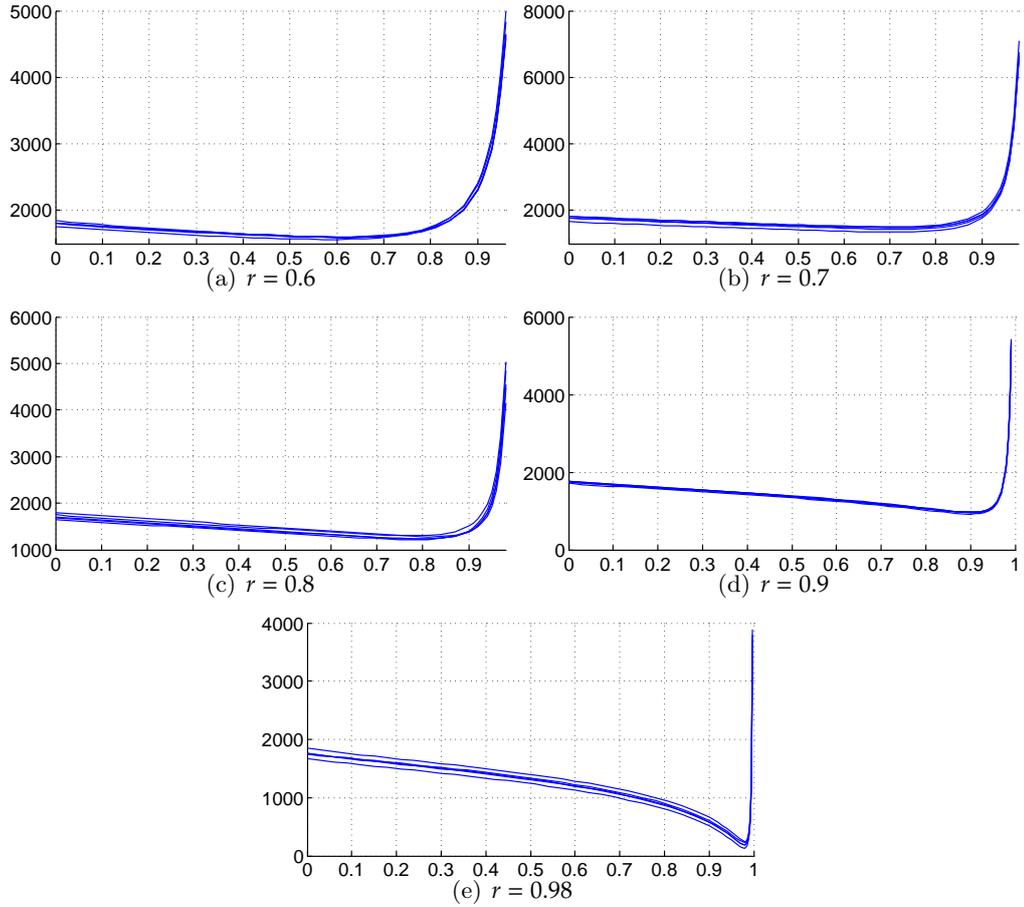


FIG. 2.5 – Courbes typiques de l’opposée de la log-vraisemblance $-l(\mathbf{x}; (\widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, r))$ en fonction de r ($q = 2$, $n = 500$).

$f_{2D}(\mathbf{x})$ et N est le nombre de simulations de Monte Carlo.

2.2.2 Estimation par la méthode des moments

Principe

Cette section rappelle succinctement le principe de la méthode des moments. Considérons une fonction $\mathbf{h}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^L$ et la statistique s_n de taille L définie comme suit :

$$\mathbf{s}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{h}(x^i), \quad (2.15)$$

où $\mathbf{h}(\cdot)$ est habituellement choisi de sorte que s_n soit constituée de moments empiriques d’ordre entier. Soit

$$\mathbf{f}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{s}_n] = \mathbb{E}[\mathbf{h}(x^1)]. \quad (2.16)$$

Un estimateur des moments de $\boldsymbol{\theta}$ est construit comme suit :

$$\widehat{\boldsymbol{\theta}}_{\text{Mo}} = \mathbf{g}(\mathbf{s}_n), \quad (2.17)$$

où $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Par exemple, en considérant la fonction

$$\mathbf{h}(\mathbf{x}) = (x_1, x_2, x_1^2, x_2^2, x_1x_2),$$

on obtient le résultat suivant :

$$\mathbf{f}(\boldsymbol{\theta}) = [m_1, m_2, m_1^2(1 + q^{-1}), m_2^2(1 + q^{-1}), m_1m_2(1 + rq^{-1})]. \quad (2.18)$$

Les paramètres inconnus (m_1, m_2, r) peuvent s'exprimer en fonction de $\mathbf{f}(\boldsymbol{\theta}) = (f_1, f_2, f_3, f_4, f_5)$. Par exemple, les relations suivantes sont vérifiées

$$m_1 = f_1, \quad m_2 = f_2, \quad r = \frac{f_5 - f_1f_2}{\sqrt{(f_3 - f_1^2)(f_4 - f_2^2)}}, \quad (2.19)$$

ce qui donne les estimateurs

$$\widehat{m}_{1\text{Mo}} = \bar{x}_1, \quad \widehat{m}_{2\text{Mo}} = \bar{x}_2, \quad \widehat{r}_{\text{Mo}} = \frac{\sum_{i=1}^n (x_1^i - \bar{x}_1)(x_2^i - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_1^i - \bar{x}_1)^2} \sqrt{\sum_{i=1}^n (x_2^i - \bar{x}_2)^2}}. \quad (2.20)$$

On retrouve dans ce cas l'expression des moyennes empiriques et du coefficient de corrélation empirique. Les estimateurs des moyennes sont donc les mêmes que ceux obtenus par maximum de vraisemblance, et seul l'estimateur \widehat{r}_{Mo} du coefficient de corrélation est différent de l'EMV \widehat{r}_{ML} . On remarque que l'estimateur \widehat{r}_{Mo} défini par 2.20 est l'estimateur habituel du coefficient de corrélation. Il est bien sûr possible de construire par la méthode des moments d'autres estimateurs qui utilisent la forme spécifique des moments d'une loi Gamma bivariée. Cependant les simulations ont montré que les performances obtenues notamment sur des données réelles sont sensiblement équivalentes. Dans la suite, on n'étudie donc que l'estimateur des moments \widehat{r}_{Mo} . Cet estimateur a en effet le mérite d'être convergent même pour des lois qui n'appartiennent pas à la famille des lois Gamma bivariées.

Performance

Les performances asymptotiques de l'estimateur $\widehat{\boldsymbol{\theta}}_{\text{Mo}}$ peuvent être obtenues en imitant la démarche proposée dans [PF89]. Un point clé des résultats obtenus dans cet article est l'hypothèse $\mathbf{s}_n \xrightarrow{a.s.} \mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ qui est vérifiée ici en appliquant la loi forte des grands nombres à (2.15). En conséquence, l'erreur quadratique moyenne asymptotique de $\widehat{\boldsymbol{\theta}}_{\text{Mo}}$ peut se calculer comme suit :

$$\lim_{n \rightarrow \infty} n\mathbf{E} [(\widehat{\boldsymbol{\theta}}_{\text{Mo}} - \boldsymbol{\theta})^2] = \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta})^t, \quad (2.21)$$

où $\mathbf{G}(\boldsymbol{\theta})$ est la matrice jacobienne du vecteur $\mathbf{g}(\cdot)$ au point $\mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ et

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} n\mathbf{E} [(s_n - \mathbf{s})(s_n - \mathbf{s})^T]. \quad (2.22)$$

Dans l'exemple précédent, d'après les relations (2.19), $\mathbf{g} : \mathbb{R}^5 \rightarrow \mathbb{R}^3$ est défini comme suit :

$$\mathbf{g}(\mathbf{x}) = \left(x_1, x_2, \frac{x_5 - x_1x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}} \right). \quad (2.23)$$

Les dérivées partielles de g_1 et g_2 par rapport à $x_i, i = 1, \dots, 5$ sont triviales. En notant $\gamma = \sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}$, celles de g_3 sont

$$\begin{aligned}\frac{\partial g_3}{\partial x_1} &= -\frac{x_2}{\gamma} + \frac{x_1(x_4 - x_2^2)(x_5 - x_1x_2)}{\gamma^3}, \\ \frac{\partial g_3}{\partial x_2} &= -\frac{x_1}{\gamma} + \frac{x_2(x_3 - x_1^2)(x_5 - x_1x_2)}{\gamma^3}, \\ \frac{\partial g_3}{\partial x_3} &= \frac{(x_1x_2 - x_5)(x_4 - x_2^2)}{2\gamma^3}, \\ \frac{\partial g_3}{\partial x_4} &= \frac{(x_1x_2 - x_5)(x_3 - x_1^2)}{2\gamma^3}, \\ \frac{\partial g_3}{\partial x_5} &= \frac{1}{\gamma}.\end{aligned}\tag{2.24}$$

Les éléments de $\Sigma(\boldsymbol{\theta})$ peuvent se calculer à l'aide de l'expression (1.19) et des moments d'une loi Gamma bivariée $\mathbf{h}(\mathbf{x})$. Ceci permet d'obtenir une expression numérique des erreurs quadratiques moyennes asymptotiques (2.21) pour chaque valeur du paramètre $\boldsymbol{\theta}$.

2.2.3 Estimation sur données synthétiques

On s'intéresse ici aux performances des estimateurs des lois Gamma bivariées obtenus à la section précédente. Afin d'obtenir des données synthétiques, les lois Gamma bivariées sont simulées d'après l'algorithme 1 présenté à la section 1.2.2. Les premières simulations comparent les performances de la méthode des moments avec la méthode du maximum de vraisemblance en fonction de n . Notons que les valeurs possibles de n sont $n = (2p + 1) \times (2p + 1)$, où $p \in \mathbb{N}$ (i. e. $3 \times 3 = 9$, $5 \times 5 = 25$, ..., $25 \times 25 = 625$). Ces valeurs correspondent à la taille des fenêtres qui seront utilisées pour le recalage, comme nous l'expliquerons dans le paragraphe suivant. Le nombre de simulations de Monte Carlo est 10000 pour toutes les figures présentées dans ce rapport. Les autres paramètres pour ce premier exemple sont $m_1 = 100$, $m_2 = 200$ et $q = 1$ (images monovue). Seuls les estimateurs du coefficient de corrélation r sont comparés. En effet les estimateurs des moyennes sont identiques pour les deux méthodes. De plus, le paramètre q étant connu, seul l'estimateur de r nous intéresse car, comme cela a été expliqué à la section 2.1.2, ce paramètre est la mesure de similarité retenue pour les applications (et autres les mesures de similarité étudiées se déduisent de ce paramètre). Les figures 2.6 et 2.7 représentent le logarithme décimal des erreurs quadratiques moyennes (EQM) de l'estimateur du coefficient de corrélation en fonction du logarithme décimal de la taille n de l'échantillon, pour deux valeurs de r ($r = 0.2$ et $r = 0.8$).

Les courbes \circ correspondent à l'estimateur des moments alors que les courbes Δ correspondent à l'estimateur du maximum de vraisemblance. Ces figures montrent que l'estimateur du maximum de vraisemblance est plus performant que l'estimateur des moments pour les lois gamma bivariées. Les figures montrent également que la différence est plus accentuée pour de fortes valeurs de r , et que l'estimation est d'autant meilleure que le paramètre r est proche de 1. Ce dernier résultat était prévisible pour l'EMV \widehat{r}_{ML} d'après l'allure des log-vraisemblances négatives représentées aux figures 2.4 et 2.5 : les courbes obtenues sont effectivement très plates autour de la valeur théorique lorsque r est proche de 0 et à l'inverse beaucoup plus pointue lorsque r

est proche de 1.

Remarquons que les erreurs quadratiques moyennes asymptotiques des deux estimateurs déterminées par (2.21) sont aussi représentées sur les figures 2.6 et 2.7 (lignes continues). Les expressions théoriques des erreurs quadratiques moyennes asymptotiques sont en accord avec les résultats obtenus par simulation, même pour de faibles valeurs de n . Ceci est particulièrement vrai pour de grandes valeurs du coefficient de corrélation r .

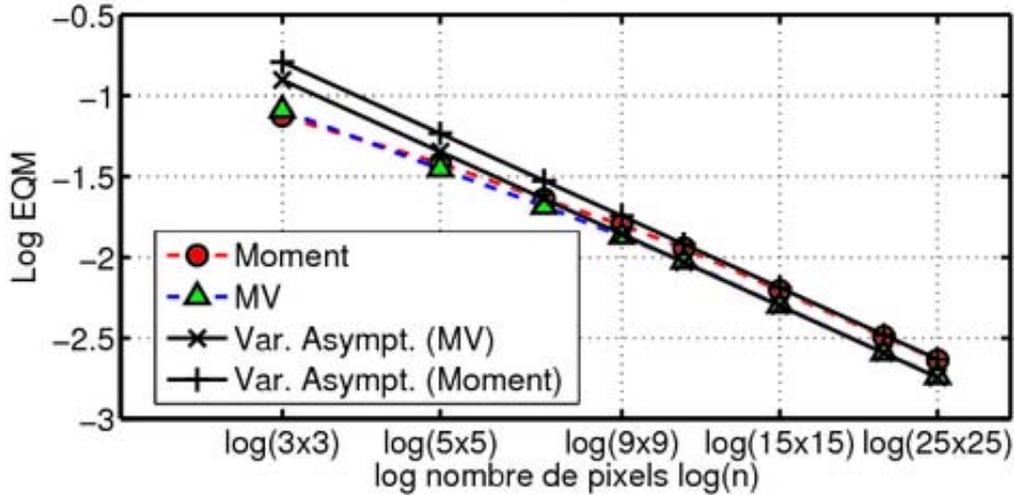


FIG. 2.6 – Erreurs quadratiques moyennes en fonction de $\log(n)$ pour $r = 0.2$.

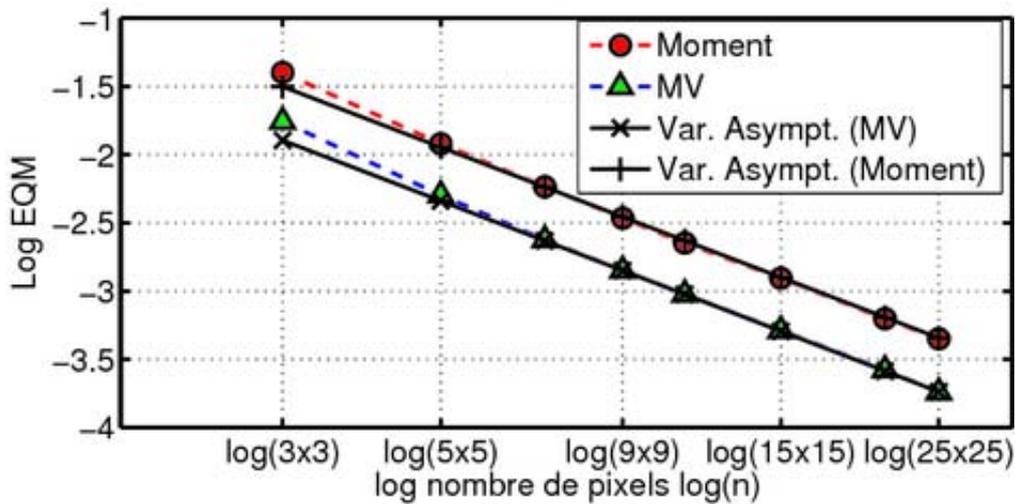


FIG. 2.7 – Erreurs quadratiques moyennes en fonction de $\log(n)$ pour $r = 0.8$.

2.2.4 Recalage pour des images réelles

Cette partie étudie une technique de recalage basée sur les lois gamma bivariées pour des images réelles. La technique de recalage étudiée est tout d'abord décrite.

Procédure de recalage

Considérons deux images dont les pixels sont notés $\{x_1^1, \dots, x_1^n\}$ et $\{x_2^1, \dots, x_2^n\}$. Étant donnée l'image de gauche x_1 , nous proposons les trois étapes suivantes :

- Étape 1 : déterminer la zone de recherche dans l'image de droite x_2 . Pour cette expérience, la zone de recherche est réduite à une ligne (composée de 10 pixels avant et de 10 pixels après le pixel courant),
- Étape 2 : pour chaque pixel x_2^j dans la zone de recherche, estimer une mesure de similarité (coefficient de corrélation ou information mutuelle) entre x_1^i et x_2^j ,
- Étape 3 : sélectionner le pixel donnant la mesure de similarité la plus grande.

Ces trois étapes ont été appliquées à un couple d'images Radarsat monovues (i.e. $q = 1$) acquises avant et après l'éruption du volcan Nyiragongo qui a eu lieu en janvier 2002. Ces images Radarsat sont représentées sur les figures 2.8(a) (avant éruption) et 2.8(b) (après éruption). On notera qu'on peut observer clairement quelques changements dus à l'éruption, en particulier sur le haut de la piste d'atterrissage représentée sous la forme d'une ligne noire. La figure 2.8(c) est le masque qui indique les pixels de l'image qui ont été touchés par l'éruption (pixels blancs). Ce masque a été construit par des photo-interpréteurs à partir des images RSO ici étudiées, d'images optiques et d'observations sur le terrain.

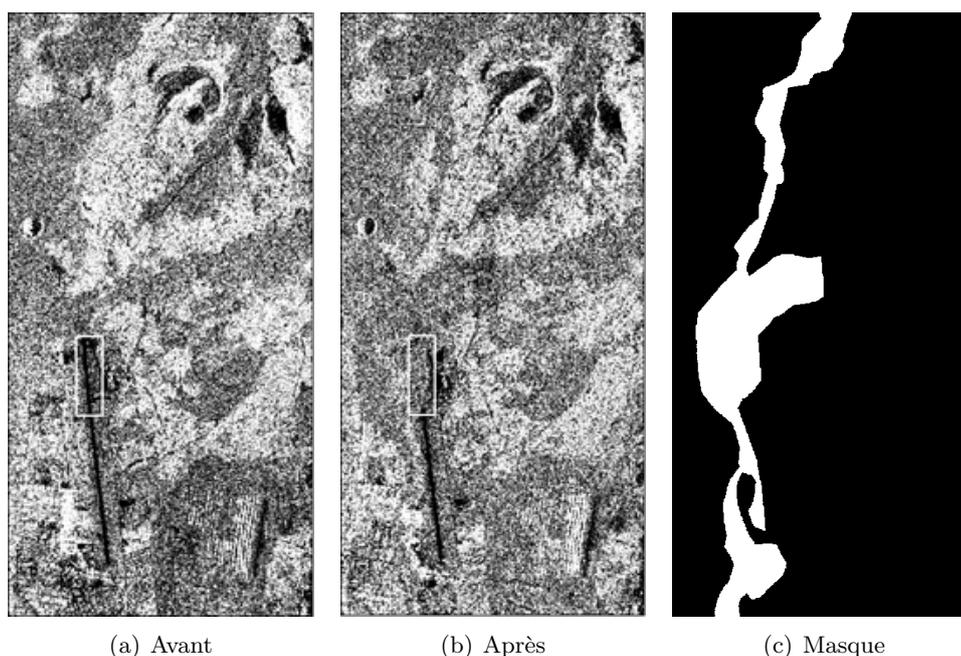


FIG. 2.8 – Images Radarsat du volcan Nyiragongo.

Estimation des coefficients de corrélation

Les figures 2.9(a), 2.9(b) et 2.9(c) montrent une moyenne des estimées du coefficient de corrélation calculée à partir des pixels noirs affectés par l'éruption pour différentes tailles de fenêtres. Plus précisément, pour chaque pixel noir, x_1^i de l'image de gauche, on considère une fenêtre de taille $n = (2p + 1) \times (2p + 1)$ centrée autour de x_1^i . La même fenêtre est aussi considérée

dans l'image de droite autour du pixel x_2^j . Le coefficient de corrélation entre les pixels x_1^i et x_2^j est estimé en utilisant les $n = (2p + 1) \times (2p + 1)$ couples de pixels situés dans les fenêtres de droite et de gauche. Cette opération est répétée pour différents pixels centraux x_2^j appartenant à la zone de recherche (i.e. les 21 pixels de $[x_2^{i+\tau}]_{-10 \leq \tau \leq 10} = [x_2^{i-10}, \dots, x_2^{i-1}, x_2^i, x_2^{i+1}, \dots, x_2^{i+10}]$), où τ est le décalage entre la fenêtre de droite et la fenêtre de gauche. Les résultats sont moyennés sur tous les pixels noirs contenus dans le masque 2.8(c). La moyenne des estimations des coefficients de corrélation est maximale quand $j = i$ (ou de manière équivalente lorsque $\tau = 0$), i.e. quand les fenêtres de droite et de gauche sont situées à la même position. Ce résultat indique que le coefficient de corrélation peut être utilisé pour le recalage d'images.

On définit par la suite la sélectivité comme l'amplitude relative du pic de la courbe par rapport à celle du plateau. Il est alors important d'étudier comment la sélectivité de chaque estimateur varie lorsque la taille de la fenêtre varie. Selon les courbes présentées à la figure 2.9, l'estimateur du maximum de vraisemblance possède donc une meilleure sélectivité que l'estimateur des moments. Cependant d'autres critères d'évaluation des performances du recalage effectué peuvent être définis. On peut par exemple considérer, pour chaque couple de pixels, les décalages τ pour lesquels la corrélation estimée est maximale. La figure 2.10 représente alors l'histogramme de la distribution des décalages τ obtenus lorsque la fenêtre d'estimation est de taille $n = 15 \times 15$. Le critère ainsi construit est donc différent de celui présenté précédemment qui est basé sur les moyennes des coefficients de corrélation en fonction du décalage τ . La figure 2.10 montre finalement que pour ce nouveau critère, les sélectivités des deux estimateurs étudiés sont similaires.

Même si les différentes méthodes d'estimation fournissent des résultats similaires pour le recalage, il est important de noter que le modèle paramétrique des lois Gamma bivariées permet de définir la loi jointe du vecteur (x_1^i, x_2^j) . Cette loi pourrait être utilisée pour d'autres applications, comme la segmentation d'images ou la classification.

La même opération est effectuée sur une zone rectangulaire composée de pixels affectés par l'éruption (pixels blancs du masque 2.8(c)) et représentée par un rectangle sur les figures 2.8(a) et 2.8(b). Les résultats de la figure 2.11 montrent clairement que les coefficients de corrélation estimés sont plus faibles quand ils sont calculés sur une zone touchée par l'éruption par rapport à une zone non affectée par l'éruption et qu'il n'y a plus de pic pour le décalage $\tau = 0$ qui pourrait être utilisé pour le recalage. Ces résultats sont intéressants, car ils montrent que le coefficient de corrélation est une mesure de similarité qui peut être utilisée pour la détection de changements. Ceci fait l'objet de la partie suivante de ce chapitre.

On peut par ailleurs remarquer à partir des figures 2.9 ou même 2.11 que les valeurs moyennes des coefficients de corrélation estimés par maximum de vraisemblance et par la méthode des moments sont sensiblement différentes. En effet pour le modèle des lois Gamma bivariées, le coefficient de corrélation r est toujours positif. Cette contrainte est en particulier vérifiée par l'estimateur du maximum de vraisemblance \widehat{r}_{ML} , puisque la loi n'existe pas dans le cas contraire, et donc les valeurs moyennes de cet estimateur sont nécessairement positives. Ce n'est pas le cas de l'estimateur \widehat{r}_{Mo} obtenu par la méthode des moments, qui vérifie juste la contrainte $\widehat{r}_{ML} \in [-1, 1]$. Pour cette raison, les valeurs moyennes de cet estimateur sont proches de 0 (notamment si les images ne sont pas correctement recalées). Cependant si l'on imposait la contrainte de positivité à ce dernier estimateur (en seillant par exemple les valeurs négatives trouvées à 0), la courbe obtenue pour l'estimateur \widehat{r}_{Mo} serait réhaussée et le niveau du plateau serait similaire à celui obtenu pour la courbe des moyennes de l'estimateur du maximum de vraisemblance \widehat{r}_{ML} .

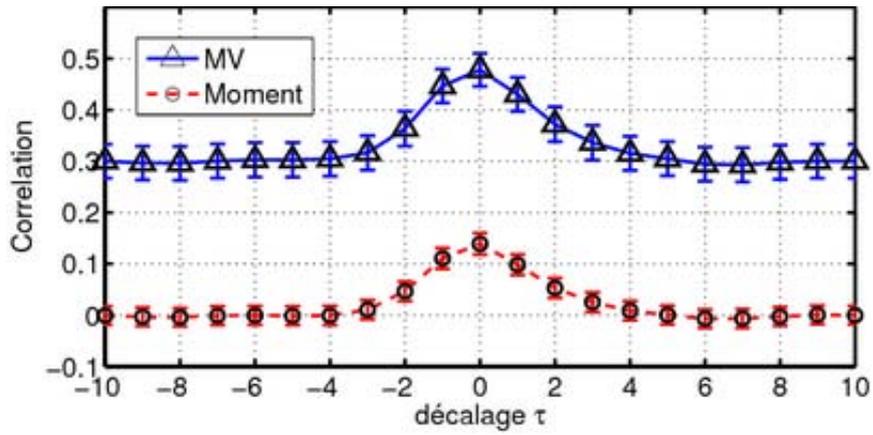
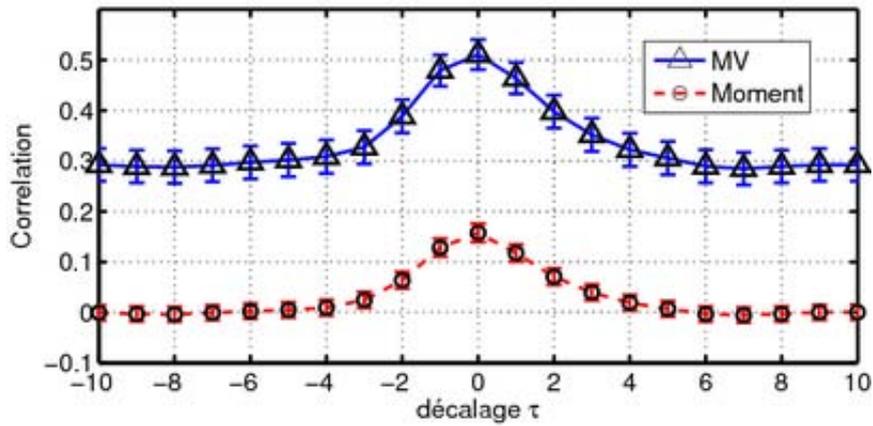
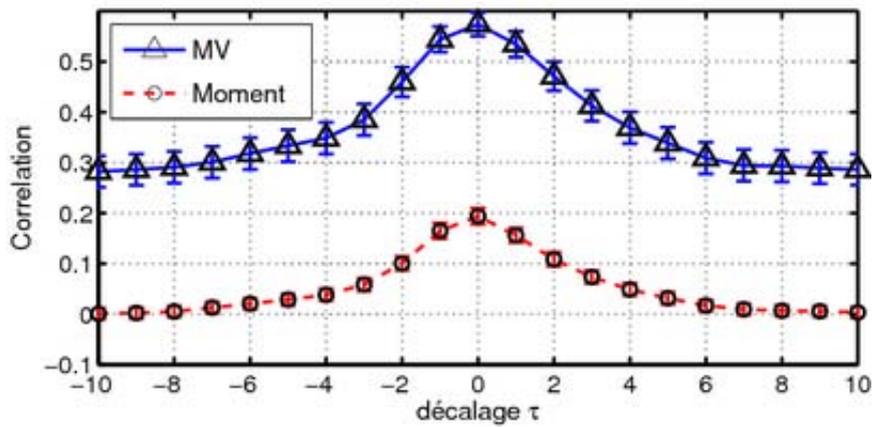
(a) Window Size = 7×7 (b) Window Size = 9×9 (c) Window Size = 15×15

FIG. 2.9 – Moyennes des coefficients de corrélation estimés en fonction de τ et barres d'erreurs pour les pixels non affectés par l'éruption (pixels noirs de la figure 2.8(c)) (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments) pour différentes tailles de fenêtres.

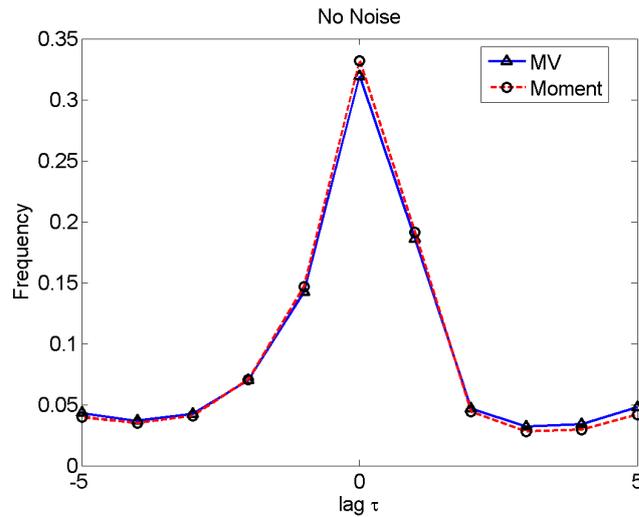


FIG. 2.10 – Histogramme des décalages estimés τ pour les pixels non affectés par l'éruption (pixels noirs du masque 2.8(c)) des images 2.8(a) et 2.8(b) (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).

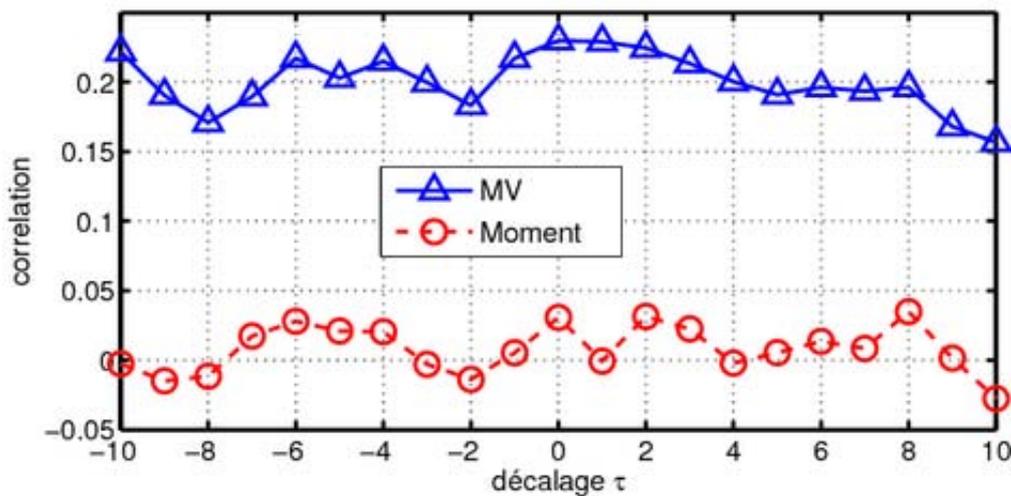


FIG. 2.11 – Moyennes des coefficients de corrélation estimés en fonction de τ pour les pixels blancs appartenant à la région rectangulaire (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).

Influence d'un bruit additif Cette section examine l'effet d'un bruit additif gaussien sur les performances d'estimation. On considère ici, pour chaque couple de pixels, le décalage τ pour lequel la corrélation estimée est maximale. La figure 2.12 représente alors l'histogramme de la distribution des décalages τ obtenus pour une fenêtre d'estimation de taille $n = 15 \times 15$ lorsque les données sont corrompues par un bruit blanc gaussien. Il apparaît clairement que même pour un rapport signal à bruit très faible (ici -20dB), les performances de recalage pour les deux méthodes d'estimation restent très bonnes. Ceci s'explique car si (X_1, X_2) est un couple de v.a.

et B_1, B_2 des v.a. mutuellement indépendantes et indépendantes de (X_1, X_2) , alors :

$$\text{cov}(X_1 + B_1, X_2 + B_2) = \text{cov}(X_1 + B_1, X_2) = \text{cov}(X_1, X_2).$$

La covariance entre deux pixels est donc insensible à un bruit additif, même si $\text{var}(B_i) \gg \text{var}(X_i)$ pour $i \in \{1, 2\}$. Par conséquent, les variations relatives du coefficient de corrélation sont conservées, puisque :

$$r(X_1 + B_1, X_2 + B_2) = \frac{\text{cov}(X_1 + B_1, X_2 + B_2)}{\sqrt{\text{var}(X_1 + B_1) \text{var}(X_2 + B_2)}} = \frac{\text{cov}(X_1, X_2)}{\prod_{i=1}^2 \sqrt{\text{var}(X_i) + \text{var}(B_i)}}.$$

Il faut cependant nuancer ce résultat car les variances des estimateurs de r en présence ou en l'absence de bruit peuvent différer de manière significative.

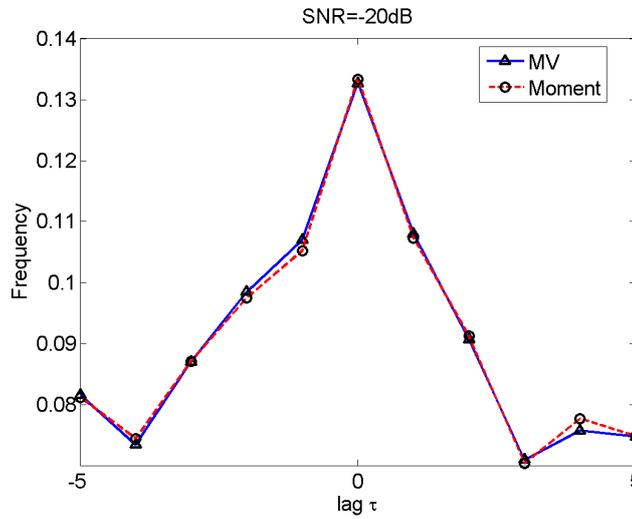


FIG. 2.12 – Histogramme des décalages estimés τ lorsque les images 2.8(a) et 2.8(b) sont corrompues par un bruit additif gaussien (MV : estimateur du maximum de vraisemblance, Moment : estimateur des moments).

2.2.5 Détection de changements

Données synthétiques

On considère des vecteurs $\mathbf{x} = (x_1, x_2)^T$ (provenant d'images synthétiques de taille 128×128) distribués suivant des lois gamma bivariées de coefficients de corrélation $r = 0.3$ et $r = 0.65$ indiquant respectivement la présence et l'absence de changements. Le coefficient de corrélation r de chaque vecteur $\mathbf{x}^{(i,j)} = (x_1^{(i,j)}, x_2^{(i,j)})^T$ (pour $i, j = 1 \dots 128$) est estimé à l'aide des vecteurs appartenant à des fenêtres de taille $n = (2p + 1) \times (2p + 1)$ centrées autour du pixel (i, j) des deux images analysées. Puisqu'il a été montré que les différentes mesures de similarité étudiées sont des fonctions bijectives de r lorsque le paramètre de forme q , i.e. le nombre de vues des images RSO, est connu, on considère donc le test d'hypothèses suivant :

$$\begin{aligned} \mathcal{H}_0 \quad (\text{absence de changement}) : \quad & r > \lambda, \\ \mathcal{H}_1 \quad (\text{présence de changement}) : \quad & r < \lambda, \end{aligned} \tag{2.25}$$

où λ est un seuil et r est le coefficient de corrélation (dont les estimées sont obtenues à l'aide de la méthode des moments ou avec la méthode du maximum de vraisemblance). Les performances du détecteur de changements (2.25) peuvent être mesurées à l'aide des probabilités de détection et de fausse alarme [Tre68, p. 34]

$$\begin{aligned} P_D &= P[\text{accepter } \mathcal{H}_1 | \mathcal{H}_1 \text{ est vraie}] = P[\widehat{r} < \lambda | \mathcal{H}_1 \text{ est vraie}], \\ P_{FA} &= P[\text{accepter } \mathcal{H}_1 | \mathcal{H}_0 \text{ est vraie}] = P[\widehat{r} > \lambda | \mathcal{H}_0 \text{ est vraie}]. \end{aligned} \quad (2.26)$$

Un couple (P_{FA}, P_D) peut donc être défini pour chaque valeur du seuil λ . Ce seuil λ dépend donc de la probabilité de fausse alarme et est généralement calculé pour obtenir une probabilité de fausse alarme P_{FA} donnée. Les courbes représentant P_D en fonction de P_{FA} sont appelées caractéristiques opérationnelle de réception (COR) et sont des outils classiques pour déterminer les performances associées à un problème de détection [Tre68, p. 38].

Les courbes COR pour le problème (2.25) sont représentées sur les figures 2.13(a), 2.13(b) et 2.13(c) pour trois tailles de fenêtres correspondant à $p \in \{9, 15, 21\}$. L'estimateur du maximum de vraisemblance donne clairement de meilleures performances que l'estimateur des moments pour ces exemples. Cependant, il est intéressant de noter que les deux estimateurs donnent des performances similaires pour de grandes tailles de fenêtres. Ceci s'explique car la résolution caractéristique des changements introduits dans les images synthétiques est dans ce cas inférieure à la fenêtre d'estimation de taille $n = 21 \times 21$, ce qui limite les performances de la détection.

Données réelles

Cette partie considère deux images ERS monovues (i.e. $q = 1$) acquises à deux dates différentes autour de Gloucester (en Angleterre) avant et pendant des inondations (le 9 septembre 2000 et le 21 octobre 2000 respectivement). Les images ainsi que le masque indiquant les pixels ayant été modifiés par les inondations sont montrés sur les figures 2.14(a), 2.14(b) et 2.14(c). Le masque, qui correspond à la vérité terrain a été obtenu par des photo-interpréteurs à partir des images RSO étudiées et à partir d'une autre carte de vérité terrain déduite d'images Landsat et SPOT acquises un jour avant les images radar. Ces images sont de taille 3049×1170 pixels, chaque pixel correspondant à une résolution d'environ une dizaine de mètres.

Cette partie compare les performances de plusieurs détecteurs de changements

- Le détecteur développé par Touzi [TLB89, RvZ93], dénommé *Ratio Edge*, qui a été beaucoup utilisé pour les images RSO,
- le détecteur basé sur le coefficient de corrélation entre les deux images, où \widehat{r} dans (2.25) a été estimé avec la méthode des moments (appelé "*Correlation Moment*"),
- le détecteur basé sur le coefficient de corrélation entre les deux images, où \widehat{r} dans (2.25) a été estimé par la méthode du maximum de vraisemblance pour les lois gamma bivariées (appelé "*Correlation ML*").

Le détecteur *ratio edge* est basé sur une mesure de similarité construite à partir du rapport des moyennes de chaque image estimées sur un voisinage du pixel d'intérêt, ici une fenêtre glissante. Ce rapport permet de prendre en compte la nature multiplicative du bruit de speckle. La mesure associée à ce détecteur est souvent normalisée de la façon suivante :

$$\tau_{RE} = 1 - \min\left(\frac{m_2}{m_1}, \frac{m_1}{m_2}\right).$$

Pour le modèle des lois Gamma bivariées, les estimateurs des moyennes étudiées sont les moyennes

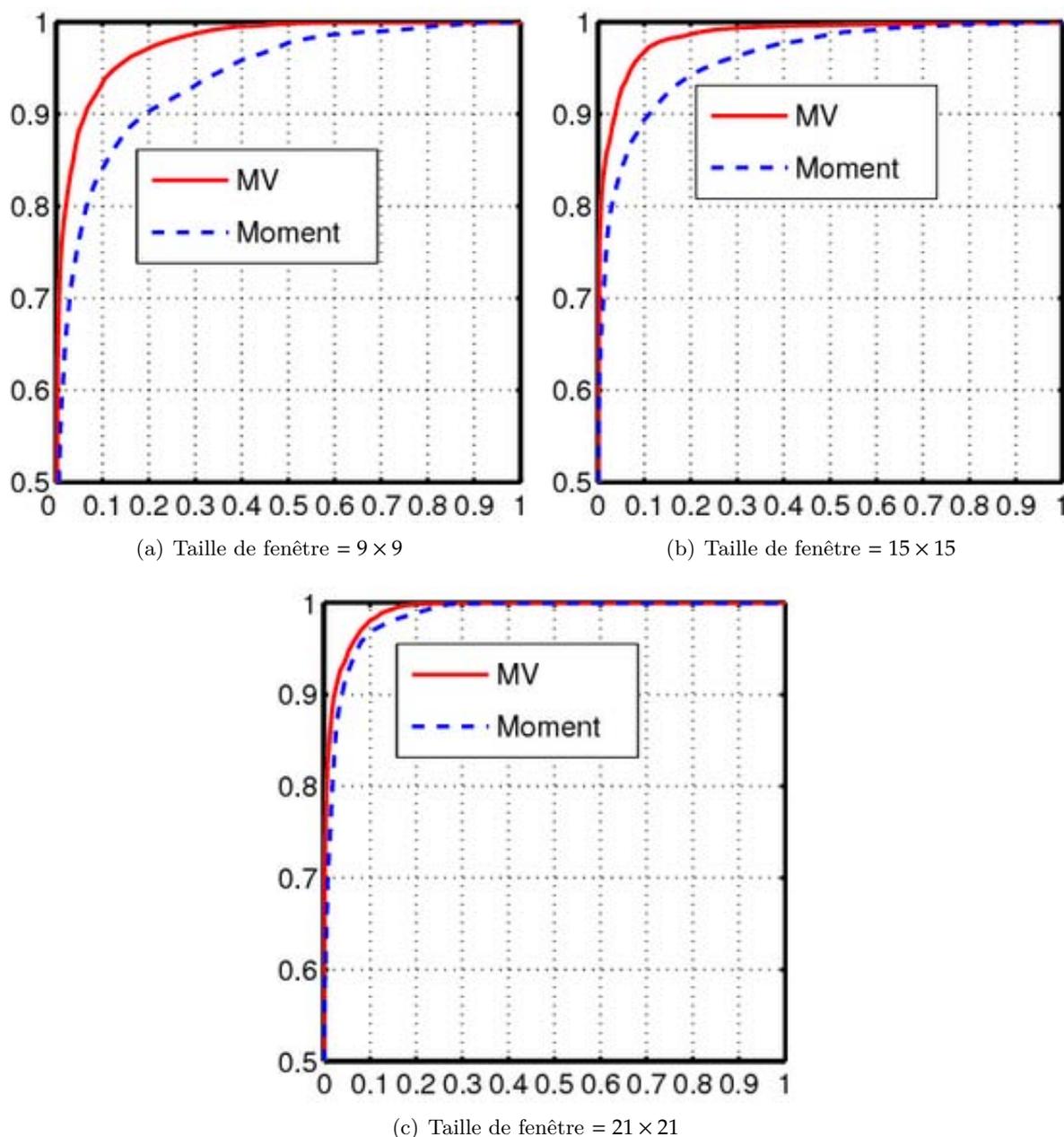


FIG. 2.13 – Courbes COR pour données synthétiques pour différentes tailles de fenêtres.

empiriques. On trouve finalement que la statistique de tests du détecteur *ratio edge* a pour expression :

$$\tau_{RE} = 1 - \min\left(\frac{\bar{x}_2}{\bar{x}_1}, \frac{\bar{x}_1}{\bar{x}_2}\right).$$

Il faut noter que la mesure de similarité ainsi obtenue représente une approche antagoniste par rapport aux mesures de similarité déduites du modèle des ϕ -divergences. En effet ce détecteur ne considère que les changements d'échelles entre les images. À l'inverse, l'information mutuelle, la mesure d'indépendance du χ^2 ou encore le coefficient de corrélation mesurent la dépendance

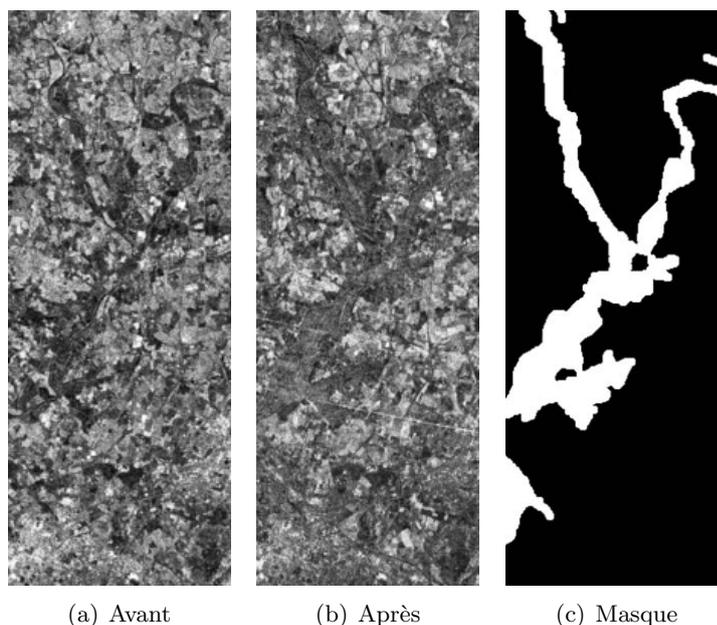


FIG. 2.14 – Images Radarsat de Gloucester avant et après inondations.

entre les images et ne prennent pas en compte les différences d'échelle d'après la proposition 2.1.2.

Les courbes COR pour ce problème de détection de changements sont représentées sur les figures 2.15(a), 2.15(b) et 2.15(c) pour différentes tailles de fenêtres. D'après ces figures, le détecteur *ratio edge* est performant pour de faibles probabilités de fausses alarmes, mais beaucoup moins pour de grandes probabilités de fausses alarmes. Ceci s'explique car la réflectivité moyenne de la scène observée a été modifiée par l'inondation. La réflectivité des sols gorgés d'eau a en effet globalement diminuée, ce qui entraîne beaucoup de fausses alarmes pour une probabilité de détection élevée donnée. Il faut noter enfin qu'il est courant en traitement d'image de considérer des niveaux de fausses alarmes élevées, i.e. de l'ordre de $P_{FA} = 0.5$ ou 0.6 , car les cartes de changement obtenues peuvent ensuite subir des post-traitements afin, par exemple, d'éliminer les groupes de pixels isolés détectés. En définitive, le détecteur basé sur l'estimateur du coefficient de corrélation par la méthode du maximum de vraisemblance pour les lois gamma bivariées donne les meilleurs résultats sur ce jeu de données. Ceci souligne la pertinence du modèle de lois Gamma bivariées pour la détection de changement.

Les résultats précédents ont été obtenus en utilisant le fait que les images sont monovues (i.e. $q = 1$). Cependant afin de pouvoir comparer des images entre elles, les images doivent être pré-traitées. L'opération qui consiste à corriger les effets dus à la géométrie d'acquisition, à la courbure de la terre et à sa rotation ou au relief s'appelle *l'orthorectification*. Cette étape est nécessaire afin d'obtenir des images superposables. Les corrections sont alors effectuées localement : les déformations sont classiquement approchées par des modèles de fractions rationnelles [HTC04, Yan00], et les pixels sont ensuite interpolés (par des méthodes bilinéaires, ou par des fonctions bsplines) à partir du modèle de terrain obtenu. Les interpolations ainsi effectuées ne sont pas toujours distribuées suivant une loi Gamma. Par exemple la somme de deux lois Gamma indépendantes $\mathcal{G}(q, p_1)$ et $\mathcal{G}(q, p_2)$ n'est plus une loi Gamma si les paramètres d'échelle p_1 et p_2 sont différents (il suffit de calculer la transformée de Laplace de la somme pour s'en convaincre).

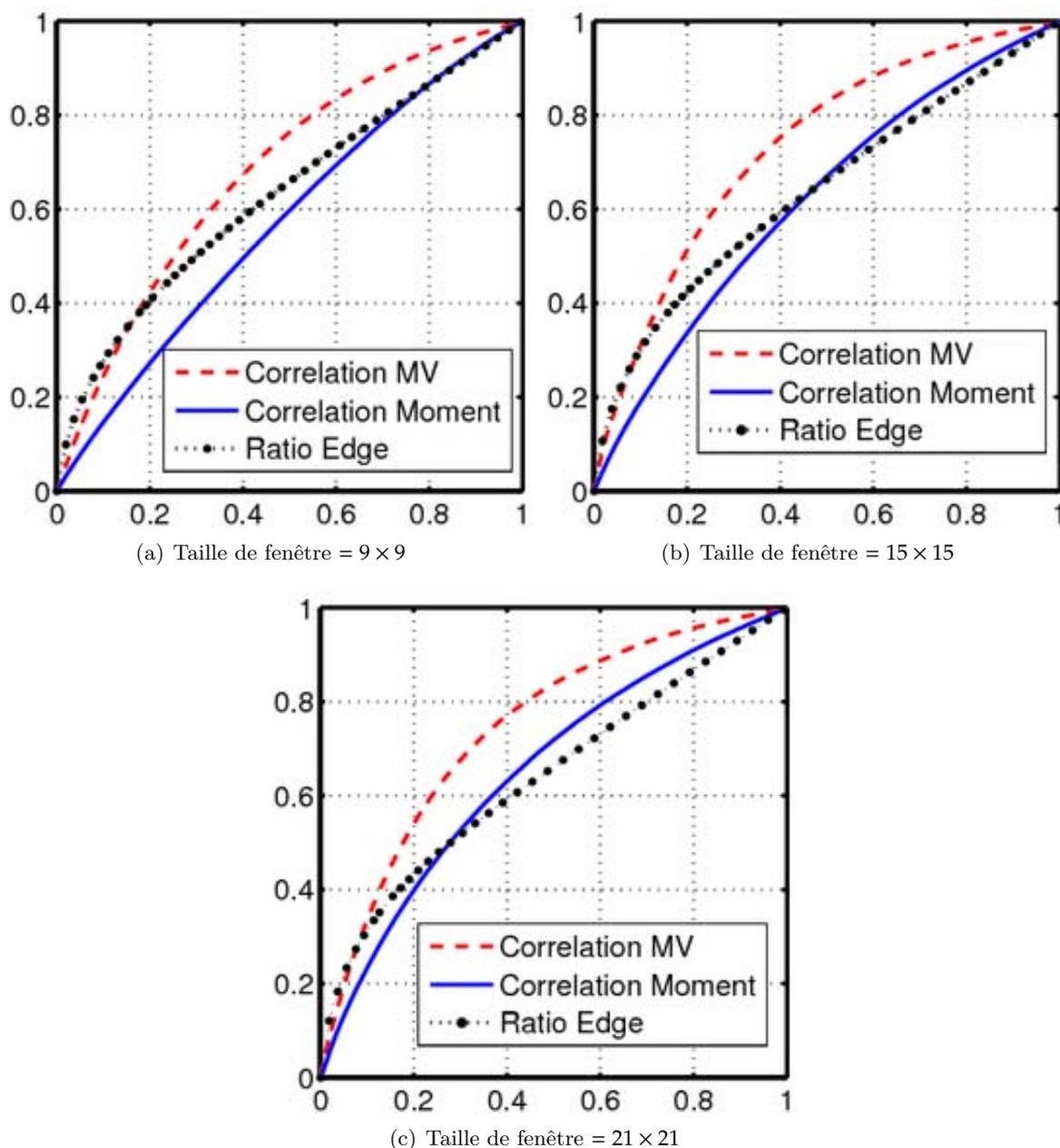


FIG. 2.15 – Courbes COR pour les images de Gloucester pour différentes tailles de fenêtres.

On peut cependant montrer qu'en pratique, une somme de lois Gamma de paramètre de forme identiques q peut être approchée de manière très satisfaisante par une loi Gamma de paramètre de forme $q' > q$. On parle alors pour les images RSO de nombre de vue équivalent (NVE). On peut donc se demander quel est le NVE suite au pré-traitement des images RSO. Nous nous proposons alors d'étudier l'influence du paramètre q sur les performances de détection. Les figures 2.16 montrent les courbes COR obtenues pour différentes valeurs de q pour une taille de fenêtre 15×15 tandis que les courbes 2.17 représentent les cartes de corrélation correspondantes. Comme on peut le voir sur la figure 2.16, les meilleures performances sont obtenues pour $q = 1$ et $q = 1.5$, i.e. lorsque le NVE est similaire aux nombres de vues des images RSO avant orthorectification.

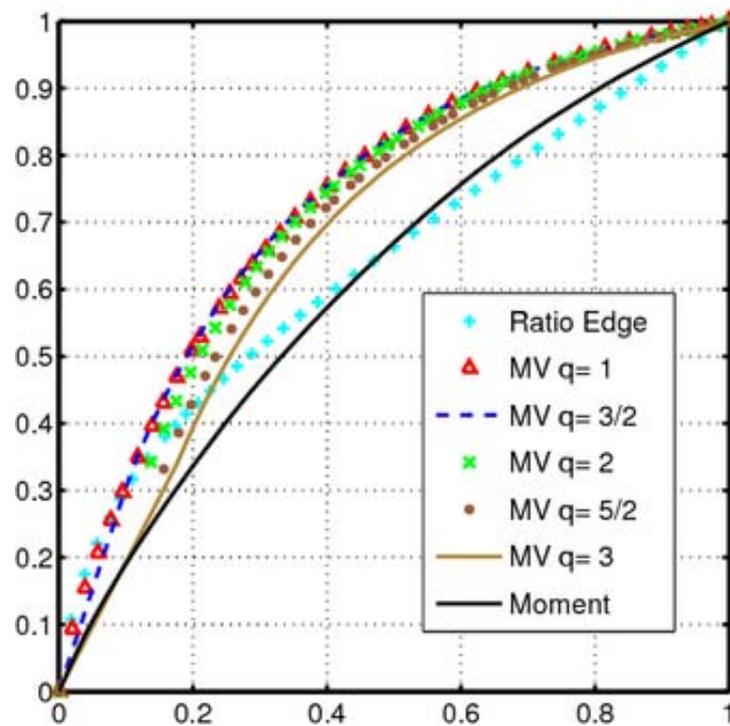


FIG. 2.16 – Courbes COR pour différentes valeurs du paramètre q ($n = 15 \times 15$).

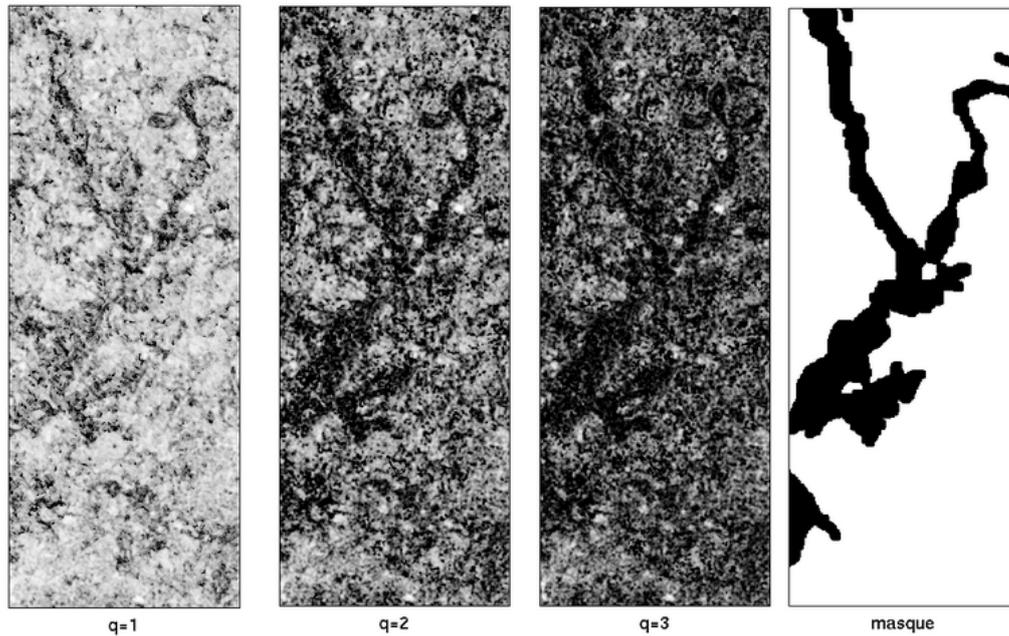


FIG. 2.17 – Cartes de corrélations pour différentes valeurs du paramètre q (de gauche à droite : $q = 1$, $q = 2$, $q = 3$ et le masque des pixels affectés en noir).

Cependant, les modifications étant locales et spécifiques à chaque image, il n'y a pas de raison, pour un pixel donné, que le NVE obtenu pour l'image de référence corresponde au NVE obtenu pour l'image secondaire. Par conséquent, les lois marginales des pixels sont modélisées par de lois Gamma de paramètre de forme différents. Ce cas de figure fait l'objet du chapitre suivant.

Les derniers résultats montrent l'avantage d'utiliser l'information mutuelle pour la détection de changements. Considérons le détecteur de changement suivant basé sur l'information mutuelle

$$\mathcal{H}_0 : \text{IM}_q(\widehat{r}) > \lambda_{PFA}, \quad (2.27)$$

$$\mathcal{H}_1 : \text{IM}_q(\widehat{r}) < \lambda_{PFA}, \quad (2.28)$$

où $\text{IM}_q(\widehat{r})$ est l'information mutuelle dont le calcul est détaillé à la section 2.1.2. Les courbes COR obtenues avec les détecteurs (2.25) et (2.27) sont identiques, ce qui est en accord avec le fait qu'il existe une relation bijective entre \widehat{r} et $\text{IM}_q(\widehat{r})$ d'après la section 2.1.2. Cependant, l'avantage de l'information mutuelle pour la détection de changements est illustré sur la figure 2.18, qui montre la probabilité d'erreur moyenne $Pe = \frac{1}{2}(P_{ND} + P_{FA})$ (où $P_{ND} = 1 - P_D$ est la probabilité de non détection) en fonction du seuil λ pour les détecteurs (2.25) et (2.27). Cette figure indique qu'il y a plus de valeurs du seuil λ vérifiant $P_{ND} + P_{FA} \leq Pe$, où Pe est une probabilité d'erreur maximale fixée. En conséquence, le seuil est plus facile à ajuster pour le détecteur basé sur l'information mutuelle (2.27) que pour le détecteur basé sur le coefficient de corrélation (2.25).

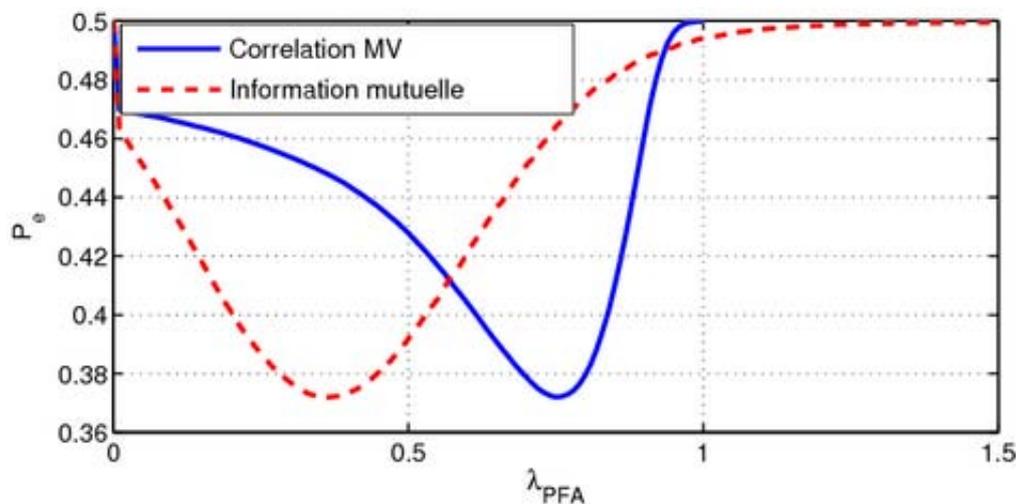


FIG. 2.18 – Probabilité d'erreur moyenne $Pe = \frac{1}{2}(P_{FA} + P_{ND})$ en fonction du seuil λ pour $n = 9 \times 9$.

2.3 Conclusion

Ce chapitre avait pour but d'évaluer les performances de techniques de recalage d'images et de détection de changements basées sur les lois gamma bivariées. Nous avons montré que, sous l'hypothèse de speckle pleinement développé, le modèle des lois Gamma bivariées est adapté aux images RSO et permet de modéliser la dépendance temporelle ces images. Pour ce modèle, les mesures de similarité entre les images étudiées se déduisent des estimateurs du coefficient de

corrélation entre les deux images. L'estimateur du maximum de vraisemblance donne de bons résultats pour cette estimation. Nous avons ensuite montré, au travers de résultats théoriques et à l'aide de simulations effectuées sur des données synthétiques et réelles, que les lois gamma bivariées étaient des outils prometteurs pour le recalage d'images et la détection de changements.

Cependant le modèle des lois Gamma bivariées étudiées dans ce chapitre impose que les lois marginales des couples issus des deux images soient des lois Gamma de même paramètre de forme. Ceci peut être un facteur limitant si les images RSO étudiées ont des nombres de vues différents. Il faut dans ce cas adapter le modèle, ce qui est l'objet du prochain chapitre.

Détection de changements pour des images RSO multi-capteurs

Sommaire

3.1 Lois Gamma multivariées multi-capteurs	78
3.1.1 Définition	78
3.1.2 La loi Gamma bivariée multi-capteurs	79
3.1.3 Moments d'une LGBMu	80
3.1.4 Densité de Probabilité d'une LGBMu	81
3.1.5 Mesures de similarités pour une LGBMu	81
3.2 Estimation des paramètres des LGBMu	83
3.2.1 Méthode du Maximum de Vraisemblance (MV)	83
3.2.2 Fonctions d'Inférence par les Marges (FIM)	85
3.2.3 Méthode de Moments	88
3.3 Données synthétiques	89
3.3.1 Méthodes du maximum de vraisemblance et des moments	89
3.3.2 Fonctions d'inférence par les marges (FIM) et méthode du maximum de vraisemblance (MV)	91
3.4 Détection de Changements	92
3.4.1 Images synthétiques	92
3.4.2 Images Réelles	94
3.5 Comparaison des détecteurs mono et multi-capteurs	97
3.5.1 Estimation à partir de vecteurs synthétiques	97
3.5.2 Détection de changements à l'aide d'images synthétiques et réelles	98
3.6 Estimation des nombres de vues équivalents	100
3.7 Conclusion	104

Dans un contexte général, la fusion de données est devenue une approche très populaire qui permet de tirer profit des avantages respectifs de différents types de capteurs. En outre, dans beaucoup d'applications opérationnelles comme la surveillance terrestre, il n'est souvent pas possible de choisir les données, seules celles déjà existantes et archivées étant exploitables. Il est donc particulièrement important de pouvoir exploiter les données issues de différents capteurs RSO, notamment pour les applications de détection de changements. Ce chapitre étudie une nouvelle famille de lois gamma multivariées dont les marges sont des lois gamma univariées de paramètres de forme deux à deux distincts. Ces lois qui généralisent les familles de lois

Gamma multivariées présentées à la section 1.3.2 sont dénommées *lois gamma multivariées multi-capteurs*. Ces lois permettent en effet de modéliser des images acquises par des capteurs RSO dont les nombres de vues sont différents. Dans ce cas, les lois marginales sont des lois Gamma dont les paramètres de forme sont deux à deux différents d'après la proposition 2.1.1 (l'hypothèse de speckle pleinement développée est supposée vérifiée).

Ce chapitre présente tout d'abord les propriétés des lois gamma multivariées multi-capteurs que nous nous proposons d'utiliser pour la détection de changements. Les estimateurs des paramètres de ces lois sont ensuite étudiés. L'application aux problèmes de détection de changement en imagerie RSO est étudiée dans la troisième partie du chapitre. Les performances obtenues sont comparées à celles obtenues pour le modèle des lois Gamma multivariées dans la quatrième partie. Enfin, le cas où les nombres de vues associés à chaque image RSO sont estimés est étudié dans la dernière partie de ce chapitre.

3.1 Lois Gamma multivariées multi-capteurs

3.1.1 Définition

Les lois Gamma multivariées multi-capteurs sont obtenues à partir des lois Gamma multivariées et peuvent être vues comme une généralisation de ces dernières. Si $P(\mathbf{z})$ est un polynôme affine sur \mathbb{R}^d avec $\mathbf{z} = (z_1, \dots, z_d)^T$, on note par la suite p_i les coefficients des monômes z_i du polynôme $P(\mathbf{z})$ pour $1 \leq i \leq d$.

Définition 3.1.1. *Un vecteur aléatoire $\mathbf{Y} = (Y_1, \dots, Y_d)^T$ suit une loi gamma multivariée multi-capteurs (LGMMu) de paramètres le polynôme affine P et le vecteur $\mathbf{q} = (q_1, \dots, q_d)$, notée $\mathbf{Y} \sim \mathcal{G}(\mathbf{q}, P)$, s'il peut être construit de la manière suivante :*

$$\begin{aligned} Y_1 &= X_1, \\ Y_i &= X_i + Z_i, \quad 2 \leq i \leq d. \end{aligned} \tag{3.1}$$

où

- $\mathbf{X} = (X_1, \dots, X_d)^T$ est un vecteur aléatoire qui suit une LGM sur \mathbb{R}_+^d avec pour paramètre de forme q_1 et pour paramètre d'échelle P , i.e. $\mathbf{X} \sim \mathcal{G}(q_1, P)$,
- Z_2, \dots, Z_d sont des v.a. sur \mathbb{R} , indépendantes et distribuées suivant des lois gamma univariées $Z_i \sim \mathcal{G}(q_i - q_1, p_i)$ avec $q_i \geq q_1$ et la convention $Z_i = 0$ lorsque $q_i - q_1 = 0$.
- le vecteur $\mathbf{Z} = (Z_2, \dots, Z_d)^T$ à valeur dans \mathbb{R}^{d-1} est indépendant de \mathbf{X} .

À partir de cette définition, on obtient facilement la transformée de Laplace qui caractérise une telle distribution. En effet, par indépendance des vecteurs \mathbf{X} et \mathbf{Z} , la transformée de Laplace du vecteur $\mathbf{Y} \sim \mathcal{G}(\mathbf{q}, P)$ ainsi construite a pour expression :

$$\begin{aligned} \mathcal{L}_{\mathbf{Y}}(\mathbf{z}) &= \mathbb{E} \left[e^{-\sum_{i=1}^d z_i Y_i} \right] = \mathbb{E} \left[e^{-\sum_{i=1}^d z_i X_i} \right] \prod_{i=2}^d \mathbb{E} \left[e^{-z_i Z_i} \right] = \mathcal{L}_{\mathbf{X}}(\mathbf{z}) \prod_{i=2}^d \mathcal{L}_{Z_i}(z_i), \\ &= [P(\mathbf{z})]^{-q_1} \prod_{i=2}^d (1 + p_i z_i)^{-(q_i - q_1)}. \end{aligned} \tag{3.2}$$

Une autre conséquence de la définition ci-dessus est que la loi de la variable Y_i est la loi Gamma univariée $\mathcal{G}(q_i, p_i)$. En posant en effet $z_j = 0$ pour $1 \leq j \neq i \leq d$ dans (3.2), on observe en effet

que la transformée de Laplace de Y_i est

$$\mathcal{L}_{Y_i}(z_i) = \frac{1}{(1 + p_i z_i)^{q_1}} \frac{1}{(1 + p_i z_i)^{q_i - q_1}} = \frac{1}{(1 + p_i z_i)^{q_i}},$$

ce qui correspond bien à l'expression de la transformée de Laplace de la loi $\mathcal{G}(q_i, p_i)$.

Remarques :

1. Une condition suffisante pour que la LGMMu $\mathcal{G}(\mathbf{q}, P)$ existe est que la LGM $\mathcal{G}(q_1, P)$ existe. On note de plus qu'une LGMMu se ramène à une LGM de paramètre le polynôme affine P et de paramètre de forme $q = q_1$ lorsque $q_i = q_1$ pour tout $2 \leq i \leq d$.
2. Toutes les marges univariées d'une LGMMu sont des lois Gamma de paramètres de formes deux à deux différents. Ces paramètres de forme modélisent les nombres de vues associés à chaque image RSO. C'est pourquoi on parle de *loi Gamma multivariée multi-capteurs*. Le vecteur \mathbf{q} est par la suite appelé le paramètre de forme de la LGMMu.
3. Ce modèle suppose que la première marge Y_1 possède un paramètre de forme q_1 plus petit que tous les autres paramètres de forme q_i pour tout $2 \leq i \leq d$. Cette hypothèse n'est pas restrictive car on peut toujours réordonner les marges du vecteur de façon à satisfaire cette hypothèse. Cependant, cette hypothèse montre que la loi multivariée obtenue n'est pas symétrique puisque Y_1 joue un rôle particulier dans la définition de cette loi.
4. Tous les sous-vecteurs issus de \mathbf{Y} sont eux-mêmes distribués selon une LGMMu. Cette propriété découle de la proposition 1.3.1 sur les LGM et de la construction donnée ci-dessus des LGMMu.

Enfin, un résultat important sur ces familles de lois est qu'elles ne peuvent pas être vues comme des familles exponentielles. En effet d'après la proposition B.1.2 et l'expression des transformées de Laplace d'une LGMMu (3.2), une mesure de probabilité ν qui appartient à la FEN engendrée par la loi $\mathcal{G}(\mathbf{q}, P)$ définie sur \mathbb{R}^d a pour transformée de Laplace :

$$\mathcal{L}_\nu(\mathbf{z}) = \frac{\mathcal{L}_{\mathcal{G}(\mathbf{q}, P)}(\boldsymbol{\theta} + \mathbf{z})}{\mathcal{L}_{\mathcal{G}(\mathbf{q}, P)}(\boldsymbol{\theta})} = (P_\boldsymbol{\theta}(\mathbf{z}))^{-q_1} \prod_{i=2}^n \left(1 + \frac{p_i}{1 + p_i \theta_i} z_i \right)^{-(q_i - q_1)},$$

où $P_\boldsymbol{\theta}(\mathbf{z}) = \frac{P(\boldsymbol{\theta} + \mathbf{z})}{P(\boldsymbol{\theta})}$. Le polynôme $P_\boldsymbol{\theta}(\mathbf{z})$ est bien par construction un polynôme affine, mais le coefficient de son monôme z_i est en général différent de $\frac{p_i}{1 + p_i \theta_i}$ quel que soit $2 \leq i \leq d$ (et ce même en dimension $d = 2$). La loi ν n'est donc pas une LGMMu, ses marges n'étant même pas des lois Gamma. En conséquence les estimateurs du maximum de vraisemblance des moyennes d'une LGMMu ne sont pas a priori les moyennes empiriques. Ce résultat sera confirmé à la section 3.2.1 consacrée à l'estimation des paramètres d'une loi Gamma bivariée multi-capteurs.

3.1.2 La loi Gamma bivariée multi-capteurs

Une loi gamma bivariée multi-capteurs (LGBMu) correspond au cas particulier où la dimension est $d = 2$. Cette loi est donc caractérisée par la transformée de Laplace suivante :

$$\mathcal{L}_{(Y_1, Y_2)}(\mathbf{z}) = \left(\frac{1}{1 + p_1 z_2 + p_2 z_2 + p_{12} z_1 z_2} \right)^{q_1} \left(\frac{1}{1 + p_2 z_2} \right)^{q_2 - q_1}, \quad (3.3)$$

avec les contraintes suivantes :

$$p_1 > 0, p_2 > 0, p_{12} > 0, p_1 p_2 - p_{12} \geq 0 \quad \text{et} \quad q_2 \geq q_1. \quad (3.4)$$

Dans le cas bivarié, les conditions (3.4) assure que l'expression (3.3) est la transformée de Laplace d'une densité de probabilité définie sur $[0, \infty]^2$. En effet dans ce cas la loi Gamma bivariée de paramètre de forme q_1 et de polynôme affine $P(z_1, z_2) = 1 + p_1 z_2 + p_2 z_2 + p_{12} z_1 z_2$ existe d'après le théorème 1.2.1. La LGBMu peut être obtenue selon la construction donnée à la définition 3.1.1. Sous ces conditions le vecteur (Y_1, Y_2) peut donc être obtenu de la façon suivante :

$$\begin{aligned} Y_1 &= X_1, \\ Y_2 &= X_2 + Z_2, \end{aligned} \quad (3.5)$$

où $(X_1, X_2) \sim \mathcal{G}(q_1, P)$ et Z_2 est une v.a. indépendante de (X_1, X_2) et de loi $\mathcal{G}(q_2 - q_1, p_2)$.

3.1.3 Moments d'une LGBMu

Les moments de (Y_1, Y_2) sont obtenus à partir des moments de (X_1, X_2) et de Z_2 . Par exemple, grâce à l'indépendance entre (X_1, X_2) et Z_2 , on obtient les résultats suivant :

$$\begin{aligned} \mathbb{E}[Y_i] &= q_i p_i, \quad i \in \{1, 2\}, \\ \text{var}(Y_i) &= q_i p_i^2, \quad i \in \{1, 2\}, \\ \text{cov}(Y_1, Y_2) &= \text{cov}(X_1, X_2) = q_1(p_1 p_2 - p_{12}), \\ r(Y_1, Y_2) &= \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(Y_1)} \sqrt{\text{var}(Y_2)}} = \sqrt{\frac{q_1}{q_2}} \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \end{aligned} \quad (3.6)$$

Il est important de remarquer que les conditions (3.4) entraînent que le coefficient de corrélation prend ses valeurs dans l'intervalle :

$$0 \leq r(Y_1, Y_2) \leq \sqrt{\frac{q_1}{q_2}}. \quad (3.7)$$

Les bornes de l'intervalle sont atteintes pour les cas particuliers suivants :

- La borne supérieure est atteinte lorsque les v.a. X_1 et X_2 sont totalement corrélés. En effet dans ce cas le paramètre p_{12} est nul (voir le théorème 1.2.1). On peut d'ailleurs remarquer que si en plus $p_1 = p_2$, ce qui est équivalent à dire que $X_1 = X_2$, alors le vecteur (Y_1, Y_2) défini à l'équation (3.5) est distribué suivant la loi Gamma bivariée de MacKay présentée à la section 1.2.1 (p. 12)¹¹.
- Le cas où $r(Y_1, Y_2) = 0$ est atteint lorsque $p_{12} = p_1 p_2$, i.e. lorsque les marges du vecteur (X_1, X_2) sont indépendantes d'après l'équation (1.26). Selon la définition du vecteur (Y_1, Y_2) donnée à l'équation (3.5), les marges Y_1 et Y_2 sont alors indépendantes. Comme pour les lois Gamma bivariées, *absence de corrélation et indépendance sont des propriétés équivalentes pour les LGBMu.*

Plutôt que de travailler par la suite avec le coefficient de corrélation $r(Y_1, Y_2)$, il est préférable d'introduire le coefficient de corrélation normalisé, défini comme :

$$r'(Y_1, Y_2) = \sqrt{\frac{q_2}{q_1}} r(Y_1, Y_2) = \frac{p_1 p_2 - p_{12}}{p_1 p_2}, \quad (3.8)$$

¹¹On peut également montrer que la borne de Fréchet supérieure (voir [Joe97, chap. 3]) des lois bivariées dont les lois marginales sont $\mathcal{G}(q_1, p_1)$ et $\mathcal{G}(q_2, p_1)$ avec $q_2 > q_1$ est la loi de MacKay. Par conséquent, la valeur maximale du coefficient de corrélation que l'on peut espérer entre les deux lois $\mathcal{G}(q_1, p_1)$ et $\mathcal{G}(q_2, p_1)$ correspond au coefficient de corrélation de la loi de Mackay, i.e. $\sqrt{q_1/q_2}$. Cette valeur peut donc être atteinte pour le modèle des LGBMu.

Ce coefficient de corrélation normalisé correspond en fait au coefficient de corrélation de la loi Gamma bivariée (X_1, X_2) à partir de laquelle est construite la loi multi-capteurs, et peut donc prendre toutes les valeurs de l'intervalle $[0, 1]$.

Finalement, on peut remarquer que pour des valeurs données des paramètres de forme q_1 et q_2 , une LGBMu est entièrement caractérisée par le vecteur de paramètres $\boldsymbol{\theta} = (\mathbf{E}[Y_1], \mathbf{E}[Y_2], r'(Y_1, Y_2))$, puisque $\boldsymbol{\theta}$ et (p_1, p_2, p_{12}) sont liés par une transformation bijective.

3.1.4 Densité de Probabilité d'une LGBMu

Par construction, la densité du vecteur $\mathbf{Y} = (Y_1, Y_2)^T$ qui est distribuée selon une LGBMu (i.e. $\mathbf{Y} \sim \mathcal{G}(\mathbf{q}, P)$) est la convolution entre la fonction $z \mapsto f_{X_1, X_2}(x_1, z)$, où $f_{X_1, X_2}(\mathbf{x})$ est la densité de la loi Gamma bivariée de paramètre de forme q_1 dont une expression est donné à l'équation (1.35), et la densité $f_Z(z)$ de $Z \sim \mathcal{G}(q_2 - q_1, p_2)$. On obtient finalement après calculs (voir l'annexe D.2.1) :

$$f_{\mathbf{Y}}(\mathbf{y}) = \left(\frac{p_1 p_2}{p_{12}} \right)^{q_1} \frac{y_1^{q_1-1} y_2^{q_2-1} e^{-\left(\frac{p_2}{p_{12}} y_1 + \frac{p_1}{p_{12}} y_2\right)}}{p_1^{q_1} p_2^{q_2} \Gamma(q_2) \Gamma(q_1)} \Phi_3 \left(q_2 - q_1; q_2; c \frac{p_{12}}{p_2} y_2, c y_1 y_2 \right) \mathbb{I}_{[0, \infty]^2}(\mathbf{y}), \quad (3.9)$$

où $c = (p_1 p_2 - p_{12}) / p_{12}^2$ et où Φ_3 est une fonction appelée fonction de Horn. Les fonctions de Horn regroupent vingt séries hypergéométriques confluentes convergentes d'ordre deux [AET81, 224–229] dont la fonction Φ_3 définie par :

$$\Phi_3(a; b; x, y) = \sum_{m, n=0}^{\infty} \frac{(a)_m}{(b)_{m+n}} \frac{x^m y^n}{m! n!}, \quad \forall x, y \in \mathbb{R}, \quad (3.10)$$

où $(a)_m$ est le symbole de Pochhammer défini en (1.21).

Il est intéressant de constater que la relation

$$f_q(c y_1 y_2) = \Phi_3 \left(0; q; c \frac{p_{12}}{p_2} y_2, c y_1 y_2 \right) / \Gamma(q)$$

permet de montrer que la densité d'une LGBMu se ramène à celle d'une loi Gamma bivariée (1.35) lorsque $q_1 = q_2 = q$.

3.1.5 Mesures de similarités pour une LGBMu

Comme cela a été remarqué à la section 3.1.3, lorsque les paramètres de forme $q_2 \geq q_1 > 0$ sont connus, la loi du vecteur (Y_1, Y_2) peut être paramétrée par le vecteur $\boldsymbol{\theta} = (m_1, m_2, r')$ où m_1 et m_2 sont les moyennes respectives de Y_1 et Y_2 et r' est le coefficient de corrélation normalisé défini à l'équation (3.8). Selon la proposition 2.1.2, l'information mutuelle et la mesure d'indépendance du χ^2 ne dépendent pas des paramètres d'échelle. Finalement lorsque $q_2 \geq q_1 > 0$ sont fixés, ces mesures de similarité ne dépendent que de r' et sont même des fonctions croissantes de r' qui s'annule si et seulement si $r' = 0$. Par conséquent, les performances entre les méthodes de détection de changement basées sur ces mesures de similarités et celles basées directement sur le coefficient de corrélation seront identiques à $q_2 \geq q_1 > 0$ fixés.

Ce résultat n'est plus vrai si les nombres de vues des images RSO sont inconnus, car alors les paramètres q_1 et q_2 doivent être estimés. Il est alors intéressant d'obtenir une expression de

ces mesures de similarité. Malheureusement, il est difficile d'obtenir une expression analytique simple, ou même une valeur approchée lorsque $r' \rightarrow 1$, de l'information mutuelle associée à une LGBMu. Par contre, la mesure d'indépendance du χ^2 peut s'exprimer facilement en remarquant que les LGBMu sont des probabilités de Lancaster, dont une caractérisation a été donnée page 22. En effet, à partir de la transformée de Laplace (3.3) d'une LGBMu, et avec un calcul similaire à celui effectué à l'annexe C.1.3 pour les lois Gamma bivariées, on obtient l'expression suivante de la densité d'une LGBMu (on suppose ici que les moyennes de la loi sont $m_1 = q_1$ et $m_2 = q_2$) :

$$p_{(Y_1, Y_2)}(y_1, y_2) = \mathcal{G}_{q_1, 1}(y_1) \mathcal{G}_{q_2, 1}(y_2) \sum_{k=0}^{\infty} \rho_k P_k^{(q_1-1)}(y_1) P_k^{(q_2-1)}(y_2), \quad (3.11)$$

où la fonction $x \mapsto \mathcal{G}_{q, 1}(x)$ correspond à la densité de la loi Gamma univariée $\mathcal{G}(q, 1)$ et où $(P_k^{(q-1)})_{k \geq 0}$ est la famille de polynôme orthonormale définie à l'équation (1.46) avec

$$\rho_k = r'^k \sqrt{\frac{(q_1)_k}{(q_2)_k}}. \quad (3.12)$$

D'après l'équation (1.41), la mesure d'indépendance du χ^2 a pour expression :

$$I_{\chi^2}(q_1, q_2, r') = -1 + \sum_{k \geq 0} \frac{(q_1)_k}{(q_2)_k} (r'^2)^k = -1 + {}_2F_1(1, q_1; q_2; r'^2) \quad \forall q_2 \geq q_1 > 0, r' \in [0, 1], \quad (3.13)$$

où ${}_2F_1(a, b; c; x)$ est la fonction hypergéométrique standard [AS64, p. 559] :

$${}_2F_1(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}, \quad \forall x \in [0, 1].$$

Il faut noter que nous avons obtenu une nouvelle expression analytique de la densité d'une LGBMu. De plus, cette expression correspond à la loi Gamma bivariée proposée par Sarmanov [Sar66] lors de sa construction de familles de lois bivariées de marges données. Cependant, cette représentation ne peut être exploitée par la suite, car le calcul de l'expression (3.11) est numériquement instable. En pratique lorsque $r' > 0.5$, la série

$$\sum_{k=0}^{\infty} \rho_k P_k^{(q_1-1)}(y_1) P_k^{(q_2-1)}(y_2)$$

a une valeur strictement positive, mais très proche de 0 lorsque $y_1 > 0$ et $y_2 > 0$ sont sensiblement différents. Les termes de la série n'étant pas de signe constant, il est alors très difficile de calculer numériquement cette valeur (par exemple, l'évaluation numérique donne souvent des résultats négatifs). Pour ces raisons, l'expression de la densité utilisée notamment pour les problèmes d'estimation sera celle donnée à l'équation (3.9). Cette expression est en effet une série de termes positifs et son évaluation numérique (voir annexe D.2.1) ne pose pas problème.

En conclusion, coefficient de corrélation, information mutuelle et mesure d'indépendance du χ^2 sont des mesures de similarité locales équivalentes pour la détection de changements, puisqu'elles se déduisent l'une de l'autre par une transformation bijective lorsque les nombres de vues associées aux images sont fixés. Dans ce cas la mesure de similarité étudiée sera donc le coefficient de corrélation. Cependant, lorsque les nombres de vues sont inconnus, la mesure de l'indépendance du χ^2 dont une expression analytique simple est disponible sera préférable.

3.2 Estimation des paramètres des LGBMu

Ce chapitre étudie différentes méthodes d'estimation des paramètres d'une LGBMu¹². Les notations suivantes sont utilisées dans la suite du chapitre :

$$m_1 = \mathbf{E}[Y_1], \quad m_2 = \mathbf{E}[Y_2], \quad r' = r(Y_1, Y_2) \sqrt{\frac{q_2}{q_1}},$$

et par conséquent le vecteur des paramètres inconnus est $\boldsymbol{\theta} = (m_1, m_2, r')^T$. Dans ce chapitre, on s'intéresse à l'estimation du vecteur $\boldsymbol{\theta}$ à partir de n vecteurs $\mathbf{Y} = (\mathbf{Y}^1, \dots, \mathbf{Y}^n)$, où $\mathbf{Y}^i = (Y_1^i, Y_2^i)$ suit une loi GBMu de paramètre le vecteur $\boldsymbol{\theta}$. Il est important de remarquer que les paramètres de forme q_1 et q_2 sont supposés connus, comme c'est le cas dans la plupart des applications pratiques. Cette hypothèse peut cependant être relâchée.

3.2.1 Méthode du Maximum de Vraisemblance (MV)

Principe

La méthode du maximum de vraisemblance peut être appliquée à l'échantillon \mathbf{Y} , une expression explicite de la densité étant disponible. Dans ce cas particulier, après avoir supprimé les termes qui ne dépendent pas de $\boldsymbol{\theta}$, la fonction de log-vraisemblance de \mathbf{Y} s'écrit :

$$l(\mathbf{Y}; \boldsymbol{\theta}) = -nq_1 \log(1 - r') - n \left(\sum_{j=1}^2 q_j \log m_j + \frac{q_j}{m_j(1 - r')} \bar{Y}_j \right) + \sum_{i=1}^n \log \Phi_3(q_2 - q_1; q_2; dY_2^i, cY_1^i Y_2^i), \quad (3.14)$$

où $c = \frac{r'q_1q_2}{m_1m_2(1-r')^2}$, $d = \frac{r'q_2}{m_2(1-r')}$ et $\bar{Y}_1 = \frac{1}{n} \sum_{i=1}^n Y_1^i$, $\bar{Y}_2 = \frac{1}{n} \sum_{i=1}^n Y_2^i$ sont les moyennes empiriques de Y_1 et Y_2 . En dérivant la log-vraisemblance par rapport à $\boldsymbol{\theta}$, l'estimateur du maximum de vraisemblance (EMV) de $\boldsymbol{\theta}$ est obtenu comme la solution du système :

$$\mathbf{u}(\mathbf{Y}; \boldsymbol{\theta}) = \left(\frac{\partial l(\mathbf{Y}; m_1, m_2, r')}{\partial m_1}, \frac{\partial l(\mathbf{Y}; m_1, m_2, r')}{\partial m_2}, \frac{\partial l(\mathbf{Y}; m_1, m_2, r')}{\partial r'} \right)^T = \mathbf{0}^T, \quad (3.15)$$

où $\mathbf{u}(\mathbf{Y}; \boldsymbol{\theta}) = \frac{\partial l(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ est appelée la fonction de score. Le calcul effectué à l'annexe (D.2.4) montre alors que l'EMV de m_2 a pour expression :

$$\widehat{m}_{2\text{MV}} = \bar{Y}_2. \quad (3.16)$$

Il n'existe par contre pas d'expression simple des EMV des paramètres m_1 et r' . En particulier m_1 ne correspond plus à la moyenne empirique \bar{Y}_1 , mis à part lorsque les marges sont indépendantes (i.e. $r' = 0$). Ce résultat illustre les deux points suivants qui ont déjà fait l'objet de remarques :

- la dissymétrie du modèle des LGBMu : les estimateurs du maximum de vraisemblance des moyennes n'ont pas une expression similaire.

¹²Les résultats présentés ici peuvent être étendus au cas général des LGMMu en utilisant le concept de vraisemblance composite. Voir la section 4.3.2 pour plus de précisions.

- les familles de LGBMu ne sont pas des familles exponentielles naturelles : les EMV des moyennes ne sont pas en général les moyennes empiriques.

Les EMV de m_1 et r' sont donc obtenus en remplaçant m_2 par $\widehat{m}_{2\text{MV}}$ dans (3.14) et en minimisant l'opposée de la log-vraisemblance $l(\mathbf{Y}; (m_1, \widehat{m}_{2\text{ML}}, r'))$ ainsi obtenue par rapport à m_1 et r' . Cette minimisation est effectuée en utilisant une méthode de type quasi-Newton avec les contraintes $m_1 > 0$ et $r' \in [0, 1[$, une expression analytique du gradient de la log-vraisemblance pouvant être obtenue. Les points clés sur la manière de calculer numériquement la fonction de Horn et ses dérivées sont exposés dans les annexes D.2.2 et D.2.3. Finalement, l'EMV du coefficient de corrélation r est déduit par invariance fonctionnelle :

$$\widehat{r}_{\text{ML}} = \sqrt{\frac{q_1}{q_2}} \widehat{r}'_{\text{ML}}.$$

Afin de prouver l'existence des estimateurs du maximum de vraisemblance, il faudrait montrer que l'opposé de la log-vraisemblance (3.14) admet un unique minimum sur l'ensemble des paramètres $\boldsymbol{\theta} \in [0, \infty[^2 \times [0, 1[$ mais il est difficile de montrer théoriquement ce résultat. Plus de détails seront donnés à la section 3.2.2 sur l'estimateur obtenu par la méthode des fonctions d'inférences par les marges qui sera l'estimateur retenu par la suite pour les problèmes de détection de changements.

Performance

Les propriétés de l'EMV $\widehat{m}_{2\text{ML}}$ sont obtenues directement à partir de la loi gamma univariée $\mathcal{G}(q_2, p_2)$. Cet estimateur est évidemment sans biais et convergent. Cet estimateur est de plus efficace, car sous les hypothèses de régularité qui sont vérifiées pour ce modèle, les estimateurs du maximum de vraisemblance sont asymptotiquement efficaces. Par construction, $\text{var}(\widehat{m}_{2\text{ML}}) = \frac{K_2}{n}$, avec $K_2 = m_2^2/q_2$. Cette variance est donc égale à la borne de Cramer-Rao de m_2 et $\widehat{m}_{2\text{ML}}$ est efficace. Les performances de $\widehat{m}_{1\text{ML}}$ et \widehat{r}_{ML} sont plus difficiles à étudier, car aucune expression analytique de ces estimateurs n'est disponible. Ces estimateurs étant obtenus par la méthode du maximum de vraisemblance, ils sont asymptotiquement sans biais et asymptotiquement efficaces. Les erreurs quadratiques moyennes (EQM) de $\widehat{m}_{1\text{ML}}$ et \widehat{r}_{ML} peuvent donc être approchées pour des échantillons de grande taille par la borne de Cramer-Rao (BCR). Pour des estimateurs non biaisés, la BCR est obtenue en inversant la matrice d'information de Fisher \mathcal{I} définie comme :

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E} \left[\frac{\partial \mathbf{u}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right].$$

Le calcul de \mathcal{I} nécessite de déterminer l'opposé des espérances des dérivées partielles d'ordre deux de $l(\mathbf{Y}; \boldsymbol{\theta})$ par rapport à m_1 , m_2 et r dans (3.14). Des expressions analytiques explicites des éléments de \mathcal{I} sont difficiles à obtenir à cause du terme $\log \Phi_3$. Comme dans le cas des Lois Gamma bivariées (voir section 2.2.1), les éléments de la matrice sont donc approchés par des méthodes de type Monte-Carlo. Ceci fournit des approximations appréciables des EQM des EMV (voir le chapitre suivant sur les résultats de simulations).

3.2.2 Fonctions d'Inférence par les Marges (FIM)

Principe

Cette méthode d'estimation a été proposée par Shih et Louis [SL95], avant d'être généralisée par Xu [Xu96], afin d'estimer les paramètres de lois bivariées construites à partir de copules. La méthode d'estimation FIM se décompose en deux étapes dont les concepts principaux sont présentés par exemple dans [Joe97, chap. 10]. La première étape consiste à estimer séparément les paramètres des lois marginales à partir des vraisemblances de ces lois univariées. Les paramètres de dépendance sont ensuite estimés à partir des estimateurs précédents et de la vraisemblance jointe de la loi multivariée. Dans le cadre des LGBMu, ces deux étapes sont explicitées ci-dessous :

1. Estimation des paramètres inconnus m_1 et m_2 à partir des lois marginales de Y_1 and Y_2 . Cette estimation est effectuée en maximisant les vraisemblances marginales $l(Y_1; m_1)$ et $l(Y_2; m_2)$ par rapport à m_1 et m_2 respectivement,
2. Estimation du paramètre r' en maximisant la vraisemblance jointe $l(\mathbf{Y}; \widehat{m}_{1\text{FIM}}, \widehat{m}_{2\text{FIM}}, r')$ par rapport à r' . Il faut noter que les paramètres m_1 and m_2 ont été remplacés dans la vraisemblance jointe par leur estimation effectuée dans la première étape de la méthode FIM.

Cette méthode est qualifiée de fonctions d'inférence par les marges, car les fonctions d'inférence, ou d'estimation, correspondent aux fonctions scores des vraisemblances des marges et de la vraisemblance jointe. L'intérêt de la procédure d'estimation FIM est d'être souvent beaucoup plus simple d'un point de vue calculatoire que la méthode MV qui estime tous les paramètres simultanément à partir de la vraisemblance jointe. En effet une optimisation numérique par rapport à plusieurs variables est bien plus coûteuse en terme de calculs que plusieurs optimisations par rapport à moins de paramètres. Dans notre cas, les lois marginales d'une LGBMu sont des lois gamma univariées de paramètres de forme q_i et de moyennes m_i , pour $i = \{1, 2\}$. Par conséquent, les estimateurs FIM de m_1, m_2, r' sont obtenus comme étant les solutions de :

$$\mathbf{g}(\mathbf{Y}; \boldsymbol{\theta}) = \left(\frac{\partial l_1(Y_1; m_1)}{\partial m_1}, \frac{\partial l_2(Y_2; m_2)}{\partial m_2}, \frac{\partial l(\mathbf{Y}; m_1, m_2, r')}{\partial r'} \right)^T = \mathbf{0}^T, \quad (3.17)$$

où l_i est la log-vraisemblance associée à la variable Y_i , pour $i = \{1, 2\}$, et l est la log-vraisemblance jointe définie à l'équation (3.14). Les estimateurs FIM de m_1 et m_2 sont obtenus de façon classique à l'aide des propriétés des lois gamma univariées (ces lois forment en effet une famille exponentielle, et donc l'EMV de la moyenne est la moyenne empirique) :

$$\widehat{m}_{1\text{FIM}} = \bar{Y}_1, \quad \widehat{m}_{2\text{FIM}} = \bar{Y}_2. \quad (3.18)$$

L'estimateur FIM de r' est obtenu en remplaçant m_1 et m_2 par \bar{Y}_1 et \bar{Y}_2 dans (3.14) et en minimisant la log-vraisemblance ainsi obtenue par rapport à r' . Cette dernière minimisation est effectuée par une méthode de type quasi-Newton sous contrainte (la contrainte étant $r' \in [0, 1]$), une expression analytique du gradient étant disponible.

Il faut noter que la méthode MV présentée précédemment nécessite d'optimiser la log-vraisemblance $l(\mathbf{Y}; m_1, \widehat{m}_{2\text{ML}}, r')$ par rapport à m_1 et r' , tandis que la méthode FIM nécessite seulement d'optimiser $l(\mathbf{Y}; \bar{Y}_1, \bar{Y}_2, r')$ par rapport à une unique variable r' . La procédure d'optimisation est par conséquent bien moins coûteuse en temps de calcul pour la méthode FIM que pour la méthode MV. Par ailleurs, l'estimateur de m_2 est le même pour les deux méthodes MV et FIM. Finalement, il est intéressant de souligner que la vraisemblance jointe est égale au

produit des lois gamma univariées lorsque $r' = 0$. Par conséquent, les estimateurs MV et FIM des moyennes sont identiques lorsque $r' = 0$.

Existence

Les estimateurs des moyennes obtenus par la méthode FIM sont bien définis pour une loi LGBMu, ces paramètres correspondant à l'unique valeur pour laquelle la vraisemblance de chaque marge est maximum. Prouver l'existence des estimateurs FIM nécessiterait de montrer que la fonction $r' \in [0, 1[\mapsto -l(\mathbf{Y}; \widehat{m}_{1\text{FIM}}, \widehat{m}_{2\text{FIM}}, r')$ admet un minimum unique. Cependant, même dans le cas où le vecteur $\mathbf{X} = (X_1, X_2)$ à partir duquel est construit le vecteur \mathbf{Y} est distribué suivant une loi Gamma bivariée standard, la log-vraisemblance obtenue n'est pas concave. En effet la fonction $r' \in [0, 1[\mapsto \Phi_3(a, b, \alpha r', \beta r')$ n'est pas toujours log-concave suivant les valeurs de $\alpha > 0$ et $\beta > 0$. Il n'est alors pas possible de démontrer l'unicité du minimum par des arguments de concavité. Le cas général est encore plus difficile à démontrer. Plusieurs réalisations de l'opposée de la log-vraisemblance à minimiser en fonction de r' sont représentées sur les figures 3.1 et 3.2. Ces réalisations sont obtenues pour des échantillons de taille $n = 500$ distribués suivant une LGBMu de paramètres de forme $\mathbf{q} = (q_1, q_2) = (2, 4)$, de moyennes $m_1 = m_2 = 1$ et pour différentes valeurs du coefficient de corrélation normalisé $r' \in [0, 1[$. Pour de grandes valeurs du coefficient de corrélation normalisé telle que $r' = 0.98$, on constate sur la figure 3.2(h) que le critère à minimiser n'est pas concave. Cependant, au vu des différentes réalisations obtenues pour le critère à minimiser, il existe bien un minimum unique situé autour de la valeur théorique et il est clair que tout algorithme de gradient convergera vers cet unique minimum.

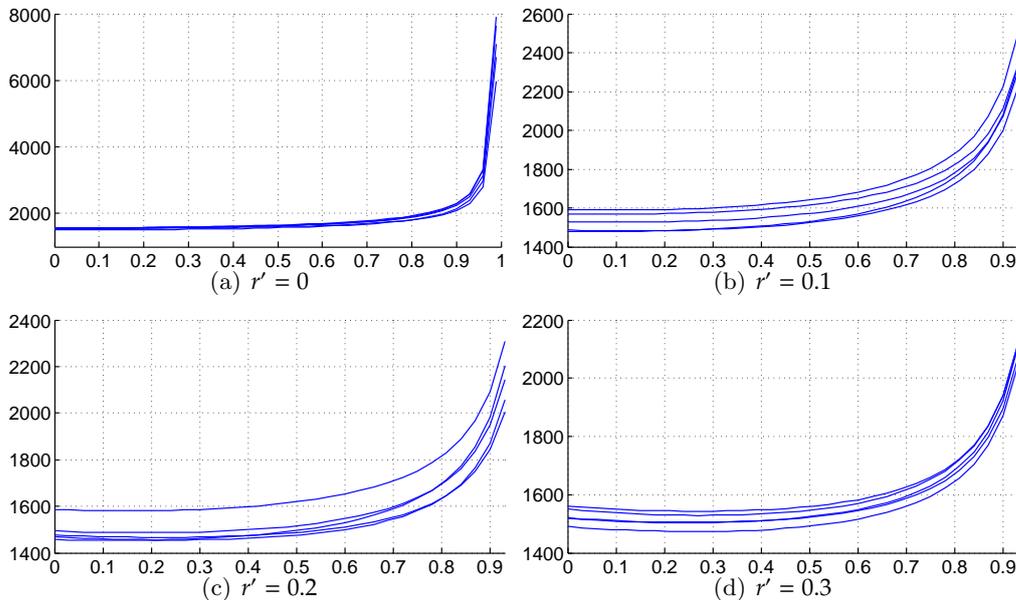


FIG. 3.1 – Courbes typiques de la log-vraisemblance négative en fonction de r' ($q_1 = 2$, $q_2 = 4$, $n = 500$) pour différents échantillons.

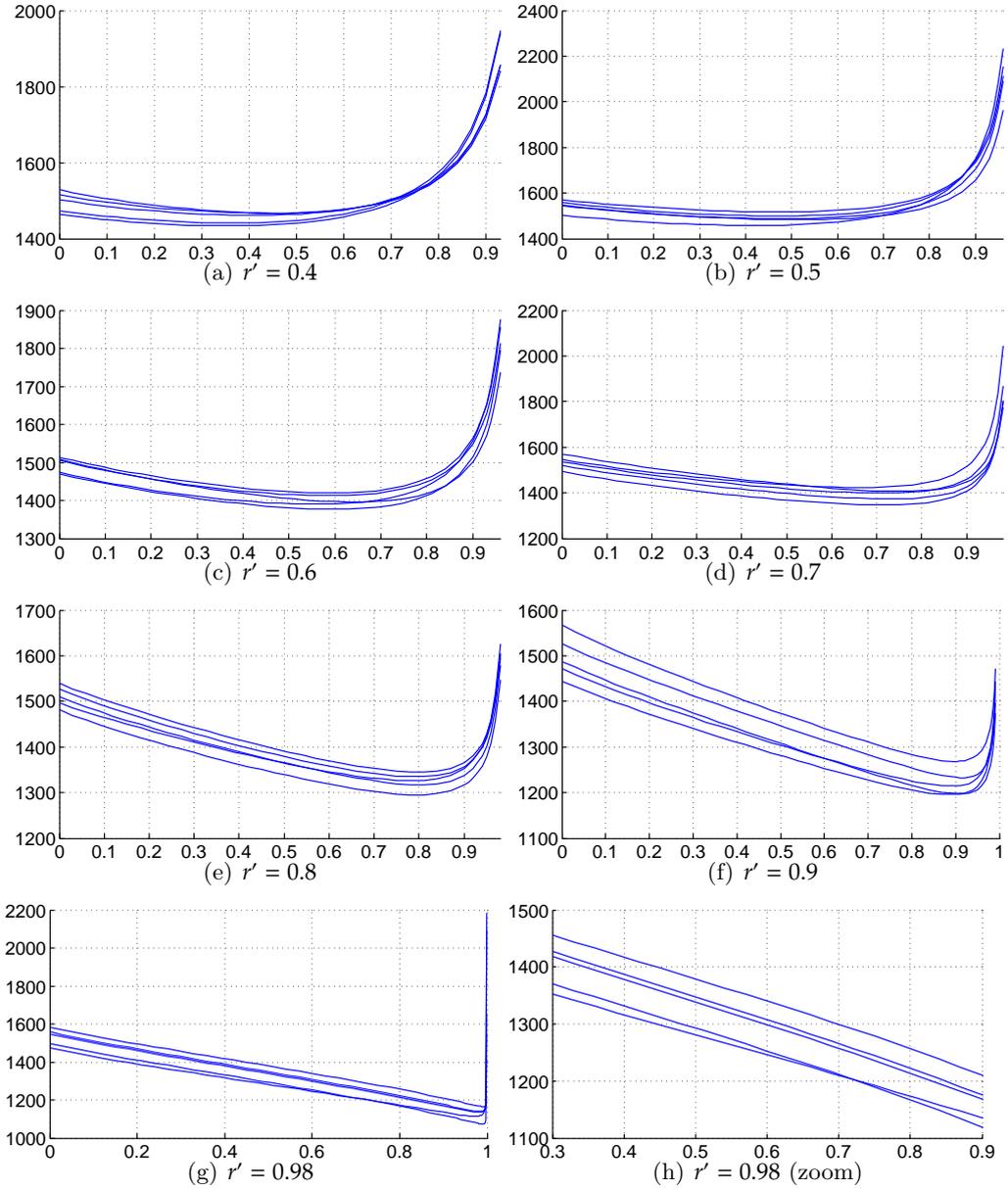


FIG. 3.2 – Courbes typiques de la log-vraisemblance négative en fonction de r' ($q_1 = 2$, $q_2 = 4$, $n = 500$) pour différents échantillons.

Performance

Les propriétés asymptotiques des estimateurs FIM peuvent être obtenues à l'aide du système des fonctions d'inférences $\mathbf{g}(\mathbf{Y}; \boldsymbol{\theta})$ sous les hypothèses de régularité usuelles requises par l'EMV (voir [Joe97] pour plus de précisions). En particulier, l'estimateur FIM de $\boldsymbol{\theta}$ noté $\widehat{\boldsymbol{\theta}}_{\text{FIM}}$ est tel que $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{FIM}} - \boldsymbol{\theta})$ converge en loi vers la loi normale $\mathcal{N}(\mathbf{0}, \mathbf{V})$, où la matrice de covariance asymptotique \mathbf{V} est la matrice d'information de Godambe inverse, définie par :

$$\mathbf{V} = \mathbf{D}_g^{-1} \mathbf{M}_g \mathbf{D}_g^{-T} \quad (3.19)$$

où

$$D_g = \mathbb{E}[\partial g(\mathbf{Y}; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}], \quad M_g = \mathbb{E}[g(\mathbf{Y}; \boldsymbol{\theta}) g^T(\mathbf{Y}; \boldsymbol{\theta})].$$

Des calculs directs mènent aux expressions suivantes pour les matrices D_g et M_g [Joe05] :

$$D_g = \begin{pmatrix} \mathcal{J}_{11} & 0 & 0 \\ 0 & \mathcal{J}_{22} & 0 \\ \mathcal{I}_{13} & \mathcal{I}_{23} & \mathcal{I}_{33} \end{pmatrix}, \quad M_g = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} & 0 \\ \mathcal{J}_{12} & \mathcal{J}_{22} & 0 \\ 0 & 0 & \mathcal{I}_{33} \end{pmatrix}$$

où

- \mathcal{I}_{ij} sont les entrées de la matrice d'information de Fisher, $\mathcal{I} = (\mathcal{I}_{ij})_{1 \leq i, j \leq 3}$,
- \mathcal{J}_{11} et \mathcal{J}_{22} sont les informations de Fisher associées aux marges Y_1 et Y_2 respectivement,
- $\mathcal{J}_{12} = \mathbb{E}[g_1(\mathbf{Y}; \boldsymbol{\theta}) g_2(\mathbf{Y}; \boldsymbol{\theta})] = \mathbb{E}\left[\frac{\partial l_1(\mathbf{Y}_1; m_1)}{\partial m_1} \frac{\partial l_2(\mathbf{Y}_2; m_2)}{\partial m_2}\right]$.

Les termes \mathcal{J}_{11} , \mathcal{J}_{22} , \mathcal{J}_{12} associés aux LGMMu sont facilement obtenus en considérant les log-vraisemblances univariées $l_1(\mathbf{Y}_1; m_1)$ et $l_2(\mathbf{Y}_2; m_2)$:

$$\mathcal{J}_{11} = \frac{q_1}{m_1^2}, \quad \mathcal{J}_{22} = \frac{q_2}{m_2^2}, \quad \mathcal{J}_{12} = \frac{q_1}{m_1 m_2} r'.$$

Comme expliqué précédemment en 3.2.1, les entrées de la matrice d'information de Fisher n'ont pas d'expression analytique explicite. Aussi, ces termes ont-ils été calculés en utilisant une méthode numérique d'intégration (quadrature de Simpson). Notons que cette méthode de calcul permet de contrôler l'erreur d'approximation, ce qui n'est pas le cas lors de l'estimation par tirages de Monte-Carlo. Les performances asymptotiques obtenues pour cette méthode d'estimation seront comparées aux bornes de Cramer-Rao à la prochaine section.

3.2.3 Méthode de Moments

Les estimateurs de (m_1, m_2, r) étudiés dans ce chapitre sont comparés aux estimateurs standards basés sur la méthode des moments

$$\widehat{m}_{1\text{Mo}} = \bar{X}_1, \quad \widehat{m}_{2\text{Mo}} = \bar{X}_2, \tag{3.20}$$

$$\widehat{r}_{\text{Mo}} = \frac{\sum_{i=1}^n (X_1^i - \bar{X}_1)(X_2^i - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_1^i - \bar{X}_1)^2} \sqrt{\sum_{i=1}^n (X_2^i - \bar{X}_2)^2}}. \tag{3.21}$$

On peut remarquer que ces estimateurs ne dépendent pas de la loi du couple (Y_1, Y_2) , et restent convergents même si le vecteur (Y_1, Y_2) n'est pas distribué suivant une LGBMu. On peut noter que ces estimateurs correspondent aux estimateurs obtenus par la méthode des moments étudiés au chapitre précédent.

Les performances asymptotiques de ces estimateurs $\hat{\boldsymbol{\theta}}_{\text{Mo}} = (\widehat{m}_{1\text{Mo}}, \widehat{m}_{2\text{Mo}}, \widehat{r}_{\text{Mo}})^T$ peuvent être obtenues lorsque le vecteur (Y_1, Y_2) suit une LGBMu en imitant les résultats établis à la section 2.2.2 du chapitre précédent.

3.3 Données synthétiques

D'après la définition 3.1.1 donnée à la page 78, un vecteur \mathbf{Y} distribué suivant une loi GBMu peut être généré en ajoutant une variable aléatoire Z de loi gamma univariée à un vecteur \mathbf{X} distribué suivant une loi GBMo. La génération du vecteur \mathbf{X} a été décrite dans l'algorithme 1 présenté page 27. Toutes les lois GBMu vérifiant les conditions (3.4) peuvent ainsi être générées. À partir des échantillons ainsi générés, les performances des estimateurs obtenus peuvent être comparées.

3.3.1 Méthodes du maximum de vraisemblance et des moments

Les premières simulations comparent les performances de la méthode du maximum de vraisemblance avec la méthode des moments en fonction de la taille n de l'échantillon. Notons que les valeurs possibles de n sont $n = (2p+1) \times (2p+1)$ où $p \in \mathbb{N}$. Ces valeurs correspondent à la taille des fenêtres qui seront utilisées pour la détection de changements, comme nous l'expliquerons dans le paragraphe suivant. Le nombre de simulations de Monte Carlo est 10000 pour toutes les figures présentées dans ce paragraphe. Les autres paramètres pour ce premier exemple sont $m_1 = 100$, $m_2 = 100$, $q_1 = 1$ (nombre de vues associé à la première image) et $q_2 = 2$ (nombre de vues associé à la deuxième image). Les figures 3.3, 3.4 et 3.5 montrent les erreurs quadratiques moyennes (EQM) de l'estimateur du coefficient de corrélation normalisé pour différents valeurs de r' ($r' = 0.2$, $r' = 0.5$ et $r' = 0.8$). Les courbes obtenues sont représentées pour une échelle log-log (log étant ici le logarithme décimal). Les courbes \diamond correspondent à l'estimateur des moments alors que les courbes Δ correspondent à l'estimateur du maximum de vraisemblance.

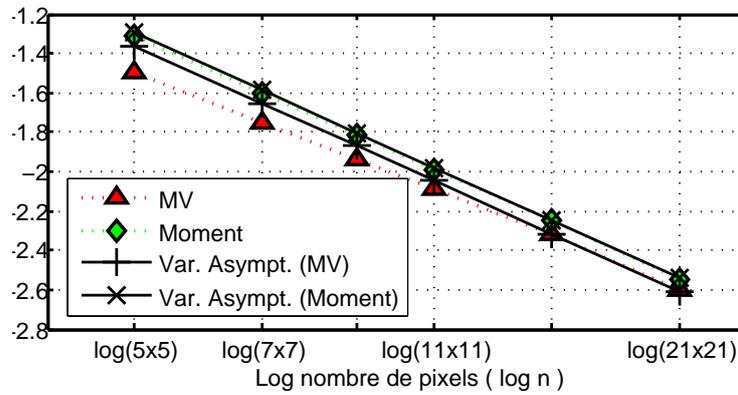


FIG. 3.3 – log EQM en fonction de log n pour le paramètre r ($r' = 0.2$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$).

Ces figures montrent que l'estimateur du maximum de vraisemblance est plus performant pour ce problème que l'estimateur des moments, en particulier pour les valeurs élevées du coefficient de corrélation normalisé r' . Remarquons que les erreurs quadratiques moyennes asymptotiques des deux estimateurs sont aussi représentées (lignes continues). Les expressions théoriques des erreurs quadratiques moyennes asymptotiques sont en accord avec les résultats obtenus par simulation, même pour de faibles valeurs de n . Finalement, ces figures montrent que des estimations “fiables” de r' peuvent être obtenues pour des valeurs de n supérieures à 9×9 , i.e. même pour

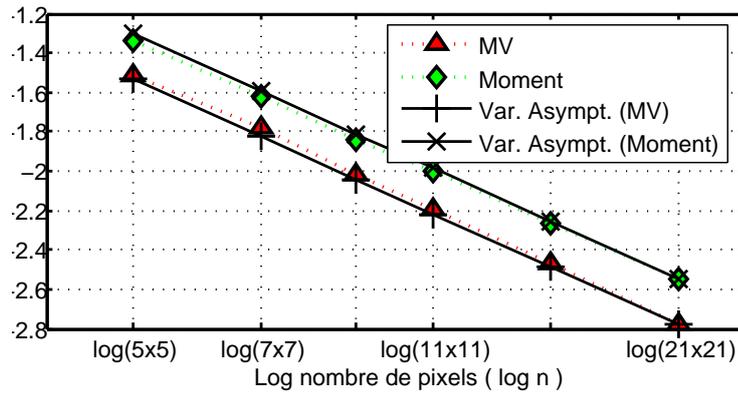


FIG. 3.4 – log EQM en fonction de log n pour le paramètre r ($r' = 0.5$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$).

des fenêtres relativement petites.

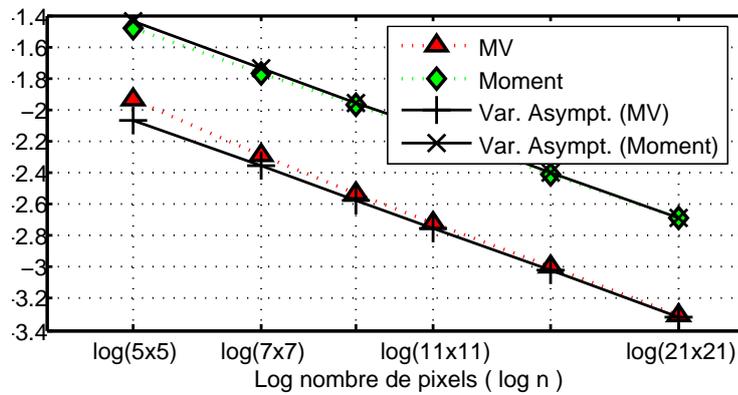


FIG. 3.5 – log EQM en fonction de log n pour le paramètre r ($r' = 0.8$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$).

Les figures 3.6 et 3.7 comparent les erreurs quadratiques moyennes obtenues sur l'estimation de la moyenne m_1 par maximum de vraisemblance et par la méthode des moments pour deux valeurs de r' ($r' = 0.8$ et $r' = 0.9$). Les deux estimateurs affichent des performances similaires sur ce paramètre, l'estimateur du maximum de vraisemblance étant légèrement plus performant pour les valeurs élevées de r' . Notons que les estimateurs de m_2 obtenus par les méthodes de maximum de vraisemblance et de moments sont identiques. Par conséquent, les erreurs quadratiques moyennes correspondantes ne sont pas présentées dans ce chapitre.

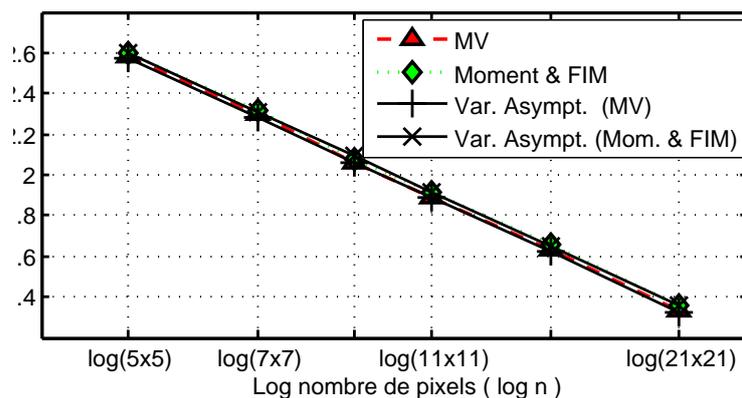


FIG. 3.6 – \log EQM en fonction de $\log n$ pour le paramètre m_1 ($r' = 0.8$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$).

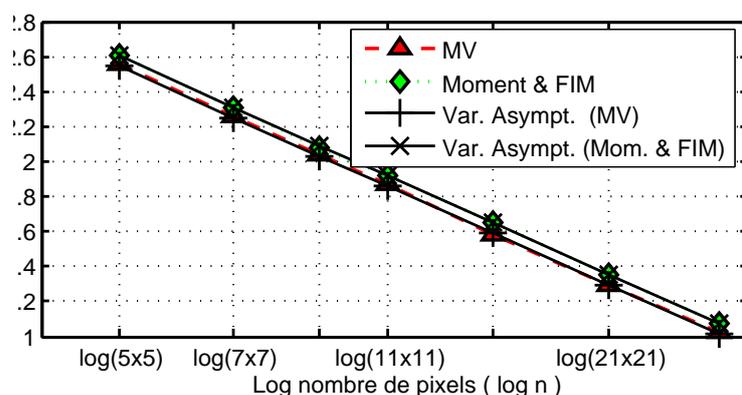


FIG. 3.7 – \log EQM en fonction de $\log n$ pour le paramètre m_1 ($r' = 0.9$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$).

3.3.2 Fonctions d'inférence par les marges (FIM) et méthode du maximum de vraisemblance (MV)

Ce paragraphe compare les performances des estimateurs MV et FIM des paramètres r et m_1 . La figure 3.8 illustre les performances asymptotiques des deux estimateurs en représentant le rapport de leurs variances asymptotiques, appelé rapport d'efficacité asymptotique (REA), en fonction du paramètre r' . Cette figure montre que les estimateurs MV et FIM ont des performances quasi-identiques lorsque r' n'est pas trop proche de 1. Ce résultat est confirmé à la figure 3.9 qui montre les erreurs quadratiques moyennes obtenues pour les deux méthodes (notées dans les légendes *ML* pour le maximum de vraisemblance et *IFM* pour la méthode FIM) pour différentes valeurs de la taille de l'échantillon n (les autres paramètres de cette simulation sont $q_1 = 1$, $q_2 = 2$, $m_1 = 100$, $m_2 = 100$ et $r' = 0.9$).

La figure 3.8 montre aussi que les performances asymptotiques des estimateurs MV et FIM pour le paramètre m_1 diffèrent de manière significative lorsque r' tend vers 1. Cependant, ce

FIG. 3.8 – Rapport d'efficacité asymptotique (REA) pour les estimateurs de m_1 et r ($q_1 = 1$, $q_2 = 5$, $m_1 = 1$ et $m_2 = 1$) en fonction de r' .

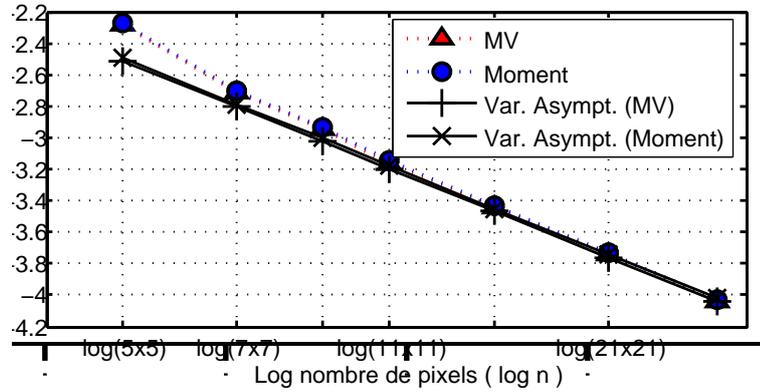
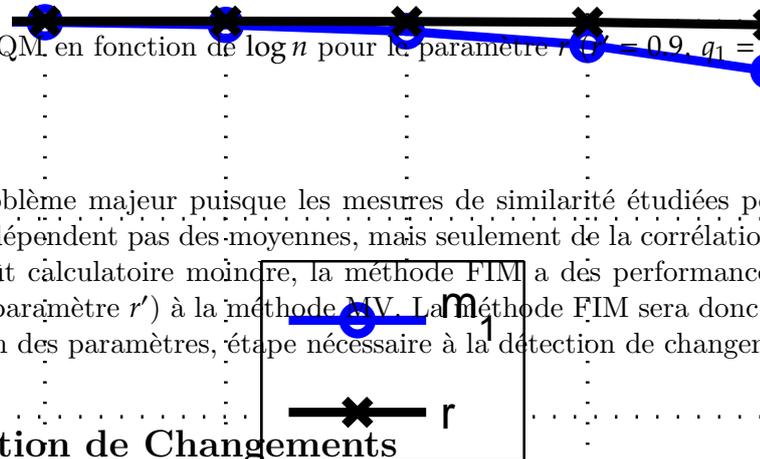


FIG. 3.9 – \log EQM en fonction de $\log n$ pour le paramètre $r = 0.9$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$ et $m_2 = 100$.



n'est pas un problème majeur puisque les mesures de similarité étudiées pour la détection de changement ne dépendent pas des moyennes, mais seulement de la corrélation r' . Il ressort donc que pour un coût calculatoire moindre, la méthode FIM a des performances identiques (pour l'estimation du paramètre r') à la méthode MV. La méthode FIM sera donc choisie par la suite pour l'estimation des paramètres, étape nécessaire à la détection de changements.

3.4 Détection de Changements

3.4.1 Images synthétiques

On travaille dans ce paragraphe sur des vecteurs synthétiques $\mathbf{x} = (x_1, x_2)^T$ (provenant de deux images synthétiques de taille 762×292) distribués suivant des lois GBMu dont les valeurs du coefficient de corrélation normalisé $r' = 0.3$ et $r' = 0.7$ modélisent respectivement la présence et l'absence de changements entre les images. Le coefficient de corrélation r de chaque vecteur bivarié $\mathbf{x}^{(i,j)} = (x_1^{(i,j)}, x_2^{(i,j)})^T$ (pour $1 \leq i \leq 762$, $1 \leq j \leq 292$) est estimé localement à partir des pixels appartenant aux deux fenêtres de taille $n = (2p + 1) \times (2p + 1)$ centrées sur les pixels de coordonnées (i, j) des deux images étudiées.

Comme dans le cas mono-capteur (voir section 2.2.5), lorsque les nombres de vues associés à chaque image sont connus (ou égaux), le problème de détection de changement dans des images RSO multi-capteurs se résume alors au test d'hypothèses binaires suivant :

$$\begin{aligned} \mathcal{H}_0 \quad (\text{absence de changement}) : \quad & r > \lambda, \\ \mathcal{H}_1 \quad (\text{présence de changement}) : \quad & r \leq \lambda, \end{aligned} \tag{3.22}$$

où λ est un seuil et r est le coefficient de corrélation. Les courbes COR associées à ce problème de détection de changements obtenues pour les différents estimateurs de r sont représentées figures

3.10(a), 3.10(b) et 3.10(c) pour trois valeurs représentatives du couple (q_1, q_2) , et deux tailles de fenêtre d'estimation $n = (9 \times 9)$ et $n = (21 \times 21)$.

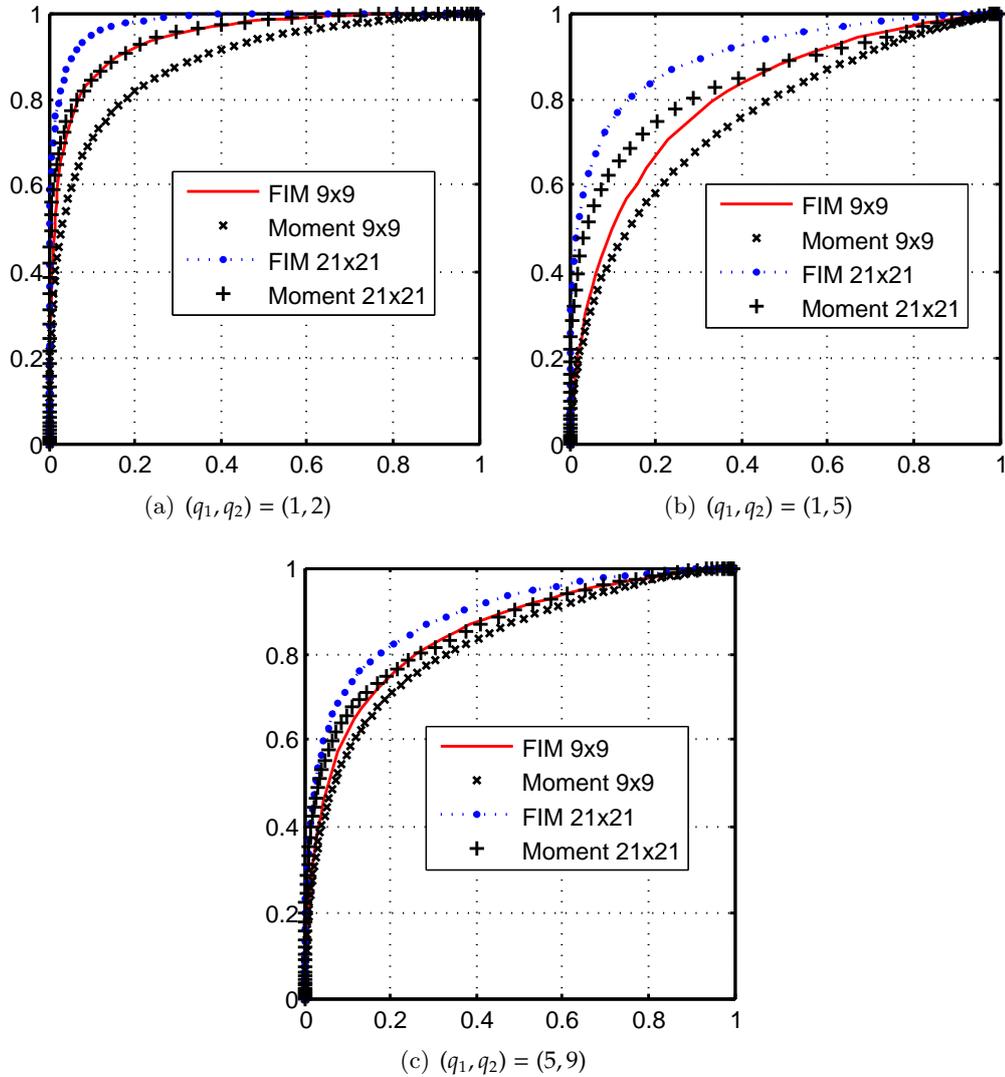


FIG. 3.10 – Courbes COR obtenues à partir des données synthétiques pour différents couples de paramètres de formes (q_1, q_2) et pour différentes tailles de fenêtre ($n = 9 \times 9$ ou $n = 21 \times 21$) .

L'estimateur FIM se révèle être bien plus performant que l'estimateur des moments sur ces exemples. Les figures 3.10(a) et 3.10(b) montrent aussi que les performances de détection semblent décroître lorsque $q_2 - q_1$ augmente, i.e. quand la différence entre les nombres de vues des deux images augmente. Afin de confirmer cette observation, une expression théorique des courbes COR asymptotiques a été calculée. Cette expression repose sur la normalité asymptotique du coefficient de corrélation. Dans le cas de la détection de changements, en notant r_0 et r_1 les vraies valeurs du coefficient de corrélation sous les hypothèses H_0 et H_1 , on obtient les résultats

suivants :

$$\begin{aligned} P_D &= P[\widehat{r} < \lambda \mid \mathcal{H}_1 \text{ est vraie}] = P[\widehat{r} < \lambda \mid \widehat{r} \sim \mathcal{N}(r_1, \sigma_1^2)], \\ P_{FA} &= P[\widehat{r} < \lambda \mid \mathcal{H}_0 \text{ est vraie}] = P[\widehat{r} < \lambda \mid \widehat{r} \sim \mathcal{N}(r_0, \sigma_0^2)], \end{aligned}$$

où σ_0^2 et σ_1^2 sont les variances asymptotiques des estimateurs FIM du coefficient de corrélation (calculées à partir de l'inverse de la matrice d'information de Godambe définie en (3.19)). En notant $\Phi(x)$ la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$, on obtient le résultat suivant :

$$P_D = \Phi \left[\frac{r_0 - r_1}{\sigma_1} + \frac{\sigma_0}{\sigma_1} \Phi^{-1}(P_{FA}) \right]. \quad (3.23)$$

Ce résultat fournit une expression théorique asymptotique des courbes COR associées au problème de détection (3.22) et permet d'analyser les performances de détection en fonction des paramètres de forme de la loi GBMu. Par exemple, la figure 3.11 représente la probabilité de détection P_D en fonction de q_1 et q_2 , pour une probabilité de fausse alarme donnée $P_{FA} = 0.3$. Cette figure confirme que la probabilité de détection pour une probabilité de fausse alarme donnée est une fonction décroissante du paramètre $q_2 - q_1$.

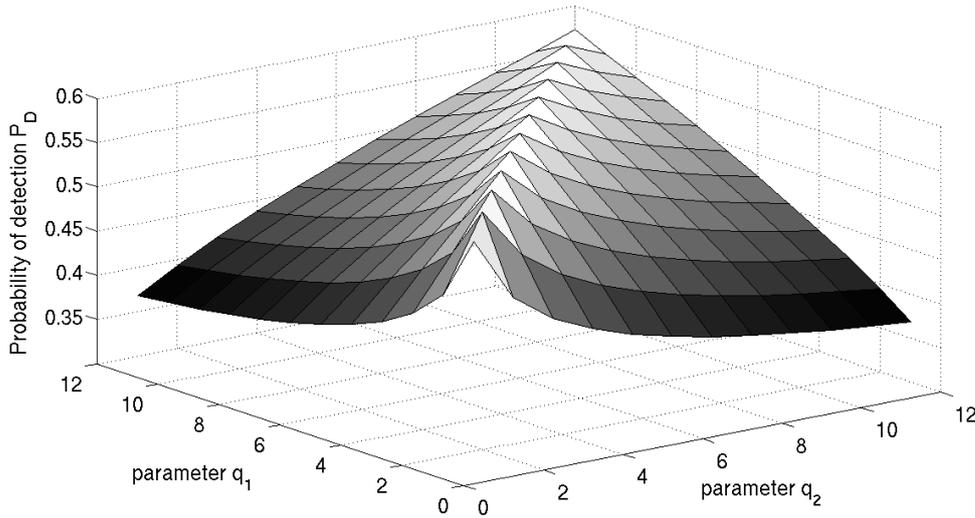


FIG. 3.11 – P_D en fonction des paramètres de forme q_1 et q_2 , ($P_{FA} = 0.3$, $n = 1$).

3.4.2 Images Réelles

On étudie dans ce paragraphe des images acquises à différentes dates à côté de Gloucester (Angleterre) avant et après une inondation (le 9 sept. 2000 et le 21 oct. 2000 respectivement). Ces images ainsi que la vérité terrain qui indique les changements survenus ont été présentées à la figure 2.14 du chapitre précédent (voir page 72). Les images originales, qui sont monovues, ont été transformées en des images ayant un nombre de vues plus élevé en remplaçant les pixels par leur moyenne sur un voisinage donné. Sous l'hypothèse que les distributions des pixels voisins sont indépendantes et de lois Gamma identiques, la moyenne de ces pixels est une loi Gamma dont le paramètre de forme est la somme des paramètres de forme des pixels. Afin d'obtenir un nombre de vues q_2 égal à 5 pour l'image secondaire, les 5 pixels voisins appartenant à chaque clique,

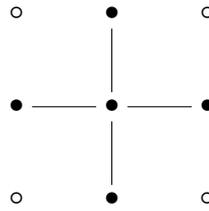


FIG. 3.12 – clique de 5 pixels

dont la forme est représentée à la figure 3.4.2, de l'image ont été moyennés. L'image secondaire étant en théorie monovue, on trouve donc que le nombre de vue de l'image obtenue est 5 ce qui correspond au paramètre de forme $q_2 = 5$. Il faut noter que ce moyennage introduit des corrélations spatiales dans l'image. Aussi afin de conserver l'hypothèse d'indépendance spatiale entre les pixels des images à traiter (hypothèse nécessaire pour que les échantillons à partir desquels sont réalisées les estimations ne contiennent que des réalisations indépendantes), les images ont été sous-échantillonnées suivant un pas de 4 en abscisse et ordonnée. On obtient finalement, les images de taille 762×292 présentées à la figure 3.13.

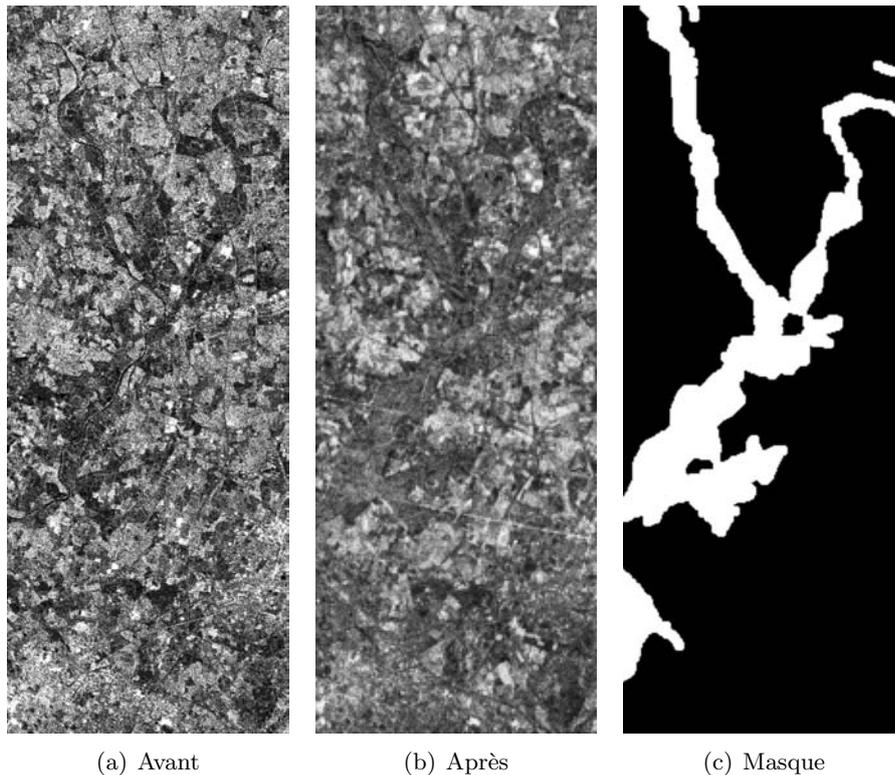


FIG. 3.13 – Images Radarsat de Gloucester avant et après inondation.

Les différents détecteurs étudiés sur ces images sont :

- le détecteur du rapport des moyennes (noté “*ratio edge*” dans les légendes). Ce détecteur a été présenté à la section 2.2.5 (p. 70) du chapitre précédent, et compare les rapports des moyennes entre les images.
- le détecteur de changements basé sur la corrélation, où \widehat{r} in (2.25) a été estimé grâce à

l'estimateur des moments (noté "Correlation Moment" dans les légendes),

- le détecteur de changements basé sur la corrélation, où \widehat{r} in (2.25) a été estimé grâce à la méthode FIM pour des lois GBMu (noté "Correlation IFM" dans les légendes).

Les deux premiers détecteurs sont identiques à ceux proposés au chapitre précédent lorsque les images ont le même nombre de vue. Les courbes COR obtenues pour ce problème de détection de changement sont représentées sur les figures 3.14(a), 3.14(b) et 3.14(c) pour différentes tailles de fenêtres ($n = 9 \times 9$, $n = 15 \times 15$ et $n = 21 \times 21$). Le nombre de vues associées aux deux images sont $q_1 = 1$ pour l'image avant inondation et $q_2 = 5$ pour l'image après inondation. Il apparaît clairement que le détecteur basé sur la corrélation estimée par la méthode FIM fournit les meilleurs performances de détection.

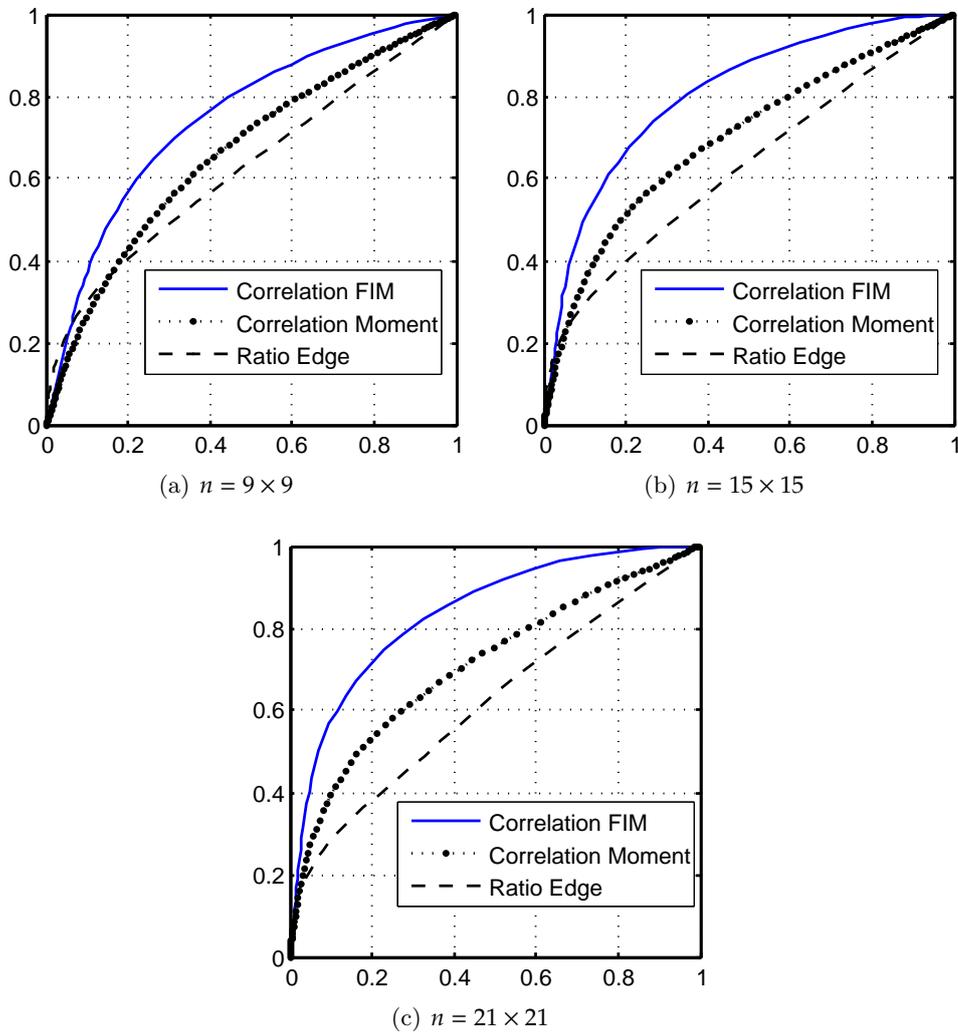


FIG. 3.14 – Courbes COR pour les images de Gloucester ($q_1 = 1$, $q_2 = 5$).

3.5 Comparaison des détecteurs mono et multi-capteurs

Afin d'apprécier l'intérêt des lois GBMu pour des images multi-capteurs, les performances d'estimation et de détection associées à ces lois sont comparées avec celles obtenues lorsque l'on modélise des données multi-capteurs par une loi Gamma bivariée (LGM). Les LGM étudiées au chapitre précédent seront appelées *lois gamma multivariées mono-capteur* et notées LGBMo dans cette section et la suivante par opposition aux lois multi-capteurs. Les lois mono-capteur possèdent l'avantage d'être moins lourdes au niveau calculatoire, puisque leur densité ne fait pas intervenir la fonction de Horn Φ_3 mais uniquement une fonction de Bessel modifiée (1.33). Même si le modèle théorique ne correspond pas à ces lois GBMo, il est intéressant de comparer leurs performances avec les lois GBMu. Cette comparaison permet alors d'apprécier les performances et la robustesse des procédures d'estimation et de détection associées aux lois GBMu et donc leur intérêt par rapport au modèle plus simple des lois GBMo.

3.5.1 Estimation à partir de vecteurs synthétiques

Les figures 3.15 et 3.16 représentent les EQM obtenues pour l'estimation du coefficient de corrélation r à partir d'un échantillon de n réalisations d'un vecteur distribué selon une loi GBMu, pour deux valeurs du coefficient de corrélation normalisé $r' = 0.3$ et $r' = 0.7$. Dans la figure 3.15, les nombres de vues associées à la loi GBMu de l'échantillon sont $q_1 = 1$ et $q_2 = 2$. Les EQM obtenues avec la loi GBMu correspondante (notée "Mu" dans les légendes) et avec les lois GBMo dont le nombre de vues est $q = 1$ et $q = 2$ (notées "Mo 1", "Mo 2" respectivement) sont comparées. Dans la figure 3.16, les nombres de vues associées à la loi GBMu de l'échantillon sont $q_1 = 1$ et $q_2 = 5$. Les EQM obtenues avec la loi GBMu correspondante (notée "Mu" dans les légendes) et avec les lois GBMo dont les nombres de vues sont $q = 1$, $q = 3$ et $q = 5$ (notées "Mo 1", "Mo 3", "Mo 5" respectivement) sont comparées. Les autres paramètres sont $m_1 = 100$, $m_2 = 100$ et le nombre de simulations est égal à 10000.

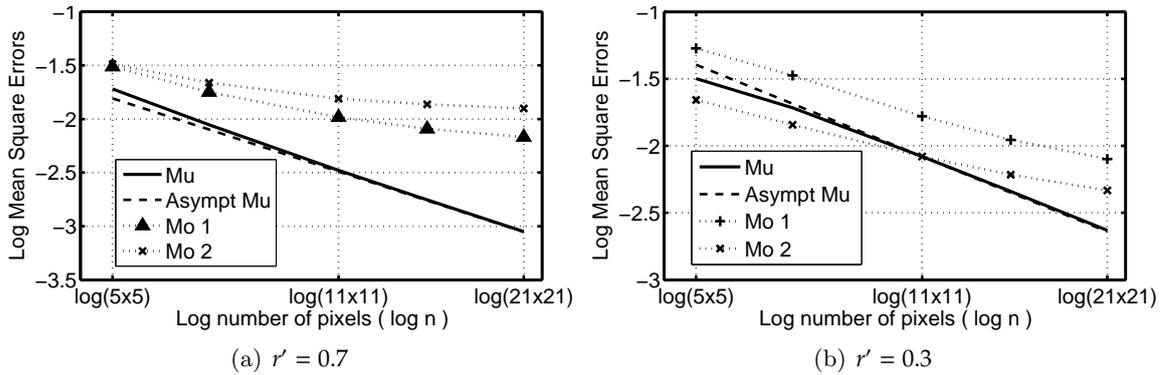
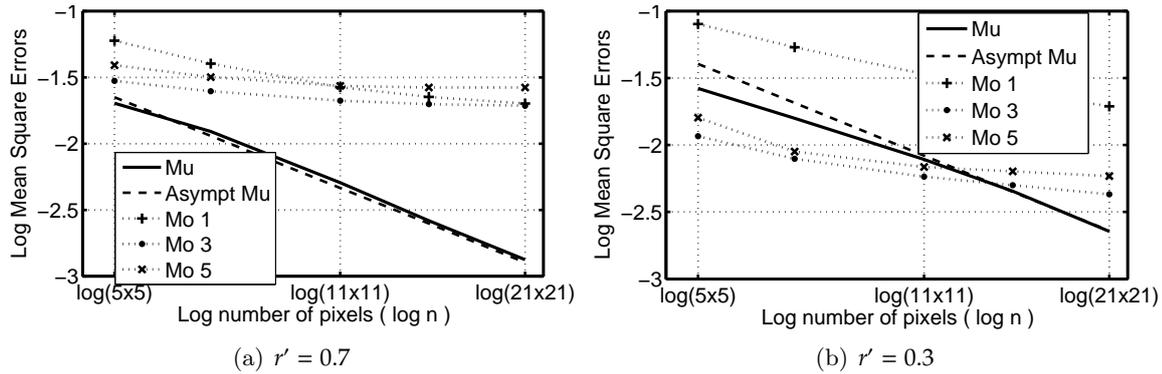


FIG. 3.15 – log EQM en fonction de $\log(n)$ pour le paramètre r' ($q_1 = 1$, $q_2 = 2$)

À part dans quelques cas particuliers (petites tailles d'échantillon n et faibles valeurs du coefficient de corrélation), le modèle basé sur les lois GBMu fournit les meilleures performances.


 FIG. 3.16 – log EQM en fonction de $\log(n)$ pour le paramètre r' ($q_1 = 1, q_2 = 5$)

3.5.2 Détection de changements à l'aide d'images synthétiques et réelles

Lors de l'estimation des paramètres à partir d'images (synthétiques ou réelles), l'homogénéité des données n'est pas vérifiée. En effet les fenêtres d'estimations peuvent contenir des mélanges de pixels correspondant à des zones avec et sans changement. Ce phénomène pose alors un problème de robustesse des méthodes d'estimations. Les courbes COR présentées figure 3.17 comparent les performances de détection obtenues pour les modèles de lois GBMu et GBMo sur des images synthétiques multi-capteurs du même type que celles présentées au paragraphe 3.4.1, pour des fenêtres d'estimation de taille $n = 15 \times 15$ et différents jeux de nombres de vues ((1,2), (1,5) et (5,9)). Ces simulations montrent que les lois GBMu sont moins robustes à ces problèmes d'inhomogénéité que certaines lois GBMo. En effet les meilleures performances de détection sont atteintes pour des lois GBMo dont les nombres de vue sont sous-estimés (en prenant le plus petit des nombres de vue associés aux deux capteurs).

Lorsque la taille des fenêtres d'estimation est faible, le nombre de fenêtres contenant des mélanges de pixels inhomogènes (pixels affectés par le changement ou non) diminue. La robustesse aux inhomogénéités est alors moins cruciale. Les courbes COR obtenues sur les mêmes images synthétiques que précédemment mais pour une taille de fenêtre d'estimation $n = 5 \times 5$ sont présentées à la figure 3.18. Les meilleures performances de détection sont alors atteintes par les lois GBMu.

Les résultats obtenus sur images synthétiques sont confirmés sur données réelles, comme le montrent les courbes COR présentées aux figures 3.19 obtenues pour différentes tailles de fenêtre d'estimation ($n = 5 \times 5, n = 9 \times 9, n = 15 \times 15$) à partir des images ERS de Gloucester présentées à la figure 2.14, p. 72. Les nombres de vue théoriques de ces images sont $q_1 = 1, q_2 = 5$. Les performances obtenues pour la loi GBMu avec $(q_1, q_2) = (1, 5)$ (notée "Mu") sont comparées aux performances obtenues pour des lois GBMo dont les nombres de vue sont 1, 3 et 5 (notées "Mo 1", "Mo 3", "Mo 5" respectivement). Il apparaît sur ces figures que les lois GBMu ont parfois de moins bonnes performances que des lois de type GBMo. Ceci est probablement dû à un manque de robustesse vis à vis des inhomogénéités, problème évoqué précédemment.

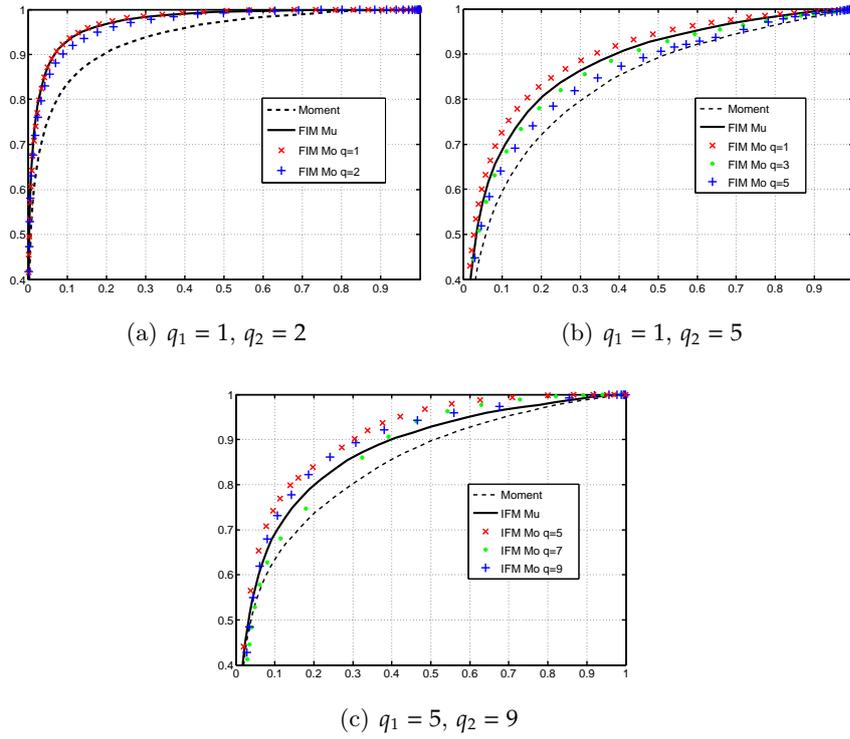


FIG. 3.17 – Courbes COR synthétiques pour une taille de fenêtre d'estimation $n = 15 \times 15$

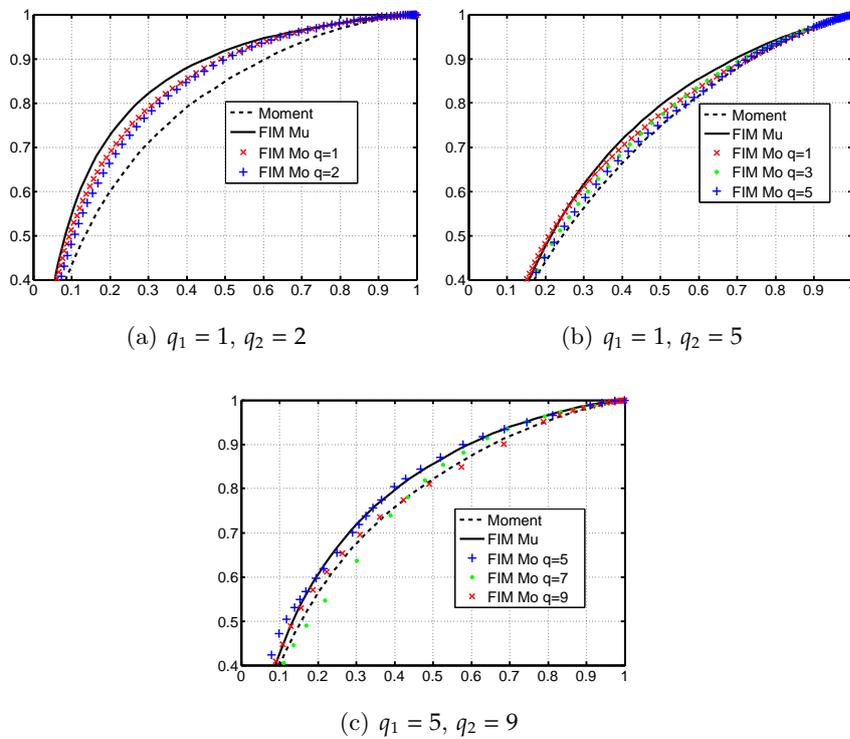
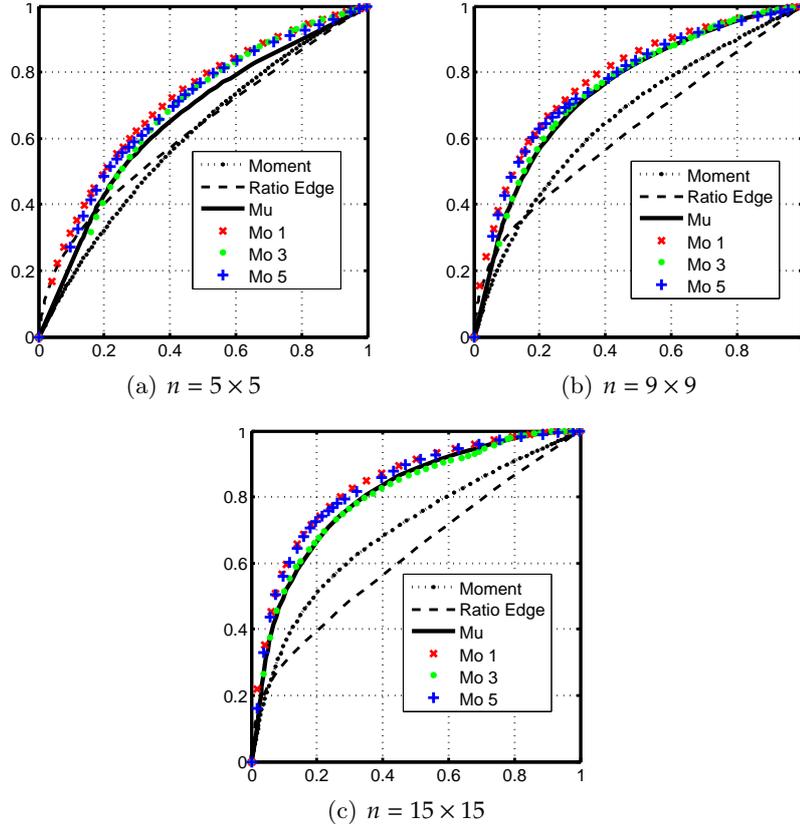


FIG. 3.18 – Courbes COR synthétiques pour une taille de fenêtre d'estimation $n = 5 \times 5$


 FIG. 3.19 – Courbes COR pour les images de Gloucester ($q_1 = 1$, $q_2 = 5$)

3.6 Estimation des nombres de vues équivalents

Les paramètres de forme q_1 et q_2 sont théoriquement connus, car ils correspondent aux nombres de vues de chaque image. Cependant, cette hypothèse n'est pas vraiment valide, car les images à traiter ont subies des pré-traitements afin d'être superposables. Le processus d'orthorectification, qui a été présenté p. 72, produit une modification des statistiques locales des images. Les distributions des pixels sont toujours modélisées par des loi Gamma, mais les paramètres de forme de ces lois peuvent être différents. Les valeurs de ces paramètres de forme sont appelés les nombre de vues équivalents (NVE).

Il est alors intéressant d'étudier les performances obtenues lorsque les NVE, i.e. les paramètres de forme, sont estimés. Dans ce cas le vecteur des paramètres inconnus est $\theta = (m_1, q_1, m_2, q_2, r')^T$. De manière analogue au cas où les paramètres de forme q_1 et q_2 sont connus, les paramètres sont estimés par la méthode d'inférence fonctionnelle par les marges (méthode FIM). On obtient alors les deux étapes suivantes :

1. Estimation des moyennes m_1 et m_2 et des paramètres de forme q_1 et q_2 à partir des lois marginales de Y_1 and Y_2 . Cette estimation est effectuée en maximisant les vraisemblances marginales $l(Y_1; m_1, q_1)$ et $l(Y_2; m_2, q_2)$ par rapport aux couples (m_1, q_1) et (m_2, q_2) respectivement,

2. Estimation du paramètre r' en maximisant la vraisemblance jointe

$$l(\mathbf{Y}; \widehat{m}_{1\text{FIM}}, \widehat{q}_{1\text{FIM}}, \widehat{m}_{2\text{FIM}}, \widehat{q}_{2\text{FIM}}, r')$$

par rapport à r' . Il faut noter que les moyennes m_1 and m_2 et les paramètres de forme q_1 et q_2 ont été remplacés dans la vraisemblance jointe par leurs estimés obtenus dans la première étape de la méthode FIM.

On trouve alors de façon classique à l'aide des propriétés des lois gamma univariées :

$$\widehat{m}_{1\text{FIM}} = \bar{Y}_1, \quad \widehat{m}_{2\text{FIM}} = \bar{Y}_2, \quad (3.24)$$

et $\widehat{q}_{1\text{FIM}}$ et $\widehat{q}_{2\text{FIM}}$ étant obtenus en maximisant les vraisemblances $l(Y_1; \bar{Y}_1, q_1)$ et $l(Y_2; \bar{Y}_2, q_2)$ respectivement. L'estimation des paramètres de forme est cependant délicate et les mesures de similarité dépendent fortement des valeurs estimées de ces paramètres. Dans les simulations suivantes, les paramètres $\widehat{q}_{1\text{FIM}}$ et $\widehat{q}_{2\text{FIM}}$ ont été estimés sur des fenêtres de taille $n = 31 \times 31$ plus grandes que les fenêtres d'estimation utilisées pour les autres paramètres. De plus, la procédure d'évaluation numérique rapide de la fonction de Horn Φ_3 présentée à l'annexe D.2.2 nécessite que les paramètres q_2 et q_1 soient entiers. Les valeurs des estimés de q_1 et q_2 ont donc été arrondies à l'entier non nul le plus proche. Les cartes des nombres de vues équivalents obtenus pour les images de Gloucester sont représentées sur la figure 3.20 Il apparaît déjà clairement sur ces

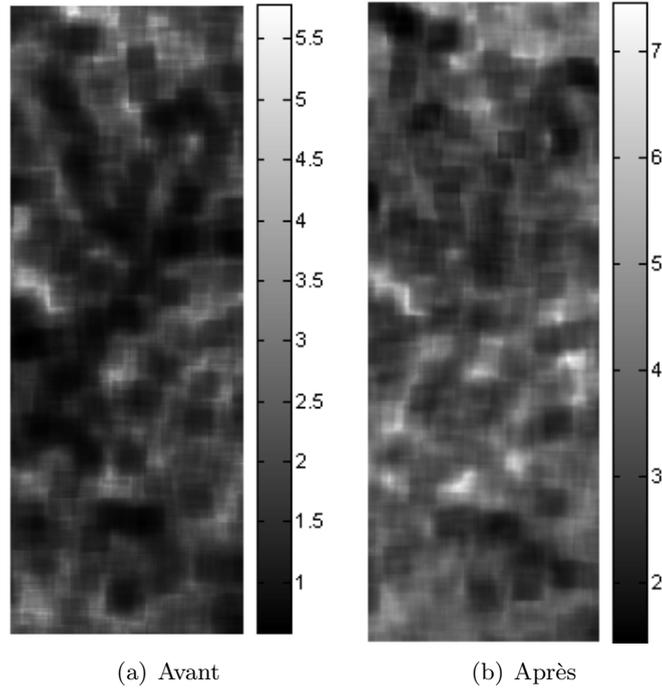


FIG. 3.20 – NVE estimés pour les images de Gloucester ($n = 31 \times 31$)

cartes qu'une information sur les changements survenus entre les images est disponible à travers les différences entre les NVE de chaque image. Ceci montre bien que les mesures de similarité doivent prendre en compte les valeurs de ces NVE. Dans ce cas, la mesure d'indépendance du χ^2 n'est plus une fonction bijective du coefficient de corrélation r , car elle dépend également des paramètres q_1 et q_2 . Par conséquent, le coefficient de corrélation r' et la mesure d'indépendance

du χ^2 sont deux mesures de similarité distinctes pour lesquelles les performances de détection seront différentes. Les cartes de détection obtenues pour les différentes mesures de similarité étudiées lorsque la fenêtre d'estimation est de taille $n = 15 \times 15$ sont présentées sur la figure 3.22. Les cartes obtenues lorsque les NVE sont estimés sont dénommées "Correlation Mu" lorsque la mesure de similarité est le coefficient de corrélation (fig. 3.22(a)), ou " χ^2 Mu" lorsque la mesure de similarité est la mesure de l'indépendance du χ^2 (fig. 3.22(b)). La carte 3.22(c), dénommée "Correlation Mo $q = 1$ " est obtenue à partir du coefficient de corrélation estimé par maximum de vraisemblance en supposant que les images ont le même nombre de vue égal à 1. Il a été vu à la section précédente que les meilleurs résultats sont en effet obtenus pour cette mesure lorsque les nombres de vues sont supposés uniformes sur toute l'image et de valeur fixée. Enfin les cartes obtenues lorsque le coefficient de corrélation est estimé par la méthode des moments et pour la mesure de similarité du "ratio edge" sont présentées respectivement aux figures 3.22(d) et 3.22(e). Il faut noter que la mesure d'indépendance du χ^2 n'étant pas une grandeur bornée

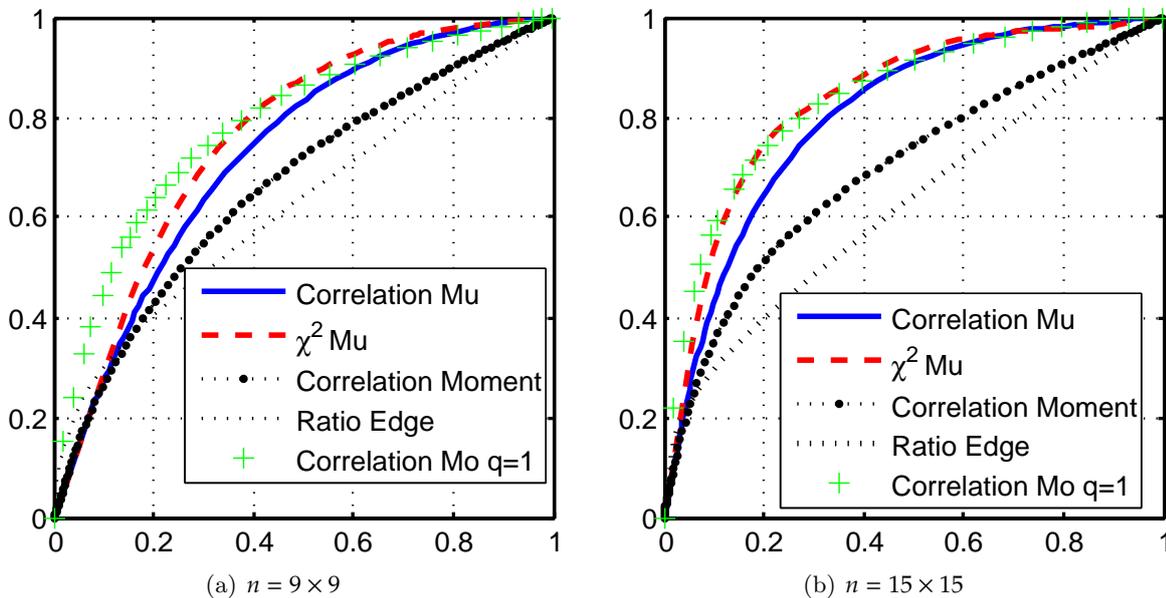


FIG. 3.21 – Courbes COR pour les images de Gloucester

contrairement aux autres mesures, la carte 3.22(b) a été seuillée afin que les contrastes soient perceptibles.

Finalement, les performances de détection associées à ces cartes sont présentées à la figure 3.21 pour deux tailles de fenêtre d'estimation : $n = 9 \times 9$ (fig. 3.21(a)) et $n = 15 \times 15$. (fig. 3.21(b)). Il apparaît que pour les niveaux de fausses alarmes qui nous intéressent, par exemple $0.4 \leq P_{FA} \leq 0.6$, le détecteur basé sur la mesure de l'indépendance du χ^2 lorsque les NVE sont estimés fournit les meilleurs résultats, juste devant la mesure construite à partir du coefficient de corrélation estimés par la méthode MV lorsque les images sont supposées monovues. Ceci souligne l'intérêt du modèle des lois Gamma multivariées multi-capteurs pour la détection de changement en imagerie RSO.

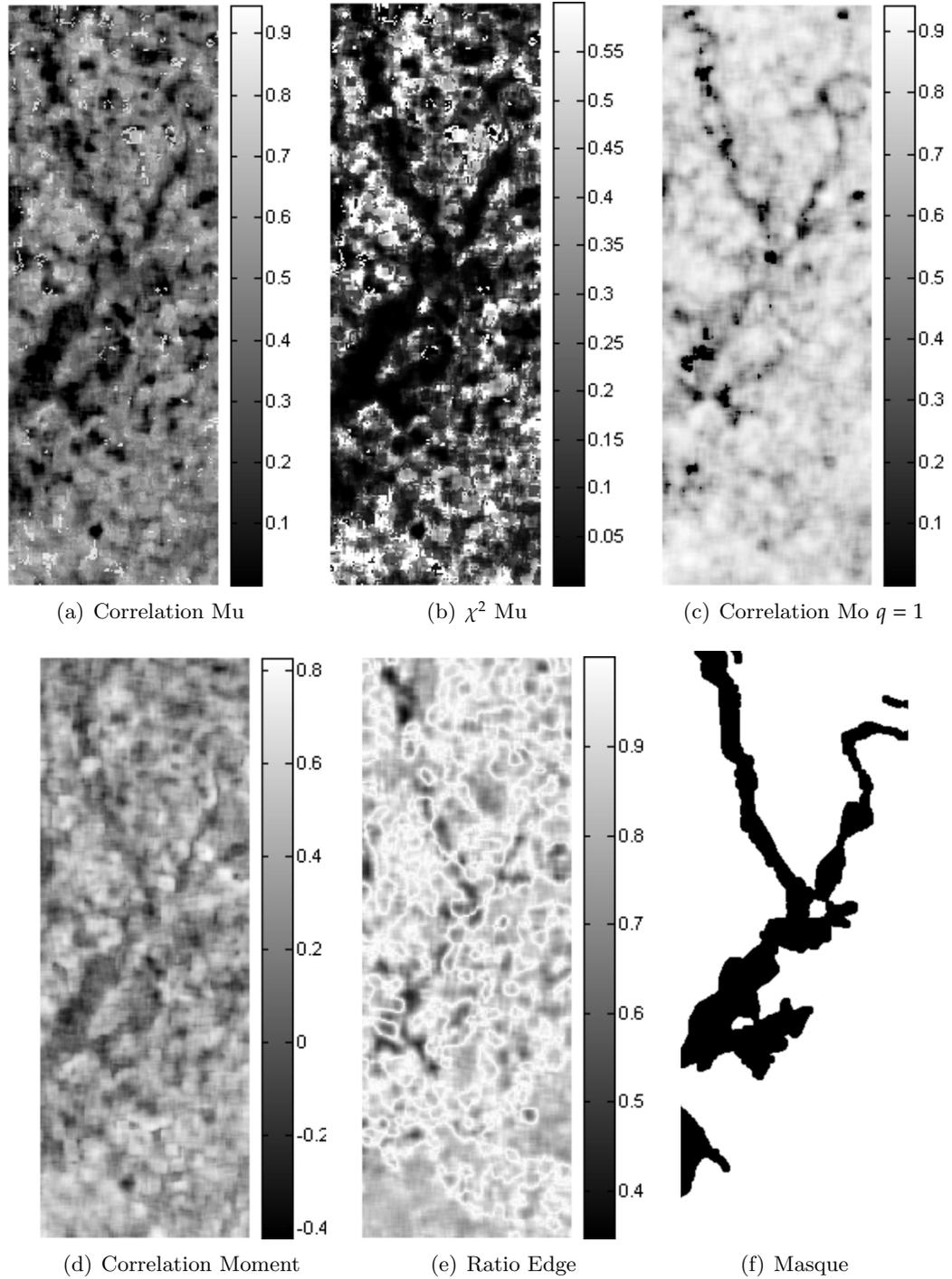


FIG. 3.22 – Cartes de détection pour les différentes mesures de similarités ($n = 15 \times 15$)

3.7 Conclusion

Une nouvelle famille de distributions basée sur des lois gamma multivariées (notées lois GMMu) a été étudiée dans ce chapitre. Ces lois servent à modéliser les images RSO lorsque les nombres de vues associés à ces images sont différents et, définissent ainsi un cadre statistique afin d'étudier les propriétés de ces images. Des algorithmes d'estimation des paramètres de ces lois reposant sur la méthode du maximum de vraisemblance, la méthode d'inférence fonctionnelle par les marges et la méthode des moments ont été étudiés. Cette étude a montré que les estimateurs du coefficient de corrélation de ces lois sont utiles pour la détection de changements sur des images possédant des nombres de vues distincts.

L'estimation des paramètres des lois GMMu s'avère moins robuste aux inhomogénéités présentes dans les fenêtres d'estimation que les lois GMMo, même si ce résultat reste à vérifier sur de "vraies" images multi-capteurs. Les algorithmes de détection de changements basés sur les lois GMMu peuvent toutefois intervenir dans un deuxième temps afin d'améliorer les performances de détection lorsqu'une carte de changement a déjà été obtenue pour une méthode donnée. Dans ce cas, les fenêtres d'estimation pourront être adaptées en fonction de la carte de changement afin de réduire le nombre de fenêtres comportant des pixels affectés et non affectés par un changement. Les mesures de similarités estimées grâce au modèle des lois GMMu pourront permettre d'affiner la carte de détection initiale. Cette famille de loi a en outre le mérite de poser un cadre statistique très précis permettant d'étudier avec finesse les images RSO. Ce modèle pourrait être appliqué, par exemple, à des problèmes de classification lorsqu'une série d'images d'une même scène est disponible.

Enfin, être capable de travailler sur des images dont les nombres de vues sont différents est intéressant, non seulement lorsque les images ont été acquises par des capteurs différents, mais aussi lorsque les capteurs sont identiques ou ont le même nombre de vues. En effet, en raison des pré-traitements nécessaires afin de comparer les images, des différences entre les nombres de vues équivalents de chaque image existent, même si les images ont été acquises par le même capteur et sous un mode d'enregistrement similaire. Le modèle des lois GMMu permet alors de prendre en compte ce phénomène, et permet de plus de mettre en œuvre des stratégies de détection qui ont montré leur efficacité.

Lois Gamma multivariées pour l'imagerie polarimétrique

Sommaire

4.1 Imagerie polarimétrique	107
4.1.1 Principe d'un système imageur en imagerie polarimétrique cohérente	107
4.1.2 Lumière totalement polarisée	109
4.1.3 Lumière partiellement polarisée	109
4.2 Imagerie polarimétrique "fort-flux"	112
4.2.1 Distribution du vecteur des intensités	113
4.2.2 Estimation de la matrice de polarisation et du DdP	114
4.2.3 Estimation sur données synthétiques	122
4.2.4 Estimation à partir d'images synthétiques	126
4.3 Imagerie polarimétrique "faible-flux"	129
4.3.1 Modèle statistique	129
4.3.2 Estimation de la matrice de polarisation et du DdP	131
4.3.3 Estimation sur données synthétiques	136
4.3.4 Estimation à partir d'images synthétiques	139
4.4 Conclusion	141

La plupart des techniques d'imagerie visent à quantifier la réflectivité des objets présents dans le champ de vision du système imageur. Ces systèmes se limitent souvent à l'analyse des niveaux d'intensité lumineuse reçue. C'est en particulier le cas des systèmes d'imagerie RSO étudiées dans les deux chapitres précédents. Pourtant, de nombreuses techniques d'imagerie telles que l'échographie, l'imagerie par résonance magnétique ou justement l'imagerie RSO permettent de mesurer d'autres grandeurs qui ne relèvent pas des seules propriétés radiométriques des matériaux.

L'imagerie polarimétrique consiste à étudier l'état de la polarisation de la lumière reçue par le système qui forme l'image. En effet, l'interaction de la lumière avec les surfaces diffusantes des matériaux imagés influe sur son état de polarisation. Par conséquent, l'état de polarisation de la lumière rétrodiffusée reflète les propriétés polarimétriques des éléments de la scène imagée. L'information alors apportée peut s'avérer très pertinente pour caractériser un objet. Par exemple, les matériaux métalliques dépolarisent moins la lumière que ceux en plastique, et en général les objets manufacturés dépolarisent moins la lumière que le fond naturel (végétation,

...) d'une scène. Aussi, contrairement à l'imagerie d'intensité classique, cette technique permet de distinguer les éléments d'une scène dont les niveaux de réflectivité sont similaires mais dont les propriétés polarimétriques diffèrent. Cette propriété est illustrée sur la figure 4.1 pour une scène composée d'éléments en bois, métal et papier et illuminée par un laser. À cause de la faible qualité de l'illumination, il est très difficile de distinguer les divers éléments, dont les niveaux de réflectivité sont relativement similaires, sur l'image d'intensité 4.1(a). Les différents éléments de la scène apparaissent néanmoins clairement à partir de l'analyse polarimétrique de la scène présentée sur la figure 4.1(b). Pour ces raisons, l'imagerie polarimétrique tend à se répandre dans

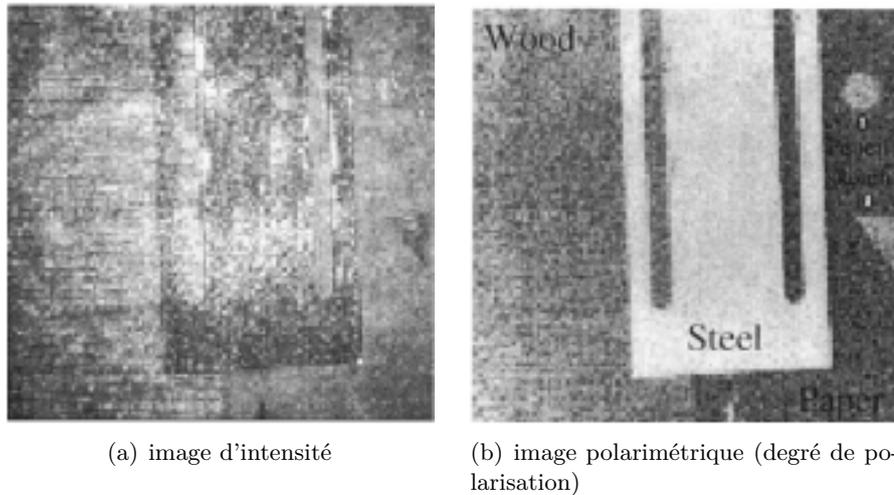


FIG. 4.1 – Images d'une scène composée d'éléments en bois, métal et papier illuminée par un laser[BC00]

de nombreuses applications, en particulier dans les domaines de l'imagerie militaire, médicale [JJN⁺99, JRL02], ou de la vision industrielle [Wol90, Wol94].

Ce chapitre se décompose en trois parties. La première partie expose brièvement le principe de la polarimétrie dans le cas de l'imagerie active et cohérente, qui sera le cadre de notre étude.

La deuxième partie traite le cas de l'imagerie polarimétrique *fort-flux*, i.e. lorsque l'intensité lumineuse de la lumière rétrodiffusée est suffisamment importante pour être mesurée directement.

La dernière partie est consacrée à l'étude de l'imagerie polarimétrique *faible-flux*, i.e. lorsque les niveaux d'intensité de la lumière rétrodiffusée sont très faibles et ne peuvent plus être mesurés directement. Ce cas oblige à considérer un bruit supplémentaire au phénomène de speckle, appelé bruit de Poisson. Ce cas oblige à considérer la théorie semi-classique de l'optique où les mesures correspondent aux nombres de photons détectés par des gamma-caméras. Ces nombres de photons collectés sont alors distribués selon des lois de Poisson mélangées multivariées (LPMM) (voir définition 1.3.5, p. 42), dont la loi de mélange est la loi de l'intensité lumineuse.

4.1 Imagerie polarimétrique

Par opposition à l'imagerie passive où l'on analyse la lumière naturelle diffusée par la scène, l'imagerie active consiste à illuminer la scène. On peut ainsi parfaitement contrôler les propriétés radiométriques et polarimétriques de la lumière incidente. L'analyse de la lumière rétrodiffusée permet alors de déterminer l'influence précise des matériaux pour une lumière incidente donnée. Par ailleurs, on considère dans notre étude que la source d'émission est cohérente, comme c'est le cas avec un laser. L'intérêt du laser, qui émet une onde quasi-monochromatique, est de produire des faisceaux lumineux très directifs, permettant d'imager à longue distance. Sous ces conditions, le champ électrique \mathbf{E} de la lumière peut être modélisé par une onde plane quasi-monochromatique de direction de propagation l'axe \mathbf{z} , qui a pour expression en un point donné :

$$\mathbf{E} = [A_x \mathbf{x} + A_y \mathbf{y}] e^{-2i\pi\nu t}, \quad (4.1)$$

où \mathbf{x} et \mathbf{y} représentent respectivement les vecteurs unitaires des axes horizontaux et verticaux, et ν la fréquence centrale de l'onde. L'état de polarisation de la lumière est caractérisé par les amplitudes et les déphasages respectifs des termes complexes A_x et A_y . Le vecteur $\mathbf{A} = (A_x, A_y) \in \mathbb{C}^2$, appelé le *vecteur de Jones* permet donc de décrire l'état de polarisation de la lumière. Une représentation équivalente de ce vecteur complexe est le vecteur réel $\mathbf{S} = (S_0, S_1, S_2, S_3)^T \in \mathbb{R}^4$, défini comme :

$$\begin{aligned} S_0 &= |A_x|^2 + |A_y|^2, \\ S_1 &= |A_x|^2 - |A_y|^2, \\ S_2 &= 2\Re(A_x A_y^*), \\ S_3 &= 2\Im(A_x A_y^*), \end{aligned} \quad (4.2)$$

où $\Re(z)$ et $\Im(z)$ désignent respectivement la partie réelle et la partie imaginaire du complexe z , et où z^* est le conjugué de z . Le vecteur \mathbf{S} obtenu est appelé le *vecteur instantané de Stokes* [Bro98, p. 102]. Ce vecteur contient toutes les informations sur l'état de polarisation du champ électrique \mathbf{E} défini à l'équation (4.1), exceptée la phase absolue du champ qui est sans intérêt pour l'étude polarimétrique. Par conséquent, la mesure de ce vecteur permet de déterminer l'état de polarisation de la lumière.

4.1.1 Principe d'un système imageur en imagerie polarimétrique cohérente

Le schéma présenté sur la figure 4.2 illustre le principe de l'imagerie polarimétrique cohérente [BC00, Fad06]. La lumière est ici émise par un laser qui illumine la scène d'intérêt. La lumière rétrodiffusée est alors captée par un télescope qui la concentre et la dirige vers des capteurs CCD. Deux ensembles d'analyseurs optiques, composés chacun d'un retardateur (ou déphaseur) et d'un polariseur, sont intercalés sur les chemins optiques des ondes incidentes et rétrodiffusées. Le premier permet de contrôler l'état de polarisation de la lumière incidente. Le second celui de la lumière rétrodiffusée juste avant qu'elle ne soit mesurée par le capteur. Afin de mesurer les propriétés de la lumière rétrodiffusée pour une onde incidente donnée, quatre mesures d'intensités I_1, I_2, I_3, I_4 de la lumière rétrodiffusée sont nécessaires et suffisantes [Bro98, p. 340–341] :

- I_1 est obtenue pour une polarisation suivant la direction horizontale
- I_2 est obtenue pour une polarisation suivant la direction verticale
- I_3 est obtenue pour une polarisation de $\pi/4$ par rapport à l'axe horizontal

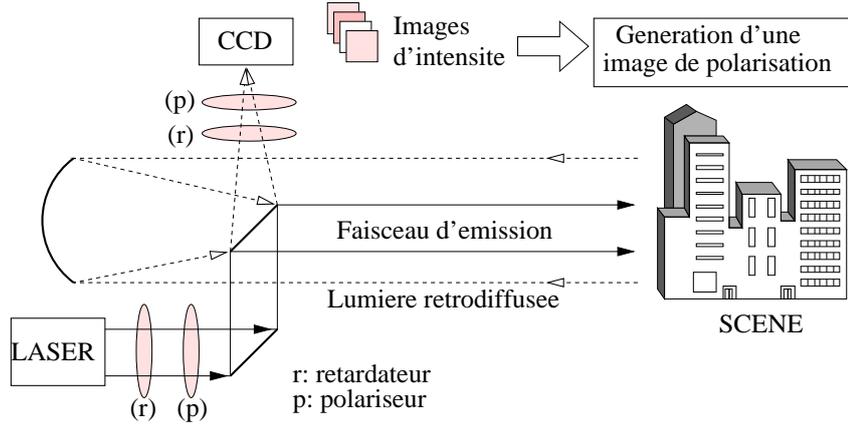


FIG. 4.2 – Principe d'un système imageur pour l'imagerie polarimétrique cohérente [BC00, Fad06].

- I_4 est obtenue en retardant la lumière rétrodiffusée de $\pi/2$, puis pour une polarisation de $\pi/4$ par rapport à l'axe horizontal

Avec le formalisme des vecteurs et des matrices de Jones, les matrices de polarisation sont des matrices dont les termes sont déterministes et qui correspondent à des projections [Hua94] :

$$P_\theta = \begin{pmatrix} \cos^2(\theta) & \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) & \sin^2(\theta) \end{pmatrix},$$

θ étant l'angle entre la direction du polariseur et l'axe horizontal. De même, la matrice de Jones correspondant à un retardateur (ou un déphaseur) quart d'onde (i.e. qui retarde la phase de $\pi/2$) est une matrice déterministe dont l'expression est :

$$M_{\lambda/4} = e^{i\pi/4} \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}.$$

Les intensités I_1, I_2, I_3, I_4 sont alors les normes au carré (pour la norme issue du produit scalaire hermitien : $\langle \mathbf{z}, \mathbf{y} \rangle = \mathbf{z}^\dagger \mathbf{y} \quad \forall \mathbf{z}, \mathbf{y} \in \mathbb{C}^2$ où \mathbf{z}^\dagger est la transposée du conjugué du vecteur \mathbf{z}) des vecteurs de Jones respectifs :

$$\begin{aligned} \mathbf{A}_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} A_x \\ 0 \end{pmatrix}, & \mathbf{A}_2 &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} 0 \\ A_y \end{pmatrix}, \\ \mathbf{A}_3 &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} \frac{A_x+A_y}{2} \\ \frac{A_x+A_y}{2} \end{pmatrix}, & \mathbf{A}_4 &= e^{i\pi/4} \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = e^{i\pi/4} \begin{pmatrix} \frac{A_x+A_y}{2} \\ -i \frac{A_x+A_y}{2} \end{pmatrix}. \end{aligned}$$

On trouve finalement l'expression suivante des intensités :

$$\begin{aligned} I_1 &= |A_x|^2, & I_2 &= |A_y|^2, \\ I_3 &= \frac{|A_x|^2 + |A_y|^2}{2} + \Re(A_x A_y^*), & I_4 &= \frac{|A_x|^2 + |A_y|^2}{2} + \Im(A_x A_y^*). \end{aligned} \quad (4.3)$$

Le vecteur des intensités $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$ et le vecteur instantané de Stokes introduit à l'équation (4.2) sont donc liés par une application linéaire et inversible. Par conséquent, la mesure du

vecteur \mathbf{I} permet de caractériser l'état de polarisation de la lumière. Dans la suite dans ce chapitre, on travaillera de préférence à partir du vecteur des intensités plutôt qu'à partir du vecteur de Stokes. Les images d'intensité acquises par le système imageur correspondent en définitive aux quatre composantes du vecteur \mathbf{I} . Remarquons que les quatre mesures correspondant au vecteur \mathbf{I} caractérisent l'interaction d'un matériau pour une polarisation donnée de la lumière incidente. Comme nous venons de le voir, un champ électrique de polarisation quelconque est caractérisé par les quatre intensités données à l'équation (4.3) obtenues pour quatre polarisations distinctes de champ. Aussi faudrait-il, afin d'être exhaustif, effectuer 16 mesures d'intensité (4 mesures d'intensité de l'onde rétrodiffusée pour chacune des 4 polarisations de l'onde incidente) pour déterminer l'interaction d'un matériau avec un champ électrique quelconque¹³. Dans la suite de ce chapitre, nous nous restreindrons cependant à l'étude de la polarisation du champ électrique rétrodiffusé pour un champ incident donné.

4.1.2 Lumière totalement polarisée

Lorsque les termes A_x et A_y sont déterministes, on parle alors de *lumière totalement polarisée*. Les deux composantes horizontales et verticales, du champ électrique \mathbf{E} évoluent dans ce cas de manière proportionnelle au cours du temps. Une manière classique de représenter l'état de polarisation de la lumière consiste alors à projeter dans le plan transverse $(O, \mathbf{x}, \mathbf{y})$ la trajectoire au cours du temps de l'extrémité du vecteur \mathbf{E} . En raison de la relation de proportionnalité entre les termes $A_x e^{-2i\pi\nu t}$ et $A_y e^{-2i\pi\nu t}$, on obtient la trajectoire elliptique présentée sur la figure 4.3. L'angle $\Psi \in [0, \pi[$ représente le décalage de l'ellipse de polarisation par rapport à l'axe horizontal, tandis que l'angle $\chi \in [-\pi/4, \pi/4[$ caractérise son ellipticité. On peut faire les remarques suivantes :

- lorsque $\chi = \pm\pi/4$, i.e. lorsque les complexes A_x et A_y ont le même module et ont une différence de phase égale à $\pm\pi/2$, l'ellipse de polarisation devient un cercle et l'on parle de *polarisation circulaire*.
- lorsque $\chi = 0$, i.e. lorsque les complexes A_x et A_y ont la même phase, l'ellipse de polarisation se réduit à une ligne et l'on parle donc de *polarisation rectiligne*.

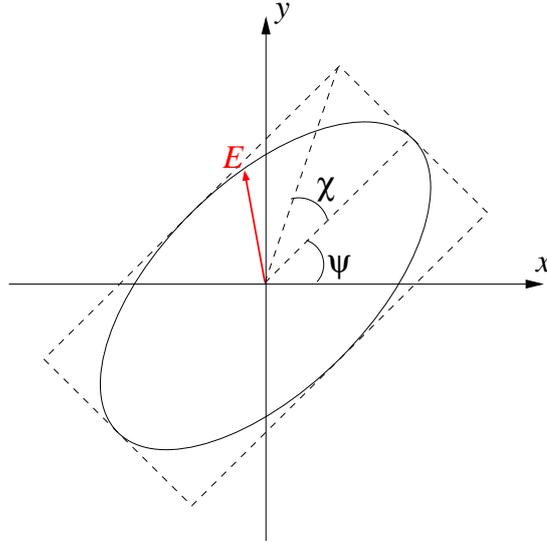
4.1.3 Lumière partiellement polarisée

Lorsqu'on s'intéresse à des milieux diffusants, la lumière que l'on doit étudier présente souvent une polarisation partielle. Dans ce cas, la représentation déterministe introduite à la section précédente ne suffit plus, et les composantes A_x et A_y du vecteur de Jones sont modélisées par des variables aléatoires.

Matrice de polarisation et degré de polarisation

Matrice de polarisation : Désormais, les variables aléatoires A_x et A_y ne sont plus proportionnelles car elles ne sont en général pas totalement corrélées. Afin d'étudier l'évolution de ces variables aléatoires complexes, une approche classique en polarimétrie [Bro98, p. 107] est de

¹³Cette étude correspondrait à déterminer la matrice de Mueller, notée M , d'un matériau [Hua94, p. 127]. Les matrices de Mueller sont en effet des matrices réelles de dimension 4×4 permettant de modéliser toutes les transformations linéaires subies par un champ électrique quelconque lors de son interaction avec un milieu diffusant. On a alors la relation : $\mathbf{S}_{\text{out}} = M\mathbf{S}_{\text{in}}$ où \mathbf{S}_{in} et \mathbf{S}_{out} sont respectivement les vecteurs de Stokes des champs incident et émergent.


 FIG. 4.3 – Ellipse de polarisation du champ électrique \mathbf{E} (diagramme de Lissajous)

considérer que le vecteur aléatoire de Jones $\mathbf{A} = (A_x, A_y)^T$ est centré et d'étudier alors la matrice de covariance de ce vecteur :

$$\Gamma = \begin{pmatrix} \mathbb{E}[A_x A_x^*] & \mathbb{E}[A_x A_y^*] \\ \mathbb{E}[A_y A_x^*] & \mathbb{E}[A_y A_y^*] \end{pmatrix} = \begin{pmatrix} a_1 & a_3 + ia_4 \\ a_3 - ia_4 & a_2 \end{pmatrix} \quad (4.4)$$

où a_1, a_2, a_3 et a_4 sont des réels tels que la matrice de covariance soit semi-définie positive :

$$\begin{aligned} a_1 &\geq 0, & a_2 &\geq 0, \\ a_1 a_2 &\geq a_3^2 + a_4^2. \end{aligned} \quad (4.5)$$

La matrice ainsi obtenue est appelée la *matrice de polarisation*, ou encore *matrice de cohérence*. Les propriétés statistiques du champ électrique sont donc étudiées à partir de ses moments d'ordre deux. Par construction, la matrice de polarisation Γ est une matrice hermitienne (i.e. $\Gamma = \Gamma^\dagger$) semi-définie positive et admet donc deux valeurs propres réelles positives. Lorsque cette matrice admet une valeur propre nulle, les variables complexes A_x et A_y sont totalement corrélées. Dans ce cas, la lumière est totalement polarisée. Par exemple :

- si $\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, alors le champ électrique est totalement polarisé et admet une polarisation rectiligne suivant la direction horizontale,
- si $\Gamma = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$, alors le champ électrique est totalement polarisé et admet une polarisation circulaire.

De plus, la trace de la matrice de la polarisation Γ correspond à l'intensité totale moyenne du champ électrique. Cette grandeur représente donc la réflectivité des matériaux imagés lorsqu'aucune polarisation n'est effectuée sur la lumière rétrodiffusée. Enfin, lorsque les paramètres a_3 et a_4 sont nuls, les composantes complexes A_x et A_y du vecteur de Jones sont décorrélées, et la dépolarisation est uniquement due aux différences entre les amplitudes de ces composantes. Les matériaux sont alors qualifiés de *dépolarisateur purs*, ce qui est en généralement le cas des matériaux métalliques.

Degré de polarisation : Puisque la lumière peut être partiellement polarisée, il devient nécessaire de définir une grandeur quantifiant son niveau de polarisation. La matrice Γ étant hermitienne et semi-définie positive, elle admet des valeurs propres réelles et positives notées $\lambda_1 \geq \lambda_2 \geq 0$. On définit alors le scalaire P , appelé *degré de polarisation* [Bro98, p. 111], comme

$$P = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}, \quad \text{ou de manière équivalente} \quad P = \sqrt{1 - 4 \frac{\det(\Gamma)}{\text{tr}(\Gamma)^2}}. \quad (4.6)$$

Le degré de polarisation P est donc une grandeur comprise dans l'intervalle $[0, 1]$:

- si $P = 1$ alors la lumière est totalement polarisée (on retrouve bien dans ce cas que la matrice de polarisation Γ possède une valeur propre λ_2 nulle et donc que les v.a. complexes A_x et A_y sont proportionnelles)
- si $P = 0$, on parle de lumière *totalement dépolarisée*
- si $0 < P < 1$, on parle de lumière *partiellement polarisée*

Par la suite, nous travaillerons avec le degré de polarisation (DdP) au carré, noté P^2 , qui est une mesure équivalente au degré de polarisation, puisque les deux grandeurs sont liées par une transformation bijective, et qui a le mérite d'avoir une expression plus simple. Avec les notations introduites à l'équation (4.4), il vient :

$$P^2 = 1 - 4 \frac{a_1 a_2 - a_3^2 - a_4^2}{(a_1 + a_2)^2}. \quad (4.7)$$

Il est important de noter que le degré de polarisation ne caractérise pas entièrement la matrice de polarisation du champ électrique. En effet, si l'on note

$$\Omega = \{(a_1, a_2, a_3, a_4) \in \mathbb{R}^4 \mid a_1, a_2 \geq 0 \text{ et } a_1 a_2 \geq a_3^2 + a_4^2\} \quad (4.8)$$

l'ensemble des valeurs admissibles des éléments de la matrice Γ , on remarque que la fonction

$$(a_1, a_2, a_3, a_4) \in \Omega \mapsto P^2$$

n'est pas injective. Cependant, le degré de polarisation a le mérite d'être une mesure scalaire simple qui s'avère suffisante pour caractériser les éléments d'une scène dans la plupart des problèmes de polarimétrie.

Phénomène de speckle

Le phénomène de speckle a déjà été introduit à la section 2.1.1 (p. 49) pour un champ électrique scalaire et donc totalement polarisé. Ce phénomène explique le comportement aléatoire de l'amplitude du champ électrique en imagerie cohérente. Il a été montré à la section 2.1.1 que sous les hypothèses de speckle pleinement développé, le champ scalaire est distribué suivant une loi gaussienne complexe circulaire.

Comme dans le cas de l'imagerie RSO, les hypothèses de speckle pleinement développé seront supposées vérifiées dans notre étude de l'imagerie polarimétrique. Le résultat trouvé pour un champ scalaire, se généralise alors au cas d'un champ vectoriel. On rappelle au préalable la définition d'un vecteur gaussien complexe circulaire (voir [Pic93, section 4.6] pour une discussion détaillée sur le sujet) :

Définition 4.1.1. Soit un vecteur aléatoire complexe $\mathbf{Z} \in \mathbb{C}^d$. On note respectivement \mathbf{X} et \mathbf{Y} les parties réelles et imaginaires de \mathbf{Z} , et $\mathbf{U} \in \mathbb{R}^{2d}$ le vecteur aléatoire de \mathbb{R}^{2d} suivant :

$$\mathbf{U} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}.$$

Par construction, le vecteur \mathbf{U} admet la matrice de covariance C_U suivante :

$$C_U = \begin{pmatrix} \mathbb{E}[\mathbf{X}\mathbf{X}^T] & \mathbb{E}[\mathbf{X}\mathbf{Y}^T] \\ \mathbb{E}[\mathbf{Y}\mathbf{X}^T] & \mathbb{E}[\mathbf{Y}\mathbf{Y}^T] \end{pmatrix} = \begin{pmatrix} C_{X,X} & C_{X,Y} \\ C_{X,Y}^T & C_{Y,Y} \end{pmatrix}.$$

Le vecteur aléatoire complexe $\mathbf{Z} \in \mathbb{C}^d$ est distribué suivant une loi gaussienne complexe circulaire si $\mathbf{U} \in \mathbb{R}^{2d}$ est un vecteur gaussien centré et si les sous-matrices $C_{X,X}$, $C_{Y,Y}$ et $C_{X,Y}$ de la matrice de covariance C_U vérifient les conditions :

$$C_{X,X} = C_{Y,Y}, \quad \text{et} \quad C_{X,Y}^T = -C_{X,Y}.$$

Finalement, on obtient le résultat suivant dans le cas d'un champ électrique vectoriel [Goo06].

Proposition 4.1.1. Sous les hypothèses de speckle pleinement développé, le vecteur de Jones $\mathbf{A} = (A_x, A_y)^T \in \mathbb{C}^2$ est distribué suivant une loi gaussienne complexe circulaire bivariée.

Une conséquence de ce résultat est que la loi du vecteur de Jones \mathbf{A} est entièrement caractérisée par sa matrice de covariance complexe qui est par définition (voir (4.4)) la matrice de polarisation Γ . On obtient par ailleurs à partir de l'expression (4.4) de la matrice Γ que le vecteur $\mathbf{U} = (\Re(A_x), \Re(A_y), \Im(A_x), \Im(A_y))^T \in \mathbb{R}^4$ est un vecteur gaussien centré dont la matrice de covariance C_U est de la forme :

$$C_U = \frac{1}{2} \begin{pmatrix} a_1 & a_3 & 0 & -a_4 \\ a_3 & a_2 & a_4 & 0 \\ 0 & a_4 & a_1 & a_3 \\ -a_4 & 0 & a_3 & a_2 \end{pmatrix}. \quad (4.9)$$

En définitive, l'étude de la polarimétrie du champ électrique se ramène à l'estimation du degré de polarisation P^2 sous l'hypothèse que le vecteur de Jones \mathbf{A} est un vecteur gaussien complexe circulaire de matrice de covariance Γ .

4.2 Imagerie polarimétrique "fort-flux"

On suppose dans cette section que les niveaux d'intensité sont suffisamment élevés pour que le vecteur \mathbf{I} des intensités, introduit à l'équation (4.3), puisse être mesuré directement. Comme cela a été expliqué précédemment, ce vecteur permet de caractériser entièrement l'état de la lumière. Cependant, en raison de la modélisation statistique de la lumière partiellement polarisée, il devient nécessaire d'étudier les propriétés statistiques de ce vecteur afin de pouvoir estimer les éléments de la matrice de polarisation. Bien que quatre mesures d'intensités, et donc quatre images, soient nécessaires pour estimer de façon exhaustive la matrice de polarisation Γ , le degré de polarisation peut être estimé à partir d'un nombre réduit de mesures. Les intérêts d'une telle démarche sont multiples :

- réduction du matériel, de son encombrement et du coût du système imageur
- réduction du temps d'acquisition

L'estimation du degré de polarisation par une méthode des moments basée sur deux images d'intensité a par exemple été récemment étudiée [Fad06, RR06]. Bien sûr, la réduction du nombre de mesures entraîne également une diminution des performances d'estimation par rapport au cas où quatre images sont disponibles. Il est donc important d'obtenir des estimateurs performants afin que ces méthodes basées sur un nombre réduit d'image soient opérationnelles. Cette section étudie des méthodes d'estimation par maximum de vraisemblance. Les estimateurs du maximum de vraisemblance ont en effet de bonnes propriétés, comme l'efficacité et la normalité asymptotique, sous des conditions de régularité assez faibles. Les performances des estimateurs ainsi obtenus seront comparées à celles obtenues pour des méthodes de moments.

4.2.1 Distribution du vecteur des intensités

La transformée de Laplace du vecteur des intensités \mathbf{I} peut facilement être obtenue en considérant la matrice aléatoire complexe $\mathbf{S} = \mathbf{A}\mathbf{A}^\dagger$ où \mathbf{A} est le vecteur de Jones et, où $(s_i, \theta_i) \in \mathbb{R}^2$ pour tout $i = 1, \dots, 4$. Un résultat classique [Goo63, GLM03] dû au fait que le vecteur de Jones \mathbf{A} est distribué suivant la loi gaussienne circulaire complexe de matrice de covariance Γ , est que la matrice aléatoire \mathbf{S} suit une loi de Wishart complexe dont la transformée de Laplace est :

$$\mathcal{L}_{\mathbf{S}}(\boldsymbol{\theta}) = \mathbb{E}[\exp[-\text{trace}(\mathbf{S}\boldsymbol{\theta})]] = |\mathbb{I}_2 + \Gamma\boldsymbol{\theta}|^{-1}, \quad (4.10)$$

où \mathbb{I}_2 est la matrice identité de $\mathbb{R}^{2 \times 2}$ et $\boldsymbol{\theta}$ une matrice 2×2 telle que $\Gamma^{-1} + \boldsymbol{\theta}$ soit une matrice hermitienne définie positive. En utilisant les notations suivantes :

$$\mathbf{S} = \begin{pmatrix} s_1 & s_3 + is_4 \\ s_3 - is_4 & s_2 \end{pmatrix} = \begin{pmatrix} |A_X|^2 & A_X A_Y^* \\ A_Y A_X^* & |A_Y|^2 \end{pmatrix},$$

et

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 & \theta_3 + i\theta_4 \\ \theta_3 - i\theta_4 & \theta_2 \end{pmatrix},$$

on obtient

$$\text{trace}(\mathbf{S}\boldsymbol{\theta}) = s_1\theta_1 + s_2\theta_2 + 2s_3\theta_3 + 2s_4\theta_4.$$

En outre, le vecteur des intensités \mathbf{I} et le vecteur aléatoire $\mathbf{s} = (s_1, \dots, s_4)^T$ sont liés par une transformation linéaire bijective, puisque d'après l'équation (4.3) :

$$\mathbf{s} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/2 & -1/2 & 1 & 0 \\ -1/2 & -1/2 & 0 & 1 \end{pmatrix} \mathbf{I} = \mathbf{M}\mathbf{I}, \quad (4.11)$$

Les équations (4.10) et (4.11) permettent d'obtenir l'expression suivante de la transformée de Laplace du vecteur des intensités :

$$\mathcal{L}_{\mathbf{I}}(\boldsymbol{\theta}) = \mathbb{E} \left[\exp \left(- \sum_{j=1}^4 \theta_j I_j \right) \right] = \frac{1}{P(\boldsymbol{\theta})}, \quad (4.12)$$

où $P(\boldsymbol{\theta})$ est le polynôme affine suivant :

$$P(\boldsymbol{\theta}) = 1 + \mathbf{m}^T \boldsymbol{\theta} + k[2\theta_1\theta_2 + \theta_3\theta_4 + (\theta_1 + \theta_2)(\theta_3 + \theta_4)], \quad (4.13)$$

le vecteur $\mathbf{m} \in \mathbb{R}^4$ et le scalaire $k \in \mathbb{R}$ étant des termes qui dépendent des paramètres réels a_1, a_2, a_3, a_4 :

$$\begin{aligned} \mathbf{m} &= \left[a_1, a_2, \frac{a_1 + a_2}{2} + a_3, \frac{a_1 + a_2}{2} + a_4 \right]^T, \\ k &= \frac{1}{2}(a_1 a_2 - a_3^2 - a_4^2). \end{aligned} \quad (4.14)$$

On reconnaît finalement la transformée de Laplace d'une loi Gamma multivariée, d'après la définition 1.3.2 (p. 29).

Proposition 4.2.1. *Sous les hypothèses de speckle pleinement développé, le vecteur des intensités $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$ est distribué selon une loi Gamma multivariée de paramètres : le paramètre de forme $q = 1$ et le polynôme affine P introduit aux équations (4.13) et (4.14).*

Les sous-vecteurs de \mathbf{I} sont donc eux aussi distribués suivant une loi Gamma multivariée (voir proposition 1.3.1, p. 30), et l'on obtient grâce à l'expression de la transformée de Laplace multivariée (4.12), les lois marginales univariées des intensités :

$$\begin{aligned} I_1 &\sim \mathcal{G}(1, a_1), & I_2 &\sim \mathcal{G}(1, a_2), \\ I_3 &\sim \mathcal{G}\left(1, \frac{a_1 + a_2}{2} + a_3\right), & I_4 &\sim \mathcal{G}\left(1, \frac{a_1 + a_2}{2} + a_4\right), \end{aligned}$$

et le vecteur \mathbf{m} défini à l'équation (4.14) correspond à la moyenne du vecteur \mathbf{I} . Une autre constatation directement obtenue à partir de l'expression transformée de Laplace (4.12) est que le degré de polarisation, qui dépend des variables a_1, a_2 et $\rho = a_3^2 + a_4^2$ d'après l'équation (4.7), peut être estimé à partir de seulement 2 ou 3 images. En effet, les expressions des transformées de Laplace des sous-vecteurs de \mathbf{I} (la transformée de Laplace du couple $(I_1, I_2)^T$ est par exemple déduite de $\mathcal{L}_{\mathbf{I}}(\boldsymbol{\theta})$ en posant $\theta_3 = \theta_4 = 0$) montrent que le vecteur des paramètres $(a_1, a_2, \rho)^T$ est identifiable pour les lois des marges suivantes :

- le couple $(I_1, I_2)^T$,
- les triplets $(I_1, I_2, I_3)^T$ et $(I_1, I_2, I_4)^T$.

Ces marges sont par ailleurs les seules, avec le vecteur $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$, pour lesquelles le vecteur $(a_1, a_2, \rho)^T$, et donc le degré de polarisation, sont identifiables.

4.2.2 Estimation de la matrice de polarisation et du DdP

On suppose dans cette section que l'on dispose d'un échantillon empirique de n réalisations $\mathbf{I}^1, \dots, \mathbf{I}^n$ (en pratique ces réalisations correspondent aux pixels appartenant à la fenêtre d'estimation, sous les hypothèses habituelles de stationnarité et d'ergodicité). Les estimateurs du degré de polarisation (DdP) obtenus par la méthode du maximum de vraisemblance et par des méthodes de moments sont alors étudiés et comparés lorsque 4, 3 ou seulement 2 images sont disponibles. Ces estimateurs sont obtenus pour des images monovue ($q = 1$), mais les résultats présentés s'étendent sans difficulté lorsque plusieurs vues de chaque image d'intensité sont moyennées.

Cas à quatre images

Dans ce cas, les quatre images d'intensité correspondent aux réalisations du vecteur $\mathbf{I} = (I_1, I_2, I_3, I_4)^T \in \mathbb{R}^4$. La transformée de Laplace (4.12) indique que le vecteur des paramètres

$\mathbf{a} = (a_1, a_2, a_3, a_4)$ est alors identifiable, et il donc possible d'estimer la matrice de polarisation Γ . Dans le but d'obtenir une expression des estimateurs du maximum de vraisemblance de ces paramètres, on démontre dans cette section que la loi du vecteur \mathbf{I} décrit une famille exponentielle naturelle.

Afin d'étudier les propriétés de la loi du vecteur \mathbf{I} , on s'intéresse à la loi du vecteur de Jones $\mathbf{A} = (A_x, A_y)^T$ à partir duquel est construit \mathbf{I} . En raison des hypothèses de speckle pleinement développé, le vecteur réel $\mathbf{U} = (\Re(A_x), \Re(A_y), \Im(A_x), \Im(A_y))^T \in \mathbb{R}^4$ est un vecteur gaussien dont la mesure de probabilité définie sur \mathbb{R}^4 a pour expression :

$$f_{\mathbf{U}}(d\mathbf{u}) = e^{-\frac{1}{2}\mathbf{u}^T C_U^{-1} \mathbf{u}} \frac{d\mathbf{u}}{\det(C_U)^{1/2} (2\pi)^2}$$

où C_U est la matrice de covariance présentée à l'équation (4.9). En remarquant :

$$\mathbf{u}^T C_U^{-1} \mathbf{u} = \text{tr}(\mathbf{u}^T C_U^{-1} \mathbf{u}) = \text{tr}(C_U^{-1} \mathbf{u} \mathbf{u}^T),$$

la mesure précédente peut se réécrire sous la forme :

$$f_{\mathbf{U}}(d\mathbf{u}) = e^{\text{tr}\left(\frac{-C_U^{-1}}{2} \mathbf{u} \mathbf{u}^T\right)} \frac{d\mathbf{u}}{\det(C_U)^{1/2} (2\pi)^2}. \quad (4.15)$$

On désigne par H l'espace vectoriel de dimension 4 des matrices réelles 4×4 de la forme :

$$\begin{pmatrix} A & B \\ B^T & A \end{pmatrix},$$

où A et B sont des matrices réelles 2×2 respectivement symétrique et antisymétrique. Il est montré en annexe D.3.1 que si une matrice inversible est dans H , son inverse aussi. Or la matrice de covariance C_U appartient à l'espace vectoriel H d'après la section 4.1.3. C'est donc également le cas de la matrice $\boldsymbol{\theta} = -\frac{1}{2}C_U^{-1}$. Mais, d'après la définition donnée à l'équation (4.3) du vecteur \mathbf{I} , on obtient en notant $\mathbf{V} = \mathbf{U}\mathbf{U}^T$:

$$\mathbf{I} = (\text{tr}(\mathbf{B}_1 \mathbf{V}), \text{tr}(\mathbf{B}_2 \mathbf{V}), \text{tr}(\mathbf{B}_3 \mathbf{V}), \text{tr}(\mathbf{B}_4 \mathbf{V}))^T,$$

où

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \mathbf{B}_4 = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{-1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{-1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Les matrices $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \mathbf{B}_4$ forment une famille libre des matrices de $\mathbb{R}^{4 \times 4}$ et appartiennent de plus à l'espace vectoriel H de dimension 4. Ces matrices forment donc une base de H , et la matrice $\boldsymbol{\theta}$ peut être décomposée comme suit :

$$\boldsymbol{\theta} = \theta_1 \mathbf{B}_1 + \theta_2 \mathbf{B}_2 + \theta_3 \mathbf{B}_3 + \theta_4 \mathbf{B}_4,$$

où les paramètres $\theta_1, \theta_2, \theta_3$ et θ_4 sont des scalaires réels. Finalement, la mesure associée à la loi de probabilité de \mathbf{I} est de la forme :

$$f_{\mathbf{I}}(d\mathbf{I}) = \exp\left(\sum_{i=1}^4 \theta_i I_i\right) \frac{\det(-\boldsymbol{\theta})^{1/2}}{\pi^2} \nu(d\mathbf{I}), \quad (4.16)$$

où la mesure ν est l'image de la mesure de Lebesgue sur \mathbb{R}^4 par l'application $\phi : \mathbb{R}^4 \rightarrow \mathbb{R}_+^4$ définie par $\phi(\mathbf{U}) = \mathbf{I}$ (notons que cette mesure n'est pas concentrée sur un hyperplan affine, sinon les composantes du vecteur \mathbf{I} seraient liées par une combinaison affine, ce qui n'est pas le cas). D'après la définition des familles exponentielles naturelles donnée à la section B.1, p. 158, on obtient le résultat suivant¹⁴ :

Théorème 4.2.2. *La loi du vecteur \mathbf{I} décrit une famille exponentielle naturelle sur \mathbb{R}^4 dont le paramètre $-\boldsymbol{\theta}$ appartient à l'ensemble ouvert des matrices définies positives de H (qui est non vide puisque toutes les matrices réelles diagonales 4×4 dont les éléments sont strictement positifs sont dans cet ensemble).*

En raison de ce résultat et du théorème B.6.3, l'estimateur du maximum de vraisemblance du vecteur des moyennes $\mathbf{m} = \mathbf{E}[\mathbf{I}]$ correspond à la moyenne empirique :

$$\widehat{\mathbf{m}}_{\text{MV}} = \frac{1}{n} \sum_{j=1}^n \mathbf{I}^j, \quad (4.17)$$

et cet estimateur est sans biais et efficace. Mais d'après l'équation (4.11), le vecteur des moyennes $\mathbf{m} = (m_1, m_2, m_3, m_4)^T$ et le vecteur des paramètres $\mathbf{a} = (a_1, a_2, a_3, a_4)^T$ de la matrice de polarisation Γ sont liés par la transformation linéaire bijective suivante :

$$\mathbf{a} = M\mathbf{m}.$$

Le principe d'invariance fonctionnelle assure alors que l'EMV du vecteur \mathbf{a} est :

$$\widehat{\mathbf{a}}_{\text{MV}} = M\widehat{\mathbf{m}}_{\text{MV}}. \quad (4.18)$$

Cet estimateur est également sans biais et efficace, puisqu'il se déduit de l'estimateur $\widehat{\mathbf{m}}_{\text{ML}}$ par une transformation linéaire. Il est important de remarquer que l'EMV ainsi obtenu correspond à l'estimateur classique en imagerie polarimétrique [Bro98, p. 340] obtenu par la méthode des moments (voir aussi [Réf02, p. 237] pour une démonstration de ce résultat par une approche différente). Les résultats démontrés dans cette section permettent d'affirmer que cet estimateur des éléments de la matrice de polarisation Γ est optimal dans le sens où il est non biaisé et efficace. En outre, la borne de Cramer-Rao peut être facilement obtenue puisqu'elle correspond à la matrice de covariance du vecteur $\widehat{\mathbf{a}}_{\text{MV}}$:

$$\text{cov}(\widehat{\mathbf{a}}_{\text{MV}}) = \frac{1}{n} M \text{cov}(\mathbf{I}) M^T,$$

où $\text{cov}(\mathbf{I})$ est la matrice de covariance du vecteur \mathbf{I} . La matrice $\text{cov}(\mathbf{I})$ se calcule sans difficulté grâce à l'expression des moments d'une loi Gamma bivariée donnée à l'équation (1.23) (p. 18) puisque chaque couple (I_i, I_j) , pour tout $1 \leq i < j \leq 4$, suit une loi Gamma bivariée dont les paramètres se déduisent de la transformée de Laplace (4.12). On trouve alors :

$$\text{cov}(\widehat{\mathbf{a}}_{\text{MV}}) = \frac{1}{n} \begin{pmatrix} a_1^2 & a_3^2 + a_4^2 & a_1 a_3 & a_1 a_4 \\ a_3^2 + a_4^2 & a_2^2 & a_2 a_3 & a_2 a_4 \\ a_1 a_3 & a_2 a_3 & c_{3,3} & a_3 a_4 \\ a_1 a_4 & a_2 a_4 & a_3 a_4 & c_{4,4} \end{pmatrix}, \quad (4.19)$$

¹⁴On pourrait déterminer l'expression analytique explicite de la loi du vecteur \mathbf{I} en effectuant le changement de variable. Cependant cette expression n'est pas utile pour la suite de l'étude, le fait que ces lois appartiennent à une FEN étant suffisant.

où $c_{3,3} = (a_1 a_2 + a_3^2 - a_4^2)/2$ et $c_{4,4} = (a_1 a_2 - a_3^2 + a_4^2)/2$.

Finalement les estimateurs ainsi obtenus sont remplacés dans l'expression (4.7) du degré de polarisation afin d'obtenir l'estimateur du maximum de vraisemblance de P^2 :

$$\widehat{P}_4^2 = 1 - \frac{4[\widehat{a}_1 \widehat{a}_2 - (\widehat{a}_3^2 + \widehat{a}_4^2)]}{(\widehat{a}_1 + \widehat{a}_2)^2},$$

où \widehat{a}_i est la i -ème composante du vecteur $\widehat{\mathbf{a}}_{\text{MV}}$ pour $i = 1, \dots, 4$. Il faut noter que l'estimateur ainsi obtenu du DdP est a priori biaisé (ce qui sera confirmé dans les simulations). Néanmoins, d'après les propriétés des EMV, cet estimateur est asymptotiquement non biaisé et efficace. La variance asymptotique de \widehat{P}_4^2 , qui correspond donc à la borne de Cramer-Rao du paramètre P^2 , a finalement pour expression :

$$\text{var}_A(\widehat{P}_4^2) = G_4^T \text{cov}(\widehat{\mathbf{a}}_{\text{MV}}) G_4 = \frac{2(1 - P^2)^2 P^2}{n}, \quad (4.20)$$

où

$$G_4(\mathbf{a}) = \left(4 \frac{a_1 a_2 - a_2^2 - 2a_3^2 - 2a_4^2}{(a_1 + a_2)^3}, 4 \frac{a_1 a_2 - a_1^2 - 2a_3^2 - 2a_4^2}{(a_1 + a_2)^3}, 8 \frac{a_3}{(a_1 + a_2)^2}, 8 \frac{a_4}{(a_1 + a_2)^2} \right)^T \quad (4.21)$$

est le gradient de la fonction de $g_4 : \mathbf{a} \mapsto g_4(\mathbf{a}) = P^2$. L'expression (4.20) permet de remarquer que la variance asymptotique de \widehat{P}_4^2 ne dépend des valeurs des paramètres a_1, a_2, a_3 et a_4 qu'à travers la valeur du degré de polarisation au carré P^2 . La variance asymptotique de \widehat{P}_4^2 est donc une fonction de P^2 qui, en outre, est maximale lorsque $P^2 = 1/3$.

Cas à trois images

On suppose dans ce cas que les 3 images d'intensité correspondent aux réalisations du vecteur $\widetilde{\mathbf{I}} = (I_1, I_2, I_3)^T$ (le cas où $\widetilde{\mathbf{I}} = (I_1, I_2, I_4)^T$ se traite de manière totalement symétrique en échangeant les variables I_3 et I_4 ainsi que les paramètres a_3 et a_4 dans les expressions trouvées). L'étude de ce cas est motivée par le fait que le vecteur aléatoire des quatre intensités $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$ est concentré sur la frontière d'un cône. Un calcul direct à partir de l'expression (4.3) montre en effet que ce vecteur vérifie¹⁵ :

$$(I_1 + I_2)^2 - (I_1 - I_2)^2 - \left(I_3 - \frac{I_1 + I_2}{2} \right)^2 - \left(I_4 - \frac{I_1 + I_2}{2} \right)^2 = 0, \quad (4.22)$$

et donc la loi de ce vecteur est singulière car concentrée sur ce cône. En conséquence, lorsque le vecteur aléatoire $\widetilde{\mathbf{I}} = (I_1, I_2, I_3)^T$ est connu, la v.a. I_4 n'admet que 1 ou 2 valeurs possibles suivant la valeur de $\widetilde{\mathbf{I}} = (I_1, I_2, I_3)^T$. En notant

$$\Delta = \sqrt{(I_1 + I_2)^2 - (I_1 - I_2)^2 - (I_3 - (I_1 + I_2)/2)^2},$$

¹⁵On retrouve en fait une propriété bien connue du vecteur instantané de Stokes $\mathbf{S} = (S_0, S_1, S_2, S_3)^T$ introduit à l'équation (4.2), ce vecteur se déduisant du vecteur $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$ par une transformation linéaire. Ce vecteur est en effet concentré sur le cône d'équation $S_0^2 = S_1^2 + S_2^2 + S_3^2$. Ce résultat est valide pour une image monovue ($q = 1$) mais ne l'est cependant plus pour une image multivues ($q > 1$).

on obtient en effet :

$$\begin{aligned} I_4 &= \frac{I_1 + I_2}{2} \pm \Delta && \text{si } (I_1 + I_2) > 2\Delta, \\ I_4 &= \frac{I_1 + I_2}{2} + \Delta && \text{sinon.} \end{aligned}$$

Il est donc légitime de penser que la quatrième image d'intensité I_4 n'apporte que peu d'information pour estimer les propriétés polarimétriques de la lumière lorsque le vecteur $\tilde{\mathbf{I}}$ est connu.

La transformée de Laplace du vecteur $\tilde{\mathbf{I}}$ s'obtient à partir de la formule (4.12) en posant $\theta_4 = 0$:

$$\mathcal{L}_{\tilde{\mathbf{I}}}(\tilde{\boldsymbol{\theta}}) = \mathbb{E} \left[\exp \left(- \sum_{j=1}^3 \theta_j I_j \right) \right] = \left(1 + \sum_{i=1}^3 m_i \theta_i + k [2\theta_1 \theta_2 + \theta_1 \theta_3 + \theta_2 \theta_3] \right)^{-1} = (\tilde{P}(\tilde{\boldsymbol{\theta}}))^{-1},$$

où $\tilde{\boldsymbol{\theta}} = (\theta_1, \theta_2, \theta_3)^T$. Les paramètres identifiables dans ce cas sont donc a_1, a_2, a_3 et a_4^2 . On constate cependant que le paramètre a_4 ne peut plus être estimé qu'au signe près puisque seule la grandeur a_4^2 est observable. Le vecteur des paramètres inconnus à estimer est donc dans cette section $\boldsymbol{\eta} = (a_1, a_2, a_3, a_4^2)$.

Méthode du maximum de vraisemblance : Un résultat intéressant est que la densité de la loi du vecteur $\tilde{\mathbf{I}} = (I_1, I_2, I_3)^T$ peut être déterminée. En effet, Letac et Wesolowski [LW07] ont récemment étudié les lois définies sur \mathbb{R}^n dont les transformées de Laplace sont de la forme :

$$\mathcal{L}(\boldsymbol{\theta}) = [1 + 2\mathbf{c}^T \boldsymbol{\theta} + v(\boldsymbol{\theta})]^{-q},$$

où $q > 0$, $\mathbf{c} \in \mathbb{R}^n$ et $v(\boldsymbol{\theta})$ est une forme quadratique sur \mathbb{R}^n . Les conditions pour lesquelles ces lois existent ont ainsi été déterminées et, une expression analytique explicite de leurs densités de probabilité a été calculée lorsque la loi n'est pas singulière. En posant alors $n = 3$, $q = 1$, $\mathbf{c} = \frac{1}{2} (m_1, m_2, m_3)^T$, et $v(\boldsymbol{\theta}) = k [2\theta_1 \theta_2 + \theta_1 \theta_3 + \theta_2 \theta_3]$, on obtient grâce à leur résultat, que la loi du vecteur $\tilde{\mathbf{I}}$ est régulière (i.e. absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^3) et admet la densité de probabilité suivante :

$$p(\tilde{\mathbf{I}}) = \frac{\exp\left(\frac{-\mathbf{c}'(\mathbf{a})^T \tilde{\mathbf{I}}}{2k}\right)}{k \sqrt{\pi v'(\tilde{\mathbf{I}})}} f_{\frac{1}{2}} \left(\frac{a_4^2 v'(\tilde{\mathbf{I}})}{16k^2} \right) \mathbb{I}_{\Omega}(\tilde{\mathbf{I}}), \quad (4.23)$$

où $\mathbb{I}_{\Omega}(\cdot)$ est la fonction indicatrice sur le cône Ω ,

$$\begin{aligned} \mathbf{c}'(\mathbf{a}) &= (a_2 + a_3, a_1 + a_3, -2a_3)^T, \\ v'(\tilde{\mathbf{I}}) &= (I_1 + I_2)^2 - (I_1 - I_2)^2 - (I_1 + I_2 - 2I_3)^2, \\ \Omega &= \{ \tilde{\mathbf{I}} \in \mathbb{R}^3; v'(\tilde{\mathbf{I}}) > 0 \}, \end{aligned}$$

et $f_q(z)$ est la fonction introduite à l'équation (1.32) (p. 21) qui intervient notamment dans l'expression de la densité de la loi Gamma bivariée.

L'EMV à partir de trois images du vecteur $\boldsymbol{\eta} = (a_1, a_2, a_3, a_4^2)^T$ peut être obtenu à l'aide des propriétés des familles exponentielles naturelles et de l'étude du gradient de la log-vraisemblance.

L'annexe D.3.2 montre que les estimateurs de a_1 , a_2 et a_3 ont les expressions explicites suivantes :

$$\begin{pmatrix} \widetilde{a}_1 \\ \widetilde{a}_2 \\ \widetilde{a}_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/2 & 1 & -1/2 \end{pmatrix} \begin{pmatrix} \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \end{pmatrix}, \quad (4.24)$$

où $\widetilde{m}_l = \frac{1}{n} \sum_{j=1}^n \mathbf{I}_l^j$ for $l = 1, 2, 3$. On obtient donc le résultat suivant :

$$(\widetilde{a}_1, \widetilde{a}_2, \widetilde{a}_3)^T = (\widehat{a}_1, \widehat{a}_2, \widehat{a}_3)^T$$

où les \widehat{a}_i sont les EMV des paramètres a_i , pour $1 \leq i \leq 3$, définis à l'équation (4.18) dans le cas à quatre images. Nous avons montré que ces estimateurs sont efficaces pour la loi du vecteur $\mathbf{I} = (I_1, I_2, I_3, I_4)^T$. Ces EMV sont donc également efficaces pour la loi du sous vecteur $\widetilde{\mathbf{I}} = (I_1, I_2, I_3)^T$. L'estimation du paramètre a_4^2 se révèle plus délicate puisqu'il n'existe pas d'expression analytique simple de cet estimateur. En conséquence, cet estimateur est obtenu numériquement en calculant la racine de la fonction score suivante (voir annexe D.3.2) :

$$4(\widehat{d} - a_4^2) - \frac{1}{n} \sum_{j=1}^n v'(\widetilde{\mathbf{I}}^j) \frac{f_{\frac{3}{2}}\left(\frac{a_4^2 v'(\widetilde{\mathbf{I}}^j)}{4(\widehat{d} - a_4^2)^2}\right)}{f_{\frac{1}{2}}\left(\frac{a_4^2 v'(\widetilde{\mathbf{I}}^j)}{4(\widehat{d} - a_4^2)^2}\right)} = 0, \quad (4.25)$$

où $\widehat{d} = \widehat{a}_1 \widehat{a}_2 - \widehat{a}_3^2$. Cette racine peut être calculée à l'aide d'une méthode de type Newton-Raphson, et l'on observe dans les simulations que l'algorithme converge vers la racine après peu d'itérations.

D'après le principe d'invariance fonctionnelle, l'EMV à trois images du degré de polarisation a finalement pour expression :

$$\widetilde{P}_3^2 = 1 - \frac{4 \left[\widehat{a}_1 \widehat{a}_2 - (\widehat{a}_3^2 + \widetilde{a}_4^2) \right]}{(\widehat{a}_1 + \widehat{a}_2)^2},$$

où \widetilde{a}_4^2 est la racine de l'équation non linéaire (4.25). Les propriétés asymptotiques des estimateurs du maximum de vraisemblance assurent finalement que la matrice de covariance asymptotique des EMV du vecteur des paramètres inconnus $\boldsymbol{\eta} = (a_1, a_2, a_3, a_4^2)^T$ correspond à la borne de Cramer-Rao. Cette borne peut être calculée en inversant la matrice d'information de Fisher :

$$F_3(\boldsymbol{\eta}) = -E \left[\frac{\partial^2 \log p(\widetilde{\mathbf{I}}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \right].$$

Il n'est cependant pas possible d'obtenir une expression analytique simple de cette matrice à cause du terme $\log f_{\frac{1}{2}}$ qui intervient dans l'expression de la log-densité. Comme cela a déjà été fait dans le cas bivarié, les éléments de cette matrice sont estimés par des simulations de Monte-Carlo :

$$[F_3(\boldsymbol{\eta})]_{ij} \simeq -\frac{1}{N} \sum_{k=1}^N \frac{\partial^2 \log p(\mathbf{X}_k)}{\partial \eta_i \partial \eta_j},$$

où le vecteur aléatoire \mathbf{X}_k est distribué suivant la loi Gamma trivariée dont la densité est définie à l'équation (4.23) et N est le nombre de tirages. La variance asymptotique de l'estimateur \widehat{P}_3^2 du DdP se calcule donc de la manière suivante :

$$\text{var}(\widehat{P}_3^2) = \mathbf{G}_3^T \mathbf{F}_3^{-1} \mathbf{G}_3,$$

où \mathbf{G}_3 est le gradient de la fonction $\boldsymbol{\eta} = (a_1, a_2, a_3, a_4^2) \mapsto P^2$:

$$\mathbf{G}_3 = \left(4 \frac{a_1 a_2 - a_2^2 - 2a_3^2 - 2a_4^2}{(a_1 + a_2)^3}, 4 \frac{a_1 a_2 - a_1^2 - 2a_3^2 - 2a_4^2}{(a_1 + a_2)^3}, 8 \frac{a_3}{(a_1 + a_4)^2}, \frac{4}{(a_1 + a_4)^2} \right)^T.$$

Méthode des Moments : Afin d'apprécier les performances de l'EMV obtenu ci-dessus, on étudie un estimateur reposant sur la méthode d'estimation classique des moments. Lorsque les trois intensités observées sont I_1, I_2 et I_3 , un estimateur des paramètres a_1, a_2, a_3 et a_4^2 peut être obtenu à partir des moments d'ordre 1 et 2 suivants :

$$\mathbf{E}[I_1] = a_1, \quad \mathbf{E}[I_2] = a_2, \quad \mathbf{E}[I_3] = \frac{1}{2}(a_1 + a_2) + a_3, \quad (4.26)$$

$$\mathbf{E}[I_1 I_2] = a_1 a_2 + a_3^2 + a_4^2, \quad (4.27)$$

$$\mathbf{E}[I_1 I_3] = a_1^2 + (a_2/2 + 2a_3)a_1 + (a_3^2 + a_4^2)/2, \quad (4.28)$$

$$\mathbf{E}[I_2 I_3] = a_2^2 + (a_1/2 + 2a_3)a_2 + (a_3^2 + a_4^2)/2. \quad (4.29)$$

Les estimateurs de a_1, a_2 et a_3 se déduisent directement des équations ci-dessus :

$$(\widetilde{a}_1, \widetilde{a}_2, \widetilde{a}_3)^T = (\widehat{a}_1, \widehat{a}_2, \widehat{a}_3)^T.$$

On retrouve ainsi les estimateurs du maximum de vraisemblance des paramètres obtenus pour 3 ou 4 images. Le paramètre a_4^2 peut être estimé à partir des équations (4.27), (4.28) et (4.29). Comme le système est surdéterminé, il est judicieux de pondérer chaque équation en fonction de la covariance des moments estimés, comme cela est expliqué ci-dessous.

Supposons par exemple que la moyenne de I_3 soit grande par rapport aux moyennes de I_1 et I_2 . La variance des termes $I_1 I_2$ et $I_1 I_3$ sera d'après notre modèle bien plus grande que celle de $I_1 I_2$. Prendre le même poids pour les équations (4.27), (4.28) et (4.29) pénalisera alors l'estimation par rapport à une méthode où seule l'équation (4.27) du moment $\mathbf{E}[I_1 I_2]$ est prise en compte. Afin de contourner cette difficulté, une méthode de moindres carrés non linéaires (MCNL) a été étudiée dans [PF89]. En notant

$$\mathbf{f}(\mathbf{a}) = (\mathbf{E}[I_1 I_2], \mathbf{E}[I_1 I_3], \mathbf{E}[I_2 I_3])^T,$$

\mathbf{s}_n les moments empiriques associés à \mathbf{f} et $\mathbf{C}(\mathbf{a})$ la matrice de covariance suivante :

$$\mathbf{C}(\mathbf{a}) = n \text{cov}(\mathbf{s}_n),$$

L'estimateur de a_4^2 pour la méthode MCNL est défini par :

$$\widetilde{a}_{4\text{Mo}}^2 = \arg \min_{x>0} (\widetilde{\mathbf{f}}(x) - \mathbf{s}_n)^T \widetilde{\mathbf{C}}(x)^{-1} (\widetilde{\mathbf{f}}(x) - \mathbf{s}_n),$$

où $\widetilde{f}(x) = f(\widetilde{a}_1, \widetilde{a}_2, \widetilde{a}_3, x)$ et $\widetilde{C}(x) = C(\widetilde{a}_1, \widetilde{a}_2, \widetilde{a}_3, x)$. Comme il n'existe pas d'expression simple du critère à minimiser, le calcul numérique de $\widetilde{a}_{3\text{Mo}}^2$ est effectué à l'aide d'une méthode de gradient de type Levenberg-Marquardt. On obtient finalement l'estimateur du DdP suivant :

$$\widetilde{P}_{3\text{ Mo}}^2 = 1 - \frac{4 \left[\widetilde{a}_1 \widetilde{a}_2 - (\widetilde{a}_3^2 + \widetilde{a}_{4\text{Mo}}^2) \right]}{(\widetilde{a}_1 + \widetilde{a}_2)^2}.$$

Il est finalement intéressant de noter qu'on peut obtenir la borne inférieure de la variance asymptotique de tout estimateur de P^2 , obtenu par une méthode de moments, à partir des moments introduits dans le système (4.26) (le lecteur est invité à consulter [PF89] pour plus de détails).

Cas à deux images

Le cas à deux images correspond à l'étude de la loi du couple $\underline{I} = (I_1, I_2)^T$ dont la transformée de Laplace est :

$$\mathcal{L}_{\underline{I}}(\theta_1, \theta_2) = \left[1 + a_1 \theta_1 + a_2 \theta_2 + (a_1 a_2 - a_3^2 - a_4^2) \theta_1 \theta_2 \right]^{-1},$$

On se ramène donc à l'estimation des paramètres d'une loi Gamma bivariée. D'après l'expression (1.23) (p. 18) des moments de ces lois, les moyennes sont a_1 et a_2 et le coefficient de corrélation entre les marges est $r = (a_3^2 + a_4^2)/(a_1 a_2)$. Cette loi peut être reparamétrisée de façon équivalente par le vecteur $\eta = (a_1, a_2, \rho)^T$ où $\rho = a_3^2 + a_4^2$ est le paramètre de covariance.

Méthode du maximum de vraisemblance L'estimation par maximum de vraisemblance des paramètres d'une loi Gamma bivariée a déjà été étudiée à la section 2.2.1, p. 57. Les EMV \underline{a}_1 et \underline{a}_2 des paramètres a_1 et a_2 sont les moyennes empiriques et correspondent donc aux estimateurs obtenus pour 3 et 4 images :

$$\begin{aligned} \underline{a}_1 &= \widehat{a}_1, \\ \underline{a}_2 &= \widehat{a}_2, \end{aligned}$$

D'après le principe d'invariance fonctionnelle, l'EMV $\underline{\rho}$ du paramètre de covariance ρ est :

$$\underline{\rho} = \widehat{a}_1 \widehat{a}_2 \underline{r},$$

où l'EMV du coefficient de corrélation \underline{r} est la racine de l'équation non linéaire (2.12). Finalement, l'EMV du DdP est :

$$\underline{P}_{2\text{ Mo}}^2 = 1 - \frac{4(\widehat{a}_1 \widehat{a}_2 - \underline{\rho})}{(\widehat{a}_1 + \widehat{a}_2)^2}.$$

Cet estimateur est asymptotiquement efficace et la borne de Cramer-Rao pour cet estimateur est :

$$\text{var}(\underline{P}_2^2) = G_2^T F_2^{-1} G_2,$$

où G_2 est le gradient de la fonction $\eta = (a_1, a_2, \rho)^T \mapsto P^2$:

$$G_2 = \left(4 \frac{a_1 a_2 - a_2^2 - 2\rho}{(a_1 + a_2)^3}, 4 \frac{a_1 a_2 - a_1^2 - 2\rho}{(a_1 + a_2)^3}, \frac{1}{(a_1 + a_2)^2} \right)^T. \quad (4.30)$$

et F_2^{-1} est la matrice d'information de Fisher pour le vecteur des paramètres $\eta = (a_1, a_2, \rho)^T$.

Méthode des moments Cette méthode d'estimation a été étudiée dans [RR06, RFR07a] afin d'estimer le degré de polarisation à partir de seulement 2 images. Les estimateurs des paramètres du vecteur $\boldsymbol{\eta}$ sont obtenus à partir des moments d'ordre 1 et 2 suivants :

$$\mathbf{E}[I_1] = a_1, \quad \mathbf{E}[I_2] = a_2, \quad (4.31)$$

$$\mathbf{E}[I_1 I_2] = a_1 a_2 + \rho, \quad (4.32)$$

Le système obtenu n'est pas dans ce cas surdéterminé, et les estimateurs s'obtiennent directement. L'équation (4.31) démontre :

$$(\underline{a}_1, \underline{a}_2)^T = (\widehat{a}_1, \widehat{a}_2)^T,$$

Ces estimateurs sont donc les mêmes que ceux obtenus par maximum de vraisemblance. L'estimateur de ρ est déduit de l'équation (4.27) :

$$\underline{\rho}_{\text{Mo}} = \frac{1}{n} \sum_{j=1}^n I_1^j I_2^j - \widehat{a}_1 \widehat{a}_2,$$

et l'estimateur du DdP a pour expression :

$$\underline{P}_{2 \text{ Mo}}^2 = 1 - \frac{4(\widehat{a}_1 \widehat{a}_2 - \underline{\rho}_{\text{Mo}})}{(\widehat{a}_1 + \widehat{a}_2)^2}.$$

La matrice de covariance asymptotique de l'estimateur des moments $\underline{\boldsymbol{\eta}}_{2, \text{Mo}} = (\widehat{a}_1, \widehat{a}_2, \underline{\rho}_{\text{Mo}})^T$ s'obtient facilement à partir de l'expression des moments d'une loi Gamma bivariée (1.19) :

$$\text{var}_A(\underline{\boldsymbol{\eta}}_{2, \text{Mo}}) = \frac{1}{n} \begin{pmatrix} a_1^2 & \rho & 2a_1 \rho \\ \rho & a_2^2 & 2a_2 \rho \\ 2a_1 & 2a_2 \rho & a_1^2 a_2^2 + 4a_1 a_2 \rho + 3\rho^2 \end{pmatrix}.$$

Finalement la variance asymptotique de $\underline{P}_{2 \text{ Mo}}^2$ est :

$$\begin{aligned} \text{var}_A(\underline{P}_{2 \text{ Mo}}^2) &= G_2^T \text{var}_A(\underline{\boldsymbol{\eta}}_{2, \text{Mo}}) G_2, \\ &= \frac{2(1 - P^2)^2(P^2 + 1/2)}{n} + \frac{64a_1 a_2 \rho}{n(a_1 + a_2)^4}, \end{aligned}$$

où G_2 est le gradient de la fonction $(a_1, a_2, \rho) \mapsto P^2$, définie précédemment. Cette expression permet de souligner les points suivants :

- La variance asymptotique obtenue pour 2 images est supérieure à celle obtenue à l'équation (4.20) lorsque les 4 images d'intensité sont disponibles.
- Contrairement au cas à quatre images, cette variance n'est pas une fonction du seul DdP P^2 mais dépend de tous les paramètres $(a_1, a_2, \rho)^T$.

4.2.3 Estimation sur données synthétiques

Afin de comparer les performances des différents estimateurs présentés ci-dessus, des simulations ont été effectuées en reprenant un jeu de matrices de polarisation proposé par Fade et

TAB. 4.1 – Matrices de polarisation.

	Γ_0	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8
a_1	2	15	1	16	82	18	30	2	1.25
a_2	2	6	1	3.6	17	11	14	2	26
a_3	0	0.2	0.4	0	0	7	16	0.6	0
a_4	0	0.5	$\sqrt{0.14}$	0	13	8	8	1.8	5.5
P^2	0	0.2	0.3	0.4	0.5	0.6	0.8	0.9	0.99

Roche [Fad06]. Les 9 matrices de polarisation étudiées, qui correspondent à différentes situations physiques, sont présentées à la table 4.1 et sont classées par ordre croissant de degré de polarisation.

Dans les simulations présentées, les marqueurs + correspondent à la méthode du maximum de vraisemblance (pour 2 ou 3 images), et les marqueurs \times à la méthode des moments (pour 2 ou 3 images). Les marqueurs \diamond représentent la méthode à 4 images (on rappelle que dans ce cas les méthodes du maximum de vraisemblance et des moments donnent les mêmes estimateurs). Les simulations sont obtenues pour des échantillons empiriques distribués selon les matrices Γ_i de taille $n = p \times p$ où p est un entier impair. Ces échantillons correspondent aux réalisations contenues dans une fenêtre d'estimation carrée, centrée sur le pixel d'intérêt et de longueur p , i.e. contenant $n = p \times p$ pixels, sous les hypothèses de stationnarité et d'ergodicité.

La figure 4.4 représente le logarithme des erreurs quadratiques moyennes (EQM) obtenues pour les estimateurs à 2 images en fonction du logarithme de la taille n de la fenêtre d'estimation pour les matrices Γ_2 (figures 4.4(a) et 4.4(b)) et Γ_7 (figures 4.4(c) et 4.4(d)). Les estimateurs du paramètre $\rho = a_3^2 + a_4^2$ et du degré de Polarisation au carré P^2 sont présentés pour chacune de ces matrices (on rappelle que dans le cas à deux images, le vecteur des paramètres inconnus est $\theta = (a_1, a_2, \rho)^T$, et les estimateurs de a_1 et a_2 sont identiques pour les deux méthodes étudiées). Ces EQM ont été calculées à partir de 10000 simulations. Les simulations effectuées avec la matrice Γ_2 , dont le DdP $P^2 = 0.3$ est relativement faible, montrent que les EMV de ρ , ici égal à 0.54 (fig. 4.4(a)), et P^2 (fig. 4.4(b)) sont plus performants que les estimateurs des moments correspondants, même si les EQM obtenues pour les deux méthodes sont du même ordre de grandeur (facteur 2 entre les EQM des estimateurs du DdP). Ce résultat est confirmé pour les EQM associées au paramètre ρ de la matrice Γ_7 (fig. 4.4(c)), dont la valeur théorique est $\rho = 3.243$. Notons cependant que les EQM obtenues ont des niveaux assez élevés, par exemple 1 pour la méthode des moments lorsque $n = 11^2$. En effet, le paramètre ρ , qui représente la covariance entre les intensités I_1 et I_2 , n'est pas une grandeur normée contrairement au DdP, et sa variance augmente avec sa valeur théorique. La figure 4.4(d) montre finalement que pour la matrice Γ_7 dont le DdP $P^2 = 0.9$ est grand, l'EMV de P^2 s'avère largement plus performant que l'estimateur des moments (facteur 100 entre les EQM des estimateurs du DdP). Notons que pour les matrices Γ_3 et Γ_7 considérées, les paramètres a_1 et a_2 sont égaux. Par conséquent, le coefficient de corrélation $r = \frac{a_3^2 + a_4^2}{a_1 a_2}$ entre les intensités I_1 et I_2 et le degré de polarisation au carré P^2 sont égaux. On retrouve donc un phénomène déjà observé lors de l'estimation du coefficient de corrélation d'une loi Gamma bivariée à la section 2.2.3, p. 2.2.3 : la méthode du maximum de vraisemblance est d'autant plus performante par rapport à la méthode des moments

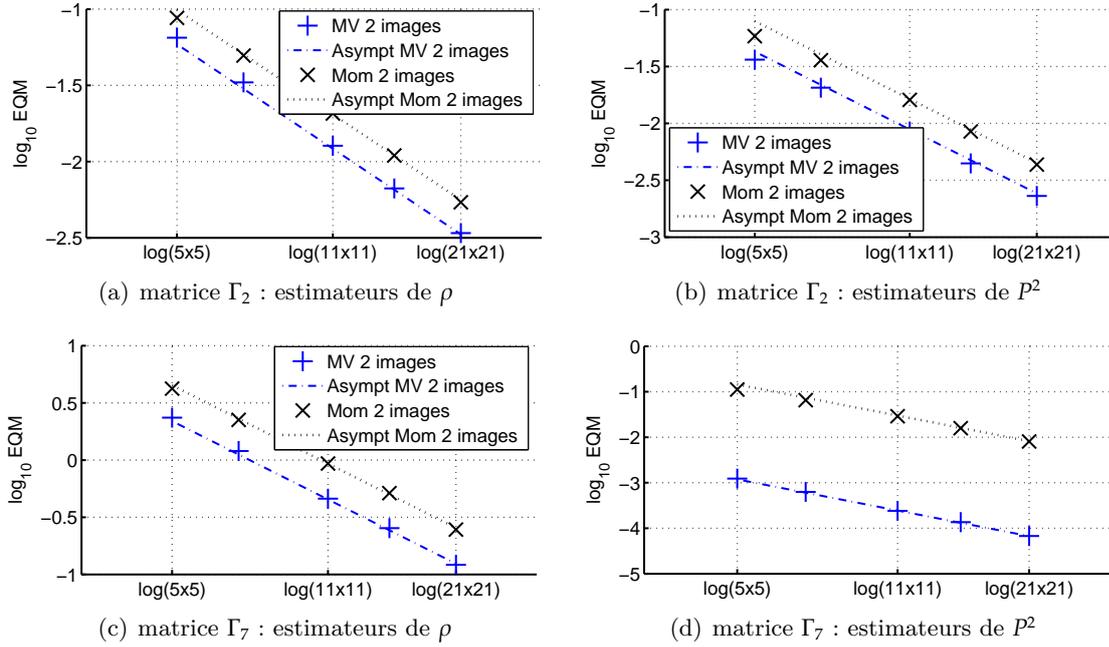


FIG. 4.4 – \log EQM des estimateurs à 2 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7 .

que le coefficient de corrélation r est proche de 1. Les valeurs des variances asymptotiques de chaque estimateur sont également représentées (lignes) sur chacune de ces figures. Les expressions asymptotiques théoriques correspondent parfaitement avec les EQM empiriques pour une taille relativement petite de la fenêtre d'estimation, par exemple $n \geq 11^2$.

La figure 4.5 représente le logarithme des EQM obtenues pour les estimateurs à 3 images en fonction du logarithme de la taille n de la fenêtre d'estimation, pour les matrices Γ_2 (fig. 4.5(a) et 4.5(b)) et Γ_7 (fig. 4.5(c) et 4.5(d)). Dans ce cas, le vecteur des paramètres inconnus est $\boldsymbol{\theta} = (a_1, a_2, a_3, a_4^2)^T$, et les estimateurs des paramètres a_1 , a_2 et a_3 , obtenus pour la méthode du maximum de vraisemblance et la méthode des moindres carrés non linéaires, sont identiques. Les estimateurs comparés sont donc les estimateurs du paramètre a_4^2 et du DdP P^2 . Les résultats obtenus dans le cas à deux images (fig. 4.4) se généralisent au cas à trois images :

- l'EMV de P^2 donne les meilleurs résultats que l'estimateur des moments, notamment pour la matrice Γ_7 dont le DdP est élevé (facteur 10 entre les EQM)
- les variances des estimateurs obtenus converge très rapidement vers leur valeurs asymptotiques théoriques.

La figure 4.6(a) présente le logarithme décimal des erreurs quadratiques moyennes des estimés du DdP à partir de 2 images en fonction de P^2 , pour l'ensemble des matrices Gamma présentées à la table 4.1. La taille de la fenêtre d'estimation est $n = 15 \times 15$. Ces EQM sont comparées à celles obtenues pour la méthode à 4 images (marqueurs \diamond). La dégradation des performances lorsqu'on observe 2 images au lieu de 4 peut être clairement observée. Les variances asymptotiques des estimateurs sont également représentées. Dans un souci de lisibilité du graphique, ces variances asymptotiques sont représentées par des courbes qui relient les valeurs calculées pour chaque matrice Γ_i . Il faut cependant bien noter que mis à part le cas à 4 images où la courbe représente la variance asymptotique exacte pour toutes les valeurs de $P^2 \in [0, 1]$,

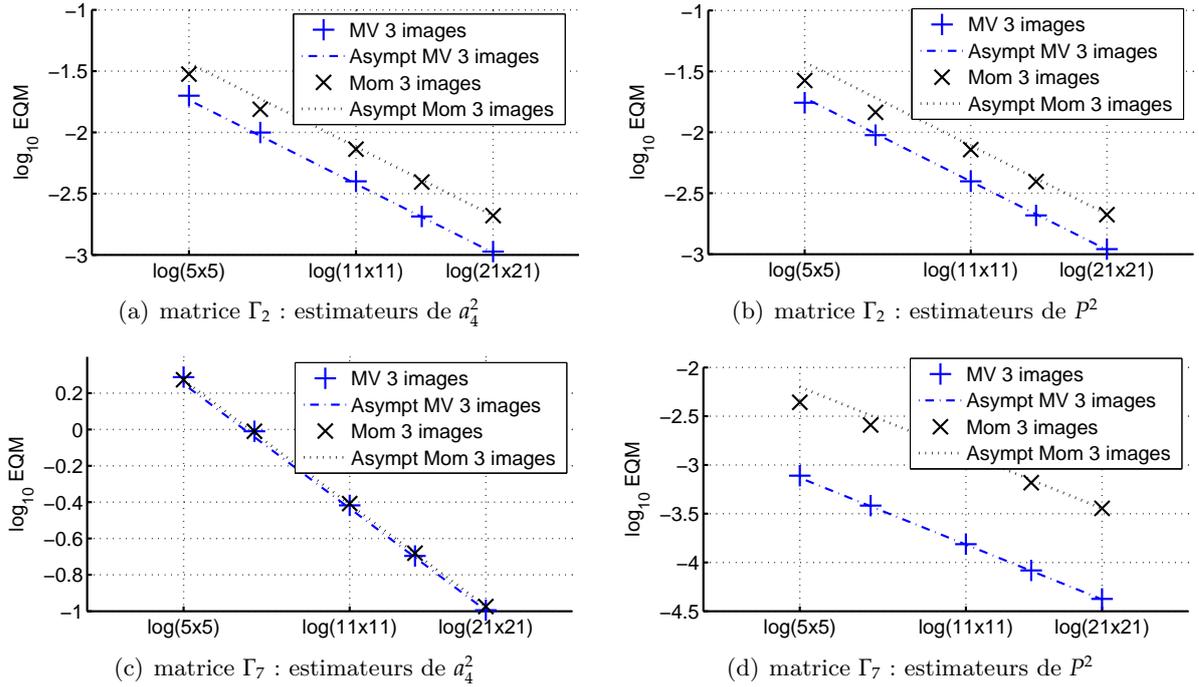


FIG. 4.5 – \log EQM des estimateurs à 3 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7 .

ces variances ne sont pas des fonctions de P^2 . Par conséquent, il est possible de trouver différentes valeurs du vecteur $(a_1, a_2, a_3, a_4)^T$ pour lesquels les DdP sont identiques mais les variances asymptotiques sont différentes. Ces performances asymptotiques correspondent bien à celles des estimées, excepté pour les matrices Γ_0 , Γ_1 et Γ_3 . Ceci s'explique en remarquant que pour les matrices Γ_0 et Γ_3 le paramètre $\rho = a_3^2 + a_4^2$ est nul. Cela signifie que le paramètre appartient à la frontière de son domaine de définition, et les propriétés asymptotiques ne sont alors plus vérifiées. Dans le cas de la matrice Γ_1 , on peut noter que le paramètre de covariance ρ est très faible par rapport aux variances des intensités a_1^2 et a_2^2 . Dans ce cas la valeur asymptotique de la variance n'est pas encore atteinte pour l'échantillon de taille $n = 15 \times 15$. L'EMV de P^2 s'avère nettement plus performant que l'estimateur des moments, particulièrement lorsque le degré de polarisation est élevé. Dans ce cas, les performances de l'EMV sont proches de celles obtenues à 4 images.

La figure 4.6(b) présente le logarithme décimal des erreurs quadratiques moyennes des estimés du DdP à partir de 3 images en fonction de P^2 , pour la méthode des moindres carrés non linéaires et pour la méthode du maximum de vraisemblance. Les variances asymptotiques sont également présentées et correspondent bien aux valeurs estimées mis à part pour les matrices Γ_0 , Γ_1 et Γ_3 (pour les mêmes raisons qu'à deux images, puisque le paramètre a_4^2 est sur la frontière de son domaine, ou très proche). L'EMV de P^2 s'avère plus performant que l'estimateur des moments, et admet pour ce jeu de matrices des performances très proches de l'estimateur à 4 images pour les grandes valeurs du DdP (par exemple $P^2 \geq 0.6$).

Enfin la comparaison des figures 4.6(a) et 4.6(b) montre que sur ce jeu de matrices, les EMV de P^2 obtenus pour 2, 3 ou 4 images ont un comportement similaire, avec un maximum près du point $P^2 = 1/3$ et une variance qui décroît rapidement lorsque P^2 tend vers 1. De plus les EMV à

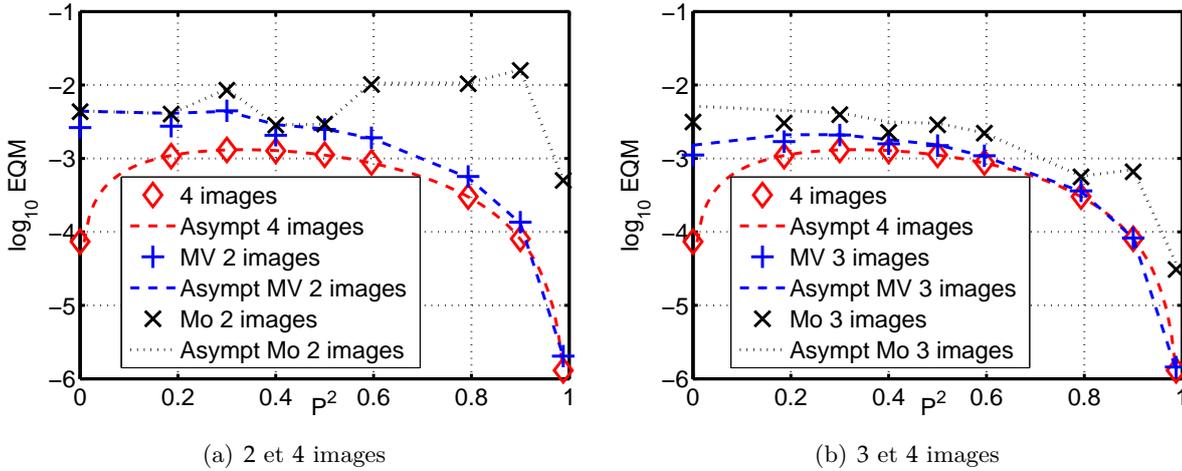


FIG. 4.6 – \log_{10} EQM des estimateurs du DdP à partir de 2, 3 et 4 images en fonction de P^2 ($n = 15 \times 15$).

trois ou même deux images ont des performances proches de celles obtenues pour quatre images. Ce résultat est d'autant plus vrai pour les grandes valeurs du DdP, ce qui est encourageant, car les matériaux que l'on cherche à détecter ont en général un degré de polarisation assez élevé par opposition au fond naturel qui est souvent très dépolarisant. Remarquons finalement que toutes les méthodes d'estimation étudiées sont déterministes si $P^2 = 1$: les données sont alors totalement corrélées et le DdP peut être déterminé de façon exacte. La variance de tous les estimateurs étudiés est donc nulle lorsque $P^2 = 1$, même dans le cas à 2 images.

4.2.4 Estimation à partir d'images synthétiques

La scène polarimétrique décrite ci-dessous a été proposée par Fade et Roche [Fad06] afin d'étudier les performances de la méthode d'estimation à deux images du DdP par la méthode des moments. Cette expérience est reprise dans cette section afin d'étudier le comportement des estimateurs précédemment obtenus sur des images simulées.

La scène polarimétrique étudiée est de dimension 512×512 pixels, et est composée de trois objets distincts. La géométrie de la scène est décrite sur la figure 4.7, et les propriétés polarimétriques des objets et du fond de la scène sont décrites dans la table 4.2. Ces propriétés ont été déterminées afin de correspondre à des situations réelles. La lumière incidente est supposée totalement polarisée suivant l'axe horizontal, et les données simulées correspondent aux quatre intensités mesurées selon le dispositif décrit à la section 4.1.1.

Puisque une seule réalisation du vecteur aléatoire des intensités \mathbf{I} est disponible pour un pixel donné, les images sont supposées localement stationnaires et ergodiques. Ces hypothèses nous permettent de calculer les estimateurs en utilisant des fenêtres d'estimation glissantes, centrées sur le pixel d'intérêt. Les images simulées sont présentées sur la figure 4.8 et sont affichées avec la même dynamique de contraste. Les images correspondant aux intensités I_1 et I_2 utilisées par les méthodes à deux images sont présentées respectivement sur les figures 4.8(a) et 4.8(b). L'intensité totale $I_1 + I_2$, qui représente l'intensité mesurée lorsqu'aucune polarisation n'est effectuée sur la lumière rétrodiffusée, est présentée à titre indicatif sur la figure 4.8(c). Les objets 1 et 2 apparaissent assez nettement sur cette image, contrairement à l'objet 3 dont

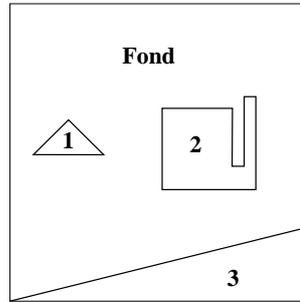


FIG. 4.7 – Composition de la scène polarimétrique.

TAB. 4.2 – Propriétés des objets de la scène polarimétrique.

Objet	Matrice de polarisation	P^2	Intensité totale	Remarques
Fond	$\begin{pmatrix} 0.79 & 0 \\ 0 & 0.98 \end{pmatrix}$	0.0115	1.77	Fond dépolarisateur pur, très dépolarisant et sombre
1	$\begin{pmatrix} 3.6 & 0 \\ 0 & 0.22 \end{pmatrix}$	0.783	3.82	Objet très lumineux et faiblement dépolarisant (par exemple métal)
2	$\begin{pmatrix} 3 & 0.1 \\ 0.1 & 0.6 \end{pmatrix}$	0.447	3.6	Objet lumineux plus dépolarisant
3	$\begin{pmatrix} 0.7 & 0.5 + 0.2i \\ 0.5 - 0.2i & 1.07 \end{pmatrix}$	0.414	1.77	Objet sombre d'intensité égale au fond. Matériau non dépolarisateur pur

la réflectivité moyenne est égale à celle du fond¹⁶. Les images des intensités I_3 , utilisées par les méthodes d'estimations à trois images, et I_4 sont présentées respectivement sur les images 4.8(d) et 4.8(e). Le degré de polarisation théorique est finalement présenté sur la figure 4.8(f). Ces images montrent que la difficulté principale de l'expérience consiste à estimer correctement

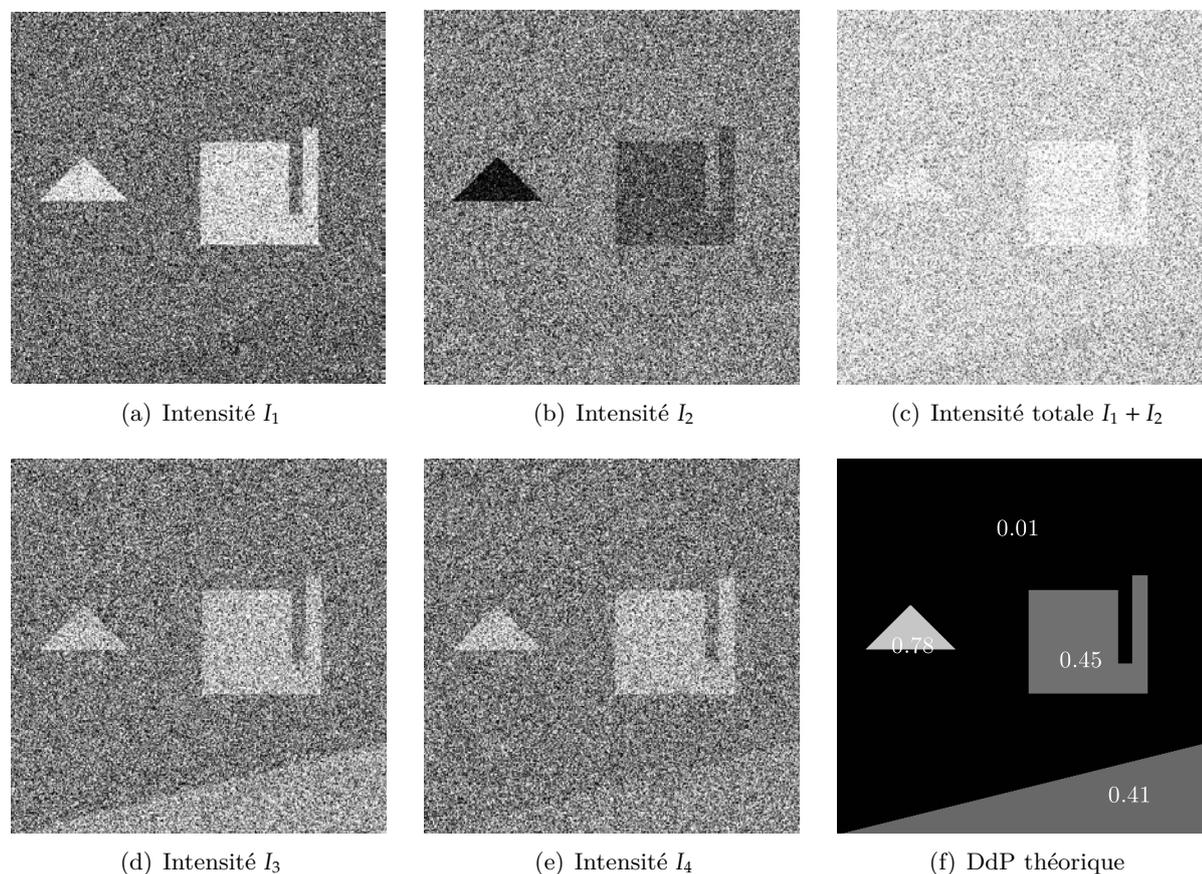


FIG. 4.8 – Images des Intensités mesurées et du DdP.

l'objet 3. Cet objet n'est par exemple pas perceptible sans une analyse statistique lorsque seules les intensités I_1 et I_2 sont disponibles.

Les résultats obtenus pour les différents estimateurs sont présentés sur la figure 4.9. Toutes les figures sont représentées avec la même dynamique de contraste. Les estimateurs ont été obtenus pour une fenêtre d'estimation carrée de taille $n = 15 \times 15$ (i.e. contenant 225 pixels). Les nombres insérés sur les images représentent les moyennes des estimés du DdP sur les objets correspondants. Ces moyennes sont plus faibles que les valeurs théoriques à cause des estimés du DdP obtenus pour les pixels proches du bord. Dans ce cas la fenêtre d'estimation contient des pixels appartenant au fond, ce qui biaise l'estimation.

¹⁶On peut cependant distinguer l'objet 3 sur l'image d'intensité totale 4.8(c). Ceci est dû au fait que les variances des intensités mesurées sur le fond de la scène et sur cet objet sont différentes. Cette observation montre qu'il est possible d'estimer des grandeurs caractéristiques de l'état de polarisation de la lumière à partir de la seule image de l'intensité totale. La loi de cette intensité correspond à la somme des marges d'une loi Gamma bivariée. Des méthodes des moments ou de maximum de vraisemblance peuvent alors être mises en œuvre afin d'estimer le DdP [RFR07b].

Méthodes à 2 images La méthode des moments (fig 4.9(a)) et la méthode du maximum de vraisemblance (fig 4.9(b)) donnent des résultats similaires sur le fond de l'image. Cependant l'objet 2 est mieux restitué pour la méthode du maximum de vraisemblance. Les valeurs estimées du DdP sont plus uniformes pour l'EMV. De plus, la méthode des moments présente des effets de bord sur chaque objet plus importants que ceux obtenus pour l'EMV, et semble donc moins robuste aux inhomogénéités. Enfin l'objet 3 est bien mieux estimé par la méthode MV : la variance des estimés est en effet plus faible et la frontière avec le fond plus nette.

Méthodes à 3 images Dans ce cas, le fond de l'image est mieux restitué pour la méthode du maximum de vraisemblance (fig 4.9(d)) que pour la méthode des moindres carrés non linéaires (fig 4.9(c)). De manière générale, l'EMV donne de meilleurs résultats. En particulier, l'objet 3 est mieux estimé pour la méthode du maximum de vraisemblance (variance plus faible, contour plus net).

L'image obtenue pour l'estimateur à partir de 4 images (fig 4.9(e)) donne finalement, les meilleurs résultats. Cependant, l'image de l'EMV pour 3 images (fig 4.9(d)) est très proche de cette dernière, et l'EMV pour 2 images fournit des résultats corrects qui peuvent être exploités. Ceci montre qu'il est possible de travailler à partir d'un nombre réduit d'images, sans une trop grande dégradation des performances.

4.3 Imagerie polarimétrique "faible-flux"

Lorsque les niveaux d'intensité sont trop faibles pour être mesurés directement, il s'avère nécessaire de considérer la théorie semi-classique de l'optique où les mesures correspondent aux nombres de photons détectés par des gamma-caméras. Ces nombres de photons collectés sont alors distribués selon des lois de Poisson mélangées multivariées (LPMM) (voir définition 1.3.5, p. 42) dont la loi de mélange est la loi du vecteur des intensités lumineuses [GRR03, RRG06a]. L'opération de mélange de la loi de Poisson par la loi des intensités est également appelée en optique la *transformation de Poisson-Mandel* des intensités [Goo85]. Cette opération a pour conséquence d'augmenter la variance des v.a. observées, c'est pourquoi on parle de *bruit de Poisson*.

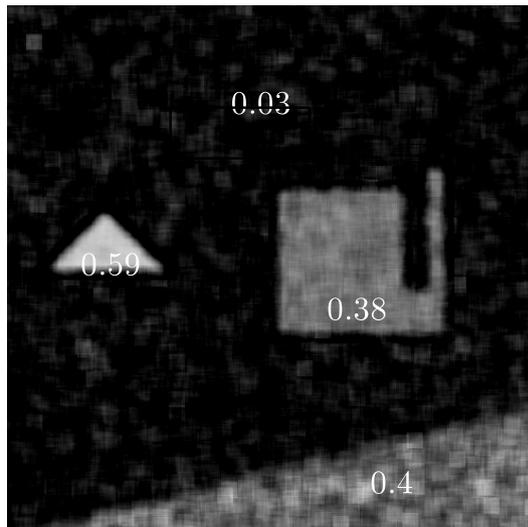
4.3.1 Modèle statistique

On note dans cette section $\mathbf{N} = (N_1, N_2, N_3, N_4)^T \in \mathbb{N}^d$ le vecteur des intensités photonisées. Il a été montré à la section 1.3.4, p. 45, que lorsque la loi de mélange est une loi Gamma multivariée $\mathcal{G}(q, P)$, le vecteur \mathbf{N} est distribué suivant une loi multinomiale négative caractérisée par sa fonction génératrice :

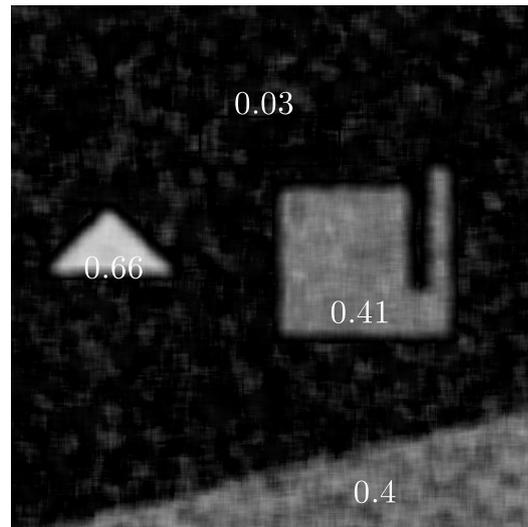
$$G_{\mathbf{N}}(\mathbf{z}) = (P(\mathbf{1} - \mathbf{z}))^{-q}.$$

D'après la transformée de Laplace (4.12) de la loi du vecteur des intensités $\mathbf{I} \in \mathbb{R}^4$, on obtient le vecteur \mathbf{N} , distribué suivant la loi multinomiale négative de paramètre de forme $q = 1$ et de polynôme affine :

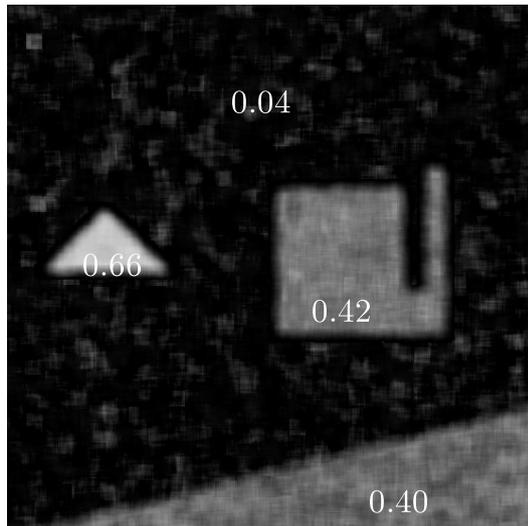
$$P_1(\mathbf{z}) = 1 + \mathbf{m}^T(\mathbf{1} - \mathbf{z}) + k[2(1 - z_1)(1 - z_2) + (1 - z_3)(1 - z_4) + (1 - z_1)(1 - z_3) + (1 - z_1)(1 - z_4) + (1 - z_2)(1 - z_4) + (1 - z_3)(1 - z_4)].$$



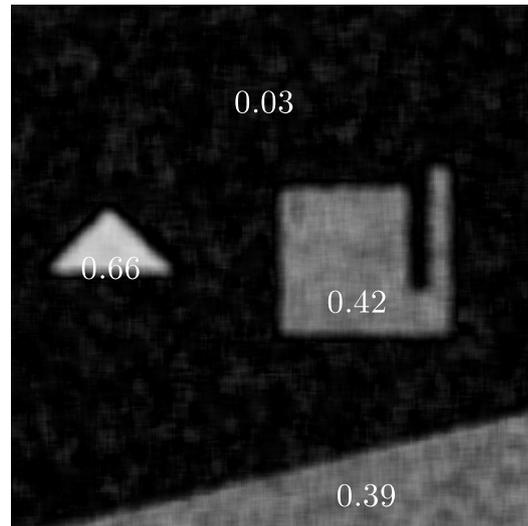
(a) Moments 2 images



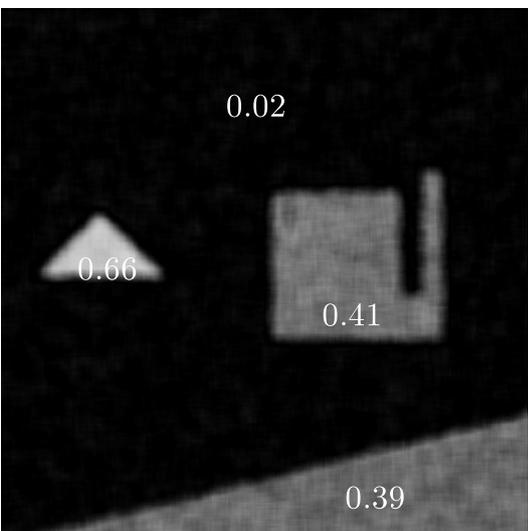
(b) MV 2 images



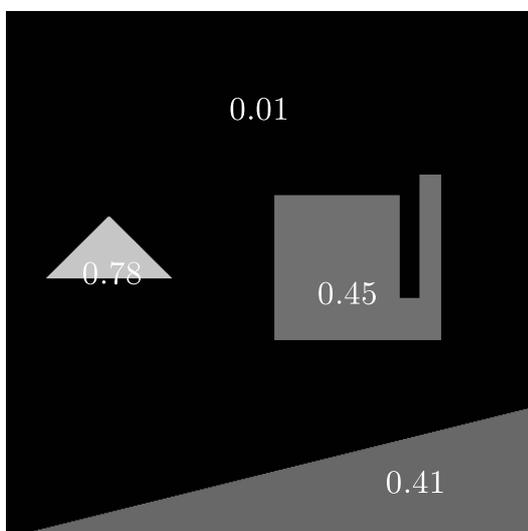
(c) Moments 3 images



(d) MV 3 images



(e) 4 images



(f) DdP théorique

FIG. 4.9 – Estimées du DdP P^2 sur la scène simulée (les nombres insérés représentent les moyennes des estimées de P^2 pour chaque objet) ($n = 15 \times 15$).

où $\mathbf{z} = (z_1, z_2, z_3, z_4)^T \in \mathbb{R}^4$, $\mathbf{m} = (a_1, a_2, \frac{a_1+a_2}{2} + a_3, \frac{a_1+a_2}{2} + a_4)^T$, et $k = (a_1 a_2 - a_3^2 - a_4^2)/2$. La distribution du vecteur \mathbf{N} est donc paramétrée par les coefficients $\mathbf{a} = (a_1, a_2, a_3, a_4)^T$ de la matrice de polarisation Γ . Selon la proposition 1.3.1, p. 30, les sous vecteurs du vecteur \mathbf{N} suivent des lois multinomiales négatives. Leur fonction génératrice est obtenue en posant $z_i = 1$ si la i -ème marge du vecteur \mathbf{N} n'appartient pas au sous-vecteur étudié, pour $1 \leq i \leq 4$. De manière analogue au cas "fort-flux", on trouve que le degré de polarisation peut être estimé à partir du vecteur \mathbf{N} , des triplets $(N_1, N_2, N_3)^T$ ou $(N_1, N_2, N_4)^T$ ou à partir du couple $(N_1, N_2)^T$.

4.3.2 Estimation de la matrice de polarisation et du DdP

Comme le montre le théorème 1.3.13, p. 44, la loi du vecteur \mathbf{N} ne décrit pas en général une famille exponentielle naturelle (les deux cas particuliers où la famille de lois obtenue est une FEN feront l'objet d'une remarque). Par conséquent, dans le modèle à 4 images, les moyennes empiriques ne sont pas a priori des estimateurs efficaces de la matrice de polarisation Γ , contrairement au modèle "fort-flux". Une autre difficulté rencontrée est qu'il n'existe pas d'expression analytique explicite des masses de probabilité d'une telle loi mis à part dans le cas bivarié. Pour ces raisons, une méthode d'estimation reposant sur le principe du maximum de vraisemblance composite est étudiée.

Maximum de vraisemblance composite

Une vraisemblance composite est une combinaison linéaire de log-vraisemblances associées à des événements marginaux ou conditionnels. Ce concept a été largement étudié dans la littérature (voir [CR03], [ZJ05], [GH87] et les références proposées) depuis l'article fondateur de Lindsay [Lin88]. Parmi les vraisemblances composites classiques, on peut citer la vraisemblance marginale, la vraisemblance des paires, ou la pseudo-vraisemblance de Besag. L'estimateur du maximum de vraisemblance composite (EMVC) est obtenu en maximisant la vraisemblance composite (VC) correspondante. L'intérêt d'utiliser une telle méthode, par rapport à une estimation du maximum de vraisemblance standard, est de réduire la complexité de la procédure d'optimisation numérique. Par conséquent, cette méthode permet d'estimer les paramètres de modèles complexes, même lorsque la densité jointe n'admet pas d'expression analytique explicite. C'est le cas de la loi du vecteur des intensités photonisées $\mathbf{N} \in \mathbb{N}^4$, étudiée dans cette section. Le choix d'une log-vraisemblance composite pour un problème donné se fait alors en fonctions des critères suivants :

1. la vraisemblance composite doit dépendre des paramètres que l'on cherche à estimer,
2. la complexité calculatoire doit être la plus petite possible.

Puisqu'il est difficile pour le problème d'imagerie "faible-flux" d'obtenir une expression des masses de probabilités jointes $\Pr(\mathbf{n}, \boldsymbol{\theta})$ en fonction du vecteur des paramètres $\boldsymbol{\theta}$, en dimension supérieure à 2, nous pouvons considérer toutes les probabilités associées aux paires (N_k, N_l) , pour $1 \leq k < l \leq d$. Une expression explicite de ces probabilités est en effet disponible d'après le théorème 1.3.8, p. 41. Cette approche conduit à étudier une vraisemblance composite classique, qualifiée de *log-vraisemblance des paires*. L'EMVC $\widehat{\boldsymbol{\theta}}_n$ est alors la valeur de $\boldsymbol{\theta}$ qui minimise la fonction suivante :

$$U_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n l(\mathbf{N}^{(i)}, \boldsymbol{\theta}), \quad (4.33)$$

où

$$l(\mathbf{n}, \boldsymbol{\theta}) = \sum_{1 \leq k < l \leq d} \log p_{k,l}(n_k, n_l, \boldsymbol{\theta}), \quad (4.34)$$

avec $p_{k,l}(n_k, n_l, \boldsymbol{\theta}) = \Pr(N_k = n_k, N_l = n_l)$ lorsque le paramètre du modèle est $\boldsymbol{\theta}$.

Propriétés asymptotiques : Cette section étudie la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance composite $\widehat{\boldsymbol{\theta}}_n$, lorsque le vecteur \mathbf{N} suit la LPMM dont la loi de mélange est une loi Gamma multivariée de paramètre $\boldsymbol{\theta}_0$ (la loi du vecteur \mathbf{N} appartient donc à une famille de lois multinomiales négatives). On considère ici le cas général où la loi est définie sur \mathbb{N}^d . Les propriétés asymptotiques sont montrées dans le cas où le paramètre de forme q , de la loi Gamma multivariée est connu, ce qui est le cas de l'imagerie polarimétrique ($q = 1$).

Sous les conditions suivantes :

1. L'espace des paramètres Θ est un compact de \mathbb{R}^p . Le vecteur $\boldsymbol{\theta}_0$ appartient à l'intérieur du compact Θ ,
2. Soit $\Delta = \{(a, b, c) \in [0, 1]^3; (1 - a)(1 - b) > c\}$ l'ensemble des valeurs admissibles des paramètres d'une loi multinomiale négative bivariée, pour la paramétrisation par le vecteur (a, b, c) présenté à la section 1.3.3, p. 40. On note alors $F_{k,l}$ la fonction de Θ vers l'ensemble Δ qui relie le vecteur des paramètres $\boldsymbol{\theta}$ aux paramètres $(a_{k,l}, b_{k,l}, c_{k,l})$ de la multinomiale négative du couple (N_k, N_l) , pour $1 \leq k < l \leq d$. La fonction $F(\boldsymbol{\theta}) = (F_{1,2}(\boldsymbol{\theta})^T, \dots, F_{d-1,d}(\boldsymbol{\theta})^T)^T$ est injective de Θ dans $\Delta^{d(d-1)/2}$,
3. Les fonctions $F_{k,l}$ sont deux fois différentiables et continues.

on peut démontrer le théorème suivant.

Théorème 4.3.1. *L'estimateur du maximum de vraisemblance composite $\widehat{\boldsymbol{\theta}}_n$ converge presque sûrement vers $\boldsymbol{\theta}_0$. De plus, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converge en loi vers la loi normale centrée, de matrice de covariance $V = I_U(\boldsymbol{\theta}_0)^{-1} \Gamma_U(\boldsymbol{\theta}_0) I_U(\boldsymbol{\theta}_0)^{-1}$, où*

$$I_U(\boldsymbol{\theta}_0)_{u,v=1,\dots,p} = - \sum_{1 \leq k < l \leq d} \mathbb{E}_{\boldsymbol{\theta}_0} \left(\frac{\partial}{\partial \boldsymbol{\theta}_u} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}_v} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \right),$$

et

$$\Gamma_U(\boldsymbol{\theta}_0)_{u,v=1,\dots,p} = \mathbb{E}_{\boldsymbol{\theta}_0} \left(\sum_{1 \leq k < l \leq d} \frac{\partial}{\partial \boldsymbol{\theta}_u} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \sum_{1 \leq r < s \leq d} \frac{\partial}{\partial \boldsymbol{\theta}_v} \log p_{r,s}(N_r, N_s, \boldsymbol{\theta}_0) \right).$$

L'hypothèse 1) n'est pas très restrictive : même si l'ensemble des paramètres Θ n'est pas compact, le théorème reste vrai sur tout compact de cet ensemble, et donc pour tout $\boldsymbol{\theta}_0$ dans l'intérieur de Θ . L'hypothèse 2) correspond aux conditions d'indentifiabilité du paramètre $\boldsymbol{\theta}$ du modèle, en fonction des lois des paires.

La démonstration de ce théorème, donnée en annexe D.4, s'appuie sur la théorie des estimateurs du minimum de contraste (voir [DCD84]). Un point clé de la démonstration est l'existence des moments d'ordre 1 et 2 d'une loi multinomiale négative bivariée. Remarquons que la matrice $-I_U(\boldsymbol{\theta}_0)$ est la somme des matrices d'information de Fisher associées aux paires (N_k, N_l) . La matrice V^{-1} est appelée *matrice d'information de Godambe* par analogie avec la matrice d'information de Fisher. Dans le cas où le vecteur \mathbf{N} est dans \mathbb{N}^2 , les deux matrices coïncident (et dans ce cas la vraisemblance de \mathbf{N} et la vraisemblance des paires sont bien sûr identiques).

La méthode du maximum de vraisemblance composite proposée permet finalement d'estimer le modèle lorsque plus de 2 images d'intensités sont disponibles. Dans cette section et dans les simulations, on s'intéresse au cas à quatre images, mais le cas à trois images se traite de façon similaire. Lorsque seulement deux images d'intensité (N_1, N_2) sont disponibles, alors la méthode du maximum de vraisemblance peut être appliquée. Le cas à deux images, plus simple, est traité en premier. La méthode du maximum de vraisemblance composite se déduit des expressions des vraisemblances trouvées dans le cas à deux images.

Cas à deux images

Dans ce cas, les images d'intensité étudiées correspondent aux réalisations des v.a. N_1 et N_2 . Comme pour l'imagerie "fort-flux", le vecteur des paramètres inconnus est $(a_1, a_2, \rho = a_3^2 + a_4^2)^T$.

Méthode du maximum de vraisemblance : Il a été montré (voir l'équation (1.79), p. 46) que la loi multinomiale négative bivariée (MNB) du vecteur des intensités photonisées (N_1, N_2) peut être reparamétrisée en fonction des coefficients (a, b, c) , introduits à l'équation (1.65), p. 40. À l'aide de l'expression (4.12), p. 113, de la transformée de Laplace du vecteur \mathbf{I} , et de la relation (1.78), p. 45, entre les paramètres d'une loi Gamma bivariée et les paramètres (a, b, c) de la loi multinomiale négative correspondante (i.e. la LPMM obtenue lorsque la loi de mélange est cette loi Gamma bivariée), on obtient :

$$\begin{aligned} a &= \frac{a_1 + a_1 a_2 - \rho}{1 + a_1 + a_2 + a_1 a_2 - \rho}, \\ b &= \frac{a_2 + a_1 a_2 - \rho}{1 + a_1 + a_2 + a_1 a_2 - \rho}, \\ c &= \frac{\rho}{(1 + a_1 + a_2 + a_1 a_2 - \rho)^2} \end{aligned} \quad (4.35)$$

Mais, lorsque le paramètre $\tau = \frac{c}{ab}$ est fixé, la loi MNB décrit une famille exponentielle naturelle (voir p. 41). Les estimateurs des moyennes sont donc les moyennes empiriques. Un raisonnement identique à celui effectué, par exemple, pour les lois Gamma bivariées (voir 2.2.1, p. 57) montre que ce résultat reste vrai lorsque le paramètre τ est inconnu. Les paramétrisations par les vecteurs $(a, b, \tau)^T$ et $(a_1, a_2, \rho)^T$ étant équivalentes, et les moyennes d'une LPMM étant égales à celles de la loi de mélange, d'après la propriété (1.73), le principe d'invariance fonctionnelle assure :

$$\widehat{a}_1 = \frac{1}{n} \sum_{j=1}^n N_1^j, \quad \widehat{a}_2 = \frac{1}{n} \sum_{j=1}^n N_2^j. \quad (4.36)$$

L'estimateur de ρ est obtenu en minimisant la log-vraisemblance, les paramètres a_1 et a_2 étant remplacés par leurs estimés. Cette log-vraisemblance se calcule grâce à l'expression (4.35) du vecteur $(a, b, c)^T$ en fonction du vecteur $\boldsymbol{\theta} = (a_1, a_2, \rho)^T$, et à l'aide de l'expression (1.69), p. 41, des masses de probabilité d'une loi MNB en fonction du triplet (a, b, c) . Comme dans le modèle "fort-flux", il n'existe pas d'expression analytique explicite de l'estimateur de ρ . Cet estimateur est donc obtenu numériquement par une méthode de gradient de type Levenberg-Marquard.

Finalement, la variance asymptotique de l'EMV du DdP, égale à la borne de Cramer-Rao, est calculée de façon similaire au cas "fort-flux", à partir de la matrice d'information de Fisher, cette dernière étant calculée par une méthode de Monte-Carlo (voir section 4.2.2).

Méthode des moments : L'expression (1.73), p. 43, des moments d'une LPMM montre que les moyennes et la covariance de cette loi sont les mêmes que celles de la loi de mélange. Les estimateurs des moments, calculés à partir des moyennes et de la covariance, sont donc identiques à ceux obtenus en "fort-flux" :

$$\widehat{a}_1 = \frac{1}{n} \sum_{j=1}^n N_{1'}^j, \quad \widehat{a}_2 = \frac{1}{n} \sum_{j=1}^n N_{2'}^j,$$

$$\widehat{\rho}_{\text{Mo}} = \frac{1}{n} \sum_{j=1}^n N_{1'}^j N_{2'}^j - \widehat{a}_1 \widehat{a}_2.$$

La matrice de covariance des estimateurs des moments peut se calculer de façon explicite à partir des formules des moments d'une LPMM (1.72) et des moments d'une loi Gamma bivariée (1.19) :

$$C = \text{var}_A \left((\widehat{a}_1, \widehat{a}_2, \widehat{\rho}_{\text{Mo}})^T \right) = \frac{1}{n} \begin{pmatrix} a_1^2 + a_1 & \rho & 2a_1\rho + \rho \\ \rho & a_2^2 + a_2 & 2a_2\rho + \rho \\ 2a_1\rho + \rho & 2a_2\rho + \rho & c_{3,3} \end{pmatrix},$$

où

$$c_{3,3} = \frac{a_1^2 a_2^2 + 4a_1 a_2 \rho + 3\rho^2 + a_1 a_2 (1 + a_1 + a_2) + 2\rho(a_1 + a_2 + 1/2)}{n}.$$

L'expression de la variance asymptotique de cet estimateur est alors donnée par la formule :

$$\text{var}_A \left(\widehat{P}^{\text{Mom}} \right) = G_2^T C G_2,$$

où G_2 est le gradient de P^2 défini à l'équation (4.30).

Cas à quatre images

Le vecteur des paramètres inconnus est maintenant $\boldsymbol{\theta} = (a_1, a_2, a_3, a_4)^T$.

Méthode du maximum de vraisemblance composite : Afin de pouvoir utiliser les résultats démontrés au théorème 4.3.1, il est nécessaire de montrer que la fonction F , associant au vecteur $\boldsymbol{\theta} = (a_1, a_2, a_3, a_4)^T$ le vecteur $((a_{k,l}, b_{k,l}, c_{k,l})_{1 \leq k < l \leq 4})$, est injective. Mais l'expression de la transformée de Laplace du vecteur \mathbf{I} (4.12) et la relation (1.79), p. 46, entre les paramètres $(a_{k,l}, b_{k,l}, c_{k,l})$ et les paramètres de la loi de mélange montrent :

$$m_1 = \frac{a_{1,2}(1 - b_{1,2}) + c_{1,2}}{(1 - a_{1,2})(1 - b_{1,2}) - c_{1,2}}, \quad m_2 = \frac{b_{1,2}(1 - a_{1,2}) + c_{1,2}}{(1 - a_{1,2})(1 - b_{1,2}) - c_{1,2}},$$

$$m_3 = \frac{a_{1,3}(1 - b_{1,3}) + c_{1,3}}{(1 - a_{1,3})(1 - b_{1,3}) - c_{1,3}}, \quad m_4 = \frac{a_{1,4}(1 - b_{1,4}) + c_{1,4}}{(1 - a_{1,4})(1 - b_{1,4}) - c_{1,4}},$$

où $\mathbf{m} = (m_1, m_2, m_3, m_4)^T = (a_1, a_2, (a_1 + a_2)/2 + a_3, (a_1 + a_2)/2 + a_4)^T$. Le vecteur $\boldsymbol{\theta} = (a_1, a_2, a_3, a_4)$ est bien identifiable à partir de la loi des marges, et la fonction F est injective. De plus, en notant :

$$p_{k,l} = \begin{cases} (a_1 a_2 - a_3^2 - a_4^2), & \text{si } (k, l) = (1, 2), \\ (a_1 a_2 - a_3^2 - a_4^2)/2, & \text{sinon.} \end{cases}$$

la relation (1.78) montre, pour tout $1 \leq k < l \leq 4$:

$$\begin{aligned} a_{k,l} &= \frac{m_k + p_{k,l}}{1 + m_k + m_l + p_{k,l}}, \\ b_{k,l} &= \frac{m_l + p_{k,l}}{1 + m_k + m_l + p_{k,l}}, \\ c_{k,l} &= \frac{m_k m_l - p_{k,l}}{(1 + m_k + m_l + p_{k,l})^2}. \end{aligned} \quad (4.37)$$

La fonction F est donc régulière, deux fois continue et différentiable. Finalement, les conditions d'application du théorème 4.3.1 sont vérifiées, et les EMVC des paramètres $\boldsymbol{\theta} = (a_1, a_2, a_3, a_4)$, obtenus en minimisant la log-vraisemblance des paires (4.33) (qui contient ici les 6 combinaisons possibles des paires du vecteur $\mathbf{N} = (N_1, N_2, N_3, N_4)$), sont convergents. Notons cependant que contrairement à l'EMV obtenu pour 2 images, aucune expression analytique des EMVC n'est disponible. La log-vraisemblance obtenue doit donc être maximisée numériquement (procédure de type Levenberg-Marquardt) par rapport aux quatre paramètres a_1, a_2, a_3 et a_4 .

Enfin, la matrice de covariance asymptotique V des estimateurs V peut être calculée en estimant les matrices I_U et Γ_U , définies au théorème 4.3.1, par des méthodes de Monte-Carlo, comme cela a été expliqué pour la matrice le calcul de la matrice de Fisher en fort-flux. Ceci permet d'obtenir la variance asymptotique de l'estimateur du DdP ainsi obtenu :

$$\text{var}_A(\widehat{P^2}) = G_4^T V G_4,$$

où G_4 est le gradient de P^2 défini à l'équation (4.21).

Méthode des moments : Comme cela a déjà été signalé, les moyennes des intensités photonisées N_i sont égales à celles des intensités "fort-flux" I_i , pour $1 \leq i \leq 4$. Par conséquent, l'estimateur des moments "faible-flux" a une expression identique à celui obtenu en fort-flux :

$$\widehat{\mathbf{a}}_{\text{Mo}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/2 & -1/2 & 1 & 0 \\ -1/2 & -1/2 & 0 & 1 \end{pmatrix} \begin{pmatrix} \widehat{m}_1 \\ \widehat{m}_2 \\ \widehat{m}_3 \\ \widehat{m}_4 \end{pmatrix},$$

où $\widehat{m}_i = \frac{1}{n} \sum_{j=1}^n N_i^j$. La covariance des estimateurs est facilement obtenue grâce à la relation (1.73), p. 43, entre la matrice de covariance d'une LPMM et la matrice de covariance de sa loi de mélange :

$$\text{cov}(\widehat{\mathbf{a}}_{\text{Mo}}) = \frac{1}{n} \begin{pmatrix} a_1^2 + a_1 & a_3^2 + a_4^2 & a_1 a_3 & a_1 a_4 \\ a_3^2 + a_4^2 & a_2^2 + a_2 & a_2 a_3 & a_2 a_4 \\ a_1 a_3 & a_2 a_3 & c_{3,3} & a_3 a_4 \\ a_1 a_4 & a_2 a_4 & a_3 a_4 & c_{4,4} \end{pmatrix}, \quad (4.38)$$

où $c_{3,3} = (a_1 + a_2 + 2a_3 + a_1 a_2 + a_3^2 - a_4^2)/2$ et $c_{4,4} = (a_1 + a_2 + 2a_4 + a_1 a_2 - a_3^2 + a_4^2)/2$. L'estimateur du DdP ainsi construit, admet alors la variance asymptotique suivante :

$$\text{var}_A(\widehat{P^2}_{\text{Mo}}) = G_4^T \text{cov}(\widehat{\mathbf{a}}_{\text{Mo}}) G_4.$$

La formule ci-dessus permet d'avoir une expression explicite de la variance asymptotique du DdP estimé par une méthode de moments lorsque 4 images d'intensités sont disponibles. Contrairement au cas "fort-flux", l'expression scalaire trouvée n'admet pas une forme simple, et est donc laissée sous forme matricielle. En particulier, cette expression n'est plus une fonction du seul degré de polarisation P^2 , mais dépend de tous les paramètres a_1, a_2, a_3 et a_4 .

Remarque :

Il est intéressant d'observer les cas particuliers où la loi du vecteur des intensités photonisées décrit une famille exponentielle. D'après le théorème 1.3.13, ces cas sont obtenus lorsque les lois marginales des intensités fort-flux sont soit indépendantes, soit totalement corrélées. Le premier cas est obtenu pour le modèle des matériaux purement dépolarisant. La matrice de polarisation associée à ces matériaux est en effet une matrice diagonale ($a_3 = a_4 = 0$), et pour ce modèle, l'état de polarisation de la lumière est caractérisé par les deux mesures fort-flux I_1 et I_2 , qui sont indépendantes. Les estimateurs efficaces des paramètres a_1 et a_2 correspondent donc aux moyennes empiriques des réalisations des v.a. N_1 et N_2 , et la matrice de covariance de ces estimateurs fournit une expression analytique simple de la borne de Cramer-Rao des estimateurs du DdP [RRG06a]. Le deuxième cas correspond à des matériaux qui ne dépolarisent pas la lumière, $P^2 = 1$, et est donc sans intérêt, puisque les propriétés polarimétriques sont totalement déterminées pour ce modèle.

4.3.3 Estimation sur données synthétiques

On reprend dans cette section le jeu de matrice étudié en "fort-flux" et présenté à la table 4.1. Une 10ème matrice est toutefois ajoutée :

$$\Gamma_9 = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}. \tag{4.39}$$

Cette matrice, qui est singulière, correspond à un DdP $P^2 = 1$, et donc à des intensités fort-flux totalement corrélées. En raison du bruit de Poisson, ce DdP ne peut plus être obtenu de manière déterministe, et doit maintenant être estimé. Sur les figures de cette section, les marqueurs + sont utilisés pour la méthode du maximum de vraisemblance pour 2 images ou la méthode de vraisemblance composite pour 4 images. Les marqueurs × correspondent à la méthode des moments (pour 2 ou 4 images). Le nombre de simulations effectuées afin de calculer les EQM empiriques, est 10000 pour les méthodes basées sur 2 images, et 1000 pour les méthodes basées sur 4 images.

La figure 4.10 représente le logarithme des erreurs quadratiques moyennes (EQM) obtenues pour les estimateurs à 2 images en fonction du logarithme de la taille n de la fenêtre d'estimation, pour les matrices Γ_2 (figures 4.10(a) et 4.10(b)) et Γ_7 (figures 4.10(c) et 4.10(d)). Les estimateurs du paramètre $\rho = a_3^2 + a_4^2$ et du DdP P^2 sont comparés pour chacune des matrices (les estimateurs de a_1 et a_2 sont identiques pour les deux méthodes étudiées). On retrouve un comportement similaire au cas "fort-flux" : l'EMV s'avère plus performant sur ces matrices, notamment pour la matrice Γ_7 dont le DdP $P^2 = 0.9$ est élevé (le DdP de la matrice Γ_3 est $P^2 = 0.3$). Les variances asymptotiques sont également représentées (lignes) et correspondent avec les EQM empiriques pour une fenêtre de taille $n \geq 11 \times 11$.

La figure 4.11 représente le logarithme des erreurs quadratiques moyennes (EQM) obtenues pour les estimateurs à 4 images en fonction du logarithme de la taille n de la fenêtre d'estimation

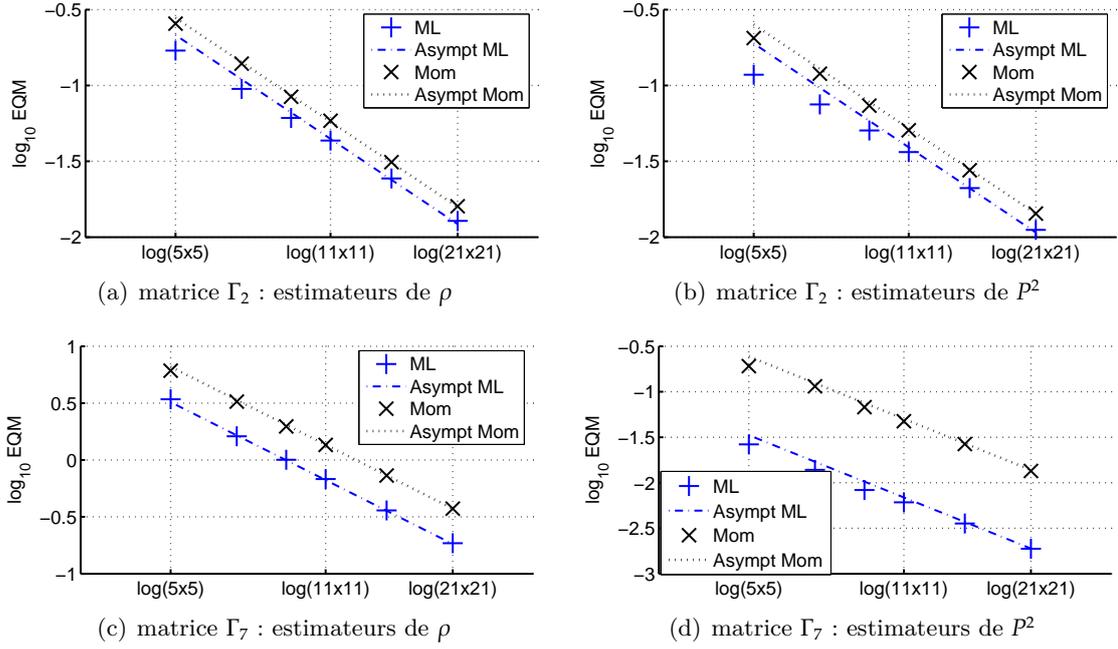


FIG. 4.10 – \log EQM des estimateurs à 2 images en fonction du logarithme de la taille des échantillons pour les matrices Γ_2 et Γ_7).

pour les matrices Γ_2 . Dans ce cas, les estimateurs du maximum de vraisemblance composite des paramètres a_1 , a_2 , a_3 et a_4 diffèrent de ceux obtenus par la méthode des moments. Les paramètres a_1 et a_2 jouant un rôle symétrique dans notre modèle, et les valeurs théoriques de ces paramètres étant égales pour les matrices Γ_2 et même Γ_7 , seules les EQM de a_1 sont représentées (fig. 4.11(a)). Les estimateurs de a_1 des deux méthodes ont des performances similaires, ce qui n'est plus le cas pour les paramètres a_3 (fig. 4.11(b)) et a_4 (fig. 4.11(c)), où l'EMVC est plus performant. Finalement, l'EMVC du DdP (fig. 4.11(d)) est plus performant, notamment pour les fenêtres de petite taille. Ces résultats sont confirmés sur la figure 4.11 pour la matrice Γ_7 . Les variances asymptotiques sont également représentées (lignes). Les variances des estimés correspondent à la variance asymptotique pour des fenêtres de taille relativement faible (par exemple $n \geq 15 \times 15$). En outre, ces variances asymptotiques confirment que la méthode des moments obtenue pour 4 images n'est pas efficace en faible-flux, puisque sa variance asymptotique est supérieure à celle obtenue pour l'EMVC.

Les logarithmes des EQM, obtenus pour les 10 matrices considérées, sont représentés sur les figures 4.13(a) (méthodes basées sur 2 images) et 4.13(b) (méthodes basées sur 4 images). La taille des fenêtres d'estimation est $n = 15 \times 15$. Les méthodes de vraisemblance donnent de meilleurs résultats que les méthodes des moments, pour les DdP élevés. Remarquons cependant que le comportement global des courbes obtenues pour ce jeu de matrice diffère de celui obtenu en fort-flux. En effet les EQM ne sont pas des "fonctions" décroissantes du DdP pour ces matrices. Un autre paramètre doit être pris en compte dans le modèle faible-flux afin d'interpréter ces résultats. Les matrices étudiées correspondent en effet à des niveaux d'intensité, i.e. $I_T = a_1 + a_2$, très disparates. En comparant par exemple les EQM pour les matrices Γ_6 ($P^2 = 0.8$, $I_T = 44$), Γ_7 ($P^2 = 0.9$, $I_T = 4$) et Γ_8 ($P^2 = 0.99$, $I_T = 27.25$), on constate que les performances sont meilleures pour les grandes intensités. En effet, dans le modèle "faible-flux", les v.a. observées

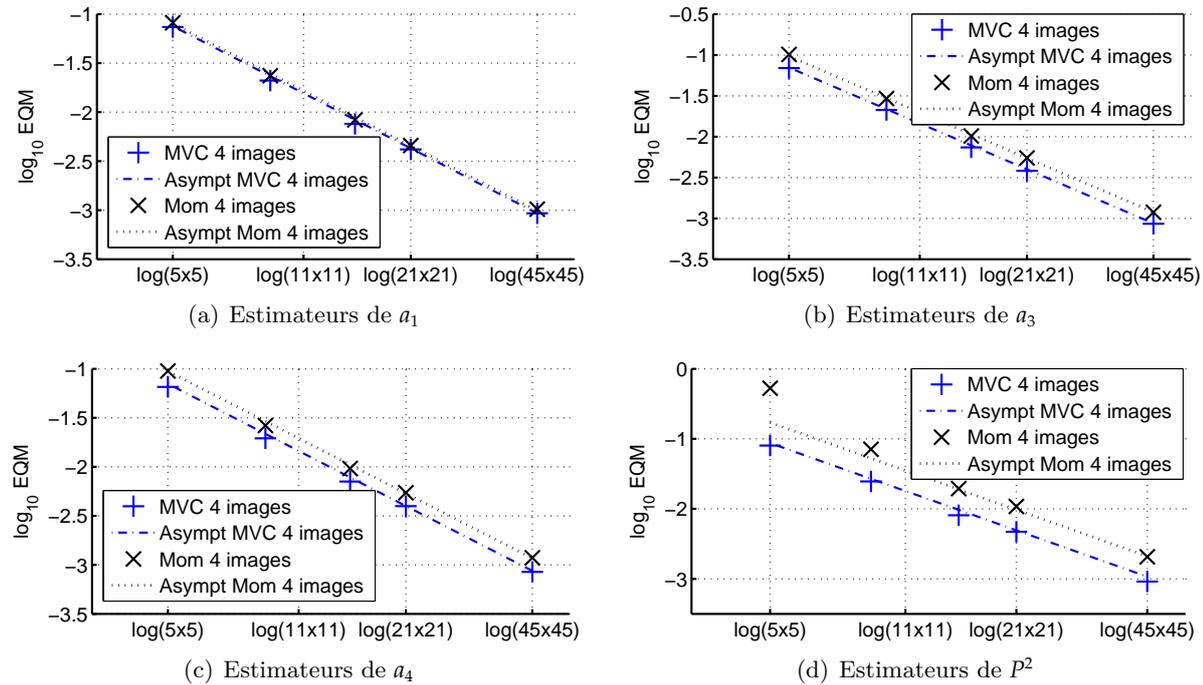


FIG. 4.11 – \log EQM des estimateurs à 4 images en fonction du logarithme de la taille des échantillons pour la matrice Γ_2 .

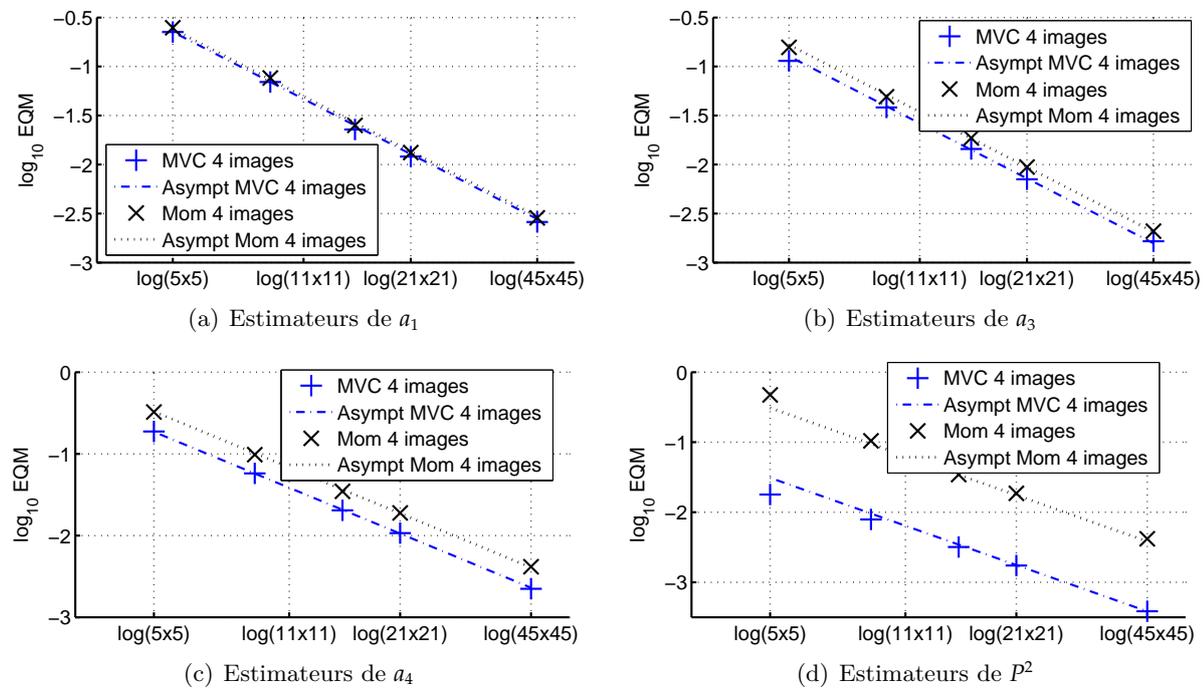


FIG. 4.12 – \log EQM des estimateurs à 4 images en fonction du logarithme de la taille des échantillons pour la matrice Γ_2 .

sont discrètes, et lorsque les niveaux des intensités photonisées sont faibles, on ne dispose que de peu d'information (beaucoup de pixels ont alors des valeurs similaires, par exemple 0, même si leurs moyennes théoriques sont différentes). Il devient alors plus difficile d'estimer les paramètres. On peut d'ailleurs remarquer que les performances pour la matrice Γ_4 ($P^2 = 0.5$) pour laquelle le niveau d'intensité $I_T = 99$ est élevé, sont très proches de celles obtenues en fort-flux. Cette remarque n'est pas vraie pour le modèle "fort-flux" où les v.a. observées sont continues, et où un changement d'échelle n'affecte pas les performances, comme le montre l'expression de la variance asymptotique du DdP (4.20) obtenue pour 4 images (qui ne dépend que de P^2), ou comme l'ont montré les résultats des simulations illustrés sur la figure 4.6.

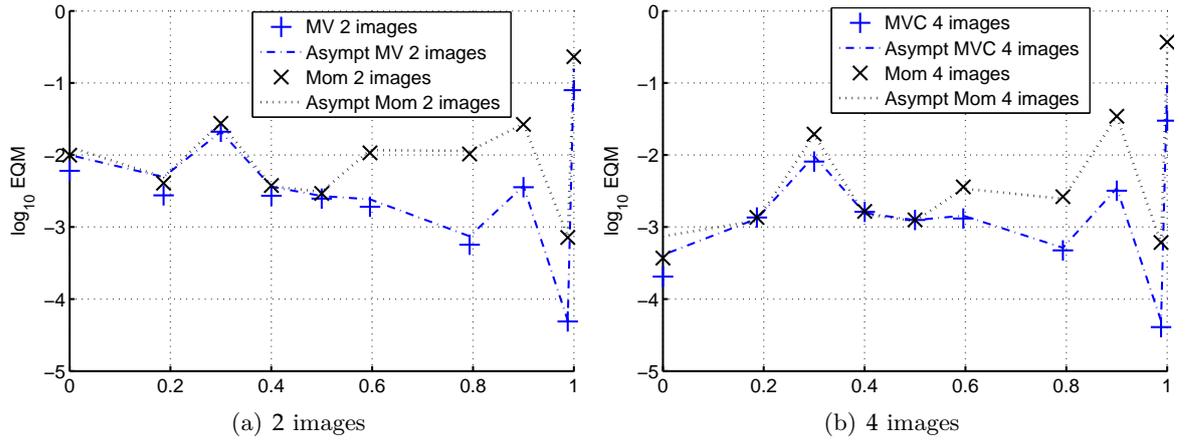


FIG. 4.13 – \log_{10} EQM des estimateurs du DdP à partir de 2 et 4 images en fonction de P^2 ($n = 15 \times 15$).

4.3.4 Estimation à partir d'images synthétiques

L'expérience d'imagerie simulée, décrite à la section 4.2.4, est ici effectuée en faible-flux. Les images d'intensité générées dans l'hypothèse de fort-flux sont maintenant corrompues par le bruit de Poisson. Notons que les valeurs des intensités photonisées ainsi obtenues sont en accord avec l'hypothèse de faible-flux. Les valeurs moyennes du nombre de photons observées, sont en effet relativement faibles : moins de 2 photons en moyenne sur les pixels appartenant au fond de la scène, et moins de 4 photons en moyenne sur les pixels les plus lumineux.

Les images des intensités photonisées sont représentées sur la figure 4.14. L'effet du bruit de Poisson est observable sur l'image des intensités totales $N_1 + N_2$, présentée sur la figure 4.14(c), où le contraste, entre le fond et les objets lumineux de l'image, est relativement faible. Les résultats obtenus pour les différents estimateurs sont présentés sur la figure 4.15. Toutes les figures sont représentées avec la même dynamique de contraste. Les DdP ont été estimés avec une fenêtre d'estimation carrée de taille $n = 15 \times 15$ (i.e. contenant 225 pixels). Enfin, les nombres insérés sur les images représentent les moyennes des estimés du DdP sur les objets correspondants.

Méthodes à 2 images : La méthode des moments (fig 4.15(a)) et la méthode du maximum de vraisemblance (fig 4.15(b)) donnent des résultats similaires sur le fond de l'image. L'objet 2 (le pseudo rectangle) est mieux restitué pour la méthode du maximum de vraisemblance. Les valeurs

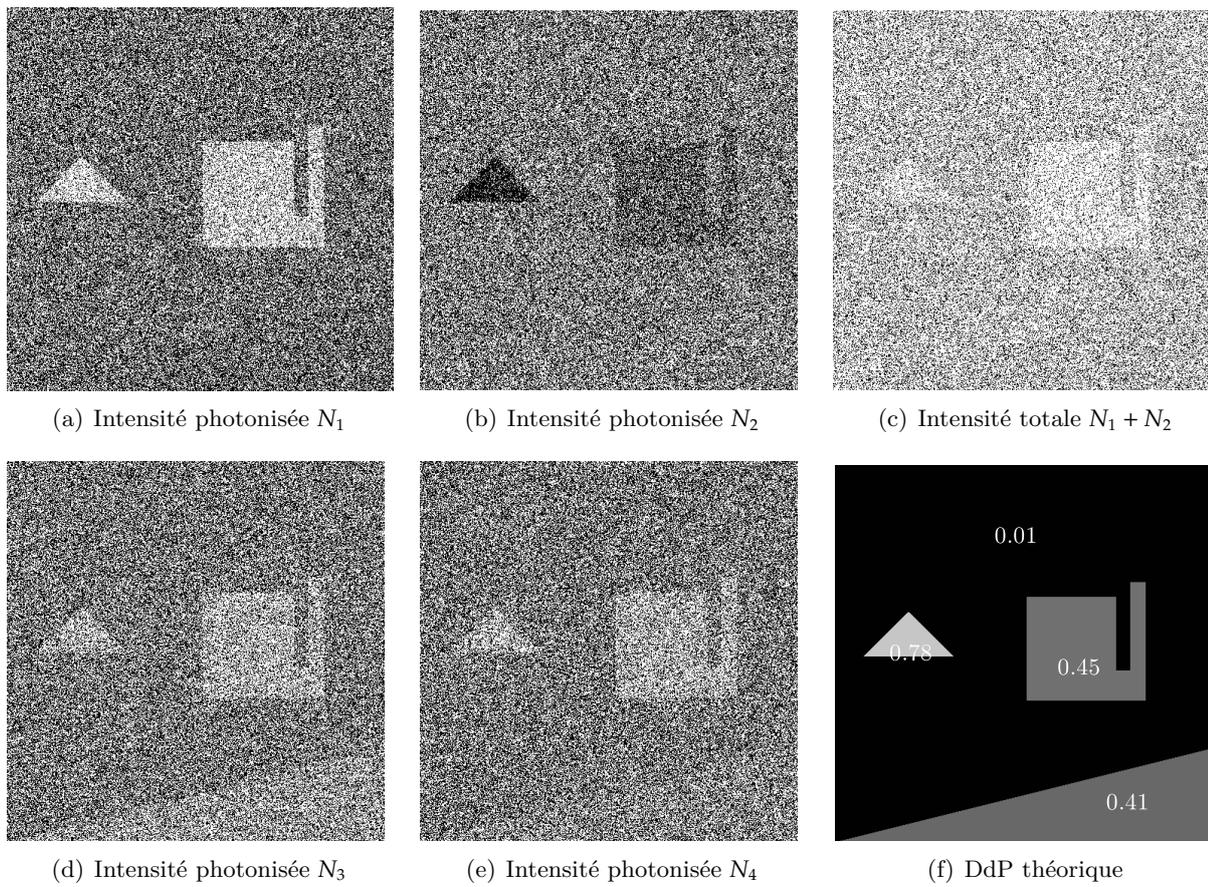


FIG. 4.14 – Intensités photonisées mesurées et DdP théorique.

estimées du DdP y sont plus uniformes pour l'EMV. De plus, on remarque que la méthode des moments présente des effets de bord sur chaque objet, plus importants que ceux obtenus pour l'EMV, et semble donc moins robuste aux inhomogénéités. Enfin, l'objet 3 (triangle inférieur) est un peu mieux estimé par la méthode MV : la variance des estimés est en effet plus faible et la frontière avec le fond un peu plus nette.

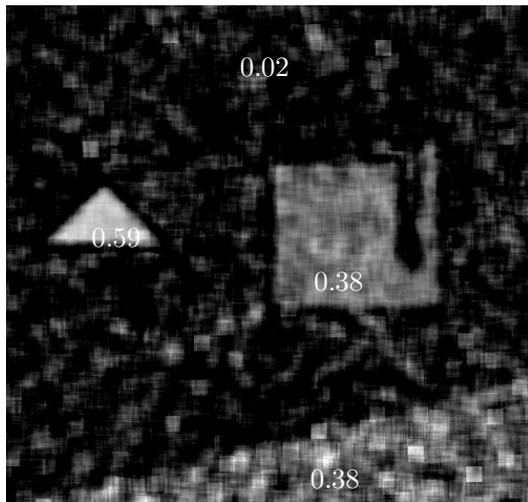
Méthodes à 4 images : Le fond de l'image est dans ce cas un peu mieux restitué pour la méthode du maximum de vraisemblance composite, fig. 4.15(d), que pour la méthode des moments, fig. 4.15(c). L'objet 3 (triangle inférieur) est bien mieux estimé pour la méthode du maximum de vraisemblance composite (variance beaucoup plus faible, contour plus net). Ceci confirme l'intérêt de l'estimateur du maximum de vraisemblance composite pour l'imagerie polarimétrique.

4.4 Conclusion

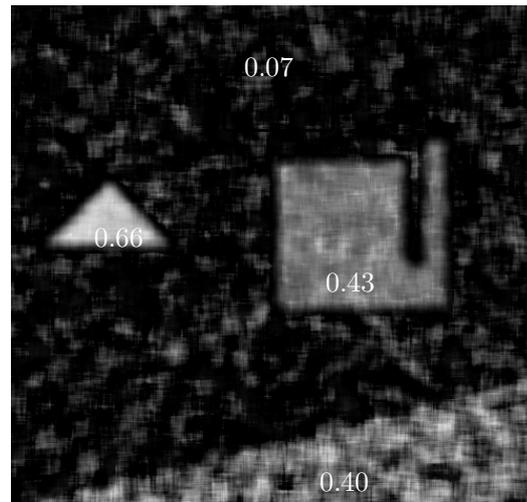
L'étude menée dans ce chapitre a permis de montrer que les intensités des images polarimétriques fort-flux sont distribuées suivant une loi Gamma multivariée. Ceci permet de démontrer que l'estimateur classique du DdP, obtenu pour 4 images, est asymptotiquement efficace, et une expression explicite de la borne de Cramer-Rao est disponible. Les résultats des simulations ont montré qu'il est possible d'estimer correctement le DdP en fort-flux, à partir d'un nombre réduit d'images, i.e. 3 ou même 2 images, à l'aide de méthodes de maximum de vraisemblance. Ces résultats sont intéressants pour des applications où la complexité du système imageur doit être réduite, ces méthodes offrant un bon compromis entre la précision des estimations et la complexité du système imageur.

Lorsque les niveaux d'intensité sont très faibles, et que l'hypothèse de faible-flux doit être prise en compte, la méthode d'estimation du maximum de vraisemblance composite s'avère performante lorsque 4 images polarimétriques sont disponibles. En raison de la dégradation des images par le bruit de Poisson, il est toutefois difficile d'estimer avec précision les propriétés polarimétriques de la scène à partir d'un nombre réduit d'images.

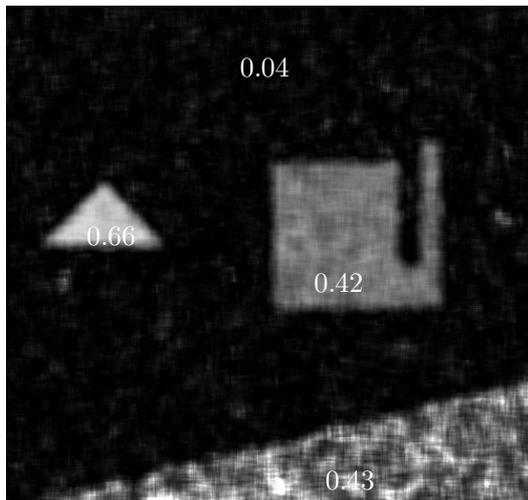
Pour finir, notons que les méthodes du maximum de vraisemblance, ou du maximum de vraisemblance composite, proposées, ne dispensent pas d'utiliser les méthodes des moments. Ces méthodes sont en effet complémentaires. Les méthodes des moments présentent l'avantage d'être quasi-instantanées, du point de vue calculatoire, du moins pour les cas à 2 ou 4 images, contrairement aux méthodes basées sur les vraisemblances, qui nécessitent d'optimiser numériquement un critère pour chaque pixel de l'image. Ces dernières méthodes peuvent donc être appliquées dans un deuxième temps, afin d'améliorer la précision des estimations obtenues sur les régions d'intérêt de l'image. Les résultats obtenus dans cette section devront toutefois être confirmés sur des images polarimétriques réelles.



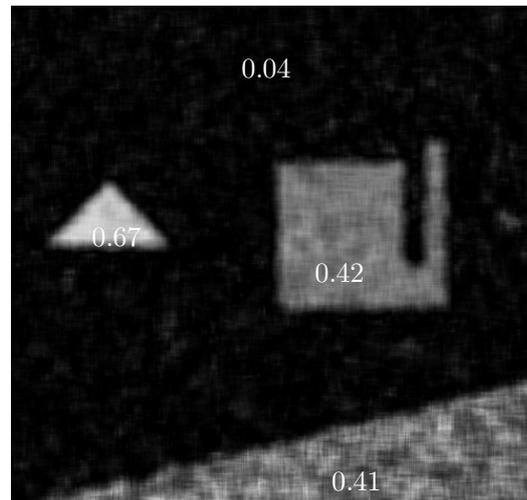
(a) Moments 2 images



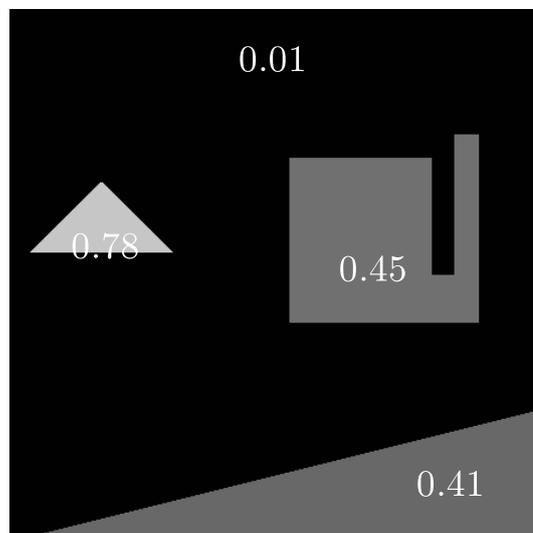
(b) MV 2 images



(c) Moments 4 images



(d) MVC 4 images



(e) DdP théorique

Conclusions

Cette thèse évalue l'apport de modèles statistiques déduits d'une famille de lois Gamma multivariées pour le traitement d'images. Développer des méthodes statistiques d'estimation performantes est un enjeu important en imagerie dans de nombreux problèmes de détection. La plupart des méthodes d'estimation existantes n'utilisent que les informations fournies par les premiers moments des statistiques observées. Ces méthodes, pourtant largement utilisées en raison de leur simplicité de mise en œuvre, et de leur coût calculatoire souvent très faible, demeurent en général sous-optimales. Il convient donc de bien évaluer l'opportunité d'utiliser ces familles de lois multivariées selon les applications étudiées.

Cette thèse étudie les perspectives offertes par le modèle des lois Gamma multivariées pour le traitement d'images issues de systèmes actifs et cohérents, tels que le radar. Sous les hypothèses dites de *speckle pleinement développé*, les statistiques des pixels des images étudiées correspondent à des lois Gamma. Les lois Gamma multivariées offrent la possibilité de poser un cadre statistique paramétrique simple et précis lorsque plusieurs images sont disponibles. Cette modélisation permet de développer des méthodes d'estimation reposant sur l'expression de vraisemblances associées à ces lois, méthodes qui sont en général performantes.

L'étude se décline en plusieurs points. Tout d'abord, nous nous sommes intéressés au problème du recalage d'image et de la détection de changements temporels entre une paire d'images RSO d'une même scène. Pour ce faire, il convient d'estimer localement des mesures de similarité entre les images. La méthode proposée consiste à modéliser les couples de pixels appartenant à chacune des deux images par une loi Gamma bivariée. Ce modèle permet de prendre en compte à la fois les contraintes strictes sur la forme des distributions des lois marginales, qui sont des lois Gamma, et la dépendance temporelle entre les images. L'estimation des paramètres des lois Gamma bivariées par la méthode du maximum de vraisemblance s'avère plus performante que les méthodes classiques d'estimation pour la détection de changements temporels, notamment sur des images réelles. Ce résultat justifie l'intérêt des lois Gamma bivariées pour la détection de changements temporels en imagerie RSO, et permet d'obtenir une méthode de référence lorsque les hypothèses de *speckle pleinement développé* sont valides.

Ce travail se prolonge lorsque les images RSO ont été acquises par des capteurs radar qui ont des caractéristiques différentes. Dans ce cas les distributions marginales des pixels de chaque image sont toujours des lois Gamma mais dont les paramètres de forme sont différents. Le modèle précédent des lois Gamma multivariées n'est alors plus approprié et doit être adapté. Nous avons alors proposé une nouvelle famille de distributions multivariées apte à modéliser la loi jointe des pixels de chaque image. L'estimation par la méthode du maximum de vraisemblance des paramètres de ces lois a été étudiée dans le cas bivarié. Les propriétés asymptotiques de ces estimateurs permettent d'analyser les performances des méthodes de détection de changements obtenues en fonction des valeurs des nombres de vues associées aux images RSO. Malheureu-

sement, pour ce modèle, le calcul des estimateurs du maximum de vraisemblance nécessite de maximiser numériquement la vraisemblance par rapport à deux variables. Afin de réduire le coût calculatoire des procédures d'estimation, une méthode reposant sur le principe des fonctions d'inférence par les marges a été mise en œuvre. Les estimateurs des mesures de similarité entre les images ainsi obtenus s'avèrent quasi-efficaces asymptotiquement, et performants pour la détection de changements temporels entre des couples d'images RSO, même si cette estimation se révèle moins robuste que pour les familles de lois Gamma bivariées précédemment étudiées. Ce modèle rend néanmoins possible l'estimation des nombres de vues équivalents de chaque image RSO, ce qui permet d'obtenir des stratégies d'estimation et de détection performantes.

Enfin, le dernier point abordé concerne l'imagerie polarimétrique active et cohérente. Ces méthodes d'imagerie analysent l'état de polarisation de la lumière rétrodiffusée par les objets imagés et apportent, dans certain cas, des informations bien plus exploitables que l'imagerie de réflectivité classique. L'estimation de l'état de polarisation de la lumière requiert d'effectuer quatre mesures d'intensité. Nous avons montré au cours de cette thèse que, sous les hypothèses de speckle pleinement développé, ces intensités suivent une loi Gamma multivariée. Ceci nous a permis de prouver que les estimateurs classiques de la matrice de polarisation sont efficaces et que l'estimateur correspondant du degré de polarisation est asymptotiquement efficace. Une expression explicite de la borne de Cramer-Rao pour ce dernier paramètre est alors disponible. En outre, ce modèle a conduit à de nouvelles méthodes d'estimation du degré de polarisation lorsque l'on ne dispose que d'un nombre réduit d'images polarimétriques. Ces méthodes fournissent finalement des performances significativement meilleures que des méthodes des moments. Nous avons mené la même étude en "faible-flux" où les niveaux d'intensité, très faibles, obligent à prendre en compte un bruit supplémentaire, le bruit de Poisson. Des méthodes originales d'estimation à partir de la log-vraisemblance des paires ont alors été développées. Ces méthodes s'avèrent particulièrement performantes dans le cas à quatre images, et permettent de montrer que l'estimateur classique des moments n'est pas efficace en "faible-flux".

Différentes perspectives se dégagent de ces travaux. Citons en quelques unes. Dans les chapitres consacrés à la détection de changements en imagerie RSO, l'étude effectuée se limite à la détection de changements abrupts entre une paire d'images d'une même scène. Un prolongement naturel de cette étude correspond à l'estimation de mesures de similarité entre les images lorsqu'une série d'images RSO multi-dates est disponible. Les changements à caractériser peuvent être abrupts, et dans ce cas deux piles d'images, avant et après, sont disponibles. Ces changements peuvent cependant être plus ténus et progressifs, lorsqu'on cherche par exemple à étudier l'évolution d'une scène. Ces différents cas de figure, correspondent alors à différentes paramétrisations d'une loi Gamma multivariée et à différentes stratégies d'estimation.

Par ailleurs, l'avènement de futurs capteurs radar soulève de nouvelles problématiques. En effet, les nouvelles générations de radar possèdent des résolutions submétriques. Ceci modifie profondément la forme du signal rétrodiffusé. En particulier, le nombre de diffuseurs au sein d'une cellule de résolution n'est plus assez important pour que les conditions d'application du théorème central limite soient vérifiées. Ceci amène à étudier de nouvelles familles de lois, qui ont des justifications physiques ou empiriques. Par exemple, en milieu urbain, la présence de nombreux réflecteurs entraîne l'utilisation de lois à queues lourdes telles que les lois de Fisher [Tup07]. La loi de Fisher ou plus généralement les lois de probabilités issues du système de Pearson ont montré de bonnes propriétés en imagerie RSO [QDM93]. Une perspective serait alors d'étendre l'étude réalisée pour le modèle des lois Gamma bivariées aux familles de lois du système de Pearson.

A

Transformées d'éléments aléatoires

La loi d'une variable aléatoire réelle ou plus généralement d'un vecteur aléatoire réel se caractérise par la fonction de répartition associée à cette variable ou ce vecteur. Il existe d'autres fonctions susceptibles de caractériser la loi d'une variable aléatoire. Dans le cas d'une variable discrète à valeur dans \mathbb{N} , on parlera de *fonction génératrice*. En général, on parlera de *fonction caractéristique* et de *transformée de Laplace*. Cette section présente quelques rappels sur ces transformées, pour une variable aléatoire, et leur extension au cas multivarié. Dans les travaux présentés par la suite, les lois étudiées seront essentiellement caractérisées par leur transformée de Laplace.

Dans ce chapitre, sont adoptées les notations vectorielles présentées ci-dessous :

- Si $\mathbf{n} = (n_1, \dots, n_d)^T$ désigne un vecteur de \mathbb{N}^d , alors on note la somme des composantes

$$|\mathbf{n}| = n_1 + \dots + n_d.$$

- La factorielle du vecteur $\mathbf{n} \in \mathbb{N}^d$ est le produit des factorielles des composantes du vecteur

$$\mathbf{n}! = n_1! \cdots n_d!.$$

- Si, de plus, $\mathbf{z} = (z_1, \dots, z_d)^T$ est un vecteur de \mathbb{R}^d , alors la puissance de ce vecteur est le produit des puissances des composantes du vecteur

$$\mathbf{z}^{\mathbf{n}} = z_1^{n_1} \cdots z_d^{n_d}.$$

- Enfin, la dérivée partielle d'ordre \mathbf{n} d'une fonction $f : \mathbf{z} \in \mathbb{R}^d \mapsto f(\mathbf{z}) \in \mathbb{R}$ au point $\mathbf{z}_0 \in \mathbb{R}^d$ est notée et définie par

$$f^{(\mathbf{n})}(\mathbf{z}_0) = \left(\frac{\partial^{n_1 + \dots + n_d}}{\partial z_1^{n_1} \cdots \partial z_d^{n_d}} f(\mathbf{z}) \right) (\mathbf{z}_0).$$

A.1 Fonction génératrice

Les principaux résultats concernant la fonction génératrice d'une variable aléatoire sont généralisés au cas de vecteurs aléatoires à valeur dans \mathbb{N}^d . Considérons donc un vecteur aléatoire $\mathbf{N} = (N_1, \dots, N_d)^T$ à valeurs discrètes dans \mathbb{N}^d .

Définition A.1.1. La fonction génératrice du vecteur aléatoire \mathbf{N} à valeurs dans \mathbb{N}^d est définie par la série entière en $\mathbf{z} \in \mathbb{R}^d$ autour du vecteur nul noté $\mathbf{0} = (0, \dots, 0)^T \in \mathbb{R}^d$:

$$G_{\mathbf{N}}(\mathbf{z}) = \sum_{\mathbf{n} \in \mathbb{N}^d} \Pr(\mathbf{N} = \mathbf{n}) \mathbf{z}^{\mathbf{n}}.$$

De manière similaire au cas univarié, la série multivariée $G_{\mathbf{N}}$ est dite absolument convergente au point \mathbf{z} si et seulement si

$$\sum_{\mathbf{n} \in \mathbb{N}^d} |\Pr(\mathbf{N} = \mathbf{n}) \mathbf{z}^{\mathbf{n}}| < \infty.$$

Les coefficients de la série entière $G_{\mathbf{N}}(\mathbf{z})$ étant tous positifs en tant que masses de probabilité, cette série est donc absolument convergente au point $\mathbf{z} = (z_1, \dots, z_d)^T$ si elle est convergente au point $\mathbf{x} = (|z_1|, \dots, |z_d|)^T \in \mathbb{R}_+^d$. L'ensemble des points pour lesquels la série multivariée est absolument convergente est alors

$$D_{\mathbf{N}} = \left\{ \mathbf{z} \in \mathbb{R}^d \mid \text{si } \mathbf{x} = (|z_1|, \dots, |z_d|)^T \text{ alors } G_{\mathbf{N}}(\mathbf{x}) < \infty \right\}. \quad (\text{A.1})$$

La fonction génératrice $G_{\mathbf{N}}$ étant en outre une série entière au voisinage de 0 par rapport à chaque variable z_i , pour $i = 1, \dots, d$, si la série multivariée $G_{\mathbf{N}}$ est absolument convergente pour un vecteur $\mathbf{x} \in \mathbb{R}^d$, elle l'est aussi pour tout vecteur $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ tel que $|y_i| \leq |x_i|$ pour tout $1 \leq i \leq d$.

Soit $\|\cdot\|_{\infty}$ la norme infinie telle que $\|\mathbf{z}\|_{\infty} = \max_{i \in \{1, \dots, d\}} |z_i|$. Le rayon de convergence pour la norme infinie de la série multivariée est alors défini¹⁷ par

$$R_{\mathbf{N}}^{\infty} = \sup \left\{ r > 0 \mid \mathbf{r} = (r, \dots, r)^T \in D_{\mathbf{N}} \right\}, \quad (\text{A.2})$$

et la série multivariée $G_{\mathbf{N}}$ est alors absolument convergente sur le disque ouvert pour la norme infinie

$$D_{\mathbf{N}}^{\infty} = \left\{ \mathbf{z} \in \mathbb{R}^d \mid \|\mathbf{z}\|_{\infty} < R_{\mathbf{N}}^{\infty} \right\}. \quad (\text{A.3})$$

Le rayon de convergence vérifie donc $R_{\mathbf{N}}^{\infty} \geq 1$, car la série $G_{\mathbf{N}}$ est absolument convergente au point $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^d$:

$$G_{\mathbf{N}}(\mathbf{1}) = \sum_{\mathbf{n} \in \mathbb{N}^d} \Pr(\mathbf{N} = \mathbf{n}) = 1 < \infty.$$

De plus, la fonction $\mathbf{z} \mapsto G_{\mathbf{N}}(\mathbf{z})$ est absolument convergente et analytique réelle sur le disque ouvert non vide $D_{\mathbf{N}}^{\infty}$.

À partir des définitions précédentes, on remarque :

¹⁷La définition donnée ici du rayon de convergence d'une série multivariée est assez grossière : si l'on considère en effet la fonction f de \mathbb{R}^2 dans \mathbb{R} telle que $f(x, y) = (1 - (x + y))^{-1}$, on trouve alors que le rayon de convergence du développement en série entière de f autour du point $(0, 0)$ est $R^{\infty} = 1/2$, la fonction f admet donc un développement en série entière sur le disque $D^{\infty} = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\|_{\infty} < 1/2\}$. Cependant on sait que cette fonction admet un développement en série entière sur l'ensemble $D = \{(x, y) \in \mathbb{R}^2 \mid |x + y| < 1\}$ qui inclut strictement l'ensemble D^{∞} . Cette notion s'avère néanmoins suffisante pour étudier les fonctions génératrices.

1. Dans le cas d'une variable aléatoire, i.e. $d = 1$, la définition A.1.1 se ramène à la définition classique de la fonction génératrice d'une variable aléatoire entière N :

$$G_N(z) = \sum_{n \geq 0} \Pr(N = n) z^n.$$

De plus, la définition du rayon de convergence R_N^∞ donnée à l'équation (A.2) correspond alors au rayon de convergence de la série entière univariée, définie ci-dessus, pour $z \in \mathbb{R}$.

2. De manière plus synthétique, la fonction génératrice du vecteur \mathbf{N} se réécrit sous la forme :

$$\forall \|\mathbf{z}\|_\infty \leq 1, \quad G_{\mathbf{N}}(\mathbf{z}) = \mathbf{E} \left[\mathbf{z}^{\mathbf{N}} \right]. \quad (\text{A.4})$$

La fonction génératrice $G_{\mathbf{N}}$ caractérise bien la loi du vecteur \mathbf{N} :

Proposition A.1.1. *La fonction génératrice G étant analytique réelle sur le disque ouvert $D_{\mathbf{N}}^\infty$ qui contient $\mathbf{0}$, toutes les dérivées partielles de G en $\mathbf{0}$ existent et d'autre part :*

$$\forall \mathbf{n} \in \mathbb{N}^d, \quad \Pr(\mathbf{N} = \mathbf{n}) = \frac{1}{\mathbf{n}!} G_{\mathbf{N}}^{(\mathbf{n})}(\mathbf{0}).$$

On appelle moment d'ordre $\mathbf{r} = (r_1, \dots, r_d)^T \in \mathbb{N}^d$ du vecteur aléatoire \mathbf{N} , la quantité $\mathbf{E}[\mathbf{N}^{\mathbf{r}}] = \mathbf{E}[N_1^{r_1} \dots N_d^{r_d}]$, et le moment factoriel d'ordre \mathbf{r} , la grandeur $\mathbf{E}[\mathbf{N}^{[\mathbf{r}]}] = \mathbf{E}[N_1^{[r_1]} \dots N_d^{[r_d]}]$ où $N_i^{[r_i]} = N_i(N_i - 1) \dots (N_i - r_i + 1)$ pour tout $1 \leq i \leq d$.

Proposition A.1.2. *La vecteur aléatoire \mathbf{N} admet un moment d'ordre $\mathbf{r} = (r_1, \dots, r_d)^T \in \mathbb{N}^d$ si et seulement si la fonction $G_{\mathbf{N}}$ admet une dérivée partielle d'ordre \mathbf{r} finie à gauche au point $\mathbf{z} = \mathbf{1}$. Dans ce cas, l'expression du moment factoriel d'ordre \mathbf{r} du vecteur \mathbf{N} est :*

$$\mathbf{E}[\mathbf{N}^{[\mathbf{r}]}] = G_{\mathbf{N}}^{(\mathbf{r})}(\mathbf{1}).$$

En particulier lorsque $R_{\mathbf{N}}^\infty > 1$, tous les moments du vecteur \mathbf{N} existent et peuvent être calculés à partir de la proposition précédente. L'expression des moments joints classiques $\mathbf{E}[\prod_{i=1}^d N_i^{r_i}]$ est obtenue à partir de celle des moments factoriels, en décomposant chaque polynôme $N_i^{r_i}$ de la variable N_i dans la base des polynômes à degrés étagés $(N_i^{[r]})_{0 \leq r \leq r_i}$.

Enfin, comme dans le cas d'une variable aléatoire, on retrouve le fait que la fonction génératrice d'une somme est le produit des fonctions génératrices.

Proposition A.1.3. *Soient \mathbf{M} et \mathbf{N} deux vecteurs aléatoires indépendants de \mathbb{N}^d de fonctions génératrices respectives $G_{\mathbf{M}}$ et $G_{\mathbf{N}}$. Alors, la fonction génératrice de la somme $\mathbf{M} + \mathbf{N}$, notée $G_{\mathbf{M}+\mathbf{N}}$ est :*

$$\forall \|\mathbf{z}\|_\infty < R_{\mathbf{M}+\mathbf{N}}^\infty, \quad G_{\mathbf{M}+\mathbf{N}}(\mathbf{z}) = G_{\mathbf{M}}(\mathbf{z})G_{\mathbf{N}}(\mathbf{z}),$$

avec $R_{\mathbf{M}+\mathbf{N}}^\infty = \min(R_{\mathbf{M}}^\infty, R_{\mathbf{N}}^\infty)$. De plus, la réciproque est vraie : si la fonction génératrice de la somme $\mathbf{M} + \mathbf{N}$ se décompose comme le produit ci-dessus, alors les vecteurs \mathbf{M} et \mathbf{N} sont indépendants.

Exemple A.1.1 : la loi Binomiale négative $\mathcal{BN}(q, p)$ de paramètres $q > 0$ et $0 < p < 1$ ¹⁸. On note $\mathcal{BN}(q, p)$ la loi de la variable entière N suivante :

$$\Pr(N = n) = \frac{\Gamma(q + n)}{\Gamma(q)} p^q \frac{(1 - p)^n}{n!} \quad \forall n \in \mathbb{N}.$$

Alors la fonction génératrice de la variable N est :

$$\begin{aligned} G_N(z) &= p^q \sum_{n \geq 0} \frac{\Gamma(q + n)}{\Gamma(q)} \frac{(z(1 - p))^n}{n!}, \\ &= \left(\frac{p}{1 - (1 - p)z} \right)^q, \end{aligned} \tag{A.5}$$

et le rayon de convergence de la série entière ainsi obtenue est $R_N^\infty = \frac{1}{1-p} > 1$. Tous les moments de N sont donc définis selon A.1.2, de plus pour tout $|1 + z| < \frac{1}{1-p} \iff |z| < \frac{p}{1-p}$:

$$\begin{aligned} f_N(z) &= G_N(1 + z) = \left(1 - \frac{1 - p}{p} z \right)^q, \\ &= \sum_{n \geq 0} \frac{\Gamma(q + n)}{\Gamma(q)} \left(\frac{1 - p}{p} \right)^n \frac{z^n}{n!} \end{aligned}$$

Comme $G_N^{(r)}(1) = f_N^{(r)}(0)$, on obtient selon la proposition A.1.2 l'expression suivante des moments factoriels de N :

$$\mathbb{E}[N(N - 1) \dots (N - r + 1)] = \frac{\Gamma(q + r)}{\Gamma(q)} \left(\frac{1 - p}{p} \right)^r \quad \forall r \in \mathbb{N}. \tag{A.6}$$

A.2 Transformée de Laplace

Pour une variable aléatoire réelle quelconque, la notion de transformée de Laplace remplace celle de fonction génératrice introduite pour les variables discrètes. Cette transformation a été introduite par le marquis Pierre Simon de Laplace en 1812, dans son ouvrage "Théorie analytique des probabilités" [Lap20], afin de caractériser diverses lois de probabilité. Les principaux résultats concernant la transformée de Laplace sont présentés pour un vecteur aléatoire à valeur dans $E = \mathbb{R}^d$. Ces résultats sont néanmoins généralisables à n'importe quel élément aléatoire à valeur dans un espace euclidien E .

On définit ci-dessous la transformée de Laplace complexe, mais elle est le plus souvent utilisée dans le cas d'un vecteur \mathbf{z} réel dans \mathbb{R}^d . On munit pour cela l'espace vectoriel \mathbb{R}^d du produit scalaire euclidien canonique :

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^d x_i y_i.$$

¹⁸Cette loi est traditionnellement définie pour un paramètre q entier. La loi négative binomiale modélise alors la loi du nombre d'échecs précédant le q -ième succès lors de la répétition d'épreuves identiques et indépendantes, de probabilité de succès p . Cependant, cette loi se généralise de manière naturelle selon la définition donnée ici, lorsque le paramètre q est un réel strictement positif, même si l'interprétation probabiliste est alors moins évidente.

Par ailleurs, si $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ est un vecteur complexe dans \mathbb{C}^d , \mathbf{x} et \mathbf{y} étant des vecteurs de \mathbb{R}^d , et si $\boldsymbol{\theta}$ est un vecteur de \mathbb{R}^d , on note $\langle \mathbf{z}, \boldsymbol{\theta} \rangle$ la grandeur complexe suivante :

$$\langle \mathbf{z}, \boldsymbol{\theta} \rangle = \langle \mathbf{x}, \boldsymbol{\theta} \rangle + i\langle \mathbf{y}, \boldsymbol{\theta} \rangle.$$

Définition A.2.1. Soit \mathbf{X} un vecteur aléatoire à valeur dans \mathbb{R}^d . La transformée de Laplace de \mathbf{X} , notée $\mathcal{L}_{\mathbf{X}}$, est définie pour un vecteur complexe $\mathbf{z} \in \mathbb{C}^d$, par :

$$\mathcal{L}_{\varepsilon, \mathbf{X}}(\mathbf{z}) = \mathbb{E} \left[e^{\varepsilon \langle \mathbf{z}, \mathbf{X} \rangle} \right],$$

où $\varepsilon = \pm 1$ suivant la convention choisie, et $\Omega_{\mathbf{X}}$ l'ensemble des vecteurs réels $\mathbf{x} \in \mathbb{R}^d$ tels que $\mathcal{L}_{\mathbf{X}}(\mathbf{x})$ existe.

Afin d'alléger les notations, la transformée de Laplace sera notée par la suite $\mathcal{L}_{\mathbf{X}}(\mathbf{z})$, la convention choisie étant précisée. À partir de la définition précédente de la transformée de Laplace, il est immédiat que $\mathbf{0}$ est dans $\Omega_{\mathbf{X}}$. De plus la convexité sur \mathbb{R}^d de la fonction $\mathbf{z} \mapsto e^{\varepsilon \langle \mathbf{z}, \mathbf{X} \rangle}$ assure que l'ensemble $\Omega_{\mathbf{X}}$ est convexe. Enfin, la transformée de Laplace complexe $\mathcal{L}_{\mathbf{X}}(\mathbf{z})$ existe pour tout vecteur $\mathbf{z} \in \mathbb{C}^d$, tel que le vecteur des parties réelles de chaque composante de \mathbf{z} , noté $\Re(\mathbf{z})$, soit dans le convexe $\Omega_{\mathbf{X}}$.

Remarques :

1. Il n'existe pas de convention arrêtée quant au signe de l'argument de l'exponentielle dans la définition de la transformée de Laplace. La transformée de Laplace peut ainsi être définie comme $\mathbb{E} \left[e^{+\langle \mathbf{z}, \mathbf{X} \rangle} \right]$ ou comme $\mathbb{E} \left[e^{-\langle \mathbf{z}, \mathbf{X} \rangle} \right]$ selon les besoins d'un auteur, les deux expressions se déduisant l'une de l'autre en remplaçant \mathbf{z} par $-\mathbf{z}$. Dans ce manuscrit, les deux conventions seront utilisées suivant les cas. La convention $\varepsilon = -1$ est en effet préférable pour l'étude des variables réelles positives, comme cela sera souligné juste après. Cependant la convention $\varepsilon = +1$ est mieux adaptée au formalisme des familles exponentielles, introduites à l'annexe suivante.
2. Lorsque la dimension est $d = 1$, on retrouve alors la définition classique de la transformée de Laplace d'une variable aléatoire réelle X :

$$\mathcal{L}_X(z) = \mathbb{E} \left[e^{\varepsilon z X} \right].$$

Dans ce cas, le domaine réel de définition Ω_X de la transformée de Laplace se réduit à un intervalle. Par conséquent, la transformée de Laplace est définie sur une bande parallèle à l'axe imaginaire dans le plan complexe.

3. On parle de la transformée de Laplace du vecteur \mathbf{X} mais il s'agit plutôt de la transformée de Laplace de la loi de \mathbf{X} notée $\mu_{\mathbf{X}}$:

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \mathbb{E} \left[e^{\varepsilon \langle \mathbf{z}, \mathbf{X} \rangle} \right] = \int_{\mathbb{R}^d} e^{\varepsilon \langle \mathbf{z}, \mathbf{x} \rangle} d\mu_{\mathbf{X}}(\mathbf{x}) = \mathcal{L}_{\mu_{\mathbf{X}}}(\mathbf{z}).$$

De plus si $\mu_{\mathbf{X}}$ possède une densité $f_{\mathbf{X}}$ par rapport à la mesure de Lebesgue, la transformée de Laplace de \mathbf{X} correspond alors à la transformée de Laplace bilatérale de la fonction de plusieurs variables $f_{\mathbf{X}}$:

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \mathbb{E} \left[e^{\varepsilon \langle \mathbf{z}, \mathbf{X} \rangle} \right] = \int_{\mathbb{R}^d} e^{\varepsilon \langle \mathbf{z}, \mathbf{x} \rangle} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathcal{L}_{f_{\mathbf{X}}}(\mathbf{z}).$$

La transformée de Laplace \mathcal{L}_X caractérise bien la loi du vecteur X .

Proposition A.2.1. *Si le convexe Ω_X admet un intérieur non vide noté Θ_X , alors la loi du vecteur X est caractérisée par sa transformée de Laplace et cette dernière est holomorphe sur l'ouvert $\{z \in \mathbb{C}^d \mid \Re(z) \in \Theta_X\}$, où $\Re(z)$ est la partie réelle du vecteur complexe z . Plus précisément, soit Y un vecteur aléatoire dans \mathbb{R}^d dont la transformée de Laplace \mathcal{L}_Y est définie sur Ω_Y . Si le convexe $\Omega_X \cap \Omega_Y$ admet un intérieur non vide et si $\mathcal{L}_X = \mathcal{L}_Y$ sur cet ensemble, alors X et Y sont de même loi.*

Les moments du vecteur X peuvent être directement obtenus à partir de sa transformée de Laplace \mathcal{L}_X .

Proposition A.2.2. *Si $\mathbf{0}$ est dans l'intérieur noté Θ_X du convexe Ω_X , alors la transformée de Laplace \mathcal{L}_X du vecteur $X = (X_1, \dots, X_d)^T$ est holomorphe sur l'ouvert $\{z \in \mathbb{C}^d \mid \Re(z) \in \Theta_X\}$ et admet le développement en série entière au voisinage de $\mathbf{0}$ suivant :*

$$\mathcal{L}_X(z) = \sum_{n \in \mathbb{N}^d} \varepsilon^{|n|} \frac{\mathbb{E}[X^n]}{n!} z^n,$$

où $z = (z_1, \dots, z_d)^T$. Par conséquent, pour tout $n \in \mathbb{N}^d$ on a :

$$\mathbb{E}[X^n] = \varepsilon^{|n|} \mathcal{L}_X^{(n)}(\mathbf{0}).$$

En raison de cette propriété, on appelle parfois la transformée de Laplace la *fonction génératrice des moments*. Cette appellation peut porter à confusion avec la notion de fonction génératrice d'une variable aléatoire à valeur entière, introduite au paragraphe précédent. Ces fonctions sont en effet distinctes, mêmes si elles sont reliées entre elles, comme le montre la proposition A.2.6 ci-après.

Exemple A.2.1 : *La loi Gamma $\mathcal{G}(q, p)$ de paramètres $(q, p) \in]0, \infty[^2$.*

On note $\mathcal{G}(q, p)$ la loi de mesure de probabilité sur \mathbb{R} suivante :

$$\mathcal{G}(q, p)(x) = \frac{e^{-\frac{x}{p}}}{\Gamma(q)} \left(\frac{x}{p}\right)^{q-1} \mathbb{I}_{[0, +\infty[}(x) \frac{dx}{p},$$

où $q > 0$ est appelé le paramètre de forme, et $p > 0$ le paramètre d'échelle. Avec la convention $\varepsilon = -1$ dans la transformée de Laplace, si X suit une loi $\mathcal{G}(q, p)$ alors sa transformée de Laplace est définie sur l'intervalle ouvert $\Omega_X = \Theta_X =]-1/p, +\infty[$. Finalement, on a pour tout $z \in]-1/p, +\infty[$:

$$\mathcal{L}_X(z) = \int_{\mathbb{R}^+} \frac{e^{-\frac{1+pz}{p}x}}{\Gamma(q)} \left(\frac{x}{p}\right)^{q-1} \frac{dx}{p},$$

En effectuant le changement de variable $u = \frac{1+pz}{p}x$, et grâce à la définition de la fonction gamma $\Gamma(q) = \int_{\mathbb{R}^+} e^{-u} u^{q-1} du$, on obtient :

$$\mathcal{L}_X(z) = (1 + pz)^{-q}. \tag{A.7}$$

Si l'on avait pris la convention $\varepsilon = +1$ dans la transformée de Laplace, on aurait obtenu de manière symétrique $\Omega_X = \Theta_X =]-\infty, 1/p[$ et sur cet intervalle :

$$\mathcal{L}_X(z) = (1 - pz)^{-q}. \quad (\text{A.8})$$

En développant maintenant en série entière la transformée de Laplace ainsi obtenue, on trouve grâce à la proposition A.2.2, l'expression suivante des moments de la loi $\mathcal{G}(q, p)$:

$$\mathbb{E}[X^n] = \frac{\Gamma(q+n)}{\Gamma(q)} p^n \quad \forall n \in \mathbb{N}. \quad (\text{A.9})$$

La transformée de Laplace d'une somme de vecteurs aléatoires indépendants est le produit des transformées de Laplace des divers vecteurs.

Proposition A.2.3. *Si \mathbf{Y} est un vecteur aléatoire indépendant de \mathbf{X} de transformée de Laplace \mathcal{L}_Y , alors la transformée de Laplace de la somme $\mathbf{X} + \mathbf{Y}$, notée $\mathcal{L}_{\mathbf{X}+\mathbf{Y}}$, est définie sur $\Omega_{\mathbf{X}+\mathbf{Y}} = \Omega_X \cap \Omega_Y$ et a pour expression :*

$$\forall \mathbf{z} \in \Omega_{\mathbf{X}+\mathbf{Y}}, \quad \mathcal{L}_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) = \mathcal{L}_X(\mathbf{z})\mathcal{L}_Y(\mathbf{z}).$$

La réciproque est vraie : si la transformée de Laplace de la somme $\mathbf{X} + \mathbf{Y}$ se décompose comme le produit ci-dessus, alors les vecteurs \mathbf{X} et \mathbf{Y} sont indépendants.

Les transformées de Laplace d'un vecteur et de son image, par une transformation affine, se déduisent l'une de l'autre.

Proposition A.2.4. *Si \mathbf{A} est une matrice inversible de $\mathbb{R}^{d \times d}$ et \mathbf{b} un vecteur \mathbb{R}^d alors $\Omega_{\mathbf{A}\mathbf{X}+\mathbf{b}} = \mathbf{A}^{-1}\Omega_X$, où $\mathbf{A}^{-1}\Omega_X$ désigne l'ensemble $\{\mathbf{A}^{-1}\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Omega_X\}$, et sur ce convexe :*

$$\mathcal{L}_{\mathbf{A}\mathbf{X}+\mathbf{b}}(\mathbf{z}) = \exp(\varepsilon(\mathbf{b}, \mathbf{z}))\mathcal{L}_X(\mathbf{A}\mathbf{z}).$$

Une autre propriété importante est la log-convexité de la transformée de Laplace.

Proposition A.2.5. *La transformée de Laplace est une fonction log-convexe sur Ω_X . En effet, si $\mathbf{z}_1, \mathbf{z}_2 \in \Omega_X$ et $\lambda \in [0, 1]$, alors :*

$$\log \mathcal{L}_X(\lambda \mathbf{z}_1 + (1 - \lambda)\mathbf{z}_2) \leq \lambda \log \mathcal{L}_X(\mathbf{z}_1) + (1 - \lambda) \log \mathcal{L}_X(\mathbf{z}_2).$$

De plus, si l'intérieur de Ω_X est non vide et si la loi de \mathbf{X} n'est pas dégénérée sur un hyperplan affine de \mathbb{R}^d , alors la log-transformée de Laplace est une fonction analytique réelle strictement convexe sur l'intérieur de Ω_X .

Démonstration. On peut supposer que $0 < \lambda < 1$, les cas $\lambda \in \{0, 1\}$ étant évidents. On utilise alors l'inégalité de Hölder : soient r et s deux réels positifs tels que $r > 1$ et $1/r + 1/s = 1$. Soient η et ξ deux variables aléatoires réelles telles que $\mathbb{E}[|\eta|^r]$ et $\mathbb{E}[|\xi|^s]$ existent, alors $\mathbb{E}[|\eta\xi|] < \mathbb{E}[|\eta|^r]^{1/r} \mathbb{E}[|\xi|^s]^{1/s}$; l'inégalité étant stricte si $|\eta|^r$ et $|\xi|^s$ ne sont pas proportionnelles presque

sûrement.

Comme

$$\mathcal{L}_X(\lambda z_1 + (1 - \lambda)z_2) = \mathbb{E} \left[\left(e^{\varepsilon \langle z_1, X \rangle} \right)^\lambda \left(e^{\varepsilon \langle z_2, X \rangle} \right)^{1-\lambda} \right],$$

En posant $r = 1/\lambda$, $\eta = \left(e^{\varepsilon \langle z_1, X \rangle} \right)^\lambda$ et $\xi = \left(e^{\varepsilon \langle z_2, X \rangle} \right)^{1-\lambda}$ il vient :

$$\mathcal{L}_X(\lambda z_1 + (1 - \lambda)z_2) \leq \mathbb{E} \left[e^{\varepsilon \langle z_1, X \rangle} \right]^\lambda \mathbb{E} \left[e^{\varepsilon \langle z_2, X \rangle} \right]^{1-\lambda} = \mathcal{L}_X(z_1)^\lambda \mathcal{L}_X(z_2)^{(1-\lambda)},$$

et l'on obtient le résultat souhaité en passant au logarithme. Si par ailleurs $z_1 \neq z_2$ et que la loi de \mathbf{X} n'est pas dégénérée sur un hyperplan affine, alors les variables $|\eta|^r$ et $|\xi|^s$ ne sont pas proportionnelles, sinon il existerait $\alpha \in \mathbb{R}$, tel que $\langle z_1 - z_2, \mathbf{X} \rangle = \alpha$ presque sûrement, et la loi de \mathbf{X} serait donc concentrée sur un hyperplan affine. L'inégalité est donc stricte dans ce cas. Si, de plus, l'intérieur de Ω_X est non vide, alors la transformée de Laplace est analytique sur cet ouvert selon A.2.1 et, par composition de fonctions analytiques réelles, le logarithme de la transformée de Laplace est elle aussi analytique réelle sur cet ouvert. \square

La fonction $z \in \mathbb{R}^d \mapsto \log \mathcal{L}_X(z)$ est parfois appelée la *fonction génératrice des cumulants* car, par définition, les cumulants sont les dérivées partielles de cette fonction (au signe près lorsque $\varepsilon = -1$).

Enfin, la transformée de Laplace d'un vecteur est reliée aux autres transformées de vecteurs aléatoires.

Proposition A.2.6. – La fonction génératrice d'un vecteur discret \mathbf{N} à valeur dans \mathbb{N}^d , définie en A.1.1, est liée à la transformée de Laplace de ce vecteur. En effet, avec la convention $\varepsilon = -1$ de la transformée de Laplace, on obtient :

$$\forall z > -\log R_N^\infty, \quad G_N(\exp(-z)) = \mathcal{L}_N(z),$$

où R_N^∞ est le rayon de convergence de la série G_N définie en (A.2). Symétriquement, avec la convention $\varepsilon = +1$, il vient :

$$\forall z < \log R_N^\infty, \quad G_N(\exp(z)) = \mathcal{L}_N(z).$$

– La fonction caractéristique du vecteur \mathbf{X} est une fonction à valeurs complexes, définie par $\phi_X(\mathbf{t}) = \mathbb{E} \left[e^{\varepsilon i \langle \mathbf{t}, X \rangle} \right]$ pour tout $\mathbf{t} \in \mathbb{R}^d$, avec $\varepsilon = \pm 1$. Par conséquent, avec la même convention sur le signe de ε dans la transformée de Laplace, on obtient :

$$\forall \mathbf{t} \in \mathbb{R}^d, \quad \phi_X(\mathbf{t}) = \mathcal{L}_X(i\mathbf{t}).$$

L'intérêt de la transformée de Laplace par rapport à la fonction caractéristique est d'être à valeurs réelles (pour un argument dont les composantes sont réelles) et donc plus facile à manipuler sur un plan pratique que la fonction caractéristique. La transformée de Laplace est, par exemple, préférable à la fonction caractéristique lorsqu'on travaille avec des vecteurs dont les composantes sont des variables aléatoires réelles positives. En effet, ces vecteurs sont stables par addition, et avec la convention $\varepsilon = -1$ dans la définition A.2.1, il est immédiat que si $\mathbf{X} \in \mathbb{R}^d$ est un tel vecteur alors $[0, \infty[^d \subset \Omega_X$ (symétriquement $] -\infty, 0]^d \subset \Omega_X$ avec la convention $\varepsilon = +1$).

Exemple A.2.2 : La loi Binomiale Négative $\mathcal{BN}(q, p)$ de paramètres $q \in]0, \infty[$ et $0 < p < 1$. La loi $\mathcal{BN}(q, p)$ introduite à l'exemple A.1.1 a pour mesure de probabilité sur \mathbb{R} :

$$\mathcal{BN}(q, p)(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)} p^q \frac{(1-p)^k}{k!} \delta_k(dx).$$

Cette loi est donc concentrée sur l'ensemble \mathbb{N} des valeurs entières et, selon l'exemple A.1.1, la fonction génératrice d'une variable N distribuée suivant cette loi est :

$$G_N(z) = \left(\frac{p}{1 - (1-p)z} \right)^q, \quad \forall |z| < R_N^\infty = \frac{1}{1-p}.$$

La proposition A.2.6 assure alors que la transformée de Laplace de N est (avec la convention $\varepsilon = -1$) :

$$\mathcal{L}_N(z) = \left(\frac{p}{1 - (1-p)e^{-z}} \right)^q, \quad \forall z > \log(1-p). \quad (\text{A.10})$$

De plus, dans ce cas particulier de dimension $d = 1$, le domaine défini ci-dessus correspond à l'ensemble $\Theta_N =]\log(1-p), +\infty[$.

De manière symétrique, on obtient avec la convention $\varepsilon = +1$:

$$\mathcal{L}_N(z) = \left(\frac{p}{1 - (1-p)e^z} \right)^q, \quad \forall z \in \Theta_N =]-\infty, -\log(1-p)[. \quad (\text{A.11})$$

Si le vecteur aléatoire réel \mathbf{X} admet une densité $f_{\mathbf{X}}$ continue par morceaux sur \mathbb{R}^d , alors cette densité peut être obtenue à partir de la transformée de Laplace $\mathcal{L}_{\mathbf{X}}$. Il suffit d'inverser la transformée de Laplace $\mathcal{L}_{\mathbf{X}}(\mathbf{z})$ par rapport à chaque composante de \mathbf{z} selon la formule présentée dans la proposition suivante :

Proposition A.2.7. (formule d'inversion de Bromvitch) : soit $F(z) = F(x + iy)$ une fonction analytique pour x appartenant à un intervalle ouvert non vide I et sommable en y pour tout $x \in I$. Alors F est une transformée de Laplace, dont l'original f est donné par :

$$\frac{f(t+0) + f(t-0)}{2} = \frac{1}{2i\pi} \int_{x-i\infty}^{x+i\infty} F(s) e^{-\varepsilon st} ds,$$

où $x \in I$.

Cette dernière intégrale se calcule souvent en utilisant le théorème des résidus. En pratique cette formule analytique est très difficilement utilisable. Aussi un moyen pour inverser une transformée de Laplace est de procéder par identification en se ramenant à des transformées de Laplace connues et tabulées dans des dictionnaires des transformées de Laplace usuelles.

Les propriétés des transformées de Laplace associées aux vecteurs aléatoires sont donc similaires à celles des variables aléatoires réelles. Le cas des matrices aléatoires réelles se traite encore une fois de façon similaire, en munissant l'espace vectoriel des matrices réelles de dimension $m \times n$, noté $\mathbb{R}^{m \times n}$, du produit scalaire suivant :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times n}, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \text{tr}(\mathbf{x}\mathbf{y}^T),$$

où $\text{tr}(\mathbf{a})$ est la trace de la matrice carrée \mathbf{a} . Si $m = 1$ ou $n = 1$, on retrouve bien la définition A.2.1 de la transformée de Laplace de vecteurs aléatoires réels. Si l'on se restreint à l'espace vectoriel des matrices réelles symétriques, ce produit scalaire devient finalement : $\langle \mathbf{x}, \mathbf{y} \rangle = \text{tr}(\mathbf{x}\mathbf{y})$. Les propriétés énoncées dans le cadre des vecteurs aléatoires sont conservées.

B

Familles Exponentielles Naturelles

Les familles exponentielles sont des ensembles de lois de probabilité très utilisées et qui ont reçu beaucoup d'attention pour deux raisons principales :

1. beaucoup de modèles statistiques sont des familles exponentielles.
2. les méthodes classiques d'estimation et de détection s'appliquent avec succès lorsque le modèle statistique est associé à une famille exponentielle.

Dans ce chapitre, nous nous concentrons sur le cas des familles exponentielles dites naturelles, les familles exponentielles généralisées, comme leur nom l'indique, en sont une extension. Les propriétés de ces dernières reposent alors sur celles des familles exponentielles naturelles qui leur sont associées.

Le formalisme et les résultats fondamentaux propres aux familles exponentielles naturelles, introduits notamment dans les deux ouvrages de référence de Barndorff-Nielsen [BN78] et Letac [Let92], ou encore exposés dans [Let] sont rappelés dans cette partie. Notons qu'un chapitre entier de [KBJ00, chap. 54] est aussi consacré aux familles exponentielles sur \mathbb{R}^d .

B.1 Définition

On se place ici dans \mathbb{R}^d ou plus généralement dans un espace vectoriel euclidien noté E , i.e. un espace vectoriel réel de dimension finie d muni d'un produit scalaire. Pour tout couple (\mathbf{x}, \mathbf{y}) de $E \times E$, on note $\langle \mathbf{x}, \mathbf{y} \rangle$ le produit scalaire canonique de l'espace euclidien E .

Dans cette section sur les familles exponentielles, les transformées de Laplace introduites, dans le cas de mesures de probabilités, à la définition A.2.1, seront toujours considérées avec la convention $\varepsilon = +1$. Soit Θ_μ l'intérieur de l'ensemble convexe Ω_μ des vecteurs $\boldsymbol{\theta}$, tels que la transformée de Laplace de la mesure μ , notée $\mathcal{L}_\mu(\boldsymbol{\theta})$ (voir la définition A.2.1 dans le cas d'une mesure de probabilité) :

$$\mathcal{L}_\mu(\boldsymbol{\theta}) = \int_E e^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle} \mu(d\mathbf{x}),$$

soit finie. De plus, Ω_μ étant convexe (voir la section A.2 dédiée à la transformée de Laplace de lois multivariées), son intérieur Θ_μ l'est aussi.

Soit $\mathcal{M}(E)$ l'ensemble des mesures de Radon μ positives sur E (i.e. les mesures positives

bornées sur tout compact de E , en particulier les mesures de probabilité) qui possèdent en outre les deux propriétés suivantes :

1. μ n'est pas concentrée sur un hyperplan affine de E , i.e. si $\mathbf{a} \in E \setminus \{\mathbf{0}\}$ et $b \in \mathbb{R}$, μ n'est pas concentrée sur un espace du type $\{\mathbf{x} \in E \mid \langle \mathbf{a}, \mathbf{x} \rangle = b\}$.
2. Θ_μ est non vide. Cette condition permet d'éviter par exemple le cas d'une distribution de Cauchy dont la mesure sur \mathbb{R} est $\mu(dx) = \frac{dx}{\pi(1+x^2)}$, et n'admet une transformée de Laplace finie que pour le singleton $\Omega_\mu = \{0\}$.

La proposition A.2.5, démontrée pour une mesure de probabilité, s'étend alors sans difficulté à toute mesure μ appartenant à $\mathcal{M}(E)$.

Proposition B.1.1. *La log-transformée de Laplace $\mathbf{x} \mapsto k_\mu(\mathbf{x}) = \log \mathcal{L}_\mu(\mathbf{x})$, encore appelée fonction génératrice des cumulants de μ , est une fonction analytique réelle strictement convexe sur Θ_μ .*

Une conséquence directe de cette proposition est que le gradient de k_μ :

$$\begin{aligned} \Theta_\mu &\rightarrow E \\ \mathbf{x} &\mapsto k'_\mu(\mathbf{x}), \end{aligned}$$

est une fonction injective, car strictement croissante.

Si $\boldsymbol{\theta} \in \Theta_\mu$, on note :

$$\begin{aligned} P_{\boldsymbol{\theta}, \mu}(d\mathbf{x}) &= \frac{e^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle}}{\mathcal{L}_\mu(\boldsymbol{\theta})} \mu(d\mathbf{x}), \\ &= e^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle - k_\mu(\boldsymbol{\theta})} \mu(d\mathbf{x}). \end{aligned} \tag{B.1}$$

Par construction, $P_{\boldsymbol{\theta}, \mu}$ est une mesure de probabilité sur E , i.e. une mesure positive telle que $P_{\boldsymbol{\theta}, \mu}(E) = 1$.

Définition B.1.1. *On appelle famille exponentielle naturelle (FEN) engendrée par μ , le modèle statistique défini par l'ensemble des probabilités sur E suivant :*

$$F(\mu) = \{P_{\boldsymbol{\theta}, \mu} \mid \boldsymbol{\theta} \in \Theta_\mu\}.$$

Une conséquence directe de cette définition est que si μ et μ' sont deux mesures de $\mathcal{M}(E)$, alors les ensembles $F(\mu)$ et $F(\mu')$ coïncident si et seulement s'il existe un couple $(\mathbf{a}, b) \in E \times \mathbb{R}$ tel que :

$$\mu'(d\mathbf{x}) = e^{\langle \mathbf{a}, \mathbf{x} \rangle + b} \mu(d\mathbf{x}).$$

Etant donnée une FEN F , l'ensemble $\mathcal{B}_F = \{\mu \in \mathcal{M}(E) \mid F = F(\mu)\}$ désigne l'ensemble des bases (ou générateurs) possibles de F . Selon la définition d'une FEN et la remarque précédente, il est clair que $F \subset \mathcal{B}_F$. Cependant, il peut exister des bases qui sont des mesures de probabilité mais qui ne sont pas dans F , par exemple lorsque le vecteur nul $\boldsymbol{\theta} = \mathbf{0}$ est dans l'ensemble Ω_μ privé de son intérieur Θ_μ , ensemble noté $\Omega_\mu \setminus \Theta_\mu$.

Une FEN peut être caractérisée grâce à la transformée de Laplace de ses éléments.

Proposition B.1.2. *Soit μ une mesure de $\mathcal{M}(E)$. Alors la FEN générée par μ est l'ensemble*

$$F(\mu) = \left\{ \nu \in \mathcal{M}(E) \mid \exists \boldsymbol{\theta}_0 \in \Theta_\mu \text{ tel que } \mathcal{L}_\nu(\mathbf{z}) = \frac{\mathcal{L}_\mu(\boldsymbol{\theta}_0 + \mathbf{z})}{\mathcal{L}_\mu(\boldsymbol{\theta}_0)} \quad \forall \mathbf{z} \in \Theta_\mu - \boldsymbol{\theta}_0 \right\},$$

où $\Theta_\mu - \boldsymbol{\theta}_0 = \{\boldsymbol{\theta} - \boldsymbol{\theta}_0 \mid \boldsymbol{\theta} \in \Theta_\mu\}$.

Cette proposition découle du fait que toute mesure de probabilité $P_{\theta,\mu}$ définie à l'équation (B.1) vérifie $\mathcal{L}_{P_{\theta,\mu}}(z) = \frac{\mathcal{L}_{\mu}(\theta+z)}{\mathcal{L}_{\mu}(\theta)}$, et de l'injectivité de la transformée de Laplace sur l'ensemble des mesures de probabilité de $\mathcal{M}(E)$ (voir proposition A.2.1). Cette proposition est valable quelle que soit la convention $\varepsilon = \pm 1$ choisie dans la transformée de Laplace. Ce résultat est utile pour savoir si une famille de distributions dont on connaît les transformées de Laplace, mais pas forcément l'expression des mesures de probabilité associées, est une FEN. Le point délicat dans une telle démarche est alors de trouver l'expression du paramètre naturel θ_0 de la FEN.

Exemple B.1.1 : *la famille exponentielle naturelle $\{\mathcal{G}(q, p); p \in]0, +\infty[$ des loi Gamma.*

Soit un paramètre réel $q > 0$, ici fixé. On considère la mesure sur \mathbb{R} suivante (cette mesure est obtenue à partir de la mesure de probabilité d'une loi Gamma $\mathcal{G}(q, p)$ en ne retenant que les termes qui ne dépendent pas du paramètre d'échelle p) :

$$\mu_q(dx) = \frac{x^{q-1}}{\Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx.$$

On trouve alors $\Theta_{\mu_q} =]-\infty, 0[$, et sur cet ensemble :

$$\mathcal{L}_{\mu_q}(\theta) = \frac{1}{(-\theta)^q} \iff k_{\mu_q}(\theta) = -q \log(-\theta).$$

Par conséquent, la FEN engendrée par μ_q est l'ensemble des lois de probabilité suivantes :

$$P_{\theta, \mu_q} = (-\theta)^q e^{\theta x} \frac{x^{q-1}}{\Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx \quad \forall \theta \in]-\infty, 0[.$$

On reconnaît la loi Gamma de paramètre de forme q et de paramètre d'échelle $p = \frac{-1}{\theta}$ notée $\mathcal{G}(q, p)$:

$$\mathcal{G}(q, p)(dx) = e^{-\frac{x}{p}} \frac{x^{q-1}}{p^q \Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx.$$

La FEN, $F = F(\mu_q)$, est la famille des lois gamma $\mathcal{G}(q, p)$, de paramètre de forme $q > 0$ fixé et de paramètre d'échelle p , à valeur dans $]0, +\infty[$. Cette FEN peut être caractérisée par la transformée de Laplace de ses éléments, qui est selon la proposition B.1.2 de la forme :

$$\forall z \in]-\infty, -\theta[\quad \mathcal{L}_{P_{\theta, \mu_q}}(z) = \left(\frac{-\theta}{-\theta - z} \right)^q = \left(1 + \frac{z}{\theta} \right)^{-q} = (1 - pz)^{-q},$$

pour $\theta = \frac{-1}{p} \in]-\infty, 0[$. On retrouve bien la formule (A.8) obtenue dans l'exemple A.2.1.

Exemple B.1.2 : *La famille exponentielle naturelle $\{\mathcal{BN}(q, p); 0 < p < 1\}$ des lois binomiales négatives.*

Le paramètre réel $q > 0$ est ici fixé. On considère la mesure génératrice suivante (cette mesure est obtenue à partir de la mesure de probabilité d'une loi binomiale négative $\mathcal{BN}(q, p)$ en ne retenant que les termes qui ne dépendent pas du paramètre de probabilité p) :

$$\mu_q(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} \delta_k(dx),$$

(cette mesure n'est pas une probabilité car $\mu(\mathbb{R}) = +\infty$, mais elle est bien dans $\mathcal{M}(\mathbb{R})$). Alors la transformée de Laplace de cette mesure est :

$$\mathcal{L}_{\mu_q}(\theta) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)} \frac{(e^\theta)^k}{k!} = (1 - e^\theta)^{-q},$$

et est définie sur $\Theta_{\mu_q} =]-\infty, 0[$. Sur cet ensemble, on a :

$$k_{\mu_q}(\theta) = -q \log(1 - e^\theta).$$

Par conséquent la FEN engendrée par μ_q est l'ensemble des lois de probabilité suivantes :

$$P_{\theta, \mu_q}(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} e^{\theta k} (1 - e^\theta)^q \delta_k(dx) \quad \forall \theta \in]-\infty, 0[.$$

En posant $p = 1 - e^\theta$, on reconnaît la loi négative binomiale $\mathcal{BN}(q, p)$, introduite notamment aux exemples A.1.1 et A.2.2. La FEN $F = F(\mu_q)$ est donc la famille des lois binomiales négatives $\mathcal{BN}(q, p)$, de paramètre de forme $q > 0$ fixé et de paramètre d'échelle p , à valeur dans $]0, 1[$. Cette FEN peut être caractérisée par la transformée de Laplace de ses éléments, qui est selon la proposition B.1.2 de la forme :

$$\forall z \in]-\infty, -\theta[\quad \mathcal{L}_{P_{\theta, \mu_q}}(z) = \left(\frac{1 + e^\theta}{1 + e^\theta e^z} \right)^q = \left(\frac{p}{1 - (1-p)e^z} \right)^{-q},$$

pour $\theta = \log(1 - p) \in]-\infty, 0[$. On retrouve bien la formule (A.11).

B.2 Moments d'une FEN

On suppose dans ce paragraphe que l'espace euclidien E est \mathbb{R}^d . Soit μ une mesure de $\mathcal{M}(\mathbb{R}^d)$. Tous les moments de la loi $P_{\theta, \mu}$ pour $\theta \in \Theta_\mu$ existent, de plus :

1. les moments de la loi $P_{\theta, \mu}$ se déduisent de la transformée de Laplace \mathcal{L}_μ de la mesure génératrice μ :

$$\forall \mathbf{n} \in \mathbb{N}^d, \quad \int_{\mathbb{R}^d} \mathbf{x}^{\mathbf{n}} P_{\theta, \mu}(d\mathbf{x}) = \frac{\mathcal{L}_\mu^{(\mathbf{n})}(\theta)}{\mathcal{L}_\mu(\theta)}.$$

2. le vecteur $\mathbf{m} \in \mathbb{R}^d$ des moyennes est le gradient de la fonction k_μ au point θ :

$$\mathbf{m} = \int_{\mathbb{R}^d} \mathbf{x} P_{\theta, \mu}(d\mathbf{x}) = k'_\mu(\theta).$$

3. la matrice de covariance associée à la loi $P_{\theta, \mu}$ est la matrice Hessienne de la fonction k_μ au point θ :

$$\forall 1 \leq i, j \leq d, \quad \int_{\mathbb{R}^d} (x_i - m_i)(x_j - m_j) P_{\theta, \mu}(d\mathbf{x}) = \frac{\partial^2 k_\mu}{\partial x_i \partial x_j}(\theta).$$

La propriété 1 découle de la proposition A.2.2, qui relie les moments de loi de probabilité $P_{\theta, \mu}$ à sa transformée de Laplace $\mathcal{L}_{P_{\theta, \mu}}$ (avec la convention $\varepsilon = +1$), et du fait que $\mathcal{L}_{P_{\theta, \mu}}(z) = \frac{\mathcal{L}_\mu(\theta+z)}{\mathcal{L}_\mu(\theta)}$. Les propriétés 2 et 3 se déduisent directement de la première propriété.

B.3 Autres paramétrisations d'une FEN

On a vu à la section précédente que le vecteur gradient $k'_\mu(\boldsymbol{\theta})$ est le vecteur moyenne de la loi $P_{\boldsymbol{\theta},\mu}$ ¹⁹. D'après la proposition B.1.1, la fonction k'_μ est injective sur Θ_μ . L'ensemble image de k'_μ , i.e. $k'_\mu(\Theta_\mu)$, s'appelle le *domaine des moyennes de F*, on le note M_F . Naturellement, l'ensemble M_F ne dépend que de F et pas de la base μ (plusieurs bases engendrent F). On note alors $\psi_\mu : M_\mu \rightarrow \Theta_\mu$ l'application réciproque de k'_μ telle que :

$$\begin{aligned} k'_\mu(\psi_\mu(\mathbf{m})) &= \mathbf{m} & \forall \mathbf{m} \in M_\mu, \\ \psi_\mu(k'_\mu(\boldsymbol{\theta})) &= \boldsymbol{\theta} & \forall \boldsymbol{\theta} \in \Theta_\mu. \end{aligned} \tag{B.2}$$

La fonction k'_μ étant analytique sur l'ouvert Θ_μ et la hessienne k''_μ étant définie positive (car la fonction k_μ est strictement convexe) donc non singulière sur Θ_μ , selon le théorème d'inversion locale, la fonction ψ_μ est analytique sur M_F . Par conséquent l'ensemble M_F est ouvert comme image réciproque d'un ouvert par une application continue.

Pour $\mathbf{m} \in M_F$, on considère la mesure de probabilité $P(\psi(\mathbf{m}), \mu)$: c'est l'élément de F de moyenne \mathbf{m} . L'injectivité de k'_μ assure qu'il est unique. Comme cet élément ne dépend pas de la mesure μ mais seulement de l'ensemble F (pour les mêmes raisons que l'ensemble M_F), on le note $P(\mathbf{m}, F)$. Par conséquent, l'application :

$$\begin{aligned} M_F &\rightarrow F \\ \mathbf{m} &\mapsto P(\mathbf{m}, F) \end{aligned}$$

est une bijection. On l'appelle la *paramétrisation de F par la moyenne*. Si μ est une base de F , i.e. $\mu \in \mathcal{B}_F$, alors l'expression des éléments de F paramétrés par leur moyenne, est :

$$P(\mathbf{m}, F) = e^{\langle \psi(\mathbf{m}), \mathbf{x} \rangle - k_\mu(\psi(\mathbf{m}))} \mu(d\mathbf{x}). \tag{B.3}$$

Les avantages de cette paramétrisation par rapport à la paramétrisation naturelle de la FEN sont les suivants :

- la paramétrisation par la moyenne est intrinsèque, elle ne dépend pas d'une base particulière.
- le paramètre \mathbf{m} a une interprétation statistique immédiate en tant que moyenne de la loi.

Un inconvénient rencontré est que l'expression de la mesure de probabilité $P(\mathbf{m}, F)$ est souvent plus compliquée.

En définitive, il existe souvent trois paramétrisations possibles d'une FEN :

1. la paramétrisation naturelle, introduite à l'équation (B.1).
2. la paramétrisation de F par la moyenne, introduite à l'équation (B.3).
3. lorsqu'il s'agit de FEN très classiques, elles ont en général, pour des raisons historiques, une paramétrisation classique qui peut n'être ni la paramétrisation naturelle, ni la paramétrisation par la moyenne. Dans ce cas, cette paramétrisation est souvent préférable.

¹⁹Cette propriété justifie le fait que l'on adopte la convention $\varepsilon = +1$ dans la transformée de Laplace (voir définition A.2.1) pour définir les FEN. En effet si l'on avait choisi la convention $\varepsilon = -1$, on obtiendrait que le vecteur $k'_\mu(\boldsymbol{\theta})$ est le vecteur moyenne à un facteur $(-1)^d$ près, selon la proposition A.2.2. Ce signe rendrait alors le formalisme associé aux FEN moins élégant.

Exemple B.3.1 : la famille exponentielle naturelle $\{\mathcal{G}(q, p); p \in]0, +\infty[$ des loi Gamma.

On a vu dans l'exemple B.1.1 qu'une base de cette FEN est la mesure μ_q , pour laquelle $\Theta_{\mu_q} =]-\infty, 0[$ et $k_{\mu_q}(\theta) = -q \log(-\theta)$ sur Θ_{μ} . Par conséquent $k'_{\mu_q}(\theta) = \frac{-q}{\theta} = m$. Cette fonction est croissante sur $]-\infty, 0[$, sa limite en $-\infty$ est 0 et sa limite à gauche en 0 est $+\infty$. Donc $M_F =]0, +\infty[$ et on trouve :

$$\psi_{\mu_q}(m) = \frac{-q}{m} \quad \forall m \in M_F. \quad (\text{B.4})$$

Finalement, on obtient les trois paramétrisations suivantes :

1. la paramétrisation naturelle :

$$\forall \theta \in]-\infty, 0[, \quad P_{\theta, \mu_q}(dx) = (-\theta)^q e^{\theta x} \frac{x^{q-1}}{\Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx,$$

2. la paramétrisation par la moyenne :

$$\forall m \in]0, +\infty[, \quad P_{m, F}(dx) = \left(\frac{q}{m}\right)^q e^{-\frac{q}{m}x} \frac{x^{q-1}}{\Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx,$$

3. la paramétrisation classique :

$$\forall p \in]0, +\infty[, \quad \mathcal{G}(q, p)(dx) = e^{-\frac{x}{p}} \frac{x^{q-1}}{p^q \Gamma(q)} \mathbb{I}_{]0, +\infty[}(x) dx.$$

Exemple B.3.2 : la famille exponentielle naturelle $\{\mathcal{BN}(q, p); 0 < p < 1\}$ des lois binomiales négatives.

On a vu dans l'exemple B.1.2 qu'une base de cette FEN est la mesure μ_q , pour laquelle $\Theta_{\mu_q} =]-\infty, 0[$ et $k_{\mu_q}(\theta) = -q \log(1 - e^\theta)$ sur Θ_{μ} . Par conséquent :

$$k'_{\mu_q}(\theta) = \frac{qe^\theta}{(1 - e^\theta)},$$

La fonction k'_{μ_q} est strictement croissante sur $\Theta_{\mu} =]-\infty, 0[$, sa limite en $-\infty$ est 0 et sa limite à gauche en 0 est $+\infty$, donc $M_F =]0, +\infty[$. En inversant la fonction k'_{μ_q} , on trouve :

$$\psi_{\mu_q}(m) = \log\left(\frac{m}{m+q}\right) \quad \forall m \in M_F.$$

Finalement, on obtient les trois paramétrisations suivantes :

1. la paramétrisation naturelle :

$$\forall \theta \in]-\infty, 0[, \quad P_{\theta, \mu_q}(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} e^{\theta k} (1 - e^\theta)^q \delta_k(dx),$$

2. la paramétrisation par la moyenne :

$$\forall m \in]0, +\infty[, \quad P_{m, F}(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} \left(\frac{m}{m+q}\right)^k \left(1 - \frac{m}{m+q}\right)^q \delta_k(dx),$$

3. la paramétrisation classique :

$$\forall p \in]0, 1[, \quad \mathcal{BN}(q, p)(dx) = \sum_{k \geq 0} \frac{\Gamma(q+k)}{\Gamma(q)k!} (1-p)^k p^q \delta_k(dx).$$

B.4 Fonction Variance d'une FEN

Tout comme $k''_{\mu}(\boldsymbol{\theta})$ est l'opérateur de covariance de $P_{\boldsymbol{\theta},\mu}(\mathbf{dx})$, on note $V_F(\mathbf{m})$ l'opérateur de covariance de $P_{\mathbf{m},F}(\mathbf{dx})$.

Soit une mesure μ dans $\mathcal{M}(\mathbb{R}^d)$. La fonction variance pour une FEN notée $F = F_{\mu}$ est définie comme :

$$\begin{aligned} V_F : M_F &\rightarrow S_+ \\ \mathbf{m} &\mapsto V_F(\mathbf{m}) = \left(\int_{\mathbb{R}^d} (x_i - m_i)(x_j - m_j) P_{\mathbf{m},F}(\mathbf{dx}) \right)_{1 \leq i, j \leq d}, \end{aligned}$$

où S_+ désigne l'ensemble des matrices $d \times d$ symétriques définies positives. La variance $V_F(\mathbf{m})$ se représente donc par une matrice symétrique définie positive, dont les entrées dépendent du vecteur des moyennes $\mathbf{m} = (m_1, \dots, m_d)^T$. A partir de la définition de la fonction ψ_F , donnée par l'équation (B.2), et de l'expression de la covariance d'une FEN, donnée à la section B.2, on obtient :

$$V_F(\mathbf{m}) = k''_{\mu}(\psi_{\mu}(\mathbf{m})) = (\psi'_{\mu}(\mathbf{m}))^{-1}. \quad (\text{B.5})$$

Cette dernière expression de la fonction variance s'étend dans le cas général où E est un espace euclidien et $\mu \in \mathcal{M}(E)$.

La fonction variance joue un rôle important dans la théorie des familles exponentielles, depuis l'article fondamental de C.-N. Morris [Mor82], où il est montré qu'elle caractérise la FEN : si F_1 et F_2 sont deux FEN de domaine des moyennes noté respectivement, M_{F_1} et M_{F_2} , et si les fonctions variance coïncident sur un ensemble non vide de $M_{F_1} \cap M_{F_2}$, alors les deux FEN sont égales : $F_1 = F_2$. La fonction variance présente généralement une forme plus simple que les mesures génératrices de la FEN, ou que leurs transformées de Laplace. Cette observation a conduit à chercher à décrire les FEN ayant une fonction variance de forme donnée. Ainsi il existe la classe de Morris des FEN sur $E = \mathbb{R}$, de fonction variance quadratique, i.e. dont la restriction de la fonction variance à M_F est un polynôme du paramètre de la moyenne, de degré inférieur ou égal à 2, qui compte six types bien connus (voir aussi [LM86]) : normal, Poisson, binomiale, binomiale négative, gamma et cosinus hyperbolique.

Exemple B.4.1 : la famille exponentielle naturelle $F = \{\mathcal{G}(q, p); p \in]0, +\infty[\}$ des lois Gamma. On a vu, à travers les exemples B.1.1 et B.3.1, que la FEN sur \mathbb{R} des lois gamma est générée par la mesure μ_q , dont la transformée de Laplace est définie sur $\Theta_{\mu} =]-\infty, 0[$, et telle que $\psi_{\mu_q}(m) = \frac{-q}{m}$ pour tout $m \in M_F =]0, +\infty[$. En dérivant la fonction ψ_{μ_q} on trouve donc, selon la formule (B.5), que la fonction variance de la FEN est :

$$V_F(m) = \frac{m^2}{q} \quad \forall m \in M_F =]0, +\infty[.$$

On obtient bien un polynôme en m de degré 2, ce qui explique pourquoi la loi Gamma de paramètre de forme q fixé est dans la classe de Morris, présentée ci-dessus.

Exemple B.4.2 : la famille exponentielle naturelle $F = \{\mathcal{BN}(q, p); p \in]0, +\infty[\}$ des lois binomiales négatives.

On a vu à travers les exemples B.1.2 et B.3.2 que la FEN sur \mathbb{R} des lois gamma est générée par la mesure μ_q , dont la transformée de Laplace est définie sur $\Theta_{\mu} =]-\infty, 0[$, et telle que

$\psi_{\mu_q}(m) = \log(m/(m+q))$ pour tout $m \in M_F =]0, +\infty[$. En dérivant la fonction ψ_{μ_q} on trouve donc selon la formule (B.5) que la fonction variance de la FEN est :

$$V_F(m) = \frac{m(m+q)}{q} \quad \forall m \in M_F =]0, +\infty[.$$

Comme pour les lois Gamma, on obtient un polynôme en m de degré 2, ce qui explique pourquoi la loi binomiale négative de paramètre de forme q fixé est dans le classe de Morris, présentée ci-dessus.

B.5 Transformation affine et convolution de FEN

Les transformations affines sont les seules transformations qui conservent la structure de FEN.

Proposition B.5.1. *Soit F une FEN de E et $\mathbf{x} \mapsto \phi(\mathbf{x}) = \mathbf{a}(\mathbf{x}) + \mathbf{b}$ une transformation affine de E , i.e. \mathbf{a} est dans le groupe général linéaire noté $GL(E)$ des automorphismes de E et $\mathbf{b} \in E$. L'ensemble :*

$$\phi(F) = \{\phi(P) \mid P \in F\}$$

est aussi une FEN. Si $\mu \in \mathcal{B}_F$, alors $\phi(F) = F(\phi(\mu))$ et :

$$\begin{aligned} \Theta_{\phi(\mu)} &= \mathbf{a}^*(\Theta_\mu) = \{\mathbf{a}^*(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta_\mu\}, \\ k_{\phi(\mu)}(\boldsymbol{\theta}) &= k_\mu(\mathbf{a}^*(\boldsymbol{\theta})) - \langle \mathbf{b}, \boldsymbol{\theta} \rangle & \forall \boldsymbol{\theta} \in \Theta_{\phi(\mu)}, \\ \phi(P_{\boldsymbol{\theta}, \mu}) &= P_{\mathbf{a}^*(\boldsymbol{\theta}), \phi(\mu)}, \\ M_{\phi(F)} &= \phi(M_F) = \{\phi(\mathbf{m}) \mid \mathbf{m} \in M_F\}, \\ V_{\phi(F)}(\tilde{\mathbf{m}}) &= \mathbf{a}V_F(\phi^{-1}(\tilde{\mathbf{m}}))\mathbf{a}^* & \forall \tilde{\mathbf{m}} \in \phi(M_F), \end{aligned}$$

où \mathbf{a}^* est l'automorphisme adjoint de \mathbf{a} défini comme : $\langle \mathbf{a}(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{a}^*(\mathbf{y}) \rangle \quad \forall (\mathbf{x}, \mathbf{y}) \in E^2$. Par exemple si $E = \mathbb{R}^d$, alors $\mathbf{a}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ où \mathbf{A} est une matrice inversible de $\mathbb{R}^{d \times d}$, et pour le produit scalaire euclidien, on a $\mathbf{a}^*(\mathbf{x}) = \mathbf{A}^T \mathbf{x}$.

La convolution est une autre opération importante pour les FEN. Soient $F(\mu)$ une FEN et n variables à valeurs dans E indépendantes $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, distribuées suivant la loi de probabilité $P_{\boldsymbol{\theta}, \mu}$. La loi jointe de ces variables est alors :

$$\exp\left(\sum_{i=1}^n \langle \boldsymbol{\theta}, \mathbf{x}_i \rangle - nk_\mu(\boldsymbol{\theta})\right) \prod_{i=1}^n \mu(d\mathbf{x}_i). \quad (\text{B.6})$$

On désigne par μ^{*n} la convolution suivante de mesures de $\mathcal{M}(E)$: μ^{*n} est l'image de la mesure $\prod_{i=1}^n \mu(d\mathbf{x}_i)$ par l'application $(\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \mathbf{x}_1 + \dots + \mathbf{x}_n$. Cette mesure est toujours dans $\mathcal{M}(E)$. La loi de la somme $\mathbf{S}_n = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n$ qui est la convolution d'ordre n de la loi $P_{\boldsymbol{\theta}, \mu}$, est notée $(P_{\boldsymbol{\theta}, \mu})^{*n}$. A partir de l'équation (B.6), on obtient le résultat suivant :

Proposition B.5.2. *La loi $(P_{\boldsymbol{\theta}, \mu})^{*n}$ est une loi exponentielle de la forme $P_{\boldsymbol{\theta}, \mu^{*n}}$ et l'ensemble ouvert des paramètres naturels de la FEN est $\Theta_{\mu^{*n}} = \Theta_\mu$. De plus, pour tout $\boldsymbol{\theta} \in \Theta_\mu$:*

$$k_{\mu^{*n}}(\boldsymbol{\theta}) = nk_\mu(\boldsymbol{\theta}), \quad \text{et par conséquent } \mathcal{L}_{\mu^{*n}}(\boldsymbol{\theta}) = (\mathcal{L}_\mu(\boldsymbol{\theta}))^n. \quad (\text{B.7})$$

La FEN $F(\mu^{*n})$ ainsi obtenue est alors appelée la $n^{\text{ième}}$ puissance ou convolution de $F(\mu)$.

Cette construction peut être généralisée. On introduit pour cela l'ensemble $\Lambda(\mu)$ des réels strictement positifs λ , pour lesquels il existe une mesure notée μ_λ , telle que :

$$\begin{aligned} \mathcal{L}_{\mu_\lambda}(\theta) &= (\mathcal{L}_\mu(\theta))^\lambda & \forall \theta \in \Theta_\mu, \\ \iff k_{\mu_\lambda}(\theta) &= \lambda k_\mu(\theta) & \forall \theta \in \Theta_\mu. \end{aligned} \quad (\text{B.8})$$

La FEN $F_\lambda = F(\mu_\lambda)$ est appelée la $\lambda^{\text{ième}}$ puissance de $F(\mu)$ et ne dépend pas de la base μ mais seulement de F . Cette FEN peut-être caractérisée par son domaine des moyennes et sa fonction variance :

$$\begin{aligned} M_{F_\lambda} &= \lambda M_F, \\ V_{F_\lambda}(\mathbf{m}) &= \lambda V_F\left(\frac{\mathbf{m}}{\lambda}\right) \quad \forall \mathbf{m} \in M_F. \end{aligned}$$

Le paramètre λ et l'ensemble $\Lambda_F = \Lambda(\mu)$ sont communément appelés le paramètre et l'ensemble de Jørgensen, ce dernier ayant souligné leur rôle dans l'étude des modèles exponentiels de dispersion [Jør87]. Il est clair, selon la proposition précédente, que pour toute FEN F , on a $\mathbb{N} \subset \Lambda_F$.

Exemple B.5.1 : la famille exponentielle naturelle $F = \{\mathcal{G}(1, p); p \in]0, +\infty[\}$ des lois exponentielles.

Les lois exponentielles sont un cas particulier des lois Gamma $\mathcal{G}(q, p)$ lorsque le paramètre de forme est $q = 1$. Comme cela a été montré dans l'exemple B.1.1, une base de la famille F est la mesure :

$$\mu(dx) = \mathbb{I}_{]0, +\infty[}(x)dx,$$

pour laquelle :

$$\mathcal{L}_\mu(\theta) = \frac{1}{-\theta} \quad \forall \theta \in \Theta_\mu =]-\infty, 0[.$$

Or pour tout $\lambda \in]0, +\infty[$, la mesure $\mu_\lambda \in \mathcal{M}(\mathbb{R})$, introduite dans l'exemple B.1.1, qui engendre la FEN $F_\lambda = \{\mathcal{G}(\lambda, p); p \in]0, +\infty[\}$ des lois gamma de paramètre de forme λ vérifie :

$$\mathcal{L}_{\mu_\lambda}(\theta) = \frac{1}{(-\theta)^\lambda} = (\mathcal{L}_\mu(\theta))^\lambda \quad \forall \theta \in \Theta_\mu =]-\infty, 0[.$$

L'ensemble de Jørgensen de la FEN F des lois exponentielles est donc $\Lambda_F =]0, +\infty[$. On obtient même un résultat plus général : l'ensemble de Jørgensen de la FEN F_q des lois Gamma $\mathcal{G}(q, p)$ de paramètre de forme $q > 0$ fixé, est $\Lambda_{F_q} = \Lambda_F =]0, +\infty[$. Une conséquence de ceci est que les lois Gamma $\mathcal{G}(q, p)$ où $(q, p) \in]0, +\infty[^2$ sont infiniment divisibles.

Exemple B.5.2 : la famille exponentielle naturelle $F = \{\mathcal{BN}(1, p); p \in]0, +\infty[\}$ des lois géométriques.

Les lois géométriques sont un cas particulier des lois binomiales négatives $\mathcal{G}(q, p)$ lorsque $q = 1$. On considère alors la mesure $\mu_\lambda \in \mathcal{M}(\mathbb{R})$, introduite dans l'exemple B.1.2, et qui engendre la FEN $F_\lambda = \mathcal{BN}\{(\lambda, p); p \in]0, 1[\}$ des lois binomiales négatives. Selon un raisonnement identique à l'exemple ci-dessus, on montre que l'ensemble de Jørgensen de la FEN F_q des lois binomiales négatives $\mathcal{BN}(q, p)$ de paramètre $q > 0$ fixé, est $\Lambda_{F_q} = \Lambda_F =]0, +\infty[$. Une conséquence de ceci est que les lois binomiales négatives où $(q, p) \in]0, +\infty[\times]0, 1[$ sont infiniment divisibles.

B.6 Estimateurs du maximum de vraisemblance

Une propriété importante des FEN est l'existence pour ces familles d'une statistique exhaustive simple.

Définition B.6.1. Soit $\{P_{\theta}; \theta \in \Theta\}$ un modèle statistique sur un espace abstrait noté (Ω, \mathcal{A}) . Une statistique T est dite exhaustive si la loi conditionnelle $\theta \mapsto P_{\theta}(\cdot | T)$ ne dépend pas de θ . Autrement dit, toute l'information disponible sur θ est contenue dans la statistique T .

Une conséquence directe est que si T est une statistique exhaustive, alors il est inutile d'observer autre chose pour obtenir une meilleure information sur θ .

On obtient le théorème suivant (voir [Let92, p. 38]) :

Théorème B.6.1. Soit une mesure $\mu \in \mathcal{M}(E)$ et $\Lambda(\mu)$ son ensemble de Jørgensen. Soient $\lambda_1, \dots, \lambda_n \in \Lambda(\mu)$, $\theta \in \Theta_{\mu}$ et $\mathbf{X}_1, \dots, \mathbf{X}_n$ des v.a. indépendantes de lois respectives $P_{\theta, \mu_{\lambda_j}}$ pour tout $1 \leq j \leq n$ (où μ_{λ_j} est la mesure qui engendre la $\lambda_j^{\text{ème}}$ puissance de la FEN $F(\mu)$). Si $\mathbf{S}_n = \mathbf{X}_1 + \dots + \mathbf{X}_n$, alors la statistique \mathbf{S}_n est exhaustive pour la loi du n -uplet $(\mathbf{X}_1, \dots, \mathbf{X}_n)$.

Un corollaire de ce théorème est que si $\mathbf{X}_1, \dots, \mathbf{X}_n$ sont des v.a. indépendantes de lois identiques $P_{\theta, \mu}$, alors la statistique $\mathbf{S}_n = \mathbf{X}_1 + \dots + \mathbf{X}_n$ est exhaustive pour ce modèle statistique, puisque $1 \in \Lambda_F$ pour toute FEN F .

Afin de déterminer l'estimateur du maximum de vraisemblance du paramètre naturel d'une FEN, on commence à étudier le cas où l'échantillon empirique n'est composé que d'une seule réalisation.

Lemme B.6.2. Soit une mesure $\mu \in \mathcal{M}(E)$. On observe \mathbf{x} qui est la réalisation d'un v.a. \mathbf{X} de loi $P_{\mu, \theta}$ où le paramètre naturel θ est inconnu. Alors l'estimateur du maximum de vraisemblance (EMV) de θ est défini si $\mathbf{x} \in M_{F(\mu)}$ et est égal à :

$$\widehat{\theta} = \psi_{\mu}(\mathbf{x}).$$

En paramétrant la FEN $F = F(\mu)$ par la moyenne \mathbf{m} , on obtient alors que l'EMV du paramètre \mathbf{m} est :

$$\widehat{\mathbf{m}} = \mathbf{x}.$$

Démonstration. On étudie la log-vraisemblance :

$$l(\theta; \mathbf{x}) = \langle \theta, \mathbf{x} \rangle - k_{\mu}(\theta).$$

La fonction $\theta \mapsto l(\theta; \mathbf{x})$ définie sur le convexe Θ_{μ} est la somme d'une fonction linéaire et d'une fonction strictement concave (car la fonction $\theta \mapsto k_{\mu}(\theta)$ est strictement convexe sur Θ_{μ} selon la proposition B.1.1). Par conséquent cette fonction est strictement concave sur Θ_{μ} . Sa différentielle est :

$$l'(\theta; \mathbf{x}) = \mathbf{x} - k'_{\mu}(\theta),$$

Si \mathbf{x} est dans M_F alors cette différentielle s'annule au point $\theta = \psi_\mu(\mathbf{x})$, et la stricte concavité de la log-vraisemblance assure que ce point est l'argument de l'unique maximum sur Θ_μ .

De plus, le paramètre θ et le vecteur des moyennes \mathbf{m} de la FEN sont reliés entre eux par la fonction bijective $\psi_\mu : M_F \rightarrow \Theta_\mu$ telle que $\theta = \psi_\mu(\mathbf{m})$. Lorsque la FEN $F = F(\mu)$ est paramétrée par la moyenne, le principe d'invariance fonctionnelle assure donc que l'EMV du paramètre \mathbf{m} est défini si $\mathbf{x} \in M_F$ et alors :

$$\widehat{\mathbf{m}} = \psi_\mu^{-1}(\psi_\mu(\mathbf{x})) = \mathbf{x}.$$

□

Le résultat précédent se généralise lorsqu'on dispose d'un échantillon empirique composé de plusieurs observations.

Théorème B.6.3. *Soit une mesure $\mu \in \mathcal{M}(E)$. On observe n réalisations $\mathbf{x}_1, \dots, \mathbf{x}_n$ de v.a. $\mathbf{X}_1, \dots, \mathbf{X}_n$ indépendants et de même loi $P_{\theta, \mu}$, où le paramètre naturel θ est inconnu. Soit $\mathbf{s}_n = \mathbf{x}_0 + \dots + \mathbf{x}_n$ la somme des réalisations. Alors l'EMV de θ est défini si $\mathbf{s}_n \in M_{F(\mu)}$ et est égal à :*

$$\widehat{\theta}_n = \psi_\mu\left(\frac{\mathbf{s}_n}{n}\right). \quad (\text{B.9})$$

Sous la même hypothèse, l'EMV du vecteur des moyennes \mathbf{m} est la moyenne empirique :

$$\widehat{\mathbf{m}}_n = \frac{\mathbf{s}_n}{n} = \overline{\mathbf{X}}_n. \quad (\text{B.10})$$

Démonstration. Selon le théorème B.6.1, la statistique $\mathbf{S}_n = \mathbf{X}_1 + \dots + \mathbf{X}_n$ est exhaustive pour θ . De plus, selon la proposition B.5.2, la loi de \mathbf{S}_n appartient à la FEN dont les lois sont $P_{\theta, \mu^{*n}}$ pour $\theta \in \Theta_\mu$. Donc d'après le lemme B.6.2 précédent, $\widehat{\theta}_n = \psi_{\mu^{*n}}(\mathbf{s}_n)$. Mais comme $k_{\mu^{*n}}(\theta) = nk_\mu(\theta)$ sur Θ_μ toujours selon la proposition B.5.2, il vient $k'_{\mu^{*n}}(\theta) = nk'_\mu(\theta)$ et donc $\psi_{\mu^{*n}}(\mathbf{m}) = \psi_\mu\left(\frac{\mathbf{m}}{n}\right)$. Finalement, on obtient les expressions du théorème. □

Remarques :

- Si $\overline{\mathbf{X}}_n \notin M_F$, les estimateurs du maximum de vraisemblance n'existent pas en général.
- Le vecteur moyenne \mathbf{m} appartient à l'ensemble ouvert M_F . Comme cela est expliqué en [Let], la loi forte des grands nombres assure qu'il existe presque sûrement une v.a. N , telle que pour tout $n \geq N$ on ait $\overline{\mathbf{X}}_n \in M_F$. Les estimateurs du maximum de vraisemblance seront donc bien définis pour un nombre suffisant d'observations.

Théorème B.6.4. *Soit une mesure $\mu \in \mathcal{M}(\mathbb{R}^d)$. L'EMV $\widehat{\mathbf{m}}_n$ défini à l'équation (B.10) du vecteur des moyennes $\mathbf{m} \in M_F$ de la FEN $F = F(\mu)$ est sans biais, de variance $\frac{1}{n}V_F(\mathbf{m})$ et efficace.*

Démonstration. Selon le théorème B.6.3, l'estimateur $\widehat{\mathbf{m}}_n$ est la moyenne empirique des observations, donc non biaisé. De plus sa variance est égale à $\frac{1}{n}\text{var}(\mathbf{X}_1)$. Or, la loi de \mathbf{X}_1 est dans la FEN $F = F(\mu)$, donc $\text{var}(\mathbf{X}_1) = V_F(\mathbf{m})$. Les hypothèses de régularité du modèle statistique sont vérifiées pour une FEN, et l'ensemble M_F des valeurs du paramètre \mathbf{m} est ouvert. Il suffit alors de montrer que la borne de Cramer Rao (BCR) est égale à la variance de notre estimateur. Pour cela on peut supposer que l'échantillon est réduit à l'observation \mathbf{x} d'un seul v.a. \mathbf{X} , donc $n = 1$.

Le vecteur score $g(\mathbf{X}, \boldsymbol{\theta})$, pour $\boldsymbol{\theta} \in \Theta_\mu$, est défini comme la différentielle par rapport à $\boldsymbol{\theta}$ de la log-vraisemblance, donc :

$$\text{var}(g(\mathbf{X}, \boldsymbol{\theta})) = \text{var}(\mathbf{X} - k'_\mu(\boldsymbol{\theta})) = \text{var}(\mathbf{X}) = V_F(\mathbf{m}).$$

La matrice d'information de Fisher pour la paramétrisation naturelle de la FEN F est donc $V_F(\mathbf{m})$. Par conséquent la BCR pour la paramétrisation naturelle est $V_F(\mathbf{m})^{-1}$. Comme $\mathbf{m} = k'_\mu(\boldsymbol{\theta})$, la BCR pour le paramètre \mathbf{m} est égale à :

$$\begin{aligned} k''_\mu(\psi_\mu(\mathbf{m}))V_F(\mathbf{m})^{-1}k''_\mu(\psi_\mu(\mathbf{m}))^T &= V_F(\mathbf{m})V_F(\mathbf{m})^{-1}V_F(\mathbf{m}), \\ &= V_F(\mathbf{m}). \end{aligned}$$

□

De plus, l'unicité de la transformée de Laplace d'une mesure de $\mathcal{M}(E)$ assure que pour le modèle statistique des FEN, la statistique \mathbf{S}_n est complète (ou encore qualifiée de *totale*). Les conditions d'application du théorème de Rao Blackwell sont donc vérifiées pour toutes les FEN définies sur $E = \mathbb{R}^d$. Il vient alors que si $f : \Theta_\mu \rightarrow \mathbb{R}$ et s'il existe une fonction $c : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\mathbf{E}[c(\mathbf{S}_n)] = f(\boldsymbol{\theta})$, alors $c(\mathbf{S}_n)$ est un estimateur de variance uniformément minimum parmi tous les estimateurs sans biais. Cependant cet estimateur n'est pas nécessairement efficace.

C

Lois Gamma Multivariées

Cette annexe regroupe les démonstrations des propriétés relatives aux lois Gamma bivariées, présentées à la section 1.2.2, aux lois Gamma multivariées, présentées à la section 1.3.2 et aux lois multinomiales négatives, présentées à la section 1.3.3.

Une identité importante, car utilisée dans beaucoup de ces démonstrations, est le développement en série entière suivant :

$$\left(\frac{1}{1-z}\right)^p = \sum_{k=0}^{\infty} \frac{p(p+1)\dots(p+k-1)}{k!} z^k = \sum_{k=0}^{\infty} \frac{(p)_k}{k!} z^k \text{ pour tout } |z| < 1, \quad (\text{C.1})$$

où $(a)_n$ désigne le symbole de Pochhammer défini aux équations (1.21) et (1.22). Cette annexe est découpée en trois sections : une première section consacrée aux lois Gamma bivariées, une deuxième aux lois Gamma multivariées et la dernière aux lois négatives multinomiales (qui sont étroitement liées aux lois Gamma).

C.1 Loi Gamma Bivariée

C.1.1 Log-convexité de la transformée de Laplace bivariée

Selon l'équation 1.16, la log-transformée de Laplace (avec la convention $\varepsilon = +1$ dans la définition A.2.1 de la transformée de Laplace) d'une loi Gamma bivariée a pour expression :

$$k(z_1, z_2) = -q \log(1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2),$$

avec $q > 0$, $p_1 > 0$ et $p_2 > 0$, et est définie au moins sur l'ensemble convexe (car les v.a. sont positives) $]0, \infty[^2$. La log-transformée de Laplace étant une fonction analytique réelle, convexe sur l'intérieur de son domaine de définition, selon la proposition A.2.5, le hessien de la fonction k doit être semi-défini positif sur l'ouvert convexe $]0, \infty[^2$. Le calcul de la matrice hessienne, notée $k''(z_1, z_2)$ donne :

$$k''(z_1, z_2) = \frac{q}{(1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^2} \begin{pmatrix} (p_1 + z_2 p_{12})^2 & p_1 p_2 - p_{12} \\ p_1 p_2 - p_{12} & (p_2 + p_{12} z_1)^2 \end{pmatrix}.$$

La trace de cette matrice est positive pour tout $(z_1, z_2) \in]0, \infty[^2$. Par conséquent, la matrice $k''(z_1, z_2)$ est semi-définie positive pour tout $(z_1, z_2) \in]0, \infty[^2$ si son déterminant est positif ou nul

pour tout $(z_1, z_2) \in]0, \infty[^2$. Le calcul du déterminant donne :

$$f(z_1, z_2) = \det k''(z_1, z_2) = q^2 \frac{p_{12}(2p_1p_2 - p_{12} + p_1p_{12}z_1 + p_{12}p_2z_2 + p_{12}^2z_1z_2)}{(1 + p_1z_1 + p_2z_2 + p_{12}z_1z_2)^3}.$$

Par composition de fonctions continues, la fonction f est aussi continue sur $]0, \infty[^2$. De plus sa limite à droite au point $(0, 0)$ existe et est égale à :

$$q^2 p_{12}(2p_1p_2 - p_{12}).$$

Cette limite est strictement négative si $p_{12} < 0$ ou si $p_{12} > 2p_1p_2$ et, dans ce cas, par continuité, la fonction f est strictement négative sur un ensemble de mesure non nulle de $]0, \infty[^2$ et la log-transformée de Laplace n'est donc pas convexe. En définitive la log-convexité de la transformée de Laplace impose :

$$0 \leq p_{12} \leq 2p_1p_2. \quad (\text{C.2})$$

C.1.2 Moments joints d'une loi Gamma bivariée

On considère la transformée de Laplace d'une loi Gamma bivariée (X, Y) introduite à l'équation 1.16 :

$$\mathcal{L}_{(X,Y)}(z_1, z_2) = \mathbb{E} \left[e^{-z_1X - z_2Y} \right] = \left(\frac{1}{1 + p_1z_1 + p_2z_2 + p_{12}z_1z_2} \right)^q, \quad (\text{C.3})$$

avec $q > 0$, $p_1 > 0$, $p_2 > 0$ et $0 < p_{12} \leq p_1p_2$. Cette transformée peut être reparamétrisée en fonction de θ_1 , θ_2 et $c = \frac{p_1p_2 - p_{12}}{p_{12}^2}$, tels que²⁰ :

$$p_1 = \frac{\theta_1}{\theta_1\theta_2 - c}, \quad p_2 = \frac{\theta_2}{\theta_1\theta_2 - c}, \quad p_{12} = \frac{1}{\theta_1\theta_2 - c}.$$

On a donc :

$$\theta_1 = \frac{p_1}{p_{12}} > 0, \quad \theta_2 = \frac{p_2}{p_{12}} > 0, \quad \theta_1\theta_2 - c = \frac{1}{p_{12}} > 0, \quad (\text{C.4})$$

et on obtient l'expression suivante de la transformée de Laplace :

$$\begin{aligned} \mathcal{L}_{(X,Y)}(x, y) &= \left(\frac{\theta_1\theta_2 - c}{(\theta_1 + x)(\theta_2 + y) - c} \right)^q \\ &= \left(\frac{\theta_1\theta_2 - c}{(\theta_1 + x)(\theta_2 + y)} \right)^q \underbrace{\left(\frac{1}{1 - \frac{c}{(\theta_1 + x)(\theta_2 + y)}} \right)^q}_A. \end{aligned}$$

Par ailleurs, $\frac{c}{\theta_1\theta_2} = \frac{p_1p_2 - p_{12}}{p_1p_2} < 1$, car pour toute loi gamma bivariée : $0 < p_{12} \leq p_1p_2$, et donc $\frac{c}{(\theta_1 + x)(\theta_2 + y)} < 1$ pour tout $x, y \geq 0$. On peut donc appliquer le développement en série entière présenté à l'équation (C.1), et l'on obtient :

$$A = \sum_{k=0}^{\infty} \frac{(q)_k}{k!} \frac{c^k}{(\theta_1 + x)^k (\theta_2 + y)^k}.$$

²⁰Cette paramétrisation correspond à la paramétrisation canonique des FEN des lois Gamma bivariées obtenues à q et c fixés, comme cela est expliqué à la section 1.2.2

D'où pour tout $(x, y) \in [0, \theta_1] \times [0, \theta_2]$:

$$\begin{aligned} \mathcal{L}_{(X,Y)}(x, y) &= (\theta_1\theta_2 - c)^q \sum_{k=0}^{\infty} \frac{(q)_k}{k!} \frac{c^k}{(\theta_1 + x)^{q+k}(\theta_2 + y)^{q+k}} \\ &= (\theta_1\theta_2 - c)^q \sum_{k=0}^{\infty} \frac{(q)_k c^k}{k!} \frac{1}{(\theta_1)^{q+k}} \sum_{i=0}^{\infty} \frac{(q+k)_i}{i!} \left(-\frac{x}{\theta_1}\right)^i \frac{1}{(\theta_2)^{q+k}} \sum_{j=0}^{\infty} \frac{(q+k)_j}{j!} \left(-\frac{y}{\theta_2}\right)^j \\ &= \left(\frac{\theta_1\theta_2 - c}{\theta_1\theta_2}\right)^q \sum_{k,i,j \geq 0} \frac{(q)_k}{k!} \left(\frac{c}{\theta_1\theta_2}\right)^k (q+k)_i (q+k)_j \frac{1}{i!} \left(-\frac{x}{\theta_1}\right)^i \frac{1}{j!} \left(-\frac{y}{\theta_2}\right)^j. \end{aligned}$$

Il vient donc selon la proposition A.2.2 et grâce à l'unicité du développement en série entière que pour tout $(m, n) \in \mathbb{N}^2$:

$$\mathbb{E}[X^m Y^n] = \left(\frac{\theta_1\theta_2 - c}{\theta_1\theta_2}\right)^q \frac{1}{\theta_1^m \theta_2^n} \sum_{k \geq 0} \frac{(q)_k}{k!} \left(\frac{c}{\theta_1\theta_2}\right)^k (q+k)_m (q+k)_n.$$

Or :

$$(q)_k (p+k)_m (q+k)_n = (q)_m (q)_n \frac{(q+m)_k (q+n)_k}{(q)_k}$$

, d'où selon (C.4) :

$$\begin{aligned} \mathbb{E}[X^m Y^n] &= \left(\frac{p_{12}}{p_1 p_2}\right)^q \frac{p_{12}^{m+n}}{p_1^m p_2^n} (p)_m (p)_n \sum_{k \geq 0} \frac{(q+m)_k (q+n)_k}{(q)_k} \frac{r^k}{k!} \\ &= \left(\frac{p_{12}}{p_1 p_2}\right)^q \frac{p_{12}^{m+n}}{p_1^m p_2^n} (q)_m (q)_n {}_2F_1(q+m, q+n; p; r), \end{aligned} \quad (\text{C.5})$$

où ${}_2F_1$ est la série hypergéométrique standard définie comme :

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \quad (\text{C.6})$$

et où r est défini comme (r est en fait le coefficient de corrélation) :

$$r = \frac{p_1 p_2 - p_{12}}{p_1 p_2}.$$

En utilisant les propriétés suivantes des fonctions hypergéométriques de la forme ${}_2F_1$:

1. La série hypergéométrique ${}_2F_1(a, b; c; z)$ converge absolument pour tout $|z| < 1$ si c n'est pas un entier négatif, et pour tout $|z| \leq 1$ si $\Re(c - a - b) > 0$,
2. ${}_2F_1(a, b; c; z) = (1 - z)^{c-a-b} {}_2F_1(c - a, c - b; c; z)$ pour $|z| < 1$,

Il vient finalement :

$$\mathbb{E}[X^m Y^n] = p_1^m p_2^n (q)_m (q)_n {}_2F_1(-m, -n; p; r), \quad (\text{C.7})$$

$$= p_1^m p_2^n (q)_m (q)_n \sum_{k=0}^{\min(m,n)} \frac{(-m)_k (-n)_k}{(q)_k} \frac{r^k}{k!}. \quad (\text{C.8})$$

C.1.3 Développement de la loi Gamma bivariée en série de polynômes de Laguerre

A partir de la caractérisation des polynômes de Laguerre donnée à l'équation (1.43), on obtient par un calcul direct :

$$\int_{\mathbb{R}^+} e^{-zx} \left(\frac{n!}{\Gamma(\alpha+n)} e^{-x} x^{\alpha-1} L_n^{(\alpha-1)}(x) \right) dx = \frac{z^n}{(1+z)^{\alpha+n}} \quad \forall n \in \mathbb{N}, \alpha > 0, z > 0. \quad (\text{C.9})$$

Cette formule fournit une expression de la transformée de Laplace inverse de la fonction $z > 0 \mapsto \frac{z^n}{(1+z)^{\alpha+n}}$. Cette expression est utile afin d'inverser la transformée de Laplace de la loi Gamma bivariée. En effet, on peut supposer que les paramètres d'échelle de la loi Gamma bivariée vérifient $p_1 = p_2 = 1$, et la transformée de Laplace (1.16) associée au couple de v.a. (X_1, X_2) devient alors pour tout $z_1, z_2 \geq 0$:

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = \left(\frac{1}{(1+z_1)(1+z_2) - rz_1 z_2} \right)^q = \left(\frac{1}{(1+z_1)(1+z_2)} \right)^q \left(\frac{1}{1 - r \frac{z_1 z_2}{(1+z_1)(1+z_2)}} \right)^q,$$

où $0 \leq r < 1$ est le coefficient de corrélation entre les marges. En remarquant que $0 \leq r \frac{z_1 z_2}{(1+z_1)(1+z_2)} < 1$ pour tout $z_1, z_2 \geq 0$, le développement en série entière présenté à l'équation (C.1) assure alors :

$$\mathcal{L}_{(z_1, z_2)} = \sum_{k=0}^{\infty} \frac{(q)_k r^k}{k!} \frac{z_1^k}{(1+z_1)^{q+k}} \frac{z_2^k}{(1+z_2)^{q+k}} \quad \forall z_1, z_2 \geq 0.$$

Grâce la formule (C.9) et à la proposition A.2.1, on obtient finalement l'expression suivante de la densité du couple (X_1, X_2) pour tout $q > 0$ et $0 \leq r < 1$:

$$p_{(X_1, X_2)}(x_1, x_2) = \prod_{i=1}^2 \frac{e^{-x_i} x_i^{q-1}}{\Gamma(q)} \sum_{k=0}^{\infty} r^k \frac{k! \Gamma(q)}{\Gamma(q+k)} L_k^{(q-1)}(x_1) L_k^{(q-1)}(x_2) \mathbb{I}_{(x_1 > 0, x_2 > 0)}(x_1, x_2).$$

En introduisant maintenant la famille de polynôme :

$$P_k^{(q)}(x) = \left(\frac{k!}{(q)_k} \right)^{1/2} L_k^{(q-1)}(x),$$

on obtient, d'après l'équation 1.44, que les polynômes $P_k^{(q)}$ sont orthonormaux pour le produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{G}(q,1)}$, où la fonction poids correspond à la densité (1.1) de la loi Gamma $\mathcal{G}(q, 1)$. De plus, la densité du couple (X_1, X_2) devient :

$$p_{(X_1, X_2)}(x_1, x_2) = \mathcal{G}_{q,1}(x_1) \mathcal{G}_{q,1}(x_2) \sum_{k=0}^{\infty} r^k P_k^{(q)}(x_1) P_k^{(q)}(x_2).$$

C.1.4 Approximation de l'information mutuelle d'une loi Gamma bivariée lorsque $r \rightarrow 1$

Comme cela est montré à la section 2.1.2, lorsque $r \rightarrow 1$, l'information mutuelle associée à une loi Gamma bivariée a pour expression :

$$\text{IM}_q(r) = -q \log(1-r) - 2 \frac{qr}{1-r} + A,$$

où :

$$A = 2\sqrt{c} \mathbb{E} \left[\sqrt{X_1 X_2} \right] - \left(\frac{q}{2} - 1 \right) (\log c + \mathbb{E} [\log X_1 + \log X_2]) + \log \left[\frac{\Gamma(q)}{2\sqrt{\pi}} \right] + o(1),$$

(X_1, X_2) étant une loi Gamma bivariée de paramètre de forme q , de moyenne (q, q) et de coefficient de corrélation entre les marges r , $c = r/(1-r)^2$, et $o(1) \rightarrow 0$ quand $r \rightarrow 1$. Il est donc nécessaire de calculer les moyennes des variables $\sqrt{X_1 X_2}$, $\log(X_1)$ et $\log(X_2)$ afin d'obtenir une expression asymptotique de l'information mutuelle.

$\mathbb{E} [\log U]$ où $U \sim \Gamma(q, p)$

L'espérance de $\log U$ peut se calculer à l'aide du simple changement de variables $V = U/p$. Alors $V \sim \Gamma(q, 1)$ et :

$$\begin{aligned} \mathbb{E} [\log V] &= \frac{1}{\Gamma(q)} \int_0^\infty \log(v) e^{-v} v^{q-1} dv = \frac{1}{\Gamma(q)} \int_0^\infty \frac{d}{dq} (e^{-v} v^{q-1}) dv, \\ &= \frac{1}{\Gamma(q)} \frac{d}{dq} \int_0^\infty e^{-v} v^{q-1} dv = \frac{1}{\Gamma(q)} [\Gamma'(q)] = \psi(q). \end{aligned}$$

On obtient finalement :

$$\mathbb{E} [\log U] = \mathbb{E} [\log V] - \log(p) = \psi(q) - \log(p). \quad (\text{C.10})$$

$\mathbb{E} \left[\sqrt{X_1 X_2} \right]$ et son approximation lorsque $r \rightarrow 1$

Un calcul direct, à partir de la densité d'une loi gamma bivariée de paramètres $p_1 = m_1/q$, $p_2 = m_2/q$ et $p_{12} = p_1 p_2 (1-r)$, donnée à l'équation (1.35), permet de montrer que l'expression des moments joints d'une loi Gamma bivariée (C.7) se généralise pour tous réels $m, n > 0$. Par conséquent, il vient :

$$\mathbb{E} \left[\sqrt{X_1 X_2} \right] = \frac{\sqrt{m_1 m_2} \Gamma(q + 1/2)^2}{q \Gamma(q)^2} {}_2F_1 \left(\frac{-1}{2}, \frac{-1}{2}; q; r \right).$$

La relation suivante :

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)},$$

permet d'obtenir :

$${}_2F_1 \left(\frac{-1}{2}, \frac{-1}{2}; q; 1 \right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2},$$

et donc le développement de Taylor au premier ordre autour de $z = 1$:

$${}_2F_1 \left(\frac{-1}{2}, \frac{-1}{2}; q; z \right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} + (z-1) {}_2F_1' \left(\frac{-1}{2}, \frac{-1}{2}; q; 1 \right) + o(1-z),$$

où $o(1-z)/(1-z)$ tend vers 0, lorsque $z \rightarrow 1$.

Comme :

$${}_2F_1'(a, b; c; z) = \frac{ab}{c} {}_2F_1(a+1, b+1; c+1; z) \quad \text{pour } |z| < 1,$$

le développement de Taylor précédent peut être écrit :

$${}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; z\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} \left(1 + \frac{z-1}{4q}\right) + o(1-z).$$

Finalement :

$$\begin{aligned} \mathbb{E}\left[\sqrt{x_1 x_2}\right] &= \frac{\sqrt{m_1 m_2}}{q} \frac{\Gamma(q+1/2)^2}{\Gamma(q)^2} \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} \left(1 - \frac{1-r}{4q}\right) + o(1-r), \\ &= \sqrt{m_1 m_2} \left(1 - \frac{1-r}{4q}\right) + o(1-r) \quad \text{pour } r \rightarrow 1. \end{aligned}$$

Enfin, en regroupant tous les termes et en remplaçant m_1 et m_2 par leur valeur q , on obtient l'expression (2.8) de l'information mutuelle.

C.2 Loi Gamma multivariée

C.2.1 LGM éuicorrélée

On s'intéresse à la densité associée au vecteur \mathbf{X} , qui admet la transformée de Laplace suivante :

$$\mathcal{L}_{\mathbf{X}}(z) = \left[\frac{-r}{1-r} + \frac{1}{1-r} \prod_{i=1}^n (1+z_i) \right]^{-q},$$

définie pour tout $r \in [0, 1[$, $q > 0$ et pour tout $z_1, \dots, z_d \geq 0$. On note alors :

$$u = \prod_{i=1}^n (1+z_i),$$

En remarquant maintenant que $0 \leq \frac{r}{u} < 1$ pour tout $r \in]0, 1[$ et $z_1, \dots, z_d \geq 0$, il vient grâce au développement en série entière (C.1) :

$$\begin{aligned} \mathcal{L}_{\mathbf{X}}(z) &= \left(\frac{1-r}{u}\right)^q \sum_{k \geq 0} \frac{(q)_k}{k!} \frac{r^k}{u^k}, \\ &= (1-r)^q \sum_{k \geq 0} \frac{(q)_k r^k}{k!} \frac{1}{u^{q+k}}, \\ &= (1-r)^q \sum_{k \geq 0} \frac{(q)_k r^k}{k!} \prod_{i=1}^d \frac{1}{(1+z_i)^{q+k}} \quad \forall z_1, \dots, z_d \geq 0. \end{aligned}$$

Grâce à l'identité suivante :

$$\frac{1}{(1+z_i)^{q+k}} = \int_{]0,\infty[} e^{-z_i x_i} e^{-x_i} \frac{x_i^{q+k-1}}{\Gamma(q+k)} dx_i, \quad \forall z_i \geq 0,$$

on obtient, selon la proposition A.2.1, que la densité associée au vecteur \mathbf{X} a l'expression suivante au point $\mathbf{x} = (x_1, \dots, x_d)^T$:

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) &= (1-r)^q \left(\prod_{i=1}^d e^{-x_i} x_i^{q-1} \right) \sum_{k \geq 0} \frac{(q)_k r^k}{k!} \frac{x_1^k}{\Gamma(q+k)} \cdots \frac{x_d^k}{\Gamma(q+k)} \mathbb{I}_{]0,\infty[^d}(\mathbf{x}), \\ &= (1-r)^q \left(\prod_{i=1}^d e^{-x_i} \frac{x_i^{q-1}}{\Gamma(q)} \right) {}_0F_{d-1}(q, \dots, q; r x_1 \dots x_d) \mathbb{I}_{]0,\infty[^d}(\mathbf{x}) \quad \forall q > 0, r \in [0, 1[, \end{aligned}$$

où ${}_0F_{d-1}$ est la fonction hypergéométrique telle que :

$${}_0F_{d-1}(a_1, \dots, a_{d-1}; x) = \sum_{k \geq 0} \frac{1}{(a_1)_k \dots (a_{d-1})_k} \frac{x^k}{k!} \quad \forall x \in \mathbb{R},$$

et $(a)_k$ est le symbole de Pochhammer défini aux équations (1.21) et (1.22).

C.2.2 LGM à structure de corrélation markovienne

On cherche dans cette section à caractériser les LGM qui satisfont à la propriété de Markov suivante :

$$\Pr(X_{i+1} \in B | X_i = x_i, \dots, X_1 = x_1) = \Pr(X_{i+1} \in B | X_i = x_i),$$

pour tout borélien $b \in \mathcal{B}(\mathbb{R}^d)$, le vecteur $\mathbf{X} = (X_1, \dots, X_d)^T$ étant un vecteur aléatoire distribué suivant une LGM de paramètre de forme $q > 0$ fixé.

Afin de caractériser ces lois, il est utile de déterminer la transformée de Laplace de la variable X_j sachant X_i , avec $j > i$. On peut supposer sans perte de généralité que la transformée de Laplace du couple (X_i, X_j) est de la forme :

$$\mathcal{L}_{(X_i, X_j)}(z_i, z_j) = \left(1 + z_i + z_j + (1 - r_{i,j}) z_i z_j \right)^{-q}, \quad (\text{C.11})$$

le paramètre $r_{i,j}$ étant le coefficient de corrélation. La densité de cette loi, notée $f_{(X_i, X_j)}$, est exprimée à l'équation (1.35), et l'on obtient l'expression de la densité de X_j sachant X_i :

$$\begin{aligned} f_{X_j}^{X_i=x_i}(x_j) &= \frac{f_{(X_i, X_j)}(x_i, x_j)}{f_{X_i}(x_i)}, \\ &= \frac{1}{(1-r_{i,j})^q} x_j^{q-1} e^{-\frac{r_{i,j}}{1-r_{i,j}} x_i - \frac{1}{1-r_{i,j}} x_j} f_q \left(\frac{r_{i,j}}{(1-r_{i,j})^2} x_i x_j \right). \end{aligned}$$

Par conséquent la transformée de Laplace de la v.a. X_j sachant X_i s'écrit :

$$\begin{aligned}
 \mathcal{L}_{X_j}^{X_i=x_i}(z_j) &= \mathbb{E} \left[e^{-z_j X_j} \mid X_i = x_i \right] = \int_0^\infty e^{-z_j x_j} f_{X_j}^{X_i=x_i}(x_j) dx_j, \\
 &= \frac{e^{-\frac{r_{i,j}}{1-r_{i,j}} x_i}}{(1-r_{i,j})^q} \sum_{k \geq 0} \frac{1}{\Gamma(q+k)k!} \left(\frac{r_{i,j} x_i}{(1-r_{i,j})^2} \right)^k \int_0^\infty e^{-\frac{1+(1-r_{i,j})z_j}{1-r_{i,j}} x_j} x_j^{q+k-1} dx_j, \\
 &= \left(1 + (1-r_{i,j})z_j \right)^{-q} e^{-\frac{r_{i,j}}{1-r_{i,j}} x_i} \sum_{k \geq 0} \frac{1}{k!} \left(\frac{r_{i,j} x_i}{(1-r_{i,j})(1+(1-r_{i,j})z_j)} \right)^k, \\
 &= \left(1 + (1-r_{i,j})z_j \right)^{-q} e^{-\frac{r_{i,j}}{1-r_{i,j}} x_i} e^{\frac{r_{i,j} x_i}{(1-r_{i,j})(1+(1-r_{i,j})z_j)}}, \\
 &= \left(1 + (1-r_{i,j})z_j \right)^{-q} e^{-\frac{r_{i,j} z_j}{1-(1-r_{i,j})z_j} x_i},
 \end{aligned}$$

et donc $\mathcal{L}_{X_j}^{X_i=x_i}(z_j) = H_{i,j}(z_j) \exp \left[-G_{i,j}(z_j) x_i \right]$, avec :

$$\begin{aligned}
 H_{i,j}(z_j) &= \left(1 + (1-r_{i,j})z_j \right)^{-q}, \\
 G_{i,j}(z_j) &= \frac{r_{i,j} z_j}{1 - (1-r_{i,j})z_j}.
 \end{aligned}$$

On cherche maintenant à prouver que la transformée de Laplace du vecteur \mathbf{X} est de la forme :

$$\mathcal{L}_{\mathbf{X}}(\mathbf{z}) = \det \left(\mathbf{I}_d + \text{Diag}(\mathbf{z}) R_{1/2} \right)^{-q} \equiv \phi_d(\mathbf{z}), \quad (\text{C.12})$$

où \mathbf{I}_d est la matrice identité d'ordre d , $\text{Diag}(\mathbf{z})$ la matrice diagonale dont les éléments diagonaux sont les composantes du vecteur \mathbf{z} , et $R_{1/2} = (a_{i,j})_{1 \leq i, j \leq d}$ la matrice de corrélation (donc symétrique) telle que :

$$\begin{aligned}
 a_{i,i} &= 1 \quad \forall 1 \leq i \leq d, \\
 a_{i,i+1} &= \sqrt{r_{i,i+1}} \quad \forall 1 \leq i \leq d-1, \\
 a_{i,j} &= \sqrt{r_{i,j}} = \sqrt{r_{i,k}} \sqrt{r_{k,j}}, \quad \forall 1 \leq i < k < j \leq d.
 \end{aligned}$$

On peut raisonner par récurrence. En dimension $d = 2$, la transformée de Laplace d'une loi Gamma bivariée donnée à l'équation (C.11) a bien pour expression :

$$\mathcal{L}_{(X_i, X_j)}(\mathbf{z}) = \det \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} z_i & 0 \\ 0 & z_j \end{pmatrix} \begin{pmatrix} 1 & \sqrt{r_{i,j}} \\ \sqrt{r_{i,j}} & 1 \end{pmatrix} \right]^{-q} = \left((1+z_i)(1+z_j) - r_{i,j} z_i z_j \right)^{-q}.$$

On suppose la propriété vraie pour la dimension $d - 1$. Grâce à la propriété de Markov, la transformée de Laplace du vecteur \mathbf{X} est alors :

$$\begin{aligned}
 \mathcal{L}_{\mathbf{X}}(\mathbf{z}) &= \mathbb{E} \left[e^{-z_1 X_1 - \dots - z_d X_d} \right] = \mathbb{E} \left[\mathbb{E} \left[e^{-z_1 X_1 - \dots - z_d X_d} \mid X_1, \dots, X_{d-1} \right] \right], \\
 &= \mathbb{E} \left[e^{-z_1 X_1 - \dots - z_{d-1} X_{d-1}} \mathbb{E} \left[e^{-z_d X_d} \mid X_1, \dots, X_{d-1} \right] \right], \\
 &= \mathbb{E} \left[e^{-z_1 X_1 - \dots - z_{d-1} X_{d-1}} \mathbb{E} \left[e^{-z_d X_d} \mid X_{d-1} \right] \right], \\
 &= H_{d-1,d}(z_d) \mathbb{E} \left[e^{-z_1 X_1 - \dots - (z_{d-1} + G_{d-1,d}(z_d)) X_{d-1}} \right], \\
 &= H_{d-1,d}(z_d) \mathcal{L}_{(X_1, \dots, X_{d-1})}(z_1, \dots, z_{d-2}, z_{d-1} + G_{d-1,d}(z_d)), \\
 &= H_{d-1,d}(z_d) \phi_{d-1}(z_1, \dots, z_{d-2}, z_{d-1} + G_{d-1,d}(z_d)).
 \end{aligned}$$

Finalement, l'expression trouvée correspond bien à $\phi_d(\mathbf{z})$ (en développant le déterminant de la matrice $(\mathbf{I}_d + \text{Diag}(\mathbf{z})R_{1/2})$ par rapport à la d -ième ligne et en utilisant la propriété de multilinéarité du déterminant pour calculer $\phi_{d-1}(z_1, \dots, z_{d-2}, z_{d-1} + G_{d-1,d}(z_d))$ on retrouve bien les mêmes expressions).

La transformée de Laplace du vecteur \mathbf{X} correspond donc à l'expression donnée à l'équation C.12. Une conséquence immédiate est que la transformée de Laplace du couple (X_i, X_j) avec $1 \leq i < i+l \leq d$ s'écrit :

$$\begin{aligned} \mathcal{L}_{(X_i, X_{i+l})}(z_i, z_{i+l}) &= \mathcal{L}_{\mathbf{X}}(0, \dots, 0, z_i, 0, \dots, 0, z_{i+l}, 0, \dots, 0), \\ &= \det \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} z_i & 0 \\ 0 & z_{i+l} \end{pmatrix} \begin{pmatrix} 1 & a_{i,i+l} \\ a_{i,i+l} & 1 \end{pmatrix} \right]^{-q}, \\ &= \left((1+z_i)(1+z_{i+l}) - a_{i,i+l}^2 z_i z_{i+l} \right)^{-q}, \end{aligned}$$

où $a_{i,i+l} = \sqrt{r_{i,i+1}r_{i+1,i+2} \dots r_{i+l-1,i+l}}$ et par conséquent :

$$r(X_i, X_{i+l}) = r(X_i, X_{i+1})r(X_{i+1}, X_{i+2}) \dots r(X_{i+l-1}, X_{i+l}) \quad \forall 1 \leq i < i+l \leq d.$$

C.2.3 Développement en série de polynômes de Laguerre de la densité d'une LGM

On s'intéresse à la densité associée à la transformée de Laplace trivariée suivante :

$$\begin{aligned} \mathcal{L}_{\mathbf{Y}}(z_1, z_2, z_3) &= ((1+z_1)(1+z_2)(1+z_3) - r_{12}z_1z_2 - r_{13}z_1z_3 - r_{23}z_2z_3 - \rho z_1z_2z_3)^{-q}, \\ &= [(1+z_1)(1+z_2)(1+z_3)]^{-q} (1 - r_{12}t_1t_2 - r_{13}t_1t_3 - r_{23}t_2t_3 - \rho t_1t_2t_3)^{-q}, \end{aligned} \quad (\text{C.13})$$

où $t_i = \frac{z_i}{1+z_i}$ pour tout $z_i \geq 0$ et $1 \leq i \leq 3$. En se restreignant à l'ensemble des z_i tels que $0 \leq r_{12}t_1t_2 + r_{13}t_1t_3 + r_{23}t_2t_3 + \rho t_1t_2t_3 < 1$ (cet ensemble existe, il suffit de prendre z_1, z_2, z_3 suffisamment grands), on trouve selon le développement en série entière (C.1) :

$$\mathcal{L}_{\mathbf{Y}}(z_1, z_2, z_3) = [(1+z_1)(1+z_2)(1+z_3)]^{-q} \sum_{n \geq 0} \frac{(q)_n}{n!} (r_{12}t_1t_2 + r_{13}t_1t_3 + r_{23}t_2t_3 + \rho t_1t_2t_3)^n.$$

Mais selon la formule du multinôme de Newton :

$$(r_{12}t_1t_2 + r_{13}t_1t_3 + r_{23}t_2t_3 + \rho t_1t_2t_3)^n = \sum_{n_1+n_2+n_3+n_4=n} \frac{n! r_{12}^{n_1} r_{13}^{n_2} r_{23}^{n_3} \rho^{n_4}}{n_1! n_2! n_3! n_4!} t_1^{N_1} t_2^{N_2} t_3^{N_3},$$

où les entiers N_i sont définis par :

$$N_1 = n - n_3, \quad N_2 = n - n_2, \quad N_3 = n - n_1,$$

on obtient finalement :

$$\mathcal{L}_{\mathbf{Y}}(z_1, z_2, z_3) = \sum_{n \geq 0} (q)_n \sum_{n_1+n_2+n_3+n_4=n} \frac{r_{12}^{n_1} r_{13}^{n_2} r_{23}^{n_3} \rho^{n_4}}{n_1! n_2! n_3! n_4!} \prod_{i=1}^3 \frac{z_i^{N_i}}{(1+z_i)^{q+N_i}}.$$

À l'aide de la formule (C.9), on reconnaît l'expression, valide pour tout $z_i \geq 0$ et tout $1 \leq i \leq 3$, de la transformée de Laplace suivante :

$$\frac{z_i^{N_i}}{(1+z_i)^{q+N_i}} = \int_0^\infty e^{-z_i y_i} \left(\frac{N_i!}{\Gamma(q+N_i)} e^{-y_i} y_i^{q-1} L_{N_i}^{(q-1)}(y_i) \right) dy_i,$$

où $L_n^{(a)}$ désigne les polynômes de Laguerre généralisés définis à la section 1.2.2. On conclut alors grâce à la proposition A.2.1 et l'on trouve l'expression suivante de la densité :

$$p_Y(y_1, y_2, y_3) = \left(\prod_{i=1}^4 e^{-y_i} \frac{y_i^{q-1}}{\Gamma(q)} \right) \sum_{n \geq 0} (q)_k \sum_{n_1+n_2+n_3+n_4=n} \frac{r_{12}^{n_1} r_{13}^{n_2} r_{23}^{n_3} \rho^{n_4}}{n_1! n_2! n_3! n_4!} \prod_{i=1}^3 \frac{N_i!}{(q)_{N_i}} L_{N_i}^{(q-1)}(y_i) \mathbb{I}_{]0, \infty[^3}(y_1, y_2, y_3).$$

C.3 Loi négative multinomiale bivariée

Cette section est divisée en deux parties. Dans la première partie, on s'attache à déterminer une expression analytique des masses de probabilité de la loi négative multinomiale bivariée . Dans la seconde partie, les conditions des valeurs des paramètres pour lesquelles la loi multinomiale négative bivariée existe, sont déterminées. La seconde partie est étroitement liée aux calculs effectués dans la première partie.

C.3.1 Masses de probabilité d'une loi négative multinomiale bivariée

On cherche ici à déterminer la mesure de probabilité sur \mathbb{N}^2 du couple $(N_1, N_2) \sim \mathcal{NM}(q, P_{a,b,c})$ dont la fonction génératrice a pour expression :

$$G_{(N_1, N_2)}(z_1, z_2) = \left[\frac{(1-a)(1-b) - c}{1 - az_1 - bz_2 + (ab-c)z_1z_2} \right]^q.$$

D'après la définition A.1.1 de fonction génératrice, on a :

$$G(z_1, z_2) = \sum_{n_1, n_2 \geq 0} \Pr(N_1 = n_1, N_2 = n_2) z_1^{n_1} z_2^{n_2}, \quad (\text{C.14})$$

et cette série est finie pour tout $|z_1| \leq 1$ et $|z_2| \leq 1$. On cherche maintenant à identifier les masses de probabilités $\Pr(N_1 = n_1, N_2 = n_2)$ à partir de la décomposition en série entière de l'expression de la fonction génératrice donnée précédemment.

Soit $g(z_1, z_2) = \frac{cz_1z_2}{(1-az_1)(1-bz_2)}$. La fonction g est continue et $g(0,0) = 0$. Il existe donc un voisinage du point $(0,0)$ noté U_1 sur lequel $|g(z_1, z_2)| < 1$. De même, la fonction $h(z_1, z_2) = (az_1, bz_2)$ étant continue et nulle au point $(0,0)$, il existe un voisinage du point $(0,0)$ noté U_2 sur lequel $|az_1| < 1$ et $|bz_2| < 1$. Les ouverts U_1 et U_2 étant non vides et contenant le point $(0,0)$, leur intersection $U = U_1 \cap U_2$ est également un ouvert non vide qui contient le point $(0,0)$. Finalement pour tout $(z_1, z_2) \in U$, on peut développer en série entière la fonction génératrice grâce à la formule (C.1) de la façon suivante :

$$\begin{aligned} G(z_1, z_2) &= \left[\frac{(1-a)(1-b) - c}{1 - az_1 - bz_2 + (ab-c)z_1z_2} \right]^q = \left[\frac{(1-a)(1-b) - c}{(1-az_1)(1-bz_2)(1 - \frac{cz_1z_2}{(1-az_1)(1-bz_2)})} \right]^q, \\ &= \left[\frac{(1-a)(1-b) - c}{(1-az_1)(1-bz_2)} \right]^q \sum_{k=0}^{\infty} \frac{(q)_k}{k!} \frac{c^k z_1^k z_2^k}{(1-az_1)^k (1-bz_2)^k}, \\ &= ((1-a)(1-b) - c)^q \sum_{k=0}^{\infty} \frac{(q)_k}{k!} \frac{c^k z_1^k z_2^k}{(1-az_1)^{(q+k)} (1-bz_2)^{(q+k)}}. \end{aligned}$$

En utilisant maintenant les décompositions en séries entières suivantes :

$$\frac{1}{(1 - az_1)^{q+k}} = \sum_{r=0}^{\infty} \frac{(q+k)_r}{r!} a^r z_1^r, \quad \frac{1}{(1 - bz_2)^{q+k}} = \sum_{s=0}^{\infty} \frac{(q+k)_s}{s!} b^s z_2^s,$$

valides pour tout $(z_1, z_2) \in U$, on obtient l'expression :

$$\begin{aligned} G_{(N_1, N_2)}(z_1, z_2) &= ((1-a)(1-b) - c)^q \sum_{k,r,s=0}^{\infty} \frac{(q)_k}{k!} \frac{(q+k)_r}{r!} \frac{(q+k)_s}{s!} a^r b^s c^k z_1^{r+k} z_2^{s+k}, \\ &= ((1-a)(1-b) - c)^q \sum_{m,n=0}^{\infty} a^m b^n \left(\sum_{k=0}^{\min(m,n)} \frac{(q)_k}{k!} \frac{(q+k)_{m-k}}{(m-k)!} \frac{(q+k)_{n-k}}{(n-k)!} \left(\frac{c}{ab}\right)^k \right) z_1^m z_2^n, \end{aligned} \quad (\text{C.15})$$

pour tout $(z_1, z_2) \in U$.

La série entière obtenue est définie sur l'ouvert non vide U , et par unicité du développement en série entière, on peut identifier les coefficients de $z_1^m z_2^n$ obtenus avec les masses de probabilité $\Pr(N_1 = m, N_2 = n)$ du couple (N_1, N_2) selon (C.14). On obtient finalement :

$$\Pr(N_1 = m, N_2 = n) = a^m b^n [(1-a)(1-b) - c]^q \sum_{k=0}^{\min(m,n)} \frac{(q)_k}{k!} \frac{(q+k)_{m-k}}{(m-k)!} \frac{(q+k)_{n-k}}{(n-k)!} \left(\frac{c}{ab}\right)^k. \quad (\text{C.16})$$

C.3.2 Valeurs admissibles des paramètres d'une loi multinomiale négative bivariée

Considérons la loi multinomiale négative bivariée du couple $(N_1, N_2) \sim \mathcal{NM}(q, P_{a,b,c})$, dont la fonction génératrice a pour expression :

$$G_{(N_1, N_2)}(z_1, z_2) = \left[\frac{(1-a)(1-b) - c}{1 - az_1 - bz_2 + (ab-c)z_1 z_2} \right]^q.$$

Dans cette section, on détermine les conditions nécessaires et suffisantes sur les paramètres $(q, P_{a,b,c})$ pour lesquelles cette loi existe.

Remarquons tout d'abord que la fonction génératrice d'une mesure de probabilité ne peut être la fonction nulle. Ceci entraîne que $(1-a)(1-b) \neq 0$. Les autres conditions nécessaires peuvent être obtenues en remarquant que les masses de probabilité $p_{m,n} = \Pr(\mathbb{N}_1 = m, \mathbb{N}_2 = n)$ sont dans l'intervalle $[0, 1]$ pour tout $(m, n) \in \mathbb{N}^2$. Le cas où le paramètre $q > 0$ est un entier pair est considéré à la fin de la section. Pour l'instant on suppose que q n'est pas un entier pair. Selon l'expression C.16 de ces probabilités, trouvée à la section précédente, on a pour tout paramètre q différent d'un entier pair :

1. $p_{0,0} = [(1-a)(1-b) - c]^q$ et donc $(1-a)(1-b) - c \geq 0$. Mais comme $(1-a)(1-b) \neq 0$, finalement $(1-a)(1-b) - c > 0$,
2. $p_{1,0} = qa[(1-a)(1-b) - c]^q$ et donc $a \geq 0$. Par symétrie, $b \geq 0$,
3. $p_{1,n} = [(1-a)(1-b) - c]^q \frac{(q)_n}{n!} b^{n-1} (qab + nc)$ conduit à la condition $c \geq 0$. En effet, pour $c < 0$, la grandeur $qab + nc$ deviendrait négative pour des valeurs suffisamment grande de n ,
4. Puisque $p_{m,0} = ((1-a)(1-b) - c)^q a^m \frac{(q)_m}{m!}$, il vient $p_{m,0} > ((1-a)(1-b) - c)^q \frac{a^m}{m}$. Cette borne inférieure tend vers l'infini lorsque m tend vers l'infini si $a > 1$. De plus le cas $a = 1$ n'est pas possible, puisque $c \geq 0$ et que la condition $(1-a)(1-b) - c > 0$ doit être vérifiée. Par symétrie on a également $b < 1$.

Prouver que ces conditions sont suffisantes requiert de montrer que 1) les coefficients des monômes $z_1^m z_2^n$ dans le développement en série entière (C.15), notés $c_{m,n}$, sont positifs, et que 2) leur somme est égale à 1. Selon les conditions nécessaires ci-dessus il est évident que les coefficients $c_{m,n}$ sont positifs quelque soit $(m,n) \in \mathbb{N}^2$, même lorsque q est un entier pair. De plus ces conditions assurent que $\left| \frac{cz_1 z_2}{(1-az_1)(1-bz_2)} \right| < 1$, $|az_1| < 1$ et $|bz_2| < 1$ pour tout $-1 \leq z_1, z_2 \leq 1$. Par conséquent le développement en série entière effectué à la section précédente est valide pour tout $(z_1, z_2) \in [-1, 1]^2$. En particulier on obtient que $\sum_{m,n \geq 0} c_{m,n} = G_{\mathbb{N}}(1, 1) = 1$. Ces conditions sont donc suffisantes quelque soit le paramètre q .

Lorsque le paramètre de forme q est un entier pair, on peut montrer que les conditions énoncées ci-dessus sont toujours nécessaires. En effet les conditions 2) et 3) restent inchangées. On peut montrer que c'est également le cas de la condition 4). En effet, même lorsque q est un entier pair on a $a \leq 1$ et $b \leq 1$, seul pose problème la démonstration du fait que b (ou a) ne peut être égal à 1. Mais si q est un entier pair alors $q \geq 2$. Par conséquent la probabilité $p_{1,n} = [(1-a)(1-b) - c]^q \frac{(q)_n}{n!} b^{n-1} (qab + nc)$ tend vers l'infini si $b = 1$ quand n tend vers l'infini. Donc $b < 1$ et par symétrie $a < 1$. Enfin, on sait que les paramètres vérifient la contrainte $(1-a)(1-b) - c \neq 0$. Démontrons que le cas $c > (1-a)(1-b)$ est impossible. En effet les conditions précédentes assurent que les coefficients $c_{m,n}$ de la série entière sont positifs. Ces coefficients correspondant aux masses d'une loi de probabilité, on doit avoir :

$$S = \sum_{m,n \geq 0} c_{m,n} = 1.$$

Reprenons alors à l'envers le calcul des coefficients effectué à la section précédente :

$$\begin{aligned} S &= ((1-a)(1-b) - c)^q \sum_{m,n=0}^{\infty} a^m b^n \left(\sum_{k=0}^{\min(m,n)} \frac{(q)_k}{k!} \frac{(q+k)_{m-k}}{(m-k)!} \frac{(q+k)_{n-k}}{(n-k)!} \left(\frac{c}{ab} \right)^k \right), \\ &= ((1-a)(1-b) - c)^q \sum_{k,r,s=0}^{\infty} \frac{(q)_k}{k!} \frac{(q+k)_r}{r!} \frac{(q+k)_s}{s!} a^r b^s c^k, \end{aligned}$$

Les sommes peuvent être permutées car tous les termes sont positifs et comme $|a| < 1$ et $|b| < 1$, on obtient :

$$S = \left(\frac{(1-a)(1-b) - c}{(1-a)(1-b)} \right)^q \sum_{k \geq 0} \frac{(q)_k}{k!} \left(\frac{c}{(1-a)(1-b)} \right)^k.$$

Mais comme $c/((1-a)(1-b)) > 1$, la série de termes positifs est divergente et tend vers l'infini quel que soit $q > 0$, ce qui n'est pas possible car $S = 1$. Ceci prouve donc $(1-a)(1-b) - c > 0$. On a montré que les conditions 1) 2) 3) 4) ci-dessus sont des conditions nécessaires et suffisantes quelque soit le paramètre de forme q .

D

Estimation des paramètres des lois Gamma multivariées

Cette annexe regroupe les démonstrations des propriétés liées à l'estimation des paramètres des lois Gamma multivariées. La première partie est consacrée aux lois Gamma bivariées étudiées au chapitre 2, la seconde au modèle des lois Gamma bivariées "multi-capteurs" étudiées au chapitre 3. Les propriétés des estimateurs des lois Gamma multivariées et des lois multinomiales négatives, rencontrées en imagerie polarimétrique, et étudiées au chapitre 4, sont démontrées dans les deux dernières parties.

D.1 Estimateur du maximum de vraisemblance d'une loi Gamma bivariée

On rappelle ici que la fonction f_q , définie à l'équation (1.32), est la série entière, absolument convergente sur \mathbb{R} , suivante :

$$f_q(z) = \sum_{k \geq 0} \frac{z^k}{\Gamma(q+k)k!}. \quad (\text{D.1})$$

D.1.1 Dérivée de la log-vraisemblance

Soit $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^n)$ n vecteurs i.i.d. de \mathbb{R}^2 tels que $\mathbf{x}^i = (x_1^i, x_2^i)$ est distribué suivant une loi gamma bivariée de paramètre de forme q , de moyennes (m_1, m_2) et de coefficient de corrélation entre les marges r . Lorsque le paramètre de forme q est connu, il est montré à la section 2.2.1 que l'estimateur du maximum de vraisemblance du paramètre r est obtenu en minimisant la fonction :

$$h(r) = 2 \frac{nq}{(1-r)} - nq \log(1-r) + \sum_{i=1}^n \log f_q(\widehat{c} x_1^i x_2^i) \quad \forall r \in [0, 1[,$$

où :

$$\widehat{c} = \frac{r}{(1-r)^2} \frac{q^2}{\bar{x}_1 \bar{x}_2}.$$

Or la fonction f_q définie à l'équation (D.1) est une série entière définie sur \mathbb{R} . Par conséquent, pour tout $z \in \mathbb{R}$:

$$\begin{aligned} f'_q(z) &= \sum_{k \geq 0} \frac{1}{\Gamma(q+k)k!} \frac{dz^k}{dz} = \sum_{k \geq 1} \frac{z^{k-1}}{\Gamma(q+k)(k-1)!} = \sum_{k \geq 0} \frac{z^k}{\Gamma(q+1+k)k!} \\ &= f_{q+1}(z). \end{aligned}$$

On obtient alors l'expression suivante de la dérivée de la fonction h :

$$h'(r) = \frac{2nq}{(1-r)^2} - \frac{nq}{1-r} - \frac{1+r}{(1-r)^3} \frac{q^2}{\bar{x}_1 \bar{x}_2} \left(\sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(cx_1^i x_2^i)}{f_q(cx_1^i x_2^i)} \right).$$

Finalement en multipliant la fonction obtenue par le terme $\frac{1+r}{(1-r)^3 nq}$, qui est strictement positif pour tout $r \in [0, 1[$, on obtient que si la fonction h admet son minimum sur $]0, 1[$ alors ce minimum est une racine de la fonction suivante :

$$g(r) = r - 1 + \frac{q}{n\bar{x}_1 \bar{x}_2} \left(\sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(cx_1^i x_2^i)}{f_q(cx_1^i x_2^i)} \right). \quad (\text{D.2})$$

D.1.2 Log-concavité de la fonction f_q

Afin de montrer la log-concavité de la fonction f_q sur \mathbb{R}^+ , il est possible de démontrer directement par le calcul que la dérivée seconde de la fonction $z \mapsto \log(f_q(z))$ est négative sur \mathbb{R}^+ . On utilise cependant ici un résultat plus général démontré dans [AVV07] :

Soit une fonction f qui admet un développement en série entière de la forme :

$$f(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!},$$

dont le rayon de convergence R est non nul. On suppose que la fonction f est strictement positive sur un intervalle $I \subset]-R, R[$. Si la suite $(a_{n+1}/a_n)_{n \geq 0}$ est strictement croissante (respectivement décroissante) alors la fonction f est strictement log-convexe (respectivement log-concave) sur I .

Par conséquent, selon la définition (D.1) de la fonction f_q :

$$\frac{a_{n+1}}{a_n} = \frac{\Gamma(q+n)}{\Gamma(q+1+n)} = \frac{1}{q+n}.$$

La suite obtenue est donc strictement décroissante et la fonction f_q est alors strictement log-concave sur \mathbb{R}^+ pour tout $q > 0$.

D.2 Loi Gamma bivariée multi-capteurs

On rappelle ici que la fonction Φ_3 introduite à l'équation (D.3), qui est une des vingt séries confluentes de Horn, est la double série entière absolument convergente sur \mathbb{R}^2 suivante :

$$\Phi_3(a; b; x, y) = \sum_{m, n \geq 0} \frac{(a)_m}{(b)_{m+n} m! n!} x^m y^n. \quad (\text{D.3})$$

D.2.1 Densité

Selon la construction de loi Gamma bivariée multi-capteurs donnée à l'équation (3.5), on a $(Y_1, Y_2) = (X_1, X_2 + Z_2)$ où $Z_2 \sim \mathcal{G}(q_2 - q_1, p_2)$ et le couple (X_1, X_2) suit une loi Gamma bivariée de transformée de Laplace :

$$\mathcal{L}_{(X_1, X_2)}(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^{-q_1}.$$

Puisque la variable Z_2 est indépendante du couple (X_1, X_2) , la densité de (Y_1, Y_2) est donc obtenue par convolution de la densité de $Z_2 \sim \mathcal{G}(q_2 - q_1, p_2)$ et de la fonction $f_{(X_1, X_2)}(x_1, \cdot)$, où $f_{(X_1, X_2)}$ est la densité du couple (X_1, X_2) , dont une expression est donnée à l'équation (1.35). On obtient donc :

$$f_{(Y_1, Y_2)}(y_1, y_2) = \frac{y_1^{q_1-1}}{p_{12}^{q_1} p_2^{q_2-q_1}} \frac{e^{-\frac{p_2}{p_{12}} y_2}}{\Gamma(q_1) \Gamma(q_2 - q_1)} \int_0^{y_2} v^{q_1-1} e^{-\frac{p_1}{p_{12}} v} e^{-\frac{y_2-v}{p_2}} (y_2 - v)^{q_2-q_1-1} f_{q_1}(c y_1 v) dv. \quad (D.4)$$

Soit $A(\mathbf{y}) = \int_0^{y_2} v^{q_1-1} e^{-\frac{p_1}{p_{12}} v} e^{-\frac{y_2-v}{p_2}} (y_2 - v)^{q_2-q_1-1} f_{q_1}(c y_1 v) dv$. En utilisant la définition (D.1) de la fonction f_q et en effectuant le changement de variable $u = \frac{y_2-v}{y_2}$, on obtient :

$$A(\mathbf{y}) = e^{-\frac{p_1}{p_{12}} y_2} y_2^{q_2-1} \sum_{k \geq 0} \frac{c^k y_1^k y_2^k}{k! \Gamma(q_1 + k)} \int_0^1 e^{u y_2 d} u^{q_2-q_1-1} (1-u)^{q_1+k-1} du, \quad (D.5)$$

où $d = c \frac{p_{12}}{p_2}$.

Soit $B(y_2) = \int_0^1 e^{u y_2 d} u^{q_2-q_1-1} (1-u)^{q_1+k-1} du$. En utilisant le développement en série entière suivant :

$$e^{u y_2 d} = \sum_{l \geq 0} \frac{d^l u^l y_2^l}{l!},$$

on obtient :

$$\begin{aligned} B(y_2) &= \sum_{l \geq 0} \frac{d^l y_2^l}{l!} \int_0^1 u^{q_2-q_1+l-1} (1-u)^{q_1+k-1} du, \\ &= \sum_{l \geq 0} \frac{d^l y_2^l}{l!} \beta(q_2 - q_1 + l, q_1 + k), \\ &= \sum_{l \geq 0} \frac{d^l y_2^l}{l!} \frac{\Gamma(q_2 - q_1 + l) \Gamma(q_1 + k)}{\Gamma(q_2 + l + k)}, \end{aligned} \quad (D.6)$$

β étant la fonction définie pour $a > 0$ et $b > 0$ par $\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

En injectant maintenant (D.5) et (D.6) dans (D.4), il vient :

$$\begin{aligned} f_{(Y_1, Y_2)}(y_1, y_2) &= \frac{y_1^{q_1-1} y_2^{q_2-1}}{p_{12}^{q_1} p_2^{q_2-q_1}} \frac{e^{-\left[\frac{p_2}{p_{12}} y_1 + \frac{p_1}{p_{12}} y_2\right]}}{\Gamma(q_1)\Gamma(q_2-q_1)} \sum_{k,l \geq 0} \frac{c^k y_1^k y_2^{k+l} d^l}{k!l!} \frac{\Gamma(q_2 - q_1 + l)}{\Gamma(q_2 + l + k)}, \\ &= \frac{y_1^{q_1-1} y_2^{q_2-1}}{p_{12}^{q_1} p_2^{q_2-q_1}} \frac{e^{-\left[\frac{p_2}{p_{12}} y_1 + \frac{p_1}{p_{12}} y_2\right]}}{\Gamma(q_1)\Gamma(q_2)} \sum_{k,l \geq 0} \frac{(q_2 - q_1)_l}{(q_2)_{k+l} k!l!} (c y_1 y_2)^k (y_2 d)^l, \\ &= \frac{y_1^{q_1-1} y_2^{q_2-1}}{p_{12}^{q_1} p_2^{q_2-q_1}} \frac{e^{-\left[\frac{p_2}{p_{12}} y_1 + \frac{p_1}{p_{12}} y_2\right]}}{\Gamma(q_1)\Gamma(q_2)} \Phi_3(q_2 - q_1; q_2; c \frac{p_{12}}{p_2} y_2, c y_1 y_2), \end{aligned} \quad (D.7)$$

où Φ_3 est la fonction de Horn définie à l'équation (D.3).

D.2.2 Calcul numérique de la fonction de Horn Φ_3

Des représentations sous forme de séries, dont les termes s'expriment grâce à des fonctions spéciales, sont utiles pour calculer des séries hypergéométriques d'ordre 2 [CG04]. Dans le cas de la fonction de Horn Φ_3 définie en (D.3), les développements suivants en terme de fonctions spéciales sont particulièrement utiles :

$$\Phi_3(a; b; x, y) = \sum_{n=0}^{\infty} \frac{y^n}{(b)_n n!} {}_1F_1[a, b+n, x], \quad (D.8)$$

$$\Phi_3(a; b; x, y) = \Gamma(b) \sum_{m=0}^{\infty} (a)_m \frac{x^m}{m!} f_{b+m}(y) = \frac{\Gamma(b)}{y^{(b-1)/2}} \sum_{m=0}^{\infty} \frac{(a)_m}{m!} \left(\frac{x}{\sqrt{y}}\right)^m I_{b+m-1}(2\sqrt{y}), \quad (D.9)$$

où ${}_1F_1$ est la série hypergéométrique confluyente d'ordre 1, i.e. ${}_1F_1[a, b, x] = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n n!} x^n$ et $I_\nu(x)$ est la fonction de Bessel modifiée [AS64, pp. 374-377]. La série hypergéométrique confluyente ${}_1F_1[a, b, x]$ peut s'exprimer ainsi [Mul01] :

$${}_1F_1[a, b, x] = \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \sum_{i \geq 0} \frac{(b-a)_i (1-a)_i}{i! x^i} F_\gamma(x; i+b-a), \quad (D.10)$$

où $F_\gamma(x; \nu)$ est la fonction de répartition de la loi gamma univariée de paramètre de forme ν et de paramètre d'échelle 1.

Notons que lorsque le paramètre a est un entier non nul, la somme en (D.10) est finie, ce qui est le cas car $a \geq 1$ est un entier (qui correspond à un nombre de vues). On en déduit l'expression suivante de Φ_3 :

$$\Phi_3(a; b; x, y) = \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \sum_{n=0}^{\infty} \frac{(y/x)^n}{n!} \sum_{i=0}^a \frac{(b+n-a)_i (1-a)_i}{i! x^i} F_\gamma(x; i+b+n-a), \quad (D.11)$$

où la dernière sommation ($0 \leq i \leq a$) est finie. L'équation (D.11) fournit alors une façon numériquement stable de calculer $\Phi_3(a; b; x, y)$ pour de grandes valeurs de x et y . A contrario, lorsque le couple (x, y) est proche de $(0, 0)$, la définition de Φ_3 , donnée en (D.3), est utilisée pour le calcul numérique. On obtient finalement, l'algorithme de calcul suivant lorsque le paramètre a est un entier :

Algorithm 3 Calcul numérique de la fonction $\Phi_3(a; b; x, y)$ pour tout $x, y > 0$

- si (x, y) est proche de $(0, 0)$, calcul à partir de la définition (D.3)
 - si $y \gg 0$ et $\frac{x}{\sqrt{y}} < 1$, calcul à partir de l'équation (D.9) (il faut alors noter qu'au lieu de calculer la fonction de Bessel $I_\nu(z)$, on calcule la fonction classique $e^{-z}I_\nu(z)$, afin d'éviter les débordements lorsque $z \gg 0$)
 - sinon, si $a \in \mathbb{N}^*$, calcul à partir de la l'équation (D.11)
-

D.2.3 Dérivées de la fonction de Horn Φ_3

A partir de la représentation de la fonction Φ_3 en double série entière, donnée en (D.3), on obtient les résultats suivants :

$$\begin{aligned} \frac{\partial}{\partial x} \Phi_3(a; b; x, y) &= \sum_{m \geq 1, n \geq 0} \frac{(a)_m}{(b)_{m+n}(m-1)!n!} x^{m-1} y^n, \\ &= \frac{\Gamma(a+1)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+1)} \sum_{m, n \geq 0} \frac{(a+1)_m}{(b+1)_{m+n} m! n!} x^m y^n, \\ &= \frac{a}{b} \Phi_3(a+1; b+1; x, y). \end{aligned} \quad (\text{D.12})$$

$$\begin{aligned} \frac{\partial}{\partial y} \Phi_3(a; b; x, y) &= \sum_{m \geq 0, n \geq 1} \frac{(a)_m}{(b)_{m+n} m! (n-1)!} x^m y^{n-1}, \\ &= \frac{\Gamma(b)}{\Gamma(b+1)} \sum_{m, n \geq 0} \frac{(a)_m}{(b+1)_{m+n} m! n!} x^m y^n, \\ &= \frac{1}{b} \Phi_3(a; b+1; x, y). \end{aligned} \quad (\text{D.13})$$

D.2.4 Estimateurs du maximum de vraisemblance

Le gradient de la log-vraisemblance exprimée à l'équation (3.14) par rapport au vecteur $\theta = (m_1, m_2, r')$, a pour expression (les dérivées partielles de la fonction de Horn Φ_3 sont données à l'annexe ci-dessus (D.2.3)) :

$$\frac{\partial l(\mathbf{Y}; \theta)}{\partial m_1} = -\frac{nq_1}{m_1} + \frac{nq_1}{m_1^2(1-r')} \bar{Y}_1 - \frac{r'}{m_1} \Delta_2, \quad (\text{D.14})$$

$$\frac{\partial l(\mathbf{Y}; \theta)}{\partial m_2} = -\frac{nq_2}{m_2} + \frac{nq_2}{m_2^2(1-r')} \bar{Y}_2 - \frac{r'}{m_2} \Delta_1 - \frac{r'}{m_2} \Delta_2, \quad (\text{D.15})$$

$$\frac{\partial l(\mathbf{Y}; \theta)}{\partial r'} = \frac{nq_1}{1-r'} - \frac{nq_1}{m_1(1-r')^2} \bar{Y}_1 - \frac{nq_2}{m_2(1-r')^2} \bar{Y}_2 + \frac{1}{1-r'} \Delta_1 + \frac{1+r'}{1-r'} \Delta_2, \quad (\text{D.16})$$

où :

$$\begin{aligned} \Delta_1 &= \frac{q_2 - q_1}{m_2(1-r')} \sum_{i=1}^n Y_2^i \frac{\Phi_3(q_2 - q_1 + 1; q_2 + 1; dY_2^i; cY_1^i Y_2^i)}{\Phi_3(q_2 - q_1; q_2; dY_2^i; cY_1^i Y_2^i)}, \\ \Delta_2 &= \frac{q_1}{m_1 m_2 (1-r')^2} \sum_{i=1}^n Y_1^i Y_2^i \frac{\Phi_3(q_2 - q_1; q_2 + 1; dY_2^i; cY_1^i Y_2^i)}{\Phi_3(q_2 - q_1; q_2; dY_2^i; cY_1^i Y_2^i)}. \end{aligned}$$

Comme cela est expliqué à la section 3.2.1, les EMV des paramètres du vecteur θ sont les valeurs qui annulent le gradient $\mathbf{u}(\mathbf{Y}; \theta) = 0$. En multipliant (D.14) par $m_1 > 0$, (D.15) par $m_2 > 0$ et (D.16) par $1 - r' > 0$, l'EMV de $\theta = (m_1, m_2, r') \in]0, +\infty[^2 \times]0, 1[$ est la solution du système :

$$-nq_1 + \frac{nq_1}{m_1(1-r')} \bar{Y}_1 - r' \Delta_2 = 0, \quad (\text{D.17})$$

$$-nq_2 + \frac{nq_2}{m_2(1-r')} \bar{Y}_2 - r'(\Delta_1 + \Delta_2) = 0, \quad (\text{D.18})$$

$$nq_1 - \frac{nq_1}{m_1(1-r')} \bar{Y}_1 - \frac{nq_2}{m_2(1-r')} \bar{Y}_2 + \Delta_1 + (1+r')\Delta_2 = 0, \quad (\text{D.19})$$

En sommant ces trois équations (i.e., (D.17)+(D.18)+(D.19)) il vient :

$$\Delta_1 + \Delta_2 = \frac{nq_2}{1-r'}. \quad (\text{D.20})$$

En remplaçant maintenant l'expression obtenue de $\Delta_1 + \Delta_2$ dans (D.18), on trouve :

$$\frac{nq_2}{1-r'} \frac{\bar{Y}_2}{m_2} - \frac{nq_2}{1-r'} = 0,$$

et donc :

$$m_2 = \bar{Y}_2.$$

D.3 Lois Gamma multivariées dont le polynôme affine est quadratique

On s'intéresse dans cette section à l'étude des estimateurs des lois Gamma multivariées sur \mathbb{R}^d dont la transformée de Laplace est de la forme :

$$\mathcal{L}_{\mathcal{G}(q,p)}(\mathbf{z}) = [P(\mathbf{z})]^{-q},$$

où P est le polynôme affine de la forme :

$$P(\mathbf{z}) = 1 + \sum_{i=1}^d p_i z_i + \sum_{1 \leq i < j \leq d} p_{ij} z_i z_j.$$

Lorsque $d = 2$, on retrouve alors la définition de la loi de Kibble. En dimension $d = 4$, ce modèle correspond à la loi du vecteur des intensités $\mathbf{I} = (I_1, I_2, I_3, I_4)$ associées à des images polarimétriques, introduite à la section 4.2.1 et dont la transformée de Laplace est donnée à l'équation (4.12), p. 113.

D.3.1 Inverse d'une matrice appartenant à l'espace vectoriel H

Afin de démontrer que la loi du vecteur $\mathbf{I} = (I_1, I_2, I_3, I_4)$ décrit une famille exponentielle, on cherche à démontrer (voir section 4.2.2, p. 114) que si une matrice symétrique et inversible est dans l'espace vectoriel H , alors son inverse aussi. On rappelle que H est l'espace vectoriel de dimension 4 des matrices réelles M de dimension 4×4 , et de la forme :

$$\begin{pmatrix} A & B \\ B^T & A \end{pmatrix},$$

où A et B sont des matrices réelles 2×2 respectivement symétrique et antisymétrique. La matrice M étant symétrique, son inverse l'est aussi. Cet inverse peut donc s'écrire :

$$M^{-1} = \begin{pmatrix} C & E \\ E^T & D \end{pmatrix},$$

où C, D et E sont des matrices réelles 2×2 et C, D des matrices symétriques. On suppose ici que la matrice A est inversible, ce qui correspond au cas étudié, puisque avec les notations introduites à la section 4.2.2, p. 114 :

$$A = \begin{pmatrix} a_1 & a_3 \\ a_3 & a_2 \end{pmatrix},$$

avec $a_3^2 + a_4^2 \leq a_1 a_2$. La matrice A n'est donc pas inversible si $a_3^2 = a_1 a_2$. Mais dans ce cas $a_4 = 0$, donc la matrice $B \in \mathbb{R}^{2 \times 2}$ est nulle, et donc la matrice $M = C_U$ n'est pas inversible. La propriété reste cependant vraie dans le cas général où la matrice $A \in \mathbb{R}^{2 \times 2}$ n'est pas inversible du moment que la matrice $M \in \mathbb{R}^{4 \times 4}$ l'est (la démonstration est laissée aux soins du lecteur). On note dans la suite respectivement I_4, I_2 et 0_2 les matrices identités de $\mathbb{R}^{4 \times 4}$, de $\mathbb{R}^{2 \times 2}$ et la matrice nulle de $\mathbb{R}^{2 \times 2}$. La relation $MM^{-1} = I_4$ nous donne le système suivant :

$$AC + BE^T = I_2, \tag{D.21}$$

$$-BE + AD = I_2, \tag{D.22}$$

$$-BC + AE^T = 0_2, \tag{D.23}$$

$$AE + BD = 0_2. \tag{D.24}$$

Les équations (D.23) et (D.24) montrent que :

$$E^T = A^{-1}BC,$$

$$E = -A^{-1}BD.$$

En remplaçant dans (D.21) et (D.22) les matrices E et E^T par les expressions trouvées ci-dessus, il vient :

$$(A + BA^{-1}B)C = I_2,$$

$$(A + BA^{-1}B)D = I_2.$$

Par conséquent les matrices C et D sont inversibles et égales. Mais alors en soustrayant les équations (D.23) et (D.24), on obtient :

$$A(E + E^T) = 0_2.$$

La matrice A étant inversible, on a $E^T + E = 0_2$, donc E est antisymétrique. Ceci montre que la matrice M^{-1} est dans H .

D.3.2 Estimateurs du maximum de vraisemblance pour le modèle des trois images d'intensité polarimétriques

Dans le cas de trois images d'intensité, la densité présentée à l'équation (4.23), p. 118, paramétrée par le vecteur $(a_1, a_2, a_3, a_4)^T \in A = \{a_1, a_2, a_3, a_4 \in \mathbb{R} \mid a_1 a_2 > a_3^2 + a_4^2\}$, peut être

reparamétrisée par le vecteur :

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, c)^T = \left(-\frac{a_2 + a_3}{2} \frac{c}{|a_4|}, -\frac{a_1 + a_3}{2} \frac{c}{|a_4|}, a_3 \frac{c}{|a_4|}, c \right)^T,$$

où $c = \frac{|a_4|}{a_1 a_2 - a_3^2 - a_4^2}$. En effet en inversant le système ainsi défini, on trouve l'expression suivante des paramètres a_1, a_2, a_3, a_4^2 en fonction de $\boldsymbol{\theta}$:

$$\begin{aligned} a_1 &= \frac{|a_4|}{c} (\theta_3 - 2\theta_2), & a_2 &= \frac{|a_4|}{c} (\theta_3 - 2\theta_1), & a_3 &= -\frac{|a_4|}{c} \theta_3, \\ |a_4| &= \frac{c}{4\theta_1\theta_2 - 2(\theta_1\theta_3 + \theta_2\theta_3) - c^2}. \end{aligned}$$

On note Θ l'ensemble ouvert des valeurs admissibles des paramètres $\boldsymbol{\theta}$ (qui est l'image de l'ensemble ouvert A par la fonction injective $(a_1, a_2, a_3, a_4^2)^T \in A \mapsto \boldsymbol{\theta}$). Pour cette nouvelle paramétrisation et avec les notations introduites à l'équation (4.23), la densité de la loi du vecteur $\widetilde{\mathbf{I}} = (I_1, I_2, I_3)^T$ a pour expression :

$$p(\widetilde{\mathbf{I}}) = e^{(\theta_1 I_1 + \theta_2 I_2 + \theta_3 I_3)} g(\boldsymbol{\theta}) \left(\pi v'(\widetilde{\mathbf{I}}) \right)^{-\frac{1}{2}} f_{\frac{1}{2}} \left(c^2 v'(\widetilde{\mathbf{I}}) / 16 \right) \mathbb{I}_{\Omega}(\widetilde{\mathbf{I}}), \quad (\text{D.25})$$

où $g(\boldsymbol{\theta}) = 4\theta_1\theta_2 - 2(\theta_1\theta_3 + \theta_2\theta_3) - c^2$. D'après la définition d'une famille exponentielle naturelle donnée à la section B.1, p. 157, lorsque le paramètre c est fixé, la loi de $\widetilde{\mathbf{I}}$ décrit une FEN dont une mesure génératrice est :

$$\mu_c(d\widetilde{\mathbf{I}}) = \left(\pi v'(\widetilde{\mathbf{I}}) \right)^{-\frac{1}{2}} f_{\frac{1}{2}} \left(c^2 v'(\widetilde{\mathbf{I}}) / 16 \right) \mathbb{I}_{\Omega}(\widetilde{\mathbf{I}}) d\widetilde{\mathbf{I}},$$

et dont les paramètres associés $(\theta_1, \theta_2, \theta_3)^T$ sont dans l'ensemble ouvert :

$$\widetilde{\Theta} = \left\{ (\theta_1, \theta_2, \theta_3)^T \mid (\theta_1, \theta_2, \theta_3, c)^T \in \Theta \right\}.$$

D'après le théorème B.6.3, les EMV des moyennes m_1, m_2 et m_3 sont les moyennes empiriques, notées respectivement $\widehat{m}_1 = \sum_{j=1}^n I_1^{(j)} / n$, $\widehat{m}_2 = \sum_{j=1}^n I_2^{(j)} / n$ et $\widehat{m}_3 = \sum_{j=3}^n I_3^{(j)} / n$. Mais si c est inconnu, la vraisemblance exprimée par rapport au vecteur $(m_1, m_2, m_3, c)^T$ est nécessairement maximum sur la droite de \mathbb{R}^4 , d'équation paramétrique $(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, c)$. Ceci montre que les EMV des moyennes de $\widetilde{\mathbf{I}}$ sont les moyennes empiriques. D'après le principe d'invariance fonctionnelle, on trouve alors l'expression donnée à l'équation (4.24) des EMV de a_1, a_2 et a_3 .

L'EMV du paramètre a_4^2 s'obtient finalement en maximisant la log-vraisemblance de $\widetilde{\mathbf{I}}$ lorsque a_1, a_2 et a_3 sont remplacés par leur EMV respectif. Après avoir éliminé les termes qui ne dépendent pas de $\mathbf{a} = (a_1, a_2, a_3, a_4^2)^T$, la log-vraisemblance s'écrit :

$$\begin{aligned} l_3(\widetilde{\mathbf{I}}^{(n)}; \mathbf{a}) &= - \sum_{j=1}^n \left[\frac{\mathbf{c}'(\mathbf{a})^T \widetilde{\mathbf{I}}^j}{2k} - \log f_{\frac{1}{2}} \left(\frac{a_4^2 v'(\widetilde{\mathbf{I}}^j)}{16k^2} \right) \right] - n \log(k), \\ &= - \frac{n}{2k} [(a_2 + a_3)\widehat{m}_1 + (a_1 + a_3)\widehat{m}_2 - 2a_3\widehat{m}_3] + \sum_{j=1}^n \log f_{\frac{1}{2}} \left(\frac{a_4^2 v'(\widetilde{\mathbf{I}}^j)}{16k^2} \right) - n \log(k). \end{aligned}$$

En dérivant cette log-vraisemblance par rapport au paramètre a_4^2 , on obtient la fonction score suivante :

$$g_3(\widehat{\mathbf{I}}^{(n)}; \mathbf{a}) = \frac{n}{4k^2} [2k - (a_2 + a_3)\widehat{m}_1 - (a_1 + a_3)\widehat{m}_2 + 2a_3\widehat{m}_3] \\ + \frac{(a_1a_2 - a_3^2 + a_4^2)}{32k^3} \sum_{j=1}^n v'(\widehat{\mathbf{I}}^j) \frac{f_{\frac{3}{2}}\left(\frac{a_4^2 v'(\widehat{\mathbf{I}}^j)}{16k^2}\right)}{f_{\frac{1}{2}}\left(\frac{a_4^2 v'(\widehat{\mathbf{I}}^j)}{16k^2}\right)}.$$

Après avoir multiplié l'expression ci-dessus par le facteur strictement négatif $-32k^3/[n(a_1a_2 - a_3^2 + a_4^2)]$, et en remplaçant les paramètres a_1 , a_2 et a_3 par leur EMV \widehat{a}_1 , \widehat{a}_2 et \widehat{a}_3 , on trouve que l'EMV du paramètre a_4^2 est la racine de la fonction donnée à l'équation (4.25).

D.4 Propriétés des estimateurs du maximum de vraisemblance composite des lois multinomiales négatives

Afin de montrer la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance composite, on peut utiliser des résultats plus généraux sur les estimations par minimum de contraste. On rappelle que l'estimateur $\widehat{\boldsymbol{\theta}}_n$ est la valeur pour laquelle le critère $U_n(\boldsymbol{\theta})$ donné à l'équation (4.33), p. 131, est minimum. D'après la loi faible des grands nombres, lorsque n tend vers l'infini, $U_n(\boldsymbol{\theta})$ converge en $P_{\boldsymbol{\theta}_0}$ -probabilité vers :

$$K(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = - \sum_{1 \leq k < l \leq d} \int \log(p_{k,l}(n_k, n_l, \boldsymbol{\theta})) p_{k,l}(n_k, n_l, \boldsymbol{\theta}_0) d\mu(n_k, n_l),$$

où μ est la mesure de comptage. Lorsque la fonction $\boldsymbol{\theta} \rightarrow K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ (de Θ dans \mathbb{R}^+) possède un unique minimum $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, U_n définit alors un contraste pour $\boldsymbol{\theta}_0$ et K . Par conséquent, $\widehat{\boldsymbol{\theta}}_n$ est appelé l'estimateur du minimum de contraste (voir [DCD84, p. 92]). Remarquons que minimiser $K(\cdot, \boldsymbol{\theta}_0)$ est équivalent à minimiser :

$$\sum_{1 \leq k < l \leq d} \int \log\left(\frac{p_{k,l}(n_k, n_l, \boldsymbol{\theta}_0)}{p_{k,l}(n_k, n_l, \boldsymbol{\theta})}\right) p_{k,l}(n_k, n_l, \boldsymbol{\theta}_0) d\mu(n_k, n_l).$$

Les propriétés de la divergence de Kullback-Leiber assurent donc que $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ est minimum pour $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. De plus, ce minimum est unique si et seulement si :

$$\mathbf{A}_0 : \quad \forall k < l; \quad p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}) = p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}_0) \quad \text{presque partout (p.p.)} \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

Consistance de l'estimateur du minimum de contraste.

Pour prouver la consistance de l'estimateur du minimum de contraste, il suffit de démontrer que les conditions suivantes sont vérifiées (voir [DCD84, p. 93]) :

\mathbf{A}_1 : Θ est un sous espace compact de \mathbb{R}^p . Les fonctions $U_n(\boldsymbol{\theta})$ et $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ sont continues pour $\boldsymbol{\theta} \in \Theta$.

A₂ : Pour tout $\eta > 0$, soit $w(n, \eta) = \sup\{|U_n(\alpha) - U_n(\beta)|; \|\alpha - \beta\| \leq \eta\}$, où $\|\cdot\|$ est la norme euclidienne. Il existe une suite $(\varepsilon_K)_{K \in \mathbb{N}}$, décroissante vers 0 telle que pour tout K :

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(w \left(n, \frac{1}{K} \right) \geq \varepsilon_K \right) = 0.$$

Normalité asymptotique de l'estimateur du minimum de contraste.

Les conditions supplémentaires suivantes sont suffisantes pour montrer la normalité asymptotique :

A₃ : Le point θ_0 appartient à l'intérieur de l'espace Θ . La fonction $U_n(\theta)$ est deux fois continue différentiable sur un voisinage V de θ_0 .

A₄ : $\sqrt{n} \nabla U_n(\theta_0)$ converge en loi vers la loi normale centrée de matrice de covariance $\Gamma_U(\theta_0)$.

A₅ : Pour tout $r > 0$ et $u, v = 1, \dots, p$,

$$1_{|\hat{\theta}_n - \theta_0| \leq r} \left[\int_0^1 \frac{\partial^2}{\partial \theta_u \partial \theta_v} U_n(\theta_0 + s(\hat{\theta}_n - \theta_0)) ds - \frac{\partial^2}{\partial \theta_u \partial \theta_v} U_n(\theta_0) \right]$$

converge en P_{θ_0} -probabilité vers zéro.

A₆ : Il existe une matrice inversible $I_U(\theta_0)$ telle que $\left(\frac{\partial^2}{\partial \theta_u \partial \theta_v} U_n(\theta_0) \right)_{u,v=1,\dots,p}$ converge en P_{θ_0} -probabilité vers $I_U(\theta_0)$.

Sous les hypothèses **A_{3,6}** et si l'estimateur du minimum de contraste est consistant, alors $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converge en loi vers la loi normale centrée de matrice de covariance :

$$I_U(\theta_0)^{-1} \Gamma_U(\theta_0) I_U(\theta_0)^{-1}$$

(voir [DCD84, p. 104]).

Pour les contrastes de la forme (4.33), tout comme pour les estimateurs du maximum de vraisemblance, on peut remplacer **A₄** et **A₆** par :

Ã₄ : Dérivation et intégration du terme $p_{k,l}$ par rapport à la mesure μ peuvent être permutées. La matrice de covariance du vecteur aléatoire $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(N_k, N_l, \theta_0)$ existe.

Ã₅ : Il existe des fonctions h_{kl} dans l'espace $L^1(P_{\theta_0})$ telles que pour tout $\theta \in V$ et $u, v = 1, \dots, p$,

$$\left| \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(n_k, n_l, \theta) \right| \leq h_{k,l}(n_k, n_l), \quad \forall (n_k, n_l) \in \mathbb{N}^2.$$

Propriétés de l'estimateur du maximum de vraisemblance composite

Afin de prouver le théorème 4.3.1, p. 132, nous devons montrer que les conditions **A₀**, **A₁**, **A₂**, **A₃**, **Ã₄**, **Ã₅** et **A₆** sont vérifiées pour le modèle étudié sous les hypothèses 1, 2 et 3. On rappelle que pour ce modèle, nous avons (voir théorème 1.3.8) :

$$p_{k,l}(n_k, n_l, \theta) = a_{k,l}(\theta)^{n_k} b_{k,l}(\theta)^{n_l} \left((1 - a_{k,l}(\theta))(1 - b_{k,l}(\theta)) - c_{k,l}(\theta) \right)^L \sum_{j=0}^{\min(n_k, n_l)} C_{L,j}^{n_k, n_l} \left(\frac{c_{k,l}(\theta)}{a_{k,l}(\theta) b_{k,l}(\theta)} \right)^j,$$

où :

$$C_{L,j}^{n_k, n_l} = \frac{(L)_j (L+j)_{n_k-j} (L+j)_{n_l-j}}{j! (n_k-j)! (n_l-j)!}.$$

Pour tout $k < l$, $F_{k,l}(\boldsymbol{\theta}) = (a_{k,l}(\boldsymbol{\theta}), b_{k,l}(\boldsymbol{\theta}), c_{k,l}(\boldsymbol{\theta}))^T$ où $F_{k,l}$ sont des fonctions de Θ vers $\Delta = \{(a_{k,l}, b_{k,l}, c_{k,l}) \in [0, 1]^3; (1 - a_{k,l})(1 - b_{k,l}) > c_{k,l}\}$, et $F(\boldsymbol{\theta}) = (F_{1,2}(\boldsymbol{\theta})^T, \dots, F_{d-1,d}(\boldsymbol{\theta})^T)^T$ est une fonction injective de Θ vers $\Delta^{p(p-1)/2}$ (hypothèse 2). De plus, les fonctions $F_{k,l}$ sont deux fois continues et différentiables (hypothèse 3).

A₀ : Puisque pour tout $k < l$, $p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}) = p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}_0)$ p.p., nous avons en particulier pour $(n_k, n_l) = (0, 0)$,

$$\left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right)^L = \left((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0) \right)^L,$$

pour $(n_k, n_l) = (1, 0)$,

$$a_{k,l}(\boldsymbol{\theta}) \left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right)^L = a_{k,l}(\boldsymbol{\theta}_0) \left((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0) \right)^L,$$

et pour $(n_k, n_l) = (0, 1)$,

$$b_{k,l}(\boldsymbol{\theta}) \left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right)^L = b_{k,l}(\boldsymbol{\theta}_0) \left((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0) \right)^L.$$

Aussi $a_{k,l}(\boldsymbol{\theta}) = a_{k,l}(\boldsymbol{\theta}_0)$, $b_{k,l}(\boldsymbol{\theta}) = b_{k,l}(\boldsymbol{\theta}_0)$ et $c_{k,l}(\boldsymbol{\theta}) = c_{k,l}(\boldsymbol{\theta}_0)$, i.e. $F_{k,l}(\boldsymbol{\theta}) = F_{k,l}(\boldsymbol{\theta}_0)$ pour tout $1 \leq k < l \leq d$. Donc $F(\boldsymbol{\theta}) = F(\boldsymbol{\theta}_0)$, ce qui implique que $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ puisque F est une fonction injective.

A₁ : Selon l'hypothèse 1, Θ est un compact de \mathbb{R}^p . La fonction $U_n(\boldsymbol{\theta})$ est continue (en tant que somme de fonctions continues) pour tout $\boldsymbol{\theta} \in \Theta$. En ce qui concerne la fonction $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$, on peut appliquer le théorème concernant les intégrales définies par un paramètre (corollaire du théorème de convergence dominée de Lebesgue). En notant :

$$A = \frac{a_{k,l}(\boldsymbol{\theta})^{n_k} b_{k,l}(\boldsymbol{\theta})^{n_l} \left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right)^L}{(1 + n_k)(1 + n_l)},$$

$$B = (1 + n_k)(1 + n_l) \sum_{j=0}^{\min(n_k, n_l)} C_{L,j}^{n_k, n_l} \left(\frac{c_{k,l}(\boldsymbol{\theta})}{a_{k,l}(\boldsymbol{\theta}) b_{k,l}(\boldsymbol{\theta})} \right)^j,$$

on obtient $p_{k,l}(n_k, n_l, \boldsymbol{\theta}) = AB$ ce qui donne $0 \leq AB \leq 1$. En raison de la contrainte sur $(a_{k,l}, b_{k,l}, c_{k,l})$ (hypothèse 2) et du fait que $(1 + n_k)(1 + n_l) C_{L,0}^{n_k, n_l} = (1 + n_k) \frac{(L)_{n_k}}{n_k!} (1 + n_l) \frac{(L)_{n_l}}{n_l!} > 1$ pour tout $L > 0$, nous avons $A < 1$ et $B > 1$. Par conséquent, $\log(A) \leq \log(AB) \leq 0$, donc $|\log(AB)| \leq |\log(A)|$, ce qui entraîne :

$$\begin{aligned} |\log(p_{k,l}(n_k, n_l, \boldsymbol{\theta}))| &\leq n_k |\log(a_{k,l}(\boldsymbol{\theta}))| + n_l |\log(b_{k,l}(\boldsymbol{\theta}))| + L \left| \log \left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right) \right| \\ &\quad + \log(1 + n_k) + \log(1 + n_l), \\ &\leq n_k \left(1 + |\log(a_{k,l}(\boldsymbol{\theta}))| \right) + n_l \left(1 + |\log(b_{k,l}(\boldsymbol{\theta}))| \right) + \\ &\quad L \left| \log \left((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}) \right) \right|. \end{aligned}$$

Les fonctions $F_{k,l}$ étant uniformément continues (en tant que fonction continues sur un compact) il vient :

$$|\log(p_{k,l}(n_k, n_l, \boldsymbol{\theta}))| \leq C_1 n_k + C_2 n_l + C_3,$$

où C_1 , C_2 et C_3 sont des constantes positives. La fonction dominée est donc P_{θ_0} -intégrable puisque tous les moments de la variable N_k , $k = 1, \dots, d$, existent. En utilisant la continuité de la fonction $p_{k,l}(n_k, n_l, \theta)$ pour $\theta \in \Theta$, on peut conclure que $K(\theta_0, \theta)$ est continue pour $\theta \in \Theta$.

A₂ : En notant $p_{k,l}(n_k, n_l, \theta) = g_{k,l}(\theta)\tilde{p}_{k,l}(n_k, n_l, \theta)$, où

$$g_{k,l}(\theta) = \left((1 - a_{k,l}(\theta))(1 - b_{k,l}(\theta)) - c_{k,l}(\theta) \right)^L,$$

$$\tilde{p}_{k,l}(n_k, n_l, \theta) = \sum_{j=0}^{\min(n_k, n_l)} C_{L,j}^{n_k, n_l} a_{k,l}(\theta)^{n_k-j} b_{k,l}(\theta)^{n_l-j} c_{k,l}(\theta)^j,$$

on obtient :

$$|U_n(\alpha) - U_n(\beta)| \leq \underbrace{\sum_{1 \leq k < l \leq d} \left| \log \left(\frac{g_{k,l}(\alpha)}{g_{k,l}(\beta)} \right) \right|}_{P_1} + \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} \left| \log \left(\frac{\tilde{p}_{k,l}(N_k^i, N_l^i, \alpha)}{\tilde{p}_{k,l}(N_k^i, N_l^i, \beta)} \right) \right|}_{P_2}.$$

Le premier terme P_1 est composé de fonctions continues sur le compact Θ et est donc uniformément continu. Donc, pour $\|\alpha - \beta\| \leq \frac{1}{K}$, il existe ε_K^1 telle que $P_1 \leq \varepsilon_K^1$, où ε_K^1 est une suite qui décroît vers zéro lorsque K tend vers l'infini. Concernant le second terme P_2 , on a :

$$P_2 \leq \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} \sup_{\theta \in \Theta} \left\| \nabla \log \left(\tilde{p}_{k,l}(N_k^i, N_l^i, \theta) \right) \right\| \|\alpha - \beta\|.$$

Comme :

$$\frac{\partial}{\partial a_{k,l}} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta) \leq \frac{N_k^i}{a_{k,l}(\theta)} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta),$$

$$\frac{\partial}{\partial b_{k,l}} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta) \leq \frac{N_l^i}{b_{k,l}(\theta)} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta),$$

$$\frac{\partial}{\partial c_{k,l}} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta) \leq \frac{N_k^i + N_l^i}{c_{k,l}(\theta)} \tilde{p}_{k,l}(N_k^i, N_l^i, \theta),$$

on obtient ($F_{k,l}$ est continûment différentiable sur un compact) pour $u = 1, \dots, p$,

$$\left| \frac{\partial}{\partial \theta_u} \log \left(\tilde{p}_{k,l}(N_k^i, N_l^i, \theta) \right) \right| \leq \frac{\partial a_{k,l}}{\partial \theta_u}(\theta) \frac{N_k^i}{a_{k,l}(\theta)} + \frac{\partial b_{k,l}}{\partial \theta_u}(\theta) \frac{N_l^i}{b_{k,l}(\theta)} + \frac{\partial c_{k,l}}{\partial \theta_u}(\theta) \frac{N_k^i + N_l^i}{c_{k,l}(\theta)}$$

$$\leq C_{k,l}(N_k^i + N_l^i),$$

où $C_{k,l}$ est une constante positive. En notant C le maximum des constantes $C_{k,l}$, $1 \leq k < l \leq d$, on obtient :

$$P_2 \leq \sqrt{p}C \|\alpha - \beta\| \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} (N_k^i + N_l^i)$$

$$\leq \frac{\sqrt{p}C}{K} \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} (N_k^i + N_l^i)$$

$$= \frac{1}{K} W_n.$$

D'après la loi faible des grands nombres, lorsque n tend vers l'infini, W_n converge en P_{θ_0} -probabilité vers $l_W = \sqrt{p} C \sum_{1 \leq k < l \leq d} \mathbf{E}(N_k + N_l) < \infty$. Soit $\varepsilon_K = \varepsilon_K^1 + \frac{2}{K} l_W$, cette suite tend vers zéro lorsque K tend vers l'infini. Finalement, puisque $w(n, 1/K) \leq P_1 + P_2 \leq \varepsilon_K^1 + \frac{1}{K} W_n$:

$$P_{\theta_0} \left(w \left(n, \frac{1}{K} \right) \geq \varepsilon_K \right) \leq \Pr_{\theta_0} (W_n - l_W > l_W),$$

converge vers zéro lorsque n tend vers l'infini puisque $l_W > 0$.

A₃ : L'hypothèse 1 implique que le point θ_0 appartient à l'intérieur du compact Θ . La fonction $U_n(\theta)$ est deux fois continue et différentiable sur Θ en tant que somme de fonctions deux fois continûment différentiables.

Ã₄ : La preuve que la dérivation et l'intégration par rapport à μ peuvent être permutées pour le terme $p_{k,l}$, repose sur les propriétés de différentiabilité des intégrales définies par un paramètre. En reprenant les calculs effectués pour **A₂**, on peut dominer le dérivées partielles de $p_{k,l}$:

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_u} p_{k,l}(n_k, n_l, \theta) \right| &= \left| \frac{\partial}{\partial \theta_u} g_{k,l}(\theta) \tilde{p}_{k,l}(n_k, n_l, \theta) + g_{k,l}(\theta) \frac{\partial}{\partial \theta_u} \tilde{p}_{k,l}(n_k, n_l, \theta) \right|, \\ &\leq (C_1 + C_2(n_k + n_l)) \tilde{p}_{k,l}(n_k, n_l, \theta^*), \end{aligned}$$

où $u = 1, \dots, p$, C_1 et C_2 sont des constantes positives et θ^* l'argument du maximum de la fonction continue $\tilde{p}_{k,l}$ définie sur le compact Θ . La fonction dominée est donc μ -intégrable. Comme $p_{k,l}$ est différentiable, dérivation et intégration par rapport à μ peuvent être permutées pour $p_{k,l}$. En particulier, ceci implique que le vecteur aléatoire $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(N_k, N_l, \theta_0)$ est centré. Afin de prouver l'existence de la matrice de covariance, on peut montrer que pour tout $u, v = 1, \dots, p$ et pour tout $k < l$:

$$\mathbf{E}_{\theta_0} \left(\left| \frac{\partial}{\partial \theta_u} \log p_{k,l}(N_k, N_l, \theta_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(N_k, N_l, \theta_0) \right| \right) < \infty.$$

Comme ci-dessus, il existe des constantes positives C_1 , C_2 et C_3 telles que :

$$\left| \frac{\partial}{\partial \theta_u} \log p_{k,l}(N_k, N_l, \theta_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(N_k, N_l, \theta_0) \right| \leq C_1 + C_2(N_k + N_l) + C_3(N_k + N_l)^2,$$

qui est fini puisque tous les moments des variables N_k , $k = 1, \dots, d$, existent.

Ã₅ : On a :

$$\left| \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(n_k, n_l, \theta) \right| \leq \left| \frac{\frac{\partial^2}{\partial \theta_u \partial \theta_v} p_{k,l}(n_k, n_l, \theta)}{p_{k,l}(n_k, n_l, \theta)} \right| + \left| \frac{\frac{\partial}{\partial \theta_u} p_{k,l}(n_k, n_l, \theta) \frac{\partial}{\partial \theta_v} p_{k,l}(n_k, n_l, \theta)}{p_{k,l}(n_k, n_l, \theta)^2} \right|.$$

Comme ci-dessus ($F_{k,l}$ est deux fois continûment différentiable), des calculs directs amènent au résultat suivant :

$$\begin{aligned} \left| \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(n_k, n_l, \theta) \right| &\leq C_1 + C_2(n_k + n_l) + C_3(n_k + n_l)^2, \\ &= h_{k,l}(n_k, n_l), \end{aligned}$$

où C_1 , C_2 et C_3 sont des constantes positives. Pour les mêmes raisons que précédemment, $h_{k,l}$ est P_{θ_0} -intégrable.

A₆ : D'après **Ã₅**, les v.a. $\frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0)$ ont une espérance finie et la loi faible des grands nombres assure que $\left(\frac{\partial^2}{\partial \theta_u \partial \theta_v} U_n(\boldsymbol{\theta}_0)\right)_{u,v=1,\dots,p}$ converge en $P_{\boldsymbol{\theta}_0}$ -probabilité vers :

$$I_U(\boldsymbol{\theta}_0)_{u,v=1,\dots,p} = \mathbb{E}_{\boldsymbol{\theta}_0} \left(\sum_{1 \leq k < l \leq d} \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \right).$$

De plus, selon **Ã₅**, la dérivation et l'intégration peuvent être permutées deux fois et selon **Ã₄** le vecteur aléatoire $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0)$ est centré ; ceci entraîne :

$$\begin{aligned} I_U(\boldsymbol{\theta}_0)_{u,v=1,\dots,p} &= - \sum_{1 \leq k < l \leq d} \mathbb{E}_{\boldsymbol{\theta}_0} \left(\frac{\partial}{\partial \theta_u} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \right), \\ &= - \sum_{1 \leq k < l \leq d} J_{F_{k,l}}(\boldsymbol{\theta}_0)^T \mathbb{E}_{\boldsymbol{\theta}_0} \left[\nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0)^T \right] J_{F_{k,l}}(\boldsymbol{\theta}_0), \end{aligned} \quad (\text{D.26})$$

où $J_{F_{k,l}}(\boldsymbol{\theta}_0)$ est la matrice Jacobienne (de taille $3 \times p$) au point $\boldsymbol{\theta}_0$. On remarque que la matrice $I_U(\boldsymbol{\theta}_0)$ est l'opposée de la somme des matrices de covariance. En notant :

$$I_U^{k,l}(\boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0} \left[\nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0)^T \right],$$

pour tout $1 \leq k < l \leq d$, l'équation (D.26) donne l'expression suivante :

$$I_U(\boldsymbol{\theta}_0) = -J_F(\boldsymbol{\theta}_0)^T \begin{pmatrix} I_U^{1,2}(\boldsymbol{\theta}_0) & & \\ 0 & \ddots & 0 \\ & & I_U^{d-1,d}(\boldsymbol{\theta}_0) \end{pmatrix} J_F(\boldsymbol{\theta}_0),$$

où $J_F(\boldsymbol{\theta}_0) = (J_{F_{1,2}}(\boldsymbol{\theta}_0)^T \dots J_{F_{d-1,d}}(\boldsymbol{\theta}_0)^T)^T$ est la matrice Jacobienne $\frac{3}{2}d(d-1) \times p$ de $F(\boldsymbol{\theta}_0)$. Comme F est une fonction définie sur Θ injective (hypothèse 2), la matrice $J_F(\boldsymbol{\theta}_0)$ est de rang p . Par conséquent, $I_U(\boldsymbol{\theta}_0)$ est inversible si la matrice diagonale par blocs, composée des sous-matrices $I_U^{k,l}(\boldsymbol{\theta}_0)$ est inversible. Il suffit donc de montrer que les matrices $I_U^{k,l}(\boldsymbol{\theta}_0)$ sont inversibles pour tout $1 \leq k < l \leq d$. Cependant, la propriété :

$$\det(I_U^{k,l}(\boldsymbol{\theta}_0)) = \det(\mathbb{E}_{\boldsymbol{\theta}_0} (\nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0)^T)) = 0,$$

est équivalente à l'existence de constantes α_1, α_2 et α_3 (non toutes nulles) telles que :

$$(\alpha_1, \alpha_2, \alpha_3) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(N_k, N_l, \boldsymbol{\theta}_0) = 0 \quad \text{p.p.} \quad (\text{D.27})$$

L'équation (D.27) implique en particulier pour $(N_k, N_l) = (0, 0)$,

$$\alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0,$$

pour $(N_k, N_l) = (1, 0)$,

$$-\alpha_1 \frac{(1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0)}{La_{k,l}(\boldsymbol{\theta}_0)} + \alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0,$$

et pour $(N_k, N_l) = (0, 1)$,

$$-\alpha_2 \frac{(1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0)}{Lb_{k,l}(\boldsymbol{\theta}_0)} + \alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0.$$

En raison des contraintes sur les paramètres $a_{k,l}(\boldsymbol{\theta}_0)$, $b_{k,l}(\boldsymbol{\theta}_0)$ et $c_{k,l}(\boldsymbol{\theta}_0)$ (voir hypothèse 2), on trouve $\alpha_1 = \alpha_2 = \alpha_3 = 0$, et donc $I_U^{k,l}(\boldsymbol{\theta}_0)$ est inversible pour tout $1 \leq k < l \leq d$. Par conséquent, la matrice $I_U(\boldsymbol{\theta}_0)$ est inversible.

Liste des publications

Articles de journaux

1. F. Chatelain, S. Lambert-Lacroix and J.-Y. Tournéret, "Composite likelihood estimation for multivariate mixed poisson models generated by gamma intensities," *Statistica Sinica*, submitted, 2007.
2. F. Chatelain, J.-Y. Tournéret, and J. Inglada, "Change detection in multisensor SAR images using multivariate gamma distributions," *IEEE Trans. Image Processing*, under second review, 2007.
3. F. Chatelain, J.-Y. Tournéret, A. Ferrari, and J. Inglada, "Bivariate gamma distributions for image registration and change detection," *IEEE Trans. Image Processing*, vol. 16, no 7, pp. 1796-1806, July 2007.

Articles de conférences

1. F. Chatelain, A. Ferrari, and J.-Y. Tournéret, "Détection d'exoplanètes à l'aide de méthodes des moments," In *Actes du XXIème Colloque GRETSI*, Troyes, France, Sept. 2007.
2. F. Chatelain, J.-Y. Tournéret, and M. Roche, "Estimating the polarization degree of polarimetric images using maximum likelihood methods," In *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Madison, Wisconsin, USA, Aug. 2007.
3. F. Chatelain and J.-Y. Tournéret, "Bivariate gamma distributions for multisensor SAR images," In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, USA, April 2007.
4. F. Chatelain, J.-Y. Tournéret, J. Inglada, and A. Ferrari, "Parameter estimation for multivariate gamma distributions, Application to image registration," In *Proc. EUSIPCO-06*, Florence, Italy, Sept. 2006.
5. F. Chatelain, A. Ferrari, and J.-Y. Tournéret, "Parameter estimation for multivariate mixed Poisson distributions," In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, vol. IV, pp. 17–20, Toulouse, France, May 2006.
6. F. Chatelain and J.-Y. Tournéret, "Composite likelihood estimation for multivariate mixed Poisson distributions," In *Proc. IEEE-SP Workshop Stat. Signal Processing*, pp. 49–54, Bordeaux, France, July 2005.

Bibliographie

- [ADFS03] K. Arwini, C.T.J. Dodson, S. Felipussi, and J. Scharcanski. Comparison of distance measures between bivariate gamma processes. Technical report, School of Mathematics, University of Manchester, 2003. available at www.maths.manchester.ac.uk/~kd/PREPRINTS/hydrology.pdf. 1.2.1
- [AET81] F. Oberhettinger A. Erdélyi, W. Magnus and F.G. Tricomi. *Higher Transcendental Functions*, volume 1. Krieger, New York, 1981. 1.2.2, 3.1.4
- [AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964. 1.2.2, 1.2.2, 1.2.2, 3.1.5, D.2.2
- [AS04] C. Aime and R. Soummer. Influence of speckle and Poisson noise on exoplanet detection with a coronagraph. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, *EUSIPCO-04*, pages 509–512, Vienna, Austria, Sept. 2004. Elsevier. 1.3.4
- [AVV07] G. D. Anderson, M. K. Vamanamurthy, and M. Vuorinen. Generalized convexity and inequalities, 2007. available at <http://arxiv.org/pdf/math/0701262>. D.1.2
- [BC00] S. Breugnot and P. Clemenceau. Modeling and performances of a polarization active imager at $\lambda = 806$ nm. *Optical Engineering*, 39(10) :2681–2688, 2000. (document), 4.1, 4.1.1, 4.2
- [Ber03] P. Bernardoff. Which negative multinomial distributions are infinitely divisible? *Bernoulli*, 9(6), 2003. 1.3.3
- [Ber06] P. Bernardoff. Which multivariate Gamma distributions are infinitely divisible? *Bernoulli*, 12(1) :169–189, 2006. 1.3.2
- [BLBE⁺94] S. K. Bar-Lev, D. Bshouty, P. Enis, G. Letac, I-Li Lu, and D. Richards. The diagonal multivariate natural exponential families and their classification. *J. of Theoret. Probab.*, 7(4) :883–929, Oct. 1994. 1.3.2, 1.3.3
- [BN52] G. E. Bates and J. Neyman. Contributions to the theory of accident proneness. 1. an optimistic model of the correlation between light and severe accidents,. *University of California Publications in Statistics*, 1 :215–254, 1952. available at <http://handle.dtic.mil/100.2/ADA311519>. 1.3.4
- [BN78] O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley, Chichester, 1978. B
- [BR81] D. J. Becker and J. J. J. Roux. A bivariate extension of the gamma distribution. *South African Statistical Journal*, 15 :1–12, 1981. 1.1.1, 1.2.1
- [Bro98] C. Brosseau. *Fundamentals of Polarized Light : a Statistical Optics Approach* Wiley. Wiley, New York, 1998. 4.1, 4.1.1, 4.1.3, 4.1.3, 4.2.2

- [CD05] M. Carpenter and N. Diawara. A multivariate gamma distribution and its characterizations. Technical report, Department of Mathematics and Statistics, Auburn University, 2005. available at <http://www.stat.auburn.edu/fsdd2006/papers/carpenter.pdf>. 1.1.1
- [CG04] F. D Colavecchia and G. Gasaneo. fl : a code to compute Appell's F_1 hypergeometric function. *Computer Physics Communications*, 157 :32–38, feb 2004. D.2.2
- [Che41] K. C. Cheriyan. A bivariate correlated gamma-type distribution function. *J Indian Math Soc*, 5 :133–144, 1941. 1.2.1
- [Cla80] R. T. Clarke. Bivariate gamma distribution for extending annual streamflow records from precipitation : Some large-sample results. *Water Resources Research*, 16(5) :863–870, 1980. 1.1.1, 1.2.1
- [CR03] D. R. Cox and N. Reid. A note on pseudolikelihood constructed from marginal densities. Technical report, available at "<http://www.utstat.utoronto.ca/reid/research/pseudo.pdf>", 2003. 4.3.2
- [Csi67] I Csiszár. Information type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica 2*, pages 299–318, 1967. 2.1.2
- [CT07] L.-S. Chen and I.-S. Tzeng. Some aspects of multivariate generalized gamma random variables. *Statistica Sinica*, 2007. to appear. 1.1.1
- [CVA03] H.-M. Chen, P. Varshney, and M. K. Arora. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, pages 2445–2454, 2003. 2.1.2
- [DB72] A. Dussauchoy and R. Berland. Lois gamma à deux dimensions. *Comptes-rendus Acad. des Sciences de Paris*, 274 :1946–1949, jun. 1972. 1.2.1
- [DCD84] D. Dacunha-Castelle and M. Duflo. *Probabilités et Statistiques, Tome 2*. Masson, 1984. 4.3.2, D.4, D.4, D.4
- [DF61] N. David and E. Fix. Rank correlation and regression in a non-normal surface. In *Proc. Fourth Berkeley Symposium*, volume 1, pages 177–197, Berkeley, California, 1961. 1.2.1
- [D'j62] Z. N. D'jachenko. On a form of bivariate γ -distribution. *Nauchne Trudy Leningradskoi Lesotekhnicheskoi Akademii*, 94 :5–17, 1962. en russe. 1.2.1, 1.2.2
- [Fad06] J. Fade. Méthodes d'estimation paramétrique du degré de polarisation en imagerie polarimétrique cohérente. Master's thesis, Université Aix-Marseille III Paul Cézanne - Institut Fresnel, juin 2006. (document), 4.1.1, 4.2, 4.2, 4.2.3, 4.2.4
- [FLT04] A. Ferrari, G. Letac, and J.-Y. Tournet. Multivariate mixed Poisson distributions. In *EUSIPCO*, Vienna, 2004. 1.3.4, 1.3.4
- [FLT07] A. Ferrari, G. Letac, and J. Y. Tournet. Exponential families of mixed poisson distributions. *J. Multivar. Anal.*, 98(6) :1283–1292, 2007. 1.3.4, 1.3.4
- [Gav70] D. P. Gaver. Multivariate gamma distributions generated by mixture. *Sankhya, Series A*, 32 :123–126, 1970. 1.2.2, 1.3.2
- [GH87] V. P. Godambe and C. C. Heyde. Quasi-likelihood and optimal estimation. *Int. Statist. Rev.*, 55 :231–234, 1987. 4.3.2
- [GL80] D. P. Gaver and P. A. W. Lewis. First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3) :727–745, Sep. 1980. 1.3.2

-
- [GLM03] P. Graczyk, G. Letac, and H. Massam. The complex wishart distribution and the symmetric group. *Ann. Statist.*, 31(1) :287–309, 2003. available at www.lsp.ups-tlse.fr/Fp/Letac/index.html. 4.2.1
- [GM86] C. Genest and J. MacKay. The joy of copulas : bivariate distributions with uniform marginals. *The American Statistician*, 40(4) :280–283, 1986. 1.1.3
- [Goo63] N. R. Goodman. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *Ann. Math. Statist.*, 34(1) :152–177, 1963. 4.2.1
- [Goo85] J. W. Goodman. *Statistical Optics*. Wiley, New York, 1985. 4.3
- [Goo06] J. W. Goodman. *Speckle phenomena in optics : Theory and Applications*. Roberts & Company, 2006. (document), 2.1.1, 2.1, 2.1.1, 2.1.1, 4.1.3
- [Gra97] J. Grandell. *Mixed Poisson Processes*. Chapman & Hall, London, 1997. 1.1.2, 1.3.4
- [Gri69] R. C. Griffiths. The canonical correlation coefficients of bivariate gamma distributions. *The Annals of Mathematical Statistics*, 40(4) :1401–1408, aug. 1969. 1.2.1, 1.2.1
- [Gri84] R. C. Griffiths. Characterization of infinitely divisible multivariable gamma distributions. *J. Multivar. Anal.*, 15(1) :13–20, 1984. 1.3.2
- [GRR03] F. Goudail, N. Roux, and P. Refregier. Performance parameters for detection in low-flux coherent images. *Optics Letters*, 28(2) :81–83, 2003. 4.3
- [GW84] A. K. Gupta and C. F. Wong. On a morgenstern-type bivariate gamma distribution. *Metrika*, 31(6) :327–332, 1984. 1.2.1, 1.2.1
- [HTC04] Y. Hu, V. Tao, and A. Croitoru. Understanding the rational function model : Methods and applications. In *Proceedings of the XXth ISPRS congress*, Istanbul, Turkey, 2004. 2.2.5
- [Hua94] S. Huard. *Polarisation de la lumière*. MASSON, 1994. 4.1.1, 13
- [IG04] J. Inglada and A. Giros. On the possibility of automatic multisensor image registration. *IEEE Trans. Geosci. Remote Sens.*, 42(10) :2104–2120, Oct. 2004. 2.1.2, 2.1.2, 2.2
- [IM07] J. Inglada and G. Mercier. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale-change analysis. *IEEE Trans. Geosci. Remote Sens.*, 45(5) :1432–1446, May 2007. 2.2
- [Ing02] J. Inglada. Similarity measures for multisensor remote sensing images. In *Proceedings of IEEE International Geosci. and Remote Sens. Symposium (IGARSS)*, volume 1, pages 104–106, 2002. 2.1.2
- [JJN⁺99] B. Johnson, R. Joseph, M. L. Nischan, A. Newbury, J. P. Kerekes, H. T. Barclay, B. C. Willard, and J. J. Zayhowski. Compact active hyperspectral imaging system for the detection of concealed targets. In *SPIE AeroSense'99*, volume 3710, pages 144–153, Orlando, FL, April 1999. 4
- [JKK92] N. L. Johnson, S. Kotz, and A. W. Kemp. *Univariate discrete distributions*. John Wiley, New York, 2nd edition, 1992. 1.3.10
- [Joe97] H. Joe. *Multivariate Models and Dependence Concepts*, volume 73. Chapman & Hall, London, May 1997. 1.1.3, 1.1.3, 11, 3.2.2, 3.2.2
- [Joe05] Harry Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivar. Anal.*, 94(2) :401–419, 2005. 3.2.2

- [Jør87] B. Jørgensen. Exponential dispersion models. *Journal of the Royal Statistical Society, Series B*, 49 :127–162, 1987. [B.5](#)
- [JRL02] S. L. Jacques, J. C. Ramella-Roman, and K. Lee. Imaging skin pathology with polarized light. *Journal of Biomedical Optics*, 7 :329–340, July 2002. [4](#)
- [KBJ00] S. Kotz, A. W. N. Balakrishnan, and N. L. Johnson. *Continuous multivariate distributions*, volume 1 : Models and Applications. Wiley, New York, 2 edition, 2000. [1.2.1](#), [1.2.1](#), [1.2.1](#), [1.2.1](#), [B](#)
- [Kib41] W. F. Kibble. A two-variate gamma-type distribution. *Sankhya*, 5 :137– 150, 1941. [1.2.1](#)
- [Kou98] A. E. Koudou. Lancaster bivariate probability distributions with poisson, negative binomial and gamma margins. *TEST*, 7(1) :95–110, 1998. [1.2.2](#)
- [KP51] A. S. Krishnamoorthy and M. Parthasarathy. A multivariate gamma-type distribution. *Annals of Mathematical Statistics*, 22(4) :549–557., Dec. 1951. [1.3.2](#), [1.3.2](#)
- [Lan58] H. O. Lancaster. The structure of bivariate distributions. *The Annals of Mathematical Statistics*, 29(3) :719–736, sep. 1958. [1.2.2](#)
- [Lan63] H. O. Lancaster. Correlations and canonical forms of bivariate distributions. *The Annals of Mathematical Statistics*, 34(2) :532–538, jun. 1963. [1.2.2](#)
- [Lap20] P.-S. Laplace. *Théorie Analytique des Probabilités*, volume 1. Courcier, Paris, 1820. Reprinted in Complete Works, vol. 7, Gauthier-Villars, Paris, 1886, available at http://gallica.bnf.fr/scripts/get_page.exe?O=77595&E=00000004&N=817&F=PDF&CD=0. [A.2](#)
- [Let] G. Letac. Encyclopaedia of mathematics : Natural exponential family of probability distributions. <http://eom.springer.de/n/n120020.htm>. [B](#), [B.6](#)
- [Let92] G. Letac. Lectures on natural exponential families and their variance functions. In *Monografias de matemática*, volume 50. IMPA, Rio de Janeiro, 1992. [B](#), [B.6](#)
- [Lin88] B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 50 :221–239, 1988. [4.3.2](#)
- [LM86] G. Letac and M. Mora. Sur les fonctions-variances des familles exponentielles sur \mathbb{R} . *C.R. Acad. Sci. Paris*, 302 :551–554., 1986. [B.4](#)
- [LS79] E. H. Lloyd and S. D. Saleem. A note on seasonal markov chains with gamma or gamma-like distributions. *Journal of Applied Probability*, 16(1) :117–128, Mar. 1979. [1.3.2](#)
- [LW07] G. Letac and J. Wesolowski. Laplace transforms which are negative powers of quadratic polynomials. *to appear in Transactions of the American Mathematical Society*, 2007. [4.2.2](#)
- [McK34] A. T. McKay. Sampling from batches. *Supplement to the Journal of the Royal Statistical Society*, 1(2) :207–216, 1934. [1.2.1](#)
- [MM91] A. M. Mathai and P. G. Moschopoulos. On a multivariate gamma. *J. Multivar. Anal.*, 39(1) :135–153, 1991. [1.3.1](#)
- [Mor56] D. Morgenstern. Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Statist.*, 8 :234–235, 1956. en allemand. [1.2.1](#)
- [Mor69] P. A. P. Moran. Statistical inference with bivariate gamma distributions. *Biometrika*, 56(3) :627–634, dec 1969. [1.2.1](#)

-
- [Mor82] C. Morris. Natural exponential families with quadratic variance functions. *Ann. Statist.*, 10 :65–80, 1982. [B.4](#)
- [MT00] G. Marsaglia and W. W. Tsang. A simple method for generating gamma variables. *ACM Trans. Math. Softw.*, 26(3) :363–372, 2000. [1.1.2](#)
- [Mui82] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, 1982. [1.3.2](#)
- [Mul01] Keith E. Muller. Computing the confluent hypergeometric function, $M(a, b, x)$. *Numerische Mathematik*, 90(1) :179–196, November 2001. [D.2.2](#)
- [Nel84] J. F. Nelson. The dirichlet-gamma-poisson model of repeated events. *Sociological Methods Research*, 12(4) :347–373, 1984. [1.3.4](#)
- [Nel99] R. B. Nelsen. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1999. [1.1.3](#)
- [NK06] S. Nadarajah and S. Kotz. Bivariate gamma distributions, sums and ratios. *Bulletin of the Brazilian Mathematical Society*, 37(2) :241–274, 2006. [1.2.1](#)
- [NKN04] S. Nomoto, Y. Kishi, and S. Nanba. Multivariate gamma distributions and their numerical evaluations for m -branch selection diversity study. *Electronics and Communications in Japan (Part I : Communications)*, 87(8) :1–12, 2004. [1.3.2](#)
- [PF89] B. Porat and B. Friedlander. Performance analysis of parameter estimation algorithms based on high-order moments. *International Journal of adaptive control and signal processing*, 3 :191–229, 1989. [2.2.2](#), [4.2.2](#)
- [Pha76] R. M. Phatarford. Some aspects of stochastic reservoir theory. *Journal of Hydrology*, 30(3) :199–217, July 1976. [1.1.1](#)
- [Pic76] P. Picard. Généralisation de l’étude sur la survenance des sinistres en assurance automobile. *Bulletin Trimestriel de l’Institut des Actuaire Français*, 385 :204–267, 1976. [1.3.4](#)
- [Pic93] Bernard Picinbono. *Random signals and systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. [4.1.3](#)
- [PS78] A. Prékopa and T. Szántai. A new multivariate gamma distribution and its fitting to empirical streamflow data. *Water Resource Research*, 14(1) :19–24, 1978. [1.2.1](#), [1.3.1](#)
- [QDM93] H.C. Quelle, Y. Delignon, and A. Marzouki. Unsupervised bayesian segmentation of sar images using the pearson system distributions. In *Proc. IGARSS’93*, volume 3, 1993. [4.4](#)
- [Ram51] V. R. Ramabhadran. A multivariate gamma-type distribution. *Sankhyâ*, 11 :45–46, 1951. [1.2.1](#)
- [RFR07a] M. Roche, J. Fade, and P. Réfrégier. Parametric estimation of the square degree of polarization from two intensity images degraded by fully developed speckle noise. *J. Opt. Soc. Am.*, 24(9) :2719–2727, Sept. 2007. [4.2.2](#)
- [RFR07b] P. Réfrégier, J. Fade, and M. Roche. Estimation precision of the degree of polarization from a single speckle intensity image. *Optics Letters*, 32(7) :739–741, 2007. [16](#)
- [RR06] M. Roche and P. Réfrégier. Comparison of the estimation of the degree of polarization from four or two intensity images degraded by speckle noise. In *Proc. EUSIPCO-06*, Florence, Italy, September 2006. elsevier. [4.2](#), [4.2.2](#)

- [RRG06a] P. Réfrégier, M. Roche, and F. Goudail. Cramer-rao lower bound for the estimation of the degree of polarization in active coherent imagery at low photon level. *Opt. Letters*, 31(24) :3565–3567, 2006. 4.3, 4.3.2
- [RRG06b] P. Réfrégier, M. Roche, and F. Goudail. Cramer-rao lower bound for the estimation of the degree of polarization in active coherent imagery at low photon levels. *Opt. Lett.*, Dec 2006. 1.3.4
- [RvZ93] E. J. M. Rignot and J. J. van Zyl. Change detection techniques for ERS-1 SAR data. *IEEE Trans. Geosci. Remote Sensing*, 31 :896–906, 1993. 2.2.5
- [Réf02] P. Réfrégier. *Théorie du bruit et applications en physique*. Editions Hermès-Lavoisier, Paris, 2002. 4.2.2
- [SA81] O. E. Smith and S. I. Adelfang. Gust model based on the bivariate gamma probability distribution. *Journal of Spacecraft and Rockets*, 18 :545–+, November 1981. 1.1.1, 1.2.1
- [Sar66] O. V. Sarmanov. Generalized normal correlation and two-dimensional fréchet classes. *Doklady (Soviet Mathematics)*, 168 :596–599, 1966. en russe. 1.2.1, 1.2.1, 1.2.1, 1.2.2, 3.1.5
- [SAT82] O. E. Smith, S. I. Adelfang, and J. D. Tubbs. A bivariate gamma probability distribution with application to gust modeling. Technical report, NASA technical report, jul 1982. available at <http://hdl.handle.net/2060/19820021218>. 1.1.1, 1.2.1
- [SL95] J. H. Shih and T. A. Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4) :1384–1399, Dec. 1995. 3.2.2
- [SL04] M. Shubina and M.-L. T. Lee. On maximum attainable correlation and other measures of dependence for the sarmanov family of bivariate distributions. *Communications in Statistics - Theory and Methods*, 33(5), 2004. 1.2.1, 1.2.2
- [SM99] D. Sarrut and S. Miguet. Similarity Measures for Image Registration. In *European Workshop on Content-Based Multimedia Indexing*, pages 263–270, Toulouse, France, 1999. IHMPT–IRIT. 2.1.2, 2.1.2
- [SPB78] W. R. Schucany, W. C. Parr, and J. E. Boyer. Correlation structure in farlie-gumbel-morgenstern distributions. *Biometrika*, 65(3) :650–653, dec. 1978. 1.2.1
- [SR89] S. J. Steel and N. J. Le Roux. A class of compound distributions of the reparameterised bivariate gamma extension. *South African Statistical Journal*, 23, 1989. 1.2.1
- [Szá86] T. Szántai. Evaluation of a special multivariate gamma distribution function. *Math. Program. Study*, 27 :1–16, 1986. 1.2.1
- [TLB89] R. Touzi, A. Lopés, and P. Bousquet. A statistical and geometrical edge detector for SAR images. *IEEE Trans. Geosci. Remote Sensing*, 26, 1989. 2.2.5
- [Tre68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory : Part I*. Wiley, New York, 1968. 2.2.5, 2.2.5
- [Tsi04] E. G. Tsionas. Bayesian inference for multivariate gamma distributions. *Statistics and Computing*, 14(3) :223–233, Aug. 2004. 1.3.1
- [Tup07] F. Tupin. *Champs de Markov sur graphes pour le traitement des images radar*. PhD thesis, Département Traitement du Signal et des Images, ENST, 2007. 4.4

-
- [War92] D. Warren. A multivariate gamma distribution arising from a markov model. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 6(3) :183–190, 1992. [1.1.1](#)
- [Wic33] S. D. Wicksell. On correlation functions of type III. *Biometrika*, 25 :121–133, 1933. [1.2.1](#)
- [Wol90] L. B. Wolff. Polarization-based material classification from specular reflection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(11) :1059–1071, 1990. [4](#)
- [Wol94] L. B. Wolff. Polarization camera for computer vision with a beam splitter. *Journal of the Optical Society of America A*, 11 :2935–2945, November 1994. [4](#)
- [Xu96] J. J. Xu. *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. PhD thesis, Department of Statistics, University of British Columbia, Canada, 1996. [3.2.2](#)
- [Yan00] X. Yang. Accuracy of rational function approximation in photogrammetry. In *Proceedings of ASPRS Annual Convention*, Washington D.C., USA, may 2000. American Society for Photogrammetry and Remote Sensing. [2.2.5](#)
- [YOB01] S. Yue, T. B. M. J. Ouarda, and B. Bobee. A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, 246 :1–18, June 2001. [1.1.1](#), [1.2.1](#)
- [Yor03] M. Yor. Sur certaines fonctionnelles exponentielles du mouvement brownien réel. *J. appl. probab.*, 29(1) :202–208, 2003. [1.1.2](#)
- [Yue01] S. Yue. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes*, 15 :1033–1045, April 2001. [1.1.1](#)
- [ZJ05] Y. Zhao and H. Joe. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, 33(3) :to appear, 2005. [4.3.2](#)
- [Zoc05] Mathias Zocher. *Multivariate Mixed Poisson Processes*. PhD thesis, Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden, 2005. available at deposit.ddb.de/cgi-bin/dokserv?idn=978087836&dok_var=d1&dok_ext=pdf&filename=978087836.pdf. [1.3.4](#)