Numéro d'ordre: XX

Habilitation à Diriger les Recherches

préparée au

Centre de Recherche en Informatique Signal et Automatique de Lille

Université de Lille, Centrale Lille, CNRS, UMR 9189 - CRIStAL

Structures algébriques et métriques pour les fonctions de croyance.

en vue d'une soutenance le 7 décembre 2017 par

John Klein

pour obtenir l' **Habilitation à Diriger les Recherches**

Spécialité
Automatique, Génie Informatique, Traitement du Signal et Image

Jury

Rapporteurs	
Thierry Denœux	Professeur des Universités, Université de Technologie de Compiègne
Didier Dubois	Directeur de Recherche, CNRS - IRIT
Arnaud Martin	Professeur des Universités, Université de Rennes 1
Examinateurs	
Fabio Cuzzolin	Professor, Oxford Brookes University
Gert de Cooman	Professor, Universiteit Gent
Michèle Rombaut	Professeur des Universités, Université Grenoble Alpes
Garant de l'HDR	
Olivier Colot	Professeur des Universités, Université Lille 1 Sciences et Technologies

Contents

Foreword

Data sciences aim at retrieving useful information from observed data. In general, this goal is formalized as the estimation problem of an unknown and potentially multidimensional variable θ. Most of the time, we do not have access to direct observations of this variable and we must deduce as much as possible from data that have an ill known link to the variable. This process is called **inference**.

This process is doomed to fail if data and variable are unrelated. The existence of a pattern between observations and the variable of interest is a prerequisite to inference and we will assume that this pattern does exist. In practice, this pattern can be highly complex and it is not always useful to understand it in details. To mitigate our lack of knowledge on the mechanism relating our data to θ, a solution is to resort to approximate reasoning. This is usually acceptable as long as one is able to quantify the approximation effects on the obtained estimation. By accepting to resort to approximate reasoning, one enters a whole new world of models that can potentially be more simple and computationally tractable as compared to the real one.

The most popular framework for this type of reasoning is probabilistic modeling. Probability theory has been extensively studied by many mathematicians for many centuries and its ability to capture efficiently random phenomenons and to represent uncertain information is remarkable. Thanks to probabilities, one can evaluate which value of θ is more likely than others. When the analysis permits it (as in Bayesian statistics), one can also determine to what extent the predictions on θ are reliable. Beyond the need for simplifying models, probabilistic modeling is also justified sometimes by the random feature of some data (presence of noise).

If data allow only to retrieve partial knowledge concerning θ, it is legitimate to wonder if they can in the mean time allow to estimate exact probabilities of θ. Obviously, there are two levels of uncertainty: uncertainty on the actual value of θ and uncertainty in the probability values of the variable values.

Several solutions are possible to take these two levels of uncertainty into account. In absolute terms, it is necessary to compute probabilities of probabilities (random measures).

The solution introduced by Dempster 1967 offers an intermediate answer and allows to define two functions that can be seen as upper and lower bounds of the probabilities of θ. The lower bound is called belief function and gave its name to the eponymous theory. Inference can be performed on the basis of the framework laid bare by Dempster and later completed by Shafer 1976 and several other authors.

Belief function based approximate reasoning is the main subject of this manuscript. Across research activities I carried out in my early associate professor career, I mainly focused on endowing the space where belief functions live with structures that allow belief function analysis, comparison and combination. My main approach consisted in deriving new tools that are consistent with the informational content of a belief function.

Résumé

La science des données a pour but principal l'extraction de connaissances à partir d'observations. De manière générale, le but est de déterminer la valeur prise par une variable inconnue θ. N'ayant pas accès directement à des observations de cette variable, il faut alors déduire un maximum d'informations sur θ à partir des données dont nous disposons. Ce processus de déduction se nomme **inférence**.

Ce processus n'a aucune chance d'aboutir si les données sont indépendantes de la variable θ. L'existence d'un lien sémantique entre les observations et la variable d'intérêt est un pré-requis qu'on supposera vérifié. En pratique, ce lien peut être d'une très grande complexité et il n'est pas forcément utile de le comprendre en détail. Pour palier notre méconnaissance du mécanisme liant nos données à θ, une solution consiste à avoir recours au raisonnement approximatif. Le tout est de pouvoir quantifier l'approximation et ses effets sur notre estimation. Si l'on accepte le recours à cette pratique, on ouvre la voie à tout un nouveau monde de modèles qui prendront potentiellement une forme beaucoup plus simple et synthétique que le modèle réel.

L'outil privilégié pour ce type raisonnement est la modélisation probabiliste. Ce formalisme largement étudié par les mathématiciens depuis plusieurs siècles a montré sa capacité à capturer efficacement des phénomènes aléatoires et à représenter des informations incertaines. Grâce aux probabilités, on peut savoir quelle valeur de θ est plus vraisemblable qu'une autre. Quand l'analyse probabiliste est poussée (approche Bayésienne) on sait également dire quel degré de confiance nous pouvons accorder à notre prédiction de la valeur de θ.

Au delà de l'intérêt du raisonnement approximatif, la modélisation probabiliste se justifie aussi parfois par le caractère incertain des données (présence de bruit).

Si les données ne nous permettent de remonter qu'à une connaissance partielle de θ, il est légitime de se demander si elles permettent de remonter à une connaissance exacte des probabilités de θ. A l'évidence, il y a un double niveau d'incertitude : incertitude sur la valeur prise par θ et incertitude sur la valeur des probabilités de θ.

Plusieurs solutions sont possibles pour prendre en compte ce double niveau. Dans l'absolu, il faut alors calculer des probabilités de probabilités (mesures aléatoires). La solution proposée par Dempster 1967 offre une réponse intermédiaire en permettant d'obtenir deux fonctions pouvant être vues respectivement comme une borne inférieure et une borne supérieure sur les valeurs des probabilités de θ. La borne inférieure est appelée fonction de croyance et a donné son nom à la théorie éponyme qui consiste à construire des processus inférentiels dans le cadre posé par Dempster puis complété par de nombreux auteurs.

Le raisonnement approximatif à partir de fonctions de croyance est l'objet principal de ce mémoire. Au cours des travaux de recherche que j'ai pu mener dans mon début de carrière d'enseignant-chercheur, je me suis principalement attaché à doter l'espace dans lequel vivent les fonctions de croyance de structures permettant leur analyse, comparaison et combinaison. Dans cette démarche, mon objectif principal a été de dériver de nouveaux outils cohérents avec le contenu informationnel d'une fonction de croyance.

List of Symbols

Notation	Description
P_X	Probability distribution of the discrete random variable X
p_X	Probability density of the continuous random variable X
Ω	An abstract space inducing probabilities in another space through a random variable
σ_Ω	The canonical σ-algebra or σ-field associated to a space Ω.
μ	A probability measure on Ω
X	A random variable representing observations
$\mathbb{X}$	The space where X takes its values
θ	The (ill-known) variable of interest
Θ	The space where the variable of interest lives
2^Θ	The power set induced by the space Θ
n	The size of Θ when this latter is finite
N	The size of 2^Θ when this latter is finite, i.e. $N = 2^n$
$\mathbb{N}$	The set of positive integers
$\mathbb{R}$	The set of reals
$\mathbb{E}$	The expectation operator
Γ	A multi-valued mapping
$\Gamma_\top^{-1}$	Upper pseudo inverse of Γ
$\Gamma_\perp^{-1}$	Lower pseudo inverse of Γ
Γ^{-1}	Pseudo inverse of Γ (in the usual sense)
$P_\top$	Upper probability induced by a multi-valued upper inverse
$P_\perp$	Lower probability induced by a multi-valued lower inverse
S	A source
bel	A belief function
pl	A plausibility function
q	A commonality function
b	An implicability function
w	A conjunctive weight function
v	A disjunctive weight function

Notation	Description
m	A mass function
$\overline{m}$	Negation of a mass function
m_B	A categorical mass function focused on the subset $B \subseteq \Theta$
m_B^w	A simple mass function with B and Θ as focal elements
m^w	A discounted mass function with discount rate w
$m_{1\|B}$ or $m_1(.\|B)$	Conditional mass function
π	A contour function or possibility distribution
N	A necessity measure
Π	A possibility measure
ν	A capacity
ν^*	The conjugate capacity of ν
1_B	The indicator function for set B
B^c	The complement of set B
proj_Θ	Set projection operator
Δ	Set symmetric difference
$\setminus$	Set difference
$\|.\|$	Set cardinality
Φ	Cdf of a centered Gaussian distribution with unit variance
$\mathrm{Unif}_{[a;b]}$	The uniform distribution on the interval $[a;b]$
min or $\wedge$	The minimum operator
max or $\vee$	The maximum operator
$\mathcal{P}_\Theta$	The simplex of probability measures on Θ
$\mathcal{P}$	A probability measure set (p.m.-set), i.e. a subset of $\mathcal{P}_\Theta$
$\mathcal{P}_\nu$	The core of the capacity ν
$\underline{\nu}_\mathcal{P}$	The lower probability induced by $\mathcal{P}$
$\overline{\nu}_\mathcal{P}$	The upper probability induced by $\mathcal{P}$
$\mathbf{vec}$	The vectorization operator
$\mathbf{m}$	A mass vector
$\mathbf{I}$	The identity matrix
$\mathbf{1}$	The all-ones matrix
$\mathbf{J}$	The matrix with null components except those on the anti-diagonal which are equal to 1
$\mathbf{K}_\alpha^\cap$	An α-specialization matrix
$\mathbf{K}_\alpha^\cup$	An α-generalization matrix

Notation	Description
$\otimes$	Product measure
$\oplus$	Dempster's rule
$\bigcirc\!\!\!\!\cap^{\alpha}$	An α-conjunctive rule
$\bigcirc\!\!\!\!\cup^{\alpha}$	An α-disjunctive rule
$\bigcirc\!\!\!\!\cap$	The conjunctive rule
$\bigcirc\!\!\!\!\cup$	The disjunctive rule
$\bigcirc\!\!\!\!\wedge$	The cautious rule
$\bigcirc\!\!\!\!\vee$	The bold rule
$\sqcap$	A distance based conjunctive operator
$\sqcup$	A distance based disjunctive operator
$\preceq$	A pre-order
$\sqsubseteq$	A partial order
Spe	Specificity measure for mass functions
Card	Expected cardinality for mass functions
κ	Dempster's degree of conflict
$\kappa\left(m\right)$	Mass assigned to $\emptyset$, i.e. $m\left(\emptyset\right)$
$\kappa_{\mathcal{A}}$	Conflict induced by the conjunctive combination of the family $\mathcal{A}$ of mass functions
$\kappa_{\mathcal{A}}\left(\boldsymbol{\alpha}\right)$	Conflict induced by the conjunctive combination of the family $\mathcal{A}$ of discounted mass functions - discount rates are the entries of $\boldsymbol{\alpha}$
ℓ	The cardinality of a family of mass functions
$\mathcal{M}$	The mass space
$\mathcal{M}_x\left(m_1\right)$	The set of mass functions that are x-included in m_1
d_J	Jousselme distance
$d_{f,k}$	L_k norm based distance between set functions of type $f \in \{pl, q, bel, b\}$
$d^{\cap,\alpha}_{mat,k}$	L_k matrix norm based distance between between α-specialization matrices
$d^{\cup,\alpha}_{mat,k}$	L_k matrix norm based distance between between α-generalization matrices
$d^{\cap,\alpha}_{opk}$	k operator matrix norm based distance between between α-specialization matrices
$d^{\cup,\alpha}_{opk}$	k operator matrix norm based distance between between α-generalization matrices

General introduction

This document has been written as part of the preparation of my *Habilitation à diriger les recherches* (HDR) degree. It is focused on my main research interest which is the theory of belief functions. In the body of this document, I thus develop a presentation of this framework and of my contributions in this field. This presentation is spread across several chapters. In the first chapter, I try to give a global picture of uncertainty theories and how the theory of belief functions places itself in this picture. In the second chapter, I present important elements of the theory. My personal contributions are sorted in the three next chapters which review the structure of the space where belief functions live. I examine order theoretic structures, metric structures and algebraic structures, hence the title of this document. The body of this document is written as if it was a monograph on belief functions and associated structures therefore I will use a conventional writing style and use ”we” instead of ”I”. However, I will spend more time on my personal contributions (providing proofs and more examples) because this is necessary for my evaluation. Of course, this should not be interpreted as a narcissist fascination for my own work.

In the next paragraphs, I will give a short presentation of belief functions and their applications. Since the evaluation of an HDR candidate is not limited to one research topic, I will also present briefly my other research activities which mainly deal with medical image processing and machine learning.

I will also give some personal and general information on my career as assistant professor in the university of Lille1. To summarize my career, I will address separately the following points: teaching activities, team leading activities and scientific activities. Concerning scientific activities, I will highlight PhD supervisions and related publications as well as other publications that were written outside the scope of these PhDs.

Belief functions in a nutshell

In the late sixties, Arthur P. Dempster published a series of papers around a mathematical model that can be viewed as a random variable which cannot be precisely observed. By precisely observed, we mean that a realization of this variable does not always translate in a single valued observation but a set valued one. Dempster's model relies on two ingredients: a probability space and a multi-valued mapping (a mapping whose values are sets). By introducing pseudo-inverses of this multi-valued mapping, he shows that the model can also be viewed as a system of lower and upper bounds for probabilities associated to the variable values.

A few years later, Dempster supervised Glenn Shafer's seminal work which introduced many results for this model. After his PhD, Shafer published a book to summarize his ideas and he defended that reasoning under uncertainty using this model makes sense and generalizes some probabilistic calculus rules but it is not always necessary to rely on an underlying probability space. The functions that share the same properties as Dempster's lower probabilities are thus called in general belief functions. Those sharing the same properties as Dempster's upper probabilities are called plausibility functions. Since plausibility functions are in bijective correspondence with belief functions, the term belief functions became after some time the default denomination of the model and the whole framework as well.

The pioneering work of Dempster and Shafer has dragged the attention of many researchers from the artificial intelligence, statistical sciences and mathematical logic communities. We have recently celebrated 50 years of research on belief functions and related topics. The next chapter are mainly focused on theoretical aspects of belief functions therefore I want to mention here that the framework has been applied in many fields (see table 1).

I personally discovered the framework of belief functions in the second year of my PhD when I had the idea to combine several likelihood functions to improve a particle filter for vehicle tracking in on-road videos. Ten years later, my interest for Dempster and Shafer's ideas is untouched. After I was recruited as assistant professor

Table 1: A short list of articles with applications of belief functions.

Application Field	Examples of approaches relying on belief functions
Machine Learning	•Denœux 1995 presents a k nearest neighbor classification algorithm where data points induce belief functions. •Xu et al. 2015 present a calibration method to turn the output of classifiers like SVMs into belief functions and build a classifier ensemble. •Denœux et al. 2015 introduce an unsupervised iterative clustering algorithm. Data point dissimilarities are used to compute cluster membership plausibilities. This approach is non-parametric and the number of clusters can be set to an arbitrary initial value.
Information Fusion	•Clemens et al. 2016 use a belief function based sensor fusion as part of a self localization and mapping algorithm. The benefits of belief functions as compared to standard probabilities is that the origins of map uncertainties are more easy to interpret (lack of information, inconsistencies between sensor outputs, etc.) •Klein et al. 2009 combine belief functions based on likelihood functions to obtain robust estimates of the probability that an object is present in an image. Each likelihood relies on different image features (texture, shape, color and movement). The object is tracked through an image sequence using a particle filter.
Intelligent Vehicles	•El Zoghby et al. 2014 introduce an approach allowing vehicles to build belief functions that characterize the uncertainty of the presence of obstacles on road. Information is exchanged with neighbor vehicles as part of vehicular ad hoc networks.
Computer Vision	•Gong and Cuzzolin 2017 address human pose estimation. Mappings from image features to poses are learned from examples. Belief functions allow to learn multi-valued mappings. The approach outperforms relevance vector machines or Gaussian process regression.
Econometrics	•Autchariyapanitkul et al. 2014 use belief functions for quantile regression in order to predict stock returns in the capital asset pricing model.
Signal & Image Processing	•Lian et al. 2017 introduce a belief function based voxel clustering technique for a segmentation in 3D medical imaging. •Labourey et al. 2015 use Denœux's belief function based k nearest neighbor classifier on sound signals to recognize human activity.
Statistics	•Martin et al. 2010 build a high dimensional statistical test by using belief function based inference. In particular, they derive a Poisson process homogeneity test which outperforms a likelihood ratio test.
Social Networks	•Dhaou et al. 2017 use belief functions to detect communities in large graphs representing connected people in a social network.
Risk analysis	•Démotier et al. 2006 assess the plausibility of various noncompliance scenarios for water treatment based on weak information.

in Lille1, I began to study several tools in the framework of belief functions around notions of conflict, distances and combinations. These works will be exposed in the next chapters. I also diversified a bit my activities as I will explain in the next section.

Other research activities

In this section, I give a brief presentation of research activities of mine that are unrelated to belief functions.

Medical image processing

I started working on a medical image processing problem during the supervision of Cyrille Feudjio's PhD. These works were focused on mammograms (X-ray imaging of woman's breast). This imaging modality is the most frequently used one around the world to detect breast cancer signs because of its moderate cost. Given the huge amount of mammograms produced during screening campaigns, radiologists are seeking computerized assistance to analyze these images.

The goal of this PhD was to design an image processing methodology that allows to prioritize patients. Physicians have established a strong correlation between the amount of dense tissues in mammary glands and cancer risk. Consequently, we addressed dense tissue segmentation in order to compute an approximate dense tissue ratio.

We rapidly realized that the problem must be tackled step by step. Indeed, there are several image regions that are irrelevant for dense tissue detection: background and pectoral muscle tissues. We thus designed specific algorithms for these preliminary tasks. Figure 1 summarizes the global approach.

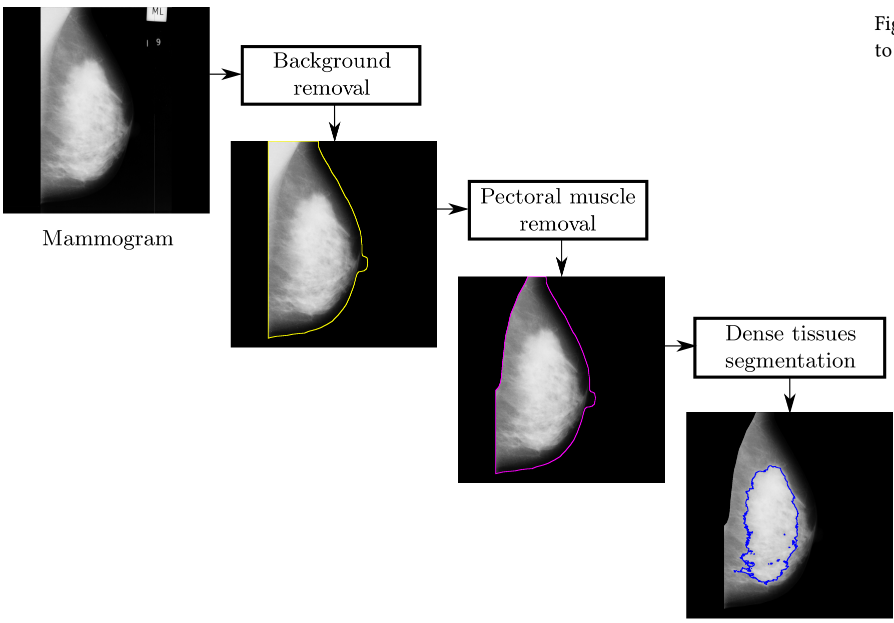


Figure 1: Analysis steps of a mammogram to detect dense tissues.

The two preliminary steps were performed using a standard segmentation technique, namely fuzzy c-means. Our contribution does not lie on the image processing side but only on the methodology consisting in using spatial information and a post-processing to obtain a really satisfying image region detection. The proposed methodologies achieves better or comparable results as compared to state-of-the-art approaches with a limited complexity.

The final step is more original from an image processing standpoint. First, dense tissues can be accurately segmented by a mere thresholding operation. The problem thus translates into a threshold estimation one. We proposed to use the same threshold for each mammogram provided they underwent a histogram specification prior to thresholding. We build an objective function to optimize in order to find the gamma correction that will best lead to the prescribed histogram which is a bimodal one (one mode for dense tissue pixels, the other one for non-dense tissue pixels). This objective function relies on a Wasserstein distance between the target histogram and the histogram obtained after gamma correction. A difficulty is that some mammogram histograms are unimodal (only dense tissues or no dense tissues at all) therefore we added a regularization term to the objective function to prevent too strong image modifications.

Machine learning

Since many information fusion questions arise when working with belief functions, I started to wonder if information fusion (on the decision side) could be useful for supervised classification or regression tasks. I discovered a vast literature on various ways of mixing learning algorithms. There are of course the famous ensemble methods (like bagging and boosting) but also more generic approaches to mix heterogeneous classifiers.

A simple idea consists in training several classifiers on the same dataset and tune their hyperparameters at best. We can then perform a majority vote on their predictions. Even this simple scheme turns out to allow lower error rates sometimes. However, the fusion of classifier predictions can obviously be improved if we take their reliabilities into account. This idea is the starting point of Mahmoud Albardan's PhD which is ongoing.

So far, we have come up with an approach relying on confusion matrices. The matrices are estimated by cross-validation. We normalize them to obtain conditional distributions of classifier predictions given the true class. We then combine these probability distributions using parametrized t-norms followed by a renormalization. The t-norm parameter allows to select our model from a continuum ranging from the independence case to the total dependence case. By total dependence, we mean that classifiers always predict the same classes. The t-norm parameter is set by grid search. Preliminary results are promising and a manuscript was submitted in june.

Career

General information

Birth date	14/09/1982
Position	Maître de conférences - associate professor (appointed on 01/09/2009)
Career duration	8 years
Institution	Université Lille1 Sciences et Technologies
Research unit	CRIStAL UMR CNRS 9189
Research team	Signals, Models and Applications (SIGMA)
Address	bat. P2, avenue Carl Gauss, cité scientifique, 59655 Villeneuve d'Ascq, France
Phone number	+33 3 20 43 48 77
Email	john.klein@univ-lille1.fr
Homepage	http://sigma.univ-lille1.fr/node/85

Education

- 2005-2008: PhD in the University of Rouen / LITIS laboratory - *Suivi robuste d'objets dans les séquences d'images par fusion de sources, application au suivi de véhicules dans des scènes routières.*
  Jury members :
  - Olivier Colot (reviewer), professor, Université Lille1 Sciences et Technologies,

  - Thierry Denoeux (reviewer), professor, Université de Technologie de Compiègne,
  - Claude Lorgeau (examiner), professor, Ecole des Mines de Paris,
  - Laurent Trassoudaine (examiner), professor, Université Blaise Pascal de Clermont Ferrand,
  - Pierre Miché (Director), professor, Université de Rouen,
  - Christèle Lecomte (Supervisor), associate professeur, Université de Rouen,
- 2002-2005: Higher Engineer School at *Ecole Nationale Supérieure d'Electronique, Informatique et Radio-communications de Bordeaux* (ENSEIRB), telecommunication department and Master degree in signal and image processing in the University of Bordeaux 1 (in 2005),
- 2000-2001: *Classes préparatoires aux grandes écoles*, lycée Corneille, Rouen, section MPSI/MP.

Teaching activities

Since I arrived at the university of Lille1, my teaching activities are mainly around:
- Signal processing (temporal and frequency analysis, analog and digital filtering),
- Computer engineering (multi-task and real time operating systems),
- Programming (C language),
- Data sciences (machine learning and data fusion),
- Mobile robotics (introduction using LEGO mindstorm robots).

The above subjects are taught to master students. I also occasionally took part in control theory, digital electronics and industrial information technologies units for bachelor students. For each unit, I completely renewed the course materials (slides, exercises and practicals).

Most of the time, I teach in the electrical engineering and automation department of the university, which is the department I am assigned to. I also teach sometimes in the computer sciences department and in another partner institution named Ecole Centrale de Lille (one of the French *grandes écoles*).

In France, the due teaching workload is 192 hours a year for associate and full professors. I accepted almost every year overtime teaching hours because the number of students joining our department keeps growing year after year while the number of teaching staff members does not. In the worst case, overtime hours reached 50% of my annual workload which impacted my scientific production.

Management

In 2012, I became the leader of the automation team of the electrical engineering and automation department of the university. This team is responsible for teaching units related to
- control theory,
- industrial and field information technologies,
- signal processing,
- computer engineering.

There are 12 tenured teaching staff members in the team (2 full professors and 10 associate professors) and from 2 to 6 temporary teaching members (teaching assistants and PhD students). The team also comprises two electronics technicians, one IT manager and one secretary.

My missions are
- treasury management,
- meeting conduction,
- interacting with faculty dean and university vice-presidents,
- recruitment management,
- assignment of teachings to team members,
- advising during periodic organization revision of the department.

Publication history

International Journals (with JCR certification)	International Conferences (with reviewing process)
7	8

My scientific production is summarized by the above table and is detailed below. My name is in bold letters so as to highlight my author position. The names of my former PhD students are underlined. Some of my recent publications do not have PhD students as co-authors. These publications are either research works I carried outside the scope of PhD supervisions or are a continuation of PhD related works after the student graduated and left our resarch unit. Also, publications in connection with my PhD are spotted with ◇ symbol.

Articles in international journals

(from most recent to oldest)

1. **J. Klein**, S. Destercke, O. Colot, *Idempotent conjunctive and disjunctive combination of belief functions by distance minimization*, in **International Journal of Approximate Reasoning** (Impact Factor: 2.69), vol. 92, pp. 32-48, 2018 https://doi.org/10.1016/j.ijar.2017.10.004

2. **J. Klein**, S. Destercke, O. Colot, *Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation*, in **International Journal of Approximate Reasoning** (Impact Factor: 2.69), vol. 71, pp. 15-33, April 2016, http://dx.doi.org/ 10.1016/j.ijar.2016.01.001.

3. S. Li, H. Wang, Y. Tian, A. Aitouche and **J. Klein**. *Direct power control of DFIG wind turbine systems based on an intelligent proportional-integral sliding mode control*, in **ISA Transactions** (Impact Factor: 2.60), vol. 64, pp. 431-439, September 2016, http://dx.doi.org/10.1016/j.isatra.2016.06.003.

4. M. Loudahi, **J. Klein**, J. M. Vannobel and O. Colot, *Evidential Matrix Metrics as Distances Between Meta-Data Dependent Bodies of Evidence*, in **IEEE Transactions on Cybernetics** (Impact Factor: 4.94), vol. 46, no. 1, pp. 109-122, Jan. 2016., doi: 10.1109/TCYB.2015.2395877

5. M. Loudahi, **J. Klein**, J.-M. Vannobel, O. Colot, *New distances between bodies of evidence based on Dempsterian specialization matrices and their consistency with the conjunctive combination rule*, in **International Journal of Approximate Reasoning** (Impact Factor: 2.69), vol. 55, issue 5, pp. 1093-1112, July 2014, http://dx.doi.org/10.1016/j.ijar.2014.02.007.

6. C. Feudjio, **J. Klein**, A. Tiedeu, O. Colot, *Automatic extraction of pectoral muscle in the MLO view of mammograms*, in **Physics in Medicine and Biology** (Impact Factor: 2.81), vol.58 , no. 23, pp.8493-515, 2013, doi: 10.1088/0031-9155/58/23/8493.

7. **J. Klein** and O. Colot, *Singular sources mining using evidential conflict analysis* in **International Journal of Approximate Reasoning** (Impact Factor: 2.69), vol. 52, pp. 1433–1451, Dec. 2011, http://dx.doi.org/10.1016/j.ijar.2011.08.005

8. ◇**J. Klein**, C. Lecomte and P. Miché, *Hierarchical and conditional combination of belief functions induced by visual tracking*, in **International Journal of Approximate Reasoning** (Impact Factor: 2.69), vol. 51, pp. 410-428, March 2010, http://dx.doi.org/10.1016/j.ijar.2009.12.001

Articles in international conferences

(from most recent to oldest)

9. **J. Klein**, S. Destercke, O. Colot. *Idempotent Conjunctive Combination of Belief Functions by Distance Minimization*, in **Belief Functions: Theory and Applications: Fourth International Conference, BELIEF 2016**, Pragua, Czech Republic, September 21-23, 2016. Lecture Notes in Computer Science, Springer. doi: 10.1007/978-3-319-45559-4_16 **Best paper award**

10. **J. Klein**, M. Loudahi, J. M. Vannobel, O. Colot, *α-Junctions of Categorical Mass Functions*, in **Belief Functions: Theory and Applications: Third International Conference, BELIEF 2014**, Oxford, UK, September 26-28, 2014. Lecture Notes in Artificial Intelligence, Springer, doi: 10.1007/978-3-319-11191-9_1

11. M. Loudahi, **J. Klein**, J. M. Vannobel, O. Colot, *Fast Computation of L_p Norm-Based Specialization Distances between Bodies of Evidence*, in **Belief Functions: Theory and Applications: Third International Conference, BELIEF 2014**, Oxford, UK, September 26-28, 2014. Lecture Notes in Artificial Intelligence, Springer, doi: 10.1007/978-3-319-11191-9_46

12. **J. Klein**, O. Colot, *A Belief Function Model for Pixel Data*, in **Belief Functions: Theory and Applications: Second International Conference, BELIEF 2012**, Compiègne, France, September 26-28, 2014. Lecture Notes in Artificial Intelligence, Springer, doi: 10.1007/978-3-642-29461-7_22

13. **J. Klein**, O. Colot, *Automatic discounting rate computation using a dissent criterion*, in **Belief Functions: Theory and Applications: First International Conference, BELIEF 2010**, Brest, France, September 26-28, 2010.

14. ◇**J. Klein**, C. Lecomte and P. Miché, *Preceding car tracking using belief functions and a particle filter*, in **IEEE International Conference on Pattern Recognition, ICPR 2008**, Tampa, FL, USA, December 8-11, 2008, doi:10.1109/ICPR.2008.4761008

15. ◇**J. Klein**, C. Lecomte and P. Miché, *Tracking objects in videos with texture features*, in **IEEE International Conference on Electronics, Circuits and Systems, ICECS 2007**, Marrakech, Morocco, December 11-14, 2007, doi:10.1109/ICECS.2007.4511049.

16. ◇**J. Klein**, C. Lecomte and P. Miché, *Fast Color-Texture Discrimination: Application to Car Tracking*, in **IEEE Intelligent Transportation Systems Conference, ITSC 2007**, Seattle, WA, USA, September 30- October 3, 2007, doi:10.1109/ITSC.2007.4357765

PhD supervision

I have completed the supervision of two PhDs and one is ongoing. I give details on these PhDs hereafter. I want to stress that all former PhD students have published articles in international journals with a JCR impact factor. Concerning the ongoing PhD of Mahmoud Albardan, a manuscript has been submitted to "Information Fusion" and another one is in preparation.

Student Name	**Mehena Loudahi**
PhD Title	*Distances matricielles dans la théorie des fonctions de croyance pour l'analyse et caractérisation des interactions entre les sources d'informations*
PhD type	Contrat doctoral de l'université Lille1
Start Date	01/09/2011
Defense Date	01/12/2014
Related Publications	2 international journals, 2 international conferences and 1 national conference
Supervisors	Olivier Colot (Director), Jean-Marc Vannobel and John Klein
Supervision rate	50%
Post graduation position	Lecturer at Université d'Artois

Student Name	**Cyrille Feudjio Kougoum**
PhD Title	*Segmentation of mammographic images for computer aided diagnosis*
PhD type	Cotutelle with the University of Yaoundé1 (Cameroon), Erasmus Mundus funding
Start Date	01/09/2012
Defense Date	05/10/2016, the defense was delayed by 8 months in the wake of exceptional administrative problems (visa denied by French authorities). The defense was initially scheduled in December 2015.
Related Publications	1 international journal
Supervisors in Lille1	Olivier Colot (Director) and John Klein
Supervisors in Yaoundé1	Alain Tiédeu (Co-Director)
Supervision rate	50%
Post graduation position	Associate professor in Buea (Cameroon)

Student Name	**Mahmoud Albardan**
PhD Title	*Fusion de signaux physiologiques pour la reconnaissance d'états émotionnels*
PhD type	Contrat doctoral de l'université Lille1
Start Date	01/09/2015
Defense Date	31/08/2018 (prevision)
Related Publications	1 paper submitted to an international journal
Supervisors	Olivier Colot (Director) and John Klein
Supervision rate	70%

Peripheral research and education activities

A researcher's professional life is not limited to article writing, proof derivation or programming. We also have to carry out a number of peripheral tasks for our institutions and scientific community. Here follows a number of these tasks.

Associate professor recruitment committee	I joined 2 committees as local member (as vice-president for one of them) and 1 committee as non-local member in Université d'Artois.
Article Review	Reviewer for IJAR (International Journal of Approximate Reasoning), IEEE Transactions on Cybernetics, PRL (Pattern Recognition Letters), ESWA (Expert Systems With Applications) and Information Sciences.

Project	Sino-French CAI YUANPEI program for the development of the sino-French laboratory of automatic control and signal processing (LaFCAS). International collaboration between Lille1/Ecole Centrale de Lille and Nanjing University of Science and Technology. Project manager: Abdel Aitouche and Haoping Wang. The Cai Yuanpei program is meant to help to start "co-tutelle" PhD between France and China. There is no global funding but depending on demand, Campus France covers travel fees for supervisors and provides a grant for a PhD student long stay in France. The project lasted for 2 years (2015 and 2016). Although, a co-tutelle could not be established due to administrative complications on the Chinese side, the PhD student (Shanzhi Li) joined CRIStAL for 2 years. The PhD was initially thought to cover automatic control and estimation under uncertainty problems for wind turbine control. I was involved for a time in the supervision of Shanzhi Li but given that he is more keen on automatic control than estimation, we decided to focus on these aspects and I withdrew from the supervising team since my expertise in this field is limited.
	European INTERREG project INCASE - this project is an education European project for the 2 seas (Channel and north sea) geographic area. This project involves 11 public institutes and universities from Belgium, France, the Netherlands and UK. The project is concerned with the production of educational material and demonstrators in the field of 4.0 industry. This covers industrial IoT, mobile robotics and smart homes. This is a three year project that started in September 2016. In total, the project budget is around 4,500,000€.
Professional development courses	I gave a 2 day lecture on Machine Learning in June 2017 for Ministry of armies staff members. This training course was performed as part of the ANR EVE project although I am not a project member.
Local responsibilities	Coordinator of invited talks for the Signal&Image team of LAGIS and for the SIGMA team of CRIStAL (till June 2015).
	Invited member of the computer sciences, electrical engineering and automation board (since 2015).
Event Organization	Member of the organizing committee of the *Third School on belief functions and their applications*, Stella Plage, France, 2015.
Collaborations	With Sebastien Destercke from Heudiasyc, Compiègne, France, With Haoping Wang from Nanjing University of Science and Technology, China. With Cyrille Feudjio and Alain Tiédeu from Yaoundé 1 University, Cameroon.

1 Belief functions and related frameworks for inference under uncertainty

Inference is about making deductions about a variable of interest based on observed data. Because of data imperfections and our lack of knowledge on the model governing the relationships between observed and hidden variables, formal frameworks for reasoning under uncertainty are needed.

One such uncertainty theory is the theory of belief functions, a.k.a. Dempster-Shafer theory or evidence theory. We present the main features of this theory in the present chapter through the prism of statistical inference in order illustrate the concepts and their interests. We will also evoke the information fusion paradigm which bears some resemblance with inference. In information fusion, we also generally aim at making deductions about an unknown variable but we do not process raw observations but instead pieces of information that can potentially have already been distilled[1] from data.

1. This can be understood as the action of a previous processing in the data analysis flowchart but more generally it can also be the action of human reasoning. Indeed Shafer illustrated many concepts with respect to human testimony combination.

We start with a short presentation of probability theory which is by far the most popular uncertainty theory. Afterwards, we present belief functions and stress their added value as compared to mainstream probabilities but also the many connections between the two theories. Finally, we also present other related uncertainty frameworks that are highly overlapping belief function theory.

Besides, we also aim at reviewing different presentations and interpretations of belief functions in an approximate chronological order. We hope this will help the reader to get a global picture of the theory as well the motivations behind it which stem from statistical sciences.

1.1 A quick overview of probability theory

In this section, we provide a modest overview of the development of probability theory across modern history.

1.1.1 *Milestones in the development of probabilities as a theory*

There are many reported mathematical approaches to games of chance in ancient civilizations but the framework known today as the probability theory is the result of the aggregation of contributions that started by the time of the Renaissance in the western civilization. Across Europe, several eminent scholars were motivated by getting deeper insights in betting games and the notion of randomness, luck or uncertainty at large.

In the 17^{th} century, Cardano analyzed dice throwing problems and pointed out the relevance of ratios of favorable outcome counts against the total number of outcomes. In the mean time, Pascal and de Fermat had a correspondence on a question raised by Chevalier de Méré concerning the odds of getting at least a double six on 24 rolls of a pair of dices. De Méré correctly inferred that the odds of getting a double six in a single throw is 1 against 36. He unfortunately mistook about the odds of getting a double six after 24 rolls which he believed was 24 out of 36. The falsity in de Méré reasoning is obvious to modern statisticians. First, following de Méré reasoning throwing 36 times the dices implies sure win which is obviously untrue. Second, performing an analysis of the underlying binomial distribution, we have

$$P\left(\text{at least one double six in 24 rolls}\right) = 1 - P\left(\text{no double six in 24 rolls}\right), \tag{1.1}$$

$$= 1 - \left(\frac{35}{36}\right)^{24}, \tag{1.2}$$

$$= 0.4914. \tag{1.3}$$

A correct analysis of the problem was deduced by Pascal and de Fermat.

In the 18^{th} century, several contributors boosted the development of probabilities as a theory bringing more nobility to the scientific questions at stake which, since then, are not anymore regarded as merely a science of dice throws. Bernoulli introduced a first version of the law of great numbers. De Moivre introduced (among other things) the concept of statistical independence, the approximation of the binomial distribution by a normal one and he laid the foundations of the central limit theorem. Another contemporary of Bernoulli and de Moivre was Bayes. This clergyman derived the premises of Bayes theorem which is pivotal to almost any statistical inference problem. The strong implications of the formula were however much studied in the 20^{th} century from which the community of Bayesian statistics arose. It is hard to say though if Bayes himself would have embraced the philosophy behind Bayesian statistics as his writings do not deal with probability interpretation.

The 19^{th} century was a turning point for probabilities with successful applications in new areas of sciences and most notably physics with Gauss deriving of the orbit of Ceres and with Boltzmann's and Gibbs' modeling of gases. The theory development was in the meantime pursued notably by Laplace who introduced the moment generating functions, the least square method[2] and hypothesis testing. He also instated the concept of inductive probabilities which has a strong Bayesian flavor.

