



**HAL**  
open science

# Modélisation stochastique possibiliste et application à la détection de changements dans les bases d'images de télédétection

Charles Lesniewska-Choquet

► **To cite this version:**

Charles Lesniewska-Choquet. Modélisation stochastique possibiliste et application à la détection de changements dans les bases d'images de télédétection. Théorie de l'information et codage [math.IT]. Université Savoie Mont Blanc, 2020. Français. NNT: . tel-03518917

**HAL Id: tel-03518917**

**<https://hal.science/tel-03518917>**

Submitted on 10 Jan 2022

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

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

# THÈSE

**pour obtenir le grade de docteur délivré par**

**l'Université Savoie-Mont-Blanc**

**Spécialité doctorale "STIC : Traitement de l'information"**

*présentée et soutenue publiquement le 23 janvier 2020 à Annecy par*

**Charles LESNIEWSKA-CHOQUET**

Thèse dirigée par : **Abdourrahmane M. ATTO**

Co-encadrée par : **Gilles Mauris**

Préparée au sein du **Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC)**

## **Modélisation stochastique possibiliste et application à la détection de changements dans les bases d'images de télédétection**

Jury composé de :

**Madame Sylvie LE HEGARAT-MASCLE**

Professeure à l'Université Paris-Sud, Présidente

**Monsieur Dominique PASTOR**

Professeur à l'IMT Atlantique, Rapporteur

**Monsieur Olivier STRAUSS**

Maître de Conférences HDR à l'Université de Montpellier, Rapporteur

**Monsieur Abdourrahmane ATTO**

Maître de Conférences à l'Université Savoie Mont Blanc, membre

**Monsieur Gilles MAURIS**

Maître de Conférences HDR à l'Université Savoie Mont Blanc, membre

**Monsieur Didier Dubois**

Directeur de Recherche Émérite à l'Institut de Recherche en Informatique de Toulouse, membre

**Monsieur Laurent FOULLOY**

Professeur à l'Université Savoie Mont Blanc, membre

**Monsieur Grégoire MERCIER**

Docteur HDR à CTO eXo maKina Digital Technologies, membre invité



---

# Remerciements

Ces travaux ont été menés au sein du laboratoire LISTIC grâce à un financement du ministère de la Recherche et de l'Enseignement Supérieur et je tiens donc tout naturellement à remercier ces deux entités sans lesquelles cette thèse n'aurait pu être possible.

Je remercie chaleureusement Monsieur Abdourrahmane Atto et Monsieur Gilles Mauris de m'avoir proposé ce sujet à la frontière entre la théorie des possibilités et traitement d'images.

Merci Abdou de t'être montré disponible et curieux envers mes idées parfois farfelues. Nos nombreuses discussions et toutes ces heures passées à noircir des tableaux d'équations resteront, pour moi, parmi les meilleurs souvenirs de cette thèse.

Merci Gilles pour ton aide précieuse et tes remarques judicieuses. Tes conseils n'ont pas toujours été évidents à intégrer sans une longue réflexion tant ils étaient riches d'implications mathématiques, mais travailler avec un expert de cette théorie aura été une expérience captivante et enrichissante.

Merci, enfin, pour la patience et le soutien que vous m'avez témoigné.

Je remercie, Madame Sylvie Le Hegarat Mascle de m'avoir fait l'honneur de présider le jury de thèse et de l'intérêt porté à mes travaux.

Je remercie également Monsieur Dominique Pastor et Monsieur Olivier Strauss d'avoir accepté de rapporter cette thèse. Vous avez suivi certains de mes travaux au cours de cette aventure et ça à été un honneur pour moi que de vous avoir comme rapporteurs.

Je remercie, d'autre part, Monsieur Didier Dubois d'avoir accepté de participer à ce jury et d'avoir pris de son temps pour améliorer ce manuscrit.

Je remercie, également, Monsieur Laurent Foulloy et Monsieur Grégoire Mercier d'avoir pris part à ce jury.

Je remercie chaleureusement mon père et ma mère Eric et Marie-Hélène pour leur soutien indéfectible. Vous avez su être là pour moi, m'encourageant sans cesse et me poussant à achever cette épreuve de longue haleine.

Je remercie, également, ma sœur et mon frère Audrey et Nicolas d'être présent à mes côtés depuis le début.

---

Enfin, la personne sans qui tout ceci n'aurait été possible, partageant au quotidien mes peines, le stress et les victoires et que je tiens à remercier tout particulièrement est ma chère et tendre Amandine. La thèse n'est pas un long fleuve tranquille, mais avec toi à mes côtés je serais presque prêt à recommencer. Indéniablement, c'est à toi que je dois la concrétisation de cette thèse et j'espère, un jour, arriver à me montrer à la hauteur de tout ce que tu m'as donné. Merci.

---

# Résumé

L'imagerie satellitaire est un domaine en pleine expansion. Chaque année, de plus en plus de satellites sont envoyés en orbite avec pour mission l'observation de notre planète Terre. Les images qu'ils renvoient, tant optiques que radar, fournissent une aide précieuse pour améliorer la gestion et l'exploitation durable de notre environnement. Cependant, les images satellites sont constituées de données imparfaites, de nombreux phénomènes perturbant le processus d'acquisition. Ainsi, l'extraction d'informations à partir des images satellites nécessite des méthodes spécifiques permettant de prendre en compte la nature de ces données.

Dans le cadre de cette thèse, nous proposons et développons de nouvelles méthodes pour modéliser les données imparfaites en nous basant sur la théorie des possibilités. Ces méthodes ont pour objectifs d'améliorer la prise en compte de l'incertitude inhérente aux données à l'aide d'un modèle que l'on appelle distribution de possibilité et qui constitue, selon nous, une alternative aux distributions de probabilité utilisées habituellement dans de nombreux domaines. Dans certains cas, une distribution de probabilité est un modèle trop précis pour décrire correctement un ensemble limité de données imparfaites. Nous construisons alors des distributions de possibilités de manière cohérente avec les probabilités.

La première partie de cette thèse présente deux travaux entrepris dans le cadre de la modélisation possibiliste univariée de données et dont les résultats ont été publiés dans deux papiers de conférences. Nous commençons par étudier les aspects théoriques liés au principe de la transformation de probabilité-possibilité en appliquant notamment les différentes transformations existantes sur des distributions de probabilité discrètes. Puis nous nous intéressons aux densités de probabilité et à leurs transformations qui nous permettent de définir des modèles paramétriques possibilistes de représentation des données imprécises. Nous définissons, ensuite, pour ces modèles une mesure de dissemblance paramétrique à partir de la divergence de Kullback-Leibler afin de pouvoir comparer leurs modélisations dans une application de détections de changements entre images satellites.

Dans la seconde partie de nos travaux, nous nous intéressons à la modélisation possibiliste multivariée de données imparfaites. Nous commençons par étendre la transformation de probabilité-possibilité univariée au cas des lois de probabilité elliptique et dérivons les formes paramétriques de plusieurs distributions de possibilité elliptique. Nous introduisons, aussi, deux formes analytiques de divergences entre des distributions de possibilité calculées à partir de la densité de probabilité gaussienne bivariée. Puis, au travers d'une application de détection de changements entre images satellites multispectrales, nous exposons l'intérêt de la modélisation possibiliste par rapport à la modélisation probabiliste lorsque les données sont perturbées par du bruit blanc gaussien. La transformation possibiliste de lois de probabilités elliptiques a fait l'objet d'une publica-

---

tion dans la revue IEEE Transactions on Fuzzy sets and Systems.

Enfin dans une dernière partie, nous présentons une nouvelle méthode pour construire un modèle possibiliste bivarié à partir de la connaissance des distributions marginales et des structures de dépendances entre elles.

**Mots-clés :** théorie des probabilités, théorie des possibilités, traitement d'images, extraction d'informations, représentation paramétrique, analyse multivariée, distribution de possibilité, mesures de dissemblance, divergence de Kullback-Leibler, transformation probabilité-possibilité

## Abstract

Satellite imagery is a rapidly expanding field. Every year, more and more satellites are sent into orbit with the mission of remotely detecting our planet Earth. The images they capture, both optical and radar, provide valuable assistance in improving the management and sustainable use of our environment. However, satellite images are made up of inaccurate data, with many phenomena disrupting the acquisition process. Thus, the extraction of information from satellite images requires specific methods to take into account the nature of this data.

In this thesis, we propose and develop new methods to model imperfect data based on possibility theory. These methods are intended to improve the consideration of data uncertainty using a model called the possibility distribution, which we believe is an alternative to the probability distributions used in many domains. In some cases, a probability distribution is too precise to describe accurately a limited set of imperfect data. We then assume that by applying a probability-possibility transformation we can define through the possibility distribution a more flexible model to better describe the data.

The first part of our work focuses on univariate possibilistic data modeling. We begin by studying the theoretical aspects related to the principle of the probability-possibility transformation by applying the various existing transformations to discrete probability distributions. Then we are interested in probability densities and their transformations which allow us to define possibilistic parametric models for representing imprecise data. We then define, for these models, a parametric dissemblance measure from the Kullback-Leibler divergence in order to be able to compare their models in an application of change detection between satellite images. In the second part of our work, we are interested in multivariate possibilistic modeling of imperfect data. We begin by extending the univariate probability-possibility transformation to the case of elliptic probability laws and derive the parametric shapes of several elliptic possibility distributions. We also

---

introduce two analytical forms of divergence measure between possibility distributions calculated from the bivariate Gaussian probability density. Then, through an application for detecting changes between multispectral satellite images, we expose the interest of possibilistic modeling compared to probabilistic modeling when the data are disturbed by white Gaussian noise.

Finally, in the last part, we present a new method for building a bivariate possibilistic model based on the knowledge of marginal distributions and the dependences structures between them.

**Key-words :** probability theory, possibility theory, image processing, information extraction, parametric representation, multivariate analysis, possibility distribution, dissemblance measure, Kullback-Leibler divergence, probability-possibility transformation



# Table des matières

<b>Table des matières</b>	<b>v</b>
<b>Liste des figures</b>	<b>vii</b>
<b>Liste des tableaux</b>	<b>xi</b>
<b>Introduction</b>	<b>1</b>
<b>1 Contextualisation</b>	<b>7</b>
1.1 Problématique de la modélisation des données . . . . .	9
1.2 Représentations des données imparfaites . . . . .	23
1.3 Construire une densité de probabilité à partir des données . . . . .	31
1.4 Conclusion du Chapitre . . . . .	45
<b>2 Transformations de probabilité-possibilité unidimensionnelle</b>	<b>47</b>
2.1 Transformation : Définition et principes . . . . .	50
2.2 Première application : Comparaison discrète de distributions de possibilité	58
2.3 Modélisation possibiliste continue . . . . .	71
2.4 Conclusion du chapitre . . . . .	90
<b>3 Distributions de possibilité multivariées</b>	<b>93</b>
3.1 Les vecteurs aléatoires réels . . . . .	96
3.2 Distributions de possibilité elliptiques . . . . .	98
3.3 Application de détection de changements sur images réelles . . . . .	106
3.4 Constructions de distributions de possibilité dominantes . . . . .	112
3.5 Construction de distributions de possibilité multivariées à partir des margi- nales . . . . .	117

3.6	Construction d'une distribution de possibilité jointe à partir de distributions marginales . . . . .	122
3.7	Conclusion . . . . .	144
<b>4</b>	<b>Conclusion &amp; perspectives</b>	<b>147</b>
<b>A</b>	<b>Annexes</b>	<b>I</b>
A.1	Calcul de la divergence de Kullback-Leibler pour des distributions de la forme de $\pi_{\mathcal{N}\mathcal{N}}$ . . . . .	I
A.2	Calcul de la divergence de Kullback-Leibler pour des distributions de la forme de $\pi_{\mathcal{W}}$ . . . . .	II
A.3	Calcul de la divergence de Kullback-Leibler pour des distributions de possibilité elliptiques de la forme de $\pi_{\mathcal{N}}$ . . . . .	III
A.4	Calcul de la divergence d'Hellinger pour des distributions de possibilité elliptiques de la forme de $\pi_{\mathcal{N}}$ . . . . .	VI

# Liste des figures

1.1	Exemple d'image en Niveau de Gris. . . . .	12
1.2	Exemple d'image RGB. . . . .	12
1.3	Histogramme obtenu à partir de deux images différentes . . . . .	14
1.4	Principe de fonctionnement d'un satellite optique. . . . .	15
1.5	Exemple de sélection d'une plage de longueur d'onde dans le domaine du visible. . . . .	16
1.6	Image RGB de la Haute-Savoie acquise par Sentinel-2B. . . . .	18
1.7	Principe d'échantillonnage du signal dans le domaine temporel et spatial. . . . .	19
1.8	Effet d'ombre et de repliement en présence de relief important. . . . .	21
1.9	Exemple de distorsion géométrique sur une image SAR de la tour Eiffel prise par le satellite TerraSAR-X. . . . .	21
1.10	Illustration d'une fonction d'appartenance . . . . .	27
1.11	Représentation d'une fonction d'appartenance triangulaire symétrique . . . . .	28
1.12	Tracé des densités de probabilités associées à la loi Normale et à la loi Gaussienne Généralisée pour une v.a.r $X$ de paramètres $\mu(X) = 2$ et $V(X) = 3$ . . . . .	38
1.13	Représentation des loi de Weibull, Rayleigh et exponentielle pour une v.a.r $X$ de paramètres $\mu = 2$ et $V(X) = 3$ . . . . .	40
2.1	Application des transformations $p \rightarrow \pi$ présentées sur un ensemble de 200 échantillons d'une loi normale de moyenne nulle et de variance égale à 1. . . . .	56
2.2	Fonctionnement de l'algorithme de détection de changements . . . . .	62
2.3	image 1 avec un bruit de chatoisement de variance $\nu = 0.1$ . . . . .	63
2.4	image 2 avec un bruit de chatoisement de variance $\nu = 0.15$ . . . . .	63
2.5	Exemple d'image de changements obtenu en utilisant la transformation TS et la distance de Jaccard pour un bruit de chatoisement de variance $\nu = 0.3$ . . . . .	65
2.6	Illustration d'une courbe ROC. . . . .	66

2.7	Résultats obtenus avec différentes mesures de dissimilarité pour chaque transformation en fonction de la variance du bruit photonique appliqué. . .	67
2.8	Résultats obtenus avec différentes mesures de dissimilarité pour chaque transformation en fonction de la variance du bruit gaussien appliqué. . . .	68
2.9	Couple d'images avant après changements du Black Rock Desert prise par TerraSAR-X en 2011 . . . . .	70
2.10	Résultats de la détection de changements avec la mesure de dissimilarité $D_{LI}$ suivant la transformation $p \rightarrow \pi$ considérée . . . . .	71
2.11	Distribution de probabilité de Weibull $p_W$ et deux transformations $p \rightarrow \pi$ continues de cette loi : $p_{W_s}$ et $p_{W_a}$ . . . . .	75
2.12	Exemple de distribution de données d'une variable aléatoire X centrée dont la valeur absolue  X  suit une loi de Weibull. . . . .	77
2.13	Distributions de possibilité obtenues avec les inégalités de Bienaymé-Chebychev et Gauss-Winckler dans le cas d'un support infini. . . . .	81
2.14	Principe de modélisation et de comparaison pour la seconde application. .	85
2.15	Résultats obtenus en comparant l'approche probabiliste et possibiliste en supposant que les données suivent une loi gaussienne . . . . .	86
2.16	Logigramme de fonctionnement de la troisième application. . . . .	88
2.17	Résultats obtenus en comparant l'approche probabiliste et possibiliste en supposant que les données centrées suivent une loi de Weibull . . . . .	88
3.1	Exemple de distributions de probabilité bidimensionnelle (première ligne) et des distributions de possibilité associées (seconde ligne). . . . .	105
3.2	Exemple de paire d'images "Avant-Après" de la ville de Beihai et le masque de vérité terrain des changements. . . . .	108
3.3	Fonctionnement de l'algorithme de détection de changements. . . . .	109
3.4	Graphique montrant les moyennes des AUC sur les 14 paires d'images en fonction du diamètre du patch d'extraction utilisé. . . . .	111
3.5	Graphique montrant les moyennes des AUC sur les 14 paires d'images en fonction de la variance du bruit gaussien appliqué. . . . .	111
3.6	Distribution de possibilité bivariée calculée à partir de l'inégalité de Chebychev multivariée. . . . .	114

3.7	Tracer de $\pi^\Delta$ pour un rayon $r = 3\sigma$ avec $\sigma = 2$ . En bleu sont tracées les projections des marginales de la distribution de possibilité obtenue par transformation de la PDF $p$ et en rouge sont tracées les projections des marginales de $\pi^\Delta$ .	116
3.8	Tracé de la copule $C(a, b) = \min(a, b)$	119
3.9	Tracé de la copule $C(a, b) = \max(0, a + b - 1)$	120
3.10	Tracé de la copule $C(a, b) = a \times b$	122
3.11	Exemple de distribution de possibilité jointe obtenue avec la copule indépendante. En bleu sont tracées les projections marginales de $\pi_{X,Y}^t(x, y)$ et en rouge les projections marginales de $\pi_{X,Y}(x, y) = \pi_X^m(x) \times \pi_Y^m(y)$ .	125
3.12	Tracé de la $t$ -conorme "union probabiliste $\pi_{X,Y}^S(x, y) = S_p(\pi_X^m(x), \pi_Y^m(y))$ .	127
3.13	Tracé de l'opérateur de moyenne géométrique $\pi_{X,Y}^{M_g}(x, y) = M_g(\pi_X^m(x), \pi_Y^m(y))$ . En rouge sont données les projections marginales de la distribution $\pi_{X,Y}^{M_g}(x, y)$ et bleu les projections de la distribution de possibilité dont la forme est donnée dans l'équation (3.22).	129
3.14	Schéma représentant l'ensemble des opérateurs d'agrégation possibiliste.	129
3.15	isocontour carré défini à partir d'un point $(x, y)$ .	133
3.16	Tracé de la distribution de possibilité construite à partir de distributions de probabilité marginales comonotones.	136
3.17	Changement de repère, rotation d'angle $\frac{\pi}{4}$ .	137
3.18	Tracé de la distribution de possibilité consistante avec des lois de probabilité marginales anticomonotones.	139
3.19	Construction de carré HIJK englobant le cercle de centre O, de rayon $R_{x,y}$ et passant par le point $(x, y)^t$ .	140
3.20	Losange ABCD inscrit dans le cercle de centre O et de rayon $R_{x,y}$ .	142
3.21	Tracer des distributions $\pi_{X,Y}^{\inf}(x, y)$ (en vert), $\pi_{X,Y}^t(x, y)$ (en noir) et $\pi_{X,Y}^{\sup}(x, y)$ (en rouge).	143
4.1	Exemple de profil temporel d'un pixel $a_{i,j}$ pris dans une série temporelle d'images satellites.	152



# Liste des tableaux

1.1	Cas particulier de la loi de Weibull : la loi de probabilité exponentielle et celle dite de Rayleigh. . . . .	39
1.2	Distances vectorielles couramment utilisées en traitement d'images . . . . .	42
1.3	Formes analytiques de la divergence de Kullback-Leibler (non symétrique) pour des lois de probabilité usuelles. . . . .	44
1.4	Formes analytiques de la distance d'Hellinger pour des lois de probabilité usuelles. . . . .	45
2.1	Degré de Consistance et Indice de spécificité pour chaque transformation appliquée à une loi normale : $\mathcal{N}(0, 1)$ . . . . .	57
2.2	Exemple de matrice de confusion. . . . .	65





# Introduction Générale

Le 3 avril 2014, l'agence spatiale européenne (ESA) mettait en orbite le satellite Sentinel 1A, le premier de la constellation des satellites Sentinel [1]. Ce satellite avait pour mission d'observer l'ensemble de la surface terrestre en une dizaine de jours, créant ainsi un flux continu d'images. Ces dernières représentent une mine considérable de données et leur exploitation est nécessaire pour extraire des informations relatives à l'évolution des structures terrestres observables par imagerie.

Ce satellite, ayant pour mission d'imager la surface terrestre, n'était bien entendu pas le premier, mais les données qu'il renvoie sont librement accessibles à la communauté scientifique, dynamisant ainsi tout le domaine de la télédétection. Le projet ANR PHOENIX dans lequel s'inscrit cette thèse a commencé en Octobre 2015 dans un objectif d'analyser les images de tels satellites.

La télédétection spatiale est le terme employé pour désigner l'ensemble des méthodes et techniques permettant d'acquérir, à distance, des données sur les caractéristiques physiques de la Terre. Grâce aux outils de mesures de plus en plus performants, embarqués à bord d'engins spatiaux (satellites, avions, ballons-sondes...), la surface de notre planète et son atmosphère sont cartographiées et analysées. Et de la météorologie jusqu'au suivi de la fonte des glaciers dans les Alpes, les applications sur les données issues de cette discipline sont nombreuses.

Cependant, les données de l'imagerie satellitaire sont difficiles à exploiter, car fortement affectées par plusieurs types de perturbations. La résolution spatiale des satellites est un exemple parlant pour illustrer un type d'imprécision dans ces données. Dans le cas du satellite 1A, la résolution spatiale est de  $5 \times 20$  mètres. Cela signifie qu'une surface de 100 mètres carrés est représentée sur l'image par la valeur d'un seul pixel. Il est donc nécessaire d'utiliser des méthodes adaptées pour traiter ces données selon leurs spécificités.

Jusque dans les années 1960, les méthodes de traitement des données étaient majoritairement ancrées dans une approche probabiliste basée sur une modélisation ponctuelle de l'incertitude. La donnée issue d'une mesure n'étant pas totalement certaine du fait de la nature imparfaite des capteurs ou des approximations réalisées par le système d'acquisition lui associer un degré de certitude entre 0 et 1 permet de préciser la confiance dans cette mesure.

Puis, suite aux travaux de Lotfi Zadeh [2] [3], une nouvelle façon de modéliser l'incertitude dans les données fit son apparition : la modélisation par intervalles flous. Cette approche est plus intuitive et plus complète pour décrire les données. La donnée n'est plus seulement une valeur unique associée à un degré de certitude, mais représentée par un intervalle pour lequel l'appartenance de cette donnée est certaine. La théorie des ensembles flous permet de compléter cette description en précisant l'intervalle certain (support) par un second intervalle plus petit auquel on attache une confiance moins grande (noyau). En particulier, son intérêt a été largement démontré pour modéliser des données qualitatives comme c'est le cas en décision multicritères [4] [5] et des données quantitatives comme c'est le cas dans le domaine de la métrologie [6].

Dans cette thèse, nous nous proposons de développer des outils à partir de la théorie des possibilités. Dans la situation où le nombre d'observations est trop faible ou bien que les données sont trop bruitées, la distribution de probabilité peut sembler trop précise pour bien décrire les données. Nous utiliserons alors des distributions de possibilité, qui sont moins précises que les distributions de probabilité, pour représenter les données issues de la télédétection. Nous pensons que la théorie des possibilités, avec son modèle de représentation par intervalles emboîtés, constitue une alternative intéressante à la théorie des probabilités pour modéliser les données imparfaites de l'imagerie satellitaire.

Le propos de cette thèse se divise alors en deux grands objectifs.

Le premier consiste à étudier, proposer et développer des méthodes univariées et multivariées issues de la théorie des possibilités pour modéliser les incertitudes de mesure dans un contexte de traitement d'images satellites.

Le second est de se comparer aux méthodes probabilistes équivalentes afin de valider l'intérêt d'outils possibilistes appliqués dans ce cas de figure.

Le premier chapitre de cette thèse sera pour nous l'occasion de présenter plus en dé-

tail le sujet et son contexte. Après une présentation synthétique sur le formalisme et les notions de données et d'informations, nous exposerons le processus d'obtention d'une image numérique dans le cas de l'imagerie satellitaire radar ou optique. Nous constaterons que les nombreuses approximations faites tout au long de ce processus d'acquisition nous obligent à considérer les valeurs des pixels dans ces images comme imparfaites, c'est-à-dire à la fois imprécises et incertaines. Puis, nous présenterons successivement les trois théories de l'incertain qui constituent, pour nous, des candidates sérieuses pour traiter ce genre de données. Nous commencerons par la théorie des probabilités, véritable socle sur lequel repose cette thèse et où la compréhension de son formalisme et de ses limitations nous est nécessaire afin de proposer de meilleurs outils. La théorie des ensembles flous, seconde théorie de l'incertain, n'est pas indispensable pour comprendre la théorie des possibilités. Mais le fait qu'elle en soit une généralisation peut permettre d'appréhender plus facilement cette dernière. De plus, certains outils que nous serons amenés à utiliser sont issus de cette théorie et il paraît alors légitime que nous en présentions les éléments principaux. Puis, nous donnerons les éléments fondamentaux de la théorie des possibilités et des distributions de possibilités. Cette description sera complétée tout au long du manuscrit par l'introduction de nouvelles notions utiles à nos développements. Dans la dernière section de ce chapitre, nous présenterons le cadre général de l'étude de données imparfaites univariées à l'aide d'outils probabilistes. En commençant par définir les grandeurs qui nous serviront à caractériser un ensemble de données ainsi que leurs méthodes d'estimations, nous expliquerons comment construire un modèle de représentation paramétrique en ayant notamment recours aux lois de probabilités. Pour finir, nous rappellerons quelques outils très souvent utilisés pour comparer des modèles probabilistes entre eux afin d'extraire des informations sur leurs ressemblances.

Les représentations probabilistes et possibilistes de données sont deux approches à la fois similaires et différentes. En effet il existe entre ces deux théories, un certain nombre de liens permettant de passer d'un formalisme à l'autre. Ainsi, dans le deuxième chapitre de cette thèse, nous nous consacrerons aux méthodes permettant de construire un modèle possibiliste à partir d'un modèle probabiliste.

Dans une première partie, nous présenterons plusieurs transformations permettant d'obtenir une distribution de possibilité à partir d'une distribution de probabilité (et inversement), définies dans la littérature scientifique. Cependant, ces transformations ne sont

pas symétriques. Dans un sens nous aurons tendance à perdre de la précision de description (sens probabilité vers possibilité). Dans l'autre nous risquons au contraire d'ajouter arbitrairement. Il nous faudra alors rappeler certains principes qui permettent de garantir que les transformations ne dénaturent pas l'information que nous voulons traiter.

Puisqu'il existe plusieurs transformations difficilement différentiables en pratique, la première contribution de cette thèse, que nous présenterons, tente d'apporter des éléments de comparaison sur un plan pratique au travers d'une première application de détection de changements entre images. Cette application nous permettra notamment d'introduire la notion de mesure de dissemblance entre deux distributions de possibilité.

Dans la deuxième partie de ce chapitre, nous étudierons le cas de la transformation continue uni-dimensionnelle de Dubois, Foulloy, Mauris et Prade [7]. Nous proposerons, pour les distributions de possibilité issues de cette transformation, une divergence de Kullback-Leibler possibiliste et dériverons plusieurs formes analytiques permettant ainsi un traitement continu des données. Nous comparerons, pour finir, la modélisation possibiliste paramétrique avec la modélisation probabiliste correspondante au travers de deux applications de détections de changements entre images.

Le troisième chapitre de cette thèse concernera l'aspect multivarié de la modélisation des données de l'imagerie satellite. L'idée n'est alors plus seulement de considérer des images à une composante mais des images constituées de plusieurs composantes reliées entre elles par des structures de corrélations.

Au travers de deux approches distinctes, nous tenterons de proposer des méthodes possibilistes pour décrire ce genre de données. La première consistera à construire un modèle multivarié à partir d'un modèle probabiliste lui aussi multivarié (méthode directe). La seconde approche proposera de passer par une modélisation individuelle de chacune des composantes de l'image puis de combiner ces lois marginales avec un opérateur mathématique nommé "Copule" qui permet de prendre en compte les dépendances pour obtenir la loi conjointe multivariée (méthode indirecte).

Ainsi, une des contributions de cette thèse consistera à étendre la transformation continue uni-dimensionnelle au cas des distributions de probabilités elliptiques multivariées. Nous présenterons le résultat de cette nouvelle transformation pour des lois de probabilité connues en précisant les formes obtenues dans le cas bidimensionnel. De plus, nous

proposerons des formes analytiques pour deux mesures de divergence appliquées à des distributions de possibilités elliptiques.

Puis, au travers d'une application de détection de changements réalisée sur une base de 14 paires d'images multispectrales, nous illustrerons l'intérêt de l'approche possibiliste proposée par rapport à l'approche probabiliste dans le cas de données bruitées .

Enfin, dans la seconde partie de ce chapitre, nous exposerons différentes méthodes permettant de construire des distributions de possibilité bivariées consistantes avec la distribution de probabilité jointe à partir des connaissances des distributions marginales et de leur dépendance. Après un court état de l'art des outils permettant de combiner deux distributions de possibilité marginales en fonction de leur dépendance, nous mettrons en évidence le fait que la distribution ainsi construite n'est pas toujours consistante et qu'il est nécessaire de définir une nouvelle méthode basée sur un empilement des régions de dispersion. La probabilité accumulée sur les régions de dispersion sera exprimée à l'aide d'une fonction copule modélisant la dépendance entre les lois de probabilité marginales.



# Chapitre 1

## Contextualisation

*« La pensée, on l'oublie trop  
souvent, est un art, c'est-à-dire un  
jeu de précision et d'imprécision,  
de flou et de rigueur. »*

---

Edgar Morin

## Introduction

Ce premier chapitre est un chapitre d'introduction et de contextualisation plus qu'un chapitre d'état de l'art à proprement parler. Il nous permet de définir mathématiquement les entités que nous allons utiliser tout au long de cette thèse, mais aussi d'expliquer plus en détail les différents aspects de la problématique en donnant des illustrations et des exemples qui permettront au lecteur d'aborder les chapitres suivants avec plus d'aisance.

Dans la première partie, nous commencerons par clarifier les notions de données et d'informations qui sont à l'heure actuelle devenues des mots polysémiques utilisés indifféremment et parfois à tort dans les médias pour parler du *so called* Big Data. Ces deux notions sont d'une grande importance pour comprendre la problématique de cette thèse. En effet tout l'enjeu de ce travail de recherche est de proposer des méthodes nouvelles pour procéder à l'extraction d'informations à partir d'une masse de données. Les données que nous considérons ici sont les valeurs des pixels des images numériques et en particulier celles issues de l'imagerie satellitaire. Cette dernière joue un rôle de plus en plus pro-éminent dans notre société en plein bouleversement. Elle permet par exemple de suivre la fonte des glaciers, mesurer la déforestation, anticiper des phénomènes météorologiques [8][9], etc. Toutefois, l'imagerie satellitaire, malgré le nombre exorbitant de données qu'elle produit chaque jour, et permettant ainsi l'application de méthodes de traitement basées sur de l'apprentissage, souffre d'un défaut majeur : ses données sont très imparfaites. En effet, les images sont parfois assez fortement bruitées, la résolution spatiale est parfois faible en raison de la limitation des moyens technologiques et tous ces facteurs compliquent le traitement de ces images. Il nous faut alors proposer de nouvelles méthodes permettant de prendre en compte avec davantage de souplesse, que les méthodes probabilistes, l'imperfection inhérente aux données.

Dans la seconde partie de ce chapitre, nous évoquerons certaines théories présentes dans la littérature scientifique et proposant des outils de modélisations des ensembles de données imparfaites. Plus particulièrement, nous présenterons la théorie des probabilités et la théorie faisant l'objet de cette thèse : la théorie des possibilités. L'intérêt de la théorie des probabilités est multiple. Historiquement et de par le nombre d'outils qu'elle propose, elle constitue un cadre de référence incontournable pour comparer de nouvelles



méthodes. De plus, le formalisme mathématique qu'elle propose est suffisamment intuitif pour être facilement implémenté et exploité dans des applications complexes de traitement d'images. Cependant, les propriétés de la théorie des probabilités imposent peut-être trop de contraintes sur les modèles les rendant parfois inutilement précis dans le cas de données imparfaites. Il faut alors nous tourner vers des théories où les contraintes sont relaxées. C'est le cas dans la théorie des ensembles-flous et dans la théorie des possibilités que nous présenterons brièvement avant de les aborder plus en détail dans les chapitres suivants.

Enfin, dans la troisième partie de ce chapitre de contextualisation, nous reviendrons sur les outils proposés par la théorie des probabilités. En effet, comme nous l'avons évoqué précédemment, il existe déjà un grand nombre de méthodes probabilistes permettant la modélisation des données imparfaites. Nous verrons ainsi comment construire un modèle probabiliste paramétrique à partir des données en nous attardant plus précisément sur la définition des différentes statistiques associées à un ensemble de données. Puis nous présenterons différents modèles univariés couramment utilisés en traitement d'images. Ces lois de probabilité ont su montrer leur efficacité dans de nombreuses applications tout en étant relativement simples à estimer. Pour terminer, nous aborderons un point important du traitement de données qui est la comparaison d'ensembles ou de modèles associés aux données. Pouvoir quantifier la ressemblance, ou au contraire la dissemblance, entre deux jeux de données joue un rôle majeur dans la plupart des applications, que ce soit pour de la classification, de la segmentation ou bien de la prédiction. Nous présenterons certaines mesures les plus usitées dans la littérature.

## 1.1 Problématique de la modélisation des données

Dans cette première section, nous allons contextualiser notre étude et tenter de définir précisément la problématique liée à la recherche d'information dans les données. Nous commencerons par définir les notions élémentaires de données, d'informations et l'interaction entre ces deux entités. Puis, nous évoquerons le processus d'acquisition d'une donnée et en particulier l'acquisition d'une donnée dans le cas de l'imagerie satellitaire. Nous nous intéresserons plus en détail au processus de discrétisation du réel, que l'on considère comme continu par nature, par un capteur, afin d'obtenir une mesure ou un

ensemble de mesures représentant avec plus ou moins de fidélité la réalité. Enfin nous mettrons en évidence les différents problèmes que pose la nature des données acquises dus aux nombreuses approximations réalisées. Cela nous conduira, en fin de section, à formuler la problématique liée à cette thèse et à laquelle nous allons tenter de répondre avec une approche novatrice par rapport à la littérature scientifique.

### 1.1.1 Données & informations

Commençons par définir dans un premier temps le formalisme que l'on adoptera tout au long de cette thèse pour parler de données et d'informations.

Une donnée désigne une valeur comme par exemple le résultat brut d'une mesure. Elle ne porte donc aucune signification. La valeur "37" est une donnée. On notera par  $x$  la fonction qui à tout élément  $\omega$  d'un ensemble de références de données  $\Omega$  lui associe une valeur dans l'ensemble valeurs des données  $\chi$ , telles que :

$$\begin{aligned} x : \Omega &\rightarrow \chi \\ \omega &\mapsto x(\omega) \end{aligned}$$

On parlera de données continues lorsque une variable peut prendre une infinité de valeur dans un ensemble de référence des données  $\Omega$  infini. Tandis que l'on parlera de données discrètes lorsque  $\Omega$  est fini et qu'il peut s'écrire de la manière suivante :  $\Omega = \{\omega_1, \dots, \omega_n\}$  où  $n \in \mathbb{N}$ . Enfin, on appellera  $\mathbb{D}$  l'ensemble des fonctions de données.

Une information est la mise en contexte d'une donnée. En reprenant l'exemple précédent, on peut associer à la valeur 37 la grandeur physique qui la caractérise pour obtenir l'information suivante : 37°C. On pourrait résumer simplement ce principe par :

$$\text{Information} := \text{donnée} \oplus \text{sens}$$

On notera par  $\varphi$  la fonction qui, à tout élément  $\theta$  de l'ensemble de référence des informations  $\Theta$ , lui associe une valeur dans l'ensemble valeurs des informations  $\Phi$ , telles que :

$$\begin{aligned} \varphi : \Theta &\rightarrow \Phi \\ \theta &\mapsto \varphi(\theta) \end{aligned}$$

De la même façon, on parlera d'informations discrètes si  $\Theta$  est dénombrable et qu'il peut s'écrire  $\Theta = \{\theta_1, \dots, \theta_n\}$  avec  $n \in \mathbb{N}$ . On parlera d'informations continues si l'ensemble  $\Theta$  est infini. On appellera  $\mathbb{I}$  l'ensemble des fonctions d'informations.

Mathématiquement, l'information est l'image d'une donnée par une fonction d'information  $\varphi$  qui peut se définir par la mesure d'une quantité physique ou par une modélisation mathématique. Cela revient à dire que l'ensemble de référence d'information  $\Theta$  est inclus dans l'ensemble valeurs des données  $\chi$ .

### 1.1.2 Les images numériques

De manière générale on définit une image numérique comme une hypermatrice de dimension  $N \times M \times C$  où chaque élément de la matrice  $N \times M$  est appelé un pixel et est constitué d'un vecteur de dimension  $C$ .  $N$  représente le nombre de pixels suivant l'axe horizontal de telle façon que la position  $i$  d'un pixel  $a$  suivant cet axe soit renseignée par  $i \in \{1, \dots, N\}$  (colonne). De la même façon,  $M$  est le nombre de pixels suivant l'axe vertical et la position  $j$  du pixel  $a$  suivant cet axe est donnée par  $j \in \{1, \dots, M\}$  (ligne), on dénotera par la suite un pixel de la façon suivante :  $a_{i,j}$ .

#### 1.1.2.1 Les types d'images numériques

Actuellement, il existe de nombreuses façons d'acquérir une image numérique. On peut par exemple citer les images obtenues à l'aide d'un microscope à force atomique [10] dont les caractéristiques en termes de composantes utilisées pour l'encodage ou en termes d'informations associées aux données mesurées diffèrent totalement d'une image prise avec un appareil numérique grand public. Toutefois, quelle que soit la manière de les obtenir on peut les classer dans deux grandes catégories suivant la nature de la méthode de traitement que l'on va utiliser : les images à une composante et les images vectorielles (ou hyper matricielles) à plusieurs composantes.

Le premier type d'image numérique que nous considérerons est donc l'image numérique qui ne possède qu'une composante. Ce sont donc des matrices de tailles  $N \times M \times 1 = N \times M$  et la valeur des pixels dépend du format d'encodage. On parlera d'image en niveau de gris (NdG) lorsque la valeur d'un pixel est comprise entre 0 et 255 (voir figure 1.1). L'analyse de ce type d'image fait appel à des méthodes de traitement univariées que nous développerons plus en détail dans la section 1.3 et dans le chapitre 2.

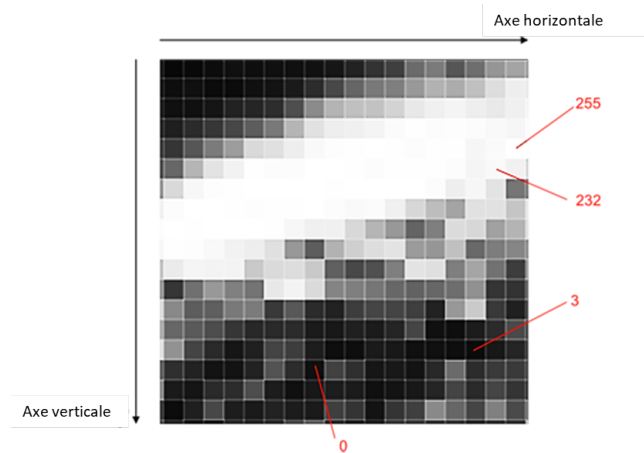


FIGURE 1.1 – Exemple d'image en Niveau de Gris.

Dans le cas où l'image est encodée sur plusieurs canaux, on parlera d'images vectorielles. En particulier, lorsque l'image possède trois canaux et représente des objets dans les longueurs d'ondes visibles on parlera d'image RGB, acronyme de Red (rouge), Green (vert) et Blue (bleu). Dans ce cas l'image est une hypermatrice de taille  $N \times M \times 3$  où chacune des composantes peut-être vue comme une image en Niveaux de Gris (voir figure 1.2). Le traitement des images vectorielles est plus compliqué que dans le cas des images à une composante. En effet, chacune des composantes de l'image représente une facette de l'information. Il est donc souvent nécessaire de prendre en compte l'ensemble des composantes simultanément. On parlera alors de méthodes de traitement multivariées et nous en aborderons certains aspects dans le chapitre 3.

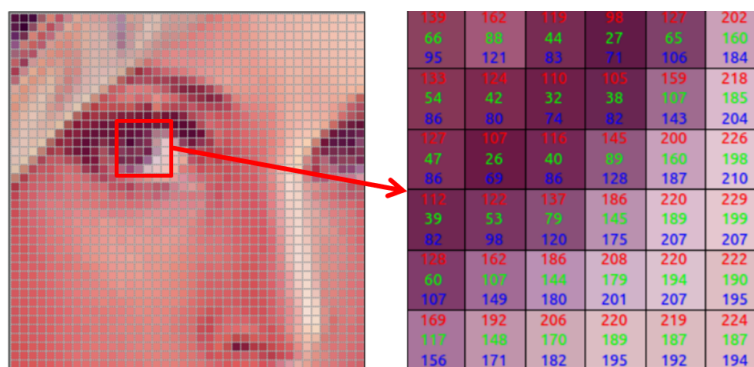


FIGURE 1.2 – Exemple d'image RGB.

Les images numériques issues de l'imagerie satellitaire aussi bien optique que radar sont essentiellement des images de la seconde catégorie [11], c'est-à-dire des images vec-

torielles. Notamment les images optiques sont constituées de 3 canaux ou plus suivant le type de capteur et les images radars sont encodées sur deux canaux : soit un canal pour l'amplitude du signal et un canal dédié à la phase, soit une composante pour la partie réelle et l'autre pour la partie imaginaire du signal. Au cours de cette thèse nous ne considérerons que des images radar amplitude/phase.

Toutefois pour les besoins des applications que nous exposerons au chapitre 2, il est possible de travailler avec des méthodes uni-variées sur des images vectorielles. En effet, dans le cas des images satellites optiques, il existe des méthodes nous permettant de convertir une image vectorielle en image à une composante [12]. Dans le cas des images radars, l'idée sera de travailler sur une composante (en amplitude) en faisant le choix délibéré d'omettre la seconde composante.

### 1.1.2.2 Histogramme d'une image numérique

L'histogramme est un outil important à présenter et souvent indispensable pour modéliser les données au sein d'une image.

L'histogramme est un estimateur de densité simple permettant de visualiser, à l'aide d'une représentation graphique, la répartition des valeurs des pixels au sein de l'image. Cet outil est beaucoup utilisé en traitement d'images puisqu'il permet de voir rapidement grâce à la distribution des pixels, les *défauts* d'une image. On peut par exemple observer qu'une image est peu contrastée, surexposée, etc. Il existe à ce titre de nombreuses méthodes permettant de corriger une image en modifiant directement son histogramme [13].

À partir d'un ensemble de  $N$  observations  $(x_1, \dots, x_N)$ , la construction d'un histogramme se fait en deux étapes : dans un premier temps, on partitionne en classes (ou cellules) l'ensemble de référence  $\Omega$  sur lequel sont ces observations et dans un second temps on compte le nombre d'observations appartenant à chaque classe.

$\Omega$  est subdivisé en  $p$  classes de longueur  $\Delta$  aussi appelée *pas* de l'histogramme. Lorsque toutes les classes ont la même longueur, on parle alors d'histogramme régulier. Pour chaque classe  $A_k$ , le nombre d'observations lui appartenant est appelé accumulateur de  $A_k$ , on le note  $A_{cc}$  et il est défini de la façon suivante :

$$A_{cc}(k) = \sum_{i=1}^N \mathbb{1}_{A_k}(x_i) \quad (1.1)$$

où  $\mathbb{1}_{A_k}(x)$  est la fonction caractéristique de la classe  $A_k$ .

Considérons une image en niveaux de gris de dimension  $N \times M$ , l'histogramme de cette image est la représentation précise de la distribution des intensités des pixels au sein de cette image. Toutefois, l'histogramme présente un inconvénient majeur : deux images différentes peuvent avoir le même histogramme comme on peut le voir dans la figure 1.3. L'histogramme ne rend donc pas compte la distribution spatiale des valeurs des pixels mais seulement de la distribution globale des intensités.

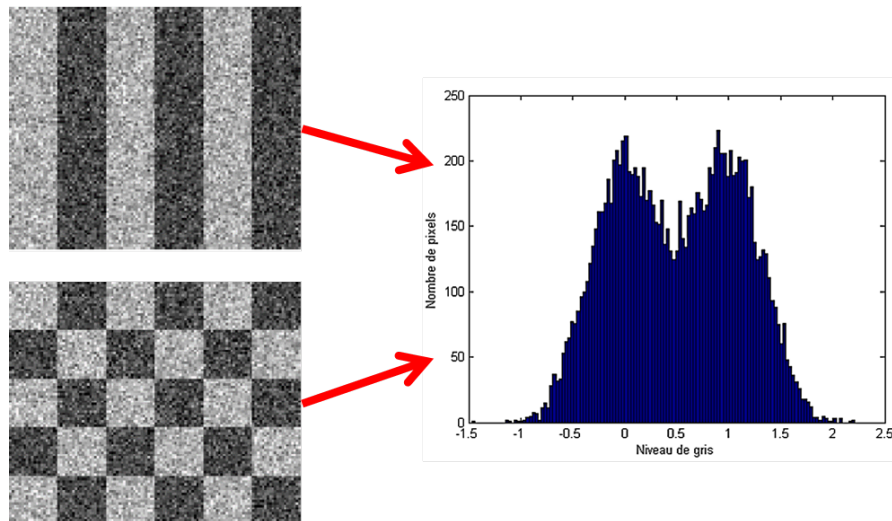


FIGURE 1.3 – Histogramme obtenu à partir de deux images différentes

Ainsi, l'histogramme est un outil simple pour la modélisation des données d'une image numérique mais incomplet sur plusieurs aspects. Nous serons par la suite amenés à utiliser des histogrammes sur des patches locaux d'une image plutôt que sur l'image globale ceci afin de diminuer l'impact de la distribution spatiale des valeurs des pixels.

### 1.1.3 Acquisition des images numériques

Nous allons maintenant nous intéresser à la manière dont sont acquises les images numériques qui constituent les données sur lesquelles reposent nos applications. Comme nous l'avons évoqué précédemment, nous distinguerons deux cas de figure sur les images numériques : si elles ont été obtenues à partir d'un satellite optique ou un satellite radar. On présentera dans ce qui suit seulement certains aspects de l'acquisition d'images numériques par satellites qui nous ont paru nécessaires pour la compréhension de la nature quelque peu particulière de ce type de données.

### 1.1.3.1 Imagerie satellitaire optique

L'acquisition d'images numériques à l'aide d'un satellite optique, ou encore appelé imagerie satellitaire optique, consiste à mesurer au moyen d'un capteur passif les rayons lumineux émis par une source externe (ici le soleil) et réfléchis par l'objet que l'on veut imager (ici la surface terrestre). À noter que tout objet réfléchissant la lumière sera imagé, y compris les perturbations atmosphériques telles que les nuages pouvant ainsi entraîner l'apparition d'artefacts parasites sur l'image finale.

Il existe différents types de capteurs que l'on distingue suivant différents critères : la taille de la plage de longueurs d'onde captée, le nombre de plages de longueurs d'onde enregistrées simultanément et la résolution spatiale de l'image [14].

La résolution spatiale d'une image satellite renseigne avec quelle précision le capteur est capable de détecter un objet au sol. Par exemple, une image satellite d'une résolution spatiale de  $15 \times 20$  mètres(m) sera telle que 1 pixel représente une surface rectangulaire de 15m par 20m de cotés. Dans le cas des images optiques la valeur de ce pixel est généralement égale à la moyenne des intensités lumineuses dans la plage de longueur d'onde voulue, reçues pour cette surface.

La plage de longueurs d'onde peut être restreinte aux longueurs d'onde perceptibles par l'œil humain comme c'est le cas pour la plupart des capteurs dans nos appareils photo, mais certains satellites sont aussi équipés de capteurs multispectraux enregistrant plusieurs bandes spectrales (environ une dizaine). D'autres encore sont équipés de capteurs hyperspectraux enregistrant jusqu'à une centaine de bandes spectrales différentes de l'ultra-violet jusqu'à l'infra rouge [15].

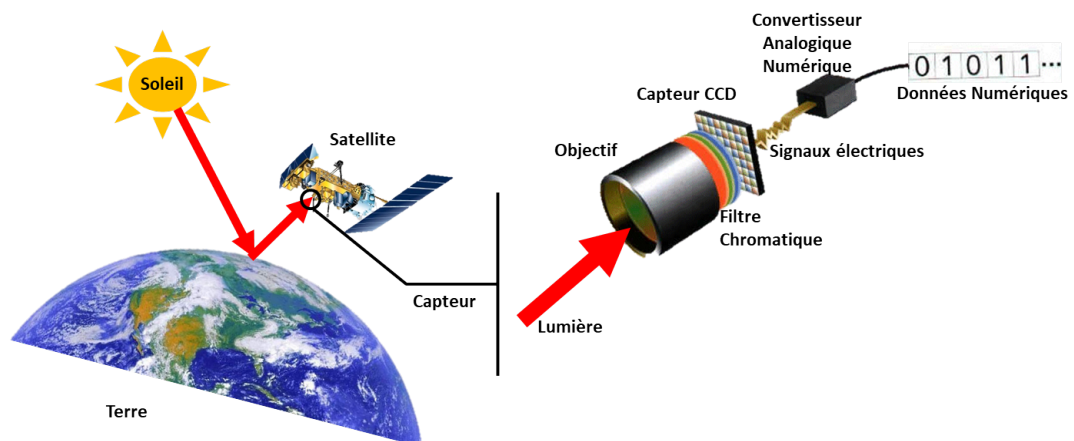


FIGURE 1.4 – Principe de fonctionnement d'un satellite optique.

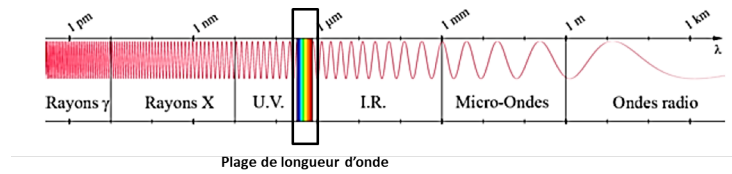


FIGURE 1.5 – Exemple de sélection d’une plage de longueur d’onde dans le domaine du visible.

Le fonctionnement du capteur est relativement simple : dans un premier temps le capteur convertit la lumière reçue en signaux électriques analogiques à l’aide d’un capteur CCD (Couple Charge Device) puis en données numériques grâce au convertisseur analogique numérique (voir figure 1.4). Toutefois, on observe une perte d’informations lors de ces deux étapes rendant les données numériques obtenues légèrement différentes de la scène photographiée et cela malgré les progrès technologiques considérables en matière de photographie numérique ces dernières années.

La première déformation intervient au niveau du capteur CCD. Le capteur CDD est assimilable à une matrice dont chacun de ses composants, appelés photosites, et a pour but de transformer, par effet photoélectrique, un photon lumineux en électrons. Cependant, différents facteurs tels que la température du circuit ou les couches entre les photosites insuffisamment isolantes ne permettent pas de garantir que chaque photon sera bien transformé en électrons. On peut par exemple montrer que l’intensité des signaux électriques n’est pas nulle, mais légèrement supérieure à zéro en l’absence de photons. C’est ce qu’on appelle le bruit photonique ou bruit blanc de manière générale.

Ainsi, pour résumer cette première déformation : si on considère qu’un signal analogique lumineux réfléchi par un objet représente la réalité, alors le signal électrique analogique est une conversion imparfaite du signal lumineux en raison du fait que tous les photons ne sont pas correctement capturés et que le capteur lui-même ajoute un bruit photonique lors de la conversion du signal.

La seconde déformation est due à la conversion des signaux électriques en données numériques par le Convertisseur Analogique Numérique. Cette conversion a pour but de transformer le signal analogique constitué d’une infinité d’amplitudes en un signal numérique contenant une quantité finie de valeurs. Elle se déroule en deux étapes : l’échantillonnage du signal électrique et la conversion du signal analogique échantillonné en signal numérique.



Lors de l'étape d'échantillonnage on va chercher à sélectionner un nombre d'échantillons suffisamment grand pour représenter le signal continu efficacement tout en étant limité par un volume de données défini par le nombre de bits sur lequel sera encodé notre signal. Deux paramètres vont ainsi intervenir : la fréquence d'échantillonnage et la résolution du signal.

La fréquence d'échantillonnage minimum est déterminée par le théorème de Shannon-Nyquist [16] qui stipule que celle-ci doit être au minimum égale à deux fois la fréquence maximale contenue dans le signal à échantillonner. On notera que si celle-ci est trop faible, les variations rapides du signal n'apparaîtront pas toutes dans le signal numérique et des phénomènes de repliement du spectre apparaissent sous la forme de fréquences parasites que l'on ne retrouve pas dans le signal original.

Le paramètre de résolution du signal détermine la précision avec laquelle le signal sera retransmis. On parle de *pas* d'échantillonnage. Plus le nombre de bits sera grand et plus le signal sera retranscrit fidèlement pour une même fréquence d'échantillonnage, mais plus celui-ci sera grand et plus le volume de données le sera aussi. L'idéal étant de trouver un compromis entre la précision désirée et le volume de données qui est limité par les solutions technologiques à disposition.

Pour résumer, cette seconde déformation le signal électrique analogique, que l'on considère déjà comme imparfait de par la première déformation, est à nouveau converti de façon imparfaite en un signal numérique au cours de l'étape d'échantillonnage.

Pour illustrer le principe d'acquisition d'image numérique par satellites optiques nous pouvons prendre comme exemple le cas du satellite Sentinel-2B mis en orbite le 7 mars 2017 [17]. Ce satellite, tout comme le satellite Sentinel-2A, est un satellite optique parmi une constellation de sept satellites envoyés en orbite autour de la Terre par l'Agence Spatiale Européenne (ESA) dans le cadre du programme Copernicus dont l'objectif est la télédétection de la surface terrestre.

Sentinel-2B est équipé d'un capteur multispectral faisant l'acquisition de treize bandes spectrales allant du visible (longueur d'onde 440nm) au moyen infrarouge (longueur d'onde 1370nm). Lors d'une acquisition, il enregistre quatre bandes spectrales (bleu, vert, rouge et proche infrarouge) avec une résolution de 10m, trois bandes (440nm, 940nm et 1370nm) avec une résolution de 60m et les six bandes restantes avec une résolution de 20m.



FIGURE 1.6 – Image RGB de la Haute-Savoie acquise par Sentinel-2B.

### 1.1.3.2 Imagerie satellitaire RADAR

Un satellite radar imageur émet des ondes électromagnétiques (longueur d'onde de l'ordre de quelques centimètres) au moyen d'une antenne sous forme de brèves impulsions et dont l'écho de ce signal, après réflexion sur une surface, est quasi instantanément enregistré par l'antenne [18]. L'imagerie satellitaire radar, contrairement à celle optique présentée précédemment, permet donc d'obtenir des images numériques d'une surface donnée à l'aide d'un capteur actif et ceci malgré la présence de nuage ou d'autres perturbations atmosphériques. En effet, les nuages étant assimilables à des gaz composés de différentes molécules dont la taille est largement inférieure à la longueur d'onde émise par le capteur, ceux-ci ne réfléchissent pas l'onde électromagnétique. Trois grandeurs fondamentales du radar sont à considérer : le temps d'aller-retour de l'onde entre le capteur et l'objet, l'intensité de l'onde reçue et le déphasage entre l'onde émise et l'onde reçue.

Le principe de l'imagerie radar est le suivant [19] : si on appelle  $c$  la vitesse de propagation des ondes électromagnétiques, que l'on peut considérer comme constante dans le milieu de propagation, alors on peut déduire la distance  $d$  d'un objet (rétrodiffuseur) au capteur en mesurant le temps de vol  $t$  qu'aura mis l'onde pour partir de l'antenne, être rétro diffusée par l'objet puis revenir à l'antenne (aller-retour). Le passage du temps de vol à la distance de l'objet au satellite étant donné par la formule suivante :

$$d = \frac{c * t}{2} \quad (1.2)$$

Ce principe est la base de tous les systèmes d'écholocation comme dans le cas de l'imagerie par satellite radar, mais aussi de l'échographie médicale et de la biologie animale (chauve-souris...) et permet de cartographier une zone en localisant chaque objet rétrodiffuseur présent dans la scène.

Toutefois, dans le but d'obtenir une image numérique de haute résolution il faut prendre en compte différents aspects : l'échantillonnage spatial et temporel du signal et la résolution spatiale du système radar.

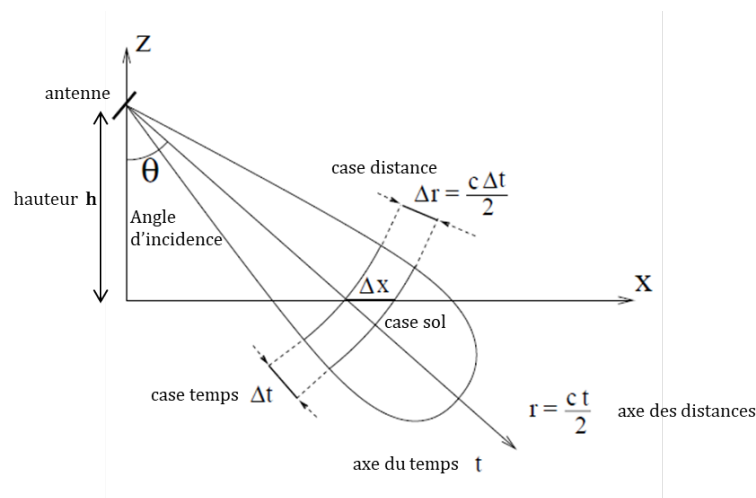


FIGURE 1.7 – Principe d'échantillonnage du signal dans le domaine temporel et spatial.

La première étape à considérer afin d'obtenir une image numérique grâce à l'écholocation est l'échantillonnage du signal reçu par l'antenne. L'échantillonnage spatial et temporel est obtenu grâce à un convertisseur CAN qui convertit le signal analogique reçu par l'antenne en signal numérique. Un signal est émis par une antenne à une hauteur  $h$  (typiquement 700km) avec un angle d'incidence  $\theta$  et est réfléchi puis échantillonné temporellement. À noter que si l'angle  $\theta$  est nul alors, dans l'hypothèse du sol parfaitement plat, l'onde incidente réfléchirait tous les objets au sol simultanément et on ne pourrait pas les distinguer. On fait alors correspondre un échantillon temporel de durée  $\Delta t$  (case temps) à un échantillon spatiale de longueur  $\Delta r = \frac{c \Delta t}{2}$  (case distance) permettant ainsi de définir une case sol de longueur  $\Delta x = \frac{\Delta r}{\sin \theta}$ . Tous les objets rétrodiffusants, contenus dans le volume défini par une case sol seront alors assimilés à un même pixel dans l'image numérique finale.

La seconde étape à prendre en compte pour obtenir une image numérique suffisamment précise afin d'être capable de distinguer deux objets voisins est la définition de la

résolution en distance du capteur. Cette résolution est directement liée à la durée d’impulsion du radar : plus celle-ci sera courte et plus il sera possible de distinguer des échos différents pour des objets proches les uns des autres. Pour déterminer la résolution en distance, on va d’abord définir un intervalle de temps  $\delta t$ , que l’on appelle aussi résolution temporelle. Cela permet de la même façon que pour l’échantillonnage de définir une résolution spatiale  $\delta r = \frac{c\delta t}{2}$  et une résolution au sol  $\delta x = \frac{c\delta t}{2\sin\theta}$ . De par le théorème de Shanon-Nyquist, on définit alors les paramètres de l’échantillonnage tel que pour distinguer deux objets distants d’une longueur  $\delta x$  on ait la relation suivante :

$$\Delta t \leq \delta t = \frac{2 \delta x \sin\theta}{c}. \quad (1.3)$$

Une seconde résolution, que l’on appelle résolution azimutale, est liée aux caractéristiques techniques de l’antenne et à la longueur d’onde émise. Plus les rétrodiffuseurs sont éloignés de l’antenne et plus le faisceau qui les atteint est large entraînant alors une baisse de résolution. De plus la largeur du faisceau est inversement proportionnelle à la taille de l’antenne, c’est-à-dire qu’une antenne de grande dimension émettra un faisceau plus fin et aura une meilleure résolution qu’une antenne de taille moindre. Cependant, la taille de l’antenne est limitée et on ne peut en choisir une aussi grande qu’on le souhaiterait.

La résolution spatiale d’une image radar comme nous venons de le décrire, dépend à la fois de la résolution en distance mais aussi de la résolution azimutale. Elle est de l’ordre de quelques kilomètres pour un pixel dans le cas d’un système radar qui serait immobile. Une technique de traitement appelée Synthetic Aperture Radar (SAR) ou radar à synthèse d’ouverture (RSO) permet d’améliorer la résolution spatiale de l’image [20]. Cette méthode utilise le déplacement du satellite autour de son orbite ainsi que l’acquisition successive de plusieurs mesures radar d’une même cible pour simuler une antenne de grande dimension. Cette technique est à la base de plusieurs applications telles que l’interferométrie [21], la polarimétrie [22] et la radargramétrie [23]. Elle permet notamment d’atteindre une résolution spatiale de l’ordre de quelques mètres (pour plus de détails, voir le livre de Curlander et McDonough [24]).

Une image SAR ainsi obtenue se présente sous la forme d’une image numérique à deux composantes : une composante pour l’amplitude du signal mesuré liée à la quantité d’énergie rétrodiffusée par le sol et une composante de phase donnant le déphasage en chaque point entre le signal émis et le signal reçu.

Cependant, dans de nombreuses applications, l'information de la composante en amplitude est suffisante. L'information de phase est utilisée simultanément avec l'information d'amplitude dans des applications d'interférométrie radar.

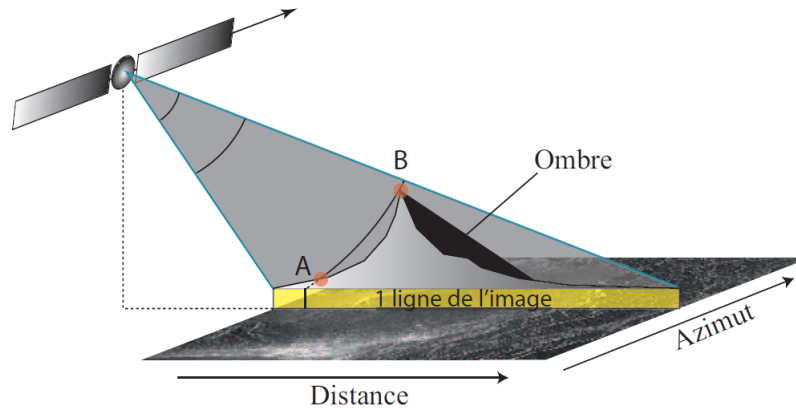


FIGURE 1.8 – Effet d'ombre et de repliement en présence de relief important.

Enfin, dû au principe d'acquisition, les images SAR souffrent de plusieurs types de déformations géométriques. Un effet en particulier est intéressant à évoquer : l'effet de repliement. Cet effet a pour cause l'angle d'inclinaison d'émission des ondes électromagnétiques et se manifeste par un effet d'aplatissement des zones pentues dirigées vers l'antenne et l'apparition d'ombre sur les zones de pentes opposées (voir figure 1.8).



FIGURE 1.9 – Exemple de distorsion géométrique sur une image SAR de la tour Eiffel prise par le satellite TerraSAR-X.

Un exemple d'image satellite SAR où l'on peut facilement apercevoir les distorsions géométriques est représenté figure 1.9. Sur cette image, la tour Eiffel nous apparaît couchée sur le côté par rapport aux autres bâtiments présents dans la scène.

### 1.1.4 Imperfection des données

Dans la section précédente, nous avons présenté les images numériques de manière générale et en particulier les images issues de l'imagerie satellitaire optique ou radar. Nous avons aussi introduit brièvement les solutions techniques permettant de passer du signal analogique qui "représente" la réalité à l'image numérique sur laquelle on se base pour extraire des informations.

Lorsque l'on travaille avec des données telles que les pixels d'une image numérique, il faut prendre en compte le fait que ces données sont imparfaites dans le sens où les différentes étapes de mesures d'un signal analogique jusqu'à son encodage en format numérique introduisent des approximations et des distorsions de la réalité.

Dans la littérature scientifique, l'imperfection dans les données peut se traduire de plusieurs façons [25] : incertitude, imprécision, incomplétude, conflit, redondance, complémentarité et contradiction. Nous pouvons réduire le type d'imperfection des données considérées dans le cadre de cette thèse à deux grandes thématiques que sont les données incertaines et les données imprécises. En effet, puisque nous considérons que les données proviennent d'une même source, nos données ne sont donc pas considérées comme conflictuelles, redondantes, complémentaires ou contradictoires ce qui nécessiterait au moins deux sources différentes mesurant le même événement. De plus, nous allons négliger l'incomplétude des données en ne choisissant de travailler que sur des bases d'images exhaustives. Ainsi, les deux principales caractéristiques de l'imperfection associées aux données sont l'incertitude et l'imprécision.

L'incertitude sur les données renseigne sur la non-conformité des données mesurées par rapport à la réalité. Elle décrit de façon non binaire à quel point une donnée est vraisemblable. L'incertitude dans les données est parfois due à l'imprécision relative sur celle-ci. Une des causes premières d'imprécision est la limitation mécanique de nos capteurs à mesurer la réalité.

Nous voyons ici qu'incertitude et imprécision sont des notions très proches allant souvent de pair. De plus, selon la communauté scientifique dans laquelle on se positionne, la distinction n'est pas toujours faite car on y utilise un même outil mathématique pour modéliser ces deux aspects de l'imperfection simultanément [26]. La théorie des possibilités que nous allons présenter et utiliser dans la suite de cette thèse est fondamentalement une théorie de l'incertain mais sa connexion étroite avec la théories des ensembles flous

en fait aussi un outil de modélisation de l'imprécision [27][26]. Ainsi, afin de simplifier, lorsque nous qualifierons dans la suite les données d'imparfaites nous ferons à la fois référence à l'incertitude et à l'imprécision à laquelle elles peuvent être sujettes.

## 1.2 Représentations des données imparfaites

Nous avons pu le voir dans la partie précédente que l'imprécision et l'incertitude étaient inhérentes aux données acquises dans le cadre de l'imagerie satellitaire. On qualifie ces données d'imparfaites et il nous faut alors les outils mathématiques adéquats afin de prendre en compte l'imperfection intrinsèque de ces données. Ainsi dans cette partie, nous nous consacrerons à la modélisation des données imparfaites au travers de deux théories largement répandues dans la littérature scientifique que sont : la théorie des probabilités et la théorie des possibilités. Nous commencerons par présenter la théorie des probabilités du fait de son utilisation massive dans de nombreux domaines. Puis nous introduirons la théorie des ensembles flous comme une base pour présenter ensuite la théorie des possibilités. À noter que d'autres théories de l'incertain existent et apportent une modélisation différente de l'imperfection des données. On peut, par exemple, citer la théorie de l'évidence de Dempster-Shafer [28][29]. Cependant son utilisation est plus pertinente dans le cadre de la combinaison de preuves comme en classification et nous ne la détaillerons donc pas dans cette thèse.

Toutefois, entre chacune de ces théories existe un point commun : attacher aux données observées une mesure de certitude rendant compte de la confiance qu'on leur accorde, compte tenu des phénomènes connus ou inconnus responsables de leur variation.

Dans ce qui suit, nous considérons que les données imparfaites sont les valeurs des pixels au sein d'une image numérique que l'on peut considérer comme la réalisation d'un événement aléatoire  $X$  à valeurs réelles dans un ensemble de référence  $\Omega$  des valeurs possibles que peut prendre un pixel. L'idée est alors de modéliser ces variations aléatoires en essayant de leur attacher une mesure de certitude propre à cette théorie : la mesure de probabilité dont nous allons rappeler les axiomes ci-après.

### 1.2.1 Théorie des probabilités

La théorie des probabilités est apparue il y a plusieurs siècles de cela. On retrouve par exemple des traces des premiers formalismes au XVII<sup>e</sup> siècle dans les travaux de Blaise Pascal et Pierre de Fermat sur les jeux de dés et de hasard. Mais c'est au XX<sup>e</sup> siècle, que Andreï N Kolmogorov et Bruno De Finetti ont mis en place le formalisme mathématique et les axiomes de la théorie des probabilités [30] [31].

Dénotons par  $\mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$  et notons  $\mathcal{A}$  un sous-ensemble de  $\mathcal{P}(\Omega)$ . On précise que dans le cas où l'ensemble de référence  $\Omega$  est fini alors  $\mathcal{A}$  regroupe l'ensemble des parties de  $\Omega$  noté alors  $2^\Omega$ . Dans le cas où  $\Omega$  est infini alors on définit  $\mathcal{A}$  comme étant une tribu.

**Définition 1.** [32]

On dit que  $\mathcal{A}$  est une tribu de  $\Omega$  si  $\mathcal{A}$  est un ensemble de parties de  $\Omega$  non vide,  $\mathcal{A}$  est stable par passage au complémentaire et par union d'une suite finie ou dénombrable d'éléments :

1.  $\mathcal{A}$  est non vide.
2. si  $A \in \mathcal{A}$  alors le complément de A que l'on note  $\bar{A} = \Omega \setminus A$  est aussi dans  $\mathcal{A}$ .
3. Si  $\{A_1, \dots, A_n\} \subset \mathcal{A}$  alors  $\cup_i A_i \subset \mathcal{A}$  pour  $i \in \{1, \dots, n\}$

Si  $\Omega$  est l'ensemble des réels  $\mathbb{R}$  alors la tribu  $\mathcal{A}$  de  $\Omega$  est la tribu engendrée par des intervalles de  $\mathbb{R}$  et on l'appelle tribu Borélienne (souvent noté  $\mathcal{B}(\mathbb{R})$ ). Le couple  $(\Omega, \mathcal{A})$  est appelé un espace mesurable.

Considérons maintenant l'espace mesurable  $(\Omega, \mathcal{A})$ , on définit une mesure de probabilité  $\mathbb{P}$  comme une fonction de  $\mathcal{B}$  dans l'intervalle  $[0, 1]$  respectant les axiomes suivant :

- $\mathbb{P}(\Omega) = 1$  et  $\mathbb{P}(\emptyset) = 0$
- Soit un événement A de  $\Omega$  alors  $\mathbb{P}(A) \in [0, 1]$
- Soit deux événements A et B de  $\Omega$  :  
 si  $A \cap B = \emptyset$  alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .
- $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$

Le premier axiome stipule que la probabilité maximale est atteinte lorsque l'événement considéré est l'ensemble de référence lui-même. Le second axiome établit que la



probabilité d'un événement pris dans l'ensemble de référence  $\Omega$  est positive et comprise entre la probabilité de l'ensemble vide et la probabilité de l'ensemble  $\Omega$ . Le troisième axiome est un axiome d'additivité. Il détermine que la probabilité de réalisation d'un événement A ou B est égale à la somme des probabilités de l'événement A et de l'événement B si et seulement si A et B sont disjoints. Sinon, la probabilité d'avoir ou A ou B est égale à la probabilité de A plus la probabilité de B moins la probabilité d'avoir A *et* B en même temps. Enfin, le dernier axiome montre que l'on peut complètement définir la connaissance de A à partir de la probabilité de l'ensemble de référence et du complémentaire de A :  $\bar{A}$ .

Dans la suite de cette thèse, on appelle espace probabilisé un espace mesurable complété par une mesure de probabilité et on le note  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Pour une mesure de probabilité  $\mathbb{P}$ , on peut définir une fonction  $F : \mathbb{R} \rightarrow [0, 1]$  caractérisant la probabilité des événements sur leur domaine de définition. Cette fonction est telle que :

$$F(x) = \mathbb{P}(\{X \leq x\}). \quad (1.4)$$

La fonction duale de la fonction de répartition est appelée fonction de survie. Elle est noté

$$\mathbb{P}(\{X \geq x\}) = 1 - F(x). \quad (1.5)$$

Et pour un intervalle A tel que  $A = ]\underline{x}, \bar{x}]$  avec  $\underline{x} \leq \bar{x}$ , nous avons

$$\mathbb{P}(A) = \mathbb{P}(\{X \in ]\underline{x}, \bar{x}]\}) = F(\bar{x}) - F(\underline{x}). \quad (1.6)$$

Si la fonction F est dérivable, on peut définir une densité de probabilité  $p$  associé à F sur un ensemble de référence  $\Omega$  à valeurs réelles.

Dans le cas où  $\Omega$  est fini et ordonné, F s'écrit :

$$F(x) = \sum_{\{\omega \in \Omega | \omega \leq x\}} p(\omega). \quad (1.7)$$

Dans le cas où F est défini sur  $\mathbb{R}$ , alors F est tel que :

$$F(x) = \int_{-\infty}^x p(\omega) d\omega. \quad (1.8)$$

F est appelé fonction de fonction de répartition ou fonction cumulative associée à  $\mathbb{P}$ .

Ainsi, pour une mesure de probabilité  $\mathbb{P}$  sur un ensemble de référence  $\Omega$ , on peut écrire :

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \quad (1.9)$$

Dans le cas où  $A = \Omega$  on retrouve alors le premier axiome :  $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$ .

Pour  $A \subseteq \mathbb{R}$ , la fonction  $p$  est alors appelée densité de probabilité et est définie de la façon suivante :

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \int_A p(\omega) d\omega \quad \text{avec} \quad \int_{\mathbb{R}} p(\omega) d\omega = 1. \quad (1.10)$$

Enfin, en 1951, Shannon dans [33] a développé un outil appelé *entropie probabiliste*, que l'on note  $H$ , d'une mesure de probabilité  $\mathbb{P}$  sur un espace mesurable  $(\Omega, \mathcal{B})$ . Cette entropie permet de définir la quantité d'information fournie par une mesure de probabilité sur un ensemble  $A \subset \Omega$  et est définie de la manière suivante :

$$H(A) = -\log(\mathbb{P}(A)). \quad (1.11)$$

Ce qui nous donne, pour une distribution ou une densité  $p$  (si elle existe) associée à  $\mathbb{P}$  selon que  $\Omega$  est un ensemble fini ou non, l'expression de l'entropie suivante :

$$H(p) = - \sum_{\omega \in \Omega} p(\omega) \log(p(\omega)) \quad (\text{cas } \Omega \text{ fini}) \quad (1.12)$$

$$H(p) = - \int_{x \in \mathbb{R}} p(x) \log(p(x)) dx \quad (\text{cas réel}). \quad (1.13)$$

Nous avons présenté ici les principaux outils de modélisation de la théorie des probabilités. Cette présentation n'est bien sûr pas exhaustive, mais des compléments viendront s'ajouter au fur et mesure de la thèse lorsque cela sera nécessaire.

### 1.2.2 Théorie des ensembles flous

Nous présentons ici brièvement la théorie des ensembles flous introduite par Lotfi Zadeh en 1965 [2]. Elle nous servira de base pour la théorie des possibilités que nous développerons dans la section suivante.

Dans cette théorie, Zadeh propose un cadre mathématique nouveau pour représenter le raisonnement humain. À l'aide d'une fonction d'appartenance  $\mu$  elle permet de remplacer l'appartenance binaire de la réalisation d'un singleton dans un ensemble par une appartenance relative à cet ensemble.

Soit un espace mesurable  $(\Omega, \mathcal{A})$ , le degré d'appartenance d'un singleton  $\omega$  à un ensemble  $A \subseteq \Omega$  est défini pour chaque singleton dans  $\Omega$  tel que :

$$\begin{aligned} \mu_A &: \Omega \rightarrow [0, 1] \\ \omega &\mapsto \mu_A(\omega). \end{aligned}$$

La fonction d'appartenance  $\mu_A$  définit formellement un ensemble flou A.

Afin d'illustrer l'utilisation des ensembles flous on peut considérer la situation suivante.

Imaginons un individu nommé Paul. On demande à une seconde personne de deviner l'âge de Paul en lui précisant que Paul est jeune. En utilisant le formalisme des ensembles flous, l'idée va être d'abord de déterminer un intervalle d'âge S dans lequel on considère qu'une personne est jeune, par exemple entre 15 et 27 ans. Cet intervalle constituera le support de notre fonction d'appartenance et toute personne dont l'âge est en dehors de cet intervalle n'est pas "jeune" :  $\forall \omega \notin S = [15, 27], \mu_A(\omega) = 0$ .

Ensuite on va déterminer le noyau K de notre ensemble flou A, c'est à dire un intervalle de valeurs indubitablement totalement jeunes, par exemple  $K = [20, 22]$ .

On aura alors  $\forall \omega \in K = [20, 22], \mu_A(\omega) = 1$ .

On obtiendra alors la fonction d'appartenance  $\mu_A$  illustré par figure 1.10.

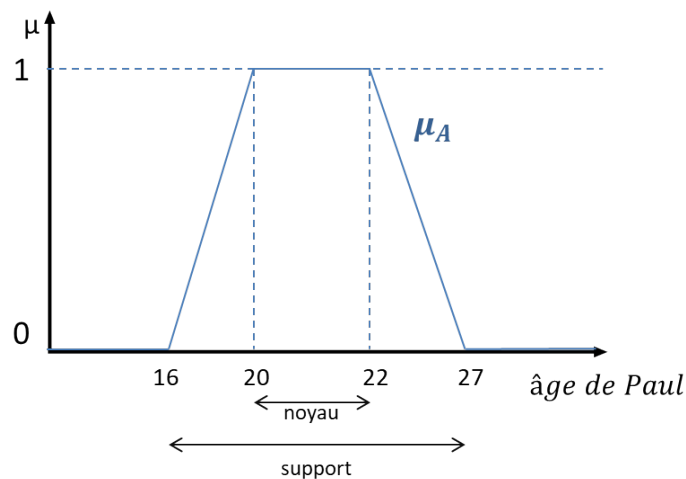


FIGURE 1.10 – Illustration d'une fonction d'appartenance

Un ensemble flou A est donc défini par les deux intervalles extrêmes plus tous les intervalles intermédiaires entre ces deux intervalles :

Noyau  $K(A) = \{\omega \mid \mu_A(\omega) = 1\}$  éléments totalement dans A

Support  $S(A) = \{\omega \mid \mu_A(\omega) > 0\}$  éléments appartenant totalement ou partiellement à A.

Dans le cas où A serait un ensemble classique alors le noyau et le support seraient confondus et la fonction d'appartenance serait égale à 1 sur tout l'ensemble A (cas de l'appartenance binaire à un ensemble).

Les fonctions d'appartenance de la théorie des ensembles flous ne sont donc pas des ensembles "classiques". La distinction entre ces deux théories se fait aussi au niveau de l'interaction entre deux ensembles. Dans le cas des ensembles flous les règles d'union, d'intersection et de complétion sont énoncées de la façon suivante :

$$\forall A, B \subseteq \Omega, \mu_{A \cup B}(\omega) = \max(\mu_A(\omega), \mu_B(\omega)), \quad (1.14)$$

$$\forall A, B \subseteq \Omega, \mu_{A \cap B}(\omega) = \min(\mu_A(\omega), \mu_B(\omega)), \quad (1.15)$$

$$\forall A \subseteq \Omega, \mu_{\bar{A}}(\omega) = 1 - \mu_A(\omega). \quad (1.16)$$

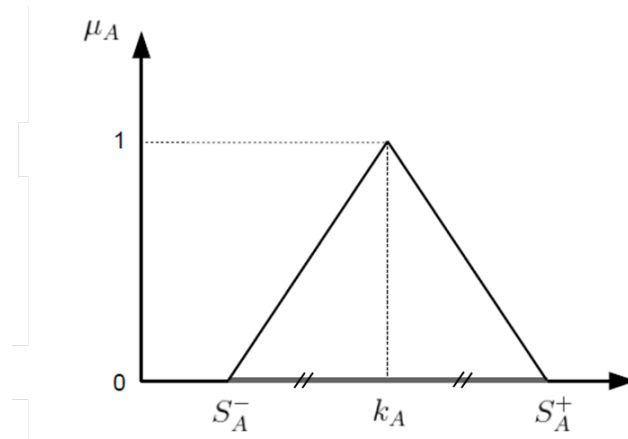


FIGURE 1.11 – Représentation d'une fonction d'appartenance triangulaire symétrique

Enfin, une fonction d'appartenance très utilisée dans la littérature est la fonction triangulaire symétrique à valeurs dans  $\mathbb{R}^+$ . Cette fonction est complètement définie par son intervalle support  $[S_A^-, S_A^+]$  et son noyau précis  $K_A$  est restreint à un unique élément  $K_A = \{k_A\}$  tel que  $[S_A^-, k_A] = [k_A, S_A^+]$ .

### 1.2.3 Théorie des possibilités

La théorie des possibilités a été introduite Lotfi Zadeh en 1978 [3] comme une restriction à la sémantique d'incertitude d'un sous-ensemble flou présentée précédemment. Développée notamment dans les travaux de Didier Dubois et Henri Prade [34][35][36], elle constitue une alternative à la théorie des probabilités pour représenter des informations imparfaites et plus particulièrement pour prendre en compte l'imprécision dans les données à l'aide d'un modèle moins contraint que ne le propose la théorie des probabilités.

Dans la théorie des possibilités il existe deux mesures duales : une mesure de possibilité  $\Pi$  associant un degré de possibilité à la réalisation d'un événement et une mesure de nécessité  $N$  décrivant la certitude de la réalisation de cet événement.

En considérant un espace mesurable  $(\Omega, \mathcal{A})$ , la mesure de possibilité  $\Pi : \Omega \rightarrow [0, 1]$  des définie de la façon suivante :

$$\Pi(\emptyset) = 0 \quad \text{et} \quad \Pi(\Omega) = 1, \quad (1.17)$$

$$\forall A, B \subseteq \Omega, \Pi(A \cup B) = \max(\Pi(A), \Pi(B)), \quad (1.18)$$

où  $\Pi(A)$  est le degré de possibilité de réalisation de l'événement  $A$ .

La mesure de nécessité  $N : \Omega \rightarrow [0, 1]$  est le dual de la mesure de possibilité telle que :

$$N(A) = 1 - \Pi(\bar{A}) \quad (1.19)$$

où  $\bar{A}$  est le complémentaire de  $A$ .

On aura alors les implications suivantes entre les deux mesures [27] :

$$\Pi(A) < 1 \Rightarrow N(A) = 0, \quad (1.20)$$

$$N(A) > 0 \Rightarrow \Pi(A) = 1. \quad (1.21)$$

De plus, il est à noter que ces deux mesures sont en lien direct avec la théorie des probabilités. En effet, quelque soit  $\Pi$  et  $N$ , il existe une mesure de probabilité  $\mathbb{P}$ , telle que :

$$\forall A \subseteq \Omega, \quad N(A) \leq \mathbb{P}(A) \leq \Pi(A), \quad (1.22)$$

et

$$\forall A \subseteq \Omega, \quad \Pi(A) = \sup_{\mathbb{P} \leq \Pi} \mathbb{P}(A). \quad (1.23)$$

Cette inégalité est d'une grande importance pour la suite de thèse, elle nous permet de définir les bornes supérieures et inférieures d'une mesure de probabilité.

Connaissant une distribution de possibilité nous avons la relation suivante  $\pi : \Omega \rightarrow [0, 1]$  satisfaisant la relation suivante :

$$\forall A \subseteq \Omega, \quad \Pi(A) = \sup_{\omega \in A} \pi(\omega). \quad (1.24)$$

La distribution de possibilité permet aussi de définir la mesure de nécessité :

$$\forall A \subseteq \Omega, \quad N(A) = \inf_{\omega \notin A} (1 - \pi(\omega)). \quad (1.25)$$

La distribution de possibilité peut être vue comme une fonction d'appartenance de la théorie des ensembles flous, elle constitue un objet essentiel dans la théorie des possibilités. La distribution  $\pi$  associe à chaque singleton  $\omega \in \Omega$  un degré dans l'intervalle  $[0, 1]$  de possibilité de réalisation de ce singleton, tel que :

$$\begin{aligned} \pi &: \Omega \rightarrow [0, 1] \\ \omega &\mapsto \pi(\omega) \end{aligned}$$

Avec ce formalisme, on peut modéliser l'ignorance totale sur la possibilité de réalisation des événements dans  $\Omega$  :  $\forall \omega \in \Omega, \pi(\omega) = 1$ .

Mais on peut aussi modéliser la connaissance complètement certaine par :  $\exists \omega_i \in \Omega, \pi(\omega_i) = 1$  et  $\forall \omega \neq \omega_i, \pi(\omega) = 0$ .

Enfin on définit le degré de spécificité (ou degré d'imprécision [37]) d'une distribution de possibilité telle que :

$$I_S(\pi) = \int_{-\infty}^{+\infty} \pi(\omega) d\omega. \quad (1.26)$$

Ce degré de spécificité permet de quantifier la quantité d'informations véhiculée par la distribution de possibilité. Plus la distribution est informative et plus ce degré sera faible.

On dira que la distribution de possibilité  $\pi_1$  est plus spécifique que  $\pi_2$  si :

$$\forall \omega \in \Omega, \quad \pi_1(\omega) \leq \pi_2(\omega). \quad (1.27)$$

On retrouve ici la notion d'inclusion floue [38].

### 1.3 Construire une densité de probabilité à partir des données

Au travers de cette dernière section du chapitre 1, nous allons expliquer comment inférer une densité de probabilité à partir d'un ensemble de données. C'est un problème qui a fait l'objet d'un nombre important de publications dans la littérature scientifique notamment parce qu'il touche de nombreux domaines de la science : biologie, mécanique, automatique, traitement du signal, etc. Grâce aux distributions de probabilité, on peut modéliser les variations aléatoires de phénomènes mesurés et ainsi observer, prédire et anticiper le comportement de systèmes physiques complexes.

Cependant, même si une densité de probabilité est un outil puissant de modélisation d'un ensemble de données, l'estimation ayant lieu sur un nombre fini d'observations la fiabilité du modèle est relative. En effet une estimation optimale d'une densité représentative est atteinte lorsque le nombre d'observations tend vers l'infini. À l'inverse plus le nombre d'observations est restreint et moins le modèle inféré est fiable. L'incertitude associée au modèle est de plus difficilement quantifiable.

Il existe, dans la littérature, deux grands types d'approches pour estimer une densité de probabilité à partir des données.

Une première approche se base sur une estimation non paramétrique de la densité de probabilité à l'aide d'une extraction d'information par noyau sommatif comme proposé par Parzen et Rosenblatt [39] ou Epanechnikov [40].

La seconde approche que nous allons développer dans cette section repose sur une estimation paramétrique de la densité.

Le principe de l'estimation paramétrique est le suivant : on suppose dans un premier temps que la distribution des observations suit une loi de probabilité particulière ayant des paramètres, puis dans un second temps on cherche à identifier les paramètres (ou moments) de la densité de probabilité paramétrique par rapport aux observations disponibles.

Les observations que nous considérerons en priorité seront les pixels d'une image mais la méthode peut s'appliquer à tous types de variables aléatoires. Nous commencerons par présenter différentes méthodes d'estimation des moments associés à une va-

riable aléatoire. Puis nous exposerons plusieurs lois de probabilité couramment utilisées par la communauté scientifique dans le domaine du traitement des images satellites. Enfin, nous terminerons en rappelant différentes méthodes permettant de comparer deux distributions entre elles.

### 1.3.1 Estimation des paramètres d'une variable aléatoire

Nous allons présenter ici différentes méthodes pour estimer les paramètres d'une distribution de probabilité modélisant une variable aléatoire.

L'estimation des statistiques d'une variable aléatoire a fait, et fait toujours, l'objet de nombreuses publications scientifiques [41][42] où l'on propose de nouvelles méthodes plus robustes pour minimiser l'écart entre l'estimation et la valeur réelle du paramètre. À ce titre il eut été intéressant pour la suite de cette thèse de pouvoir tester dans chaque application plusieurs méthodes différentes d'estimation statistique. Toutefois, l'objet de cette thèse portant davantage sur la modélisation des données plutôt que sur l'estimation statistique des variables aléatoires associées aux données, nous ne présenterons que quelques méthodes les plus usitées dont la robustesse est satisfaisante pour les besoins de nos applications.

Commençons par quelques définitions fondamentales :

**Définition 2.** [43] *Variable aléatoire réelle :*

*Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , une variable aléatoire réelle est une application mesurable de  $\Omega$  dans  $\mathbb{R}$  :*

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$$

Nous considérerons par la suite deux cas de variables aléatoires : discrètes et continues :

**Définition 3.** *Une variable aléatoire  $X$  à valeurs dans un ensemble  $\chi$  fini ou dénombrable est appelée variable aléatoire discrète (v.a.d). La loi de  $X$  est alors déterminée par :*

$$\forall x \in \chi, \mathbb{P}_X(x) = \mathbb{P}(X = x).$$

De plus si  $X$  est une variable aléatoire discrète, alors pour toute partie  $A$  de  $\chi$  on aura :  $\mathbb{P}_X(A) = \sum_{x \in A} \mathbb{P}(X = x)$  et  $\mathbb{P}_X(\chi) = 1$ .



**Définition 4.** [44] Une variable aléatoire réelle  $X$  à valeurs dans un ensemble  $\chi$  non dénombrable est appelée variable aléatoire continue (v.a.c). La loi de  $X$  est alors déterminée par l'ensemble des probabilités :

$$\forall a, b \in \mathbb{R} \text{ telles que } a < b, \quad \mathbb{P}(a < X < b).$$

Énonçons maintenant ce que nous désignerons par *variables aléatoires indépendantes* :

**Définition 5.** [45] Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , les variables aléatoires réelles  $X_1$  et  $X_2$  sont indépendantes si la condition suivante est remplie :

$$\forall x_1, x_2 \in \mathbb{R}, \quad \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1) \odot \mathbb{P}(X_2 \leq x_2)$$

où  $\odot$  représente le produit entre les deux mesures de probabilités. A noter qu'il existe plusieurs manières de définir l'indépendance entre deux variables aléatoires [46]. Toutefois, nous ne les énoncerons pas ici par souci de simplification (plus de détails dans [47]).

Puis, définissons la notion de statistique d'une variable aléatoire :

**Définition 6.** Une statistique  $S$  est une fonction des observations  $x_1, x_2, \dots, x_n$  constituant une variable aléatoire réelle :

$$\begin{aligned} S : \quad \mathbb{R}^n &\rightarrow \mathbb{R} \\ (x_1, x_2, \dots, x_n) &\mapsto S(x_1, x_2, \dots, x_n) \end{aligned}$$

Enfin, la définition d'un estimateur  $T$  d'une grandeur  $\theta$  est :

**Définition 7.** Un estimateur d'une grandeur  $\theta$  est une statistique  $T$  à valeurs dans l'ensemble des valeurs possible de  $\theta$ .

Une estimation de  $\theta$  est alors une réalisation  $t$  de l'estimateur  $T$ .

### 1.3.1.1 Espérance, moment & moyenne empirique

On définit l'espérance d'une variable aléatoire  $X$  de la manière suivante dans le cas discret puis dans le cas continu :

**Définition 8.** *Espérance mathématique :*

Soit  $X$  une variable aléatoire dans  $\mathbb{R}$  alors :

(cas discret) Si  $X$  est une variable aléatoire discrète à valeurs dans  $\chi$ , l'espérance de  $X$  est la quantité :

$$\mathbb{E}[X] = \sum_{x \in \chi} x \mathbb{P}(X = x).$$

(cas continu) Si  $X$  est une variable aléatoire continue à valeurs dans  $\chi$  et admettant une densité de probabilité  $p_X$  alors l'espérance de  $X$  est la quantité :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x p_X(x) dx$$

sous réserve de convergence de l'intégrale.

L'espérance mathématique est le cas particulier pour  $k = 1$  des moments d'une variable aléatoire. On définit les moments d'une variable de la façon suivante :

**Définition 9.** Soit  $X$  une variable aléatoire réelle. Les moments d'ordre  $k$  ( $k = 1, 2, \dots$ ) de  $X$  sont définis par :

$$\mu_k = \mathbb{E}[X^k]$$

tel que  $\mu_k = \sum_{x \in \chi} x^k \mathbb{P}(X = x)$  dans le cas où  $X$  est une variable aléatoire discrète et  $\mu_k = \int_{-\infty}^{+\infty} x^k p_X(x) dx$  lorsque  $X$  est une variable aléatoire continue.

Considérons maintenant un jeu de données  $X = (x_1, \dots, x_n)$  tel que  $X$  est une variable aléatoire discrète.

**Définition 10.** *Moyenne empirique :*

On appelle moyenne empirique, la statistique notée  $\bar{X}$  définie par :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

**Proposition 1.** Soit  $X$  une variable aléatoire réelle de moyenne  $\mu$ . On a :

$$\mathbb{E}[\bar{X}] = \mu$$

### 1.3.1.2 Moment centré & variance empirique

Définissons à présent les notions de variance et d'écart type d'une variable aléatoire. Ces grandeurs nous serviront à quantifier la dispersion d'une variable aléatoire autour de sa moyenne.

**Définition 11.** Soit  $X$  une variable aléatoire discrète telle que  $X = (x_1, \dots, x_n)$  et soit  $\bar{X}$  la moyenne de  $X$ .

On définit la variance notée  $V(X)$  de  $X$  par :

$$V(X) = \mathbb{E}[(X - \bar{X})^2].$$

On définit l'écart-type noté généralement  $\sigma$  de  $X$  par :

$$\sigma = \sqrt{V(X)}.$$

La variance est le cas particulier pour  $k = 2$  des moments centrés d'une variable aléatoire. Les moments centrés sont définis tel que :

**Définition 12.** Soit  $X$  une variable aléatoire réelle de moyenne  $\mu$ . Les moments centrés de  $X$  d'ordre  $k = (1, 2, \dots)$  sont

$$\mu'_k = \mathbb{E}[(X - \mu)^k].$$

De par la définition 11 et en appliquant le même raisonnement que pour estimer empiriquement la moyenne de  $X$ , l'estimateur empirique de la variance de  $X$  est :

**Définition 13.** Soit une  $X$  une variable aléatoire discrète telle que  $X = (x_1, \dots, x_n)$  et soit  $\bar{X}$  la moyenne de  $X$  alors la variance empirique de  $X$  est :

$$Var(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2.$$

Cependant l'estimateur présenté dans la définition 13 est biaisé. Ce qui veut dire que pour cet estimateur et avec un très grand nombre d'échantillons de  $X$  à disposition, la valeur de la variance estimée ( $Var(X)$ ) ne converge pas vers la valeur théorique ( $V(X)$ ) de la variance de  $X$ .

En effet, on peut montrer (voir par exemple [48]) que l'espérance de  $Var(X)$  est :

$$\mathbb{E}[Var(X)] = \frac{n-1}{n} V(X)$$

Ainsi, l'estimateur de la variance empirique non biaisé que nous considérerons dans la suite de cette thèse est :

**Définition 14.** Soit  $X$  une variable aléatoire discrète de moyenne  $\bar{X}$ . La variance empirique non biaisée notée  $\tilde{V}$  de  $X$  est :

$$\tilde{V} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2.$$

### 1.3.1.3 Justification de la méthode des moments

Nous avons introduit précédemment les définitions des moments et des moments centrés d'une variable aléatoire  $X$ , ainsi que certains cas particuliers de ces définitions au travers de l'espérance et de la variance mathématique. Pour la majorité des lois présentées dans la section 1.3.2, l'estimation de  $\mu$  et  $V(X)$  est généralement suffisante pour générer la densité de probabilité modélisant la variable aléatoire réelle  $X$ . Cette démarche est justifiée par la loi dite des grands nombres énoncée ci-après :

**Théorème 1.** *Loi faible des Grands Nombres (LGN) (Bernoulli 1713 [49]) :*

Soit  $X_1, X_2, \dots, X_n$   $n$  variables aléatoires réelles indépendantes et identiquement distribuées d'espérance commune  $\mu$  et de variance commune  $V < \infty$ . Alors

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq \epsilon \right) = 0.$$

Cette loi indique que lorsque l'on fait un tirage aléatoire dans une série de grande taille, plus la taille du tirage augmente, plus ses statistiques se rapprochent des statistiques de la série. Ainsi, appliqué à notre situation, cela revient à supposer que pour un échantillon de taille suffisante les moments empiriques estimés tendent vers les moments théoriques (réels) de la variable aléatoire permettant alors de poser le système d'équations de la méthode des moments.

De manière générale la méthode des moments peut s'énoncer de la manière suivante : Pour  $\theta \in \mathbb{R}$  et  $k \in \mathbb{N}^*$ , si  $\mathbb{E}[X^k] = \phi(\theta)$ , où  $\phi$  est une fonction inversible, alors l'estimateur de  $\theta$  par la méthode des moments est  $\tilde{\theta} = \phi^{-1}(\mathbb{E}[X^k])$ .

Cette méthode se généralise naturellement aux moments centrés de tous ordres.

## 1.3.2 Densités usuelles de probabilité uni-variées

Comme nous l'avons dit précédemment, la valeur d'un pixel peut être vue comme la réalisation d'une variable aléatoire  $X$  dans un ensemble de référence  $\Omega$ . Nous allons présenter ci-dessous différentes lois de probabilité univariées utilisées couramment en

traitement d'images satellites et permettant de modéliser les pixels considérés.

Nous considérerons, dans la suite, en particulier une loi symétrique définie sur  $\mathbb{R}$  et une loi asymétrique définie sur  $\mathbb{R}^+$  ainsi que les cas particuliers associés à chacune d'entre elles.

### 1.3.2.1 Loi Normale et loi Gaussienne Généralisée

La première famille de lois que nous présentons est la loi Normale, aussi appelée loi de Gauss ou loi de Laplace-Gauss.

La loi gaussienne est une loi continue symétrique et unimodale permettant de décrire le comportement d'une variable aléatoire. Elle possède deux paramètres : la moyenne  $\mu$  qui est un paramètre de position et l'écart type  $\sigma$  qui est un paramètre d'échelle. Elle sera par la suite dénotée de la façon suivante  $\mathcal{N}(\mu, \sigma)$ .

Pour tout nombre réel  $x$  pris dans l'ensemble des réels  $\mathbb{R}$ , la densité de probabilité est telle que :

$$p_{\mathcal{N}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \quad (1.28)$$

Dans le cas de la loi normale, l'espérance, la médiane et le mode de la variable aléatoire sont égales à la moyenne  $\mu$  de la distribution. La variance  $V(X)$  de la variable  $X$  est telle que :  $V(X) = \sigma^2$ .

La fonction de répartition  $F$  associée à la distribution  $p$  de la loi normale est :

$$F_{\mathcal{N}}(x) = \frac{1}{2} \left( 1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right) \quad (1.29)$$

$$\text{avec } \operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-u^2} du.$$

Il existe, d'autre part, une seconde forme de la loi normale appelée Loi Gaussienne Généralisée permettant de modéliser différemment une variable aléatoire  $X$ . Un troisième paramètre  $\beta$  permet de modifier la densité de probabilité de la loi normale de façon à prendre en compte une concentration des données autour de la moyenne (cas  $\beta$  tend vers  $\infty$ ) ou au contraire une répartition des données plus homogène sur l'ensemble de définition (cas  $\beta \leq 1$ ).

La densité de probabilité de la loi normale généralisée prend la forme suivante :

$$p_{G.\mathcal{N}}(x) = \frac{\beta}{2\theta\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left(\frac{|x-\mu|}{\theta}\right)^\beta\right) \quad (1.30)$$

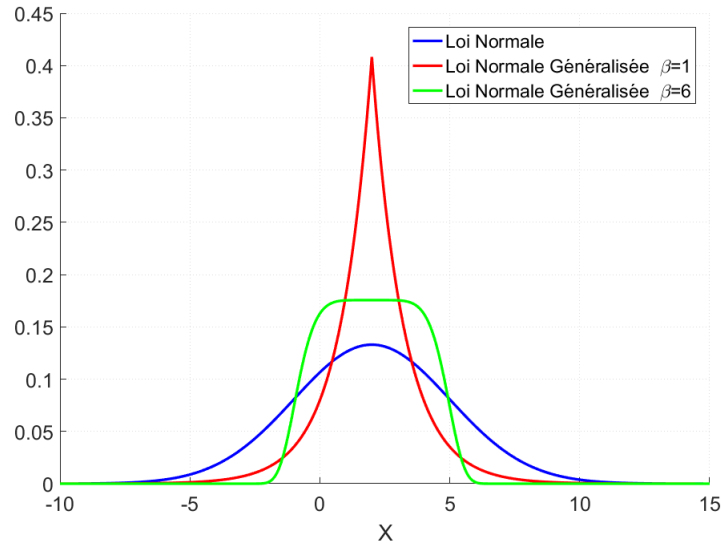


FIGURE 1.12 – Tracé des densités de probabilités associées à la loi Normale et à la loi Gaussienne Généralisée pour une v.a.r  $X$  de paramètres  $\mu(X) = 2$  et  $V(X) = 3$ .

où  $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$ .

La moyenne de cette densité de probabilité est  $\mu$  tout comme pour la loi normale classique. Cependant, la variance associée à cette loi dépend des deux paramètres  $\theta$  et  $\beta$  et est désormais  $V(X) = \frac{\theta^2 \Gamma(3/\beta)}{\Gamma(1/\beta)}$ .

Et sa fonction de répartition est :

$$F_{G.N}(x) = \frac{1}{2} + \text{sign}(x - \mu) \frac{\gamma\left[\frac{1}{\beta}, \left(\frac{|x-\mu|}{\theta}\right)^\beta\right]}{2\Gamma\left(\frac{1}{\beta}\right)} \quad (1.31)$$

où  $\gamma(a, z)$  est la fonction gamma incomplète tel que :

$$\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt, \quad (1.32)$$

et  $\text{sign}(x)$  est la fonction pour laquelle

$$\text{sign}(x) = \begin{cases} -1 & \text{si } x < 0 \\ 0 & \text{si } x = 0 \\ 1 & \text{si } x > 0. \end{cases} \quad (1.33)$$

Enfin pour illustrer ces deux lois de probabilités, nous avons tracé en figure 1.12 les densités de probabilité  $p_N$  et  $p_{G.N}$  associées à une variable aléatoire réelle ( $X$ ) qui aurait pour moyenne théorique  $\mu(X) = 2$  et pour variance théorique  $V(X) = 3$ .

### 1.3.2.2 Loi de Weibull, loi de Rayleigh et loi exponentielle

La seconde famille de lois que nous présenterons dans cette section est la loi de Weibull nommée ainsi en l'honneur du mathématicien Waloddi Weibull mort à Annecy en 1979.

La loi de Weibull est une loi de probabilité continue asymétrique définie sur  $\mathbb{R}^+$  et reposant sur deux paramètres statistiques : le facteur d'échelle  $\alpha > 0$  et le facteur de forme  $\beta > 0$ . On définit sa densité de probabilité telle que :

$$p_W(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right). \quad (1.34)$$

Pour cette loi, l'espérance vaut  $\mathbb{E}[X] = \alpha\Gamma\left(1 + \frac{1}{\beta}\right)$ , la médiane vaut  $x^m = \alpha(\ln 2)^{1/\beta}$  et son mode  $x^*$  est égal à  $\alpha\left(\frac{\beta-1}{\beta}\right)^{1/\beta}$  si  $\beta > 1$  et 0 sinon.

La fonction de répartition associée à  $p_W$  est :

$$F_W(x) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right). \quad (1.35)$$

Deux cas particuliers de la loi de Weibull sont mentionnés dans le tableau 1.1 et présentent un intérêt dans cette thèse.

Loi de probabilité	Paramètres	Densité de probabilité	Fonction de répartition
Loi de Rayleigh	$\alpha = \theta\sqrt{2}$ et $\beta = 2$	$p_R(x) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right)$	$F_R(x) = 1 - \exp\left(-\frac{x^2}{2\theta^2}\right)$
Loi exponentielle	$\beta = 1$	$p_e(x) = \alpha \exp(-\alpha x)$	$F_e(x) = 1 - \exp(-\alpha x)$

TABLEAU 1.1 – Cas particulier de la loi de Weibull : la loi de probabilité exponentielle et celle dite de Rayleigh.

Enfin une illustration des lois de Weibull, Rayleigh et exponentielle est donnée figure 1.13 pour une variable aléatoire  $X$  de moyenne  $\mu = 2$  et de variance  $V(X) = 3$ .

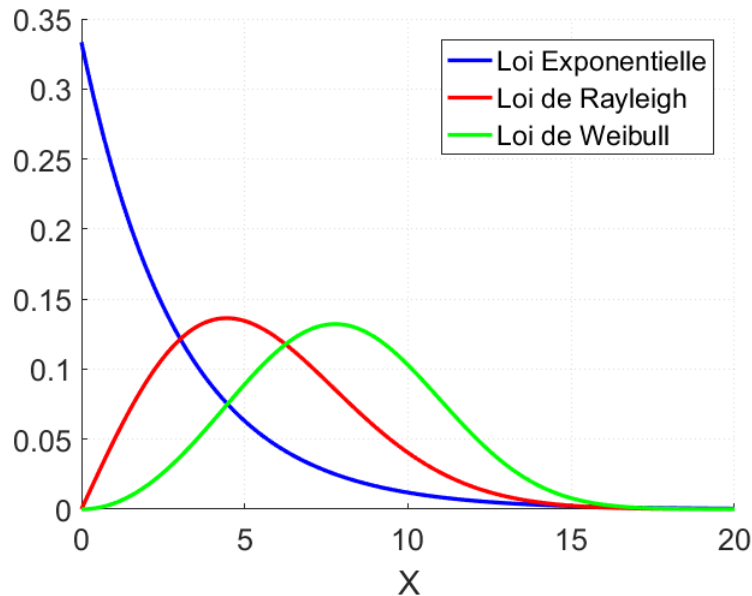


FIGURE 1.13 – Représentation des loi de Weibull, Rayleigh et exponentielle pour une v.a.r  $X$  de paramètres  $\mu = 2$  et  $V(X) = 3$ .

### 1.3.3 Comparaison de distributions de probabilité

Dans les sections précédentes nous avons vu comment déterminer une densité associée à une loi de probabilité modélisant un ensemble de données réelles. Cette partie, que l'on peut appeler étape de modélisation, est nécessaire dans l'ensemble des applications de traitements d'images. L'étape consécutive s'ajoutant naturellement à l'étape de modélisation est l'étape de comparaison des modèles ou plus particulièrement l'étape de comparaison des densités.

De nombreuses applications de traitement d'image nécessitent de pouvoir quantifier le degré de ressemblance entre deux ensembles de données. On peut par exemple citer les algorithmes de classification, de segmentation et de détection de changements. Dans chacun des exemples donnés précédemment, l'idée générale est sensiblement la même : comparer le modèle statistique représentant nos données à un modèle théorique représentant une classe pour ainsi définir un degré d'appartenance à celle-ci.

Ainsi dans les deux dernières sous-sections, nous allons définir et présenter certaines méthodes de comparaison de densités de probabilité couramment utilisées dans la communauté scientifique du traitement d'images. Dans la première section, nous verrons la notion de distance et en particulier de distance métrique entre deux fonctions indépendamment de leurs natures. Puis dans la seconde section, nous définirons la notion de di-



vergence et examinerons le cas de deux divergences en particulier pour lesquelles il existe des formes analytiques des lois précédemment présentées.

### 1.3.3.1 Distances

La distance en mathématique est à l'origine une application permettant de quantifier la longueur entre deux points. On la définit de la façon suivante pour deux points appartenant à un même ensemble.

**Définition 15.** Soit un ensemble  $E$  non vide. Une application  $d$  telle que :

$$\begin{aligned} d & : E \times E \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto d(x, y) \end{aligned}$$

est une distance sur  $E$  si elle vérifie les axiomes suivants :

- (Symétrie)  $\forall x, y \in E^2, d(x, y) = d(y, x)$
- (Séparation)  $\forall x, y \in E^2, d(x, y) = 0 \Leftrightarrow x = y$
- (Inégalité triangulaire)  $\forall x, y, z \in E^3, d(x, z) \leq d(x, y) + d(y, z)$

La notion de distance telle que l'on vient de la définir peut nous permettre de comparer deux fonctions discrètes que l'on considère sous la forme de vecteurs pris dans un espace vectoriel. Pour cela il faut définir la notion de distance dans un espace vectoriel normé.

Définissons dans un premier temps la notion de norme.

**Définition 16.** Soit  $E$  un espace vectoriel de  $\mathbb{R}$ . Une application  $\|\cdot\|$  sur  $E$  telle que :

$$\begin{aligned} \|\cdot\| & : E \rightarrow \mathbb{R}_+ \\ x & \mapsto \|x\| \end{aligned}$$

est une norme sur  $E$  si elle vérifie les axiomes suivants :

- (Séparation)  $\forall x \in E, \|x\| = 0 \Leftrightarrow x = 0$
- (homogénéité positive)  $\forall \lambda \in \mathbb{R}, \forall x \in E, \|\lambda x\| = |\lambda| \|x\|$
- (Inégalité triangulaire)  $\forall x, y \in E^2, \|x + y\| \leq \|x\| + \|y\|$

On peut alors définir la notion de distance entre deux points pris dans un espace vectoriel.

**Définition 17.** Soit un espace vectoriel normé  $(E, \|\cdot\|)$ . On définit la distance  $d$  à partir de la norme telle que :

$$d : E \times E \rightarrow \mathbb{R}_+$$

$$(x, y) \mapsto d(x, y) = \|y - x\|$$

Enfin, considérons deux points  $\mathbf{x} = (x_1, \dots, x_n)$  et  $\mathbf{y} = (y_1, \dots, y_n)$  de l'espace vectoriel  $E$ . La distance entre ces deux points peut-être calculée par chacune des distances suivantes :

Nom	Fonctions
Distance de Manhattan (norme L1)	$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n  x_i - y_i $
Distance euclidienne (norme L2)	$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Distance de Tchebychev (norme L-∞)	$d(\mathbf{x}, \mathbf{y}) = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^n  x_i - y_i ^p} = \sup_{1 \leq i \leq n}  x_i - y_i $

TABLEAU 1.2 – Distances vectorielles couramment utilisées en traitement d'images

Finalement, on peut facilement considérer une distribution associée à une loi de probabilité sur un ensemble discret sous la forme d'un vecteur et appliquer une des distances proposées dans le tableau 1.2. En revanche, dans le cas où l'on considérerait une densité associée à une loi continue, il n'existe pas toujours une forme analytique et l'on est alors obligé de discrétiser la loi considérée. Cela pose un certain nombre de problèmes liés au pas de la discrétisation notamment, pouvant entraîner d'importantes variations dans les valeurs de distances.

### 1.3.3.2 Divergences

La dernière méthode de comparaison de distributions de probabilité que nous décrivons dans cette section est basée sur la notion de divergence. Nous commençons par définir la famille des  $f$ -divergence au sens de Csiszár [50], puis considérons deux divergences en particulier : la divergence de Kullback-Leibler [51] et la distance d'Hellinger (définie d'après l'intégrale d'Hellinger [52]). Il existe d'autres types de divergences telles que les divergences au sens de Bregman [53] que nous ne présenterons pas dans cette

thèse. Nous terminons en donnant les formes analytiques de ces divergences pour certaines des lois de probabilités précédemment énoncées.

Soient deux mesures de probabilité  $\mathbb{P}$  et  $\mathbb{Q}$  sur un espace mesurable  $(\Omega, \mathcal{A})$  telle que  $\mathbb{P}$  absolument continue par rapport à  $\mathbb{Q}$ . Alors une  $f$ -divergence est une fonction  $D_f(\mathbb{P}\|\mathbb{Q})$  est définie telle que :

$$D_f(\mathbb{P}\|\mathbb{Q}) = \int_{\Omega} f\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} \quad (1.36)$$

où  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  est une fonction convexe telle que  $f(1) = 0$ .

**Proposition 2.** [54] Soit un espace mesurable  $(\Omega, \mathcal{A})$  et deux mesures de probabilité  $\mathbb{P}$  et  $\mathbb{Q}$  telle que  $\mathbb{P}$  est absolument continue par rapport à  $\mathbb{Q}$  alors :  $D_f(\mathbb{P}\|\mathbb{Q}) \geq 0$

Les mesures de divergences de la famille des  $f$ -divergences présentent deux propriétés intéressantes, elles sont toujours positives ou nulles et ne s'annulent que si  $\mathbb{P} = \mathbb{Q}$ . Cependant contrairement aux mesures de distances présentées précédemment, les  $f$ -divergences ne sont pas des mesures symétriques et elles ne respectent pas non plus l'inégalité triangulaire.

D'autre part, si  $\mathbb{P}$  (resp.  $\mathbb{Q}$ ) est absolument continue par rapport à une mesure  $\nu$   $\sigma$ -finie sur  $(\Omega, \mathcal{A})$  alors d'après le théorème de Radon-Nikodym [55] on peut écrire que :

$$d\mathbb{P} = p d\nu \quad (\text{resp. } d\mathbb{Q} = q d\nu) \quad (1.37)$$

où  $p$  (resp.  $q$ ) est la densité de probabilité associée à la mesure  $\mathbb{P}$  (resp.  $\mathbb{Q}$ ).

On peut alors réécrire équation (1.36) de la manière suivante :

$$D_f(\mathbb{P}\|\mathbb{Q}) = \int_{\Omega} f\left(\frac{p(x)}{q(x)}\right) q(x) d\nu(x) \quad (1.38)$$

Considérons maintenant un premier cas particulier pour  $f(t) = t \log(t)$ , alors la divergence ainsi obtenue est la divergence de Kullback-Leibler et on la notera  $D_{\text{KL}}(\mathbb{P}\|\mathbb{Q})$  telle que pour deux mesures  $\mathbb{P}$  et  $\mathbb{Q}$  continues de densités respective  $p$  et  $q$  on a :

$$D_{\text{KL}}(\mathbb{P}\|\mathbb{Q}) = \int_{\Omega} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx. \quad (1.39)$$

Cette divergence est aussi appelée entropie relative de  $\mathbb{P}$  par rapport à  $\mathbb{Q}$  du fait qu'on peut la réécrire sous la forme suivante :

$$\begin{aligned} D_{\text{KL}}(\mathbb{P}\|\mathbb{Q}) &= \int_{\Omega} p(x) \log(p(x)) dx - \int_{\Omega} p(x) \log(q(x)) dx \\ &= H(p, q) - H(p) \end{aligned} \quad (1.40)$$

où  $H(p, q)$  est l'entropie croisée entre  $p$  et  $q$  telle que  $H(p, q) = - \int_{\Omega} p(x) \log(q(x)) dx$ .

Il est parfois intéressant de considérer une version symétrique de la divergence de Kullback-Leibler en posant :

$$D_{KL}(\mathbb{P}, \mathbb{Q}) = D_{KL}(\mathbb{P} \parallel \mathbb{Q}) + D_{KL}(\mathbb{Q} \parallel \mathbb{P}). \quad (1.41)$$

Cette mesure que l'on appelle parfois divergence de Jeffrey [56] est toujours une divergence au sens des  $f$ -divergences mais possède en plus une propriété de symétrie :

$$D_{KL}(\mathbb{P}, \mathbb{Q}) = D_{KL}(\mathbb{Q}, \mathbb{P}).$$

Dans le tableau 1.3 sont présentées les formes analytiques de la divergence de Kullback-Leibler pour la loi Normale  $p_{G_i}(x, \mu_i, \sigma_i)$  et la loi de Weibull  $p_{W_i}(x, \alpha_i, \beta_i)$  pour  $i = 1$  ou  $2$ . Dans le cas de la loi de Weibull,  $\gamma \approx 0.5772$  est la constante d'Euler-Mascheroni. On peut aisément déduire la forme analytique de la divergence de Kullback-Leibler pour la loi de Weibull, les formes analytiques de cette divergence pour la loi exponentielle et la loi de Rayleigh.

Loi	Forme analytique
Loi Normale	$D_{KL}(p_{G_1} \parallel p_{G_2}) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$
Loi de Weibull	$D_{KL}(p_{W_1} \parallel p_{W_2}) = \log\left(\frac{\beta_1}{\alpha_1^{\beta_1}}\right) - \log\left(\frac{\beta_2}{\alpha_2^{\beta_2}}\right) + (\beta_1 - \beta_2) \left[ \log(\alpha_1) - \frac{\gamma}{\beta_1} \right] + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \Gamma\left(\frac{\alpha_2}{\alpha_1} + 1\right) - 1$

TABLEAU 1.3 – Formes analytiques de la divergence de Kullback-Leibler (non symétrique) pour des lois de probabilité usuelles.

Le second cas particulier que nous considérerons dans cette thèse est le cas  $f(t) = (\sqrt{t} - 1)^2$ . Cette divergence est appelée distance de Hellinger et est bien une  $f$ -divergence au sens de Csiszár mais possède en plus une propriété de symétrie et constitue une distance métrique sur l'espace des distributions de probabilité. On la note  $D_H(\mathbb{P}, \mathbb{Q})$ , et pour deux distributions  $p$  et  $q$  associées aux mesures  $\mathbb{P}$  et  $\mathbb{Q}$ , elle est définie telle que :

$$D_H(\mathbb{P}, \mathbb{Q}) = \sqrt{\frac{1}{2} \int_{\Omega} \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx}. \quad (1.42)$$

De même, on donne dans le tableau 1.4, les formes analytiques de la divergence d'Hellinger pour la loi normale  $p_{G_i}(x, \mu_i, \sigma_i)$  et pour la loi de Weibull  $p_{W_i}(x, \alpha_i, \beta_i)$  pour  $i = 1$  ou  $2$ . À noter que dans le cas de la loi de Weibull, la forme analytique est seulement disponible pour deux distributions ayant le même paramètre de forme  $\beta$ . On pourra tout de

même en déduire, à partir de cette forme analytique, les distances d’Hellinger associées à la loi de Rayleigh et à la loi exponentielle.

Loi	forme analytique
Loi normale	$D_H(p_{G_1}, p_{G_2}) = \sqrt{1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right)}}$
Loi de Weibull	$D_H(p_{W_1}, p_{W_2}) = \sqrt{1 - \frac{2(\alpha_1\alpha_2)^{\beta/2}}{\alpha_1^\beta + \alpha_2^\beta}}$

TABLEAU 1.4 – Formes analytiques de la distance d’Hellinger pour des lois de probabilité usuelles.

## 1.4 Conclusion du Chapitre

Au travers de ce premier chapitre, nous avons exposé le contexte dans lequel se positionne cette thèse.

Nous avons montré que les données de l’imagerie satellitaire que nous voulons exploiter sont fortement imparfaites, c’est-à-dire à la fois imprécises et incertaines.

Nous avons, ensuite, évoqué trois théories de l’incertain, parmi les différentes théories existantes dans la littérature. Elles constituent pour nous des candidates sérieuses pour traiter ce type de données grâce aux formalismes mathématiques qu’elles proposent.

Nous nous sommes, finalement, attardés sur la modélisation à l’aide de distributions de probabilité de variables aléatoires réelles et sur les mesures de dissemblances que l’on peut définir pour comparer deux distributions entre elles.

Dans le second chapitre de cette thèse, nous présenterons et proposerons des outils issus de la théorie des possibilités pour exploiter les données imparfaites de l’imagerie satellitaire.



# Chapitre 2

## Transformations de probabilité-possibilité unidimensionnelle

*« Echouer, c'est avoir la possibilité  
de recommencer de manière plus  
intelligente. »*

---

Henry Ford

### Sommaire

---

<b>1.1 Problématique de la modélisation des données</b> . . . . .	<b>9</b>
1.1.1 Données & informations . . . . .	10
1.1.2 Les images numériques . . . . .	11
1.1.2.1 Les types d'images numériques . . . . .	11
1.1.2.2 Histogramme d'une image numérique . . . . .	13
1.1.3 Acquisition des images numériques . . . . .	14
1.1.3.1 Imagerie satellitaire optique . . . . .	15
1.1.3.2 Imagerie satellitaire RADAR . . . . .	18
1.1.4 Imperfection des données . . . . .	22
<b>1.2 Représentations des données imparfaites</b> . . . . .	<b>23</b>
1.2.1 Théorie des probabilités . . . . .	24
1.2.2 Théorie des ensembles flous . . . . .	26
1.2.3 Théorie des possibilités . . . . .	28

<b>1.3 Construire une densité de probabilité à partir des données</b> . . . . .	<b>31</b>
1.3.1 Estimation des paramètres d'une variable aléatoire . . . . .	32
1.3.1.1 Espérance, moment & moyenne empirique . . . . .	33
1.3.1.2 Moment centré & variance empirique . . . . .	35
1.3.1.3 Justification de la méthode des moments . . . . .	36
1.3.2 Densités usuelles de probabilité uni-variées . . . . .	36
1.3.2.1 Loi Normale et loi Gaussienne Généralisée . . . . .	37
1.3.2.2 Loi de Weibull, loi de Rayleigh et loi exponentielle . . . . .	39
1.3.3 Comparaison de distributions de probabilité . . . . .	40
1.3.3.1 Distances . . . . .	41
1.3.3.2 Divergences . . . . .	42
<b>1.4 Conclusion du Chapitre</b> . . . . .	<b>45</b>

---



## Introduction

Dans le premier chapitre de cette thèse, nous avons évoqué des méthodes de modélisation d'ensembles de données basées sur l'utilisation d'outils probabilistes. Cette démarche repose sur l'hypothèse que chaque donnée observée est une réalisation d'une variable aléatoire et qu'on peut alors décrire le jeu de données complet grâce à une distribution de probabilité propre à cette variable aléatoire.

Deux méthodes permettant ce type de représentation ont alors été détaillées. La première méthode est la modélisation discrète des données. Elle consiste simplement à normaliser l'histogramme des valeurs des pixels afin de calculer une distribution de probabilité empirique. Cette méthode présente toutefois plusieurs inconvénients comme la conservation de toutes les irrégularités de la distribution ou la dépendance au pas d'échantillonnage de l'histogramme. La seconde méthode, sur laquelle nous nous sommes plus particulièrement attardés, est la modélisation paramétrique probabiliste continue. Elle se base sur l'estimation des moments de la variable aléatoire pour déterminer la loi de probabilité sous-jacente. En théorie, cette méthode permet de retrouver la "vraie" distribution qui est bien souvent continue. L'inconvénient de la représentation probabiliste continue est que le modèle paramétrique n'est pas assez flexible pour réellement déterminer la vraie distribution si l'échantillon à disposition est de trop faible taille ou lorsque les données sont très bruitées.

En réponse à ce problème de modélisation, nous nous sommes alors tournés vers une autre théorie de l'incertain : la théorie des possibilités.

En effet, puisque certaines contraintes propres aux probabilités sont relâchées tandis que l'essence de la modélisation probabiliste est conservée, les distributions de possibilités constituent un modèle plus souple et de ce fait plus adéquat pour modéliser les données imparfaites.

Ainsi, dans ce chapitre, nous allons étudier différentes manières de construire un modèle possibiliste à partir d'une distribution de probabilité et nous tenterons à chaque fois d'appliquer les modèles obtenus en situation afin d'observer leurs comportements. Nous commencerons dans une première section par définir les principes régissant le passage d'une modélisation à l'autre avant de présenter les différentes transformations

probabilité-possibilité proposées dans la littérature scientifique. Ces transformations seront dans un premier temps appliquées à des distributions discrètes et nous les appliquerons donc sur des distributions de probabilités empiriques. Puis, dans un second temps, nous exposerons la transformation au cœur de cette thèse : la transformation continue proposée par Dubois, Foulloy, Mauris et Prade dans [7] et nous verrons notamment que cette transformation possède des propriétés intéressantes pour nos besoins.

Par ailleurs, nous testerons en situation nos modèles possibilistes. Il est, pour cela, nécessaire de définir des mesures de comparaisons entre distributions de possibilités. Nous présenterons donc différentes mesures basées à la fois sur la théorie des ensembles flous et sur la théorie des probabilités.

## 2.1 Transformation : Définition et principes

La transformation d'une distribution de probabilité en une distribution de possibilité est une question complexe qui a pendant un temps opposé deux visions [57]. La première vision [37], qui fait aujourd'hui globalement consensus, suppose une dissymétrie dans la transformation de probabilité-possibilité. Elle suppose que, puisque le modèle possibiliste est moins informatif, le passage d'une distribution de probabilité à une distribution de possibilité implique une perte de précision et donc d'information. Dans l'autre sens, la conversion inverse implique un ajout d'information que l'on peut qualifier d'arbitraire. La seconde vision suppose davantage d'équivalence entre les modèles de probabilités et les modèles de possibilités. Elle propose alors que la transformation dans un sens comme dans l'autre doit se faire sous la condition que l'on conserve l'information totale.

De ces deux visions ont découlé des principes afin de guider le choix de la transformation probabilité-possibilité à utiliser. Ainsi nous allons définir, dans cette section, ces principes qui nous permettront de garantir que la distribution de possibilité construite depuis une distribution de probabilité vérifie les propriétés désirées par rapport à la conception de l'incertitude que l'on a dans nos données.

### 2.1.1 Principe de consistance

Le premier principe que nous allons évoquer est celui énoncé par Zadeh lui-même dans son papier séminal [3]. Il décrit de façon linguistique ce que l'on appelle la consistance de la transformation probabilité-possibilité et que l'on peut résumer ainsi :

*Un degré de probabilité élevé implique un degré de possibilité élevé et un degré faible de possibilité implique un degré faible de probabilité.*

Zadeh définit, de plus, un degré de consistance entre possibilité et probabilité par une mesure  $C$  telle que :

$$C(\pi, p) = \sum_{i=1}^n \pi(\omega_i) * p(\omega_i). \quad (2.1)$$

Cependant, Zadeh dira de ce degré de consistance qu'il est davantage une formalisation de la connexion heuristique entre probabilité et possibilité plutôt qu'un critère précis. Cette mesure peut toutefois permettre d'aider à comparer des transformations de probabilité-possibilité dans le but de choisir la transformation la plus spécifique comme nous le montrerons dans la suite de ce chapitre.

Enfin, une manière élégante de généraliser ce qui a été dit précédemment et définissant le principe fondamental de consistance des transformations probabilité-possibilité a été énoncé dans [58] par Dubois et Prade. Dans ce papier, ils montrent qu'une mesure de possibilité  $\Pi$  est équivalente à une famille  $\mathcal{P}(\Pi)$  de probabilités tel que :

$$\mathcal{P}(\Pi) = \{\mathbb{P}, \forall A \text{ mesurable}, \mathbb{P}(A) \geq \Pi(A)\}. \quad (2.2)$$

Cela revient à dire qu'une possibilité est consistante avec une probabilité si la mesure de possibilité  $\Pi$  associée domine la mesure de probabilité  $\mathbb{P}$  sur tout sous-ensemble  $A$  de  $\Omega$ . Cela garantit que l'on ne modifie pas la nature de l'information lorsque l'on applique la transformation dans le sens de "p vers  $\pi$ " ( $p \rightarrow \pi$ ).

### 2.1.2 Principe de préservation de l'ordre

Le second principe, que l'on appelle parfois principe de préservation de la préférence [59], a été défini par Dubois et Prade dans [60]. Il se distingue sous deux formes : la préservation de la préférence faible et la préservation de l'ordre forte.

Pour la préservation faible de la préférence [7], nous avons l'implication suivante entre la distribution de probabilité et la distribution de possibilité :

$$\forall \omega_i, \omega_j \in 2^\Omega, \quad p(\omega_i) > p(\omega_j) \Rightarrow \pi(\omega_i) > \pi(\omega_j). \quad (2.3)$$

Cette implication devient alors une équivalence sur l'ordre dans le cas de la préservation forte de la préférence :

$$\forall \omega_i, \omega_j \in 2^\Omega, \quad p(\omega_i) > p(\omega_j) \Leftrightarrow \pi(\omega_i) > \pi(\omega_j). \quad (2.4)$$

L'idée sous-jacente à ce principe est de garantir la conservation de la forme globale de la distribution estimée. On peut alors montrer de façon triviale pour une variable aléatoire uni-modale que si  $\omega^m$  est le mode de la distribution de probabilité alors  $\omega^m$  doit aussi être le mode de la distribution de possibilité obtenue après transformation.

### 2.1.3 Principe de conservation de l'incertitude

Le dernier principe que nous allons présenter est celui de conservation de l'incertitude énoncé par Klir [61], [62].

Avant d'énoncer ce principe, il nous faut considérer trois notions. La première, que nous avons présentée précédemment dans la section 1.2.1, est l'entropie  $H(p)$  associée à une distribution de probabilité  $p$  dont l'expression est donnée dans l'équation (1.13).

La seconde notion, définie dans [63], est la mesure de non-spécificité  $N(\pi)$  associée à une distribution de possibilité  $\pi$  :

$$N(\pi) = \sum_{i=2}^n (\pi_{\sigma(i)} - \pi_{\sigma(i+1)}) \log_2(i) \quad (2.5)$$

où  $\sigma$  est une permutation des indices  $\{1, 2, \dots, n\}$  telle que  $\pi_{\sigma(1)} > \pi_{\sigma(2)} > \dots > \pi_{\sigma(n)}$ .

Enfin, la troisième notion, énoncée elle aussi dans [63], est la mesure de discordance  $S(\pi)$  telle que :

$$S(\pi) = \sum_{i=2}^n (\pi_{\sigma(i)} - \pi_{\sigma(i+1)}) \log_2 \left( \frac{i}{\sum_{j=2}^i \pi_{\sigma(j)}} \right). \quad (2.6)$$

On définit alors le principe de préservation de l'incertitude où le but est de déterminer à l'aide d'une transformation probabilité-possibilité, une distribution de possibilité vérifiant l'égalité présentée ci-dessous :

$$H(p) = N(\pi) + S(\pi) \quad (2.7)$$

Toutefois, ce principe a été l'objet de nombreux débats dans la communauté scientifique [37] et cela pour plusieurs raisons. Une des raisons est que l'égalité au cœur de ce principe utilise des mesures d'incertitudes ensemblistes qui ne font pas consensus dans la communauté. On peut par exemple citer d'autres mesures d'information telle que la mesure de spécificité [64][65] ainsi que plus récemment la mesure d'entropie sur des ensembles flous [66]. Enfin, une seconde raison est que ce principe ne garantit pas la satisfaction de la consistance.

#### 2.1.4 Transformation $p \rightarrow \pi$

La transformation de probabilité-possibilité ( $p \rightarrow \pi$ ) est un problème qui a largement été étudié dès les premiers travaux de Zadeh en 1978 (voir [37][37][67][68]).

Dans ce qui suit, nous allons présenter les cinq transformations discrètes  $p \rightarrow \pi$  suivantes : *la transformation par normalisation* (TN), *la transformation log-interval scale* (TK), *la transformation variable* (TV), *la transformation optimale* (TO) et *la transformation symétrique* (TS).

Ces transformations ont retenu notre attention par rapport à deux critères : la simplicité d'implémentation dans une application et leur notoriété relative dans la littérature scientifique. Afin de simplifier les écritures, nous adopterons les notations suivantes dans le reste de cette section :

$$\forall i = \{1, \dots, n\}, \quad p(\omega_{\sigma(i)}) = p_i > 0, \quad p_i \geq p_{i+1}, \\ \pi(\omega_{\sigma(i)}) = \pi_i > 0, \quad \pi_i \geq \pi_{i+1}$$

où  $\sigma$  est la même permutation des indices que dans la sous-section 2.1.3.

#### **Transformation par normalisation :**

La transformation par normalisation est la plus simple parmi celles retenues. Elle est donnée par l'équation suivante :

$$\pi_i^{\text{TN}} = \frac{p_i}{p_1} \tag{2.8}$$

Cette transformation présente plusieurs avantages. Très simple à mettre en place, elle fonctionne aussi bien dans le cas discret que dans le cas où on considère un ensemble de référence  $\Omega$  infini. En ayant recours au mode de la distribution de probabilité, on peut obtenir des expressions de distributions de possibilité concises comme nous le verrons pour la distribution gaussienne (voir équation (2.41)).

Cependant, il a été démontré dans [69] que cette transformation ne satisfait généralement pas le principe de consistance énoncé dans l'équation (2.2).

### Transformation log-interval scale :

La transformation log-interval scale a été présentée par Klir dans [63] et a été construite de façon à respecter le principe de préservation de l'incertitude énoncé en section 2.1.3. Elle se définit ainsi :

$$\pi_i^{\text{TK}} = \left( \frac{p_i}{p_1} \right)^\alpha \quad (2.9)$$

avec  $\alpha$  appartenant à l'intervalle ouvert  $]0, 1[$  et vérifiant l'équation (2.7) qui prend alors la forme suivante :

$$-\sum_{i=1}^n p_i \log_2(p_i) = \sum_{i=2}^n \left( \left( \frac{p_i}{p_1} \right)^\alpha - \left( \frac{p_{i+1}}{p_1} \right)^\alpha \right) \log_2 \frac{i^2}{\sum_{j=1}^i \left( \frac{p_j}{p_1} \right)^\alpha} \quad (2.10)$$

De même que pour la transformation TN, cette transformation ne respecte pas toujours le principe de consistance comme il a été démontré dans [70]. On perd alors la propriété qu'une mesure de possibilité domine une mesure de probabilité et que la borne inférieure de la mesure de probabilité est la mesure de nécessité (voir équation (1.22)).

### Transformation variable :

La transformation variable a été proposée par Mouchaweh *et al.* [71] dans le sens  $p \rightarrow \pi$  et se présente de façon similaire à la transformation de Klir :

$$\pi_i^{\text{TV}} = \left( \frac{p_i}{p_1} \right)^{k \cdot (1-p_i)} \quad (2.11)$$

où  $k$  est une constante appartenant à l'intervalle  $0 \leq k \leq \frac{\log p_n}{(1-p_n) \log \left( \frac{p_n}{p_1} \right)}$  et où  $p_n$  est la dernière valeur de  $p$  non nulle. Le terme  $k(1-p_i)$  n'est donc pas une constante et dépend de  $p_i$  pour tout  $i \in [1, n]$ .

Dans ce même article, les auteurs montrent que leur transformation vérifie le principe de consistance suivant :

$$\forall \omega \in \Omega, \quad \pi^{\text{TV}}(\omega) \geq p(\omega). \quad (2.12)$$

Malheureusement, le respect de ce principe de consistance n'implique pas le respect du principe de consistance de Dubois et Prade énoncé dans 2.2 car celui-ci ne porte seulement que sur les singletons. Ce principe de consistance est à rapprocher de la vision de Klir où l'on suppose que l'information probabiliste et possibiliste sont comparables et que

l'on peut choisir indifféremment l'un ou l'autre de ces modèles. Or, comme nous l'avons évoqué précédemment, la vision consensuelle de la théorie des possibilités suppose une certaine dissymétrie entre les représentations : le modèle possibiliste est moins précis et donc moins informatif que le modèle probabiliste.

### **Transformation optimale :**

La transformation optimale a été proposée par Dubois, Prade et Sandri [37] en partant du postulat qu'une distribution de possibilité doit encapsuler un minimum de distributions de probabilité pour être le maximum spécifique. On obtient alors la transformation suivante de  $p \rightarrow \pi$  :

$$\pi_i^{\text{TO}} = \sum_{j=i}^n p_j. \quad (2.13)$$

Cette transformation est par définition celle qui retourne la distribution de possibilité la plus spécifique en respectant le principe de consistance[7].

### **Transformation symétrique (ou Transformation inverse de la Transformée pignistique) :**

La transformation symétrique a été énoncée par Dubois et Prade [60] puis a été redémontrée à partir de la théorie de l'évidence par Yamada [68]. Cette transformation respecte le principe de consistance ainsi que le principe de la préservation de la préférence. Elle est définie comme ceci :

$$\pi_i^{\text{TS}} = i \cdot p_i + \sum_{j=i+1}^n p_j \quad (2.14)$$

avec  $\sum_{j=n+1}^n p_j = 0$  et  $\pi_{n+1} = 0$ .

À noter que cette transformation possède une variante dans le cas où l'on considère une densité de probabilité continue. Elle s'énonce alors de la façon suivante :

$$\pi(x) = \int_{-\infty}^{+\infty} \min(p(x), p(t)) dt.$$

Cependant, nous n'utiliserons pas cette transformation dans la section concernant la modélisation possibiliste continue pour deux raisons. La première est que cette transformation est plus imprécise que la transformation optimale [57] et la seconde est que l'opérateur *min* n'est pas implémentable dans un algorithme sans repasser en discret.

### **Discussion sur les transformations $p \rightarrow \pi$**

L'intérêt d'utiliser la théorie des possibilités dans cette thèse est avant tout de proposer

des alternatives aux modèles probabilistes, que l'on juge trop précis pour représenter correctement les données satellitaires.

Dans cette optique, il semblerait intéressant de considérer seulement les transformations respectant le principe de consistance de l'équation (2.2) énoncé par Dubois et Prade. C'est-à-dire les transformations optimales (TO) et symétriques (TS). On aurait alors à disposition un modèle robuste, mais qui n'est pas toujours implémentable sans passer par une représentation discrète.

Il nous a donc semblé intéressant d'étudier aussi d'autres transformations comme la transformation par normalisation. Certes elles ne respectent pas toujours le principe de consistance, mais elles peuvent se révéler utiles pour notamment simplifier dans certains cas les formulations des densités de possibilités (voir par exemple l'équation (2.41)).

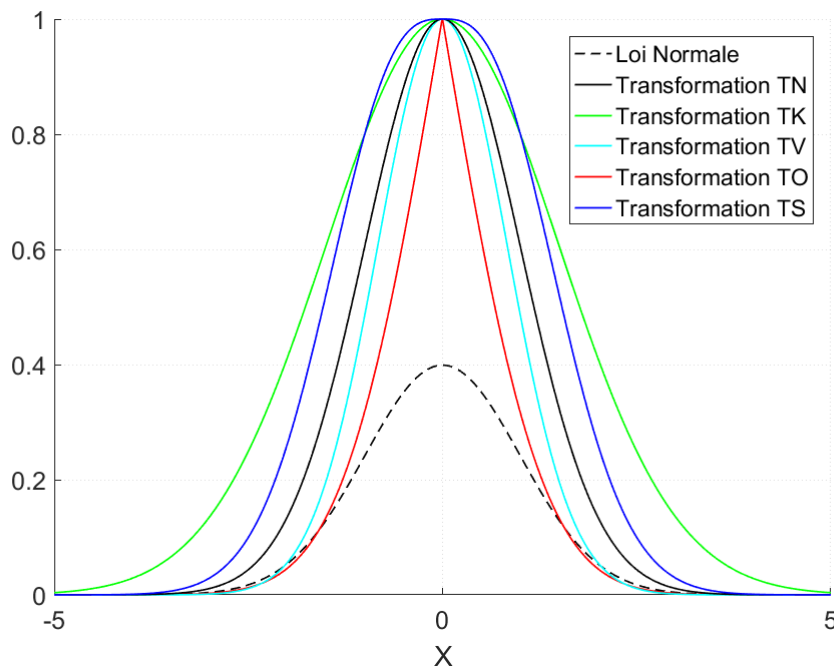


FIGURE 2.1 – Application des transformations  $p \rightarrow \pi$  présentées sur un ensemble de 200 échantillons d'une loi normale de moyenne nulle et de variance égale à 1.

Enfin, nous proposons dans la figure 2.1 une illustration de chacune des transformations précédemment évoquées appliquées à une distribution de probabilité discrète (en pointillés noirs) de moyenne nulle. Dans le cas de la transformation TK, le paramètre  $\alpha$  a été déterminé en résolvant l'équation (2.7). Pour la transformation TV, le paramètre  $k$  a été choisi comme égal à  $\frac{\log(p_n)}{(1-p_n)\log(p_n/p_1)}$ .



Nous donnons dans le tableau 2.1 à titre informatif les degrés de consistance et les indices de spécificité calculés pour chacune des transformations.

Transformations	TN	TK	TV	TO	TS
Degré de Consistance	0,7071	0,8322	0,6404	0,5014	0,8183
Indice de Spécificité	2,5066	3,7586	2,0894	1,6008	3,1915

TABLEAU 2.1 – Degré de Consistance et Indice de spécificité pour chaque transformation appliquée à une loi normale :  $\mathcal{N}(0, 1)$ .

On peut voir dans ce tableau deux choses intéressantes. La première est que l'indice de spécificité est complètement corrélé avec le degré de consistance. C'est-à-dire que si l'on ordonne les transformations suivant un ordre croissant d'indice de spécificité on retrouve le même ordre suivant les degrés de consistance. Il n'y a donc pas d'intérêt de considérer davantage le degré de consistance que l'indice de spécificité. Le second phénomène remarquable est que les transformations renvoient des résultats significativement différents. En effet, l'indice de spécificité peut être multiplié par deux suivant le choix de la transformation, comme on peut le voir entre la transformation la plus spécifique (ici TO) et la moins spécifique (ici TK).

### 2.1.5 Transformation $\pi \rightarrow p$

Nous allons présenter dans cette section, les transformées inverses associées aux transformées précédentes c'est-à-dire de  $\pi \rightarrow p$ . Ces transformations peuvent présenter de l'intérêt dans différentes situations telles que la prise de décision [72] où il est plus facile pour un expert de choisir à l'aide d'informations probabilistes [73].

À noter que chaque transformation précédemment proposée ne possède pas forcément de transformée inverse. Nous évoquerons donc seulement trois transformations  $\pi \rightarrow p$ .

Tout comme dans le sens  $p \rightarrow \pi$ , les transformations de  $\pi \rightarrow p$  sont sujettes à des principes afin de garantir que l'on ne modifie pas outre mesure l'information globale. On peut par exemple citer le principe équivalent au principe de conservation de la préférence (plus de précision dans [73]).

### Transformation par normalisation inverse :

La transformation TN inverse est un cas intéressant de transformation  $\pi \rightarrow p$ . En effet, si on applique à une distribution  $p$  la transformation TN puis à la distribution  $\pi_{\text{TN}}$  résultante la transformée TN inverse on obtient alors une distribution de probabilité  $p_{\text{TN}}$  telle que  $p = p_{\text{TN}}$ . La transformation TN inverse est donnée par cette équation :

$$p_i = \frac{\pi_i}{\sum_{j=1}^n \pi_j}. \quad (2.15)$$

### Transformation inverse de la transformée de Klir :

De même que dans le cas de la transformation TN, la transformée de Klir est symétrique dans le sens où elle propose une transformation pour passer d'une distribution de possibilité à une distribution de probabilité. Cependant, encore une fois il faut déterminer la valeur du paramètre  $\alpha$  tel que  $\alpha$  est solution du principe de conservation de l'incertitude. La transformation de Klir de  $\pi \rightarrow p$  prend la forme suivante :

$$p_i = \frac{\pi_i^{\frac{1}{\alpha}}}{\sum_{j=1}^n \pi_j^{\frac{1}{\alpha}}}. \quad (2.16)$$

À noter ici que les deux transformations  $\pi \rightarrow p$  sont basées sur un même principe de normalisation que celui nous permettant d'obtenir des distributions de probabilités empiriques à partir d'un histogramme.

### Transformation symétrique inverse (ou Transformation pignistique) :

Enfin la dernière transformation  $\pi \rightarrow p$  que nous reprendrons dans cette section a été proposée au même moment que la transformation TO proposée par Dubois et Prade. Cette transformation s'énonce ainsi :

$$p_i = \sum_{j=1}^n \frac{1}{j} (\pi_j - \pi_{j+1}) \quad (2.17)$$

avec la notation  $\pi_{n+1} = 0$ .

## 2.2 Première application : Comparaison discrète de distributions de possibilité

Nous avons présenté dans les premières parties du chapitre 2 quelques transformations de probabilité-possibilité discrètes. Puis, à l'aide des différents principes, du degré

de consistance et de moindre spécificité, nous avons mis en lumière les différences entre ces transformations sur un plan théorique.

Ce que nous proposons dans cette section, c'est d'apporter de nouveaux arguments de comparaison en contextualisant ces transformations dans une application relativement simple de détection de changements entre des paires d'images artificielles afin de contrôler au maximum le processus.

De plus, cette application nous permet d'introduire plusieurs notions : les mesures de dissemblance entre distributions de possibilité, la détection de changements entre paires d'images et la quantification de la performance en détection de changements. Ces notions nous les retrouverons d'ailleurs plusieurs fois au cours de ce chapitre 2 et du chapitre 3 sous diverses formes.

### 2.2.1 Comparaison de distributions de possibilité

La question de la comparaison de distributions de possibilité peut être abordée sous différents angles. Pour des raisons de simplicité, on va retenir dans un premier temps essentiellement deux grandes approches. D'un côté nous aurons l'approche ensembliste où l'on considère une distribution de possibilité comme un sous-ensemble flou que l'on compare à un second à l'aide d'une mesure de similarité (ou dissimilarité) [74][75]. De l'autre côté, nous aurons l'approche plus vectorielle (ou fonctionnelle) où l'on utilise des distances (cf section 1.3.3.1) afin de quantifier l'écart entre deux fonctions [76] que l'on normalise afin que ces distances respectent aussi les propriétés des mesures de dissimilarité.

Les mesures de similarité sont devenues un outil de base de la logique floue qui a été l'objet de nombreuses études dans la littérature scientifique [77][78][79]. Nous allons présenter dans ce qui suit les principales définitions ainsi que quelques mesures très souvent utilisées dans les applications.

**Définition 18.** *Mesure de similarité :*

Soit  $\mathcal{F}$  un ensemble et  $A, B$  deux sous-ensembles flous tels que  $A, B \subset \mathcal{F}$ . La fonction  $S : \mathcal{F} \times \mathcal{F} \rightarrow [0, 1]$  est appelée une mesure de similarité si et seulement si  $S$  respecte les propriétés suivantes :

— *Réflexivité :*  $S(A, B) = 1 \Leftrightarrow A = B$ ;

- *Symétrie* :  $S(A, B) = S(B, A)$  ;
- *Transitivité* : si  $A \subseteq B \subseteq C$ , alors  $S(A, B) \geq S(A, C)$  ;
- *Recouvrement* : Si  $A \cap B = \emptyset$  alors  $S(A, B) = 0$  sinon  $S \in ]0, 1]$ .

**Proposition 3.** *Mesure de dissimilarité [80] :*

*Une mesure de similarité et une mesure de dissimilarité sont corrélées deux à deux. De plus une mesure de dissimilarité  $D$  et une mesure de similarité  $S$  sont reliées entre elles par la propriété suivante :  $D + S = 1$ .*

La première mesure de dissimilarité que nous considérons est la distance de Jaccard que l'on retrouve notamment dans [70]. Elle se calcule à partir du rapport entre le cardinal de l'intersection des deux sous ensembles flous considérés  $A$  et  $B$  définis par leurs fonctions d'appartenance respective  $\mu_A, \mu_B$  et le cardinal de leur union. On l'énonce de la façon suivante :

$$\forall \omega \in \Omega \quad D_J(A, B) = 1 - \frac{\sum_{i=1}^n \min(\mu_A(\omega_i), \mu_B(\omega_i))}{\sum_{i=1}^n \max(\mu_A(\omega_i), \mu_B(\omega_i))}. \quad (2.18)$$

La seconde mesure de dissimilarité que nous présentons est inspirée d'une norme L1 (voir tableau 1.2) entre deux fonctions d'appartenance, normalisée afin de respecter les propriétés des mesures de dissimilarité [75] :

$$\forall \omega \in \Omega \quad D_{L1}(A, B) = \frac{\sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)|}{\sum_{i=1}^n (\mu_A(\omega_i) + \mu_B(\omega_i))}. \quad (2.19)$$

Enfin, la dernière mesure retenue est la distance euclidienne normalisée avec le même principe de normalisation que la mesure de dissimilarité  $D_{L1}$ . On la note  $D_{L2}$  et elle s'écrit ainsi :

$$\forall \omega \in \Omega \quad D_{L2}(A, B) = \sqrt{\frac{\sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)|^2}{\sum_{i=1}^n (\mu_A(\omega_i) + \mu_B(\omega_i))^2}}. \quad (2.20)$$

### 2.2.2 Détection de changements entre paires d'images artificielles

Nous profitons de cette première application pour présenter la détection de changements entre images. Dans cette section nous allons définir puis brièvement résumer l'état de l'art de cette discipline à part entière de la télédétection.

Avant de d'introduire cette application, nous tenons à préciser notre démarche. Tout d'abord, nous sommes conscients que les résultats des applications présentées dans ce

chapitre ne permettent pas de donner une conclusion générale, nous les considérerons donc en tant que tel c'est à dire des éléments de discussion pour répondre au problème de la modélisation des données. Ici, le propos est purement tourné vers l'illustration des méthodes que nous présentons. En effet, il nous a paru que grâce à des algorithmes de détection de changements, nous pouvions complètement exposer le processus d'extraction d'informations. C'est à dire : représenter un ensemble de données, le comparer à l'aide d'une mesure à un second ensemble et enfin en déduire des informations. De plus, comme nous le verrons dans la section suivante, la détection de changements peut se révéler un cadre idéal pour quantifier la performance de nos méthodes, mais aussi pour pouvoir évaluer le comportement de celle-ci en présence de bruit de plus ou moins forte intensité.

La détection de changements en télédétection consiste à détecter les régions (ici les pixels) ayant changé entre deux prises de vues d'une même scène à des instants différents [81]. Les changements que l'on cherche à identifier sont de nature très variée. Cela peut-être l'apparition ou la destruction d'un bâtiment [82], la fonte des glaces sur les pôles [83], la déforestation [84], etc.

Initialement, la détection de changement impliquait que les images que l'on compare soient recalées. C'est-à-dire que les images sont mises en correspondance (voir [85] pour plus de détails) afin que les pixels d'intérêts (sommet d'une montagne, routes, maisons...) aient la même position dans les deux images. Désormais, des techniques plus sophistiquées, basées notamment sur des réseaux de convolution [86], peuvent effectuer de la détection de changements sur tous types d'images y compris non recalées.

Dans cette première application, ainsi que dans la seconde application de détection de changements, nous avons voulu travailler sur des paires d'images artificielles. L'intérêt de ces images est que nous pouvons choisir la taille, la géométrie, et le nombre de changements que nous voulons détecter. De plus, nous connaissons avec une grande certitude l'emplacement de ces changements, ce qui nous aidera ensuite à évaluer la performance.

### 2.2.2.1 Fonctionnement de l'algorithme de détection de changements

L'idée de la première application est, comme nous l'avons déjà mentionné, de tenter d'ajouter des arguments afin de nous guider dans le choix de l'une ou l'autre des transformations. Ainsi nous avons voulu que l'application soit relativement simple afin de limiter les biais de comparaison.

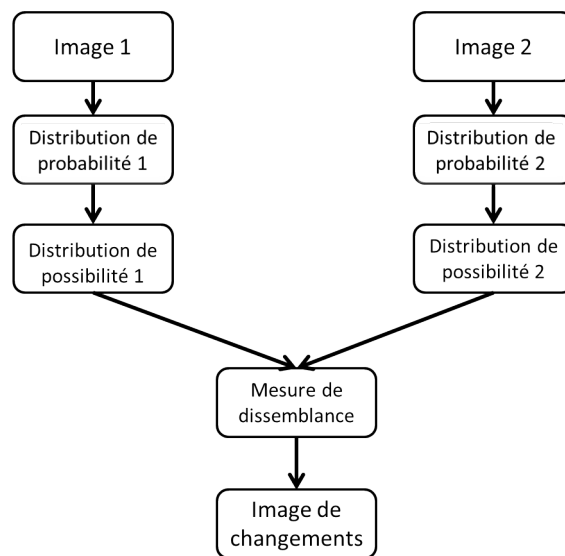


FIGURE 2.2 – Fonctionnement de l'algorithme de détection de changements

Dans cette application, on considère que l'image 1 est l'image avant qu'il ne se soit produit des changements et l'image 2 celle où les changements sont apparus.

Pour procéder à la détection de changements, on sélectionne une imagerie (ou patch) de tailles  $5 \times 5$  dans l'image 1. Cette taille de patches est choisie afin de détecter des changements très localisés dont la taille peut-être de l'ordre de quelques pixels. Puis, on construit l'histogramme des données à partir des valeurs d'intensité des 25 pixels contenus dans le patch. Cet histogramme est ensuite normalisé par la somme de ses valeurs pour obtenir une distribution de probabilité (empirique) représentant l'imagerie. On applique alors une des transformations discrètes à la distribution de probabilité pour obtenir une distribution de possibilité. On répète le même processus avec la même transformation  $p \rightarrow \pi$  afin d'obtenir une seconde distribution de possibilité modélisant une imagerie extraite de l'image 2. Puis nous comparons les distributions ainsi obtenues par l'une des mesures de dissemblance présentées en section 2.2.1. On stocke alors le résultat à la position du pixel central des patches considérés dans l'image de changement qui est une matrice de mêmes dimensions que les images originales.

Enfin, le processus est reproduit en décalant le pixel central des patchs d'un pixel jusqu'à ce que les deux images aient été complètement comparées en chaque pixel.

### 2.2.2.2 Images artificielles

Les images 1 et 2 sont des images de  $150 \times 250$  pixels à une composante dont les valeurs minimales et maximales varient entre 0 et 1 (image normalisée). Sur chacune d'entre elles sont disposées 2 ellipses servant à représenter le signal original, c'est-à-dire un élément fixe. Puis sur l'image 2 sont ajoutés des changements sous forme de disque d'un diamètre de 8 à 15 pixels dans différentes zones de l'image. Certains de ces disques recouvrent le signal original, comme on peut le voir dans les figures figure 2.3 et figure 2.4.

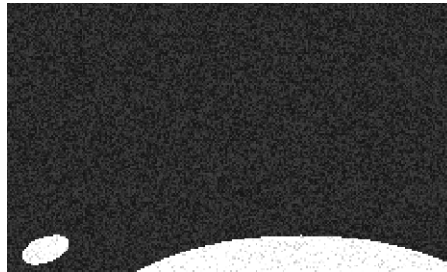


FIGURE 2.3 – image 1 avec un bruit de chatoiement de variance  $\nu = 0.1$

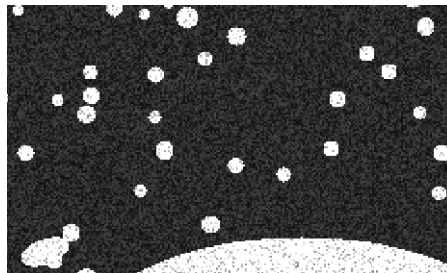


FIGURE 2.4 – image 2 avec un bruit de chatoiement de variance  $\nu = 0.15$

Chacune des images est bruitée quasi identiquement en utilisant au choix soit un bruit de type additif tel que le gaussien soit un bruit multiplicatif comme le "speckle noise" ou bruit de chatoiement. On rappelle que le bruit gaussien est un bruit que l'on retrouve fréquemment dans l'imagerie satellitaire optique tandis que le speckle noise est propre à l'intensité des images radar.

Si l'on appelle  $I$  l'image avant bruitage et  $I'$  l'image bruitée, alors pour un appliquer un bruit gaussien de moyenne  $M$  et de variance  $\nu_g$  on utilise la formule suivante :

$$I' = I + \sqrt{\nu_g} \times \text{randn}(\text{size}(I)) + M, \quad (2.21)$$

où  $\text{randn}(\text{size}(I))$  est une matrice de même dimension que  $I$  de nombre distribué aléatoirement suivant une loi normale entre 0 et 1. Dans notre application  $M = 0$  de telle sorte que la moyenne du signal ne soit pas impactée.

Pour appliquer un bruit de chatolement de moyenne nulle et de variance  $\nu_s$  on utilise l'équation suivante :

$$I' = I + I \times N, \quad (2.22)$$

où  $N$  est un bruit uniformément distribué de moyenne nulle et de variance  $\nu_s$ . Enfin, de sorte à obtenir une application suffisamment proche de la réalité, les variances des bruits appliquées aux images 1 et 2 sont légèrement différentes de telle sorte que pour un bruit gaussien on a :

$$\nu_{im1} = \nu_g,$$

$$\nu_{im2} = \nu_g + 0,05.$$

De même dans le cas du bruit de chatolement :

$$\nu_{im1} = \nu_s,$$

$$\nu_{im2} = \nu_s + 0,05.$$

### 2.2.3 Quantification de la performance : courbe ROC

Pour une transformation et une mesure de dissimilarité choisie, nous produisons en sortie de notre algorithme une image des changements détectés (cf figure 2.2). Ensuite, grâce à cette image (figure 2.5a) et à une image de la position réelle des changements, que l'on appelle masque de vérité terrain (figure 2.5b), nous pouvons évaluer un taux de détection en pourcentage reflétant la performance globale de notre méthode. Cet outil de quantification de la performance se dénomme la courbe ROC (Receiver Operating Curve) et nous en donnons les caractéristiques ci-après.

La courbe ROC est un outil couramment utilisé en statistique et en traitement d'image pour mesurer les performances d'un classifieur binaire. Ce qui est effectivement le cas dans notre application où l'image des changements renseigne en chaque pixel la valeur de dissimilarité entre l'image 1 et l'image 2 avec pour raisonnement que plus la valeur du pixel est élevée plus il y a de certitude de la présence d'un changement. Au contraire, plus la valeur est faible et moins on suppose qu'il y en ait un. Ainsi, le principe de la courbe



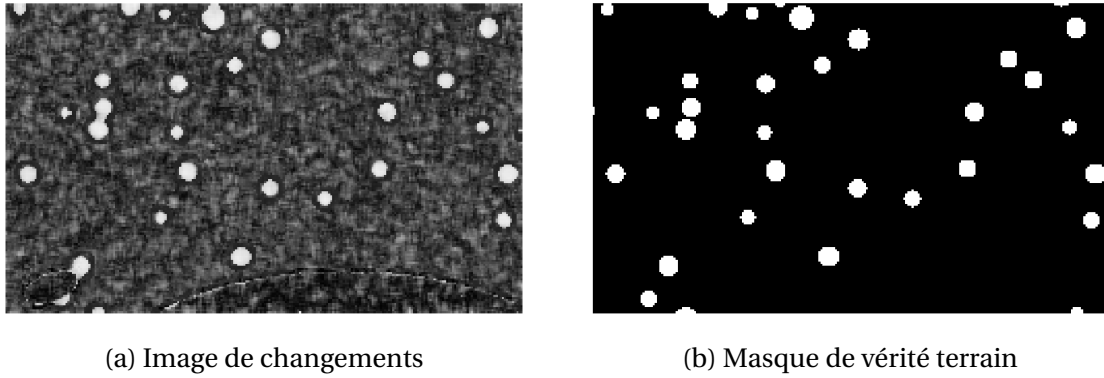


FIGURE 2.5 – Exemple d’image de changements obtenu en utilisant la transformation TS et la distance de Jaccard pour un bruit de chatoiement de variance  $\nu = 0.3$ .

ROC est le suivant : pour une certaine valeur de pixels on fixe un seuil. Avec une opération logique, on impose que toutes les valeurs au-dessus de ce seuil soient considérées comme un changement (mise à 1 de la valeur du pixel), celle en dessous comme un non-changement (mise à 0). À noter qu’on retrouve ici le classifieur binaire. Puis, avec cette image de changement seuillée et le masque de vérité terrain on construit une matrice de confusion où l’on comptabilise le nombre de vrais positifs (VP), de vrais négatifs (VN), et les mauvaises détections : faux positifs (FP) et faux négatifs (FN).

		Vérité		Total
		Positif	Négatif	
Résultat de Classification	Positif	VP	FP	VP + FP
	Négatif	FN	VN	FN + VN
Total		VP + FN	FP + VN	N

TABEAU 2.2 – Exemple de matrice de confusion.

A l’aide de cette matrice de confusion, on peut extraire de nombreuses informations sur notre classifieur telles que sa sensibilité, sa spécificité ou son taux d’erreurs (plus de précision dans [87]). Dans notre cas, pour pouvoir générer la courbe ROC on a besoin de calculer la précision  $E(s)$  et le rappel  $R(s)$  pour le seuil  $s$  considéré. La précision est aussi appelée probabilité de bonne détection et le rappel définit quant à lui la probabilité de

fausse alarme.

$$E(s) = \frac{VP(s)}{VP(s) + FP(s)} \quad (2.23)$$

$$R(s) = \frac{VP(s)}{VP(s) + FN(s)}. \quad (2.24)$$

On reproduit ces étapes pour un grand nombre de seuils et on peut alors construire la courbe ROC ( $E(s), R(s)$ ) pour  $s$  variant dans  $\mathbb{R}^+$  (ici dans  $[0, 1]$  puisque les mesures de dissimilarité prennent valeurs dans cet intervalle).

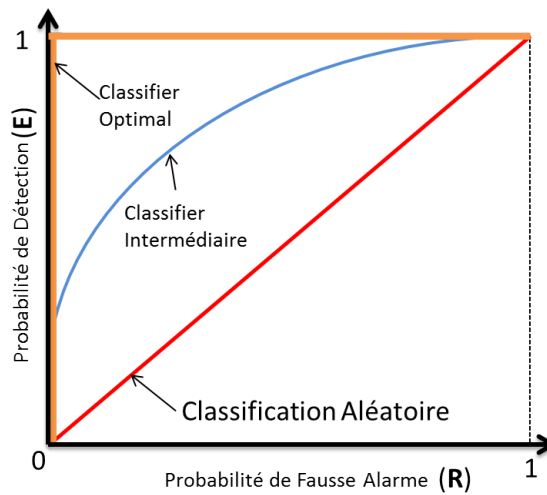


FIGURE 2.6 – Illustration d'une courbe ROC.

L'intérêt principal de la courbe ROC est la simplicité de son interprétation. Comme on peut le voir sur la figure 2.6, après évaluation du rappel et de la précision pour un grand nombre de seuils on obtient une courbe similaire à la courbe bleue. Plus cette courbe se rapproche du point (0,1) et meilleur est le classifieur. Au contraire, plus la courbe tend vers la diagonale (en rouge) et plus le résultat de la classification s'approche d'une classification aléatoire.

Mathématiquement, on peut tirer les mêmes conclusions en calculant l'aire sous la courbe ROC (AUROC). Le classifieur optimal aura alors une aire valant 1 c'est-à-dire 100% de bonne détection pour aucune fausse alarme. Le classifieur randomisé aura une aire de 0.5, ce qui signifie autant de bonnes que de mauvaises détections.

## 2.2.4 Résultats

Ci-après, nous présentons les résultats de l'application de détection de changements sous la forme de graphiques représentant en abscisse la variance du bruit appliqué (bruit photonique et bruit de chatoisement) et en ordonnée le pourcentage de bonnes détections. Le pourcentage est obtenu en multipliant par cent l'aire sous une courbe ROC.

Pour rappel, afin d'obtenir les graphiques de résultats, nous avons considéré cinq transformations  $p \rightarrow \pi$ , trois mesures de dissimilarité et deux types de bruits différents. Les résultats ont été regroupés suivant le type de bruit appliqué.

La figure 2.7 présente l'ensemble des performances des algorithmes de détection de changements calculés lorsque le bruit appliqué aux images artificielles est de type additif. Les graphiques 2.7a, 2.7b et 2.7c représentent les taux de bonne détection en fonction du bruit, obtenus respectivement avec la distance de Jaccard, la mesure  $D_{L1}$  et la mesure  $D_{L2}$  (voir section 2.2.1).

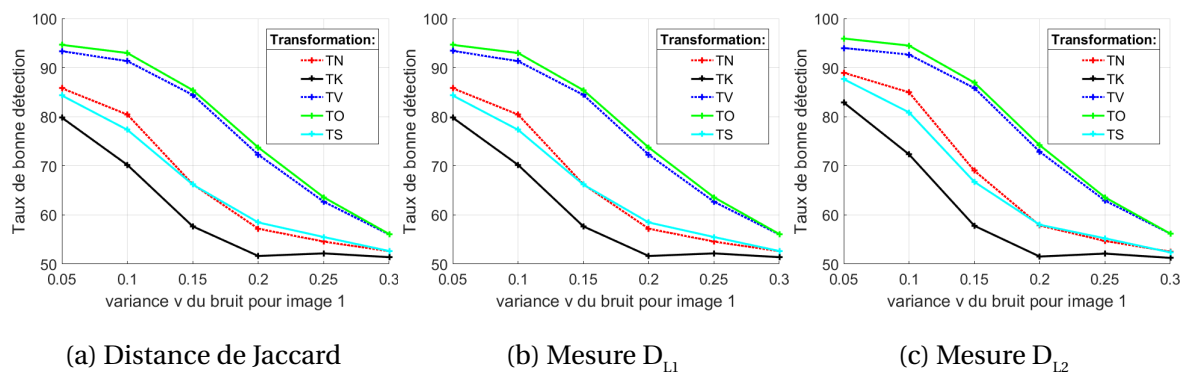


FIGURE 2.7 – Résultats obtenus avec différentes mesures de dissimilarité pour chaque transformation en fonction de la variance du bruit photonique appliqué.

Dans chacun des graphiques 2.7a, 2.7b et 2.7c, les mesures de similarités présentent des taux de performances quasi identiques pour chaque transformation. Cela semble suggérer que ces mesures de similarités observent la même dynamique et qu'il n'y aurait pas de préférence à en choisir une en particulier par rapport aux deux autres.

Les transformations TV et TO, appliquées à des distributions empiriques modélisant des données bruitées par un bruit gaussien, ont un pourcentage de bonne détection globalement supérieure aux trois autres transformations. Les transformations TK, TS et TN voient quant à elles leurs performances décroître plus rapidement que les deux autres transfor-

mations lorsque la variance du bruit gaussien appliqué augmente.

De la même façon que précédemment, la figure 2.8 présente les performances des détections de changements calculées lorsque l'on applique un bruit de chatoisement sur les images. Les graphiques 2.8a, 2.8b et 2.8c représentent les taux de bonne détection en fonction du bruit, obtenus respectivement avec la distance de Jaccard, la mesure  $D_{L1}$  et la mesure  $D_{L2}$ .

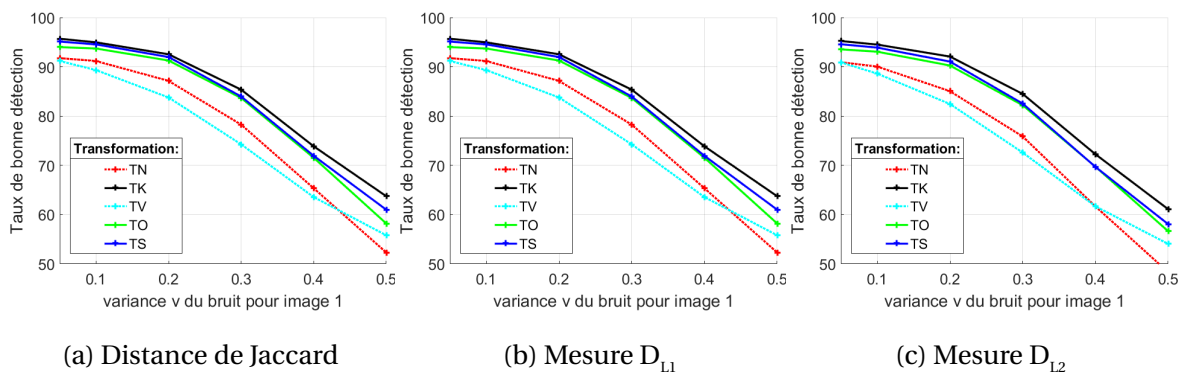


FIGURE 2.8 – Résultats obtenus avec différentes mesures de dissimilarité pour chaque transformation en fonction de la variance du bruit gaussien appliqué.

Lorsque les images artificielles sont bruitées avec un bruit de chatoisement, l'ensemble des transformations exhibent des performances comparables avec un taux de bonne détection supérieur à 90% quand la variance du bruit est faible, et une décroissance de leurs performances en fonction de l'augmentation de la variance du bruit de chatoisement. Les trois transformations TO, TK et TS semblent toutefois observer des taux de bonnes détections généralement supérieurs aux deux autres transformation TN et TV. Cette tendance semble se confirmer par le fait que les trois mesures de similarités proposent, encore une fois, une grande correspondance entre les résultats affichés.

## 2.2.5 Discussion

Quoique non exhaustive, l'idée de comparer les transformations de probabilité-possibilité au travers d'une application relativement simple de détection de changements entre images artificielles nous permet de tirer plusieurs conclusions sur les différentes transformations présentées.

Premièrement, la transformation optimale proposée par Dubois et Prade semble la plus

avantageuse à bien des égards. Sur le plan théorique, comme nous l'avons dit puis illustré précédemment, cette transformation à l'avantage de respecter le principe de consistance mais aussi d'être la plus spécifique parmi les transformations évoquées. Sur le plan applicatif, quel que soit le type de bruit appliqué, cette transformation fait partie dans les deux cas des transformations obtenant les meilleures performances. Cette transformation possède, de plus, une forme continue et sera donc la transformation que nous retiendrons pour la suite de cette thèse.

Pour les autres transformations, il est plus difficile de conclure. En effet, certaines transformations comme la transformation TV semblent mieux réagir lorsque le bruit est de type additif. Une explication possible serait à chercher du côté de l'impact du bruit sur la distribution des données. En effet, l'ajout d'un bruit photonique abouti à faire tendre la distribution des données vers une distribution de probabilité uni-modale proche de la loi gaussienne. Cela aurait pour effet de renforcer la transformation par log-normalisation dont l'exposant dépend de la valeur  $p_i$  pour chaque valeur de  $\pi_i$ .

Dans l'autre cas de figure, l'application d'un bruit multiplicatif tend à renforcer la présence des modes secondaires dans la distribution des données. Ainsi, certaines de ces transformations seraient alors plus adéquates pour modéliser ces distributions multimodales ou tout du moins, moins impactées par la présence des modes secondaires (par exemple TK ou TS).

La transformation par normalisation ne semble dans aucun des cas présenter les meilleures performances. Cependant, ses performances ne sont pas tant en deçà par rapport à la transformation optimale. Nous préconisons donc de ne pas la rejeter totalement car malgré le fait qu'elle ne respecte pas toujours le principe de consistance (parfois de peu), son implémentation est de loin la plus simple.

Enfin, les conclusions que nous tirons de cette application sur les transformations sont à relativiser car nous sommes conscients que d'autres paramètres peuvent changer la nature de ces conclusions : contenu des images, bruit... Nous pensons, toutefois, que l'application de détection de changements nous a permis d'apporter des arguments qui pourront permettre de guider les futurs utilisateurs dans le choix de la transformation  $p \rightarrow \pi$ . Pour les mesures de dissimilarité présentées, dans chacun des cas leurs performances sont très proches et cela nous a principalement permis de donner davantage de poids à nos conclusions. Sur l'utilisation de telle ou telle mesure, nous prendrons donc comme critère premier la complexité opératoire. Nous pensons que parmi les trois proposées, la

plus simple est la mesure  $D_{L1}$ .

Pour terminer cette partie, nous nous proposons d'appliquer notre algorithme de détection de changements sur des images satellites réelles (avec la mesure de dissimilarité  $D_{L1}$ ). La première image a été prise par le satellite radar TerraSAR-X en 2011 du Black Rock Desert au Nevada. L'image numéro 2 est une image de la même zone géographique prise par le même satellite pendant le festival Burning Mountain :

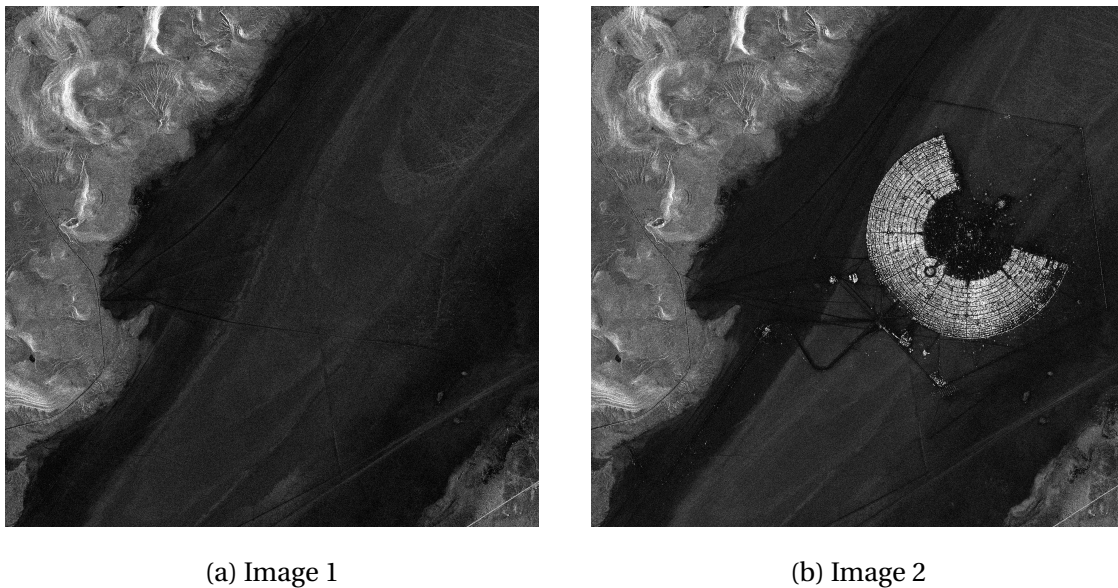
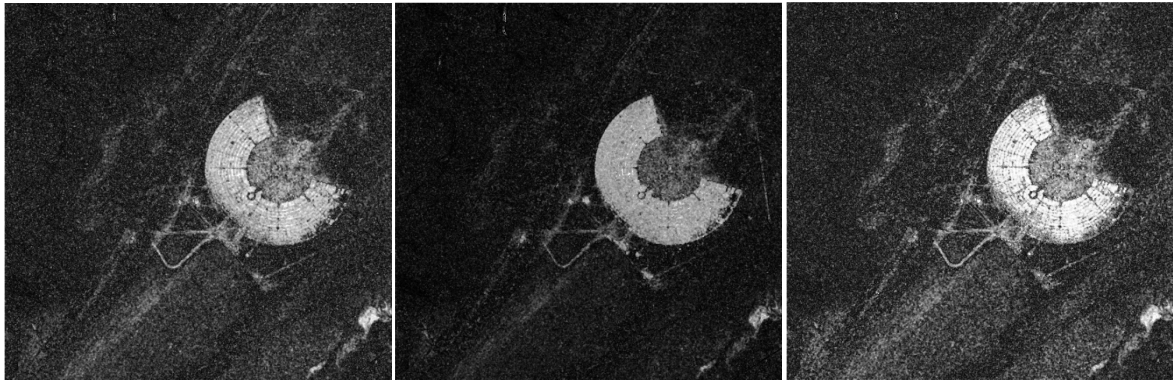


FIGURE 2.9 – Couple d'images avant après changements du Black Rock Desert prise par TerraSAR-X en 2011

La figure 2.10 montre les images de changements issues de l'application pour chacune des transformations considérées et appliquées aux images de TerraSAR-X.

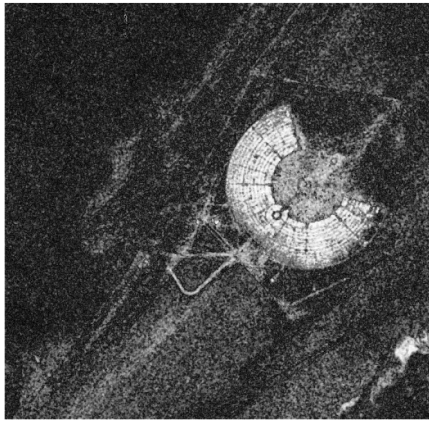
Visuellement, les images de changements présentées dans la figure 2.10 semblent confirmer les conclusions tirées de l'application de détections de changements sur des images artificielles bruitées par un bruit de chatoiement. En effet, les figures 2.10b et 2.10e, représentant les résultats des transformations TK et TS respectivement, montrent que l'on distingue correctement la structure du campement du festival présent dans la figure 2.9b tout en étant robuste à la présence de bruit dans les zones sans changements. Toutefois, il est plus difficile de trancher dans le cas des figures 2.10a 2.10d et 2.10c. La structure du festival est détectée avec plus de détails que pour les transformations TK et TS mais on constate davantage de fausses alarmes dans les zones sans changements.



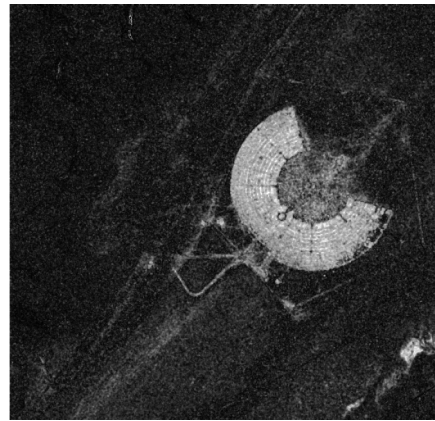
(a) Transformation TN

(b) Transformation TK

(c) Transformation TV



(d) Transformation TO



(e) Transformation TS

FIGURE 2.10 – Résultats de la détection de changements avec la mesure de dissimilarité  $D_{L1}$  suivant la transformation  $p \rightarrow \pi$  considérée .

*Cette première application a fait<sup>2</sup> l'objet d'une publication intitulée «Comparaison de transformations Probabilité-Possibilité aux moyens de mesures de Dissemblance dans un contexte de détection de changements» dans la conférence nationale LFA en 2017 [88].*

### 2.3 Modélisation possibiliste continue

Dans la suite de cette section, nous allons nous intéresser à deux méthodes permettant de définir une représentation possibiliste continue d'un ensemble de données.

La première méthode que nous allons introduire permet de définir les bases de la modélisation possibiliste et permet avec la connaissance de la densité de probabilité de construire une distribution de possibilité.

La seconde méthode, sur laquelle nous nous attarderons plus brièvement, permet de déterminer une distribution de possibilité lorsque la loi de probabilité n'est pas connue et que seules quelques informations telles que la valeur modale et l'écart-type sont connus.

Au travers de la première application, nous avons tenté de déterminer des indices nous permettant de différencier les cinq transformations  $p \rightarrow \pi$  que nous avons présentées. Les résultats de l'application ainsi que des arguments théoriques portant notamment sur le respect du principe de consistance nous ont mené à considérer en particulier la transformation optimale proposée par Dubois, Prade et Sandri.

Ainsi, après avoir présenter son fonctionnement dans un cadre de travail continu, nous allons pousser notre investigation sur cette transformation. Nous utiliserons dans une seconde et une troisième application des distributions de possibilité continues obtenues à l'aide de cette transformation pour modéliser les données et voir si elles peuvent constituer un candidat intéressant pour modéliser de façon plus nuancée les données imparfaites.

### 2.3.1 Transformation optimale continue $p \rightarrow \pi$

La transformation optimale continue a été étudiée en détails par Dubois, Foulloy, Mauris, Prade dans [7]. Elle est basée sur un principe d'empilement des intervalles de dispersion d'une distribution de probabilité pour construire la distribution de possibilité.

**Définition 19.** *La distribution de possibilité induite par une densité de probabilité continue  $p$  autour du mode  $x^m$  est la distribution (notée  $\pi^m$ ) dont les  $\alpha$ -coupe sont les intervalles fermés de dispersion  $I_\alpha^m$  de niveaux  $P(I_\alpha^m) = \alpha$  autour d'une valeur nominale  $x^m$  de  $p$ .*

La distribution de possibilité  $\pi^m$  peut alors être définie de la façon suivante :

$$\pi^m(x) = \sup\{1 - P(I_\alpha^m), x \in I_\alpha^m\}. \quad (2.25)$$

La distribution  $\pi^m$  ainsi calculée est continue et encode l'ensemble des intervalles de dispersions pour tous les niveaux de confiance. De plus, on a  $\pi^m(x^m) = 1$ .

**Théorème 2.** *(Théorème 3.1 dans [7])*

*Pour toute densité de probabilité  $p$ , la distribution de possibilité  $\pi^m$  telle que définie dans la définition 2.25 est consistante avec  $p$ . C'est à dire que pour tout ensemble mesurable  $A$  alors  $\Pi^m(A) \geq P(A)$ , où  $\Pi^m$  et  $P$  sont la mesure de possibilité et la mesure de probabilité associées respectivement à  $\pi^m$  et  $p$ .*

**Lemme 1.** *(Lemme 3.1 dans [7])*

*Pour toute densité de probabilité continue  $p$  ayant un nombre fini de modes, tous les sous-ensembles réels  $I$  de longueur minimum mesurable telle que  $P(I) = \alpha \in (0, 1]$  sont de la forme*



$\{x, p(x) \geq \beta\}$  avec  $\beta \in [0, p_{\max}]$  et  $p_{\max} = \sup(p(x))$ . Ils contiennent donc la valeur modale de  $p$ .

Le Lemme 1 démontre que les intervalles de dispersion les plus informatifs sont les  $\beta$ -coupe d'une densité de probabilité  $p$ , c'est-à-dire les intervalles de dispersion autour du mode. On en déduit alors une expression de la distribution de possibilité  $\pi^m$  de la forme suivante :

$$\pi^m(x) = 1 - \int_{\{y, p(y) \geq p(x)\}} p(y) dy. \quad (2.26)$$

À noter que l'équation (2.26) est la formulation concise de la transformation d'une densité de probabilité continue en une distribution de possibilité.

Pour le moment nous allons nous concentrer au cas où  $x \in \mathbb{R}$  et  $p$  est une densité de probabilité uni-modale uni-variée continue. C'est-à-dire que  $p$  est strictement croissante à gauche et décroissante à droite du mode  $x^m$ .

Dans ce cas, l'équation (2.26) nous conduit à la l'expression suivante :

$$\forall x \in ]-\infty, x^m], \quad \pi^m(x) = \int_{]-\infty, x]} p(y) dy + \int_{(h(x), +\infty[} p(y) dy, \quad (2.27)$$

où  $h$  est une fonction de l'intervalle  $]-\infty, x^m]$  dans  $]x^m, +\infty[$  définie par :

$$\forall x \in ]-\infty, x^m], \quad h(x) = y \geq x^m \quad (2.28)$$

où  $y$  est telle que  $p(y) = p(x)$  pour une distribution  $p$  autour du mode.

On peut alors réécrire l'équation (2.27) de la façon suivante [89][90] :

$$\forall x \in ]-\infty, x^m], \quad \pi^m(x) = F(x) + 1 - F(h(x)), \quad (2.29)$$

et

$$\forall x \in [x^m, +\infty[, \quad \pi^m(x) = \pi(h^{-1}(x)). \quad (2.30)$$

Si  $p$  est une densité de probabilité symétrique et uni-modale, on a alors l'égalité suivante :  $F(h(x)) = 1 - F(x)$ , avec  $F$  la fonction cumulative associée à  $p$ .

L'expression de la distribution de possibilité devient alors :

$$\pi^m(x) = \begin{cases} 2 \int_{-\infty}^x p(y) dy = 2F(x) & \forall x \in ]-\infty, x^m], \\ 2(1 - F(x)) & \forall x \in [x^m, +\infty[ \end{cases} \quad (2.31)$$

Et puisque la fonction cumulative  $F$  associée à  $p$  est continue et dérivable, alors  $\pi^m$  est elle aussi continue et dérivable.

Appliquons, maintenant, la transformée  $p \rightarrow \pi$  de l'équation (2.31) à quelques-unes des densités de probabilité présentées dans la section 1.3.2.

### Cas de la loi gaussienne :

En considérant une loi Gaussienne de fonction cumulative  $F_{\mathcal{N}}(x) = \frac{1}{2} \left( 1 - \operatorname{erf} \left( \frac{x-\mu}{\sigma\sqrt{2}} \right) \right)$  dont le mode est  $\mu$ , on calcule la distribution de possibilité  $\pi_{\mathcal{N}}$  suivante :

$$\pi_{\mathcal{N}}(x) = \begin{cases} 2F(x) = 1 + \operatorname{erf} \left( \frac{x-\mu}{\sigma\sqrt{2}} \right) & \forall x \in ]-\infty, \mu], \\ 2(1 - F(x)) = 1 - \operatorname{erf} \left( \frac{x-\mu}{\sigma\sqrt{2}} \right) & \forall x \in [\mu, +\infty[ \end{cases}$$

Or la fonction  $\operatorname{erf}(x)$  est négative lorsque  $x \in ]-\infty, 0]$  et positive lorsque  $x \in [0, +\infty[$ . On a alors :

$$\pi_{\mathcal{N}}(x) = 1 - \operatorname{erf} \left( \frac{|x - \mu|}{\sigma\sqrt{2}} \right), \quad (2.32)$$

où  $|\cdot|$  est la fonction valeur absolue.

### Cas de la Loi de Weibull :

Pour une loi de Weibull de fonction cumulative  $F_W(x) = 1 - \exp \left( - \left( \frac{x}{\alpha} \right)^\beta \right)$  dont le mode est noté  $x^m = \alpha \left( \frac{\beta-1}{\beta} \right)^{1/\beta}$ , on ne peut pas appliquer directement la transformation présentée dans l'équation (2.31) puisque la loi de Weibull n'est pas symétrique de part et d'autre de la valeur modale.

En effet, une application directe de la transformation continue de l'équation (2.31) nous donne la distribution de possibilité  $\pi_{W_s}$  suivante :

$$\pi_{W_s}(x) = \begin{cases} 2 \left( 1 - \exp \left( - \left( \frac{x}{\alpha} \right)^\beta \right) \right) & \forall x \in ]0, x_{\text{médiane}}], \\ 2 \exp \left( - \left( \frac{x}{\alpha} \right)^\beta \right) & \forall x \in [x_{\text{médiane}}, +\infty[ \end{cases} \quad (2.33)$$

Or cette distribution est alignée sur la médiane de la loi de Weibull et  $p_W(x_{\text{médiane}}) \leq \sup(p_w(x))$ . Cela implique que  $\pi_{W_s}$  ne respecte pas le principe de préservation de l'ordre.

On peut alors préférer utiliser une seconde transformation  $p \rightarrow \pi$  continue. Cette transformation a été proposée par Gilles Mauris dans [91] et il a été démontré qu'elle est consistante avec  $P$  tel que  $\forall A \in \Omega, \Pi(A) \geq P(A)$ .

À l'aide de cette transformation, on détermine une distribution  $p_{W_a}$  de la manière suivante :

$$\pi_{W_a}(x) = \begin{cases} \frac{F(x)}{F(x^m)} = \frac{1 - \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right)}{1 - \exp\left(-\left(\frac{x^m}{\alpha}\right)^\beta\right)} & \forall x \in ]0, x^m], \\ \frac{1 - F(x)}{1 - F(x^m)} = \frac{\exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right)}{\exp\left(-\left(\frac{x^m}{\alpha}\right)^\beta\right)} & \forall x \in [x^m, +\infty[ \end{cases} \quad (2.34)$$

Comme on peut le voir dans la figure 2.11,  $\pi_{W_a}$  à l'avantage de respecter l'asymétrie et le principe de consistance tout comme la distribution obtenue avec la transformation continue de Dubois, Prade et Sandri. Cependant, cette transformation est alignée sur le mode de la distribution de probabilité et respecte donc le principe de préservation de l'ordre contrairement à  $\pi_{W_s}$ .

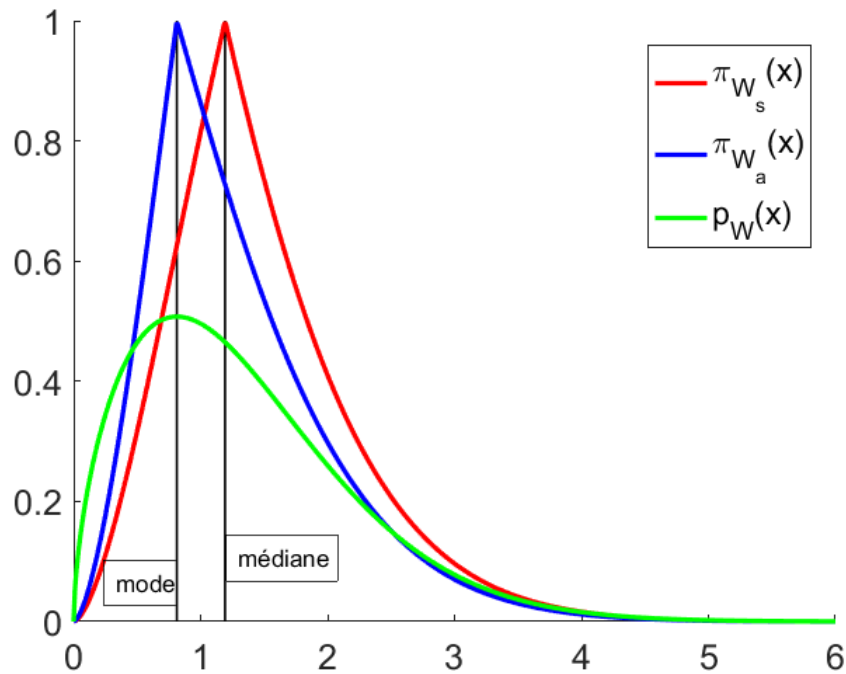


FIGURE 2.11 – Distribution de probabilité de Weibull  $p_W$  et deux transformations  $p \rightarrow \pi$  continues de cette loi :  $p_{W_s}$  et  $p_{W_a}$ .

Toutefois, que ce soit dans le cas de la loi Normale ou dans le cas de loi de Weibull,

les expressions des distributions ne sont pas propices pour l'utilisation d'une mesure de dissemblances (distances ou divergences) continue dont on pourrait calculer des expressions analytiques. On pourrait, bien sûr, appliquer des mesures de dissemblances discrètes mais on perd alors tout l'avantage d'une modélisation continue de l'information.

Nous proposons alors l'approche suivante :

Considérons une variable aléatoire  $X$  centrée en 0 et répartie de façon symétrique de part et d'autre de la moyenne (i.e.  $médiane(X) = \mu(X) = 0$ ).

Bien que particulier, ce cas de figure se présente régulièrement en traitement d'image. On retrouve par exemple ce type de distribution de donnée après une décomposition en ondelettes sur les images de coefficients d'ondelettes [92] ou après l'application d'un filtre de détection de contour tel que le Laplacien of Gaussian (plus de détails dans la section 2.3.5). Une illustration de ce type de données est représentée en figure 2.12.

Afin de modéliser ces données, on peut utiliser une loi symétrique de type gaussienne ou gaussienne généralisée. Mais on peut montrer que la valeur absolue de cette distribution, donc  $|X|$ , suit une loi asymétrique comme la loi de Weibull (voir figure 2.12d). On modélise alors l'ensemble des données par une version symétrisée de la loi de Weibull.

Considérons, maintenant, une seconde variable aléatoire  $Z$  décrite par une distribution de probabilité  $g$  telle que  $g$  est nulle sur  $\mathbb{R}_*^-$  et positive sur  $\mathbb{R}^+$ . Notons par  $p$  la distribution de probabilité décrivant la première variable aléatoire  $X$  centrée en zéro.

Enfin précisons que  $p$  est continue et tel que :

$$p(x) = \frac{1}{2}g(|x|). \quad (2.35)$$

Soit  $G$  la fonction de répartition associée à  $g$ . On calcule alors la fonction de répartition  $F$  associée à  $p$  en fonction de  $g$  de la façon suivante :

Si  $x < 0$  alors  $p(x) = \frac{1}{2}g(-x)$  et donc

$$\begin{aligned} F(x) &= \int_{-\infty}^x p(t)dt = \frac{1}{2} \int_{-\infty}^x g(-t)dt \\ &= \frac{1}{2} \int_{-\infty}^0 g(-t)dt + \frac{1}{2} \int_0^x g(-t)dt. \end{aligned}$$

On pose  $t = -u$  telle que  $dt = -du$ , on a alors

$$F(x) = -\frac{1}{2} \int_0^{-\infty} g(u)du + \frac{1}{2} \int_0^{-x} g(u)du.$$

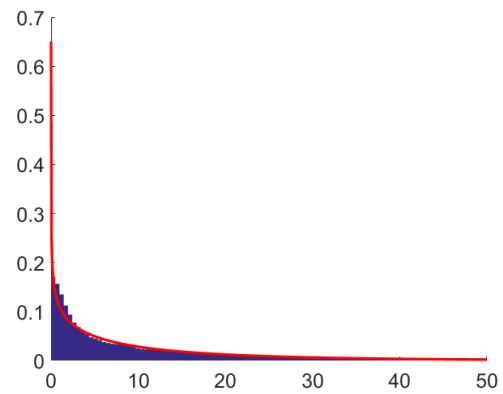
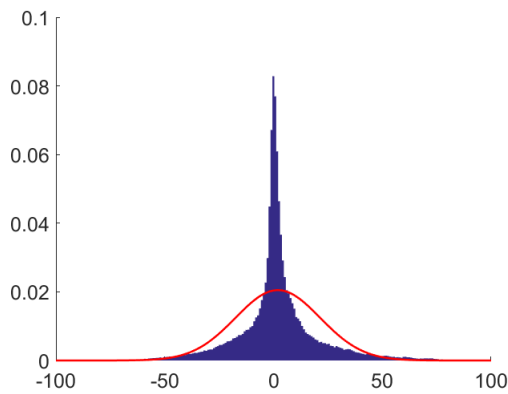
D'où

$$\forall x < 0, F(x) = \frac{1}{2} - \frac{1}{2}G(-x). \quad (2.36)$$



(a) img1 : Image originale (Lena)

(b) img2 : img1 + filtre Laplacian of Gaussian



(c) Distribution empirique de img2,  
loi gaussienne évaluée pour cette distribution

(d) Distribution de la valeur absolue de img2,  
loi de Weibull évaluée pour cette distribution

FIGURE 2.12 – Exemple de distribution de données d’une variable aléatoire  $X$  centrée dont la valeur absolue  $|X|$  suit une loi de Weibull.

Si  $x \geq 0$  alors  $p(x) = \frac{1}{2}g(x)$  et donc

$$\begin{aligned} F(x) &= \int_{-\infty}^x p(t)dt = \frac{1}{2} \int_{-\infty}^x g(t)dt \\ &= \frac{1}{2} \int_{-\infty}^0 g(t)dt + \frac{1}{2} \int_0^x g(t)dt. \end{aligned}$$

On en déduit que :

$$\forall x \geq 0, F(x) = \frac{1}{2} + \frac{1}{2}G(x). \quad (2.37)$$

Puis avec les équations (2.36) et (2.37), on écrit que :

$$F(x) = \begin{cases} \frac{1}{2} - \frac{1}{2}G(-x) & \forall x \in ]-\infty, 0[, \\ \frac{1}{2} + \frac{1}{2}G(x) & \forall x \in [0, +\infty[. \end{cases} \quad (2.38)$$

La variable aléatoire  $X$  étant symétrique et en remplaçant  $G(x)$  par la fonction de répartition associée à la loi de Weibull, on obtient alors :

$$F(x) = \begin{cases} \frac{1}{2} \exp\left(-\left(\frac{-x}{\alpha}\right)^\beta\right), & \forall x \in ]-\infty, 0[, \\ \frac{1}{2} - \frac{1}{2} \exp\left(-\left(\frac{-x}{\alpha}\right)^\beta\right). & \forall x \in [0, +\infty[. \end{cases} \quad (2.39)$$

Enfin, on applique la transformation continue de Dubois et Prade et on calcule la distribution de possibilité suivante :

$$\pi_W(x) = \exp\left(-\left(\frac{|x|}{\alpha}\right)^\beta\right). \quad (2.40)$$

Pour résumer notre démarche, dans le cas des distributions centrées, on ne peut pas utiliser directement une modélisation avec la loi de Weibull car celle-ci n'est définie que sur  $\mathbb{R}^+$ . Cependant, on peut utiliser une version symétrisée de la loi de Weibull pour modéliser adéquatement ce type de distribution.

En considérant la représentation possibiliste associée à cette version de la loi de Weibull symétrisée, on obtient alors une modélisation complète de la variable aléatoire  $X$  centrée en 0. Les paramètres de la distribution de possibilité sont alors les mêmes que les paramètres de la distribution de probabilité associée à la loi de Weibull et sont estimés par rapport à  $|X|$ .

Nous proposons, dans la suite de ce chapitre, de comparer les performances de cette modélisation possibiliste par rapport à la modélisation probabiliste dans le cas des variables aléatoires centrées.

Enfin, en remplaçant le paramètre  $\alpha$  de la distribution de possibilité par le paramètre  $\sigma\sqrt{2}$  d'une loi gaussienne et en fixant  $\beta = 2$ , on remarque que la forme de la distribution  $\pi$  est une normalisation de sa distribution de probabilité dont la moyenne serait nulle. On peut donc considérer la distribution de possibilité  $\pi_{\mathcal{N}}$  suivante

$$\pi_{\mathcal{N}}(x) = \exp\left(-\left(\frac{|x - \mu|}{\sigma\sqrt{2}}\right)^2\right). \quad (2.41)$$

Bien que cette distribution de possibilité soit obtenue par la transformation par normalisation, elle respecte le principe de consistance. On peut le démontrer de façon triviale en étudiant la fonction de différence  $h(x) = \pi_{\mathcal{N}}(x) - \pi_{\mathcal{N}}(x)$  pour  $x \in \mathbb{R}$ . Ainsi, nous nous proposons d'étudier ses performances dans un contexte de détection de changements par rapport à une modélisation probabiliste équivalente.

À noter, que la forme de la distribution de possibilité  $\pi_{\mathcal{N}}$  est à la base de nombreuses méthodes à noyau d'extraction d'informations, que l'on retrouve aussi sous le nom de radial basis function (RBF) [93].

Les deux applications que nous présenterons dans la suite de ce chapitre nous permettront d'apporter des éléments de discussion sur l'intérêt d'utiliser une approche possibiliste plutôt qu'une approche probabiliste pour modéliser des données imparfaites que l'on retrouve en traitement d'images satellites.

### 2.3.2 Construction de distribution de possibilité à partir de la connaissance seule des moments de la variable aléatoire

Lorsque l'on travaille dans le cadre de la modélisation continue, il n'est pas toujours possible de connaître la distribution de probabilité sous-jacente à un ensemble de données. Cependant, des distributions de possibilité associées peuvent être construites à partir des inégalités probabilistes [7].

Les moments d'une variable aléatoire réelle  $X$  telles que la moyenne ou la variance ne permettent pas de définir une distribution de probabilité unique, mais plutôt une famille de distributions. Il est possible d'en choisir une en particulier parmi celles appartenant à cette famille à l'aide du principe d'entropie maximale où le but est de prendre la distribution ayant la plus haute entropie au sens de Shannon, car elle contient le moins d'information par définition. Cependant un tel choix peut conduire à des intervalles de dispersion ne contenant pas forcément les intervalles déduits par l'ensemble des distributions de cette famille de distributions de probabilité [94].

L'idée est alors de reprendre ce qui est exposé dans [58], c'est-à-dire de voir un degré de possibilité comme une borne supérieure de degré de probabilité. On associe alors à toute distribution de possibilité  $\pi$  une famille  $\mathcal{P}$  non vide de mesures de probabilité dominées par la mesure de possibilité :

$$\mathcal{P}(\pi) = \{\mathbb{P}, \forall A \subset \mathbb{R}, \mathbb{P}(A) \leq \Pi(A)\}. \quad (2.42)$$

Ainsi se crée un lien évident entre probabilité et possibilité que l'on peut traduire en termes de calcul d'intervalles [95].

Dans le cas de l'unimodalité, une distribution de possibilité  $\pi$  peut se percevoir comme

un ensemble d'intervalles de dispersion. Ces intervalles sont les  $\alpha$ -coupes de  $\pi$  tels que  $\pi : [\underline{x}_\alpha, \bar{x}_\alpha] = \{x, \pi(x) \geq \alpha\}$  et sont imbriqués dans le sens où ils sont tous définis autour de la valeur modale  $x_0$  de  $\pi$ .

De plus, il a été montré dans [90] qu'empiler les intervalles de dispersion d'une distribution de probabilité  $F$  permet de déterminer une distribution de possibilité  $\pi^{x_0}$  dont le mode est  $x_0$ .

Les  $\alpha$ -coupes de la distribution  $\pi^{x_0}$  ainsi construite sont alors reliées aux intervalles de dispersion  $I_\beta$  de niveau de probabilité  $\beta = 1 - \alpha$  autour de la valeur  $x_0$ . Pour rappel, les intervalles de dispersion  $I_\beta$  sont définis par  $\mathcal{P}(X \in I_\beta) = \beta$ . Cela se traduit par l'égalité :

$$\pi^{x_0}(\underline{x}_\alpha) = \pi^{x_0}(\bar{x}_\alpha) = 1 - \mathbb{P}(X \in I_\beta).$$

En supposant que la densité de probabilité est symétrique et unimodale de mode  $x^{x_0} = x^m$  (i.e.  $\underline{x}_\alpha = x^m - t$  et  $\bar{x}_\alpha = x^m + t$ ), on peut écrire que [7] :

$$\pi^{x^m}(x^m - t) = \pi^{x^m}(x^m + t) = 1 - \mathbb{P}(|X - x^m| \leq t) = \mathbb{P}(|X - x^m| \geq t).$$

Toutefois, dans le cas où la densité de probabilité  $F$  n'est pas connue, il est intéressant de faire intervenir les inégalités probabilistes pour borner la distribution de possibilité  $\pi$  représentant la famille de probabilité  $\mathcal{P} : \pi(x) = \max_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(|X - x^m| \geq t)$ .

Les inégalités probabilistes se présentent de manière générale de la façon suivante, pour  $t \geq 0$  :

$$\mathbb{P}(|X - x_0| \geq t) \leq \lambda(F, r) \frac{\mathbb{E}_F[|X - x_0|^r]}{t^r} \quad (2.43)$$

où  $r$  est un entier non nul,  $\mathbb{E}_p[|X - x_0|^r]$  le  $r^{\text{ieme}}$  moment absolu de la distribution de probabilité  $F$  construite autour de  $X_0$  (moyenne, médiane ou mode), et  $\lambda(F, r)$  une constante dépendant de  $p$  et de  $r$ .

On peut citer deux cas particuliers de l'équation (2.43). Le premier est l'inégalité de Markov pour  $x_0 = 0$ ,  $r = 1$  et  $\lambda(F, 1) = 1$  et le second est l'inégalité de Bienaymé-Chebychev avec  $r = 2$ ,  $\lambda(F, 2) = 1$  et  $x_0 = \mathbb{E}[X]$  [96].

La forme générale des distributions de possibilité dominant la distribution  $\pi$  est alors :

$$\pi(x_0 + t) = \pi(x_0 - t) = \min\left(1, \lambda(F, r) \frac{\mathbb{E}_F[|X - x_0|^r]}{t^r}\right). \quad (2.44)$$



Cela nous permet de définir plusieurs distributions, dont les deux suivantes (en particulier) : la distribution basée sur l'inégalité de Bienaymé-Chebychev et celle basée sur l'inégalité de Gauss-Winckler.

Si seulement la moyenne  $\mu$  et l'écart-type  $\sigma$  sont connus, on applique inégalité de Bienaymé-Chebychev [97] :

$$\pi_{BC}(\mu - t) = \pi_{BC}(\mu + t) = \min\left(1, \frac{\sigma^2}{t^2}\right). \quad (2.45)$$

Si, de plus, la distribution est unimodale et symétrique alors l'inégalité précédente peut-être améliorée en utilisant l'inégalité de Gauss-Winckler [94]. Nous donnons alors l'expression de la distribution de possibilité suivante :

$$\pi_{GW}(\mu - t) = \pi_{GW}(\mu + t) = \begin{cases} 1 - \frac{t}{\sigma\sqrt{3}}, & \text{si } t \leq \frac{2}{\sqrt{3}}\sigma, \\ \frac{4\sigma^2}{9t^2} & \text{si } t \geq \frac{2}{\sqrt{3}}\sigma. \end{cases} \quad (2.46)$$

La figure suivante illustre l'utilisation des deux distributions de possibilité précédemment évoquée dans le cas d'une densité de probabilité de moyenne nulle et de variance égale à 1. À titre de comparaison, la distribution de possibilité optimale  $\pi_{\mathcal{N}}$  déterminée en section 2.3.1 est tracée en noir.

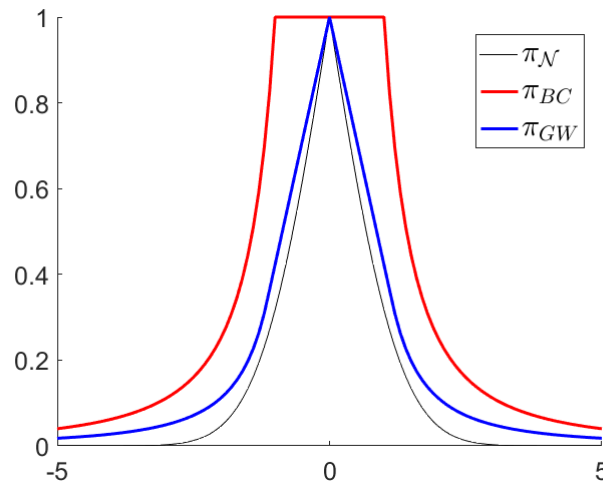


FIGURE 2.13 – Distributions de possibilité obtenues avec les inégalités de Bienaymé-Chebychev et Gauss-Winckler dans le cas d'un support infini.

### 2.3.3 Divergence de Kullback-Leibler possibiliste

Nous avons présenté dans la section 2.3.1 deux distributions de possibilité : la distribution de possibilité  $\pi_W$  associée à une version symétrisée de la loi de Weibull et la distribution  $\pi_{\mathcal{N}}$  obtenue par normalisation d'une distribution de probabilité gaussienne. Chacune des distributions sont de formes exponentielles et nous proposons donc d'utiliser une divergence de Kullback-Leibler possibiliste comme mesure de dissemblance entre distributions de possibilité du même type. Toutefois, les mesures de possibilités, contrairement aux mesures de probabilités, ne sont pas des mesures additives. Nous nous sommes appuyé sur les travaux de Vicenç Torra pour justifier l'utilisation de la divergence de Kullback-Leibler comme mesure de dissemblance continue entre distributions de possibilité.

La divergence de Kullback-Leibler possibiliste se base sur la définition de  $f$ -divergences proposées dans [66] et de la divergence KL proposée dans [98] pour des mesures non additives continues.

**Définition 20.** (Définition 27 dans [66])

Considérons l'ensemble des réels  $\mathbb{R}$  et la  $\sigma$ -algèbre de la tribu borélienne  $\mathcal{B}$ . Soient  $\mu_1$  et  $\mu_2$  deux mesures non additives sur l'espace mesurable  $(\mathbb{R}, \mathcal{B})$  qui sont des intégrales de Choquet (C) de  $\nu$  où  $\nu$  n'est pas nécessairement une mesure additive (Définition 4 dans [98]). Soit  $f$  une fonction convexe telle que  $f(1) = 0$ . Alors la  $f$ -divergence entre  $\mu_1$  et  $\mu_2$  est définie par :

$$D_{f,\nu}(\mu_1 \parallel \mu_2) = (C) \int_{\mathbb{R}} \frac{d\mu_1}{d\nu} f\left(\frac{d\mu_1/d\nu}{d\mu_2/d\nu}\right) d\nu \quad (2.47)$$

où  $d\mu_1/d\nu$  et  $d\mu_2/d\nu$  sont les dérivées de respectives de  $\mu_1$  et  $\mu_2$  par rapport à  $\nu$ .

L'intégrale de Choquet ne sera pas exploité davantage dans ce manuscrit mais le lecteur pourra trouver des compléments d'informations dans [99].

Puis en prenant pour fonction convexe le logarithme naturel tel que  $f(t) = \log(t)$  on peut alors définir la divergence de Kullback-Leibler pour des mesures non additives continues.

**Définition 21.** (Definition 8 dans [98])

Soient  $\mu_1$  et  $\mu_2$  deux mesures non additives sur l'espace mesurable  $(\mathbb{R}, \mathcal{B})$  qui sont des intégrales de Choquet (C) de  $\nu$ . Alors une divergence de Kullback-Leibler entre  $\mu_1$  et  $\mu_2$  par rapport à  $\nu$  est définie par :

$$D_{\log, \nu}(\mu_1 \| \mu_2) = (C) \int_{\mathbb{R}} \frac{d\mu_1}{d\nu} \log \left( \frac{d\mu_1/d\nu}{d\mu_2/d\nu} \right) d\nu \quad (2.48)$$

où  $d\mu_1/d\nu$  et  $d\mu_2/d\nu$  sont les dérivées de respectives de  $\mu_1$  et  $\mu_2$  par rapport à  $\nu$ .

Plusieurs expressions de la mesure  $\nu$  sont possibles. Un choix trivial, nous permettant de simplifier les calculs, est de prendre dans un premier temps une mesure  $\nu$  additive, comme nous l'avons fait dans ce travail. À noter qu'il serait intéressant de considérer dans de futurs travaux une mesure  $\nu$  non additive afin de pleinement prendre en compte les propriétés des mesures de possibilité.

Ainsi, nous avons défini la mesure  $\nu$  de la manière suivante :

$$\nu([-\infty, x]) \triangleq \nu(x) = \int_{]-\infty, x]} \pi(t) dt. \quad (2.49)$$

Cette mesure est directement liée à la "déviation standard" de la distribution de possibilité  $\pi$  [100]. On peut alors facilement montrer que  $\nu$  est une mesure additive et donc écrire l'égalité suivante pour une fonction  $\pi$  continue :

$$d\nu(x) \equiv \pi(x) dx \quad (2.50)$$

où  $dx$  peut-être assimilé à une mesure de Lebesgue.

Finalement, l'équation (2.47) devient pour une mesure  $\nu$  comme définie dans l'équation (2.50) et en considérant deux distributions de possibilité  $\pi_1$  et  $\pi_2$  continues :

$$KL(\pi_1 \| \pi_2) = \int_{\mathbb{R}} \pi_1(x) \log \left( \frac{\pi_1(x)}{\pi_2(x)} \right) dx. \quad (2.51)$$

Cependant, on peut écrire l'équation (2.51) de la façon suivante :

$$KL(\pi_1 \| \pi_2) = - \int_{\mathbb{R}} \pi_1(x) \log \left( \frac{\pi_2(x)}{\pi_1(x)} \right) dx.$$

Et en vertu de l'inégalité de Jensen, on obtient alors l'inégalité :

$$\begin{aligned} \int_{\mathbb{R}} \pi_1(x) \log \left( \frac{\pi_2(x)}{\pi_1(x)} \right) dx &\leq \log \left( \int_{\mathbb{R}} \pi_1(x) \left( \frac{\pi_2(x)}{\pi_1(x)} \right) dx \right) \\ &\leq \log \left( \int_{\mathbb{R}} \pi_2(x) dx \right) \end{aligned}$$

Or, contrairement aux distributions de probabilité on peut avoir  $\int_{\mathbb{R}} \pi_2(x) dx \geq 1$ . Et donc, dans ce cas, on aurait :

$$\log\left(\int_{\mathbb{R}} \pi_2(x) dx\right) \geq 0$$

Ce qui nous donnerait :

$$\text{KL}(\pi_1 \parallel \pi_2) \leq 0,$$

venant alors contredire les propriétés de la divergence.

Il nous faut donc considérer la divergence de Kullback-Leibler symétrique pour que notre divergence KL possibiliste respecte les propriétés des divergences comme énoncé dans la section 1.3.3.2. On considérera donc, pour la suite de cette thèse, la divergence de Kullback-Leibler possibiliste suivante :

$$\text{KL}(\pi_1, \pi_2) = \text{KL}(\pi_1 \parallel \pi_2) + \text{KL}(\pi_2 \parallel \pi_1). \quad (2.52)$$

La démonstration que l'équation (2.52) respecte la propriété de positivité des divergences est fournie en Annexe, ainsi que les démonstrations des divergences suivantes.

**Forme analytique de la divergence KL pour des distributions de possibilité de la forme de  $\pi_{\mathcal{N}N_i}(x, \mu_i, \sigma_i)$  :**

$$\text{KL}(\pi_{\mathcal{N}N_1}, \pi_{\mathcal{N}N_2}) = \frac{\sigma_1 \sqrt{2\pi}}{2} \left( \frac{\sigma_1^2}{\sigma_2^2} + \left( \frac{\mu_1 - \mu_2}{\sigma_2} \right)^2 - 1 \right) + \frac{\sigma_2 \sqrt{2\pi}}{2} \left( \frac{\sigma_2^2}{\sigma_1^2} + \left( \frac{\mu_1 - \mu_2}{\sigma_1} \right)^2 - 1 \right). \quad (2.53)$$

**Forme analytique de la divergence KL pour des distributions de possibilité de la forme de  $\pi_{W_i}(x, \alpha_i, \beta_i)$  :**

$$\text{KL}(\pi_{W_1}, \pi_{W_2}) = 2 \frac{\alpha_1}{\beta_1} \Gamma\left(\frac{1}{\beta_1}\right) \left( \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma\left(\frac{1+\beta_2}{\beta_1}\right)}{\Gamma\left(\frac{1}{\beta_1}\right)} - \frac{1}{\beta_1} \right) + 2 \frac{\alpha_2}{\beta_2} \Gamma\left(\frac{1}{\beta_2}\right) \left( \left(\frac{\alpha_2}{\alpha_1}\right)^{\beta_1} \frac{\Gamma\left(\frac{1+\beta_1}{\beta_2}\right)}{\Gamma\left(\frac{1}{\beta_2}\right)} - \frac{1}{\beta_2} \right) \quad (2.54)$$

où  $\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$ .

Les calculs des formes analytiques de la divergence de Kullback-Leibler pour les distributions de possibilités de la forme de  $\pi_{\mathcal{N}N}$  et de la forme de  $\pi_W$  sont détaillés en section A.1 et A.2 de l'annexe A.

### 2.3.4 Seconde Application : distributions de possibilité de la forme de

$$\pi_{\mathcal{N}}$$

Nous présentons dans cette section les résultats de l'application de détection de changements sur la paire d'images artificielles présentées dans les figures 2.3 et 2.4. L'idée, ici, est de comparer l'impact d'une modélisation probabiliste par rapport à une approche possibiliste. Nous allons donc mettre en opposition une modélisation à l'aide d'une distribution de probabilité gaussienne et une modélisation par distribution de possibilité obtenue en transformant la distribution de probabilité avec la transformation TN. Cela permet, ainsi, d'obtenir une distribution de possibilité de forme simple et respectant le principe de consistance.

Le principe de fonctionnement de l'application est sensiblement le même que précédemment. Pour chaque pixel des images 1 et 2, on sélectionne un patch de taille  $5 \times 5$  pixels autour de celui-ci. Puis on représente la distribution des données dans ces patches par des modèles que l'on compare deux à deux au moyen d'une mesure de dissemblance. En sortie de l'algorithme, on obtient ainsi une image des changements détectés que l'on compare à la vérité terrain pour calculer la performance globale de notre algorithme. Ce qui change dans cette application, c'est la méthode pour obtenir un modèle représen-

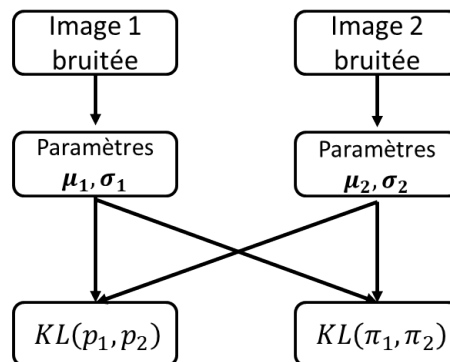


FIGURE 2.14 – Principe de modélisation et de comparaison pour la seconde application.

tatif d'un patch et la mesure de dissemblance utilisée.

D'abord, on utilise la méthode des moments pour évaluer les paramètres associés à la distribution de probabilité gaussienne (idem pour la distribution de possibilité  $\pi_{\mathcal{N}}$  puisque ce sont les mêmes paramètres). Puis, on compare la distribution considérée (de probabilité ou de possibilité) avec la distribution issue de l'image 2 en utilisant la forme analy-

tique de la divergence de Kullback-Leibler associée (voir figure 2.14).

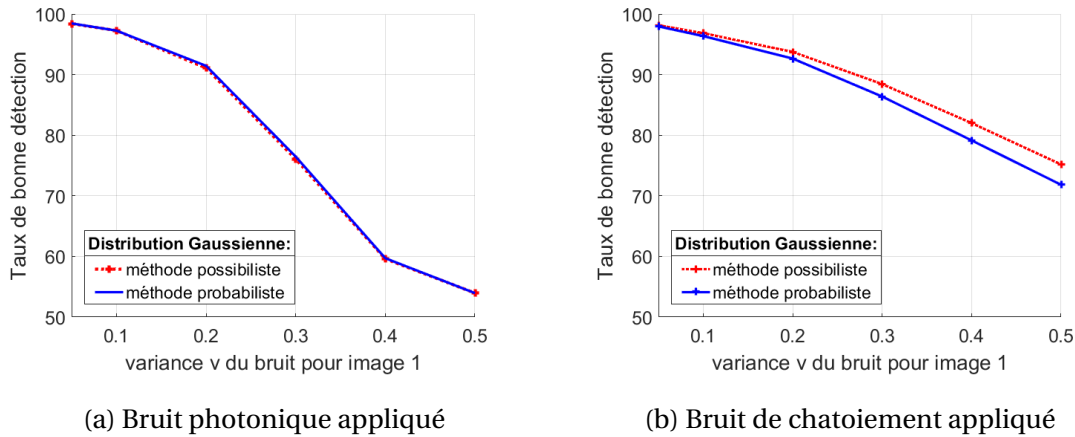


FIGURE 2.15 – Résultats obtenus en comparant l’approche probabiliste et possibiliste en supposant que les données suivent une loi gaussienne

Dans la figure 2.15 sont présentés les résultats de cette seconde application de détection de changements entre images artificielles.

On remarque que dans la figure 2.15a, les performances des deux approches sont similaires et il ne semble pas y avoir davantage à choisir une des deux approches par rapport à l’autre. Toutefois, comme nous l’avons dit précédemment, le fait d’ajouter du bruit photonique sur les images fait tendre leur distribution vers des lois gaussiennes faisant de la distribution gaussienne un candidat évident pour traiter ce type de données. De plus, la distribution  $\pi_{\mathcal{N}}$  n’est pas la plus spécifique parmi celles que l’on peut calculer à partir  $p_{\mathcal{N}}$ , et il se peut donc que le potentiel de l’approche possibiliste ne soit pas pleinement exploité expliquant ainsi la similarité des performances.

En revanche les données bruitées par un bruit de chatoisement ne semblent pas réagir de la même façon comme on peut le voir dans la figure 2.15b. En effet, bien que  $\pi_{\mathcal{N}}$  ne soit qu’une forme normalisée de  $p_{\mathcal{N}}$ , il semblerait qu’elle apporte de la robustesse face à l’augmentation de la variance de bruit de chatoisement appliqué amenant ainsi un premier résultat permettant de discuter de l’intérêt de l’approche possibiliste.

### 2.3.5 Troisième Application : distributions de possibilité de la forme de

$$\pi_W$$

La dernière application que nous présenterons dans ce chapitre a pour but d'illustrer l'utilisation de la distribution de possibilité  $\pi_W$  associée à la loi de Weibull symétrisée.

Afin de mettre en application ce cas de figure où la distribution des données est centrée et symétrique de part et d'autre de zéro, nous allons reprendre l'application de détection de changements sur les images artificielles en ajoutant une étape de filtrage. Le filtre utilisé est le Laplacian Of Gaussian (illustration en figure 2.12). C'est un filtre régulièrement utilisé en traitement d'image [101] et qui permet de mettre en évidence les régions où l'intensité des pixels varie fortement comme c'est souvent le cas pour les côtés des objets dans l'image.

Le principe de fonctionnement de l'application est alors le suivant. On ajoute, dans un premier temps, un bruit de type additif ou multiplicatif sur les images 1 et 2 et on applique le filtre Laplacian of Gaussian. Puis on reproduit la procédure de comparaison en extrayant des patches successifs afin de calculer l'image des changements détectés avec la méthode probabiliste ou possibiliste. Cependant, dans cette application, puisque la loi considérée est la loi de Weibull, on ne considère pas directement la distribution originale  $X$  des données prises dans chaque patch mais une distribution  $X'$  telle que  $X' = |X|$  afin d'estimer les paramètres  $\alpha$  et  $\beta$  associés à la loi. Le fonctionnement de l'algorithme est résumé dans la figure 2.16.

Dans la figure 2.17 sont présentés les taux de détections en fonction de la variance du bruit appliqué obtenus en considérant des données centrées suivant une loi de Weibull symétrisée.

On peut voir dans la figure 2.17a et dans la figure 2.17b que la méthode possibiliste obtient de meilleures performances que la méthode probabiliste sur ce type de données imparfaites. Toutefois, lorsque le bruit appliqué est de type additif, les deux méthodes semblent tendre rapidement vers un taux de performance d'une détection aléatoire. Tandis que dans le cas de l'application du bruit de chatoiement, la méthode possibiliste paraît clairement montrer une robustesse à l'évolution de la variance du bruit par rapport à la

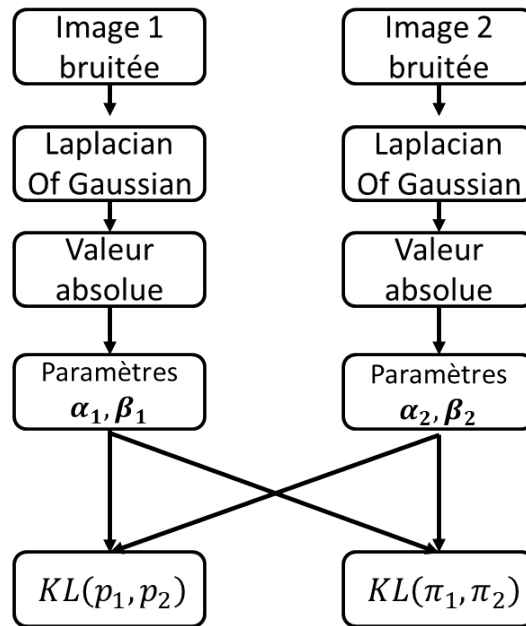
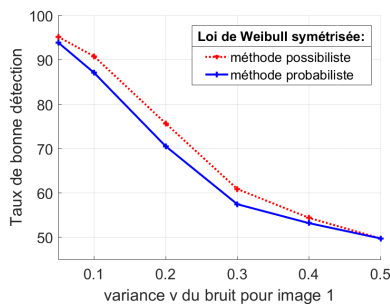
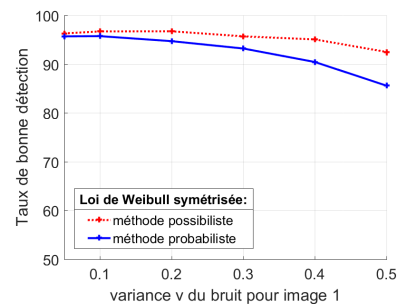


FIGURE 2.16 – Logigramme de fonctionnement de la troisième application.



(a) Bruit photonique appliqué



(b) Bruit de chatoisement appliqué

FIGURE 2.17 – Résultats obtenus en comparant l’approche probabiliste et possibiliste en supposant que les données centrées suivent une loi de Weibull

méthode probabiliste.

Ce résultat semble dû au fait que la distribution de possibilité représente de manière plus adaptée la distribution  $X$  des données de chaque patch. En effet, l’estimation des paramètres  $\alpha$  et  $\beta$  propre à la loi de Weibull est effectuée sur la distribution  $X' = |X|$ . Cela peut conduire dans certains cas à l’évaluation d’un paramètre  $\beta > 1$  impliquant ainsi un mode différent de zéro dans la distribution de probabilité modélisant  $X'$ . Or on sait que la distribution  $X$  est centrée en zéro. De ce fait la distribution de possibilité  $\pi_W$  dont la valeur modale est toujours nulle est alors un meilleur candidat pour représenter ce type de données.



*La forme analytique de la divergence de Kullback-Leibler associée à la distribution  $\pi_W$  ainsi que l'application de détection de changements permettant de la mettre en application ont été l'objet d'une publication intitulée "Image change detection by possibility distribution dissemblance" dans la conférence internationale FUZZ-IEEE 2017 qui s'est déroulée à Naples [102].*

## 2.4 Conclusion du chapitre

Le chapitre 2 de ce manuscrit recense les travaux menés dans le domaine de la modélisation uni-variée aux moyens d'outils issus de la théorie des possibilités, le fil conducteur étant de déterminer un modèle de représentation uni-varié plus nuancé que ne le sont les distributions de probabilité pour représenter les données imparfaites.

Dans la première section, nous avons effectué un état de l'art des différentes transformations de probabilité-possibilité ainsi que des principes permettant de garantir la cohérence de la transformation.

Sur le plan théorique, deux principes en particulier ont retenu notre attention afin de nous guider dans le choix de la transformation à utiliser. Le principe de consistance et le principe de préservation de l'ordre garantissent d'obtenir un modèle suffisamment proche de la distribution initiale tout en restant consistant avec celle-ci, c'est à dire en perdant le moins possible d'information. À ce titre, la transformation optimale proposée par Dubois, Prade et Sandri est la transformation la plus spécifique, donc la plus informative, qui respecte ces deux principes.

Toutefois, il nous a semblé pertinent de mettre en situation chacune de ses transformations afin d'apporter des éléments de discussion sur les performances pratiques des transformations. Grâce à l'application de changements entre images artificielles, nous avons pu tester le comportement de ces transformations par rapport à ce que l'on pourrait appeler un certain degré d'imperfection en lien avec la variance du bruit appliqué aux données. Nous avons pu mesurer ainsi à quel point il semblait nécessaire de relaxer la contrainte de précision pour augmenter les performances de notre algorithme sur ce type de données. Les résultats de l'application nous ont permis de constater plusieurs choses.

La première constatation est que, quel que soit le type de bruit appliqué, la transformation optimale présente des performances globalement supérieures aux autres transformations.

La seconde constatation est que certaines transformations telles que les transformation TK et TS réagissent très différemment suivant si le bruit est plutôt additif ou multiplicatif. Enfin, le dernier point que nous avons relevé concerne la transformation par normalisation qui, bien qu'elle ne respecte pas toujours le principe de consistance, présente des

performances moyennes tout en restant relativement simple à implémenter.

Au travers des deux applications proposées dans la section 2.3, nous avons ensuite développé deux méthodes de modélisations possibilistes continues des données imparfaites. L'idée de travailler avec des distributions continues se justifie par le fait que théoriquement il est possible de s'approcher de la distribution «vraie» des données qui est constituée d'une infinité de valeurs. De plus, elle nous permet de nous affranchir des contraintes du discret tel que le choix du pas d'échantillonnage et des temps de calcul qui en découlent.

La première modélisation est basée sur une transformation par normalisation de la loi de probabilité gaussienne. Bien que, conscients que cette transformation n'est pas la plus spécifique, nous avons préféré choisir cette transformation qui nous donne une forme simple plutôt que la forme obtenue avec la transformation continue de Dubois et Prade. De plus, ce choix nous a ensuite permis de dériver, pour cette distribution de possibilité  $\pi_{\mathcal{N}}$ , une forme analytique de la divergence de Kullback-Leibler possibiliste. Puis, en comparant la méthode de modélisation possibiliste basé sur la distribution  $\pi_{\mathcal{N}}$  avec la méthode probabiliste homologue grâce à l'application de détection de changements, nous avons pu mettre en évidence l'intérêt de l'utilisation d'un modèle moins précis pour représenter les données imparfaites.

La seconde approche continue que nous avons proposée concerne un type de distribution particulier que l'on retrouve régulièrement en traitement d'image : les données de distributions centrées en zéro, réparties symétriquement de part et d'autre de la moyenne et avec une forte concentration de valeurs en zéro. Nous savons que ce type de distribution peut être modélisé avec une loi de Weibull symétrisée et nous avons donc cherché à voir si en appliquant la transformation optimale à cette loi nous pouvions obtenir une meilleure représentation de ces données.

Les résultats de l'application de détection de changements ont semblé confirmer notre hypothèse en montrant notamment que la méthode possibiliste était plus robuste à l'augmentation de la variance des bruits appliqués.

Avant de conclure, nous tenons à préciser que plusieurs séries de tests (non présentés

ici) ont été menés sur différents jeux d'images en faisant notamment varier la taille des changements à détecter ou la taille des patches. Ces tests ont globalement montré des résultats similaires appuyant ainsi nos propos. Toutefois, en regardant en détail les images de changements à la sortie de la chaîne de traitement, il nous a paru que dans certains cas la méthode possibiliste et la méthode probabiliste ne détectaient pas les changements de la même façon. Par exemple, lorsque le bruit photonique appliqué avait une variance élevée, la méthode possibiliste détectait avec un plus grand degré de certitude les bords des objets d'intérêt que leurs centres. Inversement, avec la méthode probabiliste, le degré de certitude de la présence d'un changement était plus fort aux centres des objets que sur les bords. Nous pensons donc qu'il serait intéressant dans de futurs travaux de tendre vers une fusion des résultats des deux méthodes plutôt que de choisir une méthode par rapport à l'autre.

Enfin, comme nous l'avons évoqué dans le chapitre 1, les images satellites, aussi bien optique que radar, sont généralement constituées de plusieurs composantes. Ainsi, il nous a paru pertinent d'explorer le champ de la modélisation multivariée afin de prendre en compte les possibles corrélations entre ces différentes composantes. Nous présentons, dans le chapitre 3, deux approches pour tenter de répondre au problème de la modélisation multivariée. La première approche se basera sur une transformation directe d'un modèle probabiliste multivarié en un modèle possibiliste et la seconde, sur la combinaison de modèle univarié au moyen de copules en vue de définir une représentation multidimensionnelle.

# Chapitre 3

## Distributions de possibilité multivariées

*« Il vaut mieux croire que ne pas croire; en faisant, vous apportez tout au royaume de la possibilité. »*

---

Albert Einstein

### Sommaire

---

<b>2.1 Transformation : Définition et principes</b> . . . . .	<b>50</b>
2.1.1 Principe de consistance . . . . .	51
2.1.2 Principe de préservation de l'ordre . . . . .	51
2.1.3 Principe de conservation de l'incertitude . . . . .	52
2.1.4 Transformation $p \rightarrow \pi$ . . . . .	53
2.1.5 Transformation $\pi \rightarrow p$ . . . . .	57
<b>2.2 Première application : Comparaison discrète de distributions de possibilité</b> . . . . .	<b>58</b>
2.2.1 Comparaison de distributions de possibilité . . . . .	59
2.2.2 Détection de changements entre paires d'images artificielles . . . . .	60
2.2.2.1 Fonctionnement de l'algorithme de détection de changements . . . . .	62
2.2.2.2 Images artificielles . . . . .	63
2.2.3 Quantification de la performance : courbe ROC . . . . .	64
2.2.4 Résultats . . . . .	67
2.2.5 Discussion . . . . .	68
<b>2.3 Modélisation possibiliste continue</b> . . . . .	<b>71</b>

2.3.1	Transformation optimale continue $p \rightarrow \pi$ . . . . .	72
2.3.2	Construction de distribution de possibilité à partir de la connaissance seule des moments de la variable aléatoire . . . . .	79
2.3.3	Divergence de Kullback-Leibler possibiliste . . . . .	82
2.3.4	Seconde Application : distributions de possibilité de la forme de $\pi_{\mathcal{N}}$	85
2.3.5	Troisième Application : distributions de possibilité de la forme de $\pi_{\mathcal{W}}$	87
<b>2.4</b>	<b>Conclusion du chapitre</b> . . . . .	<b>90</b>

---

## Introduction

Dans le chapitre 2, nous avons présenté les principales méthodes permettant de construire une distribution de possibilité à partir de distributions de probabilité discrètes dans le cas de données univariées. Après une revue de l'état de l'art des transformations  $p \rightarrow \pi$  existant dans la littérature scientifique, nous les avons comparées au travers d'une première application de détection de changements entre images artificielles. Cette application a notamment fait ressortir une transformation en particulier : la transformation optimale. Puis, dans un second temps, nous nous sommes concentrés sur la modélisation possibiliste continue et en particulier sur les distributions obtenues grâce à la transformée proposée par Dubois, Prade et Sandri qui est la version continue de la transformation optimale discrète. Grâce à deux applications de détections de changements où l'on comparait les deux types de modélisations probabilistes et possibilistes, nous avons pu mettre en évidence l'intérêt de l'utilisation de modèles de représentation plus souples pour mieux décrire les données imparfaites.

L'objectif du présent chapitre est d'étendre la modélisation possibiliste univariée et continue présentée dans le chapitre précédent au cas multivarié. En effet, comme nous l'avons précisé au chapitre 1, les images satellites sont des images à plusieurs composantes et de nombreuses études [103] ont montré l'intérêt de prendre en compte l'information de la façon la plus globale possible en modélisant notamment les corrélations entre les différentes composantes [104][105].

Nous proposons deux approches distinctes pour répondre à cette problématique. La première approche consiste en une extension de la transformation de probabilité-possibilité continue univariée au cas des distributions de probabilité elliptiques dans l'esprit de [106]. Ces distributions de probabilité sont régulièrement utilisées en traitement d'images et dans le domaine de la statistique [107]. La seconde approche que nous présenterons passe par une modélisation individuelle des distributions marginales que l'on relie à l'aide d'une fonction mathématique appelée *Copule* permettant de rendre compte des structures de dépendances.

Dans le cas de la transformation directe d'une distribution de probabilité multivariée, nous donnerons les expressions des distributions de possibilité pour les trois lois de pro-

babilité multivariées suivantes : la loi gaussienne, la loi de Cauchy et la loi Student-t. Puis, une application de détection de changements sur une base de 14 images multispectrales permettra de comparer les modélisations probabilistes et possibilistes bivariées.

La construction d'un modèle possibiliste multivarié à partir de ses marginales est un problème plus complexe que ne le laisse paraître la littérature, qui ne s'intéresse pas à la consistance entre probabilité et possibilité. Après avoir rappelé comment l'on construit une probabilité jointe à l'aide d'une copule et de ses marginales par le théorème de Sklar, nous montrerons qu'une extension directe du théorème de Sklar à la théorie des possibilités n'est pas immédiate. Nous formaliserons le problème en nous limitant au cas unimodal symétrique et présenterons un certain nombre de résultats exacts d'autres approchés seulement.

### 3.1 Les vecteurs aléatoires réels

Tout d'abord, énonçons les définitions constituant le cadre mathématique dans lequel nos travaux sur l'aspect multivarié de la modélisation prennent place.

Nous noterons  $n \in \mathbb{N}$  la dimension tel que  $n \geq 1$ .

**Définition 22. Vecteur aléatoire réel**

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Soient  $n$  variables aléatoires réelles  $X_1, X_2, \dots, X_n$  définies sur  $\Omega$ . On considère l'application  $\mathbf{X}$  de  $\Omega$  dans  $\mathbb{R}^n$  définie par :

$$\begin{aligned} \mathbf{X} : \Omega &\rightarrow \mathbb{R}^n \\ \omega &\mapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \end{aligned}$$

On dit que  $\mathbf{X}$  est vecteur aléatoire réel.

Dans la suite de ce chapitre, on dénotera par  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vecteur aléatoire défini dans  $\mathbb{R}^n$ .  $X_i, i \in \{1, \dots, n\}$  est une *composante* de  $\mathbf{X}$  dans  $\mathbb{R}$  et sa loi est une *loi marginale* de la loi de  $\mathbf{X}$ .

**Définition 23. Espérance d'un vecteur aléatoire réel**

Soit un vecteur aléatoire réel  $\mathbf{X}$  sur un espace mesurable  $(\Omega, \mathcal{A})$  alors on définit l'espérance de  $\mathbf{X}$  par

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}. \quad (3.1)$$



Le vecteur  $\mu_{\mathbf{X}}$  est constitué des espérances propres à chaque composante de  $\mathbf{X}$ , mais ne prend en compte aucune relation de dépendance entre les marginales.

La structure de dépendance est décrite par la matrice de variance-covariance, aussi appelée matrice de covariance, du vecteur aléatoire  $\mathbf{X}$ .

**Définition 24. Variance d'un vecteur aléatoire réel**

Soit un vecteur aléatoire  $\mathbf{X}$  dans  $\mathbb{R}^n$ , on définit la variance de  $\mathbf{X}$  par

$$\begin{aligned} \Sigma_{\mathbf{X}} &= \mathbb{E} [(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T (\mathbf{X} - \mathbb{E}[\mathbf{X}])] \\ &= (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n} \\ &= \begin{bmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & V(X_n) \end{bmatrix}. \end{aligned} \quad (3.2)$$

La matrice  $\Sigma_{\mathbf{X}}$  est symétrique et définie positive.

À partir de la seconde ligne de l'équation (3.2), on remplit la matrice de covariance de la façon suivante :

Pour  $i = j$  on a

$$\text{Cov}(X_i, X_i) = \mathbb{E} [(X_i - \mathbb{E}[X_i])^T (X_i - \mathbb{E}[X_i])] = \mathbb{E} [(X_i - \bar{X}_i)^2] = V(X_i).$$

Pour  $i \neq j$  on obtient

$$\text{Cov}(X_i, X_j) = \mathbb{E} [(X_i - \mathbb{E}[X_i])^T (X_j - \mathbb{E}[X_j])].$$

Si  $X_i$  et  $X_j$  sont indépendantes au sens de la définition 5 du chapitre 1 alors  $\text{Cov}(X_i, X_j) = 0$ . En particulier, si  $X_1, \dots, X_n$  sont indépendantes entre elles, alors  $\Sigma_{\mathbf{X}}$  est une matrice diagonale dont les coefficients diagonaux sont les variances respectives de chacune des variables aléatoires marginales.

Pour un vecteur aléatoire  $\mathbf{X}$  dans  $\mathbb{R}^n$ , on définit la fonction de répartition jointe en tout point  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  :

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n). \quad (3.3)$$

On définit aussi, si elle existe, sa fonction de densité jointe de probabilité :

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_n}. \quad (3.4)$$

Enfin, sur des intervalles  $A_1, \dots, A_n$  de  $\mathbb{R}$  on définit la probabilité accumulée :

$$\mathbb{P}(\mathbf{X} \in A_1 \times \dots \times A_n) = \int_{A_1 \times \dots \times A_n} p_{\mathbf{X}}(t) dt. \quad (3.5)$$

## 3.2 Distributions de possibilité elliptiques

Dans cette section, nous proposons d'étendre la transformation de probabilité-possibilité continue présentée au chapitre précédent dans la section 2.3.1 au cas des distributions de probabilité elliptiques. Nous commencerons, dans un premier temps, par un rappel sur les principales distributions de probabilité elliptiques. Puis nous exposerons notre extension au cas multivarié de la transformation de probabilité-possibilité et nous dériverons des formes concises de distributions de possibilité pour les lois de Cauchy, Student-t et normale multivariée.

### 3.2.1 Rappel sur les distributions de probabilité elliptiques

**Définition 25.** (Definition 1 dans [108])

Soit  $\mathbf{X}$  un vecteur aléatoire de dimension  $n$ .  $\mathbf{X}$  est distribué elliptiquement ( $\mathbf{X} \sim \text{EC}_n(\boldsymbol{\mu}, \Sigma)$ ) si et seulement s'il existe un vecteur  $\boldsymbol{\mu} \in \mathbb{R}^n$ , une matrice  $\Sigma$  telle que  $\Sigma^{-1} \in \mathbb{R}^{n \times n}$  est semi-définie positive et une fonction  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$  telle que la fonction caractéristique  $t \mapsto \phi_{\mathbf{X}-\boldsymbol{\mu}}(t)$  de  $\mathbf{X}-\boldsymbol{\mu}$  correspond à  $t \mapsto \phi(t^T \Sigma t)$ ,  $t \in \mathbb{R}^n$ .

Considérons un vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^n$  telle que  $\mathbf{X} \sim \text{EC}_n(\boldsymbol{\mu}, \Sigma)$  de densité de probabilité  $p(\cdot, \boldsymbol{\mu}, \Sigma)$ , alors la fonction de densité  $p$  est de la forme suivante :

$$p(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \alpha_n |\Sigma|^{-1/2} g_n((\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})) \quad (3.6)$$

où  $\boldsymbol{\mu} \in \mathbb{R}^n$  est la moyenne,  $\Sigma \in \mathbb{R}^{n \times n}$  est la matrice de variance covariance semi-définie positive et symétrique,  $\alpha_n$  est un coefficient de normalisation et  $|\Sigma|$  est le déterminant associé à la matrice  $\Sigma$ .

La fonction  $g_n : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  est une fonction génératrice de densité des distributions elliptiques de  $\mathbb{R}^n$  respectant la propriété suivante :

$$\int_{\mathbb{R}} g_n(\mathbf{x}^T \mathbf{x}) d\mathbf{x} = 1. \quad (3.7)$$

**Définition 26.** Pour deux vecteurs  $\mathbf{x} \in \mathbb{R}^n$  et  $\mathbf{y} \in \mathbb{R}^n$  liés par une matrice de pondération  $\Sigma$ , la distance de Mahalanobis  $d_m$  est définie par :

$$d_m(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{\Sigma} = \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})}. \quad (3.8)$$

**Remarque 1.** Dans l'équation (3.6), le terme  $(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$  est le carré de la distance de Mahalanobis entre  $\mathbf{x}$  et  $\boldsymbol{\mu}$ .

**Remarque 2.** Dans l'équation (3.6), les contours d'iso densités définis par  $p$  sont des ellipsoïdes centrés en  $\boldsymbol{\mu}$  et dont les axes sont les valeurs propres de la matrice de variance-covariance  $\Sigma$  [109].

Ci-dessous, nous donnons, pour un vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^n$  de moyenne  $\boldsymbol{\mu} \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma \in \mathbb{R}^{n \times n}$ , les formes analytiques des densités de probabilité associées aux lois normale, Student-t et de Cauchy.

#### Loi normale multivariée :

La densité de probabilité  $p_{\mathcal{N}}$  associée à la loi normale multivariée est de la forme :

$$p_{\mathcal{N}}(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}. \quad (3.9)$$

Nous donnons aussi la forme analytique de la divergence de Kullback-Leibler non symétrique pour deux densités de probabilité  $p_{\mathcal{N}1}(\cdot, \boldsymbol{\mu}_1, \Sigma_1)$  et  $p_{\mathcal{N}2}(\cdot, \boldsymbol{\mu}_2, \Sigma_2)$  :

$$D_{\text{KL}}(p_{\mathcal{N}1} \| p_{\mathcal{N}2}) = \frac{1}{2} \left( \ln \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right) + \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \Sigma_2^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - n \right) \quad (3.10)$$

où  $\text{tr}(A)$  est la trace d'une matrice carrée  $A = (a_{ij})_{1 \leq i, j \leq n}$  et  $n$  est la dimension de vecteur aléatoire.

#### Loi Student-t multivariée :

La densité de probabilité  $p_{\mathcal{S}}$  d'un vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^n$  suivant une loi Student-t à la forme suivante :

$$p_{\mathcal{S}}(\mathbf{x}, \boldsymbol{\mu}, \Sigma, \nu) = (\pi\nu)^{-n/2} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma(\nu/2)} \frac{|\Sigma|^{-1/2}}{\left(1 + \nu^{-1}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^{(\nu+n)/2}} \quad (3.11)$$

où  $\nu$  est le degré de liberté de la distribution et  $\Gamma(\cdot)$  est la fonction gamma.

**Loi de Cauchy multivariée :**

La loi de Cauchy multivariée est un cas particulier de la loi Student-t précédente où le degré de liberté  $\nu$  vaut 1. Sa densité de probabilité  $p_{\mathcal{C}}$  est tel que :

$$p_{\mathcal{C}}(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \pi^{-n/2} \frac{\Gamma\left(\frac{1+n}{2}\right)}{\Gamma(1/2)} \frac{|\Sigma|^{-1/2}}{(1 + (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))^{(1+n)/2}}. \quad (3.12)$$

### 3.2.2 Extension de la transformation de probabilité-possibilité continue au cas multivarié

Nous nous proposons, dans cette section, d'étendre la transformation proposée dans l'équation (2.26) du chapitre 2 au cas des densités de probabilité multivariées elliptiques dont la forme est donnée dans l'équation (3.6).

De même que dans le Chapitre 2, le principe est de construire une distribution de possibilité en empilant les intervalles de dispersions calculés sur la distribution de probabilité. Cependant, il nous faut, ici, d'abord définir la probabilité accumulée dans un hypervolume délimité par un isocontour de la distribution de probabilité elliptique.

**Proposition 4.** Posons  $a_x = d_m(\mathbf{x}, \boldsymbol{\mu})$ , pour un vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^n$  de dimension  $n$  de moyenne  $\boldsymbol{\mu} \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma \in \mathbb{R}^{n \times n}$ , dont la densité de probabilité est de la forme de (3.6). La probabilité accumulée dans la région

$$A_{\Sigma} = \{\mathbf{z} \in \mathbb{R}^n \text{ tel que } d_m(\mathbf{z}, \boldsymbol{\mu}) \leq a_x\} \quad (3.13)$$

est

$$\mathbb{P}(A_{\Sigma}) := \mathbb{P}(d_m(\mathbf{z}, \boldsymbol{\mu}) \leq a_x) = \alpha_n \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \int_0^{a_x} r^{n-1} g_n(r^2) dr \quad (3.14)$$

où  $\alpha_n$  est une constante de normalisation.  $\Gamma(\cdot)$  est la fonction gamma et  $g_n$  est une fonction génératrice de densité.

*Démonstration.* Évaluons directement  $\mathbb{P}(A_{\Sigma}) = \int_{A_{\Sigma}} p(u) du$ .

Premièrement, puisque  $\Sigma$  est une matrice définie positive, nous pouvons procéder à un changement de variable dans l'intégrale en posant  $y = \Sigma^{-1/2}(u - \boldsymbol{\mu})$  dont le Jacobien est  $\frac{|\partial u|}{|\partial y|} = |\Sigma|^{1/2}$ .

Donc d'après l'équation (3.6), on obtient

$$\begin{aligned} \int_{A_{\Sigma}} p(u) du &= \alpha_n |\Sigma|^{-\frac{1}{2}} \int_{\|y\| \leq a_x} g_n(\|y\|^2) |\Sigma|^{\frac{1}{2}} dy \\ &= \alpha_n \int_{\|y\| \leq a_x} g_n(\|y\|^2) dy \end{aligned} \quad (3.15)$$

où le changement de variable change la distance de Mahalanobis en distance euclidienne  $\|\cdot\|$ .

Puis, en observant qu'il y a une invariance par rotation dans l'intégrale de l'équation (3.15) puisque c'est l'intégrale d'un ellipsoïde, nous pouvons séparer l'élément de volume  $dy$  en un élément de surface  $d\mathcal{A}$  et un élément radial  $dr$ . Ainsi, en posant  $r = \|y\|$ , on obtient une expression de (3.15) en coordonnées sphériques telles que

$$P(A_\Sigma) = \alpha_n \int_0^{a_x} \int_{S^{n-1}(r)} g_n(r^2) d\mathcal{A} dr,$$

où  $S^{n-1}(r)$  est la  $(n-1)$ -sphère de rayon  $r$ .

Et, sachant que l'aire de la surface associée à cette  $(n-1)$ -sphère peut être écrite  $\mathcal{A}_n(r) = r^{n-1} \mathcal{A}_{n-1}(1)$  où  $\mathcal{A}_{n-1}(1) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$  est la surface de la sphère unitaire [110], alors la probabilité accumulée dans la région  $A_\Sigma$  est

$$\mathbb{P}(A_\Sigma) = \alpha_n \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{a_x} r^{n-1} g_n(r^2) dr \quad (3.16)$$

□

En prenant en compte l'équation (2.26) et les résultats de la section 2.3.2, nous pouvons ainsi étendre la transformation optimale univariée au cas multivarié.

**Définition 27.** [111] Soit une densité de probabilité elliptique  $p$  de moyenne  $\mu \in \mathbb{R}^n$  et matrice de covariance  $\Sigma \in \mathbb{R}^{n \times n}$ . La distribution de possibilité multivariée de dimension  $n$  induite par la densité de probabilité  $p$  autour de la moyenne  $\mu \in \mathbb{R}^n$  est la distribution dont les  $\alpha$ -coupes sont les régions de dispersion elliptique  $A_\Sigma$  de niveau  $\mathbb{P}(A_\Sigma) = \alpha$  autour de la moyenne  $\mu$  telle que pour tout  $\mathbf{x} \in \mathbb{R}^n$  :

$$A_\Sigma = \{\mathbf{z} \in \mathbb{R}^n, \|\mathbf{z} - \mu\|_\Sigma \leq \|\mathbf{x} - \mu\|_\Sigma\} \quad (3.17)$$

où  $\|\mathbf{x} - \mu\|_\Sigma = \sqrt{(\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu)}$ .

Et

$$\begin{aligned} \pi_\Sigma(\mathbf{x}) &= \sup\{1 - P(A_\Sigma), \mathbf{x} \in A_\Sigma\} \\ &= 1 - \alpha_n \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{a_x} r^{n-1} g_n(r^2) dr \end{aligned} \quad (3.18)$$

où  $a_x = d_m(\mathbf{x}, \mu)$ .

**Théorème 3.** *Pour toute densité de probabilité elliptique  $p$ , la distribution  $\pi$  telle que définie dans la définition 27 est consistante avec  $p$ . C'est-à-dire que pour tout ensemble mesurable  $B$  on a  $\Pi(B) \geq \mathbb{P}(B)$ , où  $\Pi$  et  $\mathbb{P}$  sont les mesures de possibilité et de probabilité associée respectivement à  $\pi$  et  $p$ .*

*Démonstration.* Pour n'importe quel ensemble mesurable  $B \subseteq \mathbb{R}^n$ , on peut définir un ensemble  $C = \{\mathbf{x} \in \mathbb{R}^n, \pi(\mathbf{x}) \leq \Pi(B)\}$ . De façon évidente, on a  $B \subseteq C$  car  $\forall B$  mesurables,  $\Pi(B) = \sup_{\mathbf{x} \in B} (\pi(\mathbf{x})) = \Pi(C)$ .

D'autre part, le complémentaire  $C^c$  de  $C$  est une coupe de niveau  $\Pi(B)$  de  $\pi$ , donc  $\mathbb{P}(C^c) = 1 - \Pi(B)$ . On a alors  $\Pi(B) = \mathbb{P}(C)$ . Or  $B \subseteq C \Rightarrow \mathbb{P}(B) \leq \mathbb{P}(C)$ , donc on en conclut que  $\Pi(B) \geq \mathbb{P}(B)$ .  $\square$

**Lemme 2.** *La plus spécifique des distributions de possibilité consistante avec une distribution de probabilité  $\mathbb{P}$  de densité unimodale et continue  $p$  (dans le sens où  $\mathbb{P} \in \mathcal{P}(\Pi)$ ), et qui satisfait le principe de préservation de l'ordre est  $\pi(\mathbf{x}) = \sup (1 - \mathbb{P}(\{y \in \mathbb{R}^n | p(y) \geq p(x)\}))$ .*

*Démonstration.* La preuve de ce lemme est présentée pour  $n = 1$  dans [7], son extension pour  $n \geq 1$  est immédiate.  $\square$

De plus, on peut montrer une propriété de dominance stochastique entre distributions de possibilité de la forme de (3.18).

**Proposition 5.** *Soit  $\mathbf{X} \in \mathbb{R}^n$  un vecteur aléatoire réel de dimension  $n$  de moyenne  $\mu \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma_X \in \mathbb{R}^{n \times n}$ .*

*Soit  $\mathbf{Y} \in \mathbb{R}^n$  un second vecteur aléatoire réel de dimension  $n$  de même moyenne  $\mu \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma_Y \in \mathbb{R}^{n \times n}$ .*

*Enfin, soient  $\pi_X$  et  $\pi_Y$  deux distributions de possibilité elliptiques associées respectivement à  $\mathbf{X}$  et  $\mathbf{Y}$ .*

*Si la matrice  $M = \Sigma_Y - \Sigma_X$  est semi-définie positive, alors  $\pi_X \leq \pi_Y$ , quelque soit  $\mathbf{x}$ .*

*Démonstration.* Dénotons par  $\mathcal{C}_1$  la classe des ensembles convexes de symétrie centrale dans  $\mathbb{R}^n$  et soient  $\Sigma_X$  et  $\Sigma_Y$  deux matrices définies positives. Enfin, soient  $\mathbb{P}_{\Sigma_X}$  et  $\mathbb{P}_{\Sigma_Y}$  deux mesures de probabilité associées à des distributions de probabilité elliptiques.

Alors, d'après le théorème 5.1 dans [112], nous avons le résultat suivant :

Pour un ensemble  $C$  fermé, si  $C \in \mathcal{C}_1$  et si  $\Sigma_Y - \Sigma_X$  est une matrice semi-définie positive, alors  $\mathbb{P}_{\Sigma_Y}(C) \geq \mathbb{P}_{\Sigma_X}(C)$ .

Ainsi, on peut écrire que :

$$\begin{aligned} \mathbb{P}_{\Sigma_X}(C) \geq \mathbb{P}_{\Sigma_Y}(C) &\Leftrightarrow 1 - \mathbb{P}_{\Sigma_X}(C) \leq 1 - \mathbb{P}_{\Sigma_Y}(C) \\ &\Leftrightarrow \pi_{\Sigma_X}(x) \leq \pi_{\Sigma_Y}(x) \end{aligned}$$

□

À noter que dans le cas de vecteurs aléatoires constitués de variables aléatoires indépendantes, les matrices de covariances associées à  $\mathbf{X}$  et  $\mathbf{Y}$  sont des matrices positives diagonales. C'est-à-dire que  $\Sigma_X = \text{diag}(\sigma_{X,1}, \dots, \sigma_{X,n})$  et  $\Sigma_Y = \text{diag}(\sigma_{Y,1}, \dots, \sigma_{Y,n})$ . Ainsi, la matrice  $M = \Sigma_Y - \Sigma_X$  est définie positive si les valeurs propres de  $M$  sont positives ou nulles. En d'autres termes, nous avons :

$$\forall i \in \mathbb{N}^*, \sigma_{X,i} \leq \sigma_{Y,i} \Rightarrow \pi_X(\mathbf{x}) \leq \pi_Y(\mathbf{x}).$$

### 3.2.3 Calcul d'expressions de distributions de possibilité à partir de lois de probabilité elliptiques

L'expression générale des distributions de possibilité elliptique donnée dans l'équation (3.18) peut paraître compliquée à utiliser en l'état. Nous nous proposons donc ici de l'utiliser avec les fonctions génératrices  $g_n$  des lois de probabilité elliptique présentées dans la section 3.2.1. Nous porterons une attention particulière au cas bidimensionnel.

#### Cas de la loi normale multivariée

Par identification des termes dans l'équation (3.9), nous avons  $\alpha_n = \frac{1}{(2\pi)^{n/2}}$  et  $g_n(t) = e^{-\frac{1}{2}t}$ . Donc, d'après (3.14), l'expression de la probabilité accumulée dans la région  $A_\Sigma$  est

$$\mathbb{P}_{\mathcal{N}}(A) = \frac{2\pi^{n/2}}{(2\pi)^{n/2}\Gamma(n/2)} \int_0^{a_x} r^{n-1} e^{-\frac{r^2}{2}} dr. \quad (3.19)$$

Puis, en procédant au changement de variable  $t = \frac{r^2}{2}$ , il a été prouvé dans [111] (résultat que l'on retrouve aussi dans [113]) que la probabilité accumulée dans la région  $A_\Sigma$ , associée à la loi normale, peut alors s'écrire :

$$\mathbb{P}_{\mathcal{N}}(A_\Sigma) := \mathbb{P}_{\mathcal{N}}(d_{mah}(z, \mu) \leq a_x) = \frac{\Upsilon\left(\frac{n}{2}, \frac{a_x^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \quad (3.20)$$

où  $\gamma$  est la fonction Gamma incomplète inférieure définie par  $\gamma(a, b) = \int_0^b t^{a-1} e^{-t} dt$ .

Finalement, l'expression de la distribution de possibilité elliptique  $\pi_{\mathcal{N}}$  obtenue par transformée de la distribution de probabilité de la loi normale multivariée prend la forme :

$$\begin{aligned} \pi_{\mathcal{N}}(x) &= 1 - \frac{\gamma\left(\frac{n}{2}, \frac{a_x^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\ &= \frac{\Gamma\left(\frac{n}{2}, \frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \end{aligned} \quad (3.21)$$

où  $\Gamma(\cdot, \cdot)$  est la fonction Gamma incomplète supérieure définie par  $\Gamma(a, b) = \int_b^{+\infty} t^{a-1} e^{-t} dt$ .

On remarquera que dans le cas où  $n = 1$ , la distance de Mahalanobis est équivalente à  $d_m(x, \mu) = \|x - \mu\|_{\Sigma} = \frac{|x-\mu|}{\sigma}$  et  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ . On montre alors facilement que l'on retrouve l'équation (2.32) de la distribution de possibilité obtenue par la transformée optimale continue de la loi normale unidimensionnelle.

Dans le cas où  $n = 2$ , l'expression de la distribution  $\pi_{\mathcal{N}}$  dans l'équation (3.21) se simplifie grandement pour devenir

$$\forall x \in \mathbb{R}^2, \quad \pi_{\mathcal{N}}(x) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}. \quad (3.22)$$

On notera alors que (3.22) est un des cas particuliers où la transformation par sommation des  $\alpha$ -coupes proposée dans [7] est égale à la transformation par normalisation de la distribution de probabilité gaussienne bidimensionnelle.

**Remarque 3.** On peut calculer la forme analytique de l'indice de spécificité  $I_S$  pour toute distribution de la forme de (3.22) en calculant l'aire sous la distribution tel que

$$I_S(\pi_{\mathcal{N}}) = \int_{\mathbb{R}} \pi_{\mathcal{N}}(x) dx = 2\pi \sqrt{|\Sigma|}. \quad (3.23)$$

Et on remarque, que dans ce cas précis, si l'on considère une entropie de Shannon "possibiliste" de la forme

$$H_{\pi}(\pi_{\mathcal{N}}) = - \int_{\mathbb{R}} \pi_{\mathcal{N}}(x) \log(\pi_{\mathcal{N}}(x)) dx \quad (3.24)$$

On a alors l'égalité  $I_S(\pi_{\mathcal{N}}) = H_{\pi}(\pi_{\mathcal{N}})$

### Cas de la loi de Cauchy et Student-t :

Le cas de la loi de Cauchy et de la loi Student-t va être traité conjointement puisqu'on



retrouve l'expression de la loi de Cauchy en fixant le degré de liberté à 1.

Pour la loi Student-t, en identifiant dans l'équation (3.11) que  $\alpha_n = (\pi\nu)^{-n/2} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\nu/2)}$  et que  $g_n(t) = \frac{1}{(1+\nu^{-1}t^2)^{(\nu+n)/2}}$ , on détermine l'expression suivante de la probabilité accumulée dans la région  $A_\Sigma$  :

$$\mathbb{P}_{\mathcal{F}}(A_\Sigma) = \frac{2\nu^{-\frac{n}{2}} \Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2}) \Gamma(\frac{n}{2})} \int_0^{a_x} \frac{r^{n-1}}{\left(1 + \frac{r^2}{\nu}\right)^{(\nu+n)/2}} dr. \quad (3.25)$$

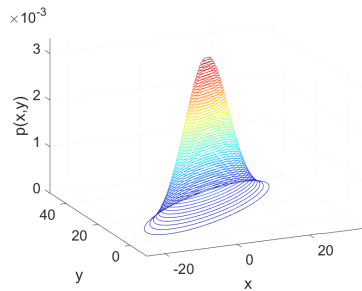
Toutefois, l'équation (3.25) ne se simplifie pas davantage sans fixer la valeur de  $n$ .

Dans le cas où  $n = 2$ , on détermine que la distribution de possibilité elliptique associée à la loi Student-t est :

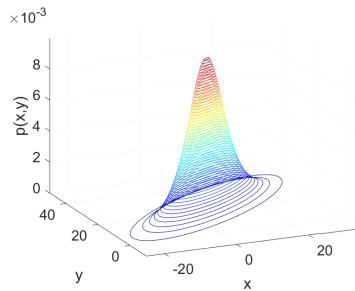
$$\forall x \in \mathbb{R}^2, \quad \pi_{\mathcal{F}}(x) = \left(1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)^{-\frac{\nu}{2}}. \quad (3.26)$$

Toujours pour  $n = 2$  et en posant  $\nu = 1$ , on en déduit l'expression de la distribution elliptique associée à la loi de Cauchy :

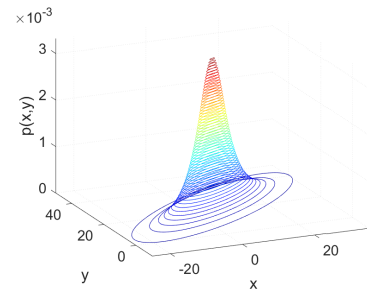
$$\forall x \in \mathbb{R}^2, \quad \pi_{\mathcal{C}}(x) = \left(1 + (x - \mu)^T \Sigma^{-1}(x - \mu)\right)^{-\frac{1}{2}}. \quad (3.27)$$



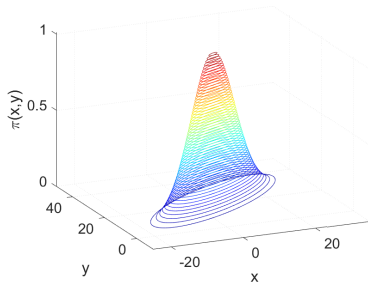
(a) PDF de la loi normale



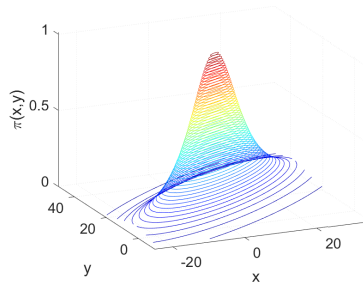
(b) PDF de la loi Student-t



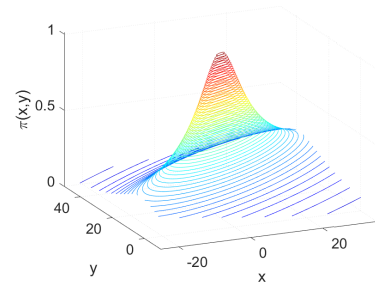
(c) PDF de la loi de Cauchy



(d)  $\pi_{\mathcal{N}}$



(e)  $\pi_{\mathcal{F}}$



(f)  $\pi_{\mathcal{C}}$

FIGURE 3.1 – Exemple de distributions de probabilité bidimensionnelle (première ligne) et des distributions de possibilité associées (seconde ligne).

Dans la figure 3.1 sont données en illustration, les différentes distributions de possibilité elliptique bidimensionnelles et la distribution de probabilité dont elles sont respectivement issues. Pour la loi Student-t le degré de liberté est  $\nu = 3$ . Dans chaque cas, le vecteur moyenne est tel que  $\mu = (5 \quad 20)^T$  et la matrice de covariance  $\Sigma$  est telle que :

$$\Sigma = \begin{bmatrix} 8 & 38.4 \\ 38.4 & 6 \end{bmatrix}.$$

### 3.3 Application de détection de changements sur images réelles

Dans cette section nous proposons une illustration de l'utilisation de la distribution de possibilité elliptique bidimensionnelle  $\pi_{\mathcal{N}}$  associée à la loi normale donnée en équation (3.22).

Pour rester cohérent avec ce qui a été proposé dans le chapitre 2 sur la modélisation possibiliste univariée, nous avons encore une fois considéré une application de détection de changements, mais cette fois-ci sur une base réelle d'images satellites multispectrales.

#### 3.3.1 Estimation de la matrice de covariance $\Sigma$

Un intérêt non négligeable de l'analyse d'images multidimensionnelles est la modélisation des structures de corrélation entre les différentes bandes dont sont composées les images. Cette modélisation est faite au travers de l'outil mathématique que nous avons évoqué précédemment et que l'on appelle matrice de covariance.

Il existe différentes méthodes pour estimer la matrice de covariance dans la littérature scientifique et on peut notamment citer la référence [114] pour des méthodes d'estimation robustes.

Toutefois, afin de garder une application relativement simple nous allons utiliser la méthode SCM (Sample Covariance Matrix) [115] particulièrement propice dans le cas où l'on suppose que la distribution des données suit une loi gaussienne multivariée.

Pour deux variables aléatoires réelles  $X = \{x_1, \dots, x_i, \dots, x_N\}$  et  $Y = \{y_1, \dots, y_i, \dots, y_N\}$  de moyenne respective  $\mu_X$  et  $\mu_Y$ , la matrice de covariance  $\Sigma \in \mathbb{R}^{2 \times 2}$  est estimée de la façon

suivante :

$$\Sigma(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y)^T. \quad (3.28)$$

### 3.3.2 Divergence possibiliste pour distribution de possibilité elliptique

Deux mesures de dissemblances ont été considérées pour cette application. La première mesure est la divergence de Kullback-Leibler que nous avons utilisée plusieurs fois dans le chapitre précédent. La seconde est la distance d'Hellinger que nous avons présenté dans le chapitre 1.

D'autres mesures auraient pu être envisagées [116] mais nous nous limitons ci-après à ces deux mesures dont nous donnons, ci-dessous, les formes analytiques dans le cas de distributions de possibilités dont la forme est donnée dans l'équation (3.22).

Le calcul de la divergence de Kullback-Leibler entre deux distributions de possibilité elliptiques  $\pi_{\mathcal{N}_1}(x, \mu_1, \Sigma_1)$  et  $\pi_{\mathcal{N}_2}(x, \mu_2, \Sigma_2)$  est analogue à celui effectué pour obtenir les formes analytiques présentées dans la section 2.3.3. Le détail du calcul est donné en Annexe A.3.

La forme analytique de la divergence de Kullback-leibler est la suivante :

$$\begin{aligned} \text{KL}(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) = & \pi \sqrt{|\Sigma_1|} \left[ \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) - 2 \right] \\ & + \pi \sqrt{|\Sigma_2|} \left[ \text{tr}(\Sigma_1^{-1} \Sigma_2) + (\mu_2 - \mu_1)^T \Sigma_1^{-1} (\mu_2 - \mu_1) - 2 \right] \end{aligned} \quad (3.29)$$

où  $\text{tr}(A)$  est la trace de la matrice  $A$ .

En suivant les travaux dans [117], la forme analytique du carré de la distance d'Hellinger entre deux distributions de possibilités de la forme de (3.22) est donnée par :

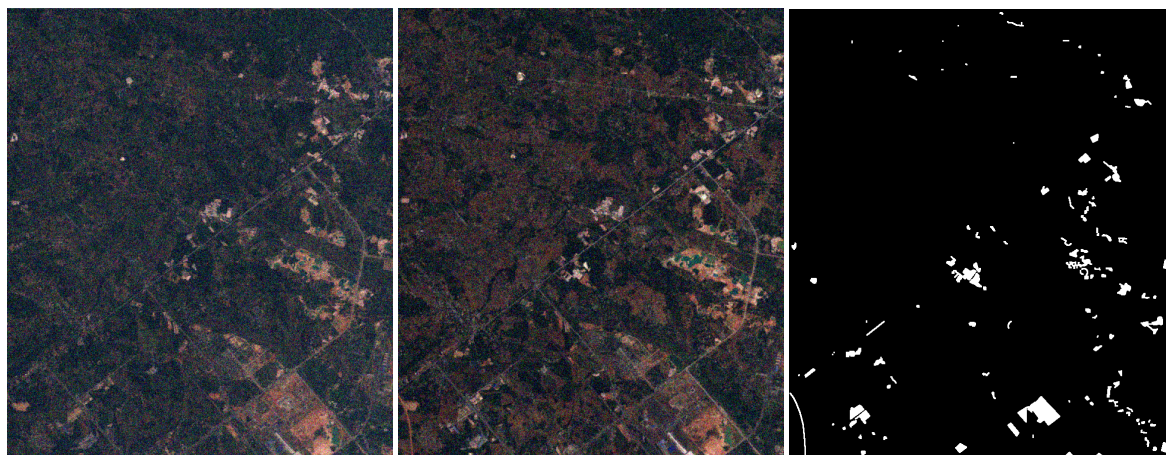
$$D_H^2(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) = \pi \left( \sqrt{|\Sigma_1|} + \sqrt{|\Sigma_2|} \right) - \frac{2\pi}{\sqrt{\left| \frac{1}{2}\Sigma_1^{-1} + \frac{1}{2}\Sigma_2^{-1} \right|}} e^{-\frac{1}{8}(\mu_1 - \mu_2)^T \left( \frac{1}{2}\Sigma_2 + \frac{1}{2}\Sigma_1 \right)^{-1} (\mu_1 - \mu_2)}. \quad (3.30)$$

Le détail des étapes de calcul est fourni en Annexe A.4.

### 3.3.3 Base d'images satellites et algorithme de détection de changements

La base d'images satellites multispectrales que nous avons considérée est composée de 14 paires d'images "Avant-Après" de différentes villes autour du globe. Ces images sont

tirées de la partie entraînement de la base d'images Onera Satellite Change Detection (OSCD) où pour chaque paire d'images le masque de vérité terrain des changements est fourni.



(a) Image de Beihai à la date  $t_1$  (b) Image de Beihai à la date  $t_2$  (c) Masque de vérité terrain

FIGURE 3.2 – Exemple de paire d'images "Avant-Après" de la ville de Beihai et le masque de vérité terrain des changements.

La base de données OSCD a été présentée dans le papier "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Network" à la conférence internationale IGARSS'2018 [118].

Elle est composée originalement de 24 paires d'images multispectrales prises par le satellite Sentinel-2 entre 2015 et 2018 de plusieurs villes un peu partout sur Terre : en Amérique du Sud, États-Unis, Europe, Moyen-Orient et Asie. Pour chaque ville, les paires d'images enregistrées sont composées de 13 bandes multispectrales variant en résolution spatiale entre dix, vingt et soixante mètres.

Le fonctionnement de l'algorithme de détection de changements est présenté dans la figure 3.3 et nous le décrivons ci-dessous.

Nous effectuons d'abord une étape de prétraitement. On commence par représenter les images au format "double precision" pour que les valeurs s'échelonnent entre 0 et 1. On procède, ensuite, à une extension symétrique proportionnellement à la taille des patches d'extraction afin d'éviter les effets de bords. Enfin, on applique un bruit additif de

moyenne nulle et de variance fixée selon les besoins. Cela permet, ainsi, de garantir une évaluation réaliste de la matrice de covariance en présence de structures de corrélation déficientes en rang; ce qui peut arriver s'il y a des pixels manquants ou si tous les pixels contenus dans le patch ont la même valeur.

Nous effectuons ensuite une étape de traitement. Pour une paire d'images, nous sélectionnons les images des bandes spectrales dans le domaine du visible (résolution spatiale de 10m). C'est-à-dire les images à trois composantes RVB et en particulier nous ne gardons que les composantes vertes et bleues de ces images. À noter que ce choix des deux composantes retenues parmi les trois est arbitraire, l'essentiel étant la structure de corrélation entre deux composantes.

Puis sur les patches extraits dans chaque image, nous évaluons les vecteurs  $\mu_1$  et  $\mu_2$  tel que la première valeur du vecteur  $\mu_1$  est la moyenne du patch extrait de la composante verte et la seconde valeur est la moyenne du patch de la composante bleue à la date  $t_1$  (réciproquement pour le vecteur  $\mu_2$ ). On estime ensuite les matrices de covariance  $\Sigma_1$  et  $\Sigma_2$  entre les bandes vertes et bleues pour chacune des dates  $t_1$  et  $t_2$ . Enfin, on calcule la forme analytique de la divergence considérée et on stocke le résultat dans l'image de changement à l'emplacement du pixel au centre du patch.

On répète ce processus jusqu'à ce que l'image des changements soit complètement remplie.

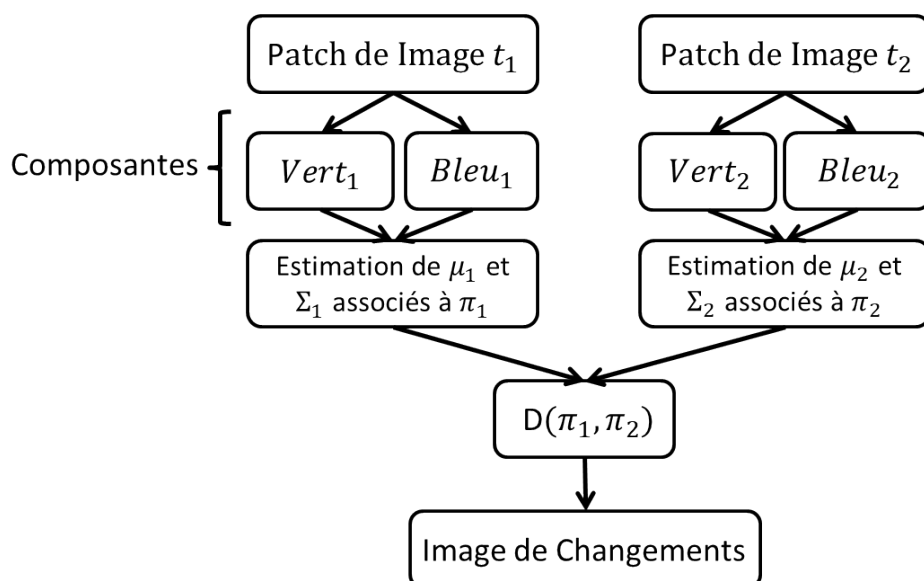


FIGURE 3.3 – Fonctionnement de l'algorithme de détection de changements.

L'évaluation de la performance de chaque algorithme est faite de la même façon que dans les applications du chapitre 2. On compare les images de changements avec le masque de vérité terrain des changements afin de calculer la courbe ROC. Puis on retient l'intégrale sous la courbe ROC comme taux de performances de l'algorithme de détection de changements.

### 3.3.4 Résultats

Deux différentes séries de tests ont été pratiquées dans cette application de détection de changements.

Dans la première série de tests, l'idée a été d'estimer l'impact de la taille des patchs sur les performances de l'algorithme. Comme on peut le voir dans la figure 3.2c, la taille des changements à détecter varie de l'ordre de quelques pixels de diamètre à quelques dizaines de pixels de diamètre, nous nous attendons donc à observer des variations dans les taux de performances suivant la taille des fenêtres d'extractions choisies. Ainsi, pour cette série, nous avons appliqué sur chaque image un bruit gaussien de moyenne nulle et de variance  $\nu = 0.01$ . Puis, nous avons exécuté l'algorithme plusieurs fois en augmentant à chaque fois la taille des patchs de  $3 \times 3$  pixels jusqu'à  $41 \times 41$  pixels et en utilisant les deux mesures de dissemblances présentées dans la section 3.3.2. Les résultats sont présentés en figure 3.4. Chaque point du graphique est la moyenne des taux de performances obtenus par l'algorithme sur les 14 paires d'images "Avant-Après".

La seconde série de tests a eu pour objectif d'évaluer l'impact de la variance du bruit additif appliqué sur les taux de performances de l'algorithme de détection de changements. La taille des patchs d'extraction a été fixée à  $5 \times 5$  pixel de diamètre, et à chaque exécution de l'algorithme nous avons augmenté la variance du bruit appliqué en allant de  $\nu = 0.01$  à  $\nu = 0.1$ .

Les résultats de cette série de tests sont présentés dans la figure 3.5. De même que dans la série de tests précédente, chaque point de cette figure représente l'AUC moyenne obtenue sur les 14 paires d'images.

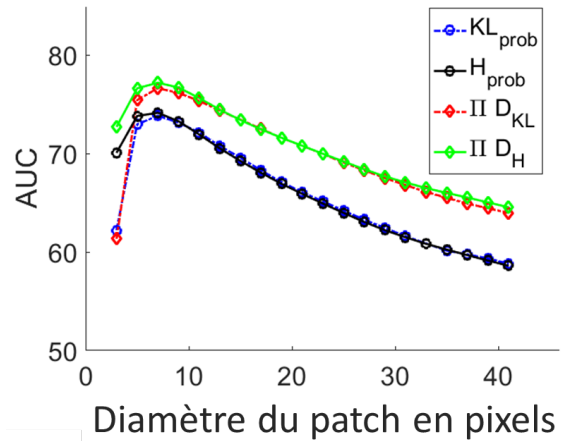


FIGURE 3.4 – Graphique montrant les moyennes des AUC sur les 14 paires d'images en fonction du diamètre du patch d'extraction utilisé.

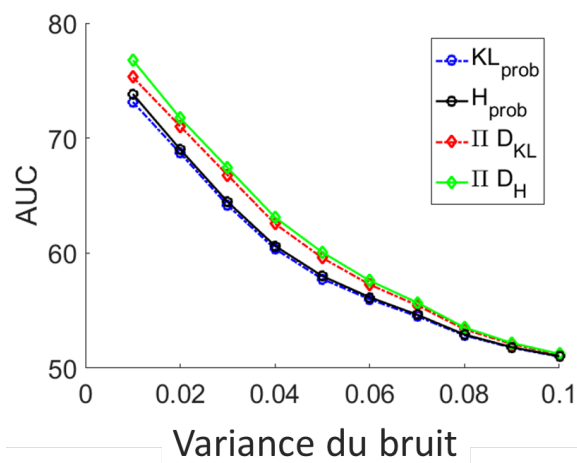


FIGURE 3.5 – Graphique montrant les moyennes des AUC sur les 14 paires d'images en fonction de la variance du bruit gaussien appliqué.

Dans les figures 3.4 et 3.5, on remarque que les taux de performances obtenus avec la divergence de Kullback-Leibler et avec la distance d'Hellinger sont globalement proches dans les deux séries de tests que ce soit avec une modélisation probabiliste ou une modélisation possibiliste. Cela indique que les performances ne sont pas favorisées par l'une ou l'autre des mesures de divergence mais qu'elles sont plutôt dépendantes du choix de la méthode de modélisation. On notera, toutefois, que dans le cas de patchs de tailles très faibles ( $3 \times 3$  pixels) lorsque l'échantillon est trop petit pour être réellement représentatif, la distance d'Hellinger semble moins impactée que la divergence de Kullback-Leibler.

Comparons, maintenant, l'approche probabiliste de l'approche possibiliste aux vues des résultats fournis par l'application de détection de changements.

Dans la figure 3.4, lorsque la taille du patch est de  $3 \times 3$  pixels, l'estimation des matrices de covariances par la méthode SCM ne peut pas être considérée comme fiable en partie dû à la présence de valeurs aberrantes dans les patches. Ainsi chaque méthode est lourdement impactée et les performances mesurées sont alors assez faibles. D'un autre côté, plus la taille du patch augmente et plus les performances diminuent à partir d'un certain seuil (10 pixels de diamètre) et ce même si la méthode SCM est davantage consistante en augmentant le nombre d'échantillons. Le meilleur compromis est trouvé pour un patch de  $7 \times 7$  pixels pour chaque méthode. Enfin, on observe globalement que quelque soit la taille du patch considéré, l'approche possibiliste présente de meilleures performances que l'approche probabiliste.

Tout comme nous avons pu l'observer dans les résultats de la section 2.3.5, la figure 3.5 nous permet à nouveau de voir que l'approche possibiliste est plus robuste à la présence de bruit que l'approche probabiliste. En effet, les taux de bonnes détections moyens de la modélisation possibiliste surpasse ceux de l'autre approche lorsque la variance du bruit appliqué est inférieur à 0.6. Cela est, de plus, confirmé par les résultats des deux mesures de dissemblances. Puis, comme on peut s'y attendre, l'écart s'atténue et les deux méthodes obtiennent une AUC moyenne de 50 lorsque la variance est égale à 0.1 indiquant alors une classification aléatoire entre zone de changements et zone de non-changements.

*Les développements théoriques présentés, ci-dessus, concernant l'extension de la transformation de probabilité-possibilité de lois de probabilité univariée au cas des lois de probabilité elliptiques ainsi que l'application de détection de changements entre images satellites multispectrales à l'aide de distributions de possibilité elliptiques ont fait l'objet d'une publication dans la revue scientifique IEEE Transactions on Fuzzy sets and Systems [111].*

### **3.4 Constructions de distributions de possibilité dominantes**

Comme nous l'avons mentionné dans le chapitre 2, il n'est pas toujours possible de connaître la loi de probabilité sous-jacente de la distribution des données. Toutefois,



comme dans le chapitre précédent, il est possible de construire une famille de distributions de possibilité à partir seulement de la connaissance de la moyenne  $\mu$  et de la covariance  $\Sigma$  d'une distribution.

Dans la suite de cette section, nous proposons deux méthodes différentes pour construire cette distribution.

### 3.4.1 Support infini

La première méthode repose sur l'inégalité de Chebychev multivariée [119] qui s'énonce de la façon suivante :

$$\forall X \in \mathbb{R}^n, \mathbb{P} \left( \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)} > t \right) \leq \frac{n}{t^2}. \quad (3.31)$$

Ce qu'énonce cette inégalité c'est que la probabilité accumulée dans la région à l'extérieur de l'isocontour elliptique de rayon  $t$  est inférieure ou égale au terme  $\frac{n}{t^2}$ . Cet isocontour elliptique délimite une région  $I_\beta$ , on a alors :

$$\mathbb{P}(X \notin I_\beta) = 1 - \mathbb{P}(X \in I_\beta) \leq \frac{n}{t^2}. \quad (3.32)$$

Commençons par définir une famille  $\mathcal{P}$ , non vide, de mesures de probabilité, telle que pour une distribution de possibilité  $\pi$ , cette famille est dominée par la mesure de possibilité  $\Pi$  sur tout intervalle  $A$  réel :

$$\mathcal{P}(\pi) = \{\mathbb{P}, \forall A \subset \mathbb{R}, \mathbb{P}(A) \leq \Pi(A)\}. \quad (3.33)$$

En supposant que la distribution des données est unimodale, on écrit que la distribution de possibilité la plus spécifique est celle dont les  $\alpha$ -coupes sont liées aux régions de dispersion  $I_\beta$  de la famille de probabilité  $\mathcal{P}$  telle que  $\beta = 1 - \alpha$ .

Si  $\pi$  est de forme elliptique de centre  $\mu$  et de matrice de covariance  $\Sigma$  alors une  $\alpha$ -coupe de  $\pi$  est une ellipse  $\mathcal{E}_\alpha$  de rayon  $t$  dont l'équation est  $\|x - \mu\|_\Sigma = t$ .

Alors la distribution de possibilité  $\pi$  représentant la famille de probabilité  $\mathcal{P}$  est :

$$\pi(x) = \max_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\|x - \mu\|_\Sigma \geq t). \quad (3.34)$$

En utilisant l'inégalité de l'équation (3.31), on détermine alors que la distribution de possibilité dominant la distribution  $\pi$  elliptique est :

$$\forall x \in \mathbb{R}^n, \pi(x) = \min \left( 1, \frac{n}{(x - \mu)^T \Sigma^{-1} (x - \mu)} \right). \quad (3.35)$$

La figure 3.6 présente un exemple de la distribution de possibilité de l'équation 3.35 dans le cas bidimensionnel.

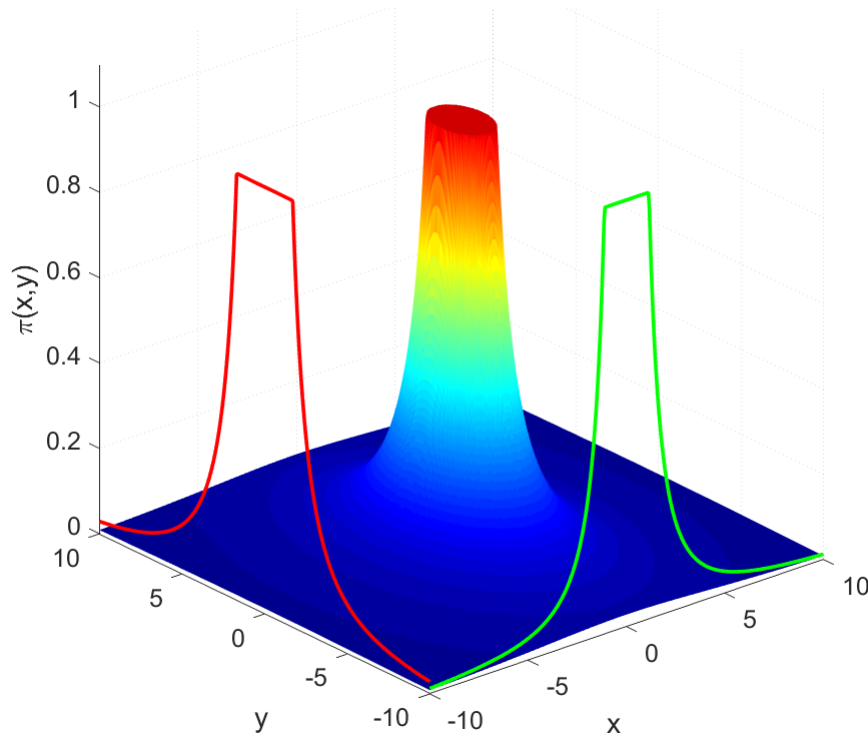


FIGURE 3.6 – Distribution de possibilité bivariable calculée à partir de l'inégalité de Chebychev multivariable.

### 3.4.2 Support borné

Le cas des distributions à support borné a été considéré par H. Akrouf, J.-F. Crouzet et O. Strauss [120] et permet de construire la distribution de possibilité la plus spécifique dominant un ensemble de distributions de probabilité sphérique bornées.

Soit  $B_r = \{x \in \mathbb{R}^n, \|x\| < r\}$  une hypersphère ouverte de dimension  $n$  et de rayon  $r$ , où  $\|\cdot\|$  est la norme euclidienne.

Soit un ensemble  $\Theta$  de distributions de probabilité sphériques bornées est définie de la manière suivante :

**Définition 28.** Une distribution de probabilité  $p$  appartient à  $\Theta$  si elle est définie comme une fonction sur l'hypersphère  $B_r$  vérifiant :

- $p$  est continue et  $\int_{B_r} p = 1$
- $p$  est radiale, i.e.  $\exists \phi_p : \mathbb{R} \rightarrow \mathbb{R}^+, p(x) = \phi_p(\|x\|)$ ,

—  $\phi_p$  est décroissante sur  $[0, r[$ ,

—  $\forall u \in [r, \infty[, \phi_p(u) = 0$ .

De façon analogue, on définit l'ensemble  $\Lambda$  des distributions de possibilité sphériques bornées :

**Définition 29.** Une distribution de possibilité  $\pi$  appartient à  $\Lambda$  si elle est définie comme une fonction sur l'hypersphère  $B_r$  vérifiant :

—  $\pi$  est continue et  $\pi(0) = 1$

—  $\pi$  est radiale, i.e.  $\exists \phi_\pi : \mathbb{R} \rightarrow [0, 1]$ ,  $\pi(x) = \phi_\pi(\|x\|)$ ,

—  $\phi_\pi$  est décroissante sur  $[0, r[$ ,

—  $\forall u \in [r, \infty[, \phi_\pi(u) = 0$ .

Soit  $\pi^\Delta \in \Lambda$  une distribution de possibilité sphérique bornée définie par :

$$\pi^\Delta(x) = \max\left(0, 1 - \left(\frac{\|x\|}{r}\right)^n\right) \quad (3.36)$$

Nous avons alors le théorème suivant (dont la preuve est dans [120]) :

**Théorème 4.** (domination optimale de  $\pi^\Delta$ )

—  $\forall p \in \Theta, \forall A \in \mathcal{L}(\mathbb{R}^n)$ ,

$$\mathbb{P}_p(A) \leq \Pi_{\pi^\Delta}(A) := \sup_{x \in A} \pi^\Delta(x)$$

—  $\Pi_{\pi^\Delta}$  est la possibilité la plus spécifique qui domine toute probabilité induite par une distribution de probabilité  $p \in \Theta$  dans le sens où, avec  $\pi \in \Lambda$  étant une distribution de possibilité :

$$(\exists x \in \mathbb{R}^n, \pi(x) < \pi^\Delta(x)) \Rightarrow (\exists A \in \mathcal{L}(B_r), \exists p \in \Theta | \mathbb{P}_p(A) > \Pi_\pi(A)). \quad (3.37)$$

Ainsi,  $\pi^\Delta$  est la distribution de possibilité, la plus spécifique, dominant toutes distributions de possibilité obtenues en transformant une distribution de probabilité  $p \in \Theta$ . Cette distribution peut alors être utilisée pour approximer une distribution de possibilité à support non bornée en choisissant pour rayon  $r$  la distance telle que  $\mathbb{P}(B_r) = 0.9$ . Dans le cas bivarié, cela correspond à  $r = 3\sigma$  où  $\sigma$  est la variance maximale de la distribution de probabilité considérée (voir figure 3.7).

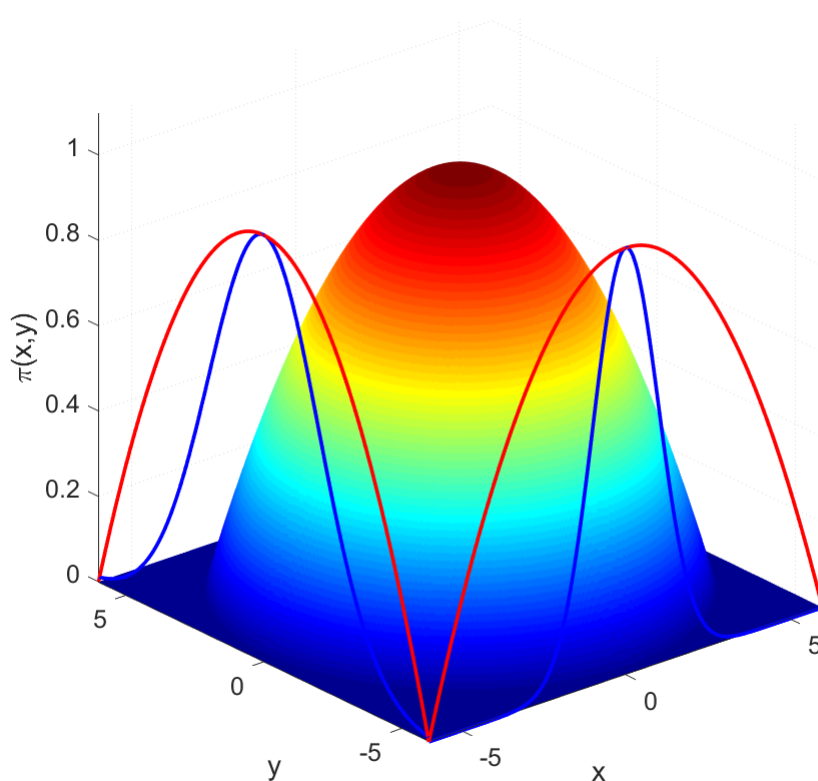


FIGURE 3.7 – Tracer de  $\pi^\Delta$  pour un rayon  $r = 3\sigma$  avec  $\sigma = 2$ . En bleu sont tracées les projections des marginales de la distribution de possibilité obtenue par transformation de la PDF  $p$  et en rouge sont tracées les projections des marginales de  $\pi^\Delta$ .

Dans la figure 3.7, nous pouvons observer que les marginales de la distribution ainsi construite (en rouge) dominent les marginales (en bleu) de la distribution de possibilité obtenue par la transformation de probabilité-possibilité de la distribution gaussienne.

## 3.5 Construction de distributions de possibilité multivariées à partir des marginales

Dans la section précédente, nous avons étendu la transformation de probabilité-possibilité optimale au cas des lois de probabilité elliptiques. Nous avons, de plus, déterminé les expressions analytiques de distributions de possibilité en  $n$ -dimensions avec l'équation (3.21) et en 2-dimensions avec les équations (3.22), (3.26) et (3.27).

Toutefois, lorsque l'on prend comme cadre de travail l'analyse multivariée de données, il n'est pas forcément aisé de déterminer les expressions de possibilité que nous venons d'évoquer. On peut alors s'attaquer au problème d'une autre manière : modéliser les lois marginales séparément puis les coupler à l'aide d'une fonction qui modélise la dépendance entre les variables en vue d'obtenir un modèle multivarié.

Dans le cadre de la théorie des probabilités, cette idée a été développée dans les travaux de Abe Sklar sous la désignation de copule du latin *copula* signifiant "lier ensemble, attacher". Ainsi, dans la suite de cette section, nous nous proposons d'étudier s'il est possible de trouver un équivalent aux copules dans le cadre de la théorie des possibilités. Nous commencerons par rappeler ce que sont les copules probabilistes ainsi que leurs principales propriétés. Puis nous présenterons un état de l'art des connecteurs de sous-ensembles flous en nous attardant plus particulièrement sur la notion de  $t$ -norme. Enfin, nous tenterons de comparer la distribution de possibilité obtenue par transformation de la loi normale multivariée avec le modèle déterminé en *couplant* des distributions de possibilité marginales équivalente chacune à une loi normale univariée.

### 3.5.1 Rappel sur les copules probabilistes

En probabilité, la théorie des copules est fondamentalement l'étude et la modélisation des dépendances entre variables aléatoires. Apparue dans les années 50 avec les travaux de Sklar [121] en se basant sur les travaux de Hoeffding [122], elle a rapidement su s'imposer dans différents domaines scientifiques tels que l'économie, la santé ou la physique.

Dans un premier temps, considérons un cadre général avec un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_n)$  à valeurs dans  $\mathbb{R}^n$  de fonction de répartition  $F(\mathbf{x})$  pour tout  $\mathbf{x} \in \mathbb{R}^n$ .

**Définition 30.** (Classes de Fréchet)

Soient  $F_1, \dots, F_n$   $n$  fonctions de répartition  $\mathbb{R} \rightarrow [0, 1]$ . La classe de Fréchet notée  $\mathcal{F}(F_1, \dots, F_n)$  est constituée de l'ensemble des fonctions de répartition de  $\mathbf{X} = (X_1, \dots, X_n)$  dont les marginales sont précisément  $F_1, \dots, F_n$ .

Cette classe de loi multivariée est souvent appelée classes de Fréchet et ces classes sont bornées.

**Proposition 6.** Pour tout  $F \in \mathcal{F}(F_1, \dots, F_n)$ ,  $\forall \mathbf{x} \in \mathbb{R}^n$ ,

$$F^-(\mathbf{x}) \leq F(\mathbf{x}) \leq F^+(\mathbf{x}) \quad (3.38)$$

où

$$F^+(\mathbf{x}) = \min\{F_i(x_i), i = 1, \dots, n\}, \quad (\text{borne supérieure de Fréchet})$$

et

$$F^-(\mathbf{x}) = \max\{0, F_1(x_1) + \dots + F_n(x_n) - (n - 1)\}. \quad (\text{borne inférieure de Fréchet})$$

À noter que  $F^+ \in \mathcal{F}(F_1, \dots, F_n)$ , mais que généralement  $F^- \notin \mathcal{F}(F_1, \dots, F_n)$  sauf dans le cas où  $n = 2$ .

Restreignons maintenant notre cadre de travail à seulement deux dimensions et introduisons la notion de copule.

**Définition 31.** (Copules en dimension 2)

Une copule  $C : [0, 1]^2 \rightarrow [0, 1]$  est une fonction de répartition dont les lois marginales sont uniformes sur  $[0, 1]$ .

**Théorème 5.** Une copule  $C : [0, 1]^2 \rightarrow [0, 1]$  bivariée est une fonction qui vérifie les propriétés suivantes

$$\text{— } \forall u_1, u_2 \in [0, 1], \quad C(u_1, 0) = C(0, u_2) = 0,$$

— Les marginales de  $C$  sont uniformes, i.e.

$$\forall u_1, u_2 \in [0, 1], \quad C(u_1, 1) = u_1 \quad \text{et} \quad C(1, u_2) = u_2,$$

—  $C$  est 2-croissante, i.e. pour tout  $0 \leq u_i \leq v_i \leq 1$ ,

$$C(v_1, v_2) + C(u_1, u_2) - C(v_1, u_2) - C(u_1, v_2) \leq 0.$$

À partir de cette définition de la copule en dimension deux, on peut suggérer un grand nombre de fonctions répondant au cahier des charges avec notamment certaines normes triangulaires ( $t$ -normes) que nous détaillerons un peu plus loin dans ce chapitre.

Considérons un vecteur aléatoire  $(X, Y)$  de fonction de répartition jointe  $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ . On peut choisir une fonction  $C$ , répondant à la définition d'une copule, afin de coupler les marginales  $F(x) = \mathbb{P}(X \leq x)$  et  $F(y) = \mathbb{P}(Y \leq y)$  entre elles. Ainsi, on écrira :

$$\mathbb{P}(X \leq x, Y \leq y) = C(\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)). \quad (3.39)$$

De la même façon, on peut lier les fonctions de survie à l'aide de la copule de survie  $C^*$  associée à  $C$  tel que :

$$\mathbb{P}(X \geq x, Y \geq y) = C^*(\mathbb{P}(X \geq x), \mathbb{P}(Y \geq y)) \quad (3.40)$$

où  $C^*$  est défini par :

$$C^*(a, b) = a + b - 1 + C(1 - a, 1 - b). \quad (3.41)$$

À noter que la notion de copule et les théorèmes énoncés ci après sont extensibles à toutes dimensions  $n \geq 2$  (avec de légères modifications). Toutefois, nous nous restreignons au cadre bidimensionnel dans un premier temps.

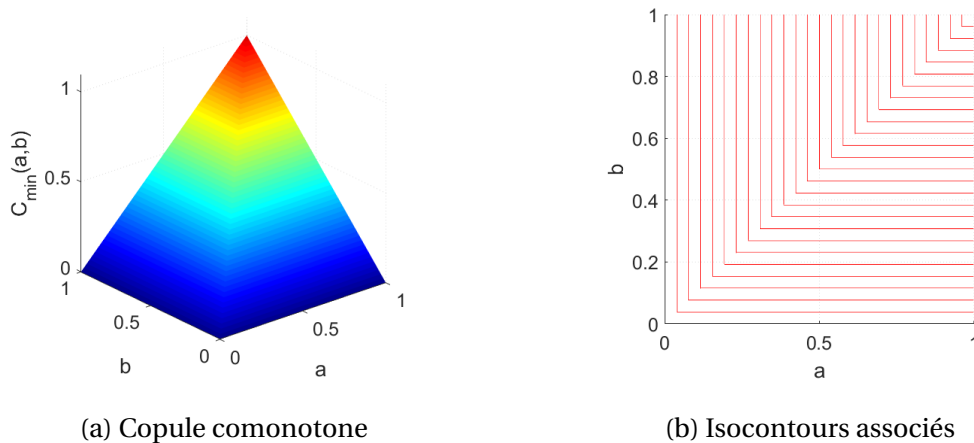


FIGURE 3.8 – Tracé de la copule  $C(a, b) = \min(a, b)$

**Théorème 6.** (Théorème d'encadrement [123])

Pour toute copule  $C$  bivariée et pour tout  $\mathbf{a} = (a_1, a_2) \in [0, 1]^2$ , la copule  $C(\mathbf{a})$  est bornée telle que :

$$C^-(\mathbf{a}) \leq C(\mathbf{a}) \leq C^+(\mathbf{a}) \quad (3.42)$$

où  $C^-$  et  $C^+$  sont les bornes de Fréchet telles que :

— borne inférieure :

$$C^-(\mathbf{a}) = \max(0, a_1 + a_2 - 1) \quad (3.43)$$

que l'on appelle aussi  $t$ -norme de Lukasiewicz ou  $t$ -norme anticomotone ( $T_L$ ),

— borne supérieure :

$$C^+(\mathbf{a}) = \min(a_1, a_2) \quad (3.44)$$

aussi dénommée  $t$ -norme minimum ou  $t$ -norme comotone ( $T_{\min}$ ).

Ce premier théorème sur les copules est un résultat important puisqu'il permet d'encadrer toutes copules par les bornes de Fréchet inférieure et supérieure.

Avant de présenter le théorème de Sklar, il est nécessaire d'introduire la notion de fonction quantile.

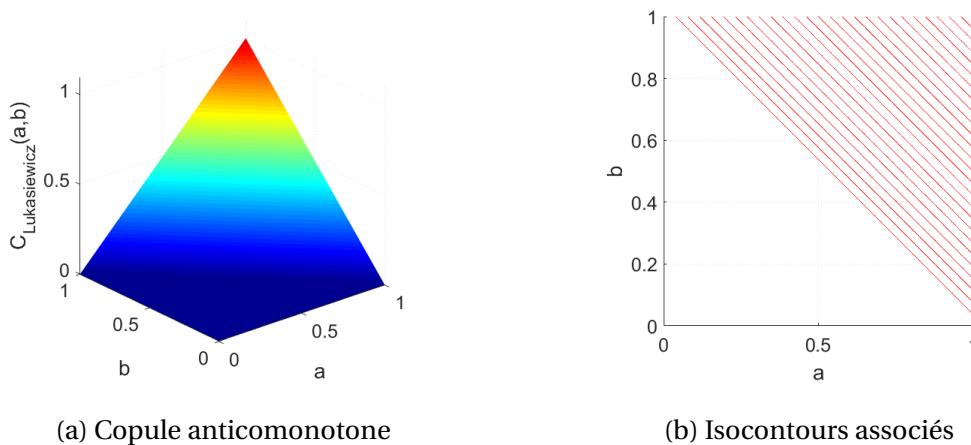


FIGURE 3.9 – Tracé de la copule  $C(a, b) = \max(0, a + b - 1)$

**Définition 32.** Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{R}$ , et  $F$  sa fonction de répartition. On appelle fonction quantile de  $X$ , la fonction notée  $F^{-1}$  de  $]0, 1[$  dans  $\mathbb{R}$ , qui à  $t \in ]0, 1[$  associe :

$$F^{-1}(t) = \inf\{x \mid F(x) = t\}.$$

**Proposition 7.** Soit  $F$  une fonction de répartition définie sur  $\mathbb{R}$ , croissante et continue à droite. Soit  $F^{-1}$  la fonction quantile associée à  $F$ . Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ .



Alors pour tout  $x \in \mathbb{R}$  et  $s \in (0, 1]$ ,

$$F(x) \geq s \Leftrightarrow x \geq F^{-1}(s).$$

**Théorème 7.** (Théorème de Sklar 1959 [121])

Si  $C$  est une copule, et  $F_1, F_2$  des fonctions de répartitions univariées, alors pour tout  $(x_1, x_2) \in \mathbb{R}^2$ ,

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \quad (3.45)$$

est une fonction de répartition de  $\mathcal{F}(F_1, F_2)$ .

Réciproquement, si  $F \in \mathcal{F}(F_1, F_2)$ , il existe une copule  $C$  satisfaisant l'équation (3.45). Cette copule n'est pas forcément unique, mais si les lois marginales  $F_1$  et  $F_2$  sont continues, elle le sera, avec, pour tout  $(a_1, a_2) \in [0, 1]^2$ ,

$$C(a_1, a_2) = F(F_1^{-1}(a_1), F_2^{-1}(a_2)), \quad (3.46)$$

où  $F_1^{-1}, F_2^{-1}$  sont les fonctions quantiles des fonctions  $F_1, F_2$ .

Le théorème de Sklar est un théorème majeur de la théorie des probabilités. Concrètement, ce théorème énonce que l'on peut créer un modèle multivarié relativement simple à partir de l'estimation seule des marginales et des structures de corrélations entre celles-ci.

Parmi les copules couramment utilisées dans la littérature, nous avons déjà présenté précédemment les bornes de Fréchet inférieure et supérieure avec la copule anticomotone et la copule comotone.

Nous introduisons maintenant la copule indépendante  $C^\perp$  comme copule liant les marginales d'un vecteur aléatoire  $X$  indépendant :

$$\forall \mathbf{a} = (a_1, a_2) \in \mathbb{R}^2 \quad C^\perp(a_1, a_2) = a_1 * a_2. \quad (3.47)$$

Cette copule est aussi une  $t$ -norme ( $T^\perp$ ).

Nous présentons ensuite la copule gaussienne  $C_g$  qui est une copule propre aux lois elliptiques telles que :

$$C(a_1, a_2) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Phi^{-1}(a_1)} \int_{-\infty}^{\Phi^{-1}(a_2)} \exp\left(\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) dx dy \quad (3.48)$$

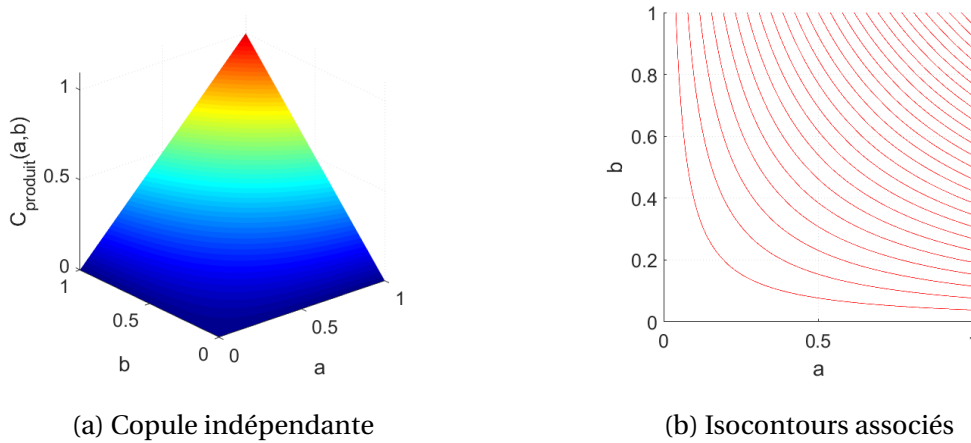


FIGURE 3.10 – Tracé de la copule  $C(a, b) = a \times b$

où  $r$  est le paramètre de corrélation tel que  $r \in [-1, 1]$  et  $\Phi$  est la fonction de répartition de la loi normale centrée réduite :

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$$

À noter qu’il existe dans la littérature un grand nombre de copules. Une famille de copules couramment utilisée est la famille des copules Archimédiennes [124] composée notamment des copules de Frank, de Gumbel et de Clayton [125]. Une seconde famille de copules est la famille des copules paramétrées qui permet de modéliser un large panel de dépendances entre variables aléatoires [126].

### 3.6 Construction d’une distribution de possibilité jointe à partir de distributions marginales

#### 3.6.1 État de l’art de la construction de distributions de possibilité bivariées à partir de distributions marginales

Dans la littérature scientifique [127], il est proposé de construire une distribution de possibilité jointe à partir des marginales et d’un opérateur d’agrégation  $h$  [128].

**Définition 33.** Une fonction  $h : [0, 1]^n \rightarrow [0, 1]$  est un opérateur d’agrégation si et seulement si il satisfait les conditions suivantes :

- $h$  est continue sur son domaine de définition,

- $h(0, 0, \dots, 0) = 0$  et  $h(1, 1, \dots, 1) = 1$  (condition de bords),
- $h$  est monotone, i.e.  $\forall i \in \mathbb{N}$  tel que  $a_i, b_i \in [0, 1]$ , si  $a_i \leq b_i$   
alors  $h(a_1, a_2, \dots, a_n) \leq h(b_1, b_2, \dots, b_n)$ ,
- $h$  est symétrique,  $h(a, b) = h(b, a)$ .

Cette fonction d'agrégation  $h$  peut se décliner sous trois principales formes. Elle peut être une  $t$ -norme  $T$ , une  $t$ -conorme  $S$  ou un opérateur de moyenne  $M$ .

**Les  $t$ -normes :**

Une  $t$ -norme se définit de la manière suivante [129] :

**Définition 34.** Une norme triangulaire ( $t$ -norme) est une opération  $T$  binaire sur l'intervalle unitaire  $[0, 1]$  qui est commutative, associative, monotone et a pour élément neutre 1, i.e., c'est une fonction  $T : [0, 1]^2 \rightarrow [0, 1]$  telle que pour tout  $x, y, z \in [0, 1]$  :

- (T1)  $T(x, y) = T(y, x)$ ,
- (T2)  $T(x, T(y, z)) = T(T(x, y), z)$ ,
- (T3) si  $y \leq z$  alors  $T(x, y) \leq T(x, z)$ ,
- (T4)  $T(x, 1) = x$ .

Les normes triangulaires permettent de définir l'intersection entre plusieurs ensembles flous, c'est-à-dire l'opération logique *ET*.

En notant  $\pi_A(x)$  et  $\pi_B(y)$  deux distributions de possibilité marginale sur les ensembles  $A$  et  $B$  respectivement, l'intersection est donnée par :

$$\pi_{(A \cap B)}(x, y) = T(\pi_A(x), \pi_B(y)). \quad (3.49)$$

Il est important de noter que nombres de copules sont des  $t$ -normes.

Ainsi, une  $t$ -norme est une copule si cette  $t$ -norme est 2-croissante (en dimension 2).

Dans l'autre sens, une copule associative et commutative est une  $t$ -norme. De façon similaire au théorème d'encadrement pour les copules, il existe un théorème d'encadrement pour les  $t$ -normes :

**Théorème 8.** Soit une  $t$ -norme  $T(a, b)$ . Pour tout  $a, b \in [0, 1]$ , on a :

$$T_D(a, b) \leq T(a, b) \leq T_{\min}(a, b), \quad (3.50)$$

où  $T_D$  est la  $t$ -norme drastique [129].

**Problème observé :**

En lien avec la partie précédente, il est intéressant de comparer la construction d'une distribution de possibilité bivariée elliptique dont les marginales seraient indépendantes à l'aide de la transformation possibiliste de lois de probabilité elliptiques et avec l'aide de la  $t$ -normes indépendante  $T^\perp$ . En effet, nous savons qu'avec la première méthode la distribution de possibilité est à la fois consistante et la plus spécifique. Elle constitue donc un point de comparaison important. Ainsi, pour illustrer l'utilisation de ce cas de figures, considérons deux distributions de probabilité gaussiennes indépendantes  $p(x)$  et  $p_Y(y)$  de moyennes respectives  $\mu_x$  et  $\mu_y$  et de variances  $\sigma_x$  et  $\sigma_y$ . On aura donc  $p_{X,Y}(x, y) = p(x)p_Y(y)$ .

Si on applique la transformation de probabilité-possibilité elliptique on obtient une distribution  $\pi_{X,Y}^t(x, y)$  de la forme de l'équation (3.22), c'est-à-dire :

$$\begin{aligned} \pi_{X,Y}^t(x, y) &= \exp \left( -\frac{1}{2} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}^T \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \right) \\ &= \exp \left( -\frac{1}{2} \frac{(x - \mu_x)^2}{\sigma_x^2} \right) \exp \left( -\frac{1}{2} \frac{(y - \mu_y)^2}{\sigma_y^2} \right). \end{aligned} \quad (3.51)$$

D'autre part, lorsque l'on applique la transformation  $p \rightarrow \pi$  1D sur chacune des marginales  $p_X(x)$  et  $p_Y(y)$  on obtient deux distributions de possibilité marginales  $\pi_X^m(x)$  et  $\pi_Y^m(y)$  de la forme de l'équation (2.32). C'est-à-dire :

$$\pi_X^m(x) = 1 - \operatorname{erf} \left( \frac{|x - \mu_x|}{\sigma_x \sqrt{2}} \right) \quad \text{et} \quad \pi_Y^m(y) = 1 - \operatorname{erf} \left( \frac{|y - \mu_y|}{\sigma_y \sqrt{2}} \right). \quad (3.52)$$

Or, comme nous l'avons dit dans le chapitre 2, la forme normalisée de la distribution gaussienne est consistante avec la distribution de probabilité gaussienne. Cependant, celle-ci n'est pas la distribution de possibilité la plus spécifique, i.e.

$$\pi_X^m(x) \leq \exp \left( -\frac{1}{2} \frac{(x - \mu_x)^2}{\sigma_x^2} \right) \quad \text{et} \quad \pi_Y^m(y) \leq \exp \left( -\frac{1}{2} \frac{(y - \mu_y)^2}{\sigma_y^2} \right). \quad (3.53)$$

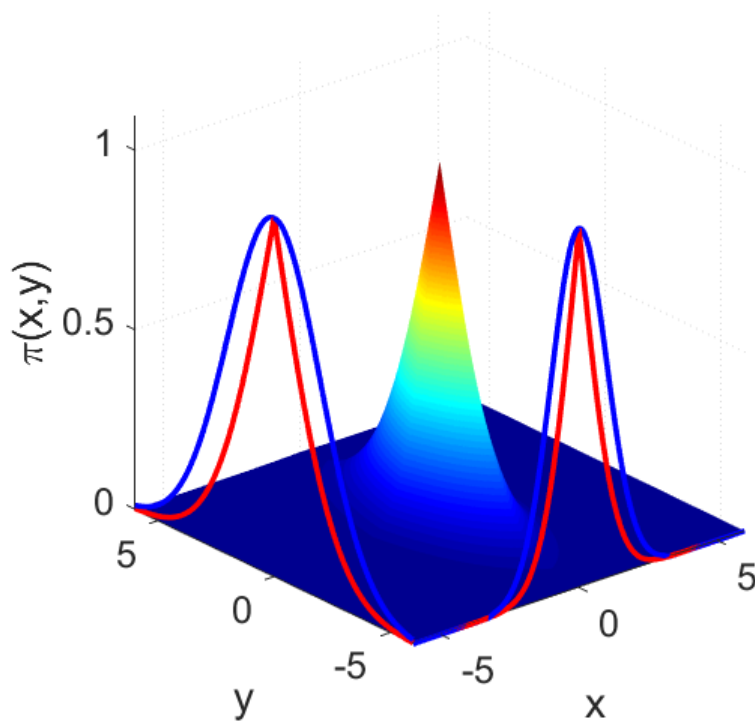


FIGURE 3.11 – Exemple de distribution de possibilité jointe obtenue avec la copule indépendante. En bleu sont tracées les projections marginales de  $\pi_{X,Y}^t(x, y)$  et en rouge les projections marginales de  $\pi_{X,Y}(x, y) = \pi_X^m(x) \times \pi_Y^m(y)$ .

On en déduit alors :

$$\pi_X^m(x)\pi_Y^m(y) \leq \pi_{X,Y}^t(x, y). \quad (3.54)$$

Ainsi, la distribution de possibilité jointe construite à partir de deux distributions de possibilité marginales indépendantes et de la copule indépendante n'est pas consistante avec la probabilité jointe. Cela constitue un problème puisque l'on ne peut pas garantir la conservation de la propriété de consistance lors de l'utilisation d'une copule comme opérateur d'agrégation.

Nous étudions, maintenant, le potentiel des autres opérateurs d'agrégation comme candidats permettant de construire des distributions de possibilité jointe consistante et suffisamment spécifique avec la distribution de probabilité jointe.

**Les  $t$ -conormes :**

Les  $t$ -conormes sont les fonctions duales associées aux  $t$ -normes. Étant donné une  $t$ -norme  $T$ , on dérive la  $t$ -conorme associée  $S$  pour tout  $a, b \in [0, 1]$  par :

$$S(a, b) = 1 - T(1 - a, 1 - b). \quad (3.55)$$

De manière générale, on définit alors les  $t$ -conormes telles que :

**Définition 35.** Une conorme triangulaire ( $t$ -conorme) est une opération  $S$  binaire sur l'intervalle unitaire  $[0, 1]$  commutative, associative, monotone et dont l'élément neutre est 0, i.e., c'est une fonction  $S : [0, 1]^2 \rightarrow [0, 1]$  telle que pour tout  $x, y, z \in [0, 1]$  :

$$(T1) S(x, y) = S(y, x),$$

$$(T2) S(x, S(y, z)) = S(S(x, y), z),$$

$$(T3) \text{ si } y \leq z \text{ alors } S(x, y) \leq S(x, z),$$

$$(T4) S(x, 0) = x.$$

Les  $t$ -conormes définissent l'opération binaire *ou* entre deux ensembles flous, elles représentent donc l'union. En reprenant les notations précédentes, on notera :

$$\pi_{(A \cup B)}(x, y) = S(\pi_A(x), \pi_B(y)). \quad (3.56)$$

Parmi les  $t$ -conormes, les opérateurs ci-dessous en sont des exemples courants :

— L'opérateur maximum  $S_{\max}$  (dual de  $T_{\min}$ ) :

$$S_{\max}(a, b) = \max(a, b) \quad (3.57)$$

— L'union probabiliste  $S_P$  (duale de  $T^\perp$ ) :

$$S_P(a, b) = a + b - ab \quad (3.58)$$

— L'union de Lukasiewicz  $S_L$  (duale de  $T_L$ ) :

$$S_L(a, b) = \min(a + b - 1, 0) \quad (3.59)$$

— L'union forte  $S_S$  :

$$S_S(a, b) = \begin{cases} \max(a, b) & \text{si } \min(a, b) = 0, \\ 1 & \text{sinon.} \end{cases} \quad (3.60)$$

Enfin, il existe un théorème d'encadrement sur les  $t$ -conormes :

**Théorème 9.** [6]

Soit une  $t$ -conorme  $S(a, b)$ , pour tout  $a, b \in [0, 1]$  :

$$S_{\max}(a, b) \leq S(a, b) \leq S_S(a, b). \quad (3.61)$$

Les  $t$ -conormes semblent, au premier abord, constituer des candidats intéressants pour obtenir une distribution de possibilité jointe consistante avec la distribution de probabilité.

Cependant, lorsque les distributions marginales sont continues, la distribution de possibilité jointe obtenue possède une infinité de modes et ne représente donc plus l'information initiale que l'on voulait modéliser (même si la distribution peut-être consistante). De plus, puisque les isocontours sont des intervalles ouverts (voir figure 3.12), l'indice de spécificité tend vers l'infini.

Dans le pire des cas, si on choisit la  $t$ -conorme  $S_S$ , la distribution jointe peut être égale à 1 en tout point. En effet, pour des distributions de possibilité marginales de la forme de l'équation (2.32) donc continues,  $\min(\pi_m(x), \pi_m(y))$  est toujours différents de 0 et donc  $S_S(\pi_m(x), \pi_m(y)) = 1$ .

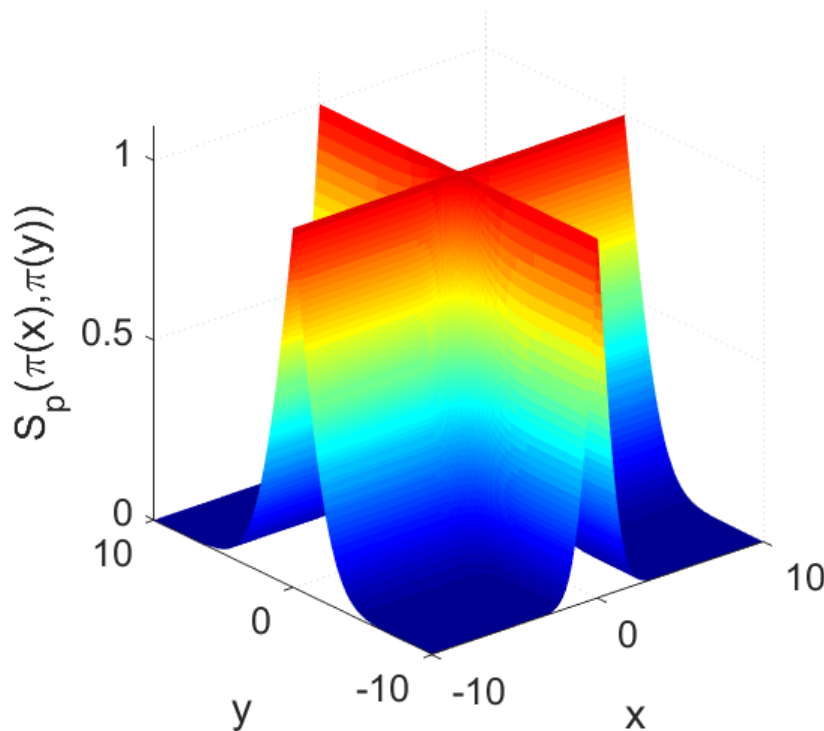


FIGURE 3.12 – Tracé de la  $t$ -conorme "union probabiliste"  $\pi_{X,Y}^S(x, y) = S_p(\pi_X^m(x), \pi_Y^m(y))$ .

**Les Opérateurs de moyennes :**

La dernière forme de l'opérateur d'agrégation est l'opérateur de moyenne  $M$  qui est défini de la manière suivante :

**Définition 36.** Une fonction  $M : [0, 1]^2 \rightarrow [0, 1]$  est un opérateur de moyenne si et seulement si  $M$  satisfait les propriétés suivantes pour tout  $x, y, z \in [0, 1]$  :

(Symétrie)  $M(x, y) = M(y, x)$ ,

(Idempotence)  $M(x, x) = x$ ,

(Monotonie) si  $y \leq z$  alors  $M(x, y) \leq M(x, z)$ .

Tout opérateur de moyenne est borné.

**Théorème 10.** [6]

Soit un opérateur de moyenne  $M(a, b)$ , pour tout  $a, b \in [0, 1]$  :

$$T_{\min}(a, b) \leq M(a, b) \leq S_{\max}(a, b). \quad (3.62)$$

Nous donnons, ci-dessous, quelques formes couramment utilisées d'opérateurs de moyenne :

— La Moyenne harmonique  $M_H$  :

$$M_H(a, b) = \frac{2ab}{a+b} \quad (3.63)$$

— La Moyenne géométrique  $M_g$  :

$$M_g(a, b) = \sqrt{ab} \quad (3.64)$$

— La Moyenne arithmétique  $M_a$  :

$$M_a(a, b) = \frac{a+b}{2} \quad (3.65)$$

Les opérateurs de moyenne constituent un compromis pour construire une distribution de possibilité jointe. En effet, la distribution jointe obtenue à l'aide de cet opérateur n'est pas consistante dans la majorité des cas mais elle se rapproche du modèle désiré.

On peut assez facilement montrer que la distribution de possibilité obtenue grâce aux opérateurs de moyenne harmonique  $M_H$  et de moyenne géométrique  $M_g$  n'est pas consistante avec la loi de probabilité jointe constituée par deux lois de probabilité marginales



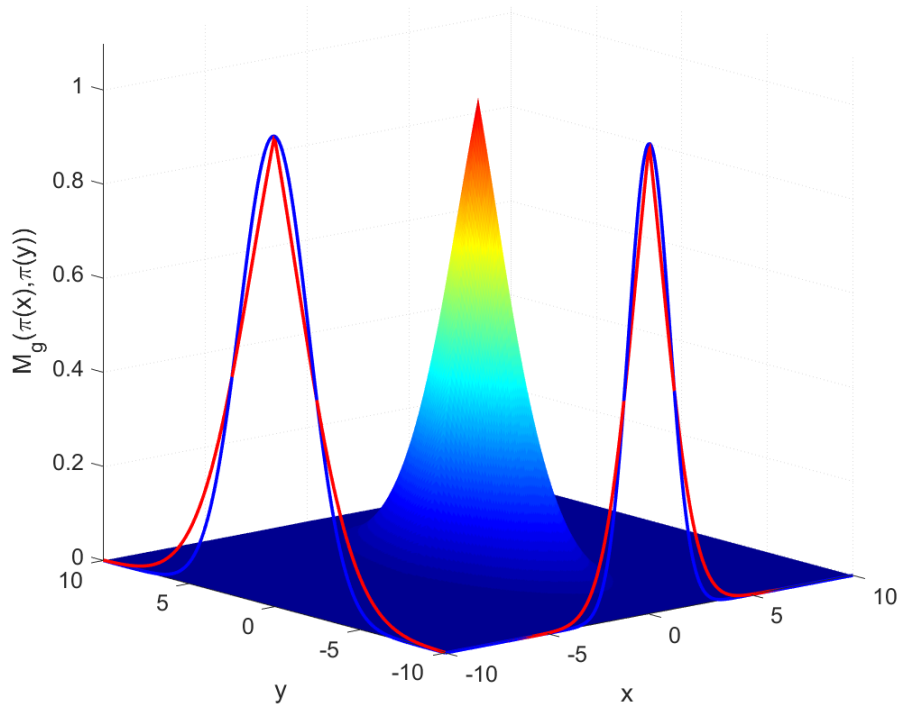


FIGURE 3.13 – Tracé de l’opérateur de moyenne géométrique  $\pi_{X,Y}^{M_g}(x, y) = M_g(\pi_X^m(x), \pi_Y^m(y))$ . En rouge sont données les projections marginales de la distribution  $\pi_{X,Y}^{M_g}(x, y)$  et bleu les projections de la distribution de possibilité dont la forme est donnée dans l’équation (3.22).

gaussiennes, indépendantes l’une de l’autre. Ils permettent toutefois de s’approcher de la distribution de possibilité obtenue en appliquant la transformation elliptique, comme on peut le voir dans la figure 3.13.

La figure 3.14 résume les différents opérateurs que l’on peut utiliser pour agréger deux distributions de possibilité marginales afin d’obtenir une distribution de possibilité jointe.

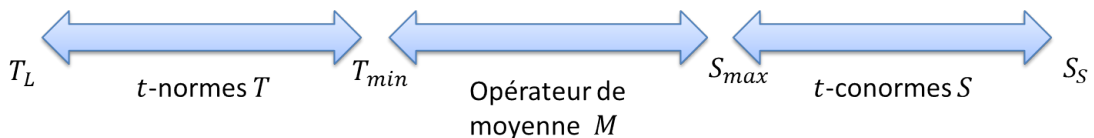


FIGURE 3.14 – Schéma représentant l’ensemble des opérateurs d’agrégation possibiliste.

Ainsi, l’idée de construire des distributions de possibilité multivariées à partir des connecteurs de la théorie des ensembles flous peut paraître séduisante et a déjà été

exploitée notamment dans les travaux de Salicone *et al* [127]. Elle présente l'avantage de permettre facilement la construction de distribution de possibilité multivariée. En revanche, l'inconvénient de cette approche est qu'elle ne permet pas de garantir que la distribution de possibilité ainsi construite soit consistante avec la distribution de probabilité équivalente.

Ci-après nous nous attaquons à ce problème par une approche basée sur les copules.

### 3.6.2 Solution proposée

Dans cette section nous proposons des solutions pour construire une distribution de possibilité consistante avec la loi de probabilité jointe à partir des distributions de possibilité équivalentes aux distributions unimodales symétriques de probabilité marginales.

Nous distinguons deux manières de procéder : soit on construit la distribution de possibilité équivalente centrée sur la valeur modale, soit on construit la distribution de possibilité équivalente centrée sur le point extrême de la fonction de répartition jointe, c'est-à-dire un point situé à l'infini.

#### 3.6.2.1 Distribution centrée sur un point extrême :

Notons  $\pi_{X,Y}(x, y) = C'(\pi_X(x), \pi_Y(y))$  la distribution de possibilité jointe et  $C'$  une fonction d'agrégation. Si chaque marginale possibiliste obtenue à partir de la distribution de probabilité marginale est centrée sur le point extrême de la fonction de répartition alors cela revient à dire que  $\pi_X(x) = F_X(x)$  et  $\pi_Y(y) = F_Y(y)$ . Le point autour duquel la distribution jointe est centrée est le point dont les coordonnées seraient  $+\infty$  sur l'axe des  $x$  et  $+\infty$  sur l'axe des  $y$ .

Dans ce cas particulier, il est évident que le théorème de Sklar, qui est vrai pour les fonctions de répartitions, reste valide pour les distributions de possibilité équivalente. En effet, si on pose  $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$ , avec  $C$  est une copule, il suffit d'avoir  $C = C'$  pour que la distribution de possibilité soit consistante avec la loi de probabilité.

On remarquera que cela reste valable pour les trois autres cas de points extrêmes, c'est-à-dire :

$$\text{— } \pi_X(x) = F_X(x) \text{ et } \pi_Y(y) = 1 - F_Y(y)$$

—  $\pi_X(x) = 1 - F_X(x)$  et  $\pi_Y(y) = F_Y(y)$

—  $\pi_X(x) = 1 - F_X(x)$  et  $\pi_X(y) = 1 - F_Y(y)$

Cependant le cas qui nous intéresse plus particulièrement est celui où la distribution de possibilité respecte le principe de consistance tout en étant centrée sur la valeur modale.

### 3.6.2.2 Distribution centrée sur la valeur modale :

Pour simplifier l'exposé, nous considérerons deux distributions de probabilité marginales symétriques et unimodales  $p_X(x)$  et  $p_Y(y)$ , centrées en zéro, de variance  $\sigma_x = \sigma_y = 1$  et de fonctions cumulatives respectives  $F_X(x)$  et  $F_Y(y)$ .

On appellera  $\pi_X(x) = \min(2F_X(x), 2(1 - F_X(x)))$  et  $\pi_Y(y) = \min(2F_Y(y), 2(1 - F_Y(y)))$ ,

les distributions de possibilité marginales équivalentes associées à  $p_X(x)$  et  $p_Y(y)$ .

L'objectif que nous nous donnons dans cette partie est de considérer l'extension du théorème de Sklar aux distributions de possibilité. Concrètement, nous recherchons une fonction d'agrégation  $C'$  permettant d'agréger les distributions équivalentes de possibilité marginales de façon à obtenir une distribution de possibilité jointe centrée et consistante avec la distribution de probabilité jointe.

Le problème d'exprimer une distribution équivalente de possibilité jointe centrée en fonction des marginales possibilistes équivalentes centrées est plus complexe dans le cas d'une distribution centrée sur un point extrême. En effet, comme nous l'avons dit précédemment, utiliser les connecteurs de la théorie des ensembles flous ne permet pas toujours d'arriver à ce résultat.

En se basant sur la théorie des copules pour combiner les fonctions de répartition marginales, notre approche consiste à exprimer la probabilité accumulée dans une région  $A_{x,y}$  délimitée par un contour qui permet une expression de la probabilité en fonction des marginales de la distribution de probabilité jointe et ceci pour tous les points  $(x, y)^T$  de l'espace.

Puis, nous construirons la distribution équivalente de possibilité à l'aide de la formule suivante :

$$\pi_{X,Y}(x, y) = 1 - \mathbb{P}(A_{x,y}). \quad (3.66)$$

À noter que si les isocontours de la région  $A_{x,y}$  coïncident avec les coupes de la densité de probabilité jointe alors d'après le théorème 3, nous savons que la distribution de possibilité, ainsi construite, est consistante avec  $\mathbb{P}$ , et d'après le lemme 2 que celle-ci est la plus spécifique.

Le problème qu'il se pose alors est celui de déterminer des formes d'isocontours suffisamment simples pour pouvoir exprimer  $\mathbb{P}(A_{x,y})$  en fonction des fonctions de répartitions marginales, tout en se rapprochant au maximum des coupes de la densité de probabilité jointe.

Les contours pour lesquels nous sommes capables d'exprimer  $\mathbb{P}(A_{x,y})$  en fonction des fonctions de répartition marginales quelques soient les distributions de probabilités sont les contours rectangulaires (droits ou inclinés) qui correspondent à la distance infinie et à la distance de Manhattan.

L'idée est alors de choisir une distance par rapport à la valeur modale de la densité de probabilité afin de pouvoir définir les dimensions de la région  $A_{x,y}$  de façon à ce que cette région soit incluse dans les coupes de la densité.

De ce fait, en construisant la distribution de possibilité de la façon suivante, nous garantissons que la distribution de possibilité jointe sera consistante avec la distribution de probabilité jointe correspondante.

En effet, si on note  $A_j$  la coupe de la distribution de probabilité jointe, nous aurons alors  $A_{x,y} \subseteq A_j$  et donc  $\mathbb{P}(A_{x,y}) \leq \mathbb{P}(A_j)$ . Cela implique que la distribution de possibilité  $\pi_{X,Y}(x, y)$  construite en considérant des régions  $A_{x,y}$  sera plus grande que la distribution de possibilité consistante et maximum spécifique obtenue à partir de la distribution de probabilité jointe.

Nous allons maintenant considérer l'extension possibiliste du théorème de Sklar pour les trois principales copules :  $C_{\min}$ ,  $C_L$  et  $C^\perp$ .

**Copule  $C_{\min}$  :**

**Proposition 8.** *Pour des régions  $A_{x,y}$  carrées, la distribution de possibilité, consistante avec une distribution de probabilité jointe, construite à partir de deux distributions marginales comonotonnes est :*

$$\pi_{X,Y}^{\min}(x, y) = \min(\pi_X(x), \pi_Y(y)). \quad (3.67)$$

*Démonstration :*

Dans un premier temps, nous supposons que les régions  $A_{x,y}$  sont carrées. Pour chaque point  $(x, y)^T$ , on définit la région carrée  $A_{x,y}^{\square}$  délimitée par l'isocontour tel que :

$$A_{x,y}^{\square} = \{u, v \in \mathbb{R}^2, \max(|u|, |v|) \leq \max(|x|, |y|)\}. \quad (3.68)$$

Comme on peut le voir sur la figure 3.15, un point quelconque  $(x, y)^T$  définit un carré (noté HIJK) qui représente l'isocontour d'une région  $A_{x,y}^{\square}$  et dont les coordonnées des coins sont  $H = \begin{pmatrix} \max(|x|, |y|) \\ \max(|x|, |y|) \end{pmatrix}$ ,  $I = \begin{pmatrix} \max(|x|, |y|) \\ -\max(|x|, |y|) \end{pmatrix}$ ,  $J = \begin{pmatrix} -\max(|x|, |y|) \\ -\max(|x|, |y|) \end{pmatrix}$  et  $K = \begin{pmatrix} -\max(|x|, |y|) \\ \max(|x|, |y|) \end{pmatrix}$ .

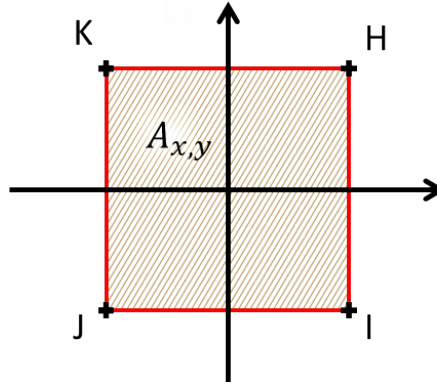


FIGURE 3.15 – isocontour carré défini à partir d'un point  $(x, y)$ .

**Remarque :**

Si l'on désire que la valeur modale soit  $\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ , alors l'intervalle  $A_{x,y}^{\square}$  défini par l'isocontour associé devient  $A_{x,y}^{\square} = \{u, v \in \mathbb{R}^2, \max(|u|, |v|) \leq \max(|x - \mu_x|, |y - \mu_y|)\}$ .

De même, si les distributions de probabilité marginales ont des paramètres d'échelles  $\sigma_x$  et  $\sigma_y$ , alors on écrit  $A_{x,y}^{\square}$  de la façon suivante :

$$A_{x,y}^{\square} = \{u, v \in \mathbb{R}^2, \max(|u|, |v|) \leq \max\left(\frac{|x - \mu_x|}{\sigma_x}, \frac{|y - \mu_y|}{\sigma_y}\right)\}.$$

Par souci de simplicité, nous considérerons, pour la suite, un isocontour centré en zéro pour des variances de distributions marginales égales à un.

Ainsi, dans le cas où la région considérée est carrée tel qu'énoncé dans l'équation (3.68), on peut écrire :

$$\begin{aligned} \mathbb{P}(A_{x,y}^{\square}) &= F_{X,Y}(\max(|x|, |y|), \max(|x|, |y|)) + F_{X,Y}(-\max(|x|, |y|), -\max(|x|, |y|)) \\ &\quad - F_{X,Y}(-\max(|x|, |y|), \max(|x|, |y|)) - F_{X,Y}(\max(|x|, |y|), -\max(|x|, |y|)), \end{aligned} \quad (3.69)$$

avec  $F_{X,Y}(x, y)$  la fonction de répartition associée à la distribution de probabilité jointe (que l'on ne connaît pas).

Cependant, d'après le théorème de Sklar (théorème 7), on a :

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)), \quad (3.70)$$

où  $C$  est la copule (unique si les fonctions de répartition marginales sont continues) permettant de lier  $F_X$  et  $F_Y$ .

On peut alors écrire l'équation (3.69) telle que :

$$\begin{aligned} \mathbb{P}(A_{x,y}^{\square}) &= C(F_X(\max(|x|, |y|)), F_Y(\max(|x|, |y|))) + C(F_X(-\max(|x|, |y|)), F_Y(-\max(|x|, |y|))) \\ &\quad - C(F_X(-\max(|x|, |y|)), F_Y(\max(|x|, |y|))) - C(F_X(\max(|x|, |y|)), F_Y(-\max(|x|, |y|))). \end{aligned}$$

Et pour une distribution de probabilité symétrique centrée en zéro, on sait que

$$F_X(-x) = 1 - F_X(x).$$

Nous avons alors :

$$\begin{aligned} \mathbb{P}(A_{x,y}^{\square}) &= C(F_X(\max(|x|, |y|)), F_Y(\max(|x|, |y|))) + C(1 - F_X(\max(|x|, |y|)), 1 - F_Y(\max(|x|, |y|))) \\ &\quad - C(1 - F_X(\max(|x|, |y|)), F_Y(\max(|x|, |y|))) - C(F_X(\max(|x|, |y|)), 1 - F_Y(\max(|x|, |y|))). \end{aligned}$$

En posant  $\forall x \in \mathbb{R}$ ,  $F(x) = F_X(x) = F_Y(x)$  et en utilisant la propriété de symétrie de la fonction copule, nous en déduisons :

$$\begin{aligned} \mathbb{P}(A_{x,y}^{\square}) &= C(F(\max(|x|, |y|)), F(\max(|x|, |y|))) + C(1 - F(\max(|x|, |y|)), 1 - F(\max(|x|, |y|))) \\ &\quad - 2C(1 - F(\max(|x|, |y|)), F(\max(|x|, |y|))). \end{aligned} \quad (3.71)$$

Il reste alors à remplacer la copule  $C$  par  $C_{\min}$  afin de modéliser la comonotonie entre les lois marginales.

La distribution de possibilité construite est alors consistante avec la loi de probabilité jointe. Elle est, de plus, maximum spécifique si les isocontours de la distribution de probabilité jointe sont des carrés de mêmes tailles. Si les régions  $A_{x,y}^{\square}$  sont par contre complètement inclus dans les isocontours de la distribution de probabilité jointe alors la distribution équivalente de possibilité est seulement consistante (démonstration triviale) mais pas obligatoirement la plus spécifique.

Ainsi, pour  $C(a, b) = \min(a, b)$ , on montre que lorsque deux distributions de probabilité sont comonotones alors la distribution de possibilité jointe  $\pi_{X,Y}^{\min}(x, y)$  est

$$\begin{aligned}
 \pi_{X,Y}^{\min}(x, y) &= 1 - \mathbb{P}(A_{x,y}^{\square}) \\
 &= 1 - \min(F(\max(|x|, |y|)), F(\max(|x|, |y|))) - \min(1 - F(\max(|x|, |y|)), 1 - F(\max(|x|, |y|))) \\
 &\quad + 2\min(1 - F(\max(|x|, |y|)), F(\max(|x|, |y|))) \\
 &= 1 - F(\max(|x|, |y|)) - 1 + F(\max(|x|, |y|)) + 2\min(1 - F(\max(|x|, |y|)), F(\max(|x|, |y|))) \\
 &= 2\min(1 - F(\max(|x|, |y|)), F(\max(|x|, |y|))) \tag{3.72}
 \end{aligned}$$

Or la fonction de répartition  $F$  est croissante sur  $\mathbb{R}$  et centrée (i.e.  $F(0) = \frac{1}{2}$ )

donc  $\forall x \in \mathbb{R}, F(|x|) \geq \frac{1}{2}$ . On en déduit que :

$$\pi_{X,Y}^{\min}(x, y) = 2(1 - F(\max(|x|, |y|))). \tag{3.73}$$

En regardant pour le cas  $x < 0$  et  $y < 0$ , on montre alors que

- pour  $x < y$  :  $\pi_{X,Y}^{\min}(x, y) = 2(1 - F_X(-x)) = 2F_X(x)$ ,
- pour  $x > y$  :  $\pi_{X,Y}^{\min}(x, y) = 2(1 - F_Y(-y)) = 2F_Y(y)$ .

Ce qui nous donne

$$\pi_{X,Y}^{\min}(x, y) = \min(2F_X(x), 2F_Y(y)) \tag{3.74}$$

Puis en regardant les trois autres cas, pour  $x < 0, y > 0$ , puis  $x > 0, y < 0$  et enfin  $x > 0, y > 0$ , on en déduit la forme finale de la distribution de possibilité jointe construite à partir de distributions de probabilités marginales comonotones :

$$\pi_{X,Y}^{\min}(x, y) = \min(\pi_X(x), \pi_Y(y)). \tag{3.75}$$

Nous obtenons, ainsi, un premier résultat important : dans le cas de lois de probabilité marginales comonotones, lorsque la dépendance est la plus forte, la distribution équivalente de possibilité est obtenue en appliquant la copule comonotone sur les distributions de possibilités marginales. Ce résultat est cohérent avec celui obtenu d'une autre manière par [130].

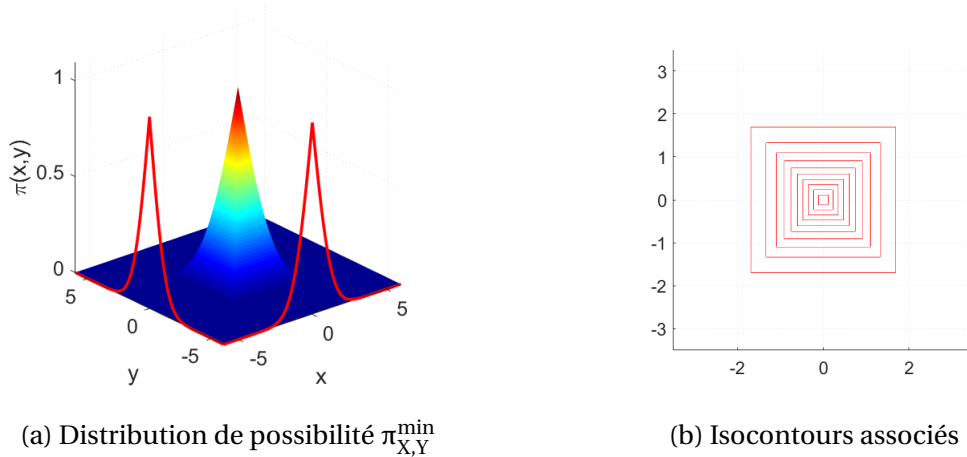


FIGURE 3.16 – Tracé de la distribution de possibilité construite à partir de distributions de probabilité marginales comonotones.

**Copule  $C_L$  :**

Lorsque la dépendance considérée entre les lois de probabilité marginales est l'anticomonotonie, il est nécessaire de considérer des formes de régions de dispersion différentes.

En effet, dans le cas particulier où les lois marginales sont des lois uniformes anticomonotones, la distribution de possibilité construite en agrégeant les quatre distributions de possibilité centrées sur les points extrêmes semble suggérer que les coupes de la distribution sont des losanges.

Ainsi, dans le cas de l'anticomonotonie, considérer des régions  $A_{x,y}$  de formes losange incluses dans les coupes de la densité de probabilité jointe permet de construire des distributions équivalentes de possibilité consistantes.

De plus, si les lois marginales sont des lois uniformes alors la distribution de possibilité, ainsi, construite est consistante et maximum spécifique.

Supposons, donc, que nous considérons des régions  $A_{x,y}$  telles que celles-ci sont des losanges droits.

Pour chaque point  $(x, y)^T$  on définit alors une région  $A_{x,y}$  délimité par l'isocontour tel que :

$$A_{x,y}^\diamond = \{u, v \in \mathbb{R}^2, |u| + |v| \leq |x| + |y|\}. \tag{3.76}$$

Cependant, lorsque les isocontours sont des losanges, on ne peut pas appliquer directe-



ment la même méthode pour calculer la probabilité accumulée sur l'intervalle  $A_{x,y}^\diamond$  que dans le cas des régions carrées.

Ce que l'on peut faire, c'est appliquer un changement de repère afin de transformer le losange droit en un carré dans un nouveau repère. Pour réaliser ceci, on effectue une rotation centrée en l'origine du repère d'un angle  $\frac{\pi}{4}$  à l'aide d'une matrice de rotation  $Rot$  telle que :

$$Rot = \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) \\ -\sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix}. \quad (3.77)$$

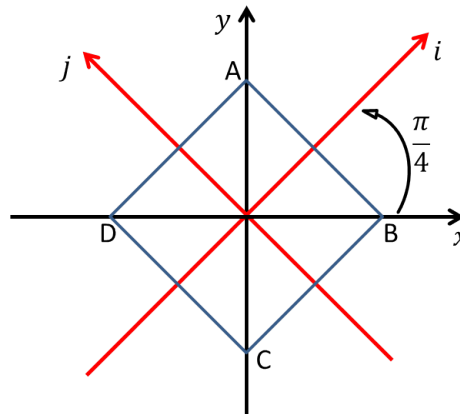


FIGURE 3.17 – Changement de repère, rotation d'angle  $\frac{\pi}{4}$ .

Dans le nouveau repère et en notant que  $\cos\left(\frac{\pi}{4}\right) = \sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}$ , les points A, B, C et D ont alors pour coordonnées (voir figure 3.17) :

$$A = \frac{\sqrt{2}}{2} \begin{pmatrix} |x| + |y| \\ |x| + |y| \end{pmatrix}, B = \frac{\sqrt{2}}{2} \begin{pmatrix} |x| + |y| \\ -(|x| + |y|) \end{pmatrix}, C = -\frac{\sqrt{2}}{2} \begin{pmatrix} |x| + |y| \\ |x| + |y| \end{pmatrix} \text{ et } D = \frac{\sqrt{2}}{2} \begin{pmatrix} -(|x| + |y|) \\ |x| + |y| \end{pmatrix}.$$

On peut, maintenant, calculer la probabilité accumulée dans l'intervalle  $A_{x,y}^\diamond$  délimité par le losange ABCD telle que :

$$\begin{aligned} \mathbb{P}(A_{x,y}^\diamond) &= F_{x,y} \left( \frac{\sqrt{2}}{2}(|x| + |y|), \frac{\sqrt{2}}{2}(|x| + |y|) \right) + F_{x,y} \left( -\frac{\sqrt{2}}{2}(|x| + |y|), -\frac{\sqrt{2}}{2}(|x| + |y|) \right) \\ &\quad - F_{x,y} \left( \frac{\sqrt{2}}{2}(|x| + |y|), -\frac{\sqrt{2}}{2}(|x| + |y|) \right) - F_{x,y} \left( -\frac{\sqrt{2}}{2}(|x| + |y|), \frac{\sqrt{2}}{2}(|x| + |y|) \right). \end{aligned} \quad (3.78)$$

Puis en réutilisant le théorème de Sklar, on remplace  $F_{X,Y}(x, y)$  par  $C(F_X(x), F_Y(y))$  et on

obtient :

$$\begin{aligned} \mathbb{P}(A_{x,y}^\diamond) &= C\left(F_X\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right), F_Y\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) + C\left(F_X\left(-\frac{\sqrt{2}}{2}(|x|+|y|)\right), F_Y\left(-\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) \\ &\quad - C\left(F_X\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right), F_Y\left(-\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) - C\left(F_X\left(-\frac{\sqrt{2}}{2}(|x|+|y|)\right), F_Y\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right). \end{aligned}$$

On pose ensuite  $\forall x \in \mathbb{R}, F(x) = F_X(x) = F_Y(x)$  et nous en déduisons :

$$\begin{aligned} \mathbb{P}(A_{x,y}^\diamond) &= C\left(F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right), F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) + C\left(1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right), 1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) \\ &\quad - 2C\left(F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right), 1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right). \end{aligned} \quad (3.79)$$

Enfin, puisque la dépendance considérée est l'anticomonotonie, nous avons

$C(a, b) = \max(0, a + b - 1)$  (copule de Lukasiewicz).

La distribution de possibilité  $\pi_{X,Y}^L(x, y)$  reflétant l'anticomonotonie entre les lois marginales est telle que :

$$\begin{aligned} \pi_{X,Y}^L(x, y) &= 1 - \mathbb{P}(A_{x,y}^\diamond) \\ &= 1 - \max\left(0, F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) + F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) - 1\right) \\ &\quad - \max\left(0, 1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) + 1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) - 1\right) \\ &\quad + 2\max\left(0, 1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) + F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) - 1\right) \\ &= 1 - \max\left(0, 2F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) - 1\right) - \max\left(0, 1 - 2F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right) \end{aligned}$$

Or, on sait que  $\forall x \in \mathbb{R}, F(|x|) \geq \frac{1}{2}$  donc  $2F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) - 1 \geq 0$  et  $1 - 2F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) \leq 0$ .

On en déduit que l'expression de la distribution de possibilité jointe consistante avec une distribution de probabilité jointe dont les marginales sont anticomonotones est :

$$\begin{aligned} \pi_{X,Y}^L(x, y) &= 1 - 2F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right) + 1 \\ &= 2\left(1 - F\left(\frac{\sqrt{2}}{2}(|x|+|y|)\right)\right), \end{aligned} \quad (3.80)$$

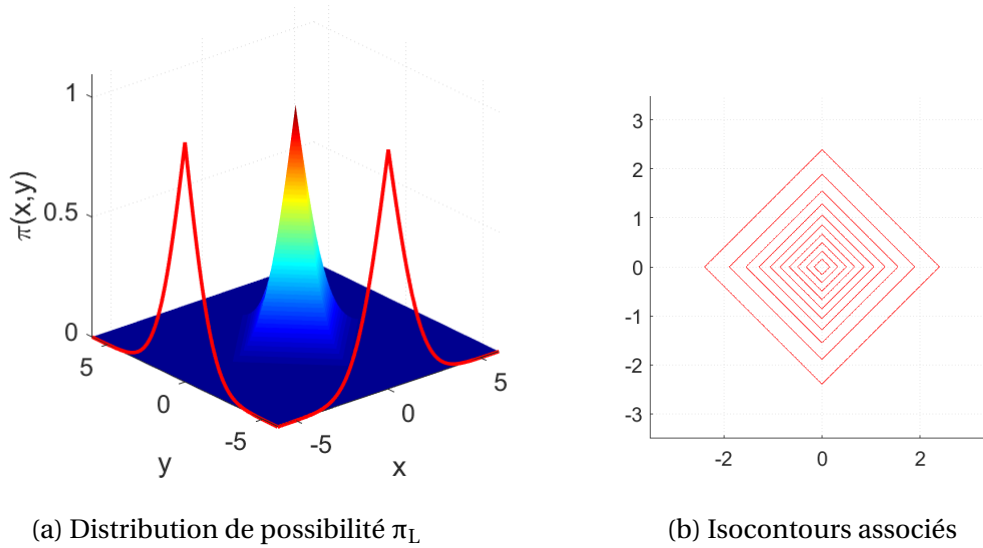


FIGURE 3.18 – Tracé de la distribution de possibilité consistante avec des lois de probabilité marginales anticomotones.

### Copule indépendante $C^\perp$ :

Dans le dernier cas considéré, nous supposons que les lois marginales sont indépendantes entre elles.

Pour construire la distribution de possibilité consistante avec la distribution de probabilité jointe, on peut reproduire ce que nous avons fait précédemment : c'est-à-dire choisir des régions inscrites dans les coupes de la distribution de probabilité jointe en délimitant les dimensions des régions à l'aide d'une distance.

Pour l'exemple, supposons que les coupes de la densité sont des cercles.

Les intervalles délimités par ces cercles sont notés pour tous points  $(x, y)^T$  de l'espace :

$$A_{x,y}^\circ = \{u, v \in \mathbb{R}^2, \sqrt{u^2 + v^2} \leq \sqrt{x^2 + y^2}\}. \quad (3.81)$$

On notera pour la suite  $R_{x,y} = \sqrt{x^2 + y^2}$  le rayon du cercle défini au point  $(x, y)^T$ .

De ce fait, les régions adéquates pour construire une distribution de possibilité consistante sont donc les carrés et les losanges inclus dans ces isocontours circulaires. Cependant, si nous considérons, en plus, les carrés englobant les isocontours hyperboliques nous pouvons déterminer une borne inférieure à la distribution de possibilité équivalente à la densité de probabilité jointe.

Ainsi, nous proposons d'encadrer le résultat que nous recherchons de la manière sui-

vante :

Premièrement, on détermine la distribution de possibilité jointe  $\pi_{X,Y}^{\text{inf}}(x, y)$  dominée par la distribution de possibilité jointe  $\pi_{X,Y}^{\perp}(x, y) = 1 - \mathbb{P}(A_{x,y}^{\circ})$ . Pour cela on calcule  $\pi_{X,Y}^{\text{inf}}(x, y)$  telle que :

$$\pi_{X,Y}^{\text{inf}}(x, y) = 1 - \mathbb{P}(A_{x,y}^{\blacksquare}), \quad (3.82)$$

où l'intervalle  $A_{x,y}^{\blacksquare}$  est délimité par un carré englobant le cercle. C'est-à-dire :

$$A_{x,y}^{\blacksquare} = \{u, v \in \mathbb{R}^2, \max(|u|, |v|) \leq R_{x,y}\}. \quad (3.83)$$

À noter, que de la façon dont est construit  $A_{x,y}^{\blacksquare}$ , on aura  $A_{x,y}^{\circ} \subseteq A_{x,y}^{\blacksquare}$ .

Ce qui implique que  $\mathbb{P}(A_{x,y}^{\circ}) \leq \mathbb{P}(A_{x,y}^{\blacksquare})$ , et donc que :

$$\pi_{X,Y}^{\text{inf}}(x, y) \leq \pi_{X,Y}^{\perp}(x, y). \quad (3.84)$$

On remarquera  $\pi_{X,Y}^{\text{inf}}(x, y)$  n'est en général pas consistante avec la distribution de probabilité jointe.

Pour calculer  $\pi_{X,Y}^{\text{inf}}$ , on commence par remarquer que pour un point  $(x, y)^T$  quelconque, les coordonnées des angles du carré HIJK englobant le cercle de centre  $(0, 0)^T$  et de rayon

$R_{x,y}$  sont (voir figure 3.19) :  $H = \begin{pmatrix} R_{x,y} \\ R_{x,y} \end{pmatrix}$ ,  $I = \begin{pmatrix} R_{x,y} \\ -R_{x,y} \end{pmatrix}$ ,  $J = \begin{pmatrix} -R_{x,y} \\ -R_{x,y} \end{pmatrix}$  et  $K = \begin{pmatrix} -R_{x,y} \\ R_{x,y} \end{pmatrix}$ .

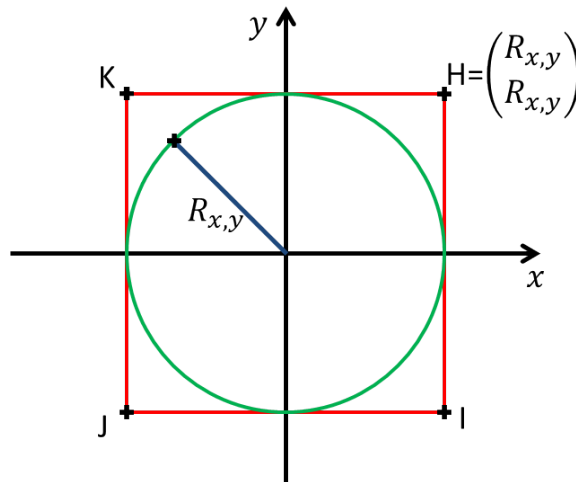


FIGURE 3.19 – Construction de carré HIJK englobant le cercle de centre O, de rayon  $R_{x,y}$  et passant par le point  $(x, y)^t$ .

Nous en déduisons alors (avec  $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$ ) :

$$\begin{aligned}
 \pi_{X,Y}^{\inf}(x, y) &= 1 - \mathbb{P}(A_{x,y}^{\blacksquare}) \\
 &= 1 - F_{X,Y}(R_{x,y}, R_{x,y}) - F_{X,Y}(-R_{x,y}, -R_{x,y}) + F_{X,Y}(-R_{x,y}, R_{x,y}) + F_{X,Y}(R_{x,y}, -R_{x,y}) \\
 &= 1 - C(F_X(R_{x,y}), F_Y(R_{x,y})) - C(F_X(-R_{x,y}), F_Y(-R_{x,y})) \\
 &\quad + C(F_X(-R_{x,y}), F_Y(R_{x,y})) + C(F_X(R_{x,y}), F_Y(-R_{x,y})). \tag{3.85}
 \end{aligned}$$

Puis en posant  $\forall x \in \mathbb{R}, F(x) = F_X(x) = F_Y(x)$  et  $C(a, b) = a \times b$  afin de refléter l'indépendance entre les lois de probabilité marginales, nous obtenons :

$$\begin{aligned}
 \pi_{X,Y}^{\inf}(x, y) &= 1 - C(F(R_{x,y}), F(R_{x,y})) - C(F(-R_{x,y}), F(-R_{x,y})) \\
 &\quad + C(F(-R_{x,y}), F(R_{x,y})) + C(F(R_{x,y}), F(-R_{x,y})) \\
 &= 1 - C(F(R_{x,y}), F(R_{x,y})) - C(1 - F(R_{x,y}), 1 - F(R_{x,y})) \\
 &\quad + C(1 - F(R_{x,y}), F(R_{x,y})) + C(F(R_{x,y}), 1 - F(R_{x,y})) \\
 &= 1 - F^2(R_{x,y}) - 1 + 2F(R_{x,y}) - F^2(R_{x,y}) + 2F(R_{x,y}) - 2F^2(R_{x,y}) \\
 &= 4F(R_{x,y})(1 - F(R_{x,y})). \tag{3.86}
 \end{aligned}$$

Nous en déduisons alors la borne inférieure de  $\pi_{X,Y}^{\perp}(x, y)$  en fonction des distributions de possibilité marginales équivalentes :

$$\pi_{X,Y}^{\inf}(x, y) = \pi_X(x) (2 - \pi_X(x)) \pi_Y(y) (2 - \pi_Y(y)). \tag{3.87}$$

Secondement, on détermine la distribution de possibilité jointe  $\pi_{X,Y}^{\sup}(x, y)$  dominant la distribution de possibilité jointe  $\pi_{X,Y}^{\perp}(x, y)$ .

Pour cela on calcule  $\pi_{X,Y}^{\sup}(x, y)$  telle que :

$$\pi_{X,Y}^{\sup}(x, y) = 1 - \mathbb{P}(A_{x,y}^{\blacklozenge}) \tag{3.88}$$

où l'intervalle  $A_{x,y}^{\blacklozenge}$  est délimité par le losange inscrit dans le cercle de centre O et rayon  $R_{x,y}$  tel que :

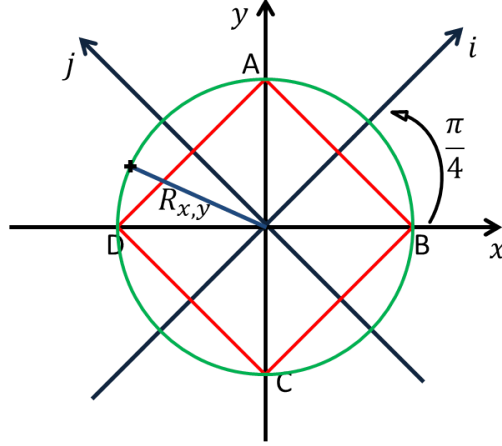
$$A_{x,y}^{\blacklozenge} = \{u, v \in \mathbb{R}^2, (|u| + |v|) \leq R_{x,y}\}. \tag{3.89}$$

Ainsi,  $A_{x,y}^{\blacklozenge}$  est construit tel que :  $A_{x,y}^{\blacklozenge} \subseteq A_{x,y}^{\circ}$ . Cela implique alors que  $\mathbb{P}(A_{x,y}^{\blacklozenge}) \leq \mathbb{P}(A_{x,y}^{\circ})$ .

On aura alors :

$$\pi_{X,Y}^{\sup}(x, y) \geq \pi_{X,Y}^{\perp}(x, y) \tag{3.90}$$

Ensuite, pour calculer  $\mathbb{P}(A_{x,y}^{\blacklozenge})$ , on remarque que pour tout point  $(x, y)^T$  du plan, les sommets d'un losange ABCD inscrit dans le cercle de centre O et de rayon  $R_{x,y}$  ont pour coordonnées (voir figure 3.20) :  $A = \begin{pmatrix} 0 \\ R_{x,y} \end{pmatrix}$ ,  $B = \begin{pmatrix} R_{x,y} \\ 0 \end{pmatrix}$ ,  $C = \begin{pmatrix} 0 \\ -R_{x,y} \end{pmatrix}$  et  $D = \begin{pmatrix} -R_{x,y} \\ 0 \end{pmatrix}$ .


 FIGURE 3.20 – Losange ABCD inscrit dans le cercle de centre O et de rayon  $R_{x,y}$ .

En appliquant la matrice de rotation  $Rot$  utiliser précédemment, on détermine les coordonnées des sommets du losange ABCD dans un nouveau repère (OIJ) telles que :

$$A = \frac{\sqrt{2}}{2} \begin{pmatrix} R_{x,y} \\ R_{x,y} \end{pmatrix}, B = \frac{\sqrt{2}}{2} \begin{pmatrix} R_{x,y} \\ -R_{x,y} \end{pmatrix}, C = -\frac{\sqrt{2}}{2} \begin{pmatrix} R_{x,y} \\ R_{x,y} \end{pmatrix} \text{ et } D = \frac{\sqrt{2}}{2} \begin{pmatrix} -R_{x,y} \\ R_{x,y} \end{pmatrix}.$$

On calcule alors la probabilité accumulée dans un losange droit ABCD comme si c'était un carré dans le repère (OIJ) telle que :

$$\begin{aligned} \mathbb{P}(A_{x,y}^\diamond) &= F_{X,Y} \left( \frac{\sqrt{2}}{2} R_{x,y}, \frac{\sqrt{2}}{2} R_{x,y} \right) + F_{X,Y} \left( -\frac{\sqrt{2}}{2} R_{x,y}, -\frac{\sqrt{2}}{2} R_{x,y} \right) \\ &\quad - F_{X,Y} \left( \frac{\sqrt{2}}{2} R_{x,y}, -\frac{\sqrt{2}}{2} R_{x,y} \right) - F_{X,Y} \left( -\frac{\sqrt{2}}{2} R_{x,y}, \frac{\sqrt{2}}{2} R_{x,y} \right) \\ &= C \left( F_X \left( \frac{\sqrt{2}}{2} R_{x,y} \right), F_Y \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \right) + C \left( 1 - F_X \left( \frac{\sqrt{2}}{2} R_{x,y} \right), 1 - F_Y \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \right) \\ &\quad - C \left( 1 - F_X \left( \frac{\sqrt{2}}{2} R_{x,y} \right), F_Y \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \right) - C \left( F_X \left( \frac{\sqrt{2}}{2} R_{x,y} \right), 1 - F_Y \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \right) \end{aligned} \quad (3.91)$$

Enfin en posant  $\forall x \in \mathbb{R}, F(x) = F_X(x) = F_Y(y)$  et  $C(a, b) = a \times b$ , nous en déduisons la forme finale de  $\pi_{X,Y}^{\text{sup}}(x, y)$  telle que :

$$\pi_{X,Y}^{\text{sup}}(x, y) = 4F \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \left( 1 - F \left( \frac{\sqrt{2}}{2} R_{x,y} \right) \right). \quad (3.92)$$

Ce qui nous donne le résultat en fonction des distributions de possibilité marginales équivalentes :

$$\pi_{X,Y}^{\text{sup}}(x, y) = \pi_X \left( \frac{\sqrt{2}}{2} x \right) \left( 2 - \pi_X \left( \frac{\sqrt{2}}{2} x \right) \right) \pi_Y \left( \frac{\sqrt{2}}{2} y \right) \left( 2 - \pi_Y \left( \frac{\sqrt{2}}{2} y \right) \right). \quad (3.93)$$

Pour résumer cette partie sur la construction d'une distribution de possibilité jointe consistante à partir de la connaissance des lois de probabilité marginales indépendantes entre elles :

- Nous n'avons pas pu exprimer  $\pi_{X,Y}^\perp(x, y)$  en fonction des marginales du fait que les isocontours sont des hyperboles et que nous ne connaissons pas de formule générale permettant d'exprimer directement la probabilité accumulée sur un isocontour circulaire en fonction des marginales.
- Nous avons, cependant, proposé la construction de deux distributions de possibilité jointes  $\pi_{X,Y}^{\text{inf}}(x, y)$  et  $\pi_{X,Y}^{\text{sup}}(x, y)$  telles que :

$$\pi_{X,Y}^{\text{inf}}(x, y) \leq \pi_{X,Y}^\perp(x, y) \leq \pi_{X,Y}^{\text{sup}}(x, y). \quad (3.94)$$

Pour finir, nous illustrons dans la figure 3.21 l'utilisation des deux distributions de possibilité jointe  $\pi_{X,Y}^{\text{inf}}(x, y)$  et  $\pi_{X,Y}^{\text{sup}}(x, y)$ . Dans la figure 3.21a, sont tracées en noir les projections marginales de  $\pi_{X,Y}^t(x, y)$  (donnée en équation (3.51)) obtenues en appliquant la transformation elliptique à une distribution de probabilité gaussienne bivariée  $p_{X,Y}(x, y) = p_X(x)p_Y(y)$  dont les distributions marginales sont indépendantes. Sur cette même figure, sont tracées en rouge les projections marginales de  $\pi_{X,Y}^{\text{sup}}(x, y)$  et en verte celle de  $\pi_{X,Y}^{\text{inf}}(x, y)$ . Nous pouvons d'ailleurs observer que les projections marginales  $\pi_X^{\text{sup}}(x) = \max_y \left( \pi_{X,Y}^{\text{sup}}(x, y) \right)$ ,  $\pi_X^{\text{inf}}(x) = \max_y \left( \pi_{X,Y}^{\text{inf}}(x, y) \right)$  et  $\pi_X^t(x) = \max_y \left( \pi_{X,Y}^t(x, y) \right)$  (respectivement  $\pi_Y^{\text{sup}}(y)$ ,  $\pi_Y^{\text{inf}}(y)$  et  $\pi_Y^t(y)$ ) suivent la relation d'ordre :

$$\pi_X^{\text{inf}}(x) \leq \pi_X^t(x) \leq \pi_X^{\text{sup}}(x). \quad (3.95)$$

Dans la figure 3.21b, sont représentés les isocontours des distributions  $\pi_{X,Y}^{\text{inf}}(x, y)$ ,  $\pi_{X,Y}^{\text{sup}}(x, y)$  et  $\pi_{X,Y}^t(x, y)$  avec le même code couleur.

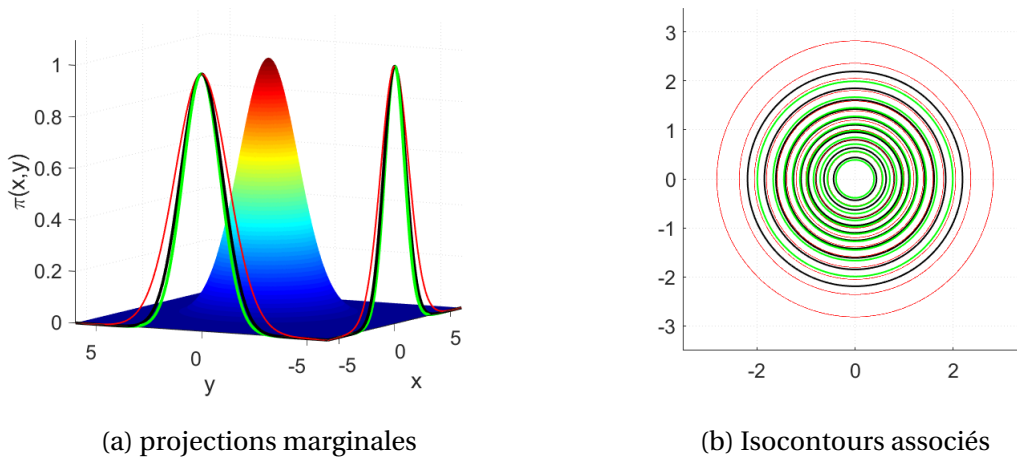


FIGURE 3.21 – Tracer des distributions  $\pi_{X,Y}^{\text{inf}}(x, y)$  (en vert),  $\pi_{X,Y}^t(x, y)$  (en noir) et  $\pi_{X,Y}^{\text{sup}}(x, y)$  (en rouge).

### 3.7 Conclusion

Dans le dernier chapitre de cette thèse, nous avons proposé plusieurs approches possibilistes pour modéliser simultanément plusieurs variables corrélées.

Notre postulat a été de considérer que les données imparfaites présentes dans chacune des composantes d'une image multidimensionnelle peut être vue comme des variables aléatoires liées entre elles. L'idée a alors été de déterminer des outils issus de la théorie des possibilités pour modéliser le plus adéquatement ces données corrélées.

Dans le but d'atteindre cet objectif, deux axes principaux d'études ont alors été présentés.

Le premier axe repose sur une transformation directe des modèles de représentation probabiliste multivariés en distributions de possibilité multivariées. Plus particulièrement, notre intérêt s'est porté sur le cas général des distributions de probabilité elliptiques dont l'utilisation est fréquente dans de nombreux domaines scientifiques.

Une transformation de probabilité-possibilité permettant de construire la distribution de possibilité consistante, la plus spécifique, avec une distribution de probabilité elliptique a été présentée. Cette transformation est une extension de la transformation de probabilité-possibilité continue présentée dans la section 2.3 dans le cas des distributions univariées. Grâce à cette transformation, nous avons notamment exprimé trois distributions de possibilité elliptiques à partir des lois de Cauchy, Student-t et normale multivariée respectivement. Nous avons, de plus, donné dans le cas bivarié, les expressions et l'allure de chacune de ces distributions.

Pour compléter notre étude, nous avons appliqué la modélisation possibiliste elliptique dans un exercice de détection de changements entre images satellites optiques multispectrales. La distribution de possibilité choisie pour cette application a été obtenue par transformation de la densité de probabilité gaussienne bivariée. Nous avons ensuite déterminé les formes analytiques de deux divergences : la divergence de Kullback-Leibler symétrique et la divergence d'Hellinger pour deux distributions de possibilité de cette forme. Puis, nous avons mis en application ce modèle de représentation et cette mesure de divergence en reprenant l'application de détection de changements du chapitre 2 sur une nouvelle base d'images satellites.

La base d'images utilisée est la base OSCD constituée de 14 paires d'images multispectrales "avant-après" ainsi que de la vérité terrain des changements pour chacune des paires. Elle



nous a permis de comparer la modélisation possibiliste des données imparfaites contenues dans ces images par rapport à une modélisation probabiliste sur deux critères : la robustesse au niveau de bruit et la taille de la fenêtre d'échantillonnage. Les résultats de la détection de changements sur les 14 paires d'images ont notamment permis de montrer que la distribution de possibilité bivariée semblait être un modèle de représentation plus adéquat que le modèle probabiliste gaussien équivalent en présence de données bruitées.

La seconde approche proposée dans ce chapitre repose sur la modélisation de chacune des dimensions d'une image séparément (marginales) avant de les assembler à l'aide d'une fonction adaptée pour en dériver une distribution de possibilité jointe représentant l'ensemble des données prises en compte.

En étudiant la littérature scientifique, nous avons remarqué que les outils d'agrégation de la théorie des ensembles flous (les  $t$ -normes et  $t$ -conormes) ne permettent en général pas de construire une distribution de possibilité jointe consistante avec la distribution de probabilité jointe à partir des connaissances seules des lois marginales et de la structure de dépendance.

Afin de remédier à ce problème, nous avons proposé de nouvelles méthodes de construction de distributions de possibilité jointes basées sur un empilement des probabilités accumulées dans les isocontours obtenus à partir des lois de probabilité marginales et d'une copule permettant de modéliser la dépendance.

Le modèle possibiliste ainsi proposé est de ce fait consistant avec la distribution de probabilité jointe lorsque l'on peut exprimer les isocontours de celle-ci en fonction des lois marginales (cas des isocontours carrés et losanges).

Finalement, dans le cas où les isocontours de la densité sont des ellipses ou des cercles, par exemple deux lois de probabilité marginales gaussiennes indépendantes, nous avons proposé un encadrement de la distribution de possibilité consistante et maximum spécifique par deux distributions de possibilité elliptiques  $\pi_{X,Y}^{\text{inf}}(x, y)$  et  $\pi_{X,Y}^{\text{sup}}(x, y)$ .

La borne inférieure de l'encadrement est une distribution de possibilité jointe construite par empilement des probabilités accumulées dans les carrés englobant les isocontours elliptiques. La borne supérieure de l'encadrement est une distribution de possibilité jointe construite par empilement des probabilités accumulées dans les intervalles délimités par les losanges inclus dans les isocontours elliptiques.



## Chapitre 4

### Conclusion & perspectives

Dans ce manuscrit de thèse, nous avons proposé un ensemble de modèles et de méthodes issues de la théorie des possibilités pour modéliser les données de l'imagerie satellitaire.

En remplaçant le modèle de représentation des données que constitue la distribution de probabilité par une distribution de possibilité, nous avons tenté de décrire avec plus de nuance la complexité des données. En effet, comme nous l'avons mentionné dans le premier chapitre, choisir une distribution de possibilité revient à choisir une famille de distributions de probabilité. C'est en jouant sur cet aspect que nous avons essayé donner de la robustesse au modèle de représentation.

Toutefois, la représentation de données imparfaites est un domaine extrêmement vaste. Il n'est pas possible d'en faire le tour en l'espace d'une thèse. Ainsi, l'axe que nous avons choisi de développer dans ce manuscrit est la représentation paramétrique de variables aléatoires puis de vecteurs aléatoires au moyen de distributions de possibilité. La valeur d'un pixel et les valeurs des pixels de son voisinage sont considérées comme des réalisations d'une variable aléatoire (ou d'un vecteur aléatoire) et l'idée est alors de décrire cette variable à l'aide d'un modèle défini par un certain nombre de paramètres.

Dans la théorie des probabilités, on peut décrire de nombreuses classes de variables aléatoires par une distribution de probabilité continue. Les paramètres qui définissent cette distribution pour une variable observée sont les moments de la variable aléatoire que l'on estime à partir des observations à notre disposition.

Or, en supposant que la loi de probabilité est bien la loi sous-jacente de la distribution des données, il se pose tout de même un problème :

peut-on considérer que les observations à notre disposition sont suffisamment nombreuses

ou suffisamment fiables pour estimer avec certitude les moments de la variable aléatoire? Puisque l'on ne peut généralement pas modifier nos données sans altérer les informations qu'elles contiennent, nous avons deux choix : considérer des estimateurs plus robustes ou choisir un modèle de représentation plus nuancé. C'est bien entendu cette seconde option que nous avons choisi d'explorer dans cette thèse avec la modélisation par distribution de possibilité.

Après avoir présenté dans le premier chapitre le contexte dans lequel se déroule cette thèse, l'objet du second chapitre a tourné autour de la modélisation possibiliste univariée appliquée au traitement d'images. Dans un premier temps, nous avons passé en revue les différentes transformations permettant de passer d'une représentation probabiliste à une représentation possibiliste. Nous avons notamment remarqué que ces transformations n'étaient pas toutes équivalentes et qu'il était nécessaire, outre le fait d'imposer certains principes pour garantir la cohérence de l'information, de pouvoir différencier ces transformations. À ce titre, une première application a été proposée afin d'observer le comportement des distributions de possibilité calculées avec chacune des transformations de la littérature par rapport à l'intensité et au type de bruit présent dans les données.

Dans un second temps, nous avons abordé la modélisation à l'aide de distributions de possibilité continues, rentrant ainsi pleinement dans le sujet, c'est-à-dire la représentation possibiliste paramétrique.

Ensuite une mesure de divergence entre deux distributions de ce type a été proposée afin de pouvoir comparer la modélisation possibiliste par rapport à la modélisation probabiliste dans une application de détection de changements entre images.

Puis nous nous sommes intéressés au traitement des attributs extraits à partir de filtre. Ce type d'extraction d'attribut est relativement courant en traitement d'images et produit généralement des distributions de données centrées en 0 et réparties de façon symétrique. Nous avons alors dérivé, pour l'étude de ce type de données, une version possibiliste symétrique de la loi de Weibull en utilisant la transformation continue approfondie par Dubois, Foulloy, Mauris et Prade.

De même que précédemment, nous avons ensuite déterminé une forme analytique d'une divergence de Kullback-Leibler possibiliste afin de pouvoir mettre en pratique cette modélisation dans un contexte de détection de changements entre images.

Dans les deux applications de ce chapitre, la méthode de modélisation basée sur des dis-

tributions de possibilité a permis d'égaliser sinon de surpasser les résultats obtenus avec les méthodes probabilistes équivalentes. Ainsi, nous avons pu mettre en évidence l'intérêt de la modélisation par distributions de possibilité comme méthode de représentation plus adéquate des données imparfaites.

Le troisième chapitre de ce mémoire a été consacré à la modélisation possibiliste multivariée.

Les images satellites sont en réalité rarement des images à une composante, mais généralement constituées de 2 composantes ou plus comme c'est le cas dans l'imagerie optique (RGB, multispectrale, hyperspectrale) mais aussi radar. Il nous a donc paru essentiel de s'attaquer à ce problème sous l'angle de la modélisation possibiliste continue. De plus, à notre connaissance, relativement peu de travaux ont concerné la modélisation possibiliste multivariée des données. Il semblait nécessaire de proposer une alternative aux lois de probabilités multivariées pour gérer ces données dans leur globalité.

La première approche que nous avons présentée est une méthode d'inférence de distributions de possibilité multivariées basée sur une extension de la transformation continue univariée au cas des lois de probabilité elliptique. Cette méthode nous a permis de calculer les formes analytiques de distributions de possibilité elliptiques issues de lois de probabilité couramment utilisées dans l'analyse multidimensionnelle telle que les lois gaussiennes, de Cauchy ou de Student multivariées.

Dans une seconde étape, nous avons déterminé les formes analytiques de deux divergences, la divergence de Kullback-Leibler et la divergence d'Hellinger, pour deux distributions de possibilité bivariées issues de la loi gaussienne.

Nous avons ensuite confronté la représentation de données à l'aide de distribution de possibilité elliptique par rapport à la représentation par modèle probabiliste en utilisant à nouveau une application de détection de changements mais en considérant cette fois une base d'images multispectrales constituée de 14 paires d'images donnant ainsi davantage de poids aux résultats de notre expérimentation. Les résultats de cette application ont démontré l'intérêt de l'utilisation d'outils possibilistes pour modéliser des données imparfaites. En outre, nous avons pu observer que les distributions de possibilité semblaient plus robustes que les distributions de probabilités lorsque du bruit est appliqué volontairement, à des niveaux d'intensité différents, aux images.

La seconde approche que nous avons exposée a pour but de déterminer la fonction d'agrè-

gation à utiliser pour lier deux distributions de possibilité marginales en fonction d'une dépendance voulue afin d'obtenir une distribution de possibilité consistante avec la distribution de probabilité jointe sous-jacente.

Le premier résultat de cette seconde partie a été de montrer que dans le cas où les distributions marginales sont comonotones, la fonction d'agrégation permettant de construire une distribution de possibilité consistante est égale à la fonction d'agrégation utilisée pour lier deux fonctions de répartition comonotones, c'est-à-dire la copule minimum.

Le second résultat, que nous avons exposé, est la forme de la fonction d'agrégation lorsque les distributions marginales sont anticomonotones.

Enfin, le dernier point abordé dans cette partie est la proposition d'une méthode permettant d'approcher la distribution de possibilité elliptique issue de deux distributions de possibilité marginales indépendantes. Le résultat consiste en un encadrement par une distribution dominant la distribution elliptique et une distribution dominée par la distribution désirée.

### **Perspectives :**

Le travail présenté dans ce manuscrit a permis de soulever plusieurs points à améliorer. Ainsi, dans ce qui suit, nous allons tenter d'exposer certaines perspectives sur lesquelles il nous semble possible d'apporter des contributions.

La première perspective sur laquelle il nous paraît intéressant de travailler se place dans le cadre de l'analyse univariée de données.

Dans la théorie des probabilités, il est possible de construire des distributions de mélanges de lois de probabilité unimodales dans le but de définir un modèle plus proche des données. On peut notamment en voir un exemple appliqué au traitement d'images dans un article [131]. Ainsi, dans la continuité des travaux présentés par Gilles Mauris [132] sur la transformation de probabilité- possibilité de distributions de probabilité bimodale, il serait pertinent d'étudier l'impact de la modélisation possibiliste multimodale univariée appliquée au traitement d'images satellites. Pour ce faire, il serait, de plus, nécessaire de pouvoir définir une mesure de dissemblance entre distributions de possibilité de ce type.

Un second point d'amélioration, toujours dans l'analyse univariée de données dans un premier temps mais qu'il serait intéressant d'étendre à l'analyse multivariée plus tard,

porterait sur la construction de distributions de possibilité à queue lourde (*heavy tail distribution* en anglais). Ces distributions ont suscité un grand intérêt dans la théorie des probabilités car elles permettent de modéliser les dépendances aux valeurs extrêmes. Cependant, ceci n'a été que légèrement abordé au cours de cette thèse dans le chapitre 2 et pourrait faire l'objet de travaux futurs d'une grande utilité pour définir un panel d'outils de modélisation possibilistes plus spécialisés suivant le type de données exploitées.

Dans le cadre de la modélisation possibiliste multivariée, une perspective immédiate à laquelle il nous semble prioritaire d'accorder de l'attention est l'extension des divergences de Kullback-Leibler symétrisée et d'Hellinger aux distributions elliptiques de dimensions  $n \geq 2$ . En effet, dans le chapitre 3 de cette thèse, seul le cas des distributions de possibilité elliptiques bivariées a été traité limitant ainsi la prise en compte globale des données comme c'est le cas pour les images multispectrales à 13 composantes de la base d'images OSCD.

A plus long terme et toujours dans le cadre de l'analyse multivariée, nous aimerions étendre la transformation de probabilité-possibilité des lois elliptiques à l'ensemble des lois de probabilités multivariées. Cela permettrait, ainsi, de déterminer des distributions de possibilité multivariées asymétriques.

Nous avons signalé à plusieurs reprises dans ce manuscrit que les modélisations probabilistes et possibilistes ne sont pas antagonistes l'une de l'autre mais au contraire complémentaires. Nous avons en effet remarqué au cours de nos expérimentations que les deux modèles ne captureraient pas exactement la même facette de l'information. Dans une perspective plus globale d'amélioration, nous aimerions pouvoir entreprendre des travaux sur la fusion des deux méthodes à l'aide d'outils issus de l'agrégation multicritère tels que par exemple les intégrales de Choquet ou de Sugeno.

À titre informatif, les premiers tests effectués semblent montrer un réel gain de performances en combinant les deux méthodes dans une application de détection de changements.

Le dernier point que nous soulignerons, dans cette partie *perspectives*, concerne la modélisation de données temporelles à l'aide d'outils possibilistes.

En effet, dans ce manuscrit, nous avons présenté d'abord des méthodes de modélisation

de variables aléatoires puis de vecteurs aléatoires, donc la suite logique serait de s'intéresser à la modélisation de processus aléatoire. Ainsi, au cours des derniers mois de cette thèse, nous avons entamé des travaux en vue d'une application d'extraction d'informations dans une série temporelle d'images satellites. La série temporelle d'images est constituée d'une vingtaine d'images radar (recalées) de la région du pont de l'Oyapock à la frontière entre la Guyane française et le Brésil. Les processus aléatoires que nous voulons modéliser sont constitués des profils d'évolution des pixels (que l'on considère comme des variables aléatoires) dans la série temporelle (voir figure 4.1). L'objectif final de cette application est de pouvoir classer les différents pixels d'une image de la série temporelle dans différentes catégories : zone ayant subi une déforestation, zone agricole, zone inondable, forêts, etc.

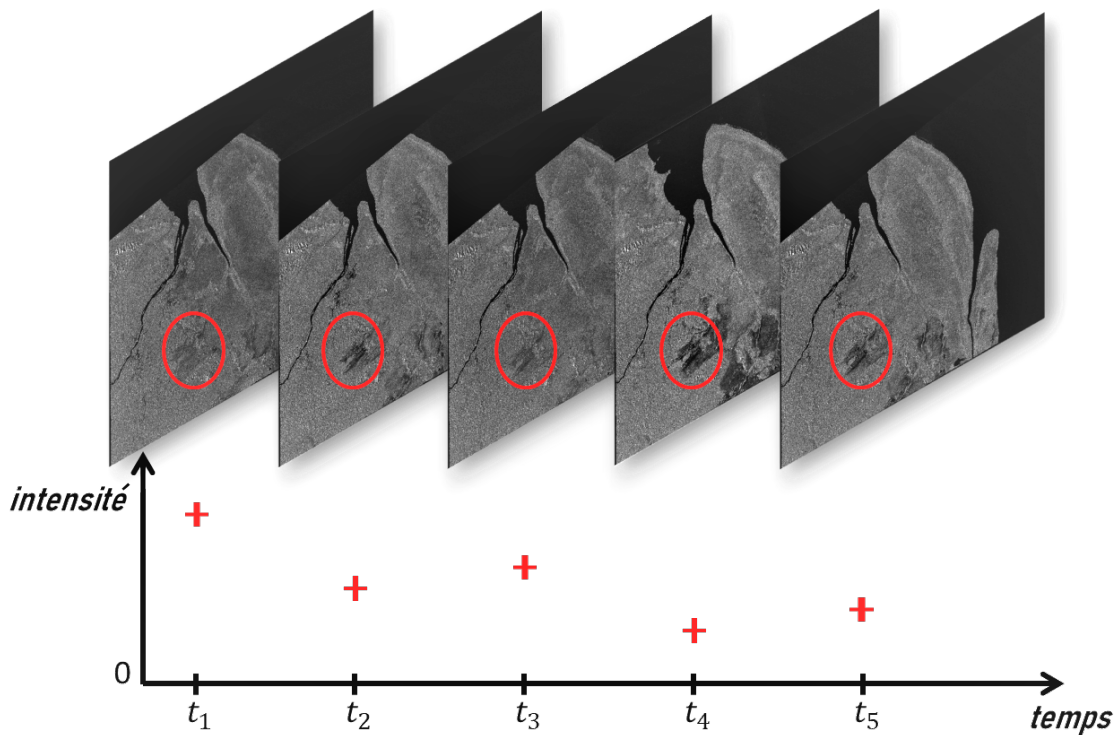


FIGURE 4.1 – Exemple de profil temporel d'un pixel  $a_{i,j}$  pris dans une série temporelle d'images satellites.

Pour répondre à cette problématique, notre approche préliminaire s'est inspirée de la modélisation par *processus gaussien*. Les processus gaussiens sont une méthode statistique utilisant le machine learning pour représenter une collection de variables aléatoires



à l'aide d'une distribution normale multivariée. L'ensemble des  $n$  points d'un processus stochastique est assimilé à un point de dimension  $n$  distribué selon une loi gaussienne multivariée. L'enjeu est alors de déterminer la distribution de probabilité représentant au mieux les données observées. Cela est réalisé en optimisant les paramètres de la fonction de covariance (fonction noyau) à l'aide d'une fonction de vraisemblance par rapport aux données (plus de détails dans [133]).

Les perspectives de nos travaux dans la modélisation des processus stochastiques sont diverses. Par rapport au processus gaussien, nous aimerions remplacer le modèle probabiliste de loi normale multivariée par la distribution de possibilité gaussienne multivariée (de l'équation (3.21)). Nous pensons qu'un tel changement pourrait permettre d'apporter de la robustesse au modèle de représentation en améliorant notamment la spécification de l'intervalle de confiance associé à chaque donnée.

---

# Publications

## Conférences nationale :

- "Comparaison de transformations de probabilité-possibilité aux moyens de mesures de dissemblance dans un contexte de détection de changements", C. Lesniewska-choquet, G. Mauris, A. Atto, 2017, conférence LFA : Rencontres francophones sur la Logique Floue et ses Applications, Amien.
- "Transformations possibiliste de lois de probabilité multi-variées elliptiques", C. Lesniewska-choquet, G. Mauris, A. Atto, 2019, conférence LFA : Rencontres francophones sur la Logique Floue et ses Applications, Alès.

## Conférence internationale :

- "Image change detection by possibility distribution dissemblance", C. Lesniewska-choquet, G. Mauris, A. Atto, 2017, conférence FUZZIEEE, Naples.

## Revue scientifique :

- "On elliptical possibility distributions", C. Lesniewska-choquet, G. Mauris, A. Atto, 2019, IEEE Transactions on Fuzzy Systems.

# Bibliographie

- [1] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, “Esa’s sentinel missions in support of earth system science,” *Remote Sensing of Environment*, vol. 120, pp. 84–90, 2012. [1](#)
- [2] L. A. Zadeh, “Fuzzy sets,” *Information and control*, vol. 8, no. 3, pp. 338–353, 1965. [2](#), [26](#)
- [3] ———, “Fuzzy sets as a basis for a theory of possibility,” *Fuzzy sets and systems*, vol. 1, no. 1, pp. 3–28, 1978. [2](#), [28](#), [51](#)
- [4] R. E. Bellman and L. A. Zadeh, “Decision-making in a fuzzy environment,” *Management science*, vol. 17, no. 4, pp. B–141, 1970. [2](#)
- [5] M. Sugeno and T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling,” *IEEE Transactions on fuzzy systems*, vol. 1, no. 1, p. 7, 1993. [2](#)
- [6] S. Salicone and M. Prioli, *Measuring uncertainty within the theory of evidence*. Springer, 2018. [2](#), [127](#), [128](#)
- [7] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, “Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities,” *Reliable computing*, vol. 10, no. 4, pp. 273–297, 2004. [4](#), [50](#), [52](#), [55](#), [72](#), [79](#), [80](#), [102](#), [104](#)
- [8] C. O. Justice, J. Townshend, B. Holben, and e. C. Tucker, “Analysis of the phenology of global vegetation using meteorological satellite data,” *International Journal of Remote Sensing*, vol. 6, no. 8, pp. 1271–1318, 1985. [8](#)
- [9] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, “Using the satellite-derived ndvi to assess ecological responses to environmental change,” *Trends in ecology & evolution*, vol. 20, no. 9, pp. 503–510, 2005. [8](#)
- [10] S. N. Magonov and M.-H. Whangbo, *Surface analysis with STM and AFM : experimental and theoretical aspects of image analysis*. John Wiley & Sons, 2008. [11](#)

- 
- [11] J. A. Richards and J. Richards, *Remote sensing digital image analysis*. Springer, 1999, vol. 3. 12
- [12] P. V. Harman, "Image processing method and apparatus," Dec. 17 2002, uS Patent 6,496,598. 13
- [13] R. Hummel, "Image enhancement by histogram transformation," *Computer graphics and image processing*, vol. 6, pp. 184–185, 1975. 13
- [14] F. Berry, "Technologies des capteurs d'images couleurs," *Ecole d'hiver sur l'image numérique couleur EHINC*, 2005. 15
- [15] C.-I. Chang, *Hyperspectral data exploitation : theory and applications*. John Wiley & Sons, 2007. 15
- [16] E. Claude and C. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng*, vol. 371, 1934. 17
- [17] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort *et al.*, "Sentinel-2 : Esa's optical high-resolution mission for gmes operational services," *Remote sensing of Environment*, vol. 120, pp. 25–36, 2012. 17
- [18] R. Bamler, "Principles of synthetic aperture radar," *Surveys in Geophysics*, vol. 21, no. 2-3, pp. 147–157, 2000. 18
- [19] F. Tupin and J. Nicolas, "Les bases de l'imagerie radar à synthèse d'ouverture," *Document de cours, Télécom ParisTech*, 2010. 18
- [20] D. R. Wehner, "High resolution radar," *Norwood, MA, Artech House, Inc., 1987, 484 p.*, 1987. 20
- [21] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in sar interferometry," *IEEE Transactions on geoscience and remote sensing*, vol. 39, no. 1, pp. 8–20, 2001. 20
- [22] S. R. Cloude and K. P. Papathanassiou, "Polarimetric sar interferometry," *IEEE Transactions on geoscience and remote sensing*, vol. 36, no. 5, pp. 1551–1565, 1998. 20
- [23] F. W. Leberl, "Radargrammetric image processing," *Artech House*, 1990. [Online]. Available : <http://infoscience.epfl.ch/record/51729> 20
- [24] J. C. Curlander and R. N. McDonough, "Synthetic aperture radar- systems and signal processing(book)," *New York : John Wiley & Sons, Inc, 1991., 1991.* 20

- [25] P. Smets, “Imperfect information : Imprecision and uncertainty,” in *Uncertainty management in information systems*. Springer, 1997, pp. 225–254. [22](#)
- [26] L. A. Zadeh, “Discussion : Probability theory and fuzzy logic are complementary rather than competitive,” *Technometrics*, vol. 37, no. 3, pp. 271–276, 1995. [22](#), [23](#)
- [27] M.-H. Masson, “Apports de la théorie des possibilités et des fonctions de croyance à l’analyse de données imprécises,” *Habilitation à diriger des Recherches soutenue*, vol. 2, 2005. [23](#), [29](#)
- [28] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42. [23](#)
- [29] A. P. Dempster, “A generalization of bayesian inference,” *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 30, no. 2, pp. 205–232, 1968. [23](#)
- [30] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. [24](#)
- [31] B. De Finetti, “La logique de la probabilité,” in *Actes du congrès international de philosophie scientifique (1935)*, vol. 4. Hermann Paris, 1936, pp. 1–9. [24](#)
- [32] R. Durrett, *Probability : theory and examples*. Cambridge university press, 2019, vol. 49. [24](#)
- [33] C. E. Shannon, “Prediction and entropy of printed english,” *Bell system technical journal*, vol. 30, no. 1, pp. 50–64, 1951. [26](#)
- [34] D. Dubois and H. Prade, “Ranking fuzzy numbers in the setting of possibility theory,” *Information sciences*, vol. 30, no. 3, pp. 183–224, 1983. [28](#)
- [35] —, “Properties of measures of information in evidence and possibility theories,” *Fuzzy sets and systems*, vol. 24, no. 2, pp. 161–182, 1987. [28](#)
- [36] D. Dubois, J. Lang, and H. Prade, “Theorem proving under uncertainty : a possibility theory-based approach,” in *Proceedings of the 10th international joint conference on Artificial intelligence-Volume 2*. Morgan Kaufmann Publishers Inc., 1987, pp. 984–986. [28](#)
- [37] D. Dubois, H. Prade, and S. Sandri, “On possibility/probability transformations,” in *Fuzzy logic*. Springer, 1993, pp. 103–112. [30](#), [50](#), [53](#), [55](#)
- [38] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel, “Towards general measures of comparison of objects,” *Fuzzy sets and systems*, vol. 84, no. 2, pp. 143–153, 1996. [30](#)

- [39] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, pp. 832–837, 1956. [31](#)
- [40] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969. [31](#)
- [41] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust regression methods for computer vision : A review," *International journal of computer vision*, vol. 6, no. 1, pp. 59–70, 1991. [32](#)
- [42] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics*. Springer, 1992, pp. 492–518. [32](#)
- [43] P. R. Halmos, *Measure theory*. Springer, 2013, vol. 18. [32](#)
- [44] D. Revuz, "Mesure et intégration," 1994. [33](#)
- [45] S. M. Ross, *Initiation aux probabilités*. PPUR presses polytechniques, 2007. [33](#)
- [46] C. Dellacherie and P.-A. Meyer, *Probabilities and potential, c : potential theory for discrete and continuous semigroups*. Elsevier, 2011, vol. 151. [33](#)
- [47] J.-Y. Ouvrard, *Probabilités : CAPES-Agrégation*. Cassini, 1998. [33](#)
- [48] G. Saporta, *Probabilités, analyse des données et statistique*. Editions Technip, 2006. [35](#)
- [49] J. Bernoulli, *Ars conjectandi*. Impensis Thurnisiorum, fratrum, 1713. [36](#)
- [50] I. Csiszár, "Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten," *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, vol. 8, pp. 85–108, 1964. [42](#)
- [51] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951. [42](#)
- [52] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen." *Journal für die reine und angewandte Mathematik*, vol. 136, pp. 210–271, 1909. [42](#)
- [53] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967. [42](#)

- [54] I. Csiszár, “I-divergence geometry of probability distributions and minimization problems,” *The Annals of Probability*, pp. 146–158, 1975. [43](#)
- [55] J. Jacod and P. Protter, *L'essentiel en théorie des probabilités*. Cassini, 2003. [43](#)
- [56] H. Jeffreys, *The theory of probability*. OUP Oxford, 1998. [44](#)
- [57] S. Sandri, “La combinaison de l’information incertaine et ses aspects algorithmiques,” Ph.D. dissertation, Toulouse 3, 1991. [50](#), [55](#)
- [58] D. Dubois and H. Prade, “When upper probabilities are possibility measures,” *Fuzzy sets and systems*, vol. 49, no. 1, pp. 65–74, 1992. [51](#), [79](#)
- [59] M. Oussalah, “On the probability/possibility transformations : a comparative analysis,” *International Journal Of General System*, vol. 29, no. 5, pp. 671–718, 2000. [51](#)
- [60] D. Dubois and H. Prade, “Unfair coins and necessity measures : towards a possibilistic interpretation of histograms,” *Fuzzy sets and systems*, vol. 10, no. 1-3, pp. 15–20, 1983. [51](#), [55](#)
- [61] J. F. Geer and G. J. Klir, “A mathematical analysis of information-preserving transformations between probabilistic and possibilistic formulations of uncertainty,” *International Journal of General System*, vol. 20, no. 2, pp. 143–176, 1992. [52](#)
- [62] G. J. Klir, “A principle of uncertainty and information invariance,” *International Journal Of General System*, vol. 17, no. 2-3, pp. 249–275, 1990. [52](#)
- [63] G. J. Klir and B. Parviz, “Probability-possibility transformations : a comparison,” *International Journal of General System*, vol. 21, no. 3, pp. 291–310, 1992. [52](#), [54](#)
- [64] D. Dubois and H. Prade, “A note on measures of specificity for fuzzy sets,” *International Journal of General System*, vol. 10, no. 4, pp. 279–283, 1985. [53](#)
- [65] R. R. Yager, “On the specificity of a possibility distribution,” *Fuzzy Sets and Systems*, vol. 50, no. 3, pp. 279–292, 1992. [53](#)
- [66] V. Torra, “Entropy for non-additive measures in continuous domains,” *Fuzzy Sets and Systems*, vol. 324, pp. 49–59, 2017. [53](#), [82](#)
- [67] T. Sudkamp, “On probability-possibility transformations,” *Fuzzy Sets and Systems*, vol. 51, no. 1, pp. 73–81, 1992. [53](#)
- [68] K. Yamada, “Probability-possibility transformation based on evidence theory,” in *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, vol. 1. IEEE, 2001, pp. 70–75. [53](#), [55](#)

- [69] D. Dubois, "On several representations of uncertain body of evidence," *Fuzzy information and decision processes*, pp. 167–181, 1982. [54](#)
- [70] D. Dubois and H. Prade, *Fuzzy sets and systems : theory and applications*. Academic press, 1980, vol. 144. [54](#), [60](#)
- [71] M. S. Mouchaweh, M. S. Bouguelid, P. Billaudel, and B. RiERA, "Variable probability-possibility transformation," *International Journal of Approximate Reasoning*, 2006. [54](#)
- [72] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 447–458, 1990. [57](#)
- [73] S. Roychowdhury and W. Pedrycz, "A survey of defuzzification strategies," *International Journal of intelligent systems*, vol. 16, no. 6, pp. 679–695, 2001. [57](#)
- [74] I. Jenhani, N. B. Amor, Z. Elouedi, S. Benferhat, and K. Mellouli, "Information affinity : A new similarity measure for possibilistic uncertain information," in *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer, 2007, pp. 840–852. [59](#)
- [75] W.-J. Wang, "New similarity measures on fuzzy sets and on elements," *Fuzzy sets and systems*, vol. 85, no. 3, pp. 305–309, 1997. [59](#), [60](#)
- [76] E. Szmids and J. Kacprzyk, "Distances between intuitionistic fuzzy sets," *Fuzzy sets and systems*, vol. 114, no. 3, pp. 505–518, 2000. [59](#)
- [77] S.-M. Chen, M.-S. Yeh, and P.-Y. Hsiao, "A comparison of similarity measures of fuzzy values," *Fuzzy sets and systems*, vol. 72, no. 1, pp. 79–89, 1995. [59](#)
- [78] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity measures of interval type-2 fuzzy sets to general type-2 fuzzy sets," in *2013 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. IEEE, 2013, pp. 1–8. [59](#)
- [79] H. Lee-Kwang, Y.-S. Song, and K.-M. Lee, "Similarity measure between fuzzy sets and between elements," *Fuzzy Sets and Systems*, vol. 62, no. 3, pp. 291–293, 1994. [59](#)
- [80] L. Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy sets and systems*, vol. 52, no. 3, pp. 305–318, 1992. [60](#)
- [81] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms : a systematic survey," *IEEE transactions on image processing*, vol. 14, no. 3, pp. 294–307, 2005. [61](#)



- [82] F. Yamazaki, Y. Yano, and M. Matsuoka, "Visual damage interpretation of buildings in bam city using quickbird images following the 2003 bam, iran, earthquake," *Earthquake Spectra*, vol. 21, no. S1, pp. 329–336, 2005. [61](#)
- [83] S. Nghiem, K. Steffen, R. Kwok, and W. Tsai, "Detection of snowmelt regions on the greenland ice sheet using diurnal backscatter change," *Journal of Glaciology*, vol. 47, no. 159, pp. 539–547, 2001. [61](#)
- [84] L. S. Guild, W. B. Cohen, and J. B. Kauffman, "Detection of deforestation and land conversion in rondonia, brazil using change detection techniques," *International Journal of Remote Sensing*, vol. 25, no. 4, pp. 731–750, 2004. [61](#)
- [85] Z. Zhang, "Le probleme de la mise en correspondance : l'etat de l'art," 1993. [61](#)
- [86] K. Sakurada and T. Okatani, "Change detection from a street image pair using cnn features and superpixel segmentation." in *BMVC*, 2015, pp. 61–1. [61](#)
- [87] D. M. Powers, "Evaluation : from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011. [65](#)
- [88] C. Lesniewska-Choquet, G. Mauris, and A. Atto, "Comparison of probability-possibility transformations by dissimilarity measures in a context of change detection." LFA, Rencontres francophones sur la Logique Floue et ses Applications, 2017. [71](#)
- [89] G. Mauris, "Possibility distributions : A unified representation of usual direct-probability-based parameter estimation methods," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1232–1242, 2011. [73](#)
- [90] G. Mauris, V. Lasserre, and L. Foulloy, "A fuzzy approach for the expression of uncertainty in measurement," *Measurement*, vol. 29, no. 3, pp. 165–177, 2001. [73](#), [80](#)
- [91] G. Mauris, "A possibilistic view of binomial parameter estimation," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2014, pp. 396–405. [75](#)
- [92] R. Kwitt and A. Uhl, "Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 933–936. [76](#)
- [93] M. D. Buhmann, *Radial basis functions : theory and implementations*. Cambridge university press, 2003, vol. 12. [79](#)

- [94] G. P. Mauris, "Fuzzy representation of incomplete knowledge about infinite support probability distributions in a measurement context," in *2007 IEEE International Fuzzy Systems Conference*. IEEE, 2007, pp. 1–6. [79](#), [81](#)
- [95] G. Mauris, "The principle of possibility maximum specificity as a basis for measurement uncertainty expression," in *2009 IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*. IEEE, 2009, pp. 5–9. [79](#)
- [96] W. Hoeffding, "Probability inequalities for sums of bounded random variables," in *The Collected Works of Wassily Hoeffding*. Springer, 1994, pp. 409–426. [80](#)
- [97] I. R. Savage, "Probability inequalities of the tchebycheff type," *Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics B*, vol. 65, no. 3, pp. 211–222, 1961. [81](#)
- [98] V. Torra, Y. Narukawa, and M. Sugeno, "On the f-divergence for non-additive measures," *Fuzzy Sets and Systems*, vol. 292, pp. 364–379, 2016. [82](#), [83](#)
- [99] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure," *Fuzzy sets and Systems*, vol. 29, no. 2, pp. 201–227, 1989. [82](#)
- [100] G. Mauris, "A review of relationships between possibility and probability representations of uncertainty in measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 3, pp. 622–632, 2012. [83](#)
- [101] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized laplacian of gaussian filter for blob detection and its applications," *IEEE transactions on cybernetics*, vol. 43, no. 6, pp. 1719–1733, 2013. [87](#)
- [102] C. Lesniewska-Choquet, A. M. Atto, G. Mauris, and G. Mercier, "Image change detection by possibility distribution dissemblance," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2017, pp. 1–6. [89](#)
- [103] J. M. Prats-Montalbán, A. De Juan, and A. Ferrer, "Multivariate image analysis : a review with applications," *Chemometrics and intelligent laboratory systems*, vol. 107, no. 1, pp. 1–23, 2011. [95](#)
- [104] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data : fusion of landsat tm and sar images," *IEEE transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 768–778, 1994. [95](#)

- [105] D. Manolakis, D. Marden, G. A. Shaw *et al.*, “Hyperspectral image processing for automatic target detection applications,” *Lincoln laboratory journal*, vol. 14, no. 1, pp. 79–116, 2003. [95](#)
- [106] J. N. Garcia, Z. Kutalik, K.-H. Cho, and O. Wolkenhauer, “Level sets and minimum volume sets of probability density functions,” *International journal of approximate reasoning*, vol. 34, no. 1, pp. 25–47, 2003. [95](#)
- [107] G. Frahm, “Generalized elliptical distributions : theory and applications,” Ph.D. dissertation, Universität zu Köln, 2004. [95](#)
- [108] S. Bhattacharyya and P. J. Bickel, “Adaptive estimation in elliptical distributions with extensions to high dimensions,” 2013. [98](#)
- [109] R. Blanc, E. Syrkina, and G. Székely, “Estimating the confidence of statistical model based shape prediction,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2009, pp. 602–613. [99](#)
- [110] E. W. Weisstein, “Hypersphere,” 2002. [101](#)
- [111] C. Lesniewska-choquet, G. Mauris, A. M. Atto, and G. Mercier, “On elliptical possibility distributions,” *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2019. [101](#), [103](#), [112](#)
- [112] M. D. Perlman, “Concentration inequalities for multivariate distributions : ii. elliptically contoured distributions,” *Lecture Notes-Monograph Series*, pp. 284–308, 1992. [102](#)
- [113] G. Gallego, C. Cuevas, R. Mohedano, and N. Garcia, “On the mahalanobis distance classification criterion for multidimensional normal distributions,” *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4387–4396, 2013. [103](#)
- [114] N. A. Campbell, “Robust procedures in multivariate analysis i : Robust covariance estimation,” *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, vol. 29, no. 3, pp. 231–237, 1980. [106](#)
- [115] J. P. Hoffbeck and D. A. Landgrebe, “Covariance matrix estimation and classification with limited training data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763–767, 1996. [106](#)
- [116] M. Higashi and G. J. Klir, “Measures of uncertainty and information based on possibility distributions,” *International Journal of General Systems*, vol. 9, no. 1, pp. 43–58, 1982. [107](#)

- [117] L. Pardo, *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2005. [107](#)
- [118] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2115–2118. [108](#)
- [119] A. W. Marshall and I. Olkin, “Multivariate chebyshev inequalities,” *The Annals of Mathematical Statistics*, pp. 1001–1014, 1960. [113](#)
- [120] H. Akrouf, J.-F. Crouzet, and O. Strauss, “Multidimensional possibility/probability domination for extending maxitive kernel based signal processing,” *Fuzzy Sets and Systems*, vol. 367, pp. 1–18, 2019. [114](#), [115](#)
- [121] A. Sklar, “Fonctions de répartition à n dimension et leurs marges,” *Université Paris*, vol. 8, pp. 229–231, 1959. [117](#), [121](#)
- [122] W. Hoeffding, “Masstabinvariante korrelationstheorie,” *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, vol. 5, pp. 181–233, 1940. [117](#)
- [123] A. Charpentier, “Copules et risques multiples,” *Approches statistiques du risque, Société Française de Statistiques, Editions TECHNIP, Paris*, 2014. [119](#)
- [124] C. Savu and M. Tiede, “Hierarchies of archimedean copulas,” *Quantitative Finance*, vol. 10, no. 3, pp. 295–304, 2010. [122](#)
- [125] C. Genest and R. J. MacKay, “Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données,” *Canadian Journal of Statistics*, vol. 14, no. 2, pp. 145–159, 1986. [122](#)
- [126] H. Joe, “Parametric families of multivariate distributions with given margins,” *Journal of multivariate analysis*, vol. 46, no. 2, pp. 262–282, 1993. [122](#)
- [127] A. Ferrero, M. Prioli, S. Salicone, and B. Vantaggi, “A 2-d metrology-sound probability–possibility transformation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 982–990, 2013. [122](#), [130](#)
- [128] R. R. Yager, “On ordered weighted averaging aggregation operators in multicriteria decisionmaking,” *IEEE Transactions on systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988. [122](#)

- [129] E. P. Klement, R. Mesiar, and E. Pap, *Triangular norms*. Springer Science & Business Media, 2013, vol. 8. [123](#), [124](#)
- [130] C. Baudrit, D. Dubois, and D. Guyonnet, “Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment,” *IEEE transactions on fuzzy systems*, vol. 14, no. 5, pp. 593–608, 2006. [135](#)
- [131] S. Varadarajan, P. Miller, and H. Zhou, “Spatial mixture of gaussians for dynamic background modelling,” in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2013, pp. 63–68. [150](#)
- [132] G. Mauris, “Transformation of bimodal probability distributions into possibility distributions,” *IEEE Transactions on Instrumentation and measurement*, vol. 59, no. 1, pp. 39–47, 2009. [150](#)
- [133] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3. [153](#)
- [134] L. Pardo, *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2018. [VI](#)



# Annexe A

## Annexes

### A.1 Calcul de la divergence de Kullback-Leibler pour des distributions de la forme de $\pi_{\mathcal{N}}$

Cette section détaille le calcul nécessaire pour déterminer l'équation (2.53) dans le Chapitre 2 section 2.3.3.

Considérons des distributions de possibilité telle que :

$$\pi_{\mathcal{N}_{N_i}}(x) = \exp\left(-\left(\frac{|x - \mu_i|}{\sigma_i \sqrt{2}}\right)^2\right), \quad (\text{A.1})$$

et la divergence de Kullback-Leibler possibiliste non-symétrique de la forme suivante :

$$\text{KL}(\pi_{\mathcal{N}_{N_1}} \parallel \pi_{\mathcal{N}_{N_2}}) = \int_{\mathbb{R}} \pi_{\mathcal{N}_{N_1}}(x) \log\left(\frac{\pi_{\mathcal{N}_{N_1}}(x)}{\pi_{\mathcal{N}_{N_2}}(x)}\right) dx. \quad (\text{A.2})$$

Alors

$$\begin{aligned} \text{KL}(\pi_{\mathcal{N}_{N_1}} \parallel \pi_{\mathcal{N}_{N_2}}) &= \int_{\mathbb{R}} \pi_{\mathcal{N}_{N_1}}(x) \log\left(\frac{\pi_{\mathcal{N}_{N_1}}(x)}{\pi_{\mathcal{N}_{N_2}}(x)}\right) dx \\ &= \int_{\mathbb{R}} \pi_{\mathcal{N}_{N_1}}(x) \log(\pi_{\mathcal{N}_{N_1}}(x)) dx - \int_{\mathbb{R}} \pi_{\mathcal{N}_{N_1}}(x) \log(\pi_{\mathcal{N}_{N_2}}(x)) dx \\ &= \int_{\mathbb{R}} \frac{1}{2} \left(\frac{x - \mu_2}{\sigma_2}\right)^2 \pi_{\mathcal{N}_{N_1}}(x) dx \\ &\quad - \int_{\mathbb{R}} \frac{1}{2} \left(\frac{x - \mu_1}{\sigma_1}\right)^2 \pi_{\mathcal{N}_{N_1}}(x) dx. \end{aligned}$$

En remarquant que  $\pi_{\mathcal{N}_{N_1}}(x) = \sigma_1 \sqrt{2\pi} p_{\mathcal{N}_1}(x)$ , on peut alors écrire que :

$$\begin{aligned} \text{KL}(\pi_{\mathcal{N}_{N_1}} \parallel \pi_{\mathcal{N}_{N_2}}) &= \sigma_1 \sqrt{2\pi} \left( \int_{\mathbb{R}} \frac{1}{2} \left(\frac{x - \mu_2}{\sigma_2}\right)^2 p_{\mathcal{N}_1}(x) dx \right. \\ &\quad \left. - \int_{\mathbb{R}} \frac{1}{2} \left(\frac{x - \mu_1}{\sigma_1}\right)^2 p_{\mathcal{N}_1}(x) dx \right). \end{aligned} \quad (\text{A.3})$$

Notons  $\alpha$  and  $\beta$  les deux intégrales de l'équation précédente tel que :

$$\begin{aligned} - \alpha &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{x - \mu_2}{\sigma_2} \right)^2 p_{\mathcal{N}_1}(x) dx \\ - \beta &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{x - \mu_1}{\sigma_1} \right)^2 p_{\mathcal{N}_1} dx. \end{aligned}$$

Soit  $\mathbb{E}(X_1)$  l'espérance de la variable aléatoire  $X_1 \in \mathbb{R}^2$ , nous avons alors :

$$\begin{aligned} \beta &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{x - \mu_1}{\sigma_1} \right)^2 p_{\mathcal{N}_1} dx \\ &= \mathbb{E}_1 \left[ \frac{1}{2} \left( \frac{x - \mu_1}{\sigma_1} \right)^2 \right] \\ &= \frac{1}{2\sigma_1^2} \mathbb{E}_1 [(x - \mu_1)^2] \\ &= \frac{1}{2\sigma_1^2} \sigma_1^2 \\ &= \frac{1}{2}. \end{aligned}$$

et

$$\begin{aligned} \alpha &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{x - \mu_2}{\sigma_1} \right)^2 p_{\mathcal{N}_1}(x) dx \\ &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{(x - \mu_1) + (\mu_1 - \mu_2)}{\sigma_1} \right)^2 p_{\mathcal{N}_1} dx \\ &= \int_{\mathbb{R}} \frac{1}{2} \left( \frac{x - \mu_1}{\sigma_2} \right)^2 p_{\mathcal{N}_1}(x) dx + \int_{\mathbb{R}} \frac{1}{2} \left( \frac{\mu_1 - \mu_2}{\sigma_2} \right)^2 p_{\mathcal{N}_1}(x) dx + \int_{\mathbb{R}} \left( \frac{(x - \mu_1)(\mu_1 - \mu_2)}{\sigma_2} \right)^2 p_{\mathcal{N}_1}(x) dx \\ &= \frac{1}{2\sigma_2^2} \mathbb{E}_1 [(x - \mu_1)^2] + \frac{\mu_1 - \mu_2}{2\sigma_2^2} \mathbb{E}_1 [(x - \mu_1)] + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2} \\ &= \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2}. \end{aligned}$$

En combinant les deux termes  $\alpha$  et  $\beta$  on en déduit la forme analytique suivante :

$$\text{KL}(\pi_{\mathcal{N}_1} \parallel \pi_{\mathcal{N}_2}) = \frac{\sigma_1 \sqrt{2\pi}}{2} \left( \frac{\sigma_1^2}{\sigma_2^2} + \left( \frac{\mu_1 - \mu_2}{\sigma_2} \right)^2 - 1 \right). \quad (\text{A.4})$$

## A.2 Calcul de la divergence de Kullback-Leibler pour des distributions de la forme de $\pi_{\mathcal{W}}$

Cette section détail le calcul nécessaire pour déterminer l'équation (2.54) dans le Chapitre 2 section 2.3.3.



Considérons des distributions de possibilité telle que :

$$\pi_{W_i}(x) = \exp\left(-\left(\frac{|x|}{\alpha_i}\right)^{\beta_i}\right), \quad (\text{A.5})$$

et la divergence de Kullback-Leibler possibiliste non-symétrique de la forme suivante :

$$\text{KL}(\pi_{W_1} \parallel \pi_{W_2}) = \int_{\mathbb{R}} \pi_{W_1}(x) \log\left(\frac{\pi_{W_1}(x)}{\pi_{W_2}(x)}\right) dx. \quad (\text{A.6})$$

Alors

$$\begin{aligned} \text{KL}(\pi_{W_1} \parallel \pi_{W_2}) &= \int_{\mathbb{R}} e^{-\left(\frac{|x|}{\alpha_1}\right)^{\beta_1}} \log\left(\frac{e^{-\left(\frac{|x|}{\alpha_1}\right)^{\beta_1}}}{e^{-\left(\frac{|x|}{\alpha_2}\right)^{\beta_2}}}\right) dx \\ &= 2 \int_{\mathbb{R}^+} e^{-\left(\frac{x}{\alpha_1}\right)^{\beta_1}} \log\left(\frac{e^{-\left(\frac{x}{\alpha_1}\right)^{\beta_1}}}{e^{-\left(\frac{x}{\alpha_2}\right)^{\beta_2}}}\right) dx \\ &= -2 \int_{\mathbb{R}^+} \left(\frac{x}{\alpha_1}\right)^{\beta_1} e^{-\left(\frac{x}{\alpha_1}\right)^{\beta_1}} - \left(\frac{x}{\alpha_2}\right)^{\beta_2} e^{-\left(\frac{x}{\alpha_1}\right)^{\beta_1}} dx. \end{aligned}$$

En procédant au changement de variable  $\left(\frac{x}{\alpha_1}\right)^{\beta_1} = y$  dans la dernière égalité, on obtient :

$$\begin{aligned} \text{KL}(\pi_{W_1} \parallel \pi_{W_2}) &= -2 \int_{\mathbb{R}^+} \frac{\alpha_1}{\beta_1} y^{\frac{1}{\beta_1}} e^{-y} - \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2+1} y^{\left(\frac{\beta_2+1}{\beta_1}-1\right)} e^{-y} dy \\ &= -2 \frac{\alpha_1}{\beta_1} \Gamma\left(1 + \frac{1}{\beta_1}\right) + 2 \frac{\alpha_1}{\beta_1} \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \Gamma\left(\frac{1+\beta_2}{\beta_1}\right) \end{aligned}$$

Ce qui devient après quelques étapes :

$$\text{KL}(\pi_{W_1} \parallel \pi_{W_2}) = 2 \frac{\alpha_1}{\beta_1} \Gamma\left(\frac{1}{\beta_1}\right) \left( \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma\left(\frac{1+\beta_2}{\beta_1}\right)}{\Gamma\left(\frac{1}{\beta_1}\right)} - \frac{1}{\beta_1} \right) \quad (\text{A.7})$$

où  $\Gamma(u) = \int_{\mathbb{R}^+} x^{u-1} e^{-x} dx$ .

### A.3 Calcul de la divergence de Kullback-Leibler pour des distributions de possibilité elliptiques de la forme de

$\pi_{\mathcal{N}}$

Cette section détail les étapes de calculs effectués pour obtenir la forme analytique de l'équation (3.29) dans la section 3.3.2.

Les distributions de possibilité elliptiques que l'on considère ici sont telles que :

$$\pi_{\mathcal{N}_i}(x) = \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right). \quad (\text{A.8})$$

Calculons d'abord la partie asymétrique de la divergence de Kullback-Leibler :

$$\begin{aligned}
\text{KL}(\pi_{\mathcal{N}_1} \parallel \pi_{\mathcal{N}_2}) &= \int_{\mathbb{R}} \pi_{\mathcal{N}_1}(x) \log\left(\frac{\pi_{\mathcal{N}_1}(x)}{\pi_{\mathcal{N}_2}(x)}\right) dx \\
&= \int_{\mathbb{R}} \pi_{\mathcal{N}_1}(x) \log(\pi_{\mathcal{N}_1}(x)) dx - \int_{\mathbb{R}} \pi_{\mathcal{N}_1}(x) \log(\pi_{\mathcal{N}_2}(x)) dx \\
&= \int_{\mathbb{R}} \frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \pi_{\mathcal{N}_1}(x) dx \\
&\quad - \int_{\mathbb{R}} \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \pi_{\mathcal{N}_1}(x) dx.
\end{aligned}$$

En remarquant que  $\pi_{\mathcal{N}_1}(x) = 2\pi\sqrt{|\Sigma_1|} p_{\mathcal{N}_1}(x)$  où  $p_{\mathcal{N}_1}$  est donnée dans l'équation 3.21 pour  $n = 2$ , on peut écrire :

$$\begin{aligned}
\text{KL}(\pi_{\mathcal{N}_1} \parallel \pi_{\mathcal{N}_2}) &= 2\pi\sqrt{|\Sigma_1|} \left( \int_{\mathbb{R}} \frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) p_{\mathcal{N}_1}(x) dx \right. \\
&\quad \left. - \int_{\mathbb{R}} \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) p_{\mathcal{N}_1}(x) dx \right). \tag{A.9}
\end{aligned}$$

Notons  $\alpha$  et  $\beta$  les deux intégrales dans l'équation précédente :

$$\begin{aligned}
- \alpha &= \int_{\mathbb{R}} \frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) p_{\mathcal{N}_1}(x) dx \\
- \beta &= \int_{\mathbb{R}} \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) p_{\mathcal{N}_1}(x) dx.
\end{aligned}$$

En notant  $\mathbb{E}(x)$  l'espérance de la variable aléatoire  $X \in \mathbb{R}^2$ , nous avons :

$$\begin{aligned}
\beta &= \int_{\mathbb{R}} \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) p_{\mathcal{N}_1}(x) dx \\
&= \mathbb{E}_1 \left[ \text{tr} \left( \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right) \right] \\
&= \frac{1}{2} \mathbb{E}_1 \left[ \text{tr} \left( (x - \mu_1)(x - \mu_1)^T \Sigma_1^{-1} \right) \right] \\
&= \frac{1}{2} \text{tr} \left[ \mathbb{E}_1 \left( (x - \mu_1)(x - \mu_1)^T \right) \Sigma_1^{-1} \right] \\
&= \frac{1}{2} \text{tr} \left( \Sigma_1 \Sigma_1^{-1} \right) \\
&= \frac{1}{2} \text{tr} (\mathbb{I}_2) \\
&= 1 \quad .
\end{aligned}$$

où  $I_2 \in \mathbb{R}^{2 \times 2}$  est la matrice identité.

Et

$$\begin{aligned}
\alpha &= \int_{\mathbb{R}} \frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) p_{\mathcal{N}_1}(x) dx \\
&= \int_{\mathbb{R}} \frac{1}{2} [(x - \mu_1) + (\mu_1 - \mu_2)]^T \Sigma_2^{-1} (x - \mu_1) + (\mu_1 - \mu_2) p_{\mathcal{N}_1}(x) dx \\
&= \int_{\mathbb{R}} \frac{1}{2} ((x - \mu_1)^T \Sigma_2^{-1} (x - \mu_1) + 2(x - \mu_1)^T \Sigma_2^{-1} (\mu_1 - \mu_2) \\
&\quad + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)) p_{\mathcal{N}_1}(x) dx \\
&= \frac{1}{2} (\mathbb{E}_1[(x - \mu_1)^T \Sigma_2^{-1} (x - \mu_1)] + \mathbb{E}_1[2(x - \mu_1)^T \Sigma_2^{-1} (\mu_1 - \mu_2)] \\
&\quad + \mathbb{E}_1[(\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)]) \\
&= \frac{1}{2} [\text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)]
\end{aligned}$$

En combinant  $\alpha$  et  $\beta$  on obtient alors :

$$\text{KL}(\pi_{\mathcal{N}_1} \| \pi_{\mathcal{N}_2}) = \pi \sqrt{|\Sigma_1|} \left[ \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) - 2 \right]. \quad (\text{A.10})$$

Finalement, on détermine (3.29) en ajoutant les deux versions asymétrique de la divergence de Kullback-Leibler.

NB : Il est notable que le terme  $\beta$  est similaire à l'entropie de Shanon d'une distribution de possibilité elliptique  $\pi_{\mathcal{N}_1}$  tel que :

$$\begin{aligned}
\beta &= \int \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) p_{\mathcal{N}_1}(x) dx \\
&= -\frac{1}{2\pi \sqrt{|\Sigma_1|}} \int \pi_1(x) \ln(\pi_{\mathcal{N}_1}(x)) dx
\end{aligned}$$

De plus  $\beta = 1$ , on a alors :

$$-\int \pi_{\mathcal{N}_1}(x) \ln(\pi_1(x)) dx = 2\pi \sqrt{|\Sigma_1|}. \quad (\text{A.11})$$

Et d'autre part, on a :

$$\int \pi_{\mathcal{N}_1}(x) dx = 2\pi \sqrt{|\Sigma_1|} \int p_{\mathcal{N}_1}(x) dx = 2\pi \sqrt{|\Sigma_1|}. \quad (\text{A.12})$$

Or  $\int p_{\mathcal{N}_1}(x) dx = 1$ .

Ainsi, on peut écrire une propriété d'invariance de la façon suivante :

$$\int_{-\infty}^{+\infty} \pi_{\mathcal{N}}(x) (1 + \ln(\pi_{\mathcal{N}}(x))) dx = 0 \quad (\text{A.13})$$

## A.4 Calcul de la divergence d'Hellinger pour des distributions de possibilité elliptiques de la forme de $\pi_{\mathcal{N}}$

Cette section détail les étapes de calculs effectués pour obtenir la forme analytique de l'équation (3.30) dans la section 3.3.2.

Les distributions de possibilité elliptiques que l'on considère ici sont telles que :

$$\pi_{\mathcal{N}_i}(x) = \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right). \quad (\text{A.14})$$

La divergence d'Hellinger est alors :

$$\begin{aligned} D_{\text{H}}^2(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) &= \frac{1}{2} \int_{\mathbb{R}} \left( \sqrt{\pi_{\mathcal{N}_1}(x)} - \sqrt{\pi_{\mathcal{N}_2}(x)} \right)^2 dx \\ &= \frac{1}{2} \int_{\mathbb{R}} \pi_{\mathcal{N}_1}(x) dx + \frac{1}{2} \int_{\mathbb{R}} \pi_{\mathcal{N}_2}(x) dx \\ &\quad - \int_{\mathbb{R}} \sqrt{\pi_{\mathcal{N}_1}(x) \pi_{\mathcal{N}_2}(x)} dx. \end{aligned}$$

avec  $\int_{\mathbb{R}} \pi_{\mathcal{N}}(x) dx = 2\pi \sqrt{|\Sigma|}$ .

On pose :

$$K_r(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) = \int_{\mathbb{R}} \pi_{\mathcal{N}_1}^r(x) \pi_{\mathcal{N}_2}^{1-r}(x) dx.$$

D'après [134], on peut récrire  $A = (x - \mu_1)^T r \Sigma_1^{-1}(x - \mu_1) + (x - \mu_2)^T (1-r) \Sigma_2^{-1}(x - \mu_2)$  par :

$$A = (x - \mu^*)^T C^{-1}(x - \mu^*) + B,$$

où  $\mu^* = (r \Sigma_1^{-1} + (1-r) \Sigma_2^{-1})^{-1} (r \Sigma_1^{-1} \mu_1 + (1-r) \Sigma_2^{-1} \mu_2)$ ,

$$C^{-1} = (r \Sigma_1^{-1} + (1-r) \Sigma_2^{-1}),$$

et  $B = (\mu_1 - \mu_2)^T (r \Sigma_1^{-1} + (1-r) \Sigma_2^{-1}) r(1-r)(\mu_1 - \mu_2)$ .

Donc  $K_r(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) = 2\pi |C^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{B}{2}\right)$ .

Puis en posant  $r = \frac{1}{2}$  on en déduit :

$$D_{\text{H}}^2(\pi_{\mathcal{N}_1}, \pi_{\mathcal{N}_2}) = \pi \left( \sqrt{|\Sigma_1|} + \sqrt{|\Sigma_2|} - |C^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{B}{2}\right) \right). \quad (\text{A.15})$$

Enfin, en remplaçant les termes C et B par leurs valeurs, on retrouve l'expression de l'équation (3.30).