Finally, during the course of the 20^{th} century, the probability theory endured a fast and accelerated development around several aspects. It has touched almost any area of sciences. Some of the main highlights of this time are the axiomatization of probabilities by Kolmogorov and the development to a general theory of stochastic processes by Kolmogorov and Kinchin (as well as many other preceding contributors like Markov and Lévy).

In the 20^{th} century (and late 19^{th} century), statistics also emerged as a science of its own although it has always coexisted with probability throughout their mutual developments. Some of the most influential statisticians are Pearson, Fisher, Jeffreys, Neyman and de Finetti.

2. Laplace actually proposed a probabilistic model for estimation errors that he believed should be Laplace distributed. By minimizing the sum of absolute deviations, one retrieves the sample median as estimator. The true least square setting (with quadratic error minimization) was introduced by Legendre but was also claimed by Gauss. Gauss asked himself what probability density and what error minimization problem yield the arithmetic average as estimator. He thereby derived the normal distribution.

1.1.2 *Frequencies and subjective probabilities*

The interpretation of probabilities has dragged the attention of many mathematicians and statisticians but since the middle of the 20$^{\text{th}}$ century the discussions in this scope have lead to different schools of thought which advocate respectively for different understandings of what probabilities are and how they should be used in practice. Two such main schools are briefly presented hereafter.

Frequentism

The first school of thought, which was for a very long time the mainstream one (if not the only one) is called frequentism. It advocates that the probability of any event is the limiting frequency of occurrence of the event, i.e. the number of observed occurrences versus the number of trials when this latter number tends to infinity. This allows to grasp randomness and the model is supported by experiments which is an essential point in sciences. Typical frequentist probabilities are those involved in the problem of Chevalier de Méré. The events to which we assign probabilities are not true or false in general. We can only evaluate how frequently they are true or false.

The second school of thought, which will be presented in the next paragraph is Bayesianism. However, it is important to stress that frequentist approaches are not forbidden to resort to Bayes theorem which also holds for frequentist probabilities.

Bayesianism

Scholars involved in the development of probability theory rapidly came to the conclusion that uncertainty is a broader notion than randomness. Randomness stems from **ontic** or aleatoric uncertainty. In contrast, uncertainty may also be **epistemic**, meaning that it arises from incomplete knowledge. Some events are not repeatable and yet, it is possible to assign them a degree of belief of being true. Such degrees of beliefs are called subjective or personal probabilities. For example, one such event can be the presence or absence of a disease for a given patient. This event is either true or false. There is no repeated trials involved in the probability judgment of this event. Yet, given a series of observed symptoms such a probability can be inferred[3].

Roughly speaking, Bayesians accept all sorts of probabilities while frequentists reject subjective probabilities. But the Bayesian approach goes far beyond this. One of its guiding principles[4] is to draw conclusions about an unknown quantity based on observed data only. In frequentist approaches, it is not rare to manipulate conditional probabilities of the data given an unseen variable which is criticized by Bayesians on the grounds that inferences should be based on the events that happened, not which might have happened. To motivate this, suppose we can choose from two experiments producing a sample allowing inference for some parameter θ. The probability to choose this or that experiment does not

3. The probabilities relating the presence of the disease to symptoms may be frequentist ones. These probabilities ought to be the frequency of occurrence of the symptoms when *this patient* was sick. If one resorts to frequency of occurrences of the symptoms when *someone* is sick, then the end result is not anymore the probability of the patient being sick given the symptoms, but the probability of someone randomly sampled from the human population being sick given the symptoms.

4. sometimes called the conditionality principle.

depend on θ. Yet, if frequentist approaches like significance tests relying on p-values, we may get different conclusions from the same data depending on the experiment protocol. This is violation of likelihood principle which states that inferences should be based on data only.

The Bayesians also view inference as a belief update. Accepting that we all have prior (personal) beliefs on the variable of interest, collected data are confronted with this prior and we update our beliefs accordingly.

Furthermore, the Bayesian approach is also deeply entwined with decision theory and aims at taking decisions based on the integrated risk (or Bayes risk). This amounts to derive a decision rule by minimizing the posterior expected loss[5]. The computation of this quantity involves integrating out all hidden variables. It is unclear if those approaches that do not resort to the integrated risk can be called Bayesian, leaving them to a "statistical no man's land".

5. The posterior expected loss is the expectation of a prescribed loss function against the posterior distribution.

1.1.3 *Measure-theoretic formalization*

After Borel and Lebesgue introduced measure theory, Kolmogorov proposed a measure-theoretic axiomatization of probabilities that is up to now mostly undisputed. The main motivation behind this axiomatization is that it unifies discrete and continuous (or mixes of those) in a single framework. In this section, we briefly recall fundamental definitions of this framework so as to introduce notations and highlight differences with other objects defined in the theory of belief functions or imprecise probabilities.

Measure theory relies on the concept of σ-algebras.

Definition 1. let Ω denote an abstract space. A **σ-algebra** σ_Ω is a subset of 2^Ω such that:

- $\sigma_\Omega \neq \emptyset$,
- for all set $A \in \sigma_\Omega$, $A^c \in \sigma_\Omega$ (closed under complementation),
- for all set sequences $(A_n)_{n\geq 0} \in \sigma_\Omega$, $\left(\bigcup\limits_{n\geq 0} A_n\right) \in \sigma_\Omega$ (closed under countable union).

A pair (Ω, σ_Ω) is called **measurable space**.

Roughly speaking, Ω is the **exhaustive** set of possible and **exclusive outcomes** of an uncertain experiment. In the dice throw example, there are six possible outcomes (no more no less). The result of the experiment is an integer ranging from 1 to 6 and nothing else. The σ-algebra related to this experiment is the set of all random events that can be defined based on the experiment, like {*having an odd number*}.

From this definition of σ-algebras follows that of measurable functions.

Definition 2. Let (Ω, σ_Ω) and (Θ, σ_Θ) be two measurable spaces. Let f be a mapping from Ω to Θ. f is a **measurable function** if the preimage of B by f is in σ_Ω: $f^{-1}(B) \in \sigma_\Omega, \forall B \subseteq \Theta$.

A measurable function is a function whose preimages are discernible events. Let us now move to the concept of probability measure (Kolmogorov's axioms).

Definition 3. Let (Ω, σ_Ω) denote a measurable space. A **probability measure** μ is a mapping from σ_Ω to the interval $[0;1]$ such that:

- $\mu(\emptyset) = 0$,
- $\mu(A \cup B) = \mu(A) + \mu(B)$, for all $A, B \in \sigma_\Omega$ such that $A \cap B = \emptyset$ (additivity),
- $\mu(\Omega) = 1$.

The triplet $(\Omega, \sigma_\Omega, \mu)$ is called **probability space**. Let us now introduce a key concept for probabilistic analysis:

Definition 4. A measurable mapping θ from a probability space $(\Omega, \sigma_\Omega, \mu)$ to a measurable space (Θ, σ_Θ) is called **random variable** (r.v.).

The term *random variable* is rather awkward since this object is in fact not a variable but a mapping. The explanation behind this terminology is that usually Θ is a subset of $\mathbb{N}$ (discrete r.v.) or $\mathbb{R}$ (continuous r.v.). The σ-algebra associated with the co-domain of the r.v. contains quantifiable events like $\{\theta \leq 2\}$. We will use the following convention to denote the distribution of some r.v.:

- $P_\theta = \mu \circ \theta^{-1}$ in the discrete case,
- p_θ for the density of $\mu \circ \theta^{-1}$ w.r.t. Lebesgue (assuming existence) in the continuous case and $P_\theta(B)$ also denotes $\mu \circ \theta^{-1}(B) = \int_B p_\theta(a)\, da$.

When there is no ambiguity the r.v. in subscript may be omitted.

Finally, we also recall the definition of the expectation.

Definition 5. Let θ denote a real r.v. from $(\Omega, \sigma_\Omega, \mu)$ to (Θ, σ_Θ). The **expectation**[6] of a function[7] f is the following integral (if it exists):

$$\mathbb{E}_\theta[f] = \int_\Theta f dP_\theta. \tag{1.4}$$

Likewise, when there is no ambiguity the r.v. in subscript may be omitted. The concept of expectation is more general than that of probability since

$$\mathbb{E}_\theta[1_B] = \int_\Theta 1_B dP_\theta, \tag{1.5}$$
$$= \int_B dP_\theta, \tag{1.6}$$
$$= P_\theta(B), \tag{1.7}$$

where 1_B is the **indicator function** for set B.

6. The general definition of the expectation does not rely on some r.v. but only on a given measure μ. When there is no r.v. in action, the expectation will be denoted by $\mathbb{E}_\mu[f] = \int f d\mu$.

7. f is real valued and measurable.

1.1.4 *Shortcomings of probabilities*

There are no mathematical inconsistency in the theory of probabilities but only interpretability shortcomings, so to speak. These may occur when we try to match the mathematical objects defined above to describe a given situation involving uncertainty. Two types of situations must be distinguished:

- representation shortcoming: the situation cannot be described by a probability distribution which is either too or less informative than needed,

- paradoxes: the set of axioms (or calculus rules derived from them) lead to a counter-intuitive result.

As we will see in the next section, there are many situations involving uncertainty that cannot be described by a single probability measure. Once one has accepted this observation, then representation shortcomings are merely mismatches between the prerequisites of probability theory and the studied situation. Belief functions have a greater representation power than probabilities although they cannot cover either all possible situations with uncertain features. In particular, probabilities find it difficult to encode "don't know" messages while belief functions can.

Paradoxes are more problematic because they may imply that some axioms should be revised or that some operations should be forbidden. Note that, one should be careful with paradoxes in probabilities as they may be incurred by an initial mistake in the problem understanding and several paradoxes were solved without jeopardizing the axioms[8].

8. The Monty Hall paradox is solved by a careful (but standard) Bayesian analysis.

Most reported paradoxes focus on the everlasting debate between frequentists and Bayesians but, to the best of our knowledge, there is no paradox resulting in a real clash of axioms. The only paradox that we highlight in this subsection is **the wine/water paradox** which is actually featuring representation shortcoming regarding ignorance therefore I am a bit uncomfortable with use of the word paradox in this case.

Example 1. For a given volume of liquid, we know that the liquid is a (tragic) mix of wine and water. We know that the ratio X of wine to water is between $\frac{1}{3}$ and 3. What is the probability that $X < 2$?

The principle of indifference (PI) (see Keynes 1921) states that in the absence of information about some quantity X, we should assign equal probability density value to any possible value of X. Following PI, $X \sim \text{Unif}_{\left[\frac{1}{3};3\right]}$ and we have

$$P\left(X < 2\right) = \frac{2 - \frac{1}{3}}{3 - \frac{1}{3}}, \tag{1.8}$$

$$= \frac{5}{8} \tag{1.9}$$

Let $Y = 1/X$ denote the ratio of water to wine. It is also known with certainty that $Y \in \left[\frac{1}{3};3\right]$. Applying PI again, $Y \sim \text{Unif}_{\left[\frac{1}{3};3\right]}$ and we have

$$P\left(X < 2\right) = P\left(Y > \frac{1}{2}\right), \tag{1.10}$$

$$= \frac{3 - \frac{1}{2}}{3 - \frac{1}{3}}, \tag{1.11}$$

$$= \frac{15}{16} \neq \frac{5}{8}. \tag{1.12}$$

The wine/water example thus refutes the legitimacy of PI not axioms of probabilities. In fact, it is acknowledged by most probabilists that the uniform distribution is not uninformative and it is ill advised to use it in situations involving epistemic uncertainty while it can be perfectly justified for aleatory uncertainty. A more conservative probabilistic model consists in introducing a distribution on the set of laws of X as usually done in Bayesian nonparametrics but this idea is not unrelated to belief functions (see comments in 1.3.3).

1.2 Belief function theory

This section is an attempt to give an objective overview of the theory of belief functions. It contains the definitions of many concepts that will be further studied in other chapters.

1.2.1 *Dempster's system of lower and upper probabilities*

Although there are prior related works, Dempster 1967 is considered as the starting point of the vast and widely used theory of belief functions. In this article, Dempster does not use the terminology "belief function". This term was later coined by Shafer and conveys an interpretation of Dempster's ideas which was not present at this stage.

Dempster's paper focuses on the introduction of a mathematical way to propagate uncertainty from a probability space $(\Omega, \sigma_\Omega, \mu)$ to some other measurable space (Θ, σ_Θ) through a so called multi-valued mapping Γ.

Definition 6. A **multi-valued mapping** (or set-valued mapping) Γ is a mapping which has values in the power set of some space Θ.

$$\Gamma : \Omega \longrightarrow 2^\Theta$$

In Dempster 1967 and other references, a multi-valued mapping $\Gamma : \Omega \longrightarrow \Theta$ is an object inducing a mapping from Ω to 2^Θ. These two notions are undifferentiated in this monograph.

Obviously, a notion of measurability for the mapping Γ is needed to make the propagation operational. This starts with the introduction of pseudo-inverses of Γ.

Definition 7. The **upper inverse** of a multi-valued mapping $\Gamma : \Omega \longrightarrow 2^\Theta$ is defined as

$$\Gamma_\top^{-1}(B) = \{\omega \mid \Gamma(\omega) \cap B \neq \emptyset\}, \forall B \subseteq \Theta.$$

Definition 8. The **lower inverse** of a multi-valued mapping $\Gamma : \Omega \longrightarrow 2^\Theta$ is defined as

$$\Gamma_\perp^{-1}(B) = \{\omega \mid \Gamma(\omega) \subseteq B\}, \forall B \subseteq \Theta.$$

Equipped with these inverses, the definition of measurability in the context of multi-valued mappings is intuitively the following.

Definition 9. A multi-valued mapping Γ is **strongly measurable** [9] iff for any $B \in \sigma_\Theta, \Gamma_\top^{-1}(B) \in \sigma_\Omega$.

Obviously, if a strongly measurable multi-valued mapping is such that for any ω, $\Gamma(\omega)$ is a singleton, then Γ is a random variable. In the general case, Γ is formally equivalent to a random set but its interpretation is completely different from that of random sets which is usually a model for a random quantity whose actual value is a set.

Under the above measurability conditions and if $\mu \circ \Gamma_\top^{-1}(\Theta) \neq 0$, Dempster's system of lower and upper probabilities is obtained as:

9. When Θ is finite, strong measurability is equivalent to either:
- $\Gamma_\perp^{-1}(B) \in \sigma_\Omega, \forall B \in \sigma_\Theta$
- $\Gamma^{-1}(B) = \{\omega \mid \Gamma(\omega) = B\} \in \sigma_\Omega, \forall B \in \sigma_\Theta$

- $P_\top = \frac{1}{\mu\circ\Gamma_\top^{-1}(\Theta)} \times \mu \circ \Gamma_\top^{-1}$ (upper probability),
- $P_\perp = \frac{1}{\mu\circ\Gamma_\top^{-1}(\Theta)} \times \mu \circ \Gamma_\perp^{-1}$ (lower probability).

The terminology "lower and upper probabilities" is justified by the fact $P_\perp(B)$ is the minimal amount of probability mass that is transfered from μ to B while $P_\top(B)$ is the maximal amount of probability transfered from μ to B. These two set functions encompass many probability measures on Θ. This probability measure set (or **p.m.-set** for short) dominates[10] the lower probability function.

10. A measure P on Θ dominates a lower probability $P_\perp$ iff $P(B) \geq P_\perp(B)$, for all $B \in \sigma_\Theta$.

This system is entirely characterized by the quadruplet $(\Omega, \sigma_\Omega, \mu, \Gamma)$ which is called a **source**. The subsets $\Gamma(\omega)$ of Θ are called the **focal sets** of the source. An important remark is that the p.m.-set induced by a source is entirely characterized by either the function $P_\top$ or $P_\perp$ because

$$P_\top(B) = 1 - P_\perp(B^c), \forall B \subseteq \Theta. \tag{1.13}$$

B^c denotes the complement of set B in Θ.

Example 2. Suppose an urn contains 10 black or white marbles. The urn is partially spilled and we observe that it contains at least 4 black marbles and 2 white ones. Now let $\Theta = \{\text{white}; \text{black}\}$ denote the set of possible colors. Let Ω denote the set of marbles. Define a multi-valued mapping from Ω to 2^Θ as described in figure 1.1. Let also μ denote the uniform probability measure on Ω. By definition, we have that

$$P_\perp(\{\text{white}\}) = 0.2 \leq P(\text{white}) \leq P_\top(\{\text{white}\}) = 0.6, \tag{1.14}$$
$$P_\perp(\{\text{black}\}) = 0.4 \leq P(\text{black}) \leq P_\top(\{\text{black}\}) = 0.8. \tag{1.15}$$

The obtained bounds completely match the obvious conclusions that anyone would draw based on the observed data. Yet, this example might be misleading in the sense that upper and lower probabilities do not always encompass the actual target distribution to infer. This is dependent on the chosen model.

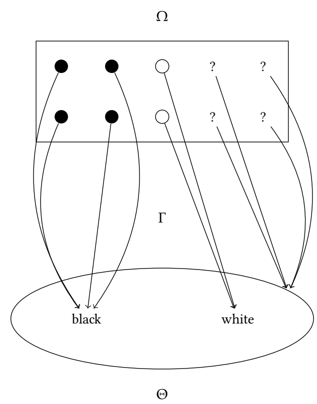


Figure 1.1: Multi-valued mapping example: 4 black marbles are revealed and mapped to $\{\text{black}\}$, 2 white ones are mapped to $\{\text{white}\}$. The remainder of the marbles are mapped to $\{\text{white}; \text{black}\} = \Theta$.

Example 3. (example 1 continued) Let $\Theta = [0; +\infty]$ and suppose Γ is a constant multi-valued mapping: $\Gamma_X(\omega) = \left[\frac{1}{3}; 3\right]$. In the wine/water paradox presented in example 1, X is a r.v. representing the ratio of wine to water and it is known with certainty that $X \in \left[\frac{1}{3}; 3\right]$. This information can be encoded by the source $(\Omega, \sigma_\Omega, \mu, \Gamma_X)$. Concerning the odds that $X < 2$, we obtain

$$P_{\perp,X}([0;2]) = 0, \tag{1.16}$$
$$P_{\top,X}([0;2]) = 1. \tag{1.17}$$

Now concerning $Y = 1/X$ the ratio of water to wine, we know that $Y \in \left[\frac{1}{3}; 3\right]$ as well. The same source can be used as an uncertainty model for Y. We can write

$$P_{\perp,Y}\left(\left[\frac{1}{2}; +\infty\right]\right) = 0 = P_{\perp,X}([0;2]), \tag{1.18}$$
$$P_{\top,Y}\left(\left[\frac{1}{2}; +\infty\right]\right) = 1 = P_{\top,X}([0;2]) \tag{1.19}$$

Consequently, contrary to the probabilistic modeling of this problem, there is no contradiction arising in the belief function framework by confronting the odds of a ratio and the odds of the inverse ratio.

We can also derive notions of lower and upper expectations through Riemann–Stieltjes integral.

Definition 10. For any measurable function $f : \Theta \longrightarrow \mathbb{R}$, we define

- $\mathbb{E}_{P_\top}[f] = \int_{-\infty}^{+\infty} v \, dP_\perp \left(f^{-1}([-\infty; v]) \right)$, **(upper expectation)**
- $\mathbb{E}_{P_\perp}[f] = \int_{-\infty}^{+\infty} v \, dP_\top \left(f^{-1}([-\infty; v]) \right)$, **(lower expectation)**

The interchange between upper and lower symbols is not intuitive but is illustrated as sidenote.[11] The sidenote also illustrates that upper and lower expectations are more general concepts than upper and lower probabilities.

11. Suppose that $f = 1_B$, the indicator function on subset $B \subseteq \Theta$. Let $(v_i)_{i=0}^{p}$ denote a series of scalar such that

$$-a = v_0 < v_1 < \ldots < v_p = a$$

with $a \in \mathbb{R}^+$. By definition of the Riemann–Stieltjes integral, we have

$$\mathbb{E}_{P_\top}[1_B] = \lim_{a \to +\infty} \lim_{\Delta \to 0} \sum_{i=1}^{p} \xi_i \left[P_\perp \left(f^{-1}([-\infty; v_i]) \right) - P_\perp \left(f^{-1}([-\infty; v_{i-1}]) \right) \right], \tag{1.20}$$

with $\xi_i \in [v_{i-1}; v_i]$ and $\Delta = \max_i v_i - v_{i-1}$.

When a is large enough there is only one interval $[v_{j-1}; v_j]$ in which the term under the sign symbol is not null and $1 \in [v_{j-1}; v_j]$. We obtain

$$\mathbb{E}_{P_\top}[1_B] = \lim_{\Delta \to 0} \xi_i \left[P_\perp \left(f^{-1}([-\infty; v_j]) \right) - P_\perp \left(f^{-1}([-\infty; v_{j-1}]) \right) \right], \tag{1.21}$$

Since $v_j > 1$, then

$$P_\perp \left(f^{-1}([-\infty; v_j]) \right) = 1$$

In addition, if one chooses $\xi_j = v_j$, as Δ gets smaller, then $\xi_j \to 1$, which gives

$$\mathbb{E}_\top[1_B] = \left[1 - P_\perp \left(f^{-1}([-\infty; v_{j-1}]) \right) \right], \tag{1.22}$$

$$= P_\top \left(f^{-1}([-\infty; v_{j-1}]) \right). \tag{1.23}$$

$$= P_\top(B). \tag{1.24}$$

Dempster also introduces in his paper a number of operations on lower and upper probabilities starting with conditioning. **Dempster's conditioning** on B consists in considering the multi-valued mapping $\Gamma \cap B$. This happens to take a usual form for upper probabilities:

$$P_\top(A \cap B) = P_\top(A|B) \, P_\top(B) \tag{1.25}$$

The corresponding result for lower probabilities does not mimic Bayes rule and is omitted.

Another interesting idea proposed by Dempster is combination via a mechanism nowadays known as **Dempster's rule**. Suppose we have two sources $S_1 = (\Omega_1, \sigma_1, \mu_1, \Gamma_1)$ and $S_2 = (\Omega_2, \sigma_2, \mu_2, \Gamma_2)$. These two sources can be combined under Dempster's rule into a single one as follows:

$$S_{1\oplus 2} = (\Omega_1 \times \Omega_2, \sigma_1 \times \sigma_2, \mu_1 \otimes \mu_2, \Gamma_{12}), \tag{1.26}$$

where we have $\Gamma_{12}(A) = \Gamma_1(A) \cap \Gamma_2(A)$. Because we use the product measure $\mu_1 \otimes \mu_2$, independence between each event in σ_1 with respect to those in σ_2 is a prerequisite to use the rule. The upper and lower probabilities induced by $S_{1\oplus 2}$ are denoted by $P_{\top,1\oplus 2}$ and $P_{\perp,1\oplus 2}$ respectively, the symbol $\oplus$ being the notation for Dempster's rule. The combination operation translates into a rather complicated expression except for upper probabilities of singletons. In particular, for each singleton $\{a\}$, we have

$$P_{\top,1\oplus 2}(a) \propto P_{\top,1}(a) \, P_{\top,2}(a), \forall a \in \Theta. \tag{1.27}$$

We will use interchangeably $P_{\top,1}(\theta)$ or $P_{\top,1}(\{\theta\})$ for any set function when dealing with singletons.

The multiplicative constant involved in the above equation stems from those probabilities that are transferred to $\emptyset$ through Γ_{12} and imply a renormalization.

When Γ_2 is a constant multi-valued mapping, Dempster's conditioning is recovered therefore Dempster's rule is often regarded as a generalization of conditioning.

Dempster also introduces a third set function for which Dempster's combination is easy to compute. This relies on a third definition of pseudo-inverse of multi-valued mappings:

$$\tilde{\Gamma}^{-1}(B) = \{\omega \mid \Gamma(\omega) \subseteq B\}, \forall B \supseteq \Theta. \tag{1.28}$$

The renormalized push forward measure $q = \frac{1}{\mu \circ \Gamma_{\top}^{-1}(\Theta)} \times \mu \circ \tilde{\Gamma}^{-1}$ is called **commonality** function. The combination of two commonality functions q_1 and q_2 induced by the sources defined above writes

$$q_{1 \oplus 2} = q_1 \odot q_2 \text{ (entrywise product).} \tag{1.29}$$

The entrywise product is also often called Hadamard product.

1.2.2 *Shafer's mathematical theory of evidence*

In his famous essay, Shafer 1976 follows a different path as Dempster and starts by formally defining belief functions and then investigate to what extent such objects are adapted to quantify degrees of belief of events and achieve inference on variables of interest. This means that he does not build belief functions upon a source but instead proposes belief functions as an alternative representation of uncertainty as compared to probabilities. In his work, Shafer also shares a vision of how belief functions should be interpreted as numbers quantifying uncertainty and what are the connections to probabilities. He also introduces many instrumental calculus rules for belief functions that are valid tools beyond the scope of his interpretation and introduction of belief functions.

In this subsection, we give a brief sketch of his pioneering work and try to outline the differences with Dempster's system of lower and upper probabilities. The analysis starts with the following definition of a belief function.

Definition 11. Let (Θ, σ_Θ) denote some measurable space. A **belief function** is a mapping $bel : \sigma_\Theta \to [0; 1]$ verifying the three following conditions

(i) $bel(\emptyset) = 0$,

(ii) $bel(\Theta) = 1$,

(iii) For any $k \geq 2$ and any collection of events $B_1, \ldots, B_k$ in σ_Θ, .

$$bel\left(\bigcup_{i=1}^{k} B_i\right) \geq \sum_{\substack{s \subseteq \{1;\ldots;k\} \\ s \neq \emptyset}} (-1)^{|s|+1}\, bel\left(\bigcap_{j \in s} B_j\right). \tag{1.30}$$

When $k = 2$ and $B_1 \subseteq B_2$, condition (iii) implies $bel(B_1) \leq bel(B_2)$, a property known as **monoticity**. More comments on this are given in 1.3.2

When inequalities in condition (iii) are equalities, then a belief function is formally equivalent to a probability measure and (1.30) is known as the inclusion-exclusion principle.

Another equivalent representation is given by plausibility functions whose definition is the following.

Definition 12. Let (Θ, σ_Θ) denote some measurable space. A **plausibility function** is a mapping $pl : \sigma_\Theta \to [0; 1]$ verifying the three following conditions

(i) $pl(\emptyset) = 0$,

(ii) $pl(\Theta) = 1$,

(iii) For any $k \geq 2$ and any collection of events $B_1, \ldots, B_k$ in σ_Θ, .

$$pl\left(\bigcap_{i=1}^{k} B_i\right) \leq \sum_{\substack{s \subseteq \{1;\ldots;k\} \\ s \neq \emptyset}} (-1)^{|s|+1}\, pl\left(\bigcup_{j \in s} B_j\right). \tag{1.31}$$

Similarly, when $k = 2$ and $B_1 \subseteq B_2$, condition (iii) implies $pl(B_1) \leq pl(B_2)$

From the above definition, it clear that, like for lower and upper probabilities, we have

$$bel\left(B\right) = 1 - pl\left(B^c\right), \forall B \in \sigma_\Theta. \tag{1.32}$$

Before reviewing Shafer's argumentation, we immediately give an important result (from Nguyen 1978) which explains the link between belief and plausibility functions with lower and upper probabilities.

Theorem 1. *Let $P_\perp$ and $P_\top$ denote respectively the lower and upper probabilities induced by the source $(\Omega, \sigma_\Omega, \mu, \Gamma)$. Then $P_\perp$ is a belief function and $P_\top$ is a plausibility function.*

In his seminal work, Shafer 1973 proves a representation theorem that states that for any belief function, one can build a probability space and a lower inverse that are a source for the belief function[12]. To avoid duplicating nomenclatures, we will now use the terms belief and plausibility functions in place of Dempster's lower and upper probabilities.

Going back to the justification of belief functions as appropriate uncertainty model, Shafer argues that the rules belief function must obey are attractive because they allow to commit a portion of belief to some subset of Θ (as in probability theory) without implying that the remainder is committed to its negation (unlike probability theory). In many situations, the collected evidence can only translate into supporting subset B (nothing more, nothing less) up to a given degree in $[0;1]$.

12. More precisely, Shafer shows that for any belief function, one can construct a probability space and a probability allocation mapping that induce the belief function. The proposed probability allocation mapping in the theorem proof matches the definition of a multi-valued mapping lower inverse.

Example 4. (example 2 continued) In the urn example, four black marbles were revealed. In light of this piece of evidence, we assigned a 0.4 degree of belief to the singleton {black}. But it would not make sense to assign $1 - 0.4 = 0.6$ degree of belief to {white}! This is not possible because there is missing information.

The baseline statistical approach would be to compute maximum likelihood estimates of the ball proportions. One thus obtains $\hat{P}\left(\text{black}\right) = \frac{2}{3}$ and $\hat{P}\left(\text{white}\right) = \frac{1}{3}$. Without denying the relevance of this frequentist approach, it is obvious that accepting these estimates as our beliefs on ball proportions is not justified based on observed data solely but also relies on additional assumptions. In particular, one assumes that the sample size is large enough to allow the discrepancies between the estimates and the expected values to be below a predefined appropriate threshold[13].

13. This idea is the one carried by Hoeffding concentration inequalities.

Shafer also argues that belief functions cannot grasp randomness but only subjective uncertainty. For instance, take a dice throw experiment. If the frequentist probabilities are known then one will adopt these frequencies as beliefs on the dice thrown outcome. However, if the frequencies are unknown, then it makes sense to adopt personal beliefs based on some available evidence and these degrees of belief are very unlikely to equate the frequencies.

Epistemic uncertainty is yet very often encoded using probability measures in the Bayesian framework and therefore they obey the additivity property while belief functions in general do not. In Shafer's view, these subjective probabilities are just a subclass of belief functions and Bayesian reasoning is compatible with the proposed framework. So belief functions are just a wider class of models in which degrees of belief may not be additive.

Another idea that is behind Shafer's approach is to derive a framework in which the degrees of belief can be constructed incrementally by application of Dempster's rule to elementary beliefs. Since Dempster's rule is operational only for belief functions, the appeal of the definition 11 is obvious. This ambition was partially achieved as Shafer proved that a large class of belief functions[14] can be decomposed in this fashion but the result does not hold in general.

14. These are called separable belief functions.

Around ten years ago, Dempster 2008 published a paper to revisit his and Shafer's ideas to present them in a more amenable way for statisticians. Dempster reaffirms his conviction that belief function calculus is a powerful framework for inference based on incomplete and uncertain evidence. He also sheds light on the fact that the framework can be regarded as a construction based on three-valued logic. In addition to usual logical state "true" and "false", a third state "unknown" is considered. Note that these are not logical states since the third one is the logical disjunction of the two first and consequently they are epistemic states. Stricto sensu, we ought to define the states as "known to be true", "known to be false" and "unknown". Belief functions assign non-negative probabilities (u, v, w) to each such state for each member of the field σ_Θ such that $u + v + w = 1$. The rules given in definition 11 are just constraining triplets to take values matching principled allocation of degrees of belief. In particular, if a set B is assigned the triplet (u, v, w) then obviously B^c must be assigned (v, u, w). In fact, since for any $C \subseteq B$, C being "true" implies B is also "true", then $u = bel(B)$. It follows that $v = bel(B^c)$ and $w = 1 - bel(B) - bel(B^c)$[15]. Example 2 is also a relevant illustration of this machinery. Dempster concludes that belief functions are ordinary textbook probabilities allowing to assign non zero probabilities to "unknown". Similar justifications are given in Dubois et al. 1996.

15. The belief function maps each member B of the σ-field to their u-probabilities. The plausibility function can also be defined in the same fashion by mapping each B to probabilities $1 - v$, i.e. the probability mass that is not committed against B. So we see that our ignorance on the truth (or falsity) of B is featured by $pl(B) - bel(B) = w$.
We can also remark that for any $B' \subseteq B$, if B' is true then it logically implies that B is also true and it follows that $bel(B') \leq bel(B)$. So we might be willing to evaluate to what extent each B' contributes to the u-probability of B. We can do that by introducing the mass function which decomposes the belief function is in this way :

$$bel(B) = \sum_{B' \subseteq B} m(B').$$

Then, we see that $bel(\Theta) = 1$ implies that $\sum_{B \subseteq \Theta} m(B) = 1$. We have also that $bel(\{a\}) = m(\{a\})$. Since bel is ∞-monotonic (1.30), by induction, we get that $m(B) \geq 0$ for any $B \subseteq \Theta$. Finally, these results entirely specifies a mass function which also specifies the belief or plausibility functions since each of these functions is in bijective correspondence with any other.

1.2.3 *Random sets*

Define the (usual) inverse of a multi-valued mapping Γ as $\Gamma^{-1}(B) = \{\omega \mid \Gamma(\omega) = B\}$. If the preimage through Γ^{-1} of any element in σ_Θ is an element of σ_Ω, then Γ abides by the (usual) definition of measurability and Γ is formally equivalent to a random set[16]. The push-forward measure $\frac{1}{\mu \circ \Gamma^{-1}(\Theta)} \times \mu \circ \Gamma^{-1}$ is the distribution of this random set. This result was proved by Nguyen 1978. He also shows that in the finite case, the distribution of the random set is the **mass function**[17].

16. Roughly speaking, a random set is a random variable whose values are sets.

17. See 2.1 for the definition of those functions. In the present chapter, a mass function can be merely regarded the set function $m = \frac{1}{\mu \circ \Gamma^{-1}(\Theta)} \times \mu \circ \Gamma^{-1}$ induced by a source.

In spite of this equivalence, there is little connection between random sets in the sense of Mathéron 1975 and Kendall 1974 and belief functions as they encode very different kinds of information. Indeed, when performing inference in the random set setting, the parameter of interest is a set while in the belief function setting it remains point-valued.

Example 5. Suppose you want to infer the spoken languages of a randomly picked person on earth: $\Theta_1 = \{\text{English}; \text{French}; \text{Spanish}\}$. Your observed data is that someone speaks English with probability $\frac{1}{2}$, French with probability $\frac{1}{10}$ and Spanish with probability $\frac{1}{3}$. It is possible to infer the spoken languages from this basic information but the end result is a distribution of subsets of Θ_1. By making an independence assumption (just for illustration), we would deduce for example that the probability that someone speaks English and French solely is:

$$\mu\left(\Gamma = \{\text{English}; \text{French}\}\right) = \frac{1}{2} \times \frac{1}{10} \times \left(1 - \frac{1}{3}\right)$$

Suppose now you want to infer the birth country of a randomly picked person on earth: $\Theta_2 = \{\text{UK}; \text{France}; \text{Spain}\}$. Now the variable of interest takes only one value at most. Based on the remark that someone speaks the home language of his birth country with probability 0.9, then we could combine all information by deconditioning or vacuously extending[18] the probability distributions to belief functions on the product space $\Theta_1 \times \Theta_2$. After marginalizing on Θ_2, the end result is this time a belief function.

18. These operations are presented in 2.4.

Complements on this question are given in Couso and Dubois 2014.

1.2.4 Random codes

A few years after publishing his book, Shafer 1981 offered a new elegant interpretation of belief functions as random codes for partial knowledge. In this interpretation, any element ω in the probability space $(\Omega, \sigma_\Omega, \mu)$ is a code representing some imprecise information about θ of the form $\{\theta \in B\}$. Exactly one of the codes is selected at random. The chance that a code ω is selected is known and is given by $\mu(\omega)$. It follows that $\mu \circ \Gamma^{-1}(B)$ is the probability that the message is $\{\theta \in B\}$ (and nothing less, nothing more).

1.2.5 Smets' transferable belief model

In an attempt to clarify the benefits of belief functions for reasoning under uncertainty, Smets and Kennes 1994 introduced the **Transferable Belief Model** (TBM). This model shares Shafer's core idea: the allocation of probability mass only to some subset B as justified by evidence without implying any support to B^c. However, the TBM distinguishes itself from Shafer's theory of evidence regarding the following essential aspects:

- a two-stage reasoning: subjective beliefs are first constructed from accumulated evidence and the quantification of the beliefs yield a belief function (credal step). Second, a probability distribution is computed from the belief function that best qualifies for an operational decision making process (pignistic[19] step). Indeed, probability distributions are instrumental to take rational decisions with minimal expected loss.
- the relaxation of the null mass requirement for $\emptyset$. In this case, we no longer have $bel(B) = 1 - pl(B^c)$ but instead $bel(B) = 1 - pl(B^c) - \kappa$ with $\kappa \in [0; 1]$. Smets interprets κ as the support given to the possibility that the true value of the variable of interest does not lie within Θ. He calls this relaxation the "open world assumption". In contrast, upholding the empty set positive mass ban is the "closed world assumption". Note that, *in fine*, the empty set mass is eliminated[20] when

19. From *pignus* ("bet").

20. The mass of the empty set cannot be eliminated if it has accumulated maximal support just like upper and lower probabilities are not defined in this case.

computing the pignistic distribution and therefore the closed world assumption is resilient. In this monograph, the mass assigned to the empty set is considered more as an algebraic convenience rather than a feature of the open world assumption.

- an axiomatic derivation of Dempster's rule of combination. This derivation is presented in details in Smets 1990 and is built upon previous works (Klawonn and Schwecke 1992; Klawonn and Smets 1992).

Smets is also a promoter of the **least commitment principle** (LCP) which states that when several belief functions qualify as uncertainty models that are consistent with evidence, then one should select a belief function with minimal degrees of belief. The notion of commitment of a belief function anticipates the definition of partial orders for belief functions that will be reviewed in chapter 3.

Another feature of the TBM claimed by Smets is that the model (at least at the credal stage) is not built upon pre-existing probability space, random variable or what so ever. More precisely, the pre-existence of a probabilistic model is not *necessary* which does not mean that building belief functions in this way is wrong. Recently, there has been a renewed interest in belief functions induced by mechanisms that are highly coupled with probability theory (Martin et al. 2010; Denœux 2014; Kanjanatarakul et al. 2014; Xu et al. 2015). My personal thoughts on this point is that belief function theory makes sense on its own as an uncertainty theory but since probability theory has been extensively studied it is desirable to explain belief functions as a construction of simpler objects with which we are familiar. It is also a mean to drag the attention of new researchers by showing that belief functions are no academic extravagance but rely on sound and simple probabilistic arguments.

1.2.6 *Fiducial inference*

Suppose the observed data can be explained by the following sampling model consisting in

(i) a φ-equation relating a data point X, the parameter θ to be inferred and a pivotal quantity U as

$$X = \varphi(\theta, U) \tag{1.33}$$

(ii) a probability measure μ defined on measurable subsets of space $\mathbb{U}$ where U takes its values.

Given this model, suppose also that observing two quantities out of the triplet (X, θ, U) uniquely determines the third one. In the fiducial[21] inference framework, X is a single sufficient statistic for the single parameter θ (a very strong prerequisite). The strength of the fiducial argument is that the uncertainty on X prior to sampling is transferred to θ after. To illustrate these concepts in action we give an example drawn from Martin et al. 2010.

21. from *fiducia* ("trust")

Example 6. Suppose θ is the first moment of a normal distribution with unit variance. Let $U \sim \text{Unif}_{[0;1]}$ and Φ denote the cdf of a centered reduced normal distribution. Finally, define the φ-equation as

$$X = \theta + \Phi^{-1}(U) \tag{1.34}$$

This procedure is actually the usual way to build many random generators. As observed by Fisher 1930, events $\{\theta \leq a\}$ and $\{U \geq \Phi(X-a)\}$ are equivalent. The **fiducial probability** of $\{\theta \leq a\}$ is thus $\Phi(X-a)$, meaning that $\theta \sim \mathcal{N}(X,1)$.

The fiducial distribution can be deemed as a form of prior-free posterior distribution. In the above example, the posterior distribution obtained from Bayesian analysis yields the same result when choosing Jeffreys flat prior.

Belief functions are a successor of fiducial inference in the sense that the specification of a prior distribution is unnecessary (but still possible if wanted). Belief functions do not resort to sufficient statistics and are not limited to inference for a scalar parameter. Indeed, a key to the generalization of fiducial inference is precisely to relax the uniqueness of θ given (X,U) and the uniqueness of U given (X,θ).

Define a multi-valued mapping $\Gamma : \mathbb{U} \to 2^{\mathbb{X}\times\Theta}$ such that

$$\Gamma(u) = \{(x,\theta) \in \mathbb{X}\times\Theta \mid X = \varphi(\theta,u)\}. \tag{1.35}$$

Using the above mapping Γ, we can define a source $(\mathbb{U},\sigma_{\mathbb{U}},\mu,\Gamma)$. The inference mechanism relies mainly on conditioning with respect to events $\{X=x\}$ or $\{\theta=a\}$.

Conditioning on θ

One can define a multi-valued mapping featuring the piece of information $\theta = a$:

$$\Gamma_a = \{(x,\theta) \in \mathbb{X}\times\Theta \mid \theta = a\}. \tag{1.36}$$

The expression is not dependent on the φ-equation or on u and the multi-valued mapping is thus constant (w.r.t. u). This constant set is just made of all pairs (x,a) that are elements of the product space $\mathbb{X}\times\Theta$. Intersecting mappings Γ and Γ_a yields a multi-valued mapping whose images are subsets of the a-cross section of $\mathbb{X}\times\Theta$. It can be projected[22] down to $2^{\mathbb{X}}$ which gives

$$\mathrm{proj}_{\mathbb{X}}(\Gamma(u)\cap\Gamma_a) = \{x \in \mathbb{X} \mid X = \varphi(a,u)\}, \tag{1.37}$$

$$= \{\varphi(a,u)\} \tag{1.38}$$

Since $\mathrm{proj}_{\mathbb{X}} \circ (\Gamma\cap\Gamma_a)$ is a point-valued mapping, the corresponding belief and plausibility functions coincide and are the sampling distribution $P_{X|\theta=a}$. This result is compliant with the fiducial model.

22. Projecting in this context is understood as dropping one of the components of pairs in the product space, since the mapping is constant over the other component.

Conditioning on X

Likewise, one can define a multi-valued mapping featuring the piece of information $X = x$:

$$\Gamma_x = \{(x,\theta) \in \mathbb{X}\times\Theta \mid X = x\}. \tag{1.39}$$

Taking intersection with Γ and projecting on x-cross sections gives the following multi-valued mapping

$$\mathrm{proj}_{\Theta}(\Gamma(u)\cap\Gamma_x) = \{a \in \Theta \mid x = \varphi(a,u)\}. \tag{1.40}$$

This mapping is, in general, set-valued and the corresponding belief and plausibility functions no longer coincide. This posterior belief function $bel_{\theta|x}$ allows us to evaluate to what extent the observed data imply some event $\{\theta \in B\}$ and it is therefore the result of the inference procedure.

Example 7. (example 6 continued) In this example, we illustrate that the belief function analysis delivers compliant results with the fiducial analysis. Suppose $X = x$ is the only observation we have to determine the mean value of a reduced normal distribution. Following the above procedure, the posterior belief function is

$$bel_{\theta|x}(B) = \mu\left(\{u \mid \mathrm{proj}_{\Theta}(\Gamma(u) \cap \Gamma_x) \subseteq B\}\right) \tag{1.41}$$

Since the φ-equation can be reversed, $X = x$ implies $\theta = x - \Phi^{-1}(u)$ and consequently, $\mathrm{proj}_{\Theta}(\Gamma(u) \cap \Gamma_x)$ are singletons. The belief function writes

$$bel_{\theta|x}(B) = \mu\left(\left\{u \mid x - \Phi^{-1}(u) \in B\right\}\right) \tag{1.42}$$

The probability that $\Phi^{-1}(U)$ falls in $x - B$ is the same as the probability that an $\mathcal{N}(x, 1)$ distributed random variable falls in B therefore we obtain the same result as in example 6.

Example 8. We provide another example to illustrate that the belief function based procedure does not necessarily resort to a single sufficient statistic. This example is given in Dempster 1966 and rephrased in Martin et al. 2010.

Let $\left\{x^{(1)}, \ldots, x^{(r)}\right\}$ denote iid samples drawn from $\mathrm{Ber}(\theta)$. Define the following φ-equation:

$$X^{(i)} = 1_{U_i \leq \theta}, \tag{1.43}$$

and $U_i \sim \mathrm{Unif}_{[0;1]}$. Taking the fiducial belief function analysis on the product space $\mathbb{X}^r \times \Theta$, a posterior belief function is obtained. Dempster shows that a closed form expression of this function for events $\{\theta \leq a\}$ can be obtained by reasoning on the random variable N_X which is the number of successes in the r Bernoulli trials.

1.2.7 *Shortcomings of belief functions*

The theory of belief functions, as any other uncertainty theory, is not exempt of limitations and it may lead to questionable results. Pearl 1990 conveyed a sharp analysis of belief functions and raised debates around several aspects of the theory:

- **Interpretation**: Pearl proposes a new interpretation of belief functions as *probabilities of provability*. He explains that the mass function can be regarded as the probability distribution of logical theories (a set of axioms). The degree $bel(A)$ is then the sum of probabilities of those theories from which A follows as logical consequence.
  Pearl argues that there are unfortunately only a few situations in which one wishes to infer such probabilities. Without denying the relevance of this interpretation[23], it does not imply that other interpretations are not equally valid.

23. This interpretation was sustained by Kohlas and Monney 1995 in their mathematical theory of hints which relies on the same probabilistic mechanisms as those introduced by Dempster 1967. Kohlas and Monney 2008 also use this interpretation to perform fiducial inference.

- **Ability to represent partial knowledge**: Pearl lists a number of examples when available evidence is insufficient to infer precise probabilities but instead probability bounds. In these examples, these bounds cannot be encoded by a belief function. Dempster 1967 had already shown that upper and lower probabilities as defined in his article are in correspondence with a subclass of p.m.-sets and there is no surprise in finding situations where belief functions do not fit prescribed probability bounds. Every uncertainty theory has a limited representation power and, in principle, we ought to choose the least general framework in which there is feasible representation for the situation we wish to analyze.
- **Updating**: Pearl also rejects belief functions as valid lower probabilities because they cannot be updated through Dempster's conditioning in any circumstances. Actually, he shows that in the "three prisoners problem" [24] Dempster's conditioning fails to retrieve the appropriate p.m.-set and instead returns a single probability distribution which does not account for some of the aspects of the toy example. This explained by the fact that we start with an initial uniform distribution and Dempster's rule cannot revise a single distribution into a set of distributions. Consequently, this is more a critic of Dempster's rule to combine upper or lower probabilities than a critic of the theory to have appropriate foundations for upper and lower probabilistic calculus. Shafer 1992 argues that Dempster's rule cannot be applied in the "three prisoners problem" because the independence assumption is not verified. Indeed, the event on which the conditioning is performed is constrained and the probability spaces inducing the belief functions are related. The aleatory choice of the jailer cannot be made without knowledge of the sentenced prisoner. Pearl 1992 is not really convinced by this argument, but at least, his example highlights that the dependent source pitfall of Dempster's rule is not easy to avoid and even Dempster's conditioning cannot be used blindly.
  Another form of conditioning was proposed by Halpern and Fagin 1992: for any $A \subseteq \Theta$ and any $B \subseteq \Theta$ such that $bel\left(B\right) > 0$ where bel is a belief function on Θ inducing the p.m.-set $\mathcal{P}$, then the conditional plausibility function is given by
$$pl\left(A|B\right) = \sup_{P \in \mathcal{P}_\Theta} P\left(A|B\right), \tag{1.47}$$
$$= \frac{pl\left(A \cap B\right)}{pl\left(A \cap B\right) + bel\left(A^c \cap B\right)}, \tag{1.48}$$
  and the conditional belief function by
$$bel\left(A|B\right) = \inf_{P \in \mathcal{P}_\Theta} P\left(A|B\right), \tag{1.49}$$
$$= \frac{bel\left(A \cap B\right)}{bel\left(A \cap B\right) + pl\left(A^c \cap B\right)}. \tag{1.50}$$
  By construction, this alternative conditioning offers a correct treatment of the three prisoners problem.
  Dubois et al. 1996 advocate that two different intellectual processes must be distinguished when performing beliefs update: focusing and

24. Among three prisoners $\{a_1, a_2, a_3\}$, one of them is randomly selected for execution using a uniform distribution. Prisoner a_1 convinces a jailer to reveal if either a_2 or a_3 are to be spared. The jailer is forbidden to reveal to a_1 what will happen for him but this information seems neutral to the jailer and he tells a_1 that a_2 will be spared. By carrying a careful probabilistic analysis, the jailer is correct and the probability that a_1 will be killed remains $\frac{1}{3}$ after conditioning on the newly acquired information. This example is reminiscent of the Monty Hall "paradox".
The lower/upper probability version of the problem is obtained by saying that the probability that the jailer gives the name a_2 or a_3 given that a_1 will be killed is unknown. Let λ denote the probability he names a_2 in the event that a_1 will be killed. We have
$$P\left(a_1|a_2 \text{ spared}\right) = \frac{P\left(a_2 \text{ spared}|a_1\right) P\left(a_1\right)}{\sum_{i=1}^{3} P\left(a_2 \text{ spared}|a_i\right) P\left(a_i\right)}, \tag{1.44}$$
$$= \frac{P\left(a_2 \text{ spared}|a_1\right)}{P\left(a_2 \text{ spared}|a_1\right) + P\left(a_2 \text{ spared}|a_3\right)}, \tag{1.45}$$
$$= \frac{\lambda}{1+\lambda}. \tag{1.46}$$
We deduce that
$$0 \leq P\left(a_1|a_2 \text{ spared}\right) \leq \frac{1}{2}.$$
Also, if $\lambda = \frac{1}{2}$ as in the precise version of the 3 prisoners puzzle, we obtain $P\left(a_1|a_2 \text{ spared}\right) = \frac{1}{3}$.

$\mathcal{P}_\Theta$ denotes the probability distribution **simplex** for Θ.

revision. Focusing consists in integrating a factual case specific piece of information which is the most frequent use case of Bayes rule. Revision consists in substituting population wise prior beliefs by new ones. So focusing is a selection process while revision is a correction one and they translate into different mathematical formulations. The authors explain that Dempster's rule is well adapted to revision while Halpern and Fagin's conditioning should be preferred for focusing.
It is hard to tell if the 3 prisoners problem translates in a focusing or revision but in the frequentist version of the problem, i.e. prisoners are sampled with equal probabilities and when the sentenced prisoner is a_1 the jailer samples from Ber (λ), then we are in the focusing framework and belief function and Bayesian solutions agree. This result can be checked by numerical simulation. If the marginal distribution of prisoners are subjective personal probabilities of prisoner a_1, then there is no ground truth as to how probability masses are redistributed after learning that a_2 is spared. Dempster's rule selects the distribution with highest entropy ($P(a_1|a_2 \text{ spared}) = \frac{1}{2}$) whereas the Bayesian solution is unchanged.

- **Evidence pooling**: Pearl acknowledges that belief functions have some merits in evidence pooling provided that each belief function summarizing a piece of evidence is combined with a prior probability distribution over the variable of interest thereby yielding a posterior distribution that is readily operational for decision making. Most of the time, the same information as the one contained in such belief functions can be encoded in likelihood functions which are computationally more easily tractable. There are however, situations in which the evidence is too imprecise to be easily translated into a likelihood function and then a belief function becomes instrumental.

Another criticism addressed to belief functions is that, in general[25], they do not have long-run frequency properties. Without this property, the statistical conclusions that can be drawn from a belief function based analysis must be carefully examined.

To overcome this difficulty, one can also build specific inference procedures where the property holds. For instance, Martin et al. 2010 use the fiducial inference based procedure and replace the pivotal measure μ with a belief function. They thus derive a less committed[26] belief function that achieves long-run frequency properties. Building upon this procedure they can safely introduce exact high-dimensional hypothesis tests.

Similarly as in the fiducial belief function analysis, Kanjanatarakul et al. 2016 use an inference model relying on a φ-equation but the end result is the predictive belief function on X induced by a likelihood-based belief function on θ. The action of the observed data on our beliefs on future occurrences of X is thus conveyed by the likelihood function solely. The proposed estimator is consistent in the sense that the predictive belief function tends to the true distribution of X as the dataset size tends to infinity.

On more practical grounds, a notorious limitation of belief function is the exponential complexity in the size n of Θ. By accepting to lose additivity, then degrees of belief cannot be summarized by means of a

25. Since belief functions are epistemic random sets they inherit the long-run frequency properties of these latter. Anyway, most of the time, belief functions encode epistemic uncertainty where there is no notion of multiple runs, since the event is either true or false but cannot be repeated.

26. This means that the p.m.-set of this function is a superset of the p.m.-set of the other one. This notion will be more formally defined in chapter 3.

distribution anymore (no free lunch). To avoid this difficulty, most of belief function practitioners use models where belief functions have a limited number of focal elements. When many variables are studied, the graphical model of Shenoy and Shafer 2008 helps to avoid unnecessary complexities when propagating beliefs.

Concerning, the specific case of continuous variables, plausibilities values may be vanishing. This calls for a notion of density function for plausibilities that is, to the best of our knowledge, only available for belief functions on Borel intervals of the real line (Smets 2005). Note that this is more an open problem than a shortcoming.

1.3 Other related uncertainty models

There are other uncertainty and partial ignorance models than probabilities or belief functions. For example, Cantor's set theory can be regarded as a model of imprecision. In this section we present three other frameworks for reasoning under uncertainty two of which are more general than belief functions and the other one being formally encompassed by belief functions.

1.3.1 *Possibility theory*

The presentation of possibility theory displayed in this subsection is inspired from Dubois et al. 1996.

Suppose a model where one sorts subsets $B \subseteq \Theta$ into three families:

- $\mathcal{T}$ the family of true propositions,
- $\mathcal{F}$ the family of false propositions,
- and $\mathcal{U}$ the family of undecided propositions.

Define two functions N and Π that respectively represent the certainty of truth and the possibility of truth. More formally,

$$\mathrm{N}(B) = 1_{\mathcal{T}}(B), \tag{1.51}$$

$$\Pi(B) = 1_{\mathcal{T}\cup\mathcal{U}}(B) \tag{1.52}$$

Take two events B_1 and B_2. If they are both certain then so is $B_1 \cap B_2$ and if at least of one of them is not certain then $B_1 \cap B_2$ is not certain either. We thus have

$$\mathrm{N}(B_1 \cap B_2) = \min\{\mathrm{N}(B_1); \mathrm{N}(B_2)\}. \tag{1.53}$$

Conversely, if at least either B_1 or B_2 is possible then so is $B_1 \cup B_2$, hence

$$\Pi(B_1 \cup B_2) = \max\{\Pi(B_1); \Pi(B_2)\}. \tag{1.54}$$

Moreover, if B is impossible then B^c is surely true, hence

$$\Pi(B) = 1 - \mathrm{N}(B^c). \tag{1.55}$$

Now, by preserving the above logical rules but allowing N and Π to take values in $[0;1]$ and to represent graded membership of families $\mathcal{T}$ and $\mathcal{T} \cup \mathcal{U}$, then we obtain a framework called possibility theory introduced by Zadeh 1978 and further elaborated by Dubois and Prade 1988.

In this framework, the function N is called necessity measure and the function Π is called possibility measure. It is easily checked that consonant belief functions[27] are formally equivalent to necessity measures. An important appealing aspect of possibility measures is that they share the same memory complexity as probability distributions. Indeed, functions N and Π can be retrieved from the possibility distribution[28] $\pi(a) = \Pi(\{a\})$.

27. See 2.2.4 for a definition of this subclass of belief functions.

28. The possibility distribution is the contour function (see 2.2.4) in the belief function framework.

1.3.2 *Imprecise probabilities*

We have already mentioned that many situations in which probability bounds are defined cannot be encoded into a belief function (like the examples given in Pearl 1990 to outline the limits of the representation power of belief functions). By relaxing the 3$^{\text{rd}}$ axiom in the definition of belief functions, we obtain objects that can encode any probability bound system and are most often called **imprecise probabilities** or probability intervals.

A pivotal notion in the imprecise probability framework is the concept of Choquet capacity[29]. We first present rapidly this concept and related notions before showing that imprecise probabilities formally encompass belief functions although in practice this will be dependent on the chosen interpretation of belief functions.

29. The term capacity is explained by the fact that Choquet introduced these objects while working on models for electrical charge of capacitors.

Definition 13. Let ν denote a set-function from 2^Θ to $\mathbb{R}$. ν is said to be a **capacity** if it has the following properties:

- $\nu(\emptyset) = 0$,
- $\nu(\Theta) = 1$,
- $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$, for any A, B in 2^Θ (monotony).

We immediately remark that any belief or plausibility function is a capacity while the converse is not true. The discrepancy between the two concepts is featured by the following property.

Definition 14. A capacity $\nu : 2^\Theta \to [0;1]$ is said to be **k-monotonic**, with $k \in \mathbb{N}^*$, if and only if for any family of k events $\mathcal{A} = (A_i)_{i=1}^k$, one has:

$$\sum_{I \subseteq \mathcal{A}} (-1)^{|I|} \nu\left(\bigcap_{A \in I} A\right) \leq \nu\left(\bigcup_{1 \leq i \leq k} A_i\right). \tag{1.56}$$

Now, we understand that belief functions are ∞-monotonic capacities and that imprecise probabilities have an intrinsic much bigger representation power.

Other instrumental definitions are:

- **super-additivity**: $\nu(A) + \nu(B) \leq \nu(A \cup B), \forall A, B \subseteq \Theta$ s.t. $A \cap B = \emptyset$,
- **sub-additivity**: $\nu(A) + \nu(B) \geq \nu(A \cup B), \forall A, B \subseteq \Theta$ s.t. $A \cap B = \emptyset$,
- **conjugate capacity**: ν^* is the conjugate capacity of ν if $\nu^*(A) = 1 - \nu(A^c), \forall A \subseteq \Theta$,
- **core**[30]: the core $\mathcal{P}_\nu$ of a capacity ν is the set of probability measures dominating ν: $\{P \in \mathcal{P}_\Theta | \nu \leq P\}$ and $\mathcal{P}_\Theta$ denotes the probability distribution **simplex** for Θ.

To conclude this short presentation of capacity theory we give the following two instrumental lemmas.

30. There is a clash of terminologies in the literature as the core of a belief function as defined by Shafer is the union of its focal elements:

$$\bigcup_{\omega \in \Omega} \Gamma(w) = \{a \in \Theta | pl(a) > 0\}.$$

In this document, the core refers only to the subset of $\mathcal{P}_\Theta$ encompassed by the belief functions. Shafer's definition is referred to as **focal core**.

Lemma 1. *All cores are convex closed subsets of $\mathcal{P}_\Theta$.* [31]

Lemma 2. *For any 2-monotonic capacity v, we have $\mathcal{P}_v \neq \emptyset$.*

The next question is how do we relate capacities with probability bounds? We will give a general[32] definition of lower and upper probabilities relying on p.m.-sets.

Definition 15. Let $\mathcal{P}_i$ denote a p.m.-set. The **lower probability** $\underline{v}_{\mathcal{P}_i}$ of $\mathcal{P}_i$ is a mapping defined as follows:

$$\begin{aligned} \underline{v}_{\mathcal{P}_i} : 2^\Theta &\rightarrow [0;1], \\ A &\rightarrow \min_{\mu \in \mathcal{P}_i} \{\mu(A)\}. \end{aligned} \tag{1.57}$$

The **upper probability** $\overline{v}_{\mathcal{P}_i}$ of $\mathcal{P}_i$ is a mapping defined as follows:

$$\begin{aligned} \overline{v}_{\mathcal{P}_i} : 2^\Theta &\rightarrow [0;1], \\ A &\rightarrow \max_{\mu \in \mathcal{P}_i} \{\mu(A)\}. \end{aligned} \tag{1.58}$$

Finally, the following important theorem achieves our goal.

Theorem 2. *Any lower (resp. upper) probability of a p.m.-set is a super-additive (resp. sub-additive) capacity.*

Furthermore, the upper probability $\overline{v}_{\mathcal{P}_i}$ of $\mathcal{P}_i$ is the conjugate of its lower probability:

$$\overline{v}_{\mathcal{P}_i} = \underline{v}^*_{\mathcal{P}_i}. \tag{1.59}$$

In the next chapter, we will see that the Möbius transform maps belief functions to convenient equivalent representations known as mass functions. The Möbius transform can also be used for capacities but it yields a mass function that is no longer guaranteed to be non-negative. This function is thus more rarely used in imprecise probabilities since for now it lacks a clear interpretation.

We will briefly comment on the interpretation of imprecise probabilities whose most widely accepted one was proposed by Walley 1991. It is a behavioral one that relies on the gambling framework of de Finetti. In this framework, the probability $P(A)$ is understood as the price one is ready to bet on A being true. The payoff is the indicator function 1_A. In this context, a lower probability is understood as the highest acceptable buying price of the gamble, and the upper probability is the lowest acceptable selling[33] price of the gamble. If the upper and lower probabilities coincide, then they jointly represent the fair price for the gamble since this is the price at which either side of the gamble are tempting.

Walley 2000 later argued that imprecise probabilities are an insufficiently expressive model of partial ignorance and uncertainty and therefore he recommends to use upper and lower coherent previsions. Roughly speaking, these latter are obtained in the same fashion as lower and upper probabilities but min and max operators are applied to expectations[34] instead of measures. Closed convex subsets of $\mathcal{P}_\Theta$ are in bijective correspondence with lower and upper previsions.

31. The contraposition is not true. There is however one-to-one correspondence between closed convex subsets of probabilities and lower and upper previsions (*cf.* Walley 2000).

32. general in the sense that it goes beyond Dempster's definition.

33. Selling the gamble is like betting against A.

34. For instance, a coherent lower prevision is given by

$$\min_{\mu \in \mathcal{P}_i} \{\mathbb{E}_\mu[g]\}, \tag{1.60}$$

where $\mathbb{E}_\mu$ is the expectation w.r.t. measure μ and g is a gamble, i.e. a bounded measurable mapping from θ to $\mathbb{R}$. So a coherent lower prevision is a gamble functional. Using gambles that are indicator functions, we retrieve the definition of a lower probability.

1.3.3 Random measures

Earlier in this chapter, we argued that there are two levels of uncertainty in inference problems: uncertainty in the values of the variable of interest and uncertainty in the degrees of belief that we can assign to each candidate value. Following this logic, each estimated probability $\hat{P}(B)$ for $B \subset \Theta$ can be regarded as a random variable in $[0;1]$.

Zooming out to the whole object $\hat{P}$, this latter is a **random measure**. This means that our uncertainty model relies on probabilities of probabilities. Random measures appear quite frequently in probabilistic analyses especially in Bayesian nonparametrics[35] or when point processes are involved. For instance, a (normalized) sum of Dirac masses each centered on a real random variable is a special type of point process (with fixed size) on the real line and it is also a random measure.

35. In such approaches, one tries to infer the posterior distribution on the set of candidate distributions for θ. It is a nonparametric method because there are infinitely many possible distributions P_θ.

Interestingly, random measures enjoy a much wider acceptance and popularity among statisticians than the other frameworks presented in this chapter. The concept of imprecise probabilities is close to a random measure with uniform distribution over its core. As a consequence, random measures are a very general model of uncertainty among those discussed. The imprecise probability model is however not formally covered by random measures with uniform distributions for the same reasons that set theory is not encompassed by probability theory. First, just like for the wine/water paradox, we have epistemic uncertainty on the true distribution and a uniform distribution on distributions is not uninformative. A possibility distribution equal to the indicator function of the core is more adapted to describe our ignorance on the true distribution. Second, in general, deriving second order probabilities requires much more information than what is necessary to obtain imprecise probabilities. Consequently, a uniform distribution on a set of distributions may be interpreted as the results of observations that the distribution happens to be one of the element of this set with equal frequency.

Random measures are however more difficult to analyze as part of some multi-valued logic or gambling protocols. Nonetheless, they resort to hardly more elaborated concepts than probability theory initially provides. This tends to show that all the frameworks evoked in this chapter are just different facets of the same objective. We are back to square one!

1.4 Conclusions

In this chapter, a (non-exhaustive) review of uncertainty theories has been conducted ranging from probability theory to belief functions or imprecise probabilities. From this analysis, it appears that all these frameworks share strong connections and that the selection of one of these frameworks is motivated by a trade-off between representation power and computation efficiency. Figure 1.2 summarizes the relationships between these frameworks. Depending on available data, it has been highlighted that some theories are too specific to produce a model fitting the data without adding additional assumptions. Going for a larger class of models is at the expense of computation load since more general frameworks

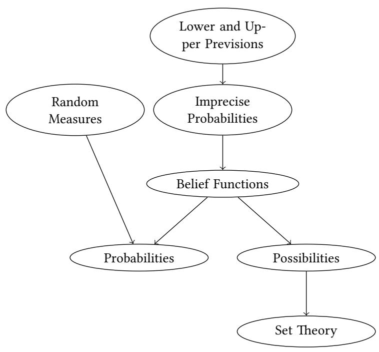


Figure 1.2: Frameworks for reasoning under uncertainty. Arrows symbolize generalization relationship between frameworks.

have fewer axioms and simplifying calculus rules.

When the data does not allow to assign a probability to some event (and in turn to its complement), then a third party comes into play: probability of ”don’t know”. The theory of belief functions appears as the simplest framework allowing to grasp such a three-valued logic and generalizing probability theory.

Our goal in the next chapters is to try to get insights as to the structure of the space where belief functions live. To best use the framework, it is indeed important to understand what operations, distances or binary relations can be defined in accordance with the spirit of the concepts presented in this chapter. Before addressing these topics, we review some important elements of belief function calculus in the next chapter.

2 Elements of belief function calculus in finite spaces

In this chapter, we give basic elements of belief function calculus defined on a finite space Θ. Others elements maybe given in later chapters where they are particularly instrumental. Readers that are familiar with the theory of belief functions may skip this chapter.

2.1 Bijective transformations

There are several alternative set functions which can encode the same level of information as belief functions. We have already seen two such functions: plausibility and commonality functions. Since belief functions can encode more general uncertainty models than probability functions, they have a larger memory complexity, namely $O(N)$ while probabilities have a complexity in $O(n)$. It thus sounds logical that all functions with an equivalent representation power as belief functions have the same complexity.

The most instrumental set functions in the belief function framework are **mass functions** denoted in general by m. There are obtained from functions bel by **Möbius transform**.

The following table gives (forward and backward) transition formulas between mass functions and other set functions. They are given in the case of allowed positive mass for $\emptyset$ and they may simplify otherwise.

$bel \to m$	$m \to bel$
$m(A) = \begin{cases} \sum\limits_{B\subseteq\Theta, B\subseteq A} (-1)^{\lvert A\rvert-\lvert B\rvert} bel(B) & \text{if } A \neq \emptyset \\ 1 - bel(\Theta) & \text{if } A = \emptyset \end{cases}$	$bel(A) = \sum\limits_{B\subseteq\Theta, B\subseteq A, B\neq\emptyset} m(B)$
$pl \to m$	$m \to pl$
$m(A) = \begin{cases} \sum\limits_{B\subseteq\Theta, B\subseteq A} (-1)^{\lvert A\rvert-\lvert B\rvert+1} pl(B^c) & \text{if } A \neq \emptyset \\ 1 - pl(\Theta) & \text{if } A = \emptyset \end{cases}$	$pl(A) = \sum\limits_{B\subseteq\Theta, B\cap A\neq\emptyset} m(B)$
$q \to m$	$m \to q$
$m(A) = \sum\limits_{B\subseteq\Theta, B\supseteq A} (-1)^{\lvert A\rvert-\lvert B\rvert} q(B)$	$q(A) = \sum\limits_{B\subseteq\Theta, B\supseteq A} m(B)$

Table 2.1: Transition relations between mass functions and belief/plausibility/commonality functions.

Mass functions were implicitly evoked in the previous chapter as it can also be retrieved through the usual pseudo-inverse of the multi-valued mapping as

$$m = \frac{1}{\mu \circ \Gamma_{\top}^{-1}(\Theta)} \times \mu \circ \Gamma^{-1}. \tag{2.1}$$

This formula highlights the fact that B is a focal element iff $m(B) > 0$.

It must also be stressed that mass functions take values in $[0;1]$ and that they sum to one

$$\sum_{A \subseteq \Theta} m(A) = 1. \tag{2.2}$$

Given the above remarks, the mass $m(A)$ is understood as how much evidence supports $\{\theta \in A\}$ being true while not allowing to decide among values in A.

Another set function type that is sometimes instrumental are **implicability** functions b. These functions are only useful when $m(\emptyset) > 0$ otherwise they coincide with belief functions. Indeed, we have

$$b(A) = bel(A) + m(\emptyset), \forall A \subseteq \Theta. \tag{2.3}$$

2.2 Subclasses of belief functions

There is a hierarchy of subclasses of belief functions which have specific properties some of which can for instance alleviate the complexity burden incurred by belief functions because they have limited number of focal elements. We review these subclasses from most simple to most general ones.

2.2.1 *Categorical belief functions*

Categorical belief functions lie at the bottom of the hierarchy and they represent imprecise but certain statements such as $\theta \in B$. Any categorical mass function has thus only one focal element B with unit mass. We denote such a categorical mass function by m_B and we have

$$m_B(A) = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{if } A \neq B \end{cases}. \tag{2.4}$$

Categorical mass functions obviously have the same representation power as sets therefore belief functions formally encompasses set theory.

2.2.2 *Bayesian belief function*

When the multi-valued mapping maps elements of Ω to singletons, the source induces a *bona fide* probability measure and the multi-valued mapping is a random variable. It is alo nonetheless a belief function which can be further processed as such. These functions are called **Bayesian belief functions**.

Bayesian and categorical belief functions are intersecting but none is included in the other. Belief functions that are both categorical and Bayesian have a mass function assigning unit mass to a singleton. They encode statements of the kind $\theta = a$ with certainty. This is the most desirable end result of an inferential procedure but in practice it can only be obtained as limiting case when one has an infinite dataset.

2.2.3 *Simple belief functions*

Simple belief functions (or simple support belief functions) were studied and introduced by Shafer 1976 who regards them as elementary pieces of information. Indeed, they represent circumstances under which evidence has only enabled to support one hypothesis $\theta \in B$ with limited probability. They have two focal elements : B and Θ. Their corresponding (simple) mass functions are denoted by m_B^w and we have

$$m_B^w(A) = \begin{cases} 1-w & \text{if } A = B \\ w & \text{if } A = \Theta \\ 0 & \text{otherwise} \end{cases} . \tag{2.5}$$

Simple belief functions generalize categorical ones as $m_B^0 = m_B$.

2.2.4 *Consonant belief functions*

When the focal elements of a belief function are **nested**[1], the belief function is said to be **consonant**. Consequently, consonant belief functions have at most n focal elements and have the same memory complexity as probability distributions. A property featuring this complexity is the following: for any consonant belief function bel, the corresponding plausibility function pl is given by

$$pl(B) = \max_{a \in B} pl(a), \forall B \subseteq \Theta. \tag{2.6}$$

Now we see that knowing plausibilities of singletons is enough to retrieve the whole plausibility function. The restriction of the plausibility function to singletons is often called the **contour function**.

More generally, consonant belief functions are such that

$$bel(A \cap B) = \min \{bel(A); bel(B)\}, \forall A, B \subseteq \Theta, \tag{2.7}$$

$$pl(A \cap B) = \max \{pl(A); pl(B)\}, \forall A, B \subseteq \Theta. \tag{2.8}$$

Consonant belief functions generalize simple ones. Indeed, for any simple mass function m_B^w, its focal elements are B and Θ and $B \subseteq \Theta$.

1. Focal elements are nested if for any two such sets B_1 and B_2, we have either $B_1 \subseteq B_2$ or $B_2 \subseteq B_1$. In other words, the inclusion is a total order for focal elements.

2.2.5 *Separable belief functions*

As previously mentioned, Shafer 1976 intended to show that belief functions represent states of beliefs that are achieved by successive combinations of elementary pieces of evidence through Dempster's rule. Such belief functions are said to be **separable**. To formalize this idea, we need to specify the elementary pieces of evidence. These pieces are embodied by simple belief functions. Consequently, a separable mass function m is such that

$$m = \bigoplus_{B \subset \Theta} m_B^{w(B)}, \tag{2.9}$$

where $w(B)$ is called the **conjunctive weight function**[2]. There are several possible such decompositions. The above becomes unique for instance when

2. In this setting, the function w takes values in $[0; 1]$. A more general version of this function was introduced by Smets in connection with another decomposition that will be detailed in 2.7.

- $m \neq m_{\Omega}$,
- each simple mass function is focused on a different subset of Θ,
- $w(B) = 1$ whenever $B \not\subseteq \{a \in \Theta | pl(a) > 0\}$.

Unfortunately, some belief functions are not separable.

Example 9. Let m denote a mass function such that $m = \beta m_A + (1-\beta) m_{A^c}$ with $\beta \in (0;1)$. A set $B \notin \{A; A^c\}$ cannot be involved in the decomposition (i.e. $w(B) = 1$) otherwise either for any $C \neq B$ we have $w(C) > 0$ and consequently B is a focal element or we would need $w(C) = 0$ for some $C \neq B$ but then all the focal elements of m are subsets of C and thus $C = \Theta$. The only remaining candidate simple mass functions for a decomposition are thus m_A^w and $m_{A^c}^{w'}$ but this implies that either Θ is a focal element (which is untrue) or $w = w' = 0$ and Dempster's rule does not apply, hence a decomposition cannot be found for m.

2.2.6 *Dogmatic belief functions*

A **dogmatic** belief function is such that Θ is not a focal element or equivalently that its corresponding mass function is such that $m(\Theta) = 0$. In practice, non-dogmatic functions are preferred because Dempster's rule always applies if one of the two operands is non-dogmatic.

A belief function is dogmatic and simple if and only if it is categorical. When a belief function is separable and dogmatic, then $\exists B$ s.t. $w(B) = 0$ and focal elements are subsets of B[3].

3. In fact, a conditioning on B has been performed and the restriction of the mass function to subsets of B is a separable mass function whose domain is 2^B.

2.2.7 *Normal belief functions*

The relaxation of the constraint $m(\emptyset) = 0$ as in Smets' TBM leads to a definition allowing to distinguish between those functions that comply with the constraint and those that do not. In this open world context, a belief function whose mass function is such that $m(\emptyset) = 0$ is said to be **normal** (or normalized). In contrast, mass functions such that $m(\emptyset) > 0$ are called **subnormal** mass functions. Finally the union of the two families is of course the entire set of mass functions, but if one desires to stress that $m(\emptyset) > 0$ is possible, then one speaks of **unnormalized** mass functions.

2.3 Discounting

When we want to assembly several sources in order to reduce our uncertainty on the value of θ, we can use a combination operation like Dempster's rule. But we cannot not always assume that each source is built upon relevant evidence. For instance, a source may be built from sensor measurements but the sensor may be uncalibrated and if the model does not take this aspect into account the corresponding belief function conveys erroneous information.

Under such circumstances, it is possible to reduce the impact of a source of information and its corresponding mass function using an operation called **discounting** introduced by Shafer 1976. Discounting a mass function m with discount rate $\alpha \in [0,1]$ is defined as:

$$m^{\alpha}(B) = \begin{cases} (1-\alpha)m(B) \text{ if } B \neq \Theta, \\ (1-\alpha)m(B)+\alpha \text{ if } B = \Theta. \end{cases} \tag{2.10}$$

The higher α is, the stronger the discounting. Thanks to discounting, the mass function induced by a source is transformed into a function closer to the vacuous mass function m_{Θ}. One may remark that a simple mass function m_A^w is the categorical function m_A discounted with rate w.

Mercier et al. 2008 proposed a refined discounting, in which discount rates are tailored for each subset $B \subset \Theta$ and each mass function. The discounting is consequently more precise and specific. It is, however, necessary to have enough information allowing subset-specific discount rate computation. Other generalized mass function correction mechanisms are introduced in Kallel and Hégarat-Mascle 2009 and Mercier et al. 2010 allowing for instance a mass function to be reinforced instead of being discounted.

As part of sequential approaches (Schubert 2010, Zhang et al. 2007, Klein and Colot 2010), it is sometimes needed to discount a mass function m step by step: $m^{\alpha_1 \circ \alpha_2} = (m^{\alpha_1})^{\alpha_2}$ with $\circ$ the composition law for successive discountings. If discountings are repeated k times with rates $(\alpha_1, ..., \alpha_k)$, one has the following property (Smets 2007) :

$$m_i^{\beta} = m_i^{\alpha_1 \circ \ldots \circ \alpha_k}, \tag{2.11}$$

$$\text{with } \beta = 1 - \prod_{i=0}^{k} (1-\alpha_i). \tag{2.12}$$

In particular, when $\forall i, \alpha_i = \alpha$, we have $\beta = 1-(1-\alpha)^k$.

2.4 Coarsening and refinement

An interesting aspect of belief functions is that they can propagate the encoded uncertainty on the variable of interest to other spaces than Θ representing different levels of granularity. For instance, suppose θ represents a real variable quantified on four levels $\Theta = \{\theta_1; \theta_2; \theta_3; \theta_4\}$ and we are able to infer a belief function on Θ. What can we deduce if the variable needs later to be quantified on 2 or 8 levels instead of 4?

It turns out that our beliefs on θ can be propagated to a new variable space through a mapping ρ that is called a **refining**.

Definition 16. A multi-valued mapping $\rho : \Theta \to 2^{\mathcal{Y}}$ is a **refining** if the family $\{\rho(a)\}_{a \in \Theta}$ is a partition of $\mathcal{Y}$.

Let m denote the mass function inferred on Θ and m' the mass function that we want to obtain on a new space denoted by $\mathcal{Y}$. Depending on the cardinality of $\mathcal{Y}$, we have that:

- $\rho : \Theta \to 2^{\mathcal{Y}}$ is a refining and $m'\left(\bigcup_{a \in A} \rho(a)\right) = m(A), \forall A \subseteq \Theta$ and all other masses are null (**refinement** case). In this case, m' is called the vacuous extension of m.

- $\rho : \mathcal{Y} \to 2^\Theta$ is a refining and we need to resort to pseudo-inverses. In this context, $\rho_\top^{-1}$ is called **outer reduction** and $\rho_\perp^{-1}$ is called **inner reduction**. For some pseudo-inverse ρ^{-1}, the new mass function m' on $\mathcal{Y}$ is obtained as

$$m'(B) = \sum_{\substack{A \subseteq \Theta, \\ \rho^{-1}(A) = B}} m(A) \text{ (\textbf{coarsening} case).} \quad (2.13)$$

  If we use an outer reduction, m' is called the marginal mass function of m.

Example 10. Let $\Theta = \{\theta_1, \theta_2\}$ and $\mathcal{Y} = \{y_1; y_2; y_3\}$ denote two spaces. Suppose there exist a refinement ρ as illustrated in figure 2.1 such that :

$$\begin{aligned} \rho(\theta_1) &= \{y_1, y_3\}, \\ \rho(\theta_2) &= \{y_2\}. \end{aligned}$$

Let us introduce the following mass function over Θ:

subset	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1, \theta_2\} = \Theta$
m	0	0.2	0.1	0.7

The mass function m' on $\mathcal{Y}$ induced by ρ from m is:

subset	$\emptyset$	$\{y_1\}$	$\{y_2\}$	$\{y_1, y_2\}$	$\{y_3\}$	$\{y_1, y_3\}$	$\{y_2, y_3\}$	$\mathcal{Y}$
m'	0	0	0.1	0	0	0.2	0	0.7

Now, suppose we want to perform the opposite operation starting from function m'. The outer and inner reductions are:

subset	$\emptyset$	$\{y_1\}$	$\{y_2\}$	$\{y_1, y_2\}$	$\{y_3\}$	$\{y_1, y_3\}$	$\{y_2, y_3\}$	$\mathcal{Y}$
$\rho_\top^{-1}$	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	Θ	$\{\theta_1\}$	$\{\theta_1\}$	Θ	Θ
$\rho_\perp^{-1}$	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	$\emptyset$	$\{\theta_1\}$	$\{\theta_1\}$	$\emptyset$	Θ

In this case, using (2.13) with any of the above pseudo-inverses yields another mass function m'' equating m:

subset	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1, \theta_2\} = \Theta$
$m'' = m$	0	0.2	0.1	0.7

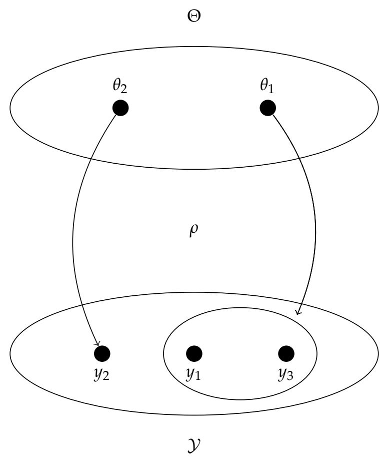


Figure 2.1: Refining mapping example: the first element of Θ is mapped to an element of $\mathcal{Y}$ and the second to a pair of elements of $\mathcal{Y}$.

2.5 Belief functions on product spaces

When performing inference, two spaces are involved: the variable space Θ and the observation space $\mathbb{X}$. The full uncertainty model thus implies the definition of a belief function on the product space $\Theta \times \mathbb{X}$ and we need to retrieve this "joint"[4] belief function from marginal or conditional ones (and vice-versa).

We have already mentioned two ways to perform conditioning and these two apply also in the case of a belief function on a product space. We will now explain how to perform marginalization. The theory of belief function also contains two principled ways to obtain a belief function on the product space from a conditional or marginal one. These latter are known as ballooning and vacuous extensions and are also presented in the following paragraphs.

4. A joint belief function actually refers to a more specific object than a belief function on a product space see 5.4 for a definition.

2.5.1 Marginalization

Let m denote a mass function on the product space $\Theta \times \mathbb{X}$. Suppose we want to compute the marginal mass function m' on one of the marginal spaces, say Θ. This amounts to a particular form of coarsening. Marginalization relies on the concept of **set projection**: $\forall B \subseteq (\Theta \times \mathbb{X})$

$$\mathrm{proj}_\Theta (B) = \bigcup_{(a,x) \in B} \{a\}, \tag{2.14}$$

$$= \{a \in \Theta | \exists x \in \mathbb{X} \text{ s.t. } (a,x) \in B\}. \tag{2.15}$$

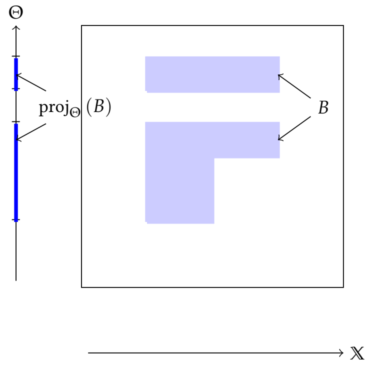


Figure 2.2: Illustration of the set projection of some focal element B in $2^{\Theta \times \mathbb{X}}$ to a subset of Θ.

The set $A = \mathrm{proj}_\Theta (B)$ is also the smallest set such that $B \subseteq (A \times \mathbb{X})$, see figure 2.2 for an illustration. Now, the **marginal mass function** m' is obtained as

$$m'(A) = \sum_{\substack{B \subseteq \Theta \times \mathbb{X} \\ \mathrm{proj}_\Theta(B) = A}} m(B), \forall A \subseteq \Theta \tag{2.16}$$

The refining map underlying the coarsening of marginalization is $\rho(A) = A \times \mathbb{X}$ for any $A \subseteq \Theta$. Using the outer reduction, m' is obtained. Moreover, we also have

$$pl'(A) = pl(A, \mathbb{X}), \forall A \subseteq \Theta. \tag{2.17}$$

2.5.2 Extension

Generally speaking, computing a "joint" belief function from a marginal or a conditional one is an ill posed problem for several joint belief functions have the same marginal or conditional distributions. Additional information or hypotheses must thus be used to select among candidate functions on the product space.

In the theory of belief functions, the **least commitment principle** (LCP) is often used to that end. It consists in selecting the function that contains as little information[5] as possible among the candidate ones. Let m denote a marginal (resp. conditional) mass function on Θ. If a mass function m' on $\Theta \times \mathbb{X}$ is an LCP solution then m' is called an **extension**. Extensions can be regarded as special kinds of refinements.

5. To make the LCP operational, a partial order on belief functions must be defined in order to sort belief functions based on their informational content. Such partial orders are presented in chapter 3

In the case where m is a marginal mass function, the LCP solution m' is known as **the vacuous extension** and is given by

$$m'(B) = \begin{cases} m(A) & \text{if } B = A \times \mathbb{X} \\ 0 & \text{otherwise} \end{cases}. \tag{2.18}$$

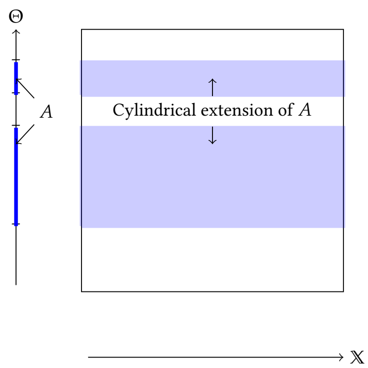


Figure 2.3: Illustration of the cylindrical extension of some focal element A in 2^Θ to a subset of $\Theta \times \mathbb{X}$.

The same refining mapping as in the marginalization process is underlying the vacuous extension. The focal elements of m' are called cylindrical extensions of those of m (see figure 2.3).

In the case where m is a mass function obtained after conditioning on some $E \subset \mathbb{X}$, the LCP solution m' is known as **the ballooning extension** (or conditional embedding) and is given by

$$m'(B) = \begin{cases} m(A) & \text{if } B = (A \times E) \cup (\Theta \times \mathbb{X} \setminus E) \\ 0 & \text{otherwise} \end{cases}. \tag{2.19}$$

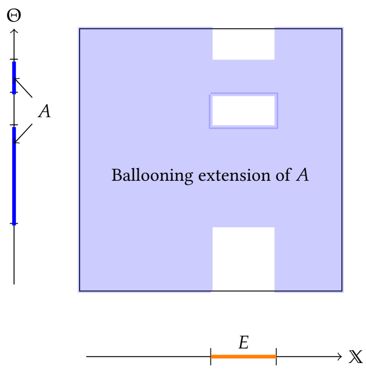


Figure 2.4: Illustration of the ballooning extension of some focal element A in 2^{Θ} to a subset of $\Theta \times \mathbb{X}$. It is assumed that Dempster's conditioning on $\Theta \times E$ with $E \subset \mathbb{X}$ was previously performed on the joint belief function.

Figure 2.4 illustrates that the ballooning extension selects the largest focal element in $2^{\Theta\times\mathbb{X}}$ such that conditioning (or intersecting) with $\Theta \times E$ and projecting on Θ gives A. Observe that the ballooning extension is the LCP solution for Dempster's conditioning only.

In both these extensions, the application of the LCP translates into selecting the largest possible focal elements for m' (subject to the constraints induced by marginalization or conditioning).

2.6 The α-junctions

In this subsection, a brief presentation of α-junctions (Smets 1997) is proposed. Suppose two sources S_1 and S_2 are induced by pieces of evidence allowing them to induce two mass functions m_1 and m_2 respectively on the same space Θ. In general, we dot no know if the pieces of evidence of S_1 and S_2 are independent, shared or a mix of those.

Evidential combination rules (such as Dempster's rule) address the problem of aggregating the two functions m_1 and m_2 into a single one synthesizing both of the initial evidence bodies. Let f be a combination operator for mass functions, i.e. $m_{12} = f(m_1, m_2)$ with m_{12} a mass function depending only on m_1 and m_2. Such an operator is an **$\boldsymbol{\alpha}$-junction** if it possesses the following properties:

- **Linearity**[6]: $\forall \lambda \in [0,1], f(m, \lambda m_1 + (1-\lambda) m_2) = \lambda f(m, m_1) + (1-\lambda) f(m, m_2)$,
- **Commutativity**: $f(m_1, m_2) = f(m_2, m_1)$,
- **Associativity**: for any additional mass function m_3, $f(f(m_1, m_2), m_3) = f(m_1, f(m_2, m_3))$,
- **Neutral element**: $\exists m_e \mid \forall m, f(m, m_e) = m$,
- **Anonymity**: for any mapping $\Upsilon : 2^{\Theta} \longrightarrow 2^{\Theta}$ such that its restriction on Θ is a permutation and $\Upsilon(B) = \bigcup_{e\in B} \Upsilon(e)$ when $|B| > 1$, we have $f(m_1 \circ \Upsilon, m_2 \circ \Upsilon) = m_{12} \circ \Upsilon$,
- **Context preservation**: $pl_1(B) = 0$ and $pl_2(B) = 0 \implies pl_{12}(B) = 0, \forall B \subseteq \Theta$.

6. The operator is linear on the vector space spanned by categorical mass functions but the output of the operator remains a mass function only in case of convex combination.

In short, α-junctions are thus linear combination rules that do not depend on the order in which pieces of evidence are processed. The justification behind these properties are detailed in Smets 1997. In the same article, Smets proves that the neutral element m_e can only be either $m_{\emptyset}$ or m_{Θ}. Depending on this, two sub-families arise: the α-disjunctive rules denoted by $\textcircled{\scriptsize$\cup$}^{\alpha}$ and the α-conjunctive rules denoted by $\textcircled{\scriptsize$\cap$}^{\alpha}$. For both of these families, Pichon and Denoeux 2009 provided the following computation formulas: $\forall E \subseteq \Theta, \forall \alpha \in [0,1]$

$$m_{1\cap^{\alpha}2}(E) = \sum_{\substack{A,B,C\subseteq\Theta,\\ (A\cap B)\cup(A^c\cap B^c\cap C)=E}} m_1(A)\, m_2(B)\, \alpha^{|C^c|}\overline{\alpha}^{|C|}, \tag{2.20}$$

$$m_{1\cup^{\alpha}2}(E) = \sum_{\substack{A,B,C\subseteq\Theta,\\ (A\Delta B)\cup(A\cap B\cap C)=E}} m_1(A)\, m_2(B)\, \alpha^{|C|}\overline{\alpha}^{|C^c|}, \tag{2.21}$$

Some simplified formulas when combining categorical mass functions are given in Klein et al. 2014.

with $\overline{\alpha} = 1 - \alpha$ and Δ the set symmetric difference.

When $\alpha = 1$, the classical conjunctive and disjunctive rules are retrieved. We denote these rules by $\textcircled{\scriptsize$\cap$} = \textcircled{\scriptsize$\cap^1 and $\textcircled{\scriptsize$\cup$} = \textcircled{\scriptsize$\cup^1. These rules have simplified expressions: $\forall E \subseteq \Theta$,

$$m_{1\cap 2}(E) = \sum_{\substack{A,B\subseteq\Theta,\\ A\cap B=E}} m_1(A)\, m_2(B)\,, \tag{2.22}$$

$$m_{1\cup 2}(E) = \sum_{\substack{A,B\subseteq\Theta,\\ A\cup B=E}} m_1(A)\, m_2(B)\,. \tag{2.23}$$

Computing an α-junctive combination has, in general, a complexity of N^3 while the conjunctive and disjunctive rules have a complexity of N^2.

Moreover, we introduce specific notations regarding the α-junctive combination of a given function m_1 with a given categorical function m_B:

- $m_{1|B} := m_1 \circledcirc m_B$, this mass function is sometimes referred to as m_1 given B,
- $m_{1\cap^\alpha B} := m_1 \textcircled{\cap}^\alpha m_B$,
- $m_{1\cup^\alpha B} := m_1 \textcircled{\cup}^\alpha m_B$,
- $m_{1|^\alpha B}$ stands for the combination of m_1 and m_B using an α-junction when distinguishing conjunctive and disjunctive cases is unnecessary.

Remember that m_B is interpreted as $\theta \in B$ with certainty. The above operations are thus often called conditioning on B.

Dempster's rule is closely related to the conjunctive rule. Indeed, since Dempster's rule was initially introduced for normal belief functions, the degree of belief assigned to $\emptyset$ is ultimately redistributed to the focal sets. We have

$$m_{1\oplus 2}(B) = \frac{1}{1-\kappa} m_{1\cap 2}(B)\,, \forall B \subseteq \Theta, \tag{2.24}$$

and $\kappa = m_{1\cap 2}(\emptyset)$ is called Dempster's **degree of conflict**.

There are plenty other ways to combine mass functions and we will present some others in chapter 5. α-junctions are presented earlier because they will also be instrumental in chapter 3.

The interpretation of α-junctions is related to information items concerning the **truthfulness** of the sources S_1 and S_2. In an information fusion context, such items are known as **meta-information** and truthfulness is a special kind of meta-information. Actually, several forms of truthfulnesses can be observed in practice but regarding α-junctions the following "adversarial" definition is retained: *S_i is untruthful if it supports the opposite of what it knows to be true.*

Depending on the truthfulness of sources, very different decisions can be made in the end, which accounts for the importance of taking meta-information into account in information fusion problems. In general, our knowledge about the truthfulness of each source is imprecise and uncertain and it is therefore expressed as a mass function on a meta-domain $\mathcal{T}_i$. Pichon 2012, explains that an element $t_C^i \in \mathcal{T}_i$ is understood as the fact that S_i is truthful when it supports $\{\theta \in C\}$ and it is untruthful when it supports $\{\theta \in C^c\}$. Let us provide a simple example where everything is deterministic:

Example 11. Suppose that $|\Theta| = 4$ and the meta-data concerning S_i is that t_C^i has probability 1 with $C = \{\theta_1, \theta_2\}$. If the source of information S_i delivers only one certain piece of evidence $\theta \in A = \{\theta_2, \theta_3\}$, then four different situations are encountered:

- The source gives support to θ_2 and can be trusted about θ_2. We conclude that θ_2 is a possible value for θ.
- The source gives no support to θ_1 and can be trusted about θ_1. We conclude that θ_1 is a not possible value for θ.
- The source gives support to θ_3 but cannot be trusted about θ_3. We conclude that θ_3 is not a possible value for θ.
- The source gives no support to θ_4 but cannot be trusted about θ_4. We conclude that θ_4 is a possible value for θ.

All in all, the testimony of the source is $\theta \in A$ but given the meta-data, the actual testimony is $\theta \in (A\Delta C)^c$.

When considering a pair of sources $(S_1; S_2)$, meta-events belong to $\mathcal{T}_1 \times \mathcal{T}_2$. Pichon also proves that:

- for α-conjunctions, the underlying meta-information is that each meta-event $\{$*either both sources are fully truthful or they both lie about* $C^c\} = \left\{\left(t^1_\Theta; t^2_\Theta\right); \left(t^1_C; t^2_C\right)\right\}$ has probability $\alpha^{|C|}\overline{\alpha}^{|C^c|}$.
- for α-disjunctions, the underlying meta-information is that each meta-event $\{$*one source is fully truthful while the other one lies at least about* $C\} = \bigcup\limits_{E \subseteq C^c} \left\{\left(t^1_\Theta; t^2_E\right); \left(t^1_E; t^2_\Theta\right)\right\}$ has probability $\alpha^{|C^c|}\overline{\alpha}^{|C|}$.

In particular, when $\alpha = 1$, the above probabilities are null whenever $C \neq \Theta$ in the conjunctive case and whenever $C \neq \emptyset$ in the disjunctive case. The meta-information thus reduces to:

- for the conjunctive rule, the event $\{$*both sources are fully truthful*$\}$ has probability 1.
- for the disjunctive rule, the event $\{$*at least one of the sources is fully truthful*$\}$ has probability 1.

Note that α-junctions are a particular case of a combination process introduced in Pichon et al. 2012 where a general framework for reasoning under various meta-information is formalized.

2.7 Decompositions

We have seen that some belief functions can be decomposed using Dempster's rule and such functions are called separable. We can think of similar decompositions with respect to other rules. Smets 1995 showed that non-dogmatic mass functions ($m(\Omega) > 0$) can be decomposed using the conjunctive rule. This decomposition of a mass function into simple ones is always unique and relies on a generalized definition of the conjunctive weight function. In this case, the codomain of conjunctive weight functions is $(0; +\infty)$ and not $[0, 1]$. Having $w(A) < 1$ is understood as the fact that some evidence has been collected allowing to support A being true. Having $w(A) > 1$ means that A is unlikely to the point that a significant amount of evidence needs to be collected before starting to support A being true. Finally, $w(A) = 1$ stands for a neutral opinion regarding event A. We refer to Denœux 2008 for more details on conjunctive weight functions. The decomposition writes

$$m = \bigcirc\!\!\!\!\cap_{B \subset \Theta} m_B^{w(B)}. \tag{2.25}$$

The simplest transition relations allowing to compute conjunctive weight functions are obtained from commonality functions as follows:

$$w(A) = \prod_{E \supseteq A} q(E)^{(-1)^{|E|-|A|+1}}, \forall A \subseteq \Theta. \tag{2.26}$$

The non-dogmatic condition prevents division by zero to happen in Equation (2.26). In practice, when a dogmatic mass function m has to be turned into a conjunctive weight function, one can discount it with a very small discount rate as compared to the minimum positive mass of m. The discount rate can be chosen as small as necessary so that the values of w stabilize to some value up to a prescribed precision threshold.

When mass functions are unnormalized ($m(\emptyset) > 0$), a dual decomposition can be obtained using **disjunctive weight functions** denoted by v:

$$m = \bigcirc\!\!\!\!\cup_{B \neq \emptyset} m^B_{v(B)}, \tag{2.27}$$

where each function $m^B_{v(B)}$ is defined as

$$m^B_{v(B)}(E) = \begin{cases} 1 - v(E) & \text{if } E = B, \\ v(E) & \text{if } E = \emptyset, \\ 0 & \text{otherwise} \end{cases}. \tag{2.28}$$

Disjunctive weight functions can be computed for instance from implicability functions as follows:

$$v(A) = \prod_{E \subseteq A} b(E)^{(-1)^{|A|-|E|+1}}, \forall A \subseteq \Theta. \tag{2.29}$$

In the same fashion as conjunctive weight functions, one turns a normalized mass function m into a disjunctive weight function by artificially assigning an infinitesimal mass value to $\emptyset$ and then renormalize so that $\sum_{E \subseteq \Theta} m(E) = 1$. Such a constraint may be perceived as less natural than $m(\Theta) > 0$, in particular under a closed-world assumption.

2.8 Matrix calculus for belief functions

Mass functions can be viewed as vectors belonging to the vector space $\mathbb{R}^N$ with categorical mass functions as base vectors. Since mass functions sum to one, the set of mass functions is the simplex $\mathcal{M}$ in that vector space whose vertices are the base vectors $\{m_A\}_{A \subseteq \Theta}$. This simplex is also called **mass space** (Cuzzolin 2008). In this paper, the following notations and conventions are used :

- Vectors are column vectors and are written in bold small letters. The operator **vec** maps a set function or a distribution to its vector form. For instance, $\mathbf{m}_i = \mathbf{vec}(m_i)$ is the mass vector corresponding to a mass function m_i. The length of mass vectors is N. The $j_A{}^{\text{th}}$ element of a mass vector $\mathbf{m}_i$ is such that $\mathbf{m}_i(j_A) = m_i(A)$ with j_A the integer index of the set A according to the binary order. The **binary order**[7] (Smets 2002) is a common way to index elements of 2^Θ without supposing any order on Θ.

7. For a given finite set $\Theta = \{\theta_1, \ldots, \theta_n\}$, the representation of subsets $A \subseteq \Theta$, known as **binary order**, is the following: each subset A is associated to a binary number made of n bits and this number has a 1 at position i if $\theta_i \in A$, and 0 otherwise. For example, when $\Theta = \{\theta_1, \theta_2, \theta_3\}$, the binary representation of the subset $\{\theta_2, \theta_3\}$ is 110. Then, considering the integer number Int_A in base 10 obtained from the binary representation of A, each subset defines a unique vector index $n_A = Int_A + 1$ starting from 1 ($\emptyset$) to 2^n (Θ).

- Matrices are written in bold capital letters. They are square and their size is $N \times N$. A matrix can be represented by $\mathbf{X} = [X(i_A, j_B)]$, or alternatively by the notation $\mathbf{X} = [X(A,B)], \forall A, B \in \Theta$. The row and column indexes i_A and j_B are those corresponding to the subsets A and B using the binary order.
- $\mathbf{I}$ is the identity matrix.
- $\mathbf{1}$ is the all-ones matrix.
- $\mathbf{J}$ is the matrix with null components except those on the anti-diagonal which are equal to 1. $\mathbf{J}$ is a permutation matrix reversing lines in case of right-handed product and reversing columns in case of left-handed product.

Matrix calculus as part of the theory of belief functions is especially interesting when it comes to mass function α-junctive combination. In Smets 2002 and Smets 1997, Smets shows that equation (2.20) and (2.21) can be written as a product between a matrix and a mass function vector. Let $\mathbf{K}^{\cap}_{1,\alpha}$ be a matrix such that $K^{\cap}_{1,\alpha}(A,B) = m_{1\cap^{\alpha}B}(A)$ and $\mathbf{K}^{\cup}_{1,\alpha}$ a matrix such that $K^{\cup}_{1,\alpha}(A,B) = m_{1\cup^{\alpha}B}(A)$. One has:

$$\mathbf{m}_{1\cap^{\alpha}2} = \mathbf{K}^{\cap}_{1,\alpha} \cdot \mathbf{m}_2, \tag{2.30}$$
$$\mathbf{m}_{1\cup^{\alpha}2} = \mathbf{K}^{\cup}_{1,\alpha} \cdot \mathbf{m}_2. \tag{2.31}$$

These matrices are also in one-to-one correspondence with the mass function m_1. We call $\mathbf{K}^{\cap}_{1,\alpha}$ and $\mathbf{K}^{\cup}_{1,\alpha}$ the **α-specialization** and **α-generalization matrices** corresponding to m_1. In general, all such matrices will be called **evidential matrices**. When it is not necessary to stress the dependency of evidential matrices on α and on the conjunctive/disjunctive cases, an evidential matrix is denoted by $\mathbf{K}_1$ for the sake of equation concision. In particular, when $\alpha = 1$, we see that the conjunctive (2.22) and disjunctive (2.23) combinations can be rewritten as a dot product.

Each element of $\mathbf{K}_1$ represents the mass assigned to a set A after learning that $\{\theta \in B\}$: $K_1(A,B) = m_{1|^{\alpha}B}(A)$. In other words, $\mathbf{K}_1$ does not only represent the current state of belief depicted by m_1 but also all reachable states from m_1 through an α-junctive conditioning. From a geometric point of view (Cuzzolin 2004), each column of an evidential matrix $\mathbf{K}_1$ corresponds to the vertex of a polytope $\mathcal{C}_1$, called the **conditional subspace** of m_1. Example 12 illustrates this latter remark.

Example 12. Let $|\Theta| = 2$ and $m_1 = \overline{\lambda} m_{\{\theta_1\}} + \lambda m_{\Theta}$ with $\lambda \in [0;1]$ and $\overline{\lambda} = 1 - \lambda$. In the conjunctive case, we have :

$$\mathbf{K}^{\cap}_{1,\alpha} = \begin{pmatrix} \overline{\lambda}\alpha + \lambda & 0 & \overline{\lambda} & 0 \\ 0 & \overline{\lambda}\alpha + \lambda & 0 & \overline{\lambda} \\ \overline{\lambda}\overline{\alpha} & 0 & \lambda & 0 \\ 0 & \overline{\lambda}\overline{\alpha} & 0 & \lambda \end{pmatrix}.$$

The four column vectors of $\mathbf{K}^{\cap}_{1,\alpha}$ are (from left to right) : $\mathbf{m}_{1\cap^{\alpha}\emptyset}$, $\mathbf{m}_{1\cap^{\alpha}\{\theta_1\}}$, $\mathbf{m}_{1\cap^{\alpha}\{\theta_2\}}$, and $\mathbf{m}_1$. By definition the polytope $\mathcal{C}_1$ is the following subset of $\mathcal{M}$:

$$\mathcal{C}_1 = \left\{ m \in \mathcal{M} \mid m = \sum_{A\subseteq\Theta} \lambda_A \mathbf{m}_{1\cap^{\alpha}A}, \sum_{A\subseteq\Theta} \lambda_A = 1, \lambda_A \in [0;1] \right\}.$$

Any mass function $m \in \mathcal{C}_1$ is the result of the combination of m_1 with another mass function using a given α-junction. Evidential matrices are consequently relevant candidates for assessing dissimilarities between bodies of evidence in compliance with α-junctions as we will see in chapter 4.

Most importantly, if $\mathbf{K}_1$ and $\mathbf{K}_2$ are two evidential matrices and if $\mathbf{K}_{12}$ denotes the matrix corresponding to the α-junction of m_1 with m_2, then one has:

$$\mathbf{K}_{12} = \mathbf{K}_1 \cdot \mathbf{K}_2. \tag{2.32}$$

Moreover, the transpose of any evidential matrix $\mathbf{K}$ is a stochastic matrix, meaning that all lines sum to one: ${}^{\mathbf{t}}\mathbf{K} \cdot \mathbf{1} = \mathbf{1}$, with ${}^{\mathbf{t}}\mathbf{K}$ the transpose matrix of $\mathbf{K}$. Finally, $\mathbf{K}_A$ will denote the evidential matrix corresponding to the categorical mass function m_A.

2.9 Inference

As we have already commented in several parts of this monograph, inference is about making deductions concerning a variable of interest $\theta \in \Theta$ based on observations that are elements of a space $\mathbb{X}$. Let $\mathcal{D} = \left\{x^{(1)}, \dots, x^{(n_x)}\right\}$ denote the collected set of observations, or dataset.

In the Bayesian probabilistic framework, we start with initial beliefs on the true value of θ which is regarded as a random variable. These beliefs are represented by a **prior** probability distribution P_0. When data arrive, each datum is regarded as a realization of a random variable X and we will update our beliefs based on a dependence model between X and θ. From Bayes rule, we write

$$P(\theta = a|\mathcal{D}) = \frac{P(\mathcal{D}|\theta = a)\, P_0(\theta = a)}{P(\mathcal{D})}. \tag{2.33}$$

The probability $P(\mathcal{D}|\theta = a)$ is called **likelihood** and evaluates how likely are the data for a given value of θ. We need a model to estimate the likelihood. This model can be obtained from information about how data points are sampled or it can be just assumed based on the shape of empirical distributions. The probability $P(\mathcal{D})$ is the evidence and usually does not need to be evaluated because it is not dependent on θ so we can still compute the **posterior distribution** $P(\theta = a|\mathcal{D})$ by multiplying the likelihood and the prior and then renormalize so that the distribution sums to one. The posterior distribution stands for our updated beliefs as to what can be inferred about θ.

In the belief function framework, we can reproduce this reasoning. In lack of any available information, we can choose a vacuous mass function as prior: $m_0 = m_\Theta$. If there is a really justified reason why we are not ignorant regarding θ, we can choose a more informative mass function and in particular, if we believe that P_0 is justified, we can choose a Bayesian mass function such that $m_0(\{a\}) = P_0(\theta = a)$ for any $a \in \Theta$. A likelihood function can also be used as basis to induce a belief function on Θ as proposed by Shafer 1976. The corresponding plausibility function is

$$pl_{\mathcal{D}}(\{a\}) = \frac{P(\mathcal{D}|\theta = a)}{\sup_{a' \in \Theta} P(\mathcal{D}|\theta = a')}, \forall a \in \Theta, \quad (2.34)$$

$$pl_{\mathcal{D}}(A) = \sup_{a' \in A} pl_{\mathcal{D}}(\{a'\}), \forall A \subseteq \Theta. \quad (2.35)$$

By construction, this belief function is consonant. Now if we combine m_0 with $m_{\mathcal{D}}$ by Dempster's rule, we obtain a posterior mass function m. In addition, if m_0 is a Bayesian mass function, the posterior mass function coincides with the posterior distribution issued by the Bayesian reasoning (see Denœux 2014).

Another inference scheme was proposed by Smets 1993. It is known as the **generalized Bayesian theorem** (GBT). Suppose we are given a set of conditional[8] mass functions on $\mathbb{X}$: $(m(.|\theta = a))_{a \in \Theta}$, a prior m_0 and for simplicity only one datum $\mathcal{D} = \{x\}$. The GBT procedure is the following

1. Compute ballooning extension of each $m(.|\theta = a)$ and the vacuous extensions of m_0 to obtain $n+1$ mass functions on $\Theta \times \mathbb{X}$;
2. Combine all of them using $⓪$;
3. Condition on $\Theta \times \{x\}$;
4. Marginalize on Θ.

If $m_0 = m_\Theta$, for any $A \subseteq \Theta$ and any $E \subseteq \mathbb{X}$, the posterior mass function obtained from the GBT procedure is such that its plausibility values write

$$pl(A|E) \propto 1 - \prod_{a \in A} (1 - pl(E|\theta = a)). \quad (2.36)$$

When we have n_x observations that are cognitively independent [9], the conjunctive combination of each posterior mass function obtained by applying the GBT to each observation is equal to the GBT solution if we had performed conditional embedding in $\Theta \times \mathbb{X}^{n_x}$ in step 1 and conditioning on $\Theta \times \left\{x^{(1)}\right\} \times \ldots \times \left\{x^{(n_x)}\right\}$ in step 3.

When all conditional mass functions and the prior mass function are Bayesian, the GBT boils down to usual Bayes rule. The GBT cannot be applied blindly as it inherits the independence assumption that is required to use $⓪$. A discussion on the limitations of the GBT is given in Dubois and Denœux 2010.

Of course, the fiducial style of inference presented in 1.2.6 is another possible way to obtain quantified beliefs regarding the actual value of θ.

8. Unless stated otherwise, $m(.|B)$ denotes a conditioning *à la Dempster.*

9. Cognitive independence holds if for any E_1 and $E_2 \subseteq \mathbb{X}$, we have

$$pl(E_1 \times E_2|\theta = a) = pl(E_1|\theta = a) \times pl(E_2|\theta = a).$$

2.10 Decision making

We have already mentioned the pignistic transform which is a central concept in the second stage of Smets' TBM. Smets has axiomatized the derivation of the pignistic probability distribution. He gives three desirable properties[10] for a candidate distribution P to qualify as the best representative of a given mass function m:

(i) $P(a)$ depends only on masses $m(B)$ where $a \in B$,
(ii) the same probability mass is assigned to any $a \in \Theta$ before or after permutation over the element indices,
(iii) if a candidate value a is ruled out by m, then $P(a) = 0$.

10. The result in Smets and Kennes 1994 is a bit more general. The authors look for any bounded map compliant with the 3 assumptions. The end result is the probability distribution (2.37)

Under the above conditions, the **pignistic distribution** necessarily writes as

$$P(a) = \sum_{\substack{A \subseteq \Theta, \\ a \in A}} \frac{1}{|A|} \frac{m(A)}{1 - m(\emptyset)}. \tag{2.37}$$

Given a probability distribution on (Θ, σ_Θ), then standard decision theory applies. Suppose a decision must be made for variable $y \in \mathbb{Y}$ and the variable is dependent on θ. The elements of $\mathbb{Y}$ are called acts and are mappings from Θ to the set of consequences $\mathcal{O}$. Decisions on what act to choose to do can be made based on an uncertainty model for θ and a reward (or utility)[11] function $r : \mathcal{O} \to \mathbb{R}$. The real $r(y(a))$ is the payoff of deciding y when $\theta = a$.

11. Some people also prefer working with loss functions. A loss function is obtained by multiplying a utility function with -1.

Example 13. Suppose a^* is an estimate of θ. An act consists in selecting an element a^* among those in Θ. The set of consequences is $\mathcal{O} = \{\text{match}; \text{mismatch}\}$. Then the act of selecting a^* is formally defined as

$$y(a) = \begin{cases} \text{match} & \text{if } a = a^* \\ \text{mismatch} & \text{if } a \neq a^* \end{cases}. \tag{2.38}$$

The set of acts $\mathbb{Y}$ is the set of such selection mappings. Furthermore, all mistakes are equally penalized (say $r(\text{mismatch}) = 0$) and all correct estimations are equally rewarded (say $r(\text{match}) = 1$), then

$$\mathbb{E}_P[r \circ y] = \sum_{a \in \Theta} r(y(a)) P(a), \tag{2.39}$$

$$= \sum_{a \in \Theta} 1_{a^*}(a) P(a), \tag{2.40}$$

$$= P(a^*) \tag{2.41}$$

The act with maximal expected reward is thus obtained for $a^* = \arg\max_{a \in \Theta} P(a)$. Observe that the result holds when P is the actual distribution of θ whereas in practice we have only access to an estimated distribution.

In general, decisions can be made when a preference preorder $\preceq$ can be obtained for the set of acts. Decision making based on probabilities is strongly supported by Savage's theorem. Savage starts by defining rationality requirements for decision in the form of seven axioms[12]. If the axioms hold for a given preorder, then he proved the following result.

12. These axioms follow a certain logic and they are presented in A.1.

Theorem 3. *A preference preorder $\preceq$ among acts with rationality requirements exists iff there is a finitely additive probability measure P and a utility function r such that*

$$y_1 \preceq y_2 \Leftrightarrow \mathbb{E}_P[r \circ y_1] \leq \mathbb{E}_P[r \circ y_2], \tag{2.42}$$

$$\Leftrightarrow \int r(y_1(a))\, dP(a) \leq \int r(y_2(a))\, dP(a). \tag{2.43}$$

Furthermore, P is unique and r is unique up to an affine transform.

We see that making a decision in a probabilistic way translates into a sound preference relation. Concerning the pignistic distribution, Smets argues that one of its main features is that it avoids sure loss (Dutch books[13]).

13. A Dutch book appears when any act incurs a negative reward with probability 1.

The second axiom of Savage is called the **sure thing principle**. By relaxing the sure thing principle into a weaker form[14], Gilboa 1987 showed that preference preorders can be obtained by substituting the measure P with a non-additive measure. Since *bel* and *pl* are such objects, they can serve as basis for decision making. For example, a preference can be derived from upper expected rewards as

$$y_1 \preceq y_2 \Leftrightarrow \mathbb{E}_{P_\top}\left[r \circ y_1\right] \leq \mathbb{E}_{P_\top}\left[r \circ y_2\right]. \tag{2.44}$$

14. The weaker form allows to prefer acts relying on known probabilities as compared to those relying on unknown probabilities. The sure thing principle is violated in these circumstances. Such a situation appears in Ellsberg's paradox.

Going back to the estimation case where all mistakes are equally penalized and all correct estimations are equally rewarded, expected upper rewards are plausibilities of singletons and therefore, the act with maximal upper expected reward is obtained for $a^* = \underset{a \in \Theta}{\arg\max}\, pl\left(a\right)$.

2.11 Conclusions

In spite of an increased complexity as compared to probabilistic calculus, belief function calculus offers a variety of operations allowing this framework to be operational for many tasks involving decision making under uncertainty. The comprehensive set of evidential tools allows practitioners to build belief functions from data, update them when necessary, perform inference and draw conclusions in a principled way.

In the next chapters, we will see other tools that allow the set of mass functions (mass space) to acquire different kinds of structure: order theoretic structure, metric structure and algebraic structure. We will also see that these three kinds of structures have intricate connections and that we can elaborate on them to derive other evidential calculus rules.

Belief spaces: from posets to lattices

3

In the previous chapter, we have encountered a number of situations where one must select a belief function from a set of candidate ones. This is notably the case when applying the least commitment principle which consists in choosing the least informative function among those examined. But on what basis can we select the solution function? We need a tool allowing to sort belief functions.

The simplest kind of such tools are pre-orders and partial orders. Endowing the belief space with these binary relations yields one of the crudest order theoretic structure and in case of a partial order, this structure is known as a poset[1]. In this chapter, we will review pre-orders and partial orders for belief functions and see how they can be used to compare two belief functions. We will also see some approaches allowing to compare belief functions relatively to a finite family of belief functions. Finally, we will see that the poset structure can be augmented to obtain a lattice for separable belief functions only. Also, the mass space structure generalizes the boolean algebra structure of set theory but the mass space is not a boolean algebra itself.

1. a short for partially ordered set.

3.1 Pairwise comparison of belief functions

A first way to characterize the underlying structure of the mass space $\mathcal{M}$ is to look for binary relations allowing pairwise comparisons of mass functions, and in particular partial orders. In this section, we give a short reminder on binary relations and corresponding structures. Afterwards, we present some frequently used partial orders in the theory of belief functions. Those orders rely either on the amount or on the coherence of the encoded information in the belief functions.

3.1.1 *Pre-orders, partial orders and all that*

A **binary relation** is a subset R of $\mathcal{M} \times \mathcal{M}$ and if a pair (m_1, m_2) is in R, we understand that m_1 is connected to m_2. A binary relation can thus also be seen as an oriented graph. One usually denotes alternatively $(m_1, m_2) \in R$ or $m_1 R m_2$.

Definition 17. A **pre-order** $\preceq$ is a binary relation on $\mathcal{M}$ with the following properties:

- reflexivity: $m \preceq m$ for all $m \in \mathcal{M}$,
- transitivity: for any triplet $(m_1, m_2, m_3) \in \mathcal{M}^3$ such that $m_1 \preceq m_2$ and $m_2 \preceq m_3$, we have $m_1 \preceq m_3$.

Definition 18. A **partial order** $\sqsubseteq$ is a pre-order with the antisymmetry property:

- for any pair (m_1, m_2) such that $m_1 \sqsubseteq m_2$ and $m_2 \sqsubseteq m_1$, we have $m_1 = m_2$.

The pair $(\mathcal{M}, \sqsubseteq)$ is called partially ordered set or **poset** for short. If in addition, each pair (m_1, m_2) in $\mathcal{M}$ is comparable, i.e. we have either $m_1 \sqsubseteq m_2$ or $m_2 \sqsubseteq m_1$, then $\sqsubseteq$ is a **total order** but no such order can be defined for belief functions (or other related set functions) if we want it to generalize set inclusion.

Using partial orders, one can define the following concepts:

- An element m of $\mathcal{M}$ such that $m \sqsubseteq m'$ for any $m' \in \mathcal{M}_0 \subset \mathcal{M}$ is called a **lower bound** of $\mathcal{M}_0$.
- Conversely, an element m of $\mathcal{M}$ such that $m' \sqsubseteq m$ for any $m' \in \mathcal{M}_0 \subset \mathcal{M}$ is called an **upper bound** of $\mathcal{M}_0$.
- A lower bound m of a subset $\mathcal{M}_0 \subset \mathcal{M}$ such that for any lower bound m' of $\mathcal{M}_0$, $m' \sqsubseteq m$ is called a **greatest lower bound**.
- An upper bound m of a subset $\mathcal{M}_0 \subset \mathcal{M}$ such that for any upper bound m' of $\mathcal{M}_0$, $m \sqsubseteq m'$ is called a **least upper bound**.
- The **minimum element** (a.k.a. bottom) $\perp$ of a poset (if it exists) is the only element that is a lower bound of any subset of $\mathcal{M}$.
- The **maximum element** (a.k.a. top) $\top$ of a poset (if it exists) is the only element that is an upper bound of any subset of $\mathcal{M}$.
- A **minimal element** m of a subset $\mathcal{M}_0 \subset \mathcal{M}$ is such that $m' \not\sqsubseteq m$ for any other $m' \in \mathcal{M}_0$.
- A **maximal element** m of a subset $\mathcal{M}_0 \subset \mathcal{M}$ is such that $m \not\sqsubseteq m'$ for any other $m' \in \mathcal{M}_0$.

From their definitions, least upper bounds and greatest lower bounds are unique when they exist.

Finally, a poset such that it is possible to find a least upper bound and a greatest lower bound for any pair (m_1, m_2) is called a **lattice**. Figure 3.1 gives an example of lattice which is a finite subset of $\mathcal{M}$.

Such qualitative relations are adapted tools to express links between belief functions. In the case of partial orders, the relation $\sqsupseteq$ can express a logical or semantic notion of "more than". For instance, one can define a partial order to model notions such as of "more informative than", "more inconsistent than", "more consonant than", etc. without the need to make every pair of belief functions comparable. This has the advantage to give clear semantics to the relation.

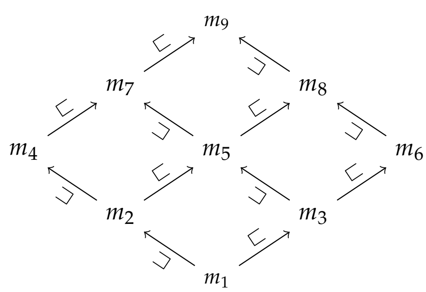


Figure 3.1: Example of a lattice. Let $\mathcal{M}_0 = \{m_1; \ldots; m_9\} \subsetneq \mathcal{M}$. We see that we can find a greatest lower bound and a least upper bound for any pair of elements in $\mathcal{M}_0$. This is no longer true for $\mathcal{M}_0 \setminus \{m_1\}$ for instance as $(m_2; m_3)$ has no lower bound at all inside $\mathcal{M}_0 \setminus \{m_1\}$. However m_2 and m_3 are two minimal elements of $\mathcal{M}_0 \setminus \{m_1\}$.

3.1.2 *Informative partial orders for belief functions*

Partial orders comparing the amount of informative contents formalize the notion of inclusion between belief functions, and play an essential role in approximation problems. Indeed, it is often desirable to compute an approximation of a general belief function such that the approximation belongs to a subclass of belief functions with smallest complexity and fewer focal elements. When the approximation is included in the initial belief function, one speaks of inner approximation. When the approximation contains the initial belief function, one speaks of outer approximation. To define such notions of inclusion, we then have to rely on partial orders. Several definitions (c.f. Yager 1986; Dubois and Prade 1986; Denœux 2008) are found in the literature:

i) m_1 is **pl-included** in m_2, denoted $m_1 \sqsubseteq_{pl} m_2$, if $pl_1(A) \leq pl_2(A)$ for all $A \in 2^\Theta$, where pl_i is the plausibility induced by m_i.
ii) m_1 is **b-included** in m_2, denoted $m_1 \sqsubseteq_b m_2$, if $b_1(A) \geq b_2(A)$ for all $A \in 2^\Theta$, where b_i is the implicability induced by m_i.
iii) m_1 is **bel-included** in m_2, denoted $m_1 \sqsubseteq_{bel} m_2$, if $bel_1(A) \geq bel_2(A)$ for all $A \in 2^\Theta$, where bel_i is the belief induced by m_i.
iv) m_1 is **q-included** in m_2, denoted $m_1 \sqsubseteq_q m_2$, if $q_1(A) \leq q_2(A)$ for all $A \in 2^\Theta$, where q_i is the commonality induced by m_i.
v) m_1 is **w-included** in m_2, denoted $m_1 \sqsubseteq_w m_2$, if $w_1(A) \leq w_2(A)$ for all $A \in 2^\Theta$, where w_i is the conjunctive weight function induced by m_i.
vi) m_1 is **v-included** in m_2, denoted $m_1 \sqsubseteq_v m_2$, if $v_1(A) \geq v_2(A)$ for all $A \in 2^\Theta$, where v_i is the disjunctive weight function induced by m_i.
vii) m_1 is **π-included** in m_2, denoted $m_1 \preceq_\pi m_2$, if $\pi_1(a) \leq \pi_2(a)$ for all $a \in \Theta$, where π_i is the contour function induced by m_i.
viii) A function m_1 is a **specialization** of m_2, denoted $m_1 \sqsubseteq_s m_2$, if there exist a non-negative $N \times N$ matrix $\mathbf{S} = [S(i,j)]$ such that

$$\text{for } j = 1, \ldots, N, \quad \sum_{i=1}^{N} S(i,j) = 1,$$
$$S(i,j) > 0 \Rightarrow A_i \subseteq A_j,$$
$$\text{for } i = 1, \ldots, N, \quad \sum_{j=1}^{N} m_2(A_j) S(i,j) = m_1(A_i).$$

The term $S(i,j) > 0$ is the mass proportion of the focal set A_j that "flows down" to focal set A_i. The order in which subsets are indexed is arbitrary.

A subclass of specialization matrices are **Dempsterian specialization matrices**. A matrix $\mathbf{D}_i$ is a Dempsterian specialization matrix if it is a specialization matrix and if for any $E_k \subseteq E_j$, one has $D_i(k,j) = m_{i|E_j}(E_k)$ for some mass function m_i. Now, one writes $m_1 \sqsubseteq_d m_2$ if m_1 is a specialization of m_2 relying on a Dempsterian matrix $\mathbf{D}_0$ which actually means that $m_1 = m_0 \circledcirc m_2$[2].

The strict version $\sqsubset$ of these inclusions is simply obtained when the inequalities are strict for at least one subset of Θ, and in the case of d- or s-inclusion when $S(i,j) > 0$ for at least one pair $A_i \subsetneq A_j$[3].

All these concepts extend classical set inclusion, in the sense that if $A \subseteq B$, then $m_A \sqsubseteq_y m_B$ for any $y \in \{v, w, d, pl, b, bel, q, s, \pi\}$ (for w- and v-inclusions see comments in the margin). It is well known that set-inclusion is a partial order on 2^Θ. Likewise, these binary relations are partial orders on $\mathcal{M}$ except $\preceq_\pi$ which is just a pre-order. These partial orders are not total orders in the sense that we may have $m_1 \not\sqsubseteq_y m_2$ and $m_2 \not\sqsubseteq_y m_1$.

Due to the duality between pl and b, pl- and b-inclusions are equivalent notions. When $m(\emptyset) = 0$, then pl-, b- and bel-inclusion coincide, and $m_1 \sqsubseteq_{pl} m_2$ is then equivalent to inclusion of their respective cores: $\mathcal{P}_1 \subseteq \mathcal{P}_2$.

The following implications hold between these notions of inclu-

The partial order $\sqsubseteq_w$ allows only non-dogmatic mass function comparisons while $\sqsubseteq_v$ allows only subnormal mass function comparisons. The definitions of the conjunctive and disjunctive weight functions can be extended to categorical mass functions. Yet, several conjunctive and disjunctive decompositions are possible for categorical mass functions. We retain the following ones:

- The conjunctive weight function w_B of m_B is such that
$$w_B(A) = \begin{cases} 0 & \text{if } A \supseteq B \\ 1 & \text{otherwise} \end{cases}. \quad (3.1)$$
- The disjunctive weight function v_B of m_B is such that
$$v_B(A) = \begin{cases} 0 & \text{if } A \subseteq B \\ 1 & \text{otherwise} \end{cases}. \quad (3.2)$$

We consider that the partial orders are also valid for this subclass of mass functions.

2. Dempsterian specialization matrices are special cases of α-conjunctive matrices.

3. For d-inclusion, this is true whenever the Dempsterian matrix is not the identity matrix, i.e. it is not in correspondence with the vacuous mass function.

sion (Dubois and Prade 1986):

$$\left.\begin{array}{c} m_1 \sqsubseteq_w m_2 \\ m_1 \sqsubseteq_v m_2 \end{array}\right\} \Rightarrow m_1 \sqsubseteq_d m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \left\{\begin{array}{c} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{array}\right\} \Rightarrow m_1 \preceq_\pi m_2. \tag{3.3}$$

Since v-, w-, d-, s-, pl- and q-inclusion are antisymmetric, that is $m_1 \sqsubseteq_y m_2$ and $m_1 \sqsupseteq_y m_2$ implies $m_1 = m_2$ for $y \in \{v, w, d, s, pl, q\}$, we also have

$$\left.\begin{array}{c} m_1 \sqsubset_w m_2 \\ m_1 \sqsubset_v m_2 \end{array}\right\} \Rightarrow m_1 \sqsubset_d m_2 \Rightarrow m_1 \sqsubset_s m_2 \Rightarrow \left\{\begin{array}{c} m_1 \sqsubset_{pl} m_2 \\ m_1 \sqsubset_q m_2 \end{array}\right. \tag{3.4}$$

3.1.3 *Coherence pre-orders for belief functions*

We can think about other orders related to important notions in evidence theory. For instance, an important notion within evidence theory is the consistency of pieces of information encoded inside a mass function, from which follows the notion of conflict between sources. There are two main ways to evaluate the consistency of a mass function, a strong and a weak one. The strong consistency measure is given by

$$\Xi(m) = \max_{a \in \Theta} \pi(a) \tag{3.5}$$

and the weak consistency measure by

$$\kappa(m) = m(\emptyset). \tag{3.6}$$

Destercke and Burger 2013 showed that $\Xi(m) = 1$ iff the focal elements of m do not have an empty intersection. This means that if we continue to receive evidence supporting the same events then we will converge to a categorical mass function and never experience total conflict[4]. Concerning weak consistency, κ can be understood as the amount of support given to incompatible events after one applied the conjunctive rule for instance.

Each measure defines a pre-order on the mass space:

i) Mass m_1 is *strongly less consistent* than m_2, denoted $m_1 \preceq_{sc} m_2$, if $\Xi(m_1) \leq \Xi(m_2)$.

ii) Mass m_1 is *weakly less consistent* than m_2, denoted $m_1 \preceq_{wc} m_2$, if $\kappa(m_1) \geq \kappa(m_2)$.

Again, strict inequalities yield strict relations. $\preceq_{sc}$ and $\preceq_{wc}$ are total pre-orders, as any two elements can be compared, but distinct elements m_1, m_2 may be equally consistent.

4. Let $\{m_i\}_{i=1}^{\ell}$ denote a family of mass functions such that they have the same focal elements $\mathcal{F} = \{E_1, \ldots, E_k\}$. Let E_0 denote the non empty intersection of the focal sets: $E_0 = \bigcap_{j=1}^{k} E_j$. We have

$$\bigoplus_{i=1}^{\ell} m_i \underset{\ell \to \infty}{\longrightarrow} m_{E_0}.$$

3.1.4 *Specificity pre-orders for belief functions*

There several circumstances[5] in which it is desirable to evaluate to what extent a belief function support small, precise hypotheses or large, imprecise hypothesis. In other words, we need to evaluate how large are focal elements.

5. See 5.4.1 where specificity measures are used to build combination rules

The specificity (Dubois and Prade. 1985) of a mass function is defined as

$$\mathrm{Spe}(m) = \sum_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} m(B) \log_2(|B|). \tag{3.7}$$

This measure also yields a pre-order for belief functions such that $m_{\{a\}}$ achieves a minimal null value and m_Θ achieves a maximal value.

Another possibility is to resort to the expected cardinality as if we were dealing with an ontic random set:

$$\mathrm{Card}\,(m) = \sum_{\substack{B\subseteq\Theta\\ B\neq\emptyset}} m\,(B)\,|B|. \tag{3.8}$$

This measure achieves a minimal null value when $m = m_\emptyset$. The maximal value is achieved when $m = m_\Theta$: $\mathrm{Card}\,(m_\Theta) = n$. This measure generalizes anyway set cardinality as $\mathrm{Card}\,(m_A) = |A|$.

The notion of specificity is also featuring informative content and consequently one can wonder which specificity pre-orders refine informative which partial orders. For example, it can be conjectured that the pre-order spanned by the expected cardinality refines $\sqsubseteq_s$. Studying these relationships is an interesting research direction.

Also, under an open world assumption, all three mentioned types of orders (informative, coherence and specificity) are somewhat overlapping notions. Indeed, we have $\lambda m_\emptyset + (1-\lambda)m_B \sqsubset_d m_B$ for any $B \neq \emptyset$ and any $\lambda \in (0;1]$. We also have that $\lambda m_\emptyset + (1-\lambda)m_B$ is more specific than m_B according to both definitions presented in the above paragraphs. Especially when B is a singleton, it hard to envisage anything more informative or more specific than m_B. Obviously, $\lambda m_\emptyset + (1-\lambda)m_B$ is less coherent than m_B therefore the specificity pre-orders and the informative partial orders seem to grasp several aspects of mass function contents.
It is not mathematically speaking incorrect to use specificity pre-orders and informative partial orders with unnormalized mass functions but one should be aware that this is going beyond their initial purposes and consequently the interpretation of the end results drawn from them must be handled with extra care.

3.2 Comparisons within families of belief functions

In an information fusion context, one may be given several pieces of information each of them summarized as a source $(\Omega, \sigma_\Omega, \mu_i, \Gamma_i)$. A family of belief functions is induced by these sources and for each pair in this family, we can use the previously mentioned pre-orders and partial orders to draw comparisons. Yet, one can wonder to what extent family based comparisons would be more interesting.

In this vein, the literature focuses on coherence pre-orders. Indeed, a very challenging open problem is the detection of irrelevant or deceptive pieces of information. A usual prerequisite postulate to tackle this problem is that belief functions with small "intra-family" coherence are a minority. If we accept this postulate, the belief functions we want to discard are significantly different from the rest of the family regarding the encoded information carried by the functions. In this context, we are now looking for a way to compare belief functions relatively to the family to which they belong.

Mining singular belief functions can be performed in several ways. Schubert 1996, 2010 proposed a criterion c_i, called the degree of falsity, that evaluates the contribution of each individual mass function m_i in the computation of the degree of conflict obtained after combining all functions in $\mathcal{A} = \{m_1, \ldots, m_\ell\}$. This degree of conflict is denoted by $\kappa_{\mathcal{A}} = \left(\bigcap_{i=1}^{\ell} m_i \right) (\emptyset)$. The degree of falsity is obtained as

$$c_i = \frac{\kappa_{\mathcal{A}} - \kappa_{\mathcal{A} \setminus \{m_i\}}}{1 - \kappa_{\mathcal{A} \setminus \{m_i\}}} \tag{3.9}$$

where $\mathcal{A} \setminus \{m_i\}$ is the set difference of $\mathcal{A}$ and $\{m_i\}$. It is clear that if m_i is the only mass function advocating for a particular solution, there will be a significant drop from $\kappa_{\mathcal{A}}$ to $\kappa_{\mathcal{A} \setminus \{m_i\}}$. Consequently, this very singular mass function will have a large degree of falsity.

Martin et al. 2008 have also introduced several criteria, called conflict measures, evaluating the conflict provoked by a mass function as compared to a set of mass functions. These criteria are defined using a distance d_J between belief functions introduced by Jousselme et al. 2001:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2} (\mathbf{m_1} - \mathbf{m_2})^T \cdot \mathbf{D} \cdot (\mathbf{m_1} - \mathbf{m_2})} \tag{3.10}$$

with $\mathbf{m}_i$ the vector form of the mass function m_i and $\mathbf{D}$ a $N \times N$ matrix whose elements are $D(A, B) = |A \cap B| \,/\, |A \cup B|$. Martin et al. propose then the following conflict measures Conf_i:

$$\mathrm{Conf}_i = \frac{1}{\ell - 1} \sum_{\substack{j=1 \\ i \neq j}}^{\ell} d_J(m_i, m_j), \tag{3.11}$$

$$\text{or } \mathrm{Conf}_i = d_J(m_i, m_\bullet), \tag{3.12}$$

with $m_\bullet$ the combination of mass functions in $\mathcal{A} \setminus \{m_i\}$. $m_\bullet$ can be obtained using different combination rules or by using the mean. Furthermore, the authors propose to tune this measure using some function f:

$$f(\mathrm{Conf}_i). \tag{3.13}$$

The heuristic for choosing f indicated by the authors is $f(x) = 1 - (1 - x^\lambda)^{1/\lambda}$ and $\lambda = 1.5$.

Klein and Colot 2011 introduced another criterion denoted ξ_i which shares the same philosophy as the degree of falsity. It is defined as

$$\xi_i = \frac{1}{\binom{\ell-1}{\ell_0-1}} \sum_{\substack{s \subseteq \mathcal{A}, \\ m_i \in s, \\ |s| = \ell_0}} \left[\kappa_s - \kappa_{s \setminus \{m_i\}} \right] \tag{3.14}$$

(personal contribution)

with $\ell_0 = \min \{|s|, s \subset \mathcal{A} \text{ such that } \kappa_s > 0\}$ the size of the smallest subset of the family $\mathcal{A}$ with a positive degree of conflict. The binomial coefficient $\binom{\ell-1}{\ell_0-1}$ guarantees that ξ_i is upper bounded by 1. A known result is that for any mass function m and any family $\mathcal{A}$,

$$\kappa_{\mathcal{A}} \leq \kappa_{\mathcal{A} \cup \{m\}}. \tag{3.15}$$

In other words, the degree of conflict can only increase when a new mass function comes into play. We call this phenomenon the **curse of conflict**.

From this result, it follows that ξ_i is non negative. Most of the time, the integer ℓ_0 is equal to 2 but it may happen that no pairwise conflict is observed while a positive degree of conflict is obtained for a triplet of mass functions. When $\ell_0 = 2$, there are $\ell - 1$ terms in the summation of equation (3.14) which becomes

$$\xi_i = \frac{1}{\ell - 1} \sum_{m_j \in \mathcal{A} \setminus \{m_i\}} \left[\kappa_{\{m_i;m_j\}} - \kappa_{\{m_j\}} \right]. \tag{3.16}$$

If in addition all mass functions are normal, then ξ_i is proportional to the average of pairwise degrees of conflict involving function m_i:

$$\xi_i = \frac{1}{\ell - 1} \sum_{m_j \in \mathcal{A} \setminus \{m_i\}} \kappa_{\{m_i;m_j\}}. \tag{3.17}$$

The computation of ξ_i in general may seem computationally more demanding. The computation load can be reduced by viewing $\kappa_{\mathcal{A}}$ as a function of discounting rates applied to each member of $\mathcal{A}$. Let α_i denote the discounted rate applied to function m_i. These rates are concatenated in a vector $\boldsymbol{\alpha}$ and $\kappa_{\mathcal{A}}(\boldsymbol{\alpha})$ denotes the degree of conflict of the family $\left\{m_1^{\alpha_1}, \ldots, m_\ell^{\alpha_\ell}\right\}$. We have

$$\xi_i = - \lim_{k \to \infty} \frac{2^{k(\ell-1)}}{\binom{\ell-1}{\ell_0-1}} \times \frac{\partial \kappa_{\mathcal{A}}}{\partial \alpha_i} \left(\left[1 - \left(\frac{1}{2} \right)^k \right] \mathbf{1} \right). \tag{3.18}$$

Consequently, when k is large, ξ_i can be approximated by a numerical partial derivative of the degree of conflict induced by identically discounted mass functions. So the cost of computing ξ_i is of the same order as the computation of two degrees of conflict between ℓ mass functions. This approximation stems from the following result which we call **conflict decomposition** [6]:

Proposition 1. *(Conflict decomposition)* $\forall \mathcal{A} \subset \mathcal{M}, |\mathcal{A}| = \ell > 1,$

(personal contribution)

$$\kappa_{\mathcal{A}} \left(\frac{1}{2} \mathbf{1} \right) = \frac{1}{2^\ell} \sum_{\substack{s \subseteq \mathcal{A}, \\ s \neq \emptyset}} \kappa_s. \tag{3.19}$$

This results states that the global conflict arising from $\mathcal{A}$ (after discounted each member by $\frac{1}{2}$) is proportional to the average of degrees of conflict arising from the sub-families of $\mathcal{A}$.

Any of the criteria presented in this section yields a total pre-order for the family $\mathcal{A}$ by using the usual order for reals on the criteria values. When $\ell = 2$, the criterion ξ_i subsumes the weak consistency measure. In practice, the criterion ξ_i has a more stable behavior w.r.t. the proportion of singular sources and is less dependent on ℓ (the size of $\mathcal{A}$) as illustrated in the following example.

Example 14. Suppose a family of mass functions $\mathcal{A}$ is the union of two sub-families $\mathcal{A}_1$ and $\mathcal{A}_2$ with respective cardinalities ℓ_1 and ℓ_2. Suppose also that $\mathcal{A}_1 = \left\{m_A^x, \ldots, m_A^x\right\}$ and $\mathcal{A}_2 = \left\{m_B^x, \ldots, m_B^x\right\}$ with $x \in [0;1]$ and $A \cap B = \emptyset$. Let $r = \frac{\ell_1}{\ell}$ denote the ratio of functions supporting A in $\mathcal{A}$.

6. Shafer 1976, theorem 3.5 is another conflict decomposition result.
If $\mathcal{A} = \{m_1, \ldots, m_\ell\}$ denotes a family of ℓ normal mass functions that can be combined by Dempster's rule denoted $m_{\mathcal{A}} = \bigoplus_{i=1}^{\ell} m_i$, then for any normal mass function $m_{\ell+1}$, we have $\kappa_{\mathcal{A} \cup \{m_{\ell+1}\}} = \kappa_{\mathcal{A}} + \kappa_{\{m_{\mathcal{A}}, m_{\ell+1}\}}$.

It can be proved (see Klein and Colot 2011) that in this case, ξ_i is linear w.r.t. r and does not depend on $\ell = \ell_1 + \ell_2$. The linearity of ξ_i as well as the non-linearity of c_i and Conf_i[7] can be seen on figure 3.2. The curves displayed in this figure are obtained for $\ell = 20$ mass functions. The variations of the criteria are given for several values of x: $x \in \{0; 0.1; \ldots 0.9\}$.

When $x = 0$ and $r = 0$, there is only one function in $\mathcal{A}_1$ giving full support to A while those in $\mathcal{A}_2$ give full support to B. There is consequently a clash between the messages encoded in the two subgroups and all three criteria reach a maximal value of 1.

Obviously, criterion c_i dynamics are concentrated around $r = 0.5$ where the majority switches from $\mathcal{A}_2$ to $\mathcal{A}_1$ while the two other criteria are decreasing more gradually. It can be noted that when $x = 0$, the degree of falsity c_i cannot be formally computed as we have $c_i = \frac{0}{0}$.

Finally, we can also mention a recent alternative approach by Schubert 2016 which relies on entropy[8] discrepancies instead of conflict discrepancies. Indeed, if removing a mass function m_i from a combination process yields a mass function with a smaller entropy, then it can be postulated that m_i is singular. The author shows that this new criterion (or a combined version of these two) outperforms c_i in the detection of deceptive messages.

7. The version of criterion Conf_i used in this experiment is computed as Jousselme distance between the examined function and the average of the remaining ones. According to our experience, this version achieves the best performances in this case.

8. The notion of entropy for a belief function whose core is $\mathcal{P}$ can be defined as $\max_{P \in \mathcal{P}} H(P)$ where H is the usual entropy function for probability distributions.

3.3 Lattices of belief functions

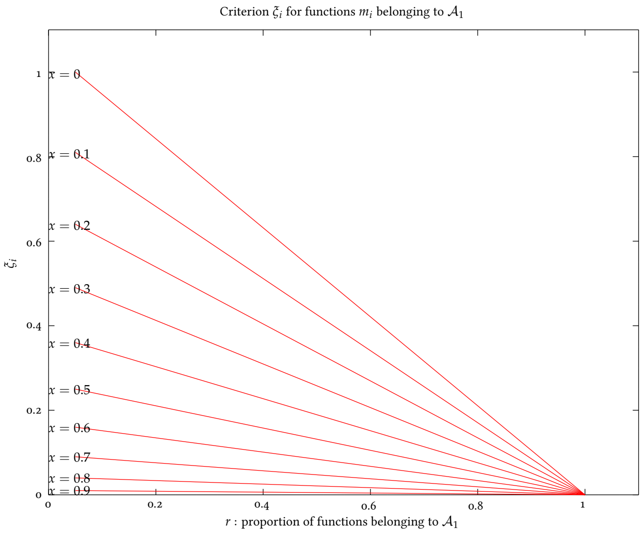
A legitimate question coming to mind is: can the mass space, regarded as a poset, be granted a richer structure by endowing it with additional operations? If for a given partial order $\sqsubseteq$, we could find a way to compute the unique maximal element among mass functions upper bounded by both m_1 and m_2 as well as the unique minimum element among mass functions lower bounded by both m_1 and m_2, then the mass space would have a **lattice** structure. These two operations are called in general a conjunction and a disjunction. In the following paragraphs, we will examine different candidate conjunction/disjunction operations but none of them achieve a lattice structure for $\mathcal{M}$. Only one reported work allows to derive a lattice structure for a subset of the mass space.

First, we see that the pair of α-junctions $(\textcircled{\cap}^\alpha ; \textcircled{\cup}^\alpha)$ are unfortunately not eligible as conjunction/disjunction operations. Indeed, from the latticial definition of a conjunction and a disjunction, we deduce that such operations are mandatorily idempotent[9]. This means that if $m_1 = m_2$, then both their conjunction and disjunction are again m_1. However, when $\alpha = 1$, for any partial order $\sqsubseteq_x$, $x \in \{w, d, s, q, pl\}$, we have that the quadruplet $(\mathcal{M}, \sqsubseteq_x, \textcircled{\cap}, \textcircled{\cup})$ generalizes[10] $(2^\Theta, \subseteq, \cap, \cup)$ which is a distributive lattice.

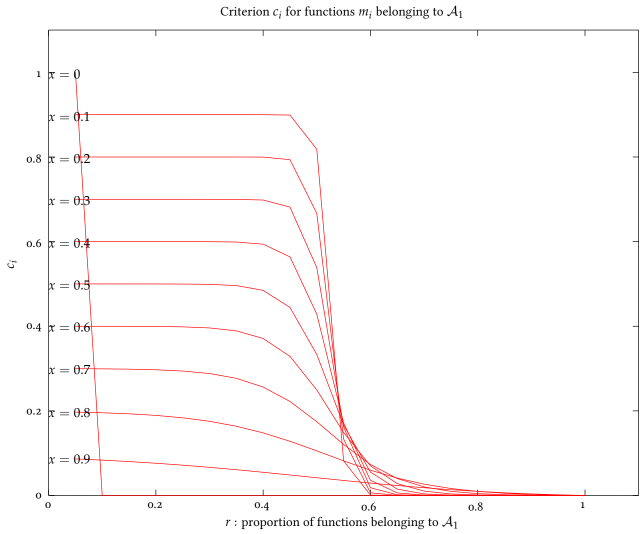
Pairs of α-junctions are not appropriate conjunction/disjunction pairs but maybe we can still find greatest lower bounds and least upper bounds for any pair $(m_1; m_2) \in (\mathcal{M}, \sqsubseteq_x)$, for some $x \in \{w, d, s, pl, q\}$ anyway. This is also untrue because, in general maximal lower bounds (or minimal upper bounds) are not unique (Dubois et al. 2001). This has to be proved for each partial order individually. Indeed, let $\mathcal{M}_x(m_1)$ denote the set of

9. If $m_1 = m_2$, there is only one maximal mass function upper bounded by m_1, i.e. m_1 itself. Likewise, m_2 is the only minimal mass function lower bounded by m_2.

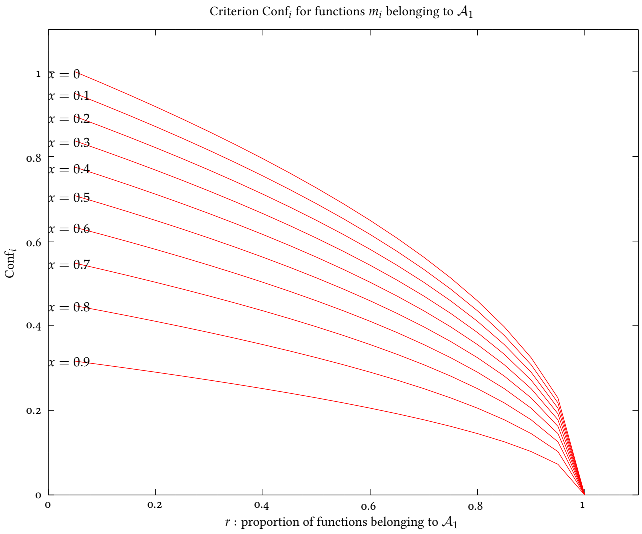
10. We mean that using $\sqsubseteq$, $\textcircled{\cap}$ or $\textcircled{\cup}$ on categorical belief functions is tantamount to using $\subseteq$, $\cap$ or $\cup$ on the subsets that are the focal elements of the categorical belief functions.



(a)



(b)



(c)

Figure 3.2: Variability of coherence criteria relatively to a family $\mathcal{A}$ of belief functions. 3 criteria are examined: ξ_i (a), c_i (b) and Conf_i (c). The family $\mathcal{A}$ is partitioned into 2 subfamilies containing simple mass functions. Each function assigns a mass $1 - x$ to either set A or B depending on which group it belongs to. The sets A and B are not intersecting. The criteria for some mass function m_i supporting A are shown as functions of r the proportion of mass functions supporting A for several values of x: $x \in \{0; 0.1; \dots 0.9\}$.

mass functions that are x-included in m_1:

$$\mathcal{M}_x\left(m_1\right) = \left\{m \in \mathcal{M} | m \sqsubseteq_x m_1\right\}.$$

The implications (3.3) imply

$$\mathcal{M}_w\left(m_1\right) \subsetneq \mathcal{M}_d\left(m_1\right) \subsetneq \mathcal{M}_s\left(m_1\right) \subsetneq \begin{cases} \mathcal{M}_q\left(m_1\right) \\ \mathcal{M}_{pl}\left(m_1\right) \end{cases}, \tag{3.20}$$

and consequently a maximal lower bound for one partial order is not necessarily also a maximal lower bound for another. We only provide a counter example for $\sqsubseteq_d$.

Example 15. Suppose $\Theta = \{a; b\}$. We want to find a greatest lower bound w.r.t. $\sqsubseteq_d$ for the pair $(m_1; m_2)$ and

$$m_1 = \frac{1}{3} m_{\{b\}} + \frac{2}{3} m_\Theta, \tag{3.21}$$

$$m_2 = \frac{1}{3} m_{\{a\}} + \frac{2}{3} m_\Theta. \tag{3.22}$$

Observe that

$$m_{1|\{a\}} \circledcap m_2 = \left(\frac{1}{3} m_\emptyset + \frac{2}{3} m_{\{a\}}\right) \circledcap m_2, \tag{3.23}$$

$$= \frac{1}{3} m_\emptyset \circledcap m_2 + \frac{2}{3} m_{\{a\}} \circledcap m_2, \tag{3.24}$$

$$= \frac{1}{3} m_\emptyset + \frac{2}{3} m_{\{a\}}, \tag{3.25}$$

$$= m_{1|\{a\}} \tag{3.26}$$

and thus $m_{1|\{a\}} \in \mathcal{M}_d\left(m_1\right) \cap \mathcal{M}_d\left(m_2\right)$. Suppose there exist $m' \in \mathcal{M}_d\left(m_1\right) \cap \mathcal{M}_d\left(m_2\right)$ and $m_{1|\{a\}} \sqsubseteq_d m'$, then we have $m' \circledcap m'' = m_{1|\{a\}}$ for some $m'' \in \mathcal{M}$. In particular, since $m' \in \mathcal{M}_d\left(m_2\right)$, there exist some m''' such that $m' = m_2 \circledcap m'''$ and it follows that

$$m_2 \circledcap m''' \circledcap m'' = m_{1|\{a\}}. \tag{3.27}$$

By conditioning on both sides by $\{a\}$, we obtain

$$m_{2|\{a\}} \circledcap m''' \circledcap m'' = m_{1|\{a\}}. \tag{3.28}$$

Since $m_{2|\{a\}} = m_{\{a\}}$, the lefthanded term is either equal to $m_{\{a\}}$ or to $m_\emptyset$, none of which are equal to $m_{1|\{a\}}$. This contradiction implies that $m_{1|\{a\}}$ is a maximal element of $\mathcal{M}_d\left(m_1\right) \cap \mathcal{M}_d\left(m_2\right)$. By notation symmetry, so is $m_{2|\{b\}}$. Since $m_{2|\{b\}} \not\sqsubseteq_d m_{1|\{a\}}$ and $m_{1|\{a\}} \not\sqsubseteq_d m_{2|\{b\}}$, there is no greatest lower bound for the pair $(m_1; m_2)$.

In the belief function literature, conjunctions and disjunctions for belief functions were introduced only for subclasses of belief functions. Kennes 1991 placed the theory of belief functions under the umbrella of **category theory**. In short, category theory is a meta-framework relying essentially on oriented graphs. States of beliefs about the variable of interest are nodes and by integration of new evidence we move to another node. This transition is embodied by an arrow relating the two nodes. Formally speaking, since several arrows can relate two nodes, then we obtain a multigraph. Examples of categories are:

- power sets: the nodes are subsets and an arrow is a pair of subsets such that one of them is included in the other one,
- probabilities: the nodes are probability distributions on a given abstract space and arrows are subsets of this space. Indeed, the usual transition[11] is conditioning which relies on a subset.
- belief functions: the nodes are belief functions and the arrow are belief functions as well. More precisely, an arrow is a belief function whose combination by Dempster's rule with the source node yield the target node.

11. Other probability kinematics are possible for instance using the total probability theorem. Suppose P_s and P_t are respectively the source and target distributions defined on Θ. Let $\mathcal{B} = (B_i)_{i=1}^{n_\mathcal{B}}$ denote a partition of Θ. A distribution P_a on the σ-field spanned by $\mathcal{B}$ is an arrow linking source and target distributions by

$$P_t(A) = \sum_{i=1}^{n_\mathcal{B}} P_s(A|B_i)\, P_a(B_i)\,. \quad (3.29)$$

The above comments highlight the fact that belief functions can be either interpreted as static objects (states of beliefs) or dynamic ones (transitions between states). An essential point in category theory is that two arrows can be "plugged" if the target of the first is the source of the second. This yields a notion of composition for arrows and the associativity of compositions is one of the axioms of the theory.

Kennes proves that a conjunction and a disjunction can be defined for separable unnormalized belief functions. We shall denote the set of such function by $\mathfrak{U} \subset \mathcal{M}$. Let m_1 and m_2 denote two unnormalized separable belief functions. Consequently, for $i \in \{1;2\}$,

$$m_i = \bigcirc\!\!\!\!\cap_{B \subseteq \Theta} m_B^{w_i(B)}.$$

We can define the following conjunction/disjunction operations

$$m_1 \overline{\wedge} m_2 = \bigcirc\!\!\!\!\cap_{B \subsetneq \Theta} m_B^{w_1(B) \wedge w_2(B)} \text{ (Conjunction)}, \quad (3.30)$$

$$m_1 \underline{\vee} m_2 = \bigcirc\!\!\!\!\cap_{B \subsetneq \Theta} m_B^{w_1(B) \vee w_2(B)} \text{ (Disjunction)}, \quad (3.31)$$

where $\wedge$ is the minimum operator for reals, and $\vee$ the maximum operator. Let m and m' denote two functions in $\mathfrak{U}$. The function m is the source of an arrow whose target is m' if there exist another unnormalized separable mass function m_0 such that $m \overline{\wedge} m_0 = m'$. This induces a partial order on $\mathfrak{U}$. We see that for non-dogmatic functions in $\mathfrak{U}$, this partial order coincides with $\sqsubseteq_w$ therefore, this partial order is denoted by $\sqsubseteq_{\tilde{w}}$[12]. If a mass function is $\tilde{w}$-included in m_1 and m_2, then it is also $\tilde{w}$-included in their conjunction as defined above. Similarly, if m_1 and m_2 are $\tilde{w}$-included in a mass function, then their disjunction as defined above is also $\tilde{w}$-included in this function. In conclusion, $(\mathfrak{U}, \sqsubseteq_{\tilde{w}}, \overline{\wedge}, \underline{\vee})$ is a lattice.

12. $m_1 \sqsubseteq_{\tilde{w}} m_2$ iff there is an arrow whose source is m_2 and target is m_1.

In addition, this lattice is distributive because $\wedge$ and $\vee$ are distributive over each other. It is also a bounded lattice whose top and bottom are respectively m_Θ and $m_\emptyset$. Indeed, the conjunctive weight function corresponding to m_Θ is constant one and $w_1(B) \wedge 1 = w_1(B), \forall B$, hence $m_1 \sqsubseteq_{\tilde{w}} m_\Theta$. The decomposition of $m_\emptyset$ is not unique. A possible decomposition for $m_\emptyset$ is obtained using a constant zero weight function. Since $w_1(B) \wedge 0 = 0, \forall B$, we see that $m_\emptyset \sqsubseteq_{\tilde{w}} m_1$.

Building upon Kennes' work, Denœux 2008 extended these results to other large classes of belief functions. Using Smets' decomposition and generalized conjunctive weight functions, for any pair $(m_1; m_2)$ of non-dogmatic belief functions, Denœux's cautious rule $\oslash\!\!\!\wedge$ is given by

$$m_1 \oslash\!\!\!\!\wedge\, m_2 = \bigcirc\!\!\!\!\cap_{B \subsetneq \Theta} m_B^{w_1(B) \wedge w_2(B)} \quad (3.32)$$

The author also introduces the bold rule $\circledvee$, which is defined for any pair $(m_1; m_2)$ of subnormal (non null mass for $\emptyset$) mass functions as

$$m_1 \circledvee m_2 = \bigcirc\!\!\!\!\cup_{A \neq \emptyset} m^A_{v_1(A) \wedge v_2(A)}, \tag{3.33}$$

where m^A_x denotes a mass function such that

$$m^A_x(E) = \begin{cases} 1-x & \text{if } E = A, \\ x & \text{if } E = \emptyset, \\ 0 & \text{otherwise} \end{cases}. \tag{3.34}$$

Now, if $\mathfrak{V}$ denotes the set of subnormal and non-dogmatic mass functions, the next question is: Is $(\mathfrak{V}, \sqsubseteq_w, \circledwedge, \circledvee)$ a lattice? Sadly, the answer is no [13] because the disjunction for $\sqsubseteq_w$ should obviously be

$$\bigcirc\!\!\!\!\cap_{B \subsetneq \Theta} m_B^{w_1(B) \vee w_2(B)}. \tag{3.35}$$

13. Formally speaking the set of non-dogmatic mass functions endowed with $\sqsubseteq_w$ and $\circledwedge$ is a meet-semilattice, while the set of subnormal mass functions endowed with $\sqsubseteq_v$ and $\circledvee$ is a join-semilattice.

Although this operation is defined (notably for unnormalized separable mass functions), it does not yield a bona fide mass function in general (see Denœux 2008 - example 5 for a counter example).

In chapter 5, we will derive other rules that share a similar philosophy as the one presented in this section. There are however not conjunctions or disjunctions in the latticial sense because they are not associative.

Finally, we could also hope that a complement operator for unnormalized separable mass functions could be defined, in which case we obtain a boolean algebra. The complement of a mass function $m \in \mathfrak{U}$ is a mass function $m^c \in \mathfrak{U}$ such that

$$m \overline{\wedge} m^c = m_\emptyset, \tag{3.36}$$

$$\text{and } m \underline{\vee} m^c = m_\Theta. \tag{3.37}$$

We cannot always find such a function in $\mathfrak{U}$ as illustrated by the following simple example.

Example 16. Let $\Theta = \{a; b\}$. We want to find a candidate complement function for the simple function $m^x_{\{a\}}$ (which is of course separable) with $0 < x < 1$. The corresponding conjunctive weight function w is given by

set	$\emptyset$	$\{a\}$	$\{b\}$
w	1	x	1

The conjunctive weight function of m_Θ is given by

set	$\emptyset$	$\{a\}$	$\{b\}$
w_Θ	1	1	1

We deduce that $w^c(\{a\}) = 1$ because we need $w(\{a\}) \vee w^c(\{a\}) = 1$ to have $m \underline{\vee} m^c = m_\Theta$. But this implies

set	$\emptyset$	$\{a\}$	$\{b\}$
$w \wedge w^c$	$w^c(\emptyset)$	x	$w^c(\{b\})$

To obtain $m \overline{\wedge} m^c = m_\emptyset$, we see that the only solution is to choose $w^c(\emptyset) = 0$ while $w^c(\{b\})$ can be set to an arbitrary value, say y. Then, this means that $m^c = m_\emptyset$ but w^c is not in line with our convention (3.1) which states that

set	$\emptyset$	$\{a\}$	$\{b\}$
$w_{\emptyset}$	0	0	0

We have to stick to a given convention otherwise the partial order $\sqsubseteq_{\tilde{w}}$ is not well defined for categorical mass functions. Choosing another convention is not a solution because the decomposition of $m_{\emptyset}$ cannot depend on x.

There is a notion of complement operator in the belief function literature known as the **negation** of a mass function. The negation of a mass function m_i is denoted by $\overline{m}_i$. The function $\overline{m}_i$ is such that $\forall A \subseteq \Theta$, $\overline{m}_i(A) = m_i(A^c)$. Some De Morgan relations can be proved for α-junctions based on mass function negations:

$$\overline{m_1 \circledcap^{\alpha} m_2} = \overline{m_1} \circledcup^{\alpha} \overline{m_2}, \tag{3.38}$$

$$\overline{m_1 \circledcup^{\alpha} m_2} = \overline{m_1} \circledcap^{\alpha} \overline{m_2}, \tag{3.39}$$

for any pair $(m_1; m_2) \in \mathcal{M}^2$ (see in Smets 2002 for a proof). A corollary result is that $(\mathcal{M}, \sqsubseteq_x, \circledcap, \circledcup, \bar{.})$, $x \in \{w, d, s, pl, q\}$, generalizes the usual boolean algebra $\left(2^{\Theta}, \subseteq, \cap, \cup, .^c\right)$.

Generally speaking, the advantage of establishing a boolean algebra structure is that it is highly coupled to propositional logic from which follows semantics and interpretations. Belief functions can however be interpreted in terms of modal logic (see Harmanec et al. 1996).

3.4 Conclusions

As illustrated in this chapter, it is not much difficult to define a partial order for belief functions so as to acquire a poset structure. In addition, all existing partial orders (or pre-orders) are easily interpretable and help to understand some of the logic behind evidential reasoning. In spite of this, there is still room for further interpretability of these tools. For example the informative partial orders allow comparisons of normal mass functions with subnormal ones. When two mass functions are equally consistent, we understand very clearly the conclusions of the partial order in terms of information content. When mass functions have different levels of inconsistency, the comparison is harder to understand as there is no information encoded by $\emptyset$. Inconsistency and information content should perhaps be captured separately and the definitions of informative partial orders may be revised accordingly.

Besides, unsurprisingly, it is a much more difficult task to define conjunction and disjunction operations in the latticial sense. Indeed, a lattice structure is a lot richer than a poset structure. When dealing with a rather general framework such as belief functions, deriving rich structures is highly challenging (if not a dead-end) because the generality of the framework is at the expense of axioms from which structures usually follow.

In chapter 5, we will see that latticial notions of conjunction or disjunction are not unrelated to that of combination rules on which magma or monoid structures rely. These structures remain, however, rather simple too.

4 Belief spaces as metric spaces

In the previous chapter, we have seen how to compare belief functions based on pre-orders and partial orders. Such tools allow to (partially) sort belief functions. Sometimes, sorting is not enough and we would like to quantify the difference between two belief functions. Distances (or metrics) are mathematical tools allowing to obtain such quantitative comparisons.

In this chapter, we will review distance and dissimilarities as part of the theory of belief functions. When endowed with a metric, the belief space is a metric space. But since belief functions live in a simplex embedded in a vector space and have specific semantics, a number of aspects of the theory should be reflected by a belief function metric. We will thus also comment on desirable properties for belief function metrics with respect to combination rules and partial orders.

4.1 Metrics for belief functions

In this section, we will review the most popular metrics (or distances) found in the belief function literature. After recalling basic notions of metric spaces, we present distances between mass vectors. Indeed, the presentation of evidential distances is more convenient under the geometric view of belief functions developed by Cuzzolin 2004, 2008, 2010a, 2010b through several papers. We also present distances between evidential matrices, i.e. α-specialization and α-generalization matrices.

4.1.1 *Distances: general concepts and definitions*

Metrics (or distances) are meant to translate the intuitive notion of *gap* between elements of a given space into mathematically sound objects. Since the work of Fréchet 1906, the universally accepted definition of a distance is the following:

Definition 19. A **distance**, or metric, in an abstract space $\mathcal{H}$ is a mapping $d : \mathcal{H} \times \mathcal{H} \to [0, a]$ with $a \in \mathbb{R}^{+*}$ that satisfies the following properties:
1. Symmetry : $d(e_1, e_2) = d(e_2, e_1)$,
2. Reflexivity : $d(e_1, e_1) = 0$,
3. Separability : $d(e_1, e_2) = 0 \Rightarrow e_1 = e_2$,
4. Triangle inequality : $d(e_1, e_2) \leq d(e_1, e_3) + d(e_3, e_2)$,

for any $e_1, e_2, e_3 \in \mathcal{H}$.

The reflexivity and separability properties are often aggregated in a global property known as definiteness.

When the scalar a is finite, the metric is said to be bounded. If a mapping d possesses a subset of the properties in the above definition, d is in general called a **dissimilarity** function. Table 4.1 gives which properties hold for widely used dissimilarities.

	(Full) Metric	Semi-metric	Quasi-metric	Pseudo-metric
(1) Symmetry	x	x		x
(2) Reflexivity	x	x	x	x
(3) Separability	x	x	x	
(4) Triangle inequality	x		x	x

Table 4.1: Properties of dissimilarities and metrics.

4.1.2 *Evidential distances*

An evidential distance is simply a distance defined on the mass space $\mathcal{M}$. It is not necessary to investigate distances in other belief spaces. By other belief spaces, we mean those spaces whose elements are other set functions that are in bijective correspondence with mass functions. Indeed, let us denote by $\mathcal{E}$ such a space and by f the bijection from $\mathcal{M}$ to $\mathcal{E}$. Then if d is a metric on $\mathcal{E}$, then $d\left(f\left(.\right), f\left(.\right)\right)$ is a metric on $\mathcal{M}$.

In this section, we start with a presentation of some evidential distances obtained by viewing $\mathcal{M}$ as a subset of the vector space $\mathbb{R}^N$. Next, we also present evidential distances obtained from matrices that are in one-to-one correspondence with mass functions. Finally, we briefly review a few dissimilarities for belief functions.

Mass vector metrics

A generic strategy to build an evidential distance is to use a distance defined on $\mathbb{R}^N$ and apply it to mass functions as if they were any vector of that space. The most popular class of distances for multidimensional real vectors are Minkowski distances:

$$d_{\mathbf{W},k}\left(m_1, m_2\right) = \left(\left(\left[\mathbf{U}\cdot\mathbf{m}_1 - \mathbf{U}\cdot\mathbf{m}_2\right]^{k/2}\right)^T \cdot \left[\mathbf{U}\cdot\mathbf{m}_1 - \mathbf{U}\cdot\mathbf{m}_2\right]^{k/2}\right)^{1/k}, \qquad (4.1)$$

The notation $\mathbf{m}^a$ represents the Hadamard (or entrywise) product of a copies of the vector $\mathbf{m}$. In contrast, $\mathbf{M}^a$ denotes the matrix product of a copies of the matrix $\mathbf{M}$.

with $\mathbf{W}$ a positive definite matrix whose Cholesky decomposition is $\mathbf{W} = \mathbf{U}^T \cdot \mathbf{U}$ and k a positive real. The matrix $\mathbf{U}$ is upper triangular.

Usual choices for parameter k are integer values and among these $k = 2$ is the most preferred one as it yields a euclidean distance (computed in different bases of $\mathbb{R}^N$). Concerning matrix $\mathbf{W}$, standard choices are matrices proportional to a matrix mapping mass vectors to belief (Cuzzolin 2011), plausibility (Denœux 2001) or commonality vectors (Klein et al. 2016b). We call those distances L_k norm based distances as they write

$$\begin{aligned} d_{f,k} : \mathcal{M}\times\mathcal{M} &\rightarrow [0,1], \\ m_1 \times m_2 &\rightarrow \frac{1}{a}\left\|\mathbf{f}_1 - \mathbf{f}_2\right\|_k, \end{aligned} \qquad (4.2)$$

where f is a generic symbol for evidential set functions ($f \in \{bel; pl; q; b; w; v\}$) and $\|.\|_k$ is the usual L_k norm:

$$\|\mathbf{f}\|_k = \left(\sum_{1\leq i\leq N} |\mathbf{f}_i|^k\right)^{\frac{1}{k}}, \forall \mathbf{f} \in \mathbb{R}^N. \qquad (4.3)$$

From Loudahi et al. 2016, the normalizing constant a is given by:

$$a = \max_{A,B \in 2^{\Theta}} d_{f,k}\left(m_A, m_B\right). \tag{4.4}$$

(personal contribution as part of M. Loudahi's PhD)

We do not get into the details of what are the values of matrices $\mathbf{W}$ underlying distances $d_{f,k}$. These matrices are given in Smets 2002. Besides, some of these distances (4.2) are equivalent. In particular, for any $k \in \mathbb{N}^*$, we have:

$$d_{b,k} = d_{pl,k}. \tag{4.5}$$

If we work with normal belief functions, we have $d_{b,k} = d_{bel,k} = d_{pl,k}$.

Other authors proposed more specific choices for matrix $\mathbf{W}$ relying on set similarity functions. In chapter 3, we already mentioned Jousselme distance (Jousselme et al. 2001):

$$d_J\left(m_1, m_2\right) = \sqrt{\frac{1}{2}\left(\mathbf{m_1} - \mathbf{m_2}\right)^T \cdot \mathbf{D} \cdot \left(\mathbf{m_1} - \mathbf{m_2}\right)} \tag{4.6}$$

The entries of matrix $\mathbf{D}$ are given by $D\left(A, B\right) = \frac{|A \cap B|}{|A \cup B|}$. This is the Jaccard index for sets A and B. Another possible choice is the Dice index $\frac{2|A \cap B|}{|A|+|B|}$.

As remarked in a survey on evidential distances by Jousselme and Maupin 2012, a key point in the choice of $\mathbf{W}$ is the ability of it to reflect the fact that the base vectors of the mass space are in correspondence with sets. Since the power set 2^{Θ} has a poset structure when endowed the inclusion partial order, the distance values between base vectors should be compliant with this structure. For instance, $m_{\{a\}}$ should be closer to $m_{\{a;b\}}$ than to $m_{\{b\}}$. In section 4.3, we will see that this principle is subsumed by a mathematical property featuring the consistency of an evidential distance with partial orders for belief functions.

The survey by Jousselme and Maupin 2012 contains many other references and details on evidential distances and is a must-read for anyone willing to use evidential distances.

Evidential matrix metrics

Any distance between mathematical objects that are in bijective correspondence with mass functions is a legitimate candidate evidential distance. We know that such a correspondence exists between mass functions and α-specialization or α-generalization matrices. A matrix distance applied to members of those matrix classes thus automatically yield other evidential metrics.

The set of $N \times N$ matrices is denoted by $\mathcal{M}at$ and has the algebraic properties of a vector space as well. Consequently, matrix distances are not much different from vector distances. Similarly as for vector distances, we can rely on norms to build distances. A matrix norm is defined as follows:

Definition 20. A **matrix norm** $\|.\|$ is a mapping defined on $\mathcal{M}at \longrightarrow \mathbb{R}^+$ satisfying the following conditions: $\forall$, $\mathbf{A}$ and $\mathbf{B} \in \mathcal{M}at$

1. $\|\mathbf{A}\| = 0 \Leftrightarrow \mathbf{A} = 0$,
2. $\|\lambda\mathbf{A}\| = |\lambda| . \|\mathbf{A}\|$, for all $\lambda \in \mathbb{R}^+$,
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$,

A matrix norm is sub-multiplicative, if in addition one has:

$$\|\mathbf{A} \cdot \mathbf{B}\| \leq \|\mathbf{A}\| \times \|\mathbf{B}\| .$$

Any norm induces a distance defined as the norm of the difference between a pair of elements. Alleging that a matrix norm is bounded for evidential matrices, Loudahi et al. 2016 introduce the following families of normalized evidential distances:

Definition 21. An **α-specialization distance** d is a mapping such that there exists a bounded matrix norm $\|.\|$ and an α-conjunction $\textcircled{\cap}^{\alpha}$ with:

(personal contribution as part of M. Loudahi's PhD)

$$\begin{aligned} d : \mathcal{M} \times \mathcal{M} &\rightarrow [0,1] , \\ m_1 \times m_2 &\rightarrow \frac{1}{a} \left\| \mathbf{K}^{\cap}_{1,\alpha} - \mathbf{K}^{\cap}_{2,\alpha} \right\| . \end{aligned} \tag{4.7}$$

$\mathbf{K}^{\cap}_{i,\alpha}$ is the α-specialization matrix corresponding to m_i and

$$a = \max_{A,B \in 2^{\Theta}} \left\| \mathbf{K}^{\cap}_{A,\alpha} - \mathbf{K}^{\cap}_{B,\alpha} \right\|$$

is a normalization factor.

Definition 22. An **α-generalization distance** d is a mapping such that there exists a bounded matrix norm $\|.\|$ and an α-disjunction $\textcircled{\cup}^{\alpha}$ with:

(personal contribution as part of M. Loudahi's PhD)

$$\begin{aligned} d : \mathcal{M} \times \mathcal{M} &\rightarrow [0,1] , \\ m_1 \times m_2 &\rightarrow \frac{1}{a} \left\| \mathbf{K}^{\cup}_{1,\alpha} - \mathbf{K}^{\cup}_{2,\alpha} \right\| . \end{aligned} \tag{4.8}$$

$\mathbf{K}^{\cup}_{i,\alpha}$ is the α-generalization matrix corresponding to m_i and

$$a = \max_{A,B \in 2^{\Theta}} \left\| \mathbf{K}^{\cup}_{A,\alpha} - \mathbf{K}^{\cup}_{B,\alpha} \right\|$$

is a normalization factor.

The family of α-specialization distances is an extension of the family introduced in Loudahi et al. 2014b which corresponds to the $\alpha = 1$ case. In this case, 1-specialization distances are just called specialization distances.

Among existing matrix norms, the most frequently used are L_k norms and operator norms. L_k matrix norms are also known as entry-wise norms. Since matrices are elements of the vector space $\mathcal{M}at$, the definition of L_k matrix norms is the following:

$$\|\mathbf{A}\|_k = \left(\sum_{1 \leq j \leq N} \sum_{1 \leq i \leq N} |\mathbf{A}_{ij}|^k \right)^{\frac{1}{k}} . \tag{4.9}$$

Both L_k vector norms and L_k matrix norms are denoted by $\|.\|_k$. They are easily distinguished since vectors are in small letters and matrices are in capital letters.

The k-operator norm $\|.\|_{opk}$, also known as induced norm, is defined for any matrix $\mathbf{A} \in \mathcal{M}at$ as follows:

$$\|\mathbf{A}\|_{opk} = \max_{\mathbf{x}\in\mathbb{R}^N,\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_k}{\|\mathbf{x}\|_k} = \max_{\mathbf{x}\in\mathbb{R}^N,\|\mathbf{x}\|_k=1} \|\mathbf{Ax}\|_k . \tag{4.10}$$

with $\|.\|_k$ the classical L_k vector norm. In particular, the 1-operator norm writes

$$\|\mathbf{A}\|_{op1} = \max_{1\leq j\leq N} \sum_{1\leq i\leq N} |\mathbf{A}_{ij}|. \tag{4.11}$$

Observe that any evidential matrix $\mathbf{K}$ is such that $\|\mathbf{K}\|_{op1} = 1$. In the sequel, $d_{mat,k}$ denotes the α-specialization or α-generalization distance relying on the L_k matrix norm. The notation d_{opk} is used when the distance relies on the k-operator norm. In the same fashion as for evidential matrices, the value of α and the conjunctive or disjunctive nature of these distances are given in exponent when necessary.

A first result to mention concerning those distances is that there is a duality between α-specialization distances and α-generalization distances that relies on the concept of mass function negation and the De Morgan relations (3.38) and (3.39). The following proposition formalizes this duality.

Proposition 2. *Suppose $\alpha \in [0,1]$. Let $d^{\cap}$ be an α-specialization distance with respect to the α-conjunctive rule $\textcircled{\cap}^{\alpha}$ and relying on an operator or L_k matrix norm. Let $d^{\cup}$ be an α-generalization distance with respect to the α-disjunctive rule $\textcircled{\cup}^{\alpha}$ and relying on the same norm. For any mass functions m_1 and m_2 on a domain Θ, one has:*

(personal contribution as part of M. Loudahi's PhD)

$$d^{\cap}(m_1, m_2) = d^{\cup}(\overline{m}_1, \overline{m}_2). \tag{4.12}$$

Proposition 2 allows us to anticipate the fact that if an α-specialization distance satisfies a given property then so does its α-generalization counterpart. It also sheds light on ties between α-specialization distances and α-generalization distances. When $\alpha \in \{0;1\}$, it appears that these ties are stronger as illustrated by Lemma 3:

Lemma 3. *Let $d^{\cap,\alpha}$ be an α-specialization distance with respect to the α-conjunctive rule $\textcircled{\cap}^{\alpha}$ and relying on an operator or L_k matrix norm. Let $d^{\cup,\alpha}$ be an α-generalization distance with respect to the α-disjunctive rule $\textcircled{\cup}^{\alpha}$ and relying on the same norm. For any mass functions m_1 and m_2 on Θ, one has:*

(personal contribution as part of M. Loudahi's PhD)

$$d^{\cap,0} = d^{\cup,0}, \tag{4.13}$$
$$d^{\cap,1} = d^{\cup,1}. \tag{4.14}$$

This result shows that, for extreme values of α, α-specialization distances and α-generalization distances coincide. Consequently, the underlying meta-information is treated in the same way in these special cases. The following example shows that this is not true when $\alpha \in]0,1[$:

Example 17. Suppose m is a mass function on Θ and B is a subset of Θ. Let m be such that:

$$m = 0.3m_B + 0.5m_{B^c} + 0.2m_{\Theta}.$$

Figure 4.1 shows α-specialization and α-generalization distances relying on the L_1 matrix norm between m and m_B when $|\Theta| = 3$ and $|B| = 2$.

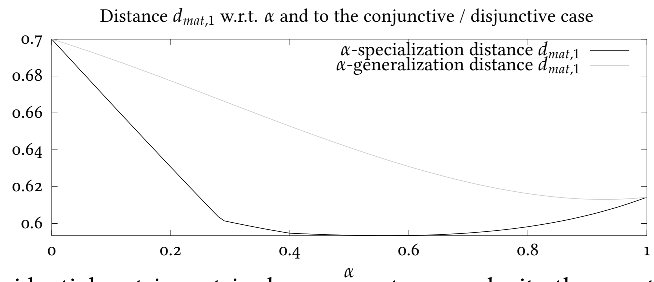


Figure 4.1: Different α-specialization and α-generalization distances relying on the L_1 matrix norm. These distances are computed between two given mass functions m_1 and m_2 such that $m_1 = m_B$ and $m_2 = 0.3m_B + 0.5m_{B^c} + 0.2m_\Theta$, with $|\Theta| = 3$ and $|B| = 2$.

Evidential matrix metrics have a greater complexity than vector metrics but their computation can be accelerated based on results in Loudahi et al. 2014a.

Dissimilarities

A large number of dissimilarities for belief functions can be found in the literature. The main ways to build dissimilarities are:

- by mapping mass functions to pignistic probability distributions (Zouhal and Denœux 1998; Tessem 1993) or to fuzzy membership functions (Han et al. 2011),
- by resorting to conflict measures (Shafer 1976; Ke et al. 2013)[1],
- by resorting to angular[2] measures (Wen et al. 2008),
- by using bi-dimensional measures whose entries are either distances or dissimilarities (Liu 2006).

Another possibility is to build a compound measure. Liu et al. 2011 build such a measure from a pignistic distribution distance and a conflict measure. Depending on the chosen components and the way they are aggregated, a compound measure may however remain a full metric. The metric introduced by Mo et al. 2016 is the convex combination of Jousselme distance d_J with a distance introduced by Sunberg and Rogers 2013. This latter distance relies on Hamming distance and allows to grasp an order relation on Θ when one such binary relation exists.

1. Burger 2014 explains that some desirable properties for conflict measures and for metrics cannot be simultaneously satisfied and that these two different notions should be characterized by different numerical measures. In spite of this, Pichon and Jousselme 2016 shows that inconsistency measures discussed in Destercke and Burger 2013 has some connection with L_∞ norm based distances between plausibilities.

2. $\frac{\mathbf{m}_1^T \cdot \mathbf{W} \cdot \mathbf{m}_2}{\|\mathbf{m}_1\|_\mathbf{W} \|\mathbf{m}_2\|_\mathbf{W}}$, where

$$\|\mathbf{m}\|_\mathbf{W} = \sqrt{\mathbf{m}^T \cdot \mathbf{W} \cdot \mathbf{m}},$$

can be interpreted as the cosine of the angles between mass vectors.

4.2 Consistency of evidential distances with combination rules

The evidential distances listed in the previous section possess all the properties of metrics and are legitimate measures to characterize how different two belief functions are. Most of them have been used successfully in a given applicative context (machine learning, belief function approximation or parameter estimation for evidential operations). On a theoretical ground, we will see that additional desirable properties can be required so that evidential distances are consistent with other components of the theory, starting in this section with combination rules.

4.2.1 *A definition of consistency between rules and distances*

A reasonable postulate is that if some information is incorporated into two mass functions by combining each of them with the same third party mass function, then they should be closer after combination than before. Loudahi et al. 2014b, 2016 formalize this idea into the following definition of consistency of an evidential distance with a combination rule:

Definition 23. Let $\odot$ be a combination rule and d an evidential distance. d is said to be **consistent** with respect to $\odot$ if for any mass functions m_1, m_2 and m_3 on Θ:

(personal contribution as part of M. Loudahi's PhD)

$$d\left(m_1 \odot m_3, m_2 \odot m_3\right) \leq d\left(m_1, m_2\right). \tag{4.15}$$

Let us focus on separable mass functions which can be decomposed into elementary pieces of evidence (simple mass functions) using Dempster's combination rule. Using a consistent evidential distance, we see that mass functions are all the closer as their decompositions involve identical elementary pieces of evidence.

Conversely, using a distance for which this property holds, then any mapping of the kind $f_0(m) = m \odot m_0$ defined on the mass space is 1-lipschitz.

4.2.2 Results on the consistency with α-junctions

We now give a collection of results from Loudahi et al. 2016 that establish the consistency of evidential matrix metrics with α-junctions. We also give in appendices some of the proofs of these results when they are not excessively long.

The first result deals with these distances when the 1-operator norm is used:

Proposition 3. *Any α-specialization or α-generalization distance d_{op1} defined using the 1-operator norm is consistent with its corresponding α-junctive combination rule.*

(personal contribution as part of M. Loudahi's PhD)
See apprendix A.2 for a proof.

Another result holds when α-specialization or α-generalization distances are defined using the L_1 matrix norm. To prove this result, it is first necessary to introduce the following lemma:

Lemma 4. *Suppose m_1 and m_2 are two mass functions defined on Θ and A and B are two subsets such that $A \subseteq B \subseteq \Theta$. Then, the following properties hold for any α-conjunctive rule $\bigcirc\!\!\!\!\cap^\alpha$ and any α-disjunctive rule $\bigcirc\!\!\!\!\cup^\alpha$*

(personal contribution as part of M. Loudahi's PhD)

$$\left\|\mathbf{m}_{1 \cap^\alpha A} - \mathbf{m}_{2 \cap^\alpha A}\right\|_1 \leq \left\|\mathbf{m}_{1 \cap^\alpha B} - \mathbf{m}_{2 \cap^\alpha B}\right\|_1, \tag{4.16}$$

$$\left\|\mathbf{m}_{1 \cup^\alpha A} - \mathbf{m}_{2 \cup^\alpha A}\right\|_1 \geq \left\|\mathbf{m}_{1 \cup^\alpha B} - \mathbf{m}_{2 \cup^\alpha B}\right\|_1. \tag{4.17}$$

From this lemma, we deduce the following corollary:

Corollary 1. *Suppose m_1 and m_2 are two mass functions defined on Θ. Let d_{op1} be an α-specialization or α-generalization distance defined using the 1-operator norm. We have:*

(personal contribution as part of M. Loudahi's PhD)

$$d_{op1}\left(m_1, m_2\right) = \frac{1}{a}\left\|\mathbf{m}_1 - \mathbf{m}_2\right\|_1. \tag{4.18}$$

This means that the 1-operator distance d_{op1} is the L_1 distance between mass vectors for any α. This corollary also implies that $a = 2$ for d_{op1}.

Equipped with these preliminary results, the following proposition regarding the consistency of distances defined using the L_1 norm can be derived:

Proposition 4. *Any α-specialization or α-generalization distance $d_{mat,1}$ defined using the L_1 matrix norm is consistent with its corresponding α-junctive combination rule.*

(personal contribution as part of M. Loudahi's PhD)
See apprendix A.2 for a proof.

One last result is available for α-specialization or α-generalization distances defined using the L_∞ matrix norm:

Proposition 5. *Any α-specialization or α-generalization distance $d_{mat,\infty}$ defined using the L_∞ matrix norm is consistent with its corresponding α-junctive combination rule.*

(personal contribution as part of M. Loudahi's PhD)
See apprendix A.2 for a proof.

The following example illustrates the inconsistency of distances d_J, $d_{mat,2}$, d_{op2} and $d_{op\infty}$ with respect to α-junctions through a numerical experiment.

Example 18. For each distance and each value of α, one iteration of the experiment carried out in this example consists in picking three random mass functions and check if inequality (4.15) is verified. The number of times that the property is verified over the number of iterations gives the **consistency rate** of the distance for a given α.

In order to provide such rates, it is necessary to generate mass functions randomly. It is sufficient to draw simple mass functions because cases of inconsistency are more frequent with such functions. Random simple mass functions are drawn uniformly using an algorithm presented in Burger and Destercke 2013 and applied to simple mass function subsimplices.

In figure 5.1, consistency rates for α-conjunctive rules and several evidential distances are shown. For this experiment, $1e4$ iterations were used. Figure 5.2 shows the same results for α-disjunctive rules.

As expected, the rates of d_{op1}, $d_{mat,1}$ and $d_{mat,\infty}$ are 100% in both the conjunctive and disjunctive cases. A rate under 100% is sufficient to prove the inconsistency of a distance. We can therefore conclude that $d_{mat,2}$, d_{op2} and $d_{op\infty}$ are inconsistent when $\alpha \neq 0$. It can be conjectured that they are consistent when $\alpha = 0$. The experiment also proves the inconsistency of d_J with any α-junction except for the disjunctive case when $\alpha = 0$. Its consistency in this latter case may also be conjectured.

4.3 Consistency of evidential distances with partial orders

Following the idea of the previous section, we now move to another consistency question between evidential distances and partial orders as defined in chapter 3.

4.3.1 A definition of consistency between orders and distances

The previous chapter highlighted the benefits of partial orders for belief functions which allow to sort belief functions in easily interpretable ways.

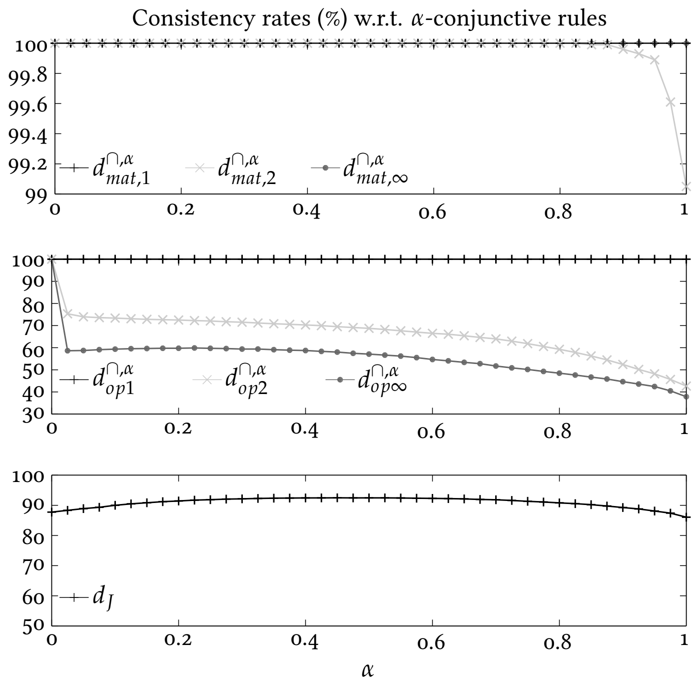


Figure 4.2: Consistency rates of several evidential distances with α-conjunctive rules with respect to parameter α.

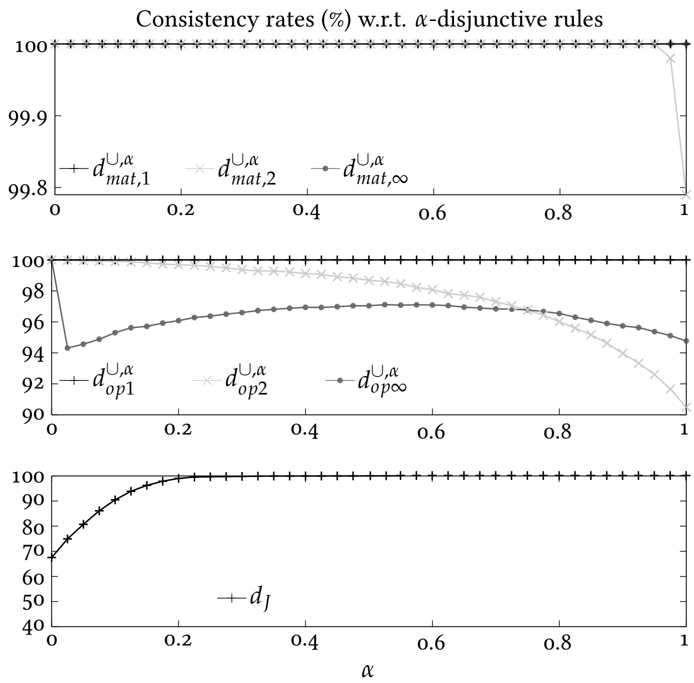


Figure 4.3: Consistency rates of several evidential distances with α-disjunctive rules with respect to parameter α.

By endowing the mass space with one of these partial orders and a distance, it makes sense to require that the chosen distance preserves this interpretation and gives distance values compliant with inclusion relations. Among other possibilities, Klein et al. 2016c formalize this notion of compliance as follows:

Definition 24. Given a partial order $\sqsubseteq_y$ defined over $\mathcal{M}$, an evidential distance (or dissimilarity) d is said to be **$\sqsubseteq_y$-compatible** if for any mass functions m_1, m_2 and m_3 such that $m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3$, we have:

(personal contribution)

$$\max\left\{d\left(m_1, m_2\right); d\left(m_2, m_3\right)\right\} \leq d\left(m_1, m_3\right), \qquad (4.19)$$

Moreover, d is said to be **$\sqsubset_y$-compatible (in the strict sense)** if $m_1 \sqsubset_y m_2 \sqsubset_y m_3$ implies a strict inequality: $\max\left\{d\left(m_1, m_2\right); d\left(m_2, m_3\right)\right\} < d\left(m_1, m_3\right)$.

In particular, if the partial order $\sqsubseteq_y$ has[3] a minimum $\perp \in \mathcal{M}$ and a maximum element $\top \in \mathcal{M}$, then satisfying strict compatibility in Definition 24 ensures that d refines the partial order $\sqsubseteq_y$ into a total preorder $\preceq_y$ defined as $m_1 \preceq_y m_2$ if $d(\perp, m_1) \leq d(\perp, m_2)$.

3. Recall that minimum and maximum elements $\perp$, $\top$ of $\sqsubseteq$ are such that for any other element x, $\perp \sqsubset x \sqsubset \top$.

Conversely, we will say that a distance is not compatible, or **incompatible**, with a partial order if Definition 24 is not satisfied for some triplet m_1, m_2, m_3, that is $m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3$ and $\max\left\{d\left(m_1, m_2\right); d\left(m_2, m_3\right)\right\} > d\left(m_1, m_3\right)$. While the compatibility of a partial order with a distance could be defined in a different way, using a $\sqsubseteq_y$-incompatible distance on a problem involving $\sqsubseteq_y$ seems ill-advised.

The trivial distance[4] is obviously compatible with any non-strict partial order and incompatible with any strict order. In general, this tends to show that $\sqsubset_y$-compatibility in the strict sense is a lot more valuable property than $\sqsubseteq_y$-compatibility.

4. For any m and m', the trivial metric equals 1 whenever $m \neq m'$.

When possible, the implications between different orders (see Equation (3.3)) can be used to avoid checking the compatibility of a distance with respect to all partial orders, as shown by the next proposition.

Proposition 6. *Consider two partial orders $\sqsubseteq_x$, $\sqsubseteq_y$ such that $\sqsubseteq_x \Rightarrow \sqsubseteq_y$ and a distance d, then*

- *if d is $\sqsubseteq_y$-compatible, then it is $\sqsubseteq_x$-compatible;*
- *if d is $\sqsubseteq_x$-incompatible, then it is $\sqsubseteq_y$-incompatible.*

(personal contribution)
See apprendix A.3 for a proof.

An immediate corollary follows concerning the strict part:

Corollary 2. *Consider two partial strict orders $\sqsubset_x$, $\sqsubset_y$ such that $\sqsubset_x \Rightarrow \sqsubset_y$ and a distance d, then*

- *if d is strictly $\sqsubset_y$-compatible, then it is strictly $\sqsubset_x$-compatible;*
- *if d is strictly $\sqsubset_x$-incompatible, then it is strictly $\sqsubset_y$-incompatible.*

(personal contribution)

4.3.2 Results on the consistency with informative partial orders

Klein et al. 2016c proved the consistency of several vector based evidential distances with information partial orders as summarized by the following proposition:

Proposition 7. *For any $k \in \mathbb{N}^* \setminus \{\infty\}$, the following assertions hold:*

(personal contribution)
See apprendix A.3 for a proof.

- *the distances $d_{b,k}$ and $d_{pl,k}$ are $\sqsubset_{pl}$ and $\prec_\pi$-compatible in the strict sense,*
- *the distance $d_{bel,k}$ is $\sqsubset_{bel}$-compatible in the strict sense,*
- *the distance $d_{q,k}$ is $\sqsubset_q$ and $\prec_\pi$-compatible in the strict sense.*
- *the pseudo-distance $d_{\pi,k}$ is $\prec_\pi$-compatible in the strict sense.*

The same results also hold for $k = \infty$ with non-strict orders.

We can easily check that the dissimilarities based on pignistic probability distribution distances are not compatible with the partial order $\sqsubseteq_w$, and therefore are also not compatible for any other partial order comparing informative content implied by $\sqsubseteq_w$. Indeed, consider the following example.

Example 19. Let $\Theta = \{a, b, c\}$ and consider three mass functions such that their corresponding conjunctive weight functions are

set	$\emptyset$	$\{a\}$	$\{b\}$	$\{a;b\}$	$\{c\}$	$\{a;c\}$	$\{b;c\}$
w_1	1	0.2	0.2	1	0.2	1	1
w_2	1	0.2	0.4	1	0.2	1	1
w_3	1	1	1	1	1	1	1

Let P_i denote the pignistic probability distributions corresponding to each mass function. We have $m_1 \sqsubseteq_w m_2 \sqsubseteq_w m_3$, but $P_1 = P_3$ and $P_2 \neq P_1$. If d_{BetP} denotes a (full) metric between pignistic probability distributions, reflexivity implies $d_{BetP}(P_1, P_3) = 0$, but separability implies $d_{BetP}(P_1, P_2) \neq 0$.

This shows that using $d_{BetP,k}$ is ill-advised in problems involving informative content, such as the approximation of belief functions.

A more surprising fact is that Jousselme distance d_J is unfortunately incompatible with $\sqsubseteq_y$ for $y \in \{w, d, s, q, pl, bel\}$, as show the next two counter-examples.

Example 20. Let $\Theta = \{a, b\}$ and consider three mass functions such that their corresponding conjunctive weight functions are

set	$\emptyset$	$\{a\}$	$\{b\}$
w_1	1	0.2	0.25
w_2	1	1	0.25
w_3	1	1	0.6

We have $m_1 \sqsubseteq_w m_2 \sqsubseteq_w m_3$ but $d_J(m_1, m_3) \approx 0.63 < d_J(m_1, m_2) \approx 0.66$, hence d_J is $\sqsubseteq_w$-incompatible. Proposition 6 gives the incompatibility with all the other mentioned partial orders except $\sqsubseteq_{bel}$.

Concerning $\sqsubseteq_{bel}$, another counter-example is required:

Example 21. Let $\Theta = \{a, b\}$ and consider the mass functions:

set	$\emptyset$	$\{a\}$	$\{b\}$	Θ
m_1	0	0.3	0.4	0.3
m_2	0	0	0.2	0.8
m_3	0.4	0	0.1	0.5
bel_1	0	0.3	0.4	1
bel_2	0	0	0.2	1
bel_3	0	0	0.1	0.6

We have $m_1 \sqsubseteq_{bel} m_2 \sqsubseteq_{bel} m_3$, but $d_J(m_1, m_3) \approx 0.36 < d_J(m_1, m_2) \approx 0.38$, hence d_J is $\sqsubseteq_{bel}$-incompatible. This counter-example also refutes the compatibility between $\sqsubseteq_{bel}$ and dissimilarities based on pignistic probability distribution distances because $P_2 = P_3 = 1_{\{b\}} \neq P_1$. If d_{BetP} denotes a (full) metric between pignistic probability distributions, reflexivity implies $d_P(P_2, P_3) = 0$, but separability implies $d_{BetP}(P_1, P_2) \neq 0$.

4.4 Conclusions

This chapter sheds light on the fact that there are plenty of relevant ways to define distances for belief functions and that the mass space has a metric space structure. Establishing this structure is important as evidential distances have many applications on both theoretical grounds and practical grounds. Reported applications of evidential distances are:

- quality assessment of an evidential algorithm (Fixsen and Mahler 1997),
- parameter optimization (Zouhal and Denœux 1998),
- as building block of an evidential operator like a conflict measure (Martin et al. 2008), an approximation algorithm (Cuzzolin 2010a, Klein et al. 2016c) or a combination rule (Klein et al. 2016b).

We also advocated that the derivation of a distance for belief functions should not be disconnected from other aspects of the theory of belief functions. By defining properties featuring the bonds between distances and partial orders or combination rules, some metrics are proved to be compliant with either an informative partial order or an α-junction.

In spite of recent advances, there are still open problems with evidential distances. First, as the reader may have remarked, there is an empty intersection among distances consistent with α-junctions and those consistent with partial orders. Deriving one for which the two properties hold (or proving the impossibility thereof) is a challenging problem. Second, other kinds of consistencies can be thought of and would require investigations. One could examine the consistency with refining mappings or with uncertainty measures for instance. Finally on the very subject of the consistency between distances (or dissimilarities) and informational content, other properties may also be desirable. In the probability theory, the informative content of a distribution is measured by means of the entropy. The Kullback–Leibler divergence is a distribution dissimilarity that achieves built-in compatibility with entropy as it is defined as the cross entropy of the compared distributions (say p_1 and p_2) minus the entropy of the first distribution (p_1). It is thus interpreted as a measure of information gain from revising one beliefs from p_2 to p_1. Obviously, the Kullback–Leibler is however not a metric as it is not symmetric. To the best of our knowledge, this track has yet hardly been explored as the generalization of entropy to belief functions is itself still an open problem (cf. Jirousek and Shenoy 2018).

Belief spaces: from magmas to monoids 5

In the first chapter, we have explained that the theory of belief functions is meant to build models for reasoning under uncertainty. One such reasoning is inference, that is, we want to make deductions about an unknown variable of interest $\theta \in \Theta$ based on data in $\mathbb{X}$. In many situations, like distributed systems, partial deductions are computed locally and must be aggregated later. In the theory of belief functions, combination rules like Dempster's rule or Smets' α-junctions allow to perform such an aggregation.

In this chapter, we will review combination rules for belief functions. We will see that depending on the rule, the belief space acquire different types of algebraic structures. In the two previous chapters, we have seen how to perform qualitative and quantitative comparisons of belief functions. We will also see that these comparisons are instrumental in the definition of many combination rules.

Several combination rules have been introduced in the previous chapters. We recall their definitions so that this chapter is self-contained.
α-junctions: for any $E \subseteq \Theta$

$$\left(m_1 \textcircled{\cap}^{\alpha} m_2\right)(E) = \sum_{\substack{A,B,C \subseteq \Theta,\\ (A\cap B)\cup(A^c\cap B^c\cap C)=E}} m_1(A)\, m_2(B)\, \alpha^{|C^c|}\overline{\alpha}^{|C|},$$

$$\left(m_1 \textcircled{\cup}^{\alpha} m_2\right)(E) = \sum_{\substack{A,B,C \subseteq \Theta,\\ (A\Delta B)\cup(A\cap B\cap C)=E}} m_1(A)\, m_2(B)\, \alpha^{|C|}\overline{\alpha}^{|C^c|},$$

Conjunctive and disjunctive rules: for any $E \subseteq \Theta$

$$\left(m_1 \textcircled{\cap} m_2\right)(E) = \sum_{\substack{A,B \subseteq \Theta,\\ A\cap B=E}} m_1(A)\, m_2(B)$$

$$\left(m_1 \textcircled{\cup} m_2\right)(E) = \sum_{\substack{A,B \subseteq \Theta,\\ A\cup B=E}} m_1(A)\, m_2(B),$$

Dempster's rule: for any $E \subseteq\neq \emptyset$

$$\left(m_1 \oplus m_2\right)(E) = \frac{1}{1-\kappa}\left(m_1 \textcircled{\cap} m_2\right)(E),$$

where $\kappa = \left(m_1 \textcircled{\cap} m_2\right)(\emptyset)$.
Cautious and bold rules:

$$m_1 \owedge m_2 = \bigcirc\!\!\!\!\cap_{B \subsetneq \Theta} m_B^{w_1(B)\wedge w_2(B)},$$

$$m_1 \ovee m_2 = \bigcirc\!\!\!\!\cup_{A \neq \emptyset} mA_{v_1(A)\wedge v_2(A)},$$

where m_x^A denotes a mass function such that

$$m_x^A(E) = \begin{cases} 1-x & \text{if } E = A, \\ x & \text{if } E = \emptyset, \\ 0 & \text{otherwise} \end{cases}.$$

5.1 Structures induced by rules

This section presents structures that are obtained by endowing the mass space with a combination rule. We also evoke structures that cannot be achieved.

5.1.1 Magmas

As already observed earlier, belief functions are quite general objects and we cannot hope for very rich algebraic structures for belief spaces. The crudest structure is the magma structure.

Definition 25. Let $\mathcal{M}$ denote some space and $\star$ denote a mapping $\mathcal{M} \times \mathcal{M} \to \mathcal{M}$. The structure $(\mathcal{M}, \star)$ is called **magma** and the mapping $\star$ is called **binary operation**.

This definition applies to the mass space by endowing it with a combination rule. Clearly for any $\alpha \in [0;1]$, both $(\mathcal{M}, \textcircled{\cap}^{\alpha})$ and $(\mathcal{M}, \textcircled{\cup}^{\alpha})$ are magmas. It does not apply for Dempster's rule $\oplus$ because the combination of two maximally conflicting mass functions is not defined. Likewise, the cautious $\owedge$ and bold $\ovee$ rules only induce a magma on subsets of $\mathcal{M}$, i.e. on the set of non-dogmatic mass functions and the set of subnormal mass functions respectively. In practice, a dogmatic mass function m_i can still be used as input for $\owedge$. Indeed, m_i can be turned into a conjunctive weight function by discounting it with a very small discount rate as compared to the minimum positive mass of m_i. The discount rate can be chosen as small as necessary so that the values of w_i stabilize to some value up to a desired precision threshold. In the same way, one turns a normalized mass function m_i into a disjunctive weight function by artificially assigning an infinitesimal mass value to $\emptyset$ and then renormalize so that $\sum_{E\subseteq\Theta} m(E) = 1$ and this function can be used as input for $\circledvee$.

When the combination rule inducing the magma has some properties, the magma inherits those thereby yielding subcategories of magmas. For instance if the rule is associative then we obtain a semigroup structure. Table 5.1 gives the list of magma subcategories thus obtained.

Rule property	Induced structure
Associativity	Semigroup
Reversibility	Quasigroup
Associativity & Commutativity & Idempotence	Semilattice
Associativity & Neutral element	Monoid
Associativity & Neutral element & Reversibility	Group
Associativity & Neutral element & Reversibility & Commutativity	Abelian Group

Table 5.1: Some subcategories of magmas.

In the rest of this section, we comment on some subcategories in the context of the theory of belief functions. For a reference on this topic, see Daniel 2004 which stems from Hájek and Valdes 1991 who introduced the algebraic notion of Dempsteroid[1].

1. This custom algebraic structure is obtained when $n = |\Theta| = 2$ by endowing the set of non categorical mass functions with Dempster's rule, a neutral element (vacuous mass function), an indempotent element (evenly distributed Bayesian mass function), a pre-order and the decombination operation for Dempster's rule although this latter operation is not defined for any mass function.

5.1.2 Monoids

A majority of rules on the belief function literature induce a commutative monoid structure for the mass space. Examples of such rules are α-junctions. The associativity and commutativity of these rules is a consequence of the commutativity and associativity of the set operations on which these rules rely (see subscripts of the sum sign in equations (2.20) and (2.21)). It is also easily shown that the neutral element[2] of any α-conjunctive rule is the vacuous mass function m_Θ while the neutral element of any α-disjunctive is the maximal conflict mass function $m_\emptyset$.

2. a neutral element is a function m_e such that $m \star m_e = m$, for any m in the magma $(\mathcal{M}, \star)$.

Conversely, it can also be proved that the only absorbing element [3] (or zero element) of any α-conjunctive rule is the maximal conflict mass function $m_\emptyset$ while the absorbing element of any α-disjunctive rule is the vacuous mass function m_Θ. As a consequence of this property, in an information fusion context, the mass assigned to $\emptyset$ cannot decrease (curse of conflict) in the conjunctive case whereas the mass assigned to Θ cannot decrease (curse of ignorance) in the disjunctive case.

3. An absorbing element m_a is such that for any element m in a magma $(\mathcal{M}, \star)$, $m_a \star m = m_a$. If a magma has an absorbing element, then it is unique.

The cautious rule does not induce a monoid structure for the set of non-dogmatic mass function because there is no neutral element for this rule. Likewise, the bold rule does not induce a monoid for the set of subnormal mass functions.

Several rules inducing a commutative monoid will be presented in the rest of this chapter. The fact that many authors strove to obtain this structure is maybe explained by the popularity of belief functions in the information fusion community. Indeed, when we combine pieces of information, most applicative contexts will require that the order with which the pieces of information are aggregated has no impact on the end result, hence commutativity and associativity. The need for a neutral element is perhaps less obvious except in the conjunctive case. In this case, any piece of information is taken for granted and if one piece stands for total ignorance, it makes perfect sense that it has no impact on the end result. So what is really desirable is not only to have a neutral element but also that this element stands for total ignorance as the vacuous mass function m_Θ does in the theory of belief functions.

5.1.3 Semilattices

If rules inducing a monoid structure are legion, those inducing semilattices are quite few. Idempotence is patently difficult to obtain even for very general families of rules as the following one.

Definition 26. Let $\psi_E : 2^\Theta \times 2^\Theta \to [0;1]$ denote some functions indexed by set $E \subseteq \Theta$ such that for any pair (m_1, m_2) of mass functions on Θ, the set function $m : 2^\Theta \to \mathbb{R}$ defined as

$$m(E) = \sum_{A,B \subseteq \Theta} \psi_E(A,B)\, m_1(A)\, m_2(B), \forall E \subseteq \Theta, \tag{5.1}$$

is a mass function.

The rule induced by functions $(\psi_E)_{E\subseteq\Theta}$ is a **quadratic rule**.

We call these rules quadratic ones as any mass value is obtained as

$$m(E) = \mathbf{m}_1^T \cdot \mathbf{\Psi}_E \cdot \mathbf{m}_2,$$

where $\mathbf{\Psi}_E$ is an $N \times N$ matrix whose entries are $\psi_E(A,B)$. The values of $\psi_E(A,B)$ can be understood as the proportion of the mass product $m_1(A)\, m_2(B)$ that flows to E through this combination. We see that α-junctions are a subfamily of the quadratic rules[4]. For example, the conjunctive rule is retrieved by setting $\psi_E = 1_E(A \cap B)$ for all $E \subseteq \Theta$. Going back to idempotence, the following result holds.

4. In the conjunctive case, if we set
$$\psi_E(A,B) = \sum_{\substack{C\subseteq\Theta,\\ (A\cap B)\cup(A^c\cap B^c\cap C)=E}} \alpha^{|C^c|}(1-\alpha)^{|C|},$$
then the corresponding quadratic rule coincides with $\textcircled{\cap}^\alpha$. More generally, the behavior-based fusion scheme of Pichon et al. 2012 also writes as a quadratic rule.

Proposition 8. *No quadratic rule is idempotent.*

(personal contribution) [unpublished]

Proof. Suppose the quadratic rule induced by functions $(\psi_E)_{E\subseteq\Theta}$ is quadratic and idempotent. Applying the rule to a categorical mass function m_C, we have

$$m_C(E) = \psi_E(C,C)\, m_C(C)\, m_C(C), \forall E \subseteq \Theta, \tag{5.2}$$
$$\Leftrightarrow 1_C(E) = \psi_E(C,C), \forall E \subseteq \Theta. \tag{5.3}$$

This means that all diagonal elements of $\mathbf{\Psi}_E$ are null except $\psi_E(E,E) = 1$. Now, let m denote a mass function such that E is the only subset in Θ that is not a focal element.

$$m(E) = 0, \tag{5.4}$$
$$\Leftrightarrow \sum_{A,B\subseteq\Theta} \psi_E(A,B)\, m(A)\, m(B) = 0 \tag{5.5}$$
$$\Leftrightarrow \sum_{\substack{A,B\subseteq\Theta\\ A\neq E, B\neq E}} \psi_E(A,B)\, m(A)\, m(B) = 0. \tag{5.6}$$

Since all other subsets are focal elements of m and the above sum is a sum of non negative terms, we deduce $\psi_E(A,B) = 0$ if $E \neq A$ or $E \neq B$.

Now, let m denote the mass function such that $m(E) = \frac{1}{N}$ for all $E \subseteq \Theta$. Again, by idempotence, we obtain

$$m(E) = \sum_{A,B\subseteq\Theta} \psi_E(A,B)\, m(A)\, m(B), \forall E \subseteq \Theta \tag{5.7}$$

$$\Leftrightarrow \frac{1}{N} = \sum_{A,B\subseteq\Theta} \psi_E(A,B)\, \frac{1}{N^2}, \forall E \subseteq \Theta, \tag{5.8}$$

$$\Leftrightarrow N = \sum_{A,B\subseteq\Theta} \psi_E(A,B), \forall E \subseteq \Theta. \tag{5.9}$$

We continue to add constraints on functions ψ_E by examining a function m' defined as

$$m'(E) = \begin{cases} \frac{1}{N-1} & \text{if } E \neq C \\ 0 & \text{if } E = C \end{cases}.$$

Applying the same reasoning as for m, we obviously have

$$N - 1 = \sum_{\substack{A,B\subseteq\Theta \\ A\neq C, B\neq C}} \psi_E(A,B), \forall E \subseteq \Theta. \tag{5.10}$$

By subtracting (5.10) to (5.9), we get

$$1 = \sum_{\substack{A\subseteq\Theta \\ A\neq C}} \psi_E(A,C) + \sum_{\substack{B\subseteq\Theta \\ B\neq C}} \psi_E(C,B) + \psi_E(C,C). \tag{5.11}$$

Remembering that $\psi_C(C,C) = 1$, we deduce that $\psi_C(A,C) = \psi_C(C,B) = 0$. The result being true for any subset C, the functions ψ_E are now fully specified as

$$\psi_E(A,B) = \begin{cases} 1 & \text{if } A = B = E \\ 0 & \text{otherwise} \end{cases}, \forall E \subseteq \Theta. \tag{5.12}$$

This set of functions ψ_E does not induce a combination rule. Indeed, take two categorical mass functions m_A and m_B as inputs such that $A \neq B$, the obtained set function by application of relation (5.1) is constant zero.

□

An obvious way to derive an idempotent rule is to resort to good old average. In fact, any convex combination kind of rule is idempotent. The average is however not associative but only **quasi-associative**. Suppose you have previously combined functions $m_1, \ldots, m_\ell$. Quasi-associativity means that there is a way to compute combinations iteratively without storing the whole set of mass functions $\{m_1, \ldots, m_\ell\}$ and restart the combination from scratch when a new function $m_{\ell+1}$ arrives. In practice, quasi-associativity is very important and the added value of associativity as compared to quasi-associativity is limited. The average rule can be classified as disjunctive because the focal core of the average of input mass functions is the union of the focal cores of these input mass functions. Consequently, any candidate solution with a positive plausibility for some input mass function also has a positive plausibility after combination. The most important drawback of averaging mass functions is that the theoretical foundations of this rule have not yet been addressed. For example, Dempster's rule rely on strong probabilistic grounds; α-junctions are understood in terms of source truthfulnesses and the cautious and bold rules relies on informative partial orders. The meaning of an average of mass functions remains, in general, more difficult to justify. A noteworthy exception is the convex combination of a mass function with the vacuous one, i.e. discouting. As we will see in 5.4.2, the discounting operation relies on a meta-information based update. Besides, the average rule does not have a neutral element.

As already remarked in chapter 3, the cautious and bold rules are idempotent and they induce a semilattice on the set of non-dogmatic mass functions and the set of subnormal mass functions respectively. Careful readers will raise objections as the definition of semilattices in this chapter is not the same as in chapter 3! In chapter 3, the notion of semilattice always comes along with a partial order whereas this is not mandatory in the algebraic version of the definition. In spite of this, the order theoretic and algebraic definitions of semilattices are equivalent, because any binary operation induces a partial order [5].

5. Let $\sqsubseteq_{\oslash}$ denote the following partial order for non dogmatic mass functions:

$$m_1 \sqsubseteq_{\oslash} m_2 \Leftrightarrow m_1 = m_1 \oslash m_2. \quad (5.13)$$

If $m_1 \sqsubseteq_w m_2$, then $w_1(A) \leq w_2(A)$, for all $A \subsetneq \Theta$. It follows that $w_1(A) \wedge w_2(A) = w_1(A)$ and thus $m_1 \oslash m_2 = m_1$.

If $m_1 \sqsubseteq_{\oslash} m_2$, then by definition of $\oslash$, $w_1(A) \wedge w_2(A) = w_1(A)$, for all $A \subsetneq \Theta$. It follows that $w_1(A) \leq w_2(A)$ and $m_1 \sqsubseteq_w m_2$.

Finally, $\sqsubseteq_{\oslash}$ and $\sqsubseteq_w$ are equivalent.

Some other idempotent rules can be derived by relying on optimization techniques as explained in section 5.3. The motivations behind the derivation of idempotent rules are mainly dealing with dependent sources when the dependence in question is hard to assess. Alleging mass functions (or a subclass of those) can be decomposed w.r.t. to an idempotent rule, then two identical elementary pieces of evidence will not be counted twice in a combination between two such functions. This cautious approach does not lead to unjustified reinforcement of the support given to focal elements. In the worst case, combining two copies of the same evidence (maximal dependency) using Dempster's rule or the conjunctive rule generates such unjustified reinforcements.

5.1.4 *Groups*

A frustrating point in the theory of belief functions is that no existing rule has achieved to induce a group structure and there is little hope that any rule can. The difficulty in deriving a group structure is that there is not always an inverse mass function, i.e. a mass function that would undo the mass reallocation incurred by a combination. If we place ourselves under the framework of category theory[6], we may a find path from a mass function m_1 to m_2 but in general we will not be able to find path back to m_1. From an artificial intelligence point of view, a model allowing to go back to a previous state of beliefs seems more in line with human reasoning. Yet belief functions were never meant to achieve reversibility in the first place.

6. See section 3.3.

Obviously, the inversion requirement is incompatible with conjunctivity and disjunctivity. If a candidate solution $a \in \Theta$ is discarded by a mass function m_i (i.e. $pl_i(\{a\}) = 0$), then this piece of information is propagated in a conjunctive[7] combination end result. So, conjunctive operators generate intrinsically irreversible mass allocations. Likewise, if a candidate solution $a \in \Theta$ is not discarded by a mass function m_i (i.e. $pl_i(\{a\}) > 0$), then this piece of information is propagated in a

7. We refer here to a more general notion of conjunction or disjunction than the one presented in 3.3.

disjunctive combination end result.

There are of course some rules that are neither conjunctive nor disjunctive like the average of mass functions or some that will be introduced in the next section. To the best of our knowledge, none of them induce a group structure. Besides, the notions of inverse mass function w.r.t to conjunctive and disjunctive rules that rely on so called **decombination**[8] (Smets 1995; Denœux 2008) should not be confused with reversibility. Let m_1 and m_2 denote two functions and m_{12} is the result of their combination using a given rule $\star$. Decombination is a mathematical process that allows to retrieve either m_1 or m_2 from m_{12}. Reversibility means that decombination is achievable by combining m_{12} with some other mass function using the same rule as the one that yielded m_{12}, i.e. $\exists m$ and m' s.t. $m_{12} \star m = m_1$, $m_{12} \star m' = m_2$.

Desirable as it may be, reversibility may also be too strong a requirement for recovering previous states of beliefs. When humans change their minds and revise their beliefs accordingly, they do not necessarily use the same process for forward and backward moves. Another possibility is that one first selects a combination operator based on meta-information and then apply it (see Klein et al. 2009 for an example of such an approach applied to a computer vision context). In conclusion, it may be sufficient that a pair of rules allow reversibility.

8. Suppose $m = m_1 \textcircled{\cap} m_2$ and we learn that m_2 turns out not to be supported by evidence. We can "remove" m_2 from m by combining m with a function $\tilde{m}$ such that

$$\tilde{q}(B) = \frac{q_1(B)}{q_2(B)}, \forall B \subseteq \Theta.$$

We usually denote $m \textcircled{\not\cap} m_2 = m \textcircled{\cap} \tilde{m}$. In general, $\tilde{m}$ is not a bona fide mass function as entrywise divisions of commonality functions are not always commonality functions and $\tilde{m}$ is not the inverse of m_2 as $m_2 \textcircled{\cap} \tilde{m} \neq m_{\Theta}$.

A similar reasoning applies in the disjunctive case using implicability functions.

5.2 Alternatives to conjunction or disjunction

Conjunctive and disjunctive rules can be regarded as two extreme information fusion tools and when the reliability of the sources is limited but not inexistent, intermediate solutions are desirable. In this section, we will present rules that are neither conjunctive nor disjunctive (like the average rule). They mainly fall in two categories: those relying on alternative policies for conflict redistribution and those that are a mix of conjunctive and disjunctive rules.

5.2.1 *Conflict redistribution*

Dempster 1967 introduces a normalizing constant in his expression of upper and lower probabilities and this constant is actually necessary because some probability mass is not assigned to non empty subsets of Θ. When we use Dempter's rule, this constant is still in action and we have seen that it is understood as $1 - \kappa$ where κ is called Dempster's degree of conflict and represents the mass arising from inconsistent focal elements of each input belief function.

One of the debated question in the wake of Dempter and Shafer's contributions is the relevance of this way to redistribute the conflict, that is, a multiplicative redistribution with equal strength for each focal element of the combination result belief function.

Yager 1987 argues that the mass assigned to non empty informative[9] subsets A should be $m_1 \textcircled{\cap} m_2(A)$ because cross-checked evidence suggests so. The usual conflict redistribution may artificially inflate the support given to A. The only focal element whose support can be increased at no

9. We mean that $A \neq \emptyset$ and $A \neq \Theta$.

risk is obviously Θ. The combination result mass function m_{yag} obtained by applying Yager's rule to input functions m_1 and m_2 is

$$m_{\text{yag}}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ m_1 \circledcirc m_2(\Theta) + m_1 \circledcirc m_2(\emptyset) & \text{if } A = \Theta \\ m_1 \circledcirc m_2(A) & \text{otherwise} \end{cases} . \quad (5.14)$$

This rule is neither conjunctive nor disjunctive[10]. Furthermore, this rule is commutative, quasi-associative and m_Θ is the neutral element.

A motivation behind Yager's rule is also to introduce an alternative rule that provides a more legitimate outcome in Zadeh 1986 "counter example" to Dempster's rule of combination.

10. If $m_1 = m_{\{a\}}$ and $m_2 = m_{\{b\}}$ then $m_{\text{yag}} = m_\Theta$, so b is an impossible candidate value for θ in m_1 but it is a possible one in m_{yag}, hence the rule is not conjunctive. If $m_1 = m_{\{a\}}$ and $m_2 = m_{\{a,b\}}$ then $m_{\text{yag}} = m_{\{a\}}$, so b is a possible candidate value for θ in m_2 but it is not a possible one in m_{yag}, hence the rule is not disjunctive.

Example 22. Let $\Theta = \{a, b, c\}$ denote some space and we need to combine the following mass functions on Θ:

$$m_1 = 0.99 m_{\{a\}} + 0.01 m_{\{b\}}, \quad (5.15)$$

$$m_2 = 0.99 m_{\{c\}} + 0.01 m_{\{b\}}. \quad (5.16)$$

In this case, we obtain $m_1 \oplus m_2 = m_{\{b\}}$. Zadeh argues that Dempster's rule combines evidence in a dubious fashion because we conclude that $\theta = b$ with certainty whereas this possibility was given little support by both sources. It can also be argued that evidence encoded in m_1 implies that c is not a possible solution and the evidence encoded in m_2 implies that a is not either. So, we are left with b only and Dempster's rule result follows this logic.

Applying Yager's rule, we obtain $m_{\text{yag}} = 0.0001 m_{\{b\}} + 0.9999 m_\Theta$, a result which is arguably more cautious than $m_1 \oplus m_2$.

Another point that dragged the attention of several authors is that conflict should maybe not be treated as a global aggregated value but instead as a collection of smaller pieces each of which stems from a pair (E_1, E_2) of non intersecting focal elements of each input mass function. Each piece of conflict can then be redistributed in a tailored way depending on E_1 and E_2.

For instance, if $m_1 = m_{\{a\}}$ and $m_2 = m_{\{b\}}$, then $m_{\text{yag}} = m_\Theta$, a result that is even more conservative than the disjunctive rule (if $n > 2$). Since the conflict in here has its roots in the incompatibility of the respective focal elements $E_1 = \{a\}$ and $E_2 = \{b\}$, Dubois and Prade 1992 suggest to redistribute this conflict on the disjunction of the focal elements. The combination result mass function m_{dp} obtained by applying Dubois and Prade's rule to input functions m_1 and m_2 is

$$m_{\text{dp}}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ m_1 \circledcirc m_2(A) + \sum\limits_{\substack{B,C \subseteq \Theta \\ B \cap C = \emptyset \\ B \cup C = A}} m_1(B)\, m_2(C) & \text{otherwise} \end{cases} . \quad (5.17)$$

Dubois and Prade's rule has the same general properties as Yager's rule[11].

Yager and Dubois and Prade rules are members of a family of rules introduced by Lefevre et al. 2002. In this setting, a set function β gives the amount of the conflict that is redistributed to each subset[12]. The set

11. In this monograph, rule properties are discussed in algebraic terms. Dubois et al. 2016 discuss rule properties from an information fusion perspective. In this scope, Dubois and Prade's rule possesses more properties than Yager's or Dempster's rule.

12. Take $\beta(\Theta) = 1$ and Yager's rule is retrieved.
Take $\beta(A) = \frac{1}{1-\kappa} \sum\limits_{\substack{B,C \subseteq \Theta \\ B \cap C = \emptyset \\ B \cup C = A}} m_1(B)\, m_2(C)$ for $A \neq \emptyset$ and Dubois and Prade's rule is retrieved. In this latter case the function β depends on m_1 and m_2 as well.

function β sums to one: $\sum_{B\subseteq\Theta}\beta(B) = 1$. The combination result mass function m_{lef} obtained by applying one such rule to input functions m_1 and m_2 is

$$m_{\text{lef}}(A) = \begin{cases} \beta(\emptyset)\, m_1 \cap\!\!\!\bigcirc m_2(\emptyset) & \text{if } A = \emptyset \\ m_1 \cap\!\!\!\bigcirc m_2(A) + \beta(A)\, m_1 \cap\!\!\!\bigcirc m_2(\emptyset) & \text{otherwise} \end{cases}. \tag{5.18}$$

Any such rule has at least the same general properties as Yager's rule except that if $\beta(\emptyset) = 1$, the corresponding rule is conjunctive because we retrieve $\cap\!\!\!\bigcirc$.

An adaptive tailored conflict redistribution was proposed by Smarandache and Dezert 2005 under the name **partial conflict redistribution** (PCR) rule. The combination result mass function m_{pcr} obtained by applying PCR rule to input functions m_1 and m_2 is

$$m_{\text{pcr}}(A) = m_1 \cap\!\!\!\bigcirc m_2(A) + \sum_{\substack{B\subseteq\Theta \\ B\cap A=\emptyset}} \left[\frac{m_1(A)^2 m_2(B)}{m_1(A)+m_2(B)} + \frac{m_1(B)\, m_2(A)^2}{m_1(B)+m_2(A)} \right], \tag{5.19}$$

Equation (5.19) is not well defined when one of the denominators is zero. Whenever this happens the corresponding fraction is actually discarded from the formula. We preferred not to make this condition explicit in (5.19) otherwise it becomes harder to read.

for $A \neq \emptyset$ and $m_{\text{pcr}}(\emptyset) = 0$.

In this case, if E_1 is a focal element of m_1, E_2 a focal element of m_2 and $E_1 \cap E_2 = \emptyset$, then a part of the conflict bounces back to E_1 and E_2. It is adaptive in the sense that the redistribution is dependent on input mass functions.

The PCR rule is commutative. It is not disjunctive because when there is no conflict it coincides with the conjunctive rule. It is not conjunctive either because if $m_1 = m_{\{a\}}$ and $m_2 = m_{\{b\}}$, then $m_{\text{pcr}} = \frac{1}{2}m_{\{a\}} + \frac{1}{2}m_{\{b\}}$. The neutral element is the vacuous mass function m_Θ because there is no conflict when one input mass function is this one and m_Θ is the neutral element of the conjunctive rule. A major difficulty is that the PCR rule is not quasi-associative and the computation complexity is not linear in the number of sources. The same remark also holds for Dubois and Prade's rule.

5.2.2 *Compound rules*

There are plenty of ways to build compound rules and we do not have the ambition to list them all. For instance, it is quite obvious that the convex combination of any set of rules is a well defined rule. This can be formalized as follows

Proposition 9. *Let m_1 and m_2 denote two mass functions on Θ. Let $m_{(1)}, \dots, m_{(\ell)}$ denote the combination results obtained by application of ℓ different rules with m_1 and m_2 as inputs. If each rule induces a magma for the mass space, then for any positive reals $(\lambda_i)_{i=1}^{\ell}$ such that $\sum_{i=1}^{\ell}\lambda_i = 1$, the rule returning the mass function $\sum_{i=1}^{\ell}\lambda_i m_{(i)}$ induces a magma as well.*

The proof of this proposition is trivial as the convex combination of mass functions is a mass function.

As an example, Martin and Osswald 2007 proposed a convex combination of Dubois and Prade's rule and PCR rule. Other authors (Delmotte et al. 1995; Delmotte et al. 1996) propose a convex combination of the conjunctive and disjunctive rules where the coefficients are derived from source reliabilities. In general, combination coefficients can be tuned by grid search on a performance measure related to the targeted application.

The approach of Florea et al. 2009 is in line with the idea of a convex combination of the conjunctive and disjunctive rules but with a minor variation in the sense that the result of the combination is enforced to be a normal mass function. This family of rules is called **robust combination rules** (RCR). The combination result mass function m_{rcr} obtained by applying one such rule to input functions m_1 and m_2 is

$$m_{\mathrm{rcr}}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \alpha(\kappa)\, m_1 \cap\!\!\!\!\bigcirc m_2(A) + \beta(\kappa)\, m_1 \cup\!\!\!\!\bigcirc m_2(A) & \text{otherwise} \end{cases}. \tag{5.20}$$

The constraint on parameters α and β allowing m_{rcr} to be a bona fide mass function is

$$(1-\kappa)\,\beta(\kappa) = 1 - \alpha(\kappa). \tag{5.21}$$

The parameters are written as functions of Dempster's degree of conflict because this degree is instrumental to tune the parameters. Any RCR rule is commutative and quasi-associative[13]. If $\beta(0) = 0$, then m_Θ is the neutral element for such RCR rules.

13. This property follows from the associativity of $\cap\!\!\!\!\bigcirc$ and $\cup\!\!\!\!\bigcirc$. One just has to keep track of the results of these combinations to compute coefficients α and β and eventually the RCR combination result.

RCR rules are subsumed by an approach by Martin and Osswald 2007 where the mixing coefficients are dependent on the focal elements of the input mass functions (both in the conjunctive and disjunctive components). This larger class of rules is known as **mix rules**. The combination result mass function m_{mix} obtained by applying one such rule to input functions m_1 and m_2 is

$$m_{\mathrm{mix}}(A) = \sum_{\substack{B,C\subseteq\Theta \\ B\cap C=A}} \psi_1(B,C)\, m_1(B)\, m_2(C) + \sum_{\substack{B,C\subseteq\Theta \\ B\cup C=A}} \psi_2(B,C)\, m_1(B)\, m_2(C). \tag{5.22}$$

As default coefficient choice, the authors propose to use Jaccard indexes for the disjunctive component and one minus this index for the conjunctive one:

$$\psi_1(B,C) = 1 - \frac{|B\cap C|}{|B\cup C|}, \tag{5.23}$$

$$\psi_2(B,C) = \frac{|B\cap C|}{|B\cup C|}. \tag{5.24}$$

Any mix rule is commutative but not necessarily quasi-associative.

5.3 Distance based rules

Dubois et al. 2016 consider fusion problem from an abstract point of view, and merely require conjunctive and disjunctive rules to satisfy the following principle: given items of information $\mathcal{I}_1, \ldots, \mathcal{I}_\ell$ and an information ordering $\sqsubseteq$ relation defined on them, a rule is conjunctive if its result $\mathcal{I}_\cap$ is such that

$$\mathcal{I}_\cap \sqsubseteq \mathcal{I}_i,\ \forall i \in \{1,\ldots,\ell\}$$

and is disjunctive if its result $\mathcal{I}_\cup$ is such that

$$\mathcal{I}_\cup \sqsupseteq \mathcal{I}_i, \ \forall i \in \{1, \ldots, \ell\}.$$

They then recommend (in absence of other information) to follow the LCP principle[14] in the conjunctive case, and the "most committed principle" in the disjunctive case to pick a combination result. This view has the advantage that it makes no a priori assumption about the shape of the rule, nor about the dependence assumption it should satisfy. The remainder of this subsection presents a recent approach from Klein et al. 2016b, 2016a where computable idempotent operators are obtained, starting with a conjunctive operator in 5.3.1 and pursuing with a disjunctive one in 5.3.4.

14. As explained in chapter 1, this principle reads: when several belief functions qualify as uncertainty models that are consistent with evidence, then one should select a belief function with minimal degrees of belief.

5.3.1 A generic way to derive conjunctive rules using partial orders

Rather than seeing a conjunctive combination as a particular operator defined either on the mass functions $m_1, \ldots, m_\ell$ or on the weight functions $w_1, \ldots, w_\ell$, a mass m^* resulting from a conjunction can just be considered as (i) more informative (in the sense of some partial order $\sqsubseteq_f$) than any $m_1, \ldots, m_\ell$ and (ii) one of the least committed elements (in terms of information) among those, in accordance with the LCP. Formally speaking, if we denote by

$$\mathcal{M}_f(m_i) := \left\{ m \in \mathcal{M} \mid m \sqsubseteq_f m_i \right\} \tag{5.25}$$

the set of all mass functions more informative than m_i, then m^* should be such that:

(i) $m^* \in \mathcal{M}_f(m_1) \cap \ldots \cap \mathcal{M}_f(m_\ell)$,

(ii) $\nexists m' \in \mathcal{M}_f(m_1) \cap \ldots \cap \mathcal{M}_f(m_\ell)$ such that $m^* \sqsubset_f m'$.

The first constraint expresses the conjunctive behavior of such an approach. The second constraint says that m^* is a maximal element (i.e. a least committed solution) for admissible solutions subject to the first constraint.

While this solution is generic and does not require any explicit model of dependence, it should be noted that the choice of the partial order to consider is not without consequence. Considering those mentioned in chapter 3, Equation (3.20) tells us that for example that $\mathcal{M}_w(m) \subseteq \mathcal{M}_{pl}(m)$, hence the space of solutions will be potentially much smaller when choosing $\sqsubseteq_w$ rather than $\sqsubseteq_{pl}$. In practice and in accordance with the LCP, it seems safer to choose the most conservative partial orders, i.e., in our case $\sqsubseteq_{pl}$ or $\sqsubseteq_q$. We will see in Section 5.3.5 that it can have an important impact on the combination results, even for simple examples.

While this definition of the cautious result of a conjunctive combination appears natural, it still faces the problem that many different solutions m^* could actually fit the two constraints, as $\sqsubseteq$ is a partial order. This means that to identify a unique solution, we need an additional criterion, that preferably leads to efficient computations. One idea to solve this problem is to use distances that are compatible with $\sqsubseteq$.

5.3.2 Distance based conjunctive operators from soft LCP

To derive new conjunctive operators, Klein et al. 2016b introduce a weaker form of least commitment principle which is called **soft LCP**. This principle states that when there are several candidate mass functions compliant with a set of constraints, the one with minimal distance value from the vacuous mass function[15] should be chosen for some $\sqsubseteq$-compatible distance. This is a soft version of the LCP in the sense that relying only on a given partial order is too demanding and this requirement is relaxed by resorting as well to a distance consistent with the order in question. From an optimization standpoint, the approach consists, however, in adding constraints ensuring uniqueness.

15. It can be argued that uniqueness of the solution is retrieved by resorting to a pre-order spanned by the functional $m \to d\left(m, m_\Theta\right)$

The resulting conjunctive operator, denoted $\sqcap$, depends on the chosen distance d, and is defined as follows

(personal contribution)

Definition 27. for any set of ℓ functions $\{m_1, .., m_\ell\}$, we have

$$m_1 \sqcap .. \sqcap m_l = \underset{m \in \mathcal{M}_f(m_1) \cap .. \cap \mathcal{M}_f(m_\ell)}{\arg\min} d\left(m, m_\Theta\right). \tag{5.26}$$

According to Klein et al. 2016c, corollary 4, the problem induced by the soft LCP is a convex optimization problem with a unique solution if the chosen distance is an L_k norm based distance $d_{f,k}$ that is $\sqsubset_f$-compatible and if $2 \leq k < \infty$. In this case, the operator is parametric and is denoted by $\sqcap_{f,k}$.

Considering results in Klein et al. 2016c, the operator $\sqcap_{f,k}$ can be applied for $f \in \{pl, q\}$. Because there is no reported evidential distance in the literature that is $\sqsubseteq_w$, $\sqsubseteq_d$ or $\sqsubseteq_s$-compatible, the operator $\sqcap_{f,k}$ cannot be easily instanciated for $f \in \{w, d, s\}$. This is not a major drawback, as those partial orders limit the space of solutions by inducing more restrictive spaces $\mathcal{M}_f\left(m_i\right)$. [16]

Concerning $\sqcap_{bel,k}$, there is no theoretical impediment but a practical one. Indeed, one may have $\mathcal{M}_{bel}\left(m_1\right) \cap \mathcal{M}_{bel}\left(m_2\right) = \emptyset$. When $f \in \{pl, q\}$, we know that such intersections are not empty because they always contain $m_\emptyset$.

Operators $\sqcap_{q,k}$ and $\sqcap_{pl,k}$ can be easily implemented using standard solvers available in scientific programming libraries because they amount to a convex minimization problem.

Observe that a soft LCP solution is an LCP solution to the problem presented in the previous subsection as long as m_Θ is the maximum of $\left(\mathcal{M}, \sqsubseteq_f\right)$[17]. Indeed condition (i) is verified by construction. Concerning condition (ii), suppose $\exists m' \in \mathcal{M}_f\left(m_1\right) \cap \ldots \cap \mathcal{M}_f\left(m_\ell\right)$ such that $m^* \sqsubset_f m'$. As d is consistent with $\sqsubset_f$, $m^* \sqsubset_f m' \sqsubset m_\Theta$ implies $d(m', m_\Theta) < d(m^*, m_\Theta)$ which is in contradiction with the very definition of m^*, hence condition (ii) is verified as well.

16. Concerning $\sqsubseteq_w$, one can derive a $\sqsubseteq_w$-compatible distance by computing $\|w_1 - w_2\|_k$. However, the minimization problem solution cannot be formulated as in (5.26) because m_Θ is not the maximum of $(\mathcal{M}, \sqsubseteq_w)$. Yet, $m_\emptyset$ is its minimum and one can look for the mass function with maximal distance from $m_\emptyset$. In this case, we obtain an LCP solution that is actually the cautious rule, i.e. $\oslash = \sqcap_{w,k}$, because this is one of the rare exceptions where the raw application of LCP leads to a unique solution anyway. The inverse pignistic (Dubois et al. 2008) is another example of an LCP problem with a unique solution.

17. This is also true if $m_\emptyset$ is the minimum of $\left(\mathcal{M}, \sqsubseteq_f\right)$ and if the operator consists in maximizing the distance to $m_\emptyset$.

Just for a quick illustration we provide the following example which is a continuation of Denœux 2008, example 2.

Example 23. Let $\Theta = \{a, b, c\}$. Here are two non-dogmatic mass functions along with their combinations under operators $\sqcap_{q,2}$, $\sqcap_{pl,2}$ and other standard approaches.

subset	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Θ
m_1	0	0	0	0.3	0	0	0.5	0.2
m_2	0	0	0.3	0	0	0	0.4	0.3
$m_1 \sqcap_{q,2} m_2$	0	0	0.2	0.1	0	0	0.5	0.2
$m_1 \sqcap_{pl,2} m_2$	0	0	0.3	0	0	0	0.4	0.3
$m_1 \textcircled{\cap} m_2$	0	0	0.6	0.12	0	0	0.2	0.08
$m_1 \owedge m_2$	0	0	0.42	0.09	0	0	0.43	0.06

We see that the conjunctive and cautious rules transfer much more mass to $\{b\}$ than operators $\sqcap_{q,2}$ and $\sqcap_{pl,2}$ do. Also, observe that $m_1 \sqcap_{pl,2} m_2 = m_2$ because $m_2 \sqsubset_{pl} m_1$.

5.3.3 *Properties of distance based conjunctive operators*

The commutativity of the set-intersection and the symmetry property of distance give that $\sqcap_{f,k}$ is commutative. Each operator $\sqcap_{f,k}$ is also idempotent: for any possible solution $m \in \mathcal{M}_f(m_1) \setminus \{m_1\}$, we have $d_{f,k}(m_1, m_\Theta) < d_{f,k}(m, m_\Theta)$ because $d_{f,k}$ is $\sqsubseteq_f$-compatible and $m \sqsubseteq_f m_1 \sqsubseteq_f m_\Theta$, hence $m_1 \sqcap_{f,k} m_1 = m_1$.

Each of these operators are also conjunctive by construction, in the sense that the output mass function is more informative than any of the initial mass functions. Indeed if m_i states that a is not a possible value of the unknown quantity ($pl_i(a) = 0$), then any function in $\mathcal{M}(m_i)$ also states so. Since the combination result belongs to $\mathcal{M}(m_i)$, then this piece of information is propagated by $\sqcap_{f,k}$.

Except for the $f = w$ case[18], these operators are, however, not associative because we can have

$$\mathcal{M}_f\left(m_1 \sqcap_{f,k} m_2\right) \subsetneq \mathcal{M}_f(m_1) \cap \mathcal{M}_f(m_2).$$

18. Remember that $\owedge = \sqcap_{w,k}$ and $\owedge$ is associative (c.f. Denœux 2008)

The above remark is illustrated by the following example in the $f = q$ case.

Example 24. Let $\Theta = \{a, b, c\}$ denote some space. Let us introduce the following mass functions on Θ:

subset	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Θ
m_1	0	0.1	0	0	0	0	0.1	0.8
m_2	0	0	0	0.1	0.1	0	0	0.8
$m_1 \sqcap_{q,2} m_2$	0	1/15	1/15	0	1/15	0	0	0.8
q_1	1	0.9	0.9	0.8	0.9	0.8	0.9	0.8
q_2	1	0.9	0.9	0.9	0.9	0.8	0.8	0.8
$q_1 \wedge q_2$	1	0.9	0.9	0.8	0.9	0.8	0.8	0.8
q_{12}	1	1/15	1/15	0.8	1/15	0.8	0.8	0.8

where q_{12} denotes the commonality function in correspondence with $m_1 \sqcap_{q,2} m_2$. Let m_3 denote the following mass function

subset	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Θ
m_3	0	0.1	0	0	0.3	0	0.2	0.4
q_3	1	0.5	0.6	0.4	0.9	0.4	0.6	0.4

We have $q_3 \leq q_{12}$ and thus $m_3 \in \mathcal{M}_q(m_1) \cap \mathcal{M}_q(m_2)$. However $m_1 \sqcap_{q,2} m_2 \not\sqsubseteq m_3$ and thus $m_3 \notin \mathcal{M}_q\left(m_1 \sqcap_{q,2} m_2\right)$.

Fortunately, when $f \in \{pl, q\}$, the constraints of the minimization problem can be stored and updated iteratively, meaning that $\sqcap_{f,k}$ is quasi-associative [19]. Let c denote a set function from 2^Θ to $[0;1]$ which is meant to store the problem constraints. Algorithm 5.1 allows to compute combinations using $\sqcap_{q,k}$ sequentially. The same algorithm works for $\sqcap_{pl,k}$. In practice, what we simply do is storing, for each set A, the lowest commonality (resp. plausibility) value encountered in $\{m_1, \ldots, m_\ell\}$.

19. Quasi-associativity is induced by the associativity of the entrywise minimum. The picture is not that simple for $f \in \{s, d\}$, and should one identify a $\sqsubseteq_s$ or $\sqsubseteq_d$-compatible distance making the corresponding operators $\sqcap_s$ and $\sqcap_d$ operational, then obtaining quasi-associativity is not ensured. Indeed, to our knowledge the characterization of the set of inner approximations of a mass function does not translate in compact constraints for these partial orders. For instance, we know from Cuzzolin 2004 that $\mathcal{M}_d(m_i)$ is a simplex with at most N vertices. Yet, in general the intersection of simplices is not a simplex but a polytope whose vertices are not easily derived and can grow in number after each iteration. Whether there is an easy way to characterize these intersections is a topic for further research.

Algorithm 5.1: Sequential combination using $\sqcap_{q,k}$

entries : $\{m_1, .., m_\ell\}, k \geq 2$.
$c \leftarrow \min\{q_1; q_2\}$ (entrywise minimum).
$m \leftarrow m_1 \sqcap_{q,k} m_2$.
for i from 3 to ℓ **do**
  $c \leftarrow \min\{c; q_i\}$ (entrywise minimum).
  $m \leftarrow \arg\min_{m'} d_{q,k}(m', m_\Theta)$ subject to $q' \leq c$.
end for
return m.

It can be argued that the choice of screening distance values from the least committed mass function in definition 27 is somewhat arbitrary. The following lemma shows that, for $\sqcap_q$ and $\sqcap_{pl}$, another relevant choice yields the same operators:

Lemma 5. *For $f \in \{q, pl\}$ and for any finite integer k such that $k \geq 2$, one has:*

(personal contribution)
See A.4 for a proof.

$$m_1 \sqcap_{f,k} .. \sqcap_{f,k} m_l = \arg\min_{m \in \mathcal{M}_f(m_1) \cap .. \cap \mathcal{M}_f(m_\ell)} d_{f,k}(m, m_\Theta) = \arg\max_{m \in \mathcal{M}_f(m_1) \cap .. \cap \mathcal{M}_f(m_\ell)} d_{f,k}(m, m_\emptyset). \quad (5.27)$$

Getting nearer to the least committed state of belief is thus equivalent to drifting apart from the most committed one for these two operators.

Another interesting property to investigate is the compatibility with Dempster's conditioning (1.25). The next proposition shows that it is retrieved as a special case of the $\sqcap_{q,k}$ conjunctive rule.

Proposition 10. *Let m_0 denote a mass function. For any finite integer k such that $k \geq 2$ and any subset $A \subseteq \Theta$, we have*

(personal contribution)
See A.4 for a proof.

$$m_0 \sqcap_{q,k} m_A = m_{0|A}. \quad (5.28)$$

Concerning $\sqcap_{pl,k}$, no result is available but simulations suggest that it may be conjectured.

Another property that can be sometimes interesting is invariance with respect to refinement. As shows the next example, the operators $\sqcap_{f,k}$ do not commute with refining mappings, for the main reasons that distances are in general not invariant with respect to refinements (for a discussion about this, see Destercke and Burger 2013).

Example 25. Let $\Theta = \{a, b, c\}$ and $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$ denote two spaces. Suppose there exist a refining mapping r such that :

$$\begin{aligned} r(a) &= \{y_1, y_4\}, \\ r(b) &= \{y_2\}, \\ r(c) &= \{y_3\}. \end{aligned}$$

Let us introduce the following mass functions on Θ:

subset	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Θ
m_1	0.1	0	0.1	0.5	0.1	0	0.1	0.1
m_2	0	0	0	0.3	0.1	0.3	0	0.3

Let us denote by m'_1 and m'_2 the mass functions on $\mathcal{Y}$ induced by r from m_1 and m_2, respectively. The mass function $m'_{1\sqcap_q 2}$ induced by r from $m_1 \sqcap_q m_2$ is not equal to $m'_1 \sqcap_q m'_2$. In particular, we have $m'_1 \sqcap_q m'_2(\{y_3\}) = 0.2$ while $m'_{1\sqcap_q 2}(\{y_3\}) = 0.1$.

We chose $k = 2$ in this example and omit it in the operator notation: $\sqcap_q = \sqcap_{q,2}$.

Although informative partial orders are preserved after refinement, the sets of more informative functions $\mathcal{M}_f(m_i)$ are different. In example 25, the hypothesis a is refined into two elements: y_1 and y_4. This implies increased freedom in the selection of the mass function minimizing the distance from the vacuous function. In general, there is no reason why this solution should be in correspondence (through mapping r) with the solution obtained without refining.

A last point that deserves investigation is the presence of a neutral element.

Proposition 11. *For any finite integer k such that $k \geq 2$ and any $f \in \{d, s, q, pl\}$, the unique neutral element of operator $\sqcap_{f,k}$ is the vacuous mass function m_Θ.*

(personal contribution) See A.4 for a proof.

5.3.4 Distance based disjunctive operators

In the same fashion as the conjunctive case, one can consider that a mass function m^* resulting from a disjunction should be (i) less informative (in the sense of some partial order $\sqsubseteq_f$) than any $m_1, \ldots, m_\ell$ and (ii) should be among the most committed elements (in terms of information) among those. This is a dual reasoning as LCP. Formally speaking, if we denote by

$$\mathcal{G}_f(m_i) := \left\{ m \in \mathcal{M} \mid m_i \sqsubseteq_f m \right\} \quad (5.29)$$

the set of all mass functions less informative than m_i, then m^* should be such that:

(i) $m^* \in \mathcal{G}_f(m_1) \cap \ldots \cap \mathcal{G}_f(m_\ell)$,

(ii) $\nexists m' \in \mathcal{G}_f(m_1) \cap \ldots \cap \mathcal{G}_f(m_\ell)$ such that $m' \sqsubset_f m^*$.

Again, such a procedure does not lead to a unique solution in general (except when $f = v$). One way to circumvent this issue is to select the mass function in $\mathcal{G}_f(m_1) \cap \ldots \cap \mathcal{G}_f(m_\ell)$ with minimal distance from the minimum of $\left(\mathcal{M}, \sqsubseteq_f\right)$ (if it exists) as long as the chosen metric is $\sqsubset_f$-compatible. The resulting disjunctive operator, denoted $\sqcup$, depends on the chosen distance d, and is defined as follows

Definition 28. for any set of ℓ functions $\{m_1, .., m_\ell\}$, we have

(personal contribution)

$$m_1 \sqcup .. \sqcup m_l = \underset{m \in \mathcal{G}_f(m_1) \cap .. \cap \mathcal{G}_f(m_\ell)}{\arg\min} d\left(m, m_\emptyset\right). \quad (5.30)$$

Let us focus on $\left(\mathcal{M}, \sqsubseteq_f\right)$, $f \in \{pl, q\}$, where a unique minimal element exists and is $m_\emptyset$. According to corollary 3 in Klein et al. 2016c, this problem is a convex optimization problem with a unique solution if the chosen L_k norm based distance $d_{f,k}$ is $\sqsubset_f$-compatible and if $2 \leq k < \infty$. The corresponding parametric operator is denoted by $\sqcup_{f,k}$.

The fact that $m_\emptyset$ is the most committed mass function for $\sqsubseteq_q$ and $\sqsubseteq_{pl}$ is not very intuitive. The following lemma delivers a better intuition as to what operators $\sqcup_q$ and $\sqcup_{pl}$ consist of, as it shows that they can be understood as the maximization of the distance from the vacuous mass function which is more intuitive.

(personal contribution)

Lemma 6. *For $f \in \{q, pl\}$ and for any finite integer k such that $k \geq 2$, one has:*

$$m_1 \sqcup_{f,k} .. \sqcup_{f,k} m_l = \underset{m \in \mathcal{G}_f(m_1) \cap .. \cap \mathcal{G}_f(m_\ell)}{\arg\min} d_{f,k}(m, m_\emptyset) = \underset{m \in \mathcal{G}_f(m_1) \cap .. \cap \mathcal{G}_f(m_\ell)}{\arg\max} d_{f,k}(m, m_\Theta). \tag{5.31}$$

The proof of this lemma is identical to that of Lemma 5. The combination operators $\sqcup_{q,k}$ and $\sqcup_{pl,k}$ have similar properties as their conjunctive counterparts. They are commutative, idempotent and quasi-associative[20] but not invariant to refinement. Quasi-associativity is achieved using algorithm 5.2 which is almost the same as algorithm 5.1.

20. Again, when $f = v$, $\sqcup_{v,k} = \circledvee$ and this rule is associative (Denœux 2008).

Algorithm 5.2: Sequential combination using $\sqcup_{q,k}$

entries : $\{m_1, .., m_\ell\}$, $k \geq 2$.
$c \leftarrow \max\{q_1; q_2\}$ (entrywise maximum).
$m \leftarrow m_1 \sqcap_{q,k} m_2$.
for i from 3 to ℓ **do**
  $c \leftarrow \max\{c; q_i\}$ (entrywise maximum).
  $m \leftarrow \arg\min_{m'} d_{q,k}(m', m_\emptyset)$ subject to $q' \geq c$.
end for
return m.

The neutral element of some of the disjunctive operators is given by the following proposition.

(personal contribution)

Proposition 12. *For any finite integer k such that $k \geq 2$ and any $f \in \{d, s, q, pl\}$, the unique neutral element of operator $\sqcup_{f,k}$ is the total conflict mass function $m_\emptyset$.*

The proof of proposition 12 is very similar as the one of proposition 12 and is thus omitted. The key point is that $m_\emptyset$ is the minimum of $\left(\mathcal{M}, \sqsubseteq_f\right)$ for $f \in \{d, s, q, pl\}$.

Just for a quick illustration we provide the following example which is a continuation of Denœux 2008, example 7.

Example 26. Let $\Theta = \{a, b, c\}$. Here are two subnormal mass functions along with their combinations under operators $\sqcup_{q,2}$, $\sqcup_{pl,2}$ and other standard approaches.

subset	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Θ
m_1	0.1	0	0	0.3	0	0	0.6	0
m_2	0.1	0	0.5	0	0	0	0.4	0
$m_1 \sqcup_{q,2} m_2$	0.1	0	0	0.3	0	0	0.6	0
$m_1 \sqcup_{pl,2} m_2$	0.1	0	0	0.3	0	0	0.6	0
$m_1 \bigcirc\!\!\!\!\!\cup\, m_2$	0.01	0	0.05	0.18	0	0	0.64	0.12
$m_1 \bigcirc\!\!\!\!\!\vee\, m_2$	0.006	0	0.0298	0.1071	0	0	0.2143	0.6429

We see that the disjunctive and bold rules transfer much more mass to Θ than operators $\sqcup_{q,2}$ and $\sqcup_{pl,2}$ do. Also, observe that $m_1 \sqcup_{q,2} m_2 = m_1 \sqcup_{pl,2} m_2 = m_1$ because $m_2 \sqsubset_q m_1$ and $m_2 \sqsubset_{pl} m_1$.

5.3.5 Related works: discussion and experiments

This section provides two quick comparisons of distance based operators and popular combination rules. They demonstrate that distance based operators allow to redistribute masses more gradually than standard approaches. The sensitivity w.r.t. parameters k and f is studied in Klein et al. 2016a, section 5.3. Implementation details when $k = 2$ are given in Klein et al. 2016a, appendix A.

A comparison with related works in the conjunctive case

As said earlier, the main motivation in obtaining an idempotent rule is to circumvent the source independence assumption when it is obviously unreasonable or hard to assess. There are many works that have addressed the problem of deriving alternatives to Dempster's rule or the conjunctive rule that do not rely on independence assumptions.

A principled and common approach is to rely on a set of axiomatic properties as done by Dubois et al. 2016 or to adapt existing rules from other frameworks as proposed by Destercke and Dubois 2011. In practice, such axioms seldom lead to a unique solution, and it is then necessary to advocate more practical solutions. A distance based rule can be seen as an instance of such an approach, where the axiom consists in using the LCP over sets of f-included masses, and the practical solution is to use a distance compliant with such an axiom. The approaches of Cattaneo 2003 and Denœux 2008 can also be seen as instances of the same principle. The former proposes to solve a conflict minimization problem rather than minimizing the informative content (thus not strictly following an LCP principle), while the latter focuses on using the set $\mathcal{M}_w(m_1) \cap \ldots \cap \mathcal{M}_w(m_\ell)$ and the order $\sqsubseteq_w$, and demonstrates that in this case there is a unique LCP solution known in closed form. Finally, an idea of combination by distance minimization is suggested but not studied in Cattaneo 2003, 2011. As will be seen in 5.4, this author pursues a different goal anyway as the constraints are on marginal mass functions.

Table 5.2 summarizes some basic theoretical properties satisfied by operators $\bigcirc\!\!\!\!\!\cap\,$, $\oplus$, $\bigcirc\!\!\!\!\!\wedge\,$ and $\sqcap_{f,k}$.

Let us illustrate the operator discrepancies on a simple situation inspired from Zadeh's counter-example (c.f. Zadeh 1986). Suppose $m_1 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{a\}}$ and $m_2 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{c\}}$ are

operator	condition for use	commutativity	associativity	idempotence	invariance w.r.t. refinement	neutral element
$\textcircled{\cap}$	none	yes	yes	no	yes	m_Θ
$\oplus$	$m_{1\textcircled{\cap}2}(\emptyset) < 1$	yes	yes	no	yes	m_Θ
$\textcircled{\wedge}$	$m_1(\Theta) > 0$ and $m_2(\Theta) > 0$	yes	yes	yes	yes	none
$\sqcap_{q,k}$	none	yes	quasi	yes	no	m_Θ
$\sqcap_{pl,k}$	none	yes	quasi	yes	no	m_Θ

Table 5.2: Summary of rule properties.

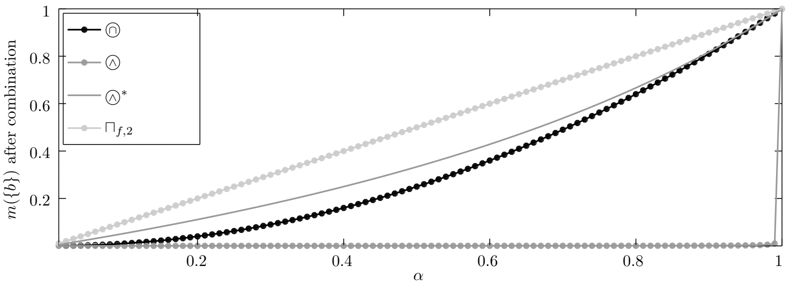


Figure 5.1: Mass assigned to $\{b\}$ after combination of $m_1 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{a\}}$ and $m_2 = \alpha m_{\{b\}} + (1-\alpha) m_{\{c\}}$ with $\textcircled{\cap}$, $\textcircled{\wedge}$ and $\sqcap_{f,2}$.

two mass functions on a frame $\Theta = \{a, b, c\}$. Figure 5.1 shows the mass assigned to $\{b\}$, the commonly supported element of m_1 and m_2, after combination by $\textcircled{\cap}$, $\textcircled{\wedge}$ and $\sqcap_{f,2}$. The same masses are obtained for $f \in \{pl, q\}$. A very small mass $\epsilon = 1e-4$ was assigned to Θ while a mass $\frac{\epsilon}{2}$ was removed from each focal element of each input mass function when using $\textcircled{\wedge}$ so as to circumvent the non-dogmatic constraint.

As could be expected, the distance based operator tries to maintain as much evidence on $\{b\}$ as possible. A striking fact is that we have obviously $m_1 \sqcap_{f,2} m_2 (\{b\}) = \alpha$. More precisely, we have $m_1 \sqcap_{f,2} m_2 = (1-\alpha)\, m_\emptyset + \alpha m_{\{b\}}$. This result can be proved for any finite $k \geq 2$ when $f = q$. Let $q_{1\wedge 2}$ denote the entrywise minimum of functions q_1 and q_2. In this particular setting, $q_{1\wedge 2}$ happens to be a valid commonality function. Consequently, $m_{1\wedge 2} \in \mathcal{M}_q(m_1) \cap \mathcal{M}_q(m_2)$. By definition of the partial order $\sqsubseteq_q$, for any function $m \in \mathcal{M}_q(m_1) \cap \mathcal{M}_q(m_2)$, we have $m \sqsubseteq_q m_{1\wedge 2}$. Since we also have $m_{1\wedge 2} \sqsubseteq_q m_\Theta$ and $d_{q,k}$ is $\sqsubseteq_q$-compatible, then $m_1 \sqcap_{q,k} m_2 = m_{1\wedge 2}$. In other words, the distance based operators coincide with the minimum rule applied to commonalities in this case. When $f = pl$, the result can also be proved. For any $m \in \mathcal{M}_{pl}(m_1) \cap \mathcal{M}_{pl}(m_2)$, the constraints $pl(\{a\}) = pl(\{c\}) = 0$ imply that only $\{b\}$ and $\emptyset$ are possible focal sets for m. More precisely, this actually implies that $\mathcal{M}_{pl}(m_1) \cap \mathcal{M}_{pl}(m_2)$ is the segment $(1-\beta)\, m_\emptyset + \beta m_{\{b\}}$ in $\mathcal{M}$ parametrized by $\beta \in [0;\alpha]$. $\sqsubseteq_{pl}$ is a total order for this segment. From relation (4.19), we obtain $m_1 \sqcap_{pl,k} m_2 = (1-\alpha)\, m_\emptyset + \alpha m_{\{b\}}$.

A closed form expression for the other rules can also be obtained. It is easy to see that $m_1 \textcircled{\cap} m_2 = \left(1-\alpha^2\right) m_\emptyset + \alpha^2 m_{\{b\}}$. Concerning the cautious rule, taking the limit $\epsilon \to 0$, we obtain

$$m_1 \textcircled{\wedge} m_2 = \begin{cases} m_\emptyset & \text{if } \alpha < 1 \\ m_{\{b\}} & \text{if } \alpha = 1 \end{cases}.$$

This example shows also that the behavior of Denœux's cautious rule $\textcircled{\wedge}$ may not be so cautious, as it keeps no mass on $\{b\}$ except when $\alpha = 1$. This is a quite bold behavior, due mainly to the fact that $\mathcal{M}_w$ induces stronger constraints than $\mathcal{M}_{pl}$ or $\mathcal{M}_q$. This clearly shows that while idempotence is a pre-requisite to have a cautious attitude towards source (in)dependence, it is not sufficient to guarantee a really cautious behavior when mass functions are not identical. Even the conjunctive rule appears to have an intermediate behavior as compared to the two others, hence could be termed as more cautious than Denœux's rule. In Denœux 2008, example 3, Denœux actually shows that using $\owedge$ can yield a mass function that is w-included in the result of the conjunctive combination. It should also be stressed that this is a limit use case for $\owedge$, hence arguably an unfavourable one.

The normalized versions of these three rules deserve also some comments. This time, we obtain $m_1 \sqcap^*_{q,k} m_2 = m_1 \sqcap^*_{pl,k} m_2 = m_1 \oplus m_2 = m_{\{b\}}$ which is the result criticized by Zadeh. In contrast, the normalized cautious rule achieves a progressive reduction of the support given to $\{b\}$ as α decreases. The normalized cautious rule appears to offer an intermediate behavior as compared to the conjunctive rule and either of the unnormalized operator $\sqcap_{q,k}$ or $\sqcap_{pl,k}$. In particular, when $\alpha = \frac{1}{2}$, $m_1 \owedge^* m_2$ is the uniform Bayesian mass function whereas operators $\sqcap_{q,k}$ and $\sqcap_{pl,k}$ are still giving some support to $\{b\}$ solely. This time, the rule $\owedge^*$ appears indeed more cautious than ours, but it could be argued to no longer be really conjunctive, as it supports every element whereas each source respectively discarded one as totally impossible.

A comparison with related works in the disjunctive case

The distance based disjunctive operators are compared to standard disjunctives rules: $\circledcup$ and $\ovee$. Table 5.3 summarizes some basic theoretical properties satisfied by operators $\circledcup$, $\ovee$ and $\sqcup_{f,k}$.

operator	condition for use	commutativity	associativity	idempotence	invariance w.r.t. refinement	neutral element
$\circledcup$	none	yes	yes	no	yes	$m_\emptyset$
$\ovee$	$m_1(\emptyset) > 0$ and $m_2(\emptyset) > 0$	yes	yes	yes	yes	none
$\sqcup_{q,k}$	none	yes	quasi	yes	no	$m_\emptyset$
$\sqcup_{pl,k}$	none	yes	quasi	yes	no	$m_\emptyset$

Table 5.3: Basic properties of operators $\circledcup$, $\ovee$ and $\sqcup_{f,k}$.

Let us illustrate the disjunctive operator discrepancies on a simple situation analogous to the experiment presented in the conjunctive case. Suppose $m_1 = \alpha m_{\{a,b\}} + (1-\alpha)\, m_{\{a,c\}}$ and $m_2 = \alpha m_{\{a,b\}} + (1-\alpha)\, m_{\{b,c\}}$ are two mass functions on a frame $\Theta = \{a,b,c\}$. Figure 5.2 shows the mass assigned to $\{a,b\}$ after combination by $\circledcup$, $\ovee$ and $\sqcup_{f,2}$. The same masses are obtained for $f \in \{pl, q\}$. A very small mass $\epsilon = 1e-4$ was assigned to $\emptyset$ while a mass $\frac{\epsilon}{2}$ was removed from each focal element of each input mass function when using $\ovee$ so as to circumvent the subnormality constraint.

The aspect of figure 5.2 is remarkably similar to the conjunctive case but the conclusions that we will draw from it are different. As could be expected, the distance based disjunctive operators try to maintain as much evidence on $\{a,b\}$ as possible. A striking fact is that we have obviously $m_1 \sqcup_{f,2} m_2 (\{a,b\}) = \alpha$. More precisely, we have $m_1 \sqcup_{f,2} m_2 = (1-\alpha)\, m_\Theta + \alpha m_{\{a,b\}}$.

This result can be proved for any finite $k \geq 2$ when $f = q$. Let $q_{1\vee 2}$ denote the entrywise maximum of functions q_1 and q_2. We have

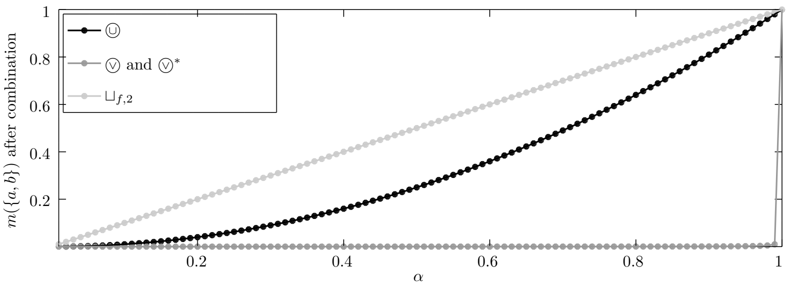


Figure 5.2: Mass assigned to $\{a,b\}$ after combination of $m_1 = \alpha m_{\{a,b\}} + (1-\alpha)\, m_{\{a,c\}}$ and $m_2 = \alpha m_{\{a,b\}} + (1-\alpha) m_{\{b,c\}}$ with $\ocup$, $\ovee$ and $\sqcup_{f,2}$.

$q_{1\vee 2}(\{a\}) = q_{1\vee 2}(\{b\}) = 1$ which implies that for any $m \in \mathcal{G}_q(m_1) \cap \mathcal{G}_q(m_2)$, only supersets of $\{a,b\}$ can be focal sets of m. In this example, this means that $m = \beta m_{\{a,b\}} + (1-\beta)\, m_\Theta$ with $\beta \in [0;1]$. Observing that if q denotes the commonality function in correspondence with m, we also have

$$q_{1\vee 2}(\{c\}) \leq q(\{c\}),$$
$$\Leftrightarrow 1-\alpha \leq \sum_{B \subseteq \{c\}} m(B).$$

Since Θ is the only set that is a superset of both $\{a,b\}$ and $\{c\}$, we deduce that $m(\Theta) \geq 1-\alpha$ or equivalently $\beta \leq \alpha$.

More precisely, this actually implies that $\mathcal{G}_q(m_1) \cap \mathcal{G}_q(m_2)$ is the segment $(1-\beta)\, m_\Theta + \beta m_{\{a,b\}}$ in $\mathcal{M}$ parametrized by $\beta \in [0;\alpha]$. $\sqsubseteq_q$ is a total order for this segment. This segment can also be seen as the set of mass functions obtained by discounting $\alpha m_{\{a,b\}} + (1-\alpha)\, m_\Theta$. From relation (4.19), we obtain $m_1 \sqcup_{q,k} m_2 = (1-\alpha)\, m_\Theta + \alpha m_{\{a,b\}}$. When $f = pl$, the same reasoning applies.

A closed form expression for the other rules can also be obtained. It is easy to see that $m_1 \ocup m_2 = \left(1-\alpha^2\right) m_\Theta + \alpha^2 m_{\{a,b\}}$. Concerning the bold rule, taking the limit $\epsilon \to 0$, we obtain

$$m_1 \ovee m_2 = \begin{cases} m_\Theta & \text{if } \alpha < 1 \\ m_{\{a,b\}} & \text{if } \alpha = 1 \end{cases}.$$

Like in the conjunctive example, the behavior of the bold rule $\ovee$ is symptomatic of the fact that $\mathcal{G}_w$ induces stronger constraints than $\mathcal{G}_{pl}$ or $\mathcal{G}_q$. The bold rule keeps no mass on $\{a,b\}$ except when $\alpha = 1$. Finally, the disjunctive rule appears to have an intermediate behavior as compared to the two others. Also, this time all normalized versions of these rules coincide with their unnormalized counterparts.

5.4 Combination on product spaces

In the previous section, combination rules are seen as binary operations on $\mathcal{M} \times \mathcal{M}$ but this does not mean that on the way from a pair $(m_1, m_2) \in \mathcal{M}^2$ to a combined mass function $m \in \mathcal{M}$ we cannot visit larger spaces than $\mathcal{M}$. In particular, when we want to combine two sources $\left(\Omega_1, \sigma_{\Omega_1}, \mu_1, \Gamma_1\right)$ and $\left(\Omega_2, \sigma_{\Omega_2}, \mu_2, \Gamma_2\right)$ where the codomain of Γ_1 and Γ_2 is Θ, we can be interested in the set of mass functions on

$\Theta \times \Theta$ whose marginals are the mass functions induced by each source. Let us introduce the following set of mass functions

$$\mathcal{J}_{12}^{\Theta\times\Theta} = \left\{ m \in \mathcal{J}^{\Theta\times\Theta} | \sum_{E\subseteq\Theta} m(E,B) = m_1(B) \text{ and } \sum_{F\subseteq\Theta} m(B,F) = m_2(B), \forall B \subseteq \Theta \right\},$$

where $\mathcal{J}^{\Theta\times\Theta}$ is the set of functions on $\Theta \times \Theta$ such that their focal elements are cross products $E \times F$ for E and F subsets of Θ. In this case we note $m(E \times F) = m(E,F)$ and the function m is called a **joint mass function**. We have

$$\mathcal{J}^{\Theta\times\Theta} \subsetneq \mathcal{M}^{\Theta\times\Theta}$$

because subsets of $\Theta \times \Theta$ can have more complicated forms[21]. For any joint mass function m, the marginalization (2.16) boils down to a summation

$$\sum_{\substack{B\subseteq\Theta_1\times\Theta_2 \\ \text{proj}_{\Theta_1}(B)=A}} m(B) = \sum_{E\subseteq\Theta_2} m(A,E), \forall A \subseteq \Theta_1.$$

Here we add subscripts to space Θ so as to specify the corresponding source but $\Theta_1 = \Theta_2 = \Theta$.

This section presents a number of approaches that resort to such joint mass functions to produce the combination result.

In this section, we need to specify the space $\mathcal{Z}$ spanning the mass space, which is thus now denoted as $\mathcal{M}^{\mathcal{Z}}$.

21. see figure 2.2 for an example of a subset that is not a cross product.

5.4.1 Approximate joint mass functions

Alleging we know the joint mass function m whose marginals are m_1 and m_2 induced by sources with unknown dependencies, Cattaneo 2003 derives the conjunctively combined mass function Conj $[m]$ as

$$\text{Conj}[m](B) = \begin{cases} \frac{1}{1-K(m)} \sum\limits_{\substack{E,F\subseteq\Theta \\ E\cap F=B}} m(E,F) & \text{if } B \neq \emptyset \\ 0 & \text{if } B = \emptyset \end{cases}, \tag{5.32}$$

where $K(m) = \sum\limits_{\substack{E,F\subseteq\Theta \\ E\cap F=\emptyset}} m(E,F)$. If the joint mass function factorizes as the product of the marginals, then Dempster's rule is retrieved. Destercke and Dubois 2011 use the same mechanism in an unnormalized context and the conjunctive rule is retrieved under the same assumption. This observation is one way to better understand what are necessary conditions to apply Dempster's rule or the conjunctive rule.

The real challenge in these approaches is obviously to obtain the joint mass function from marginals which is an ill-posed problem for infinitely many joint mass functions share identical marginals. Among other possibilities, Cattaneo 2003 proposes to obtain the joint mass function through conflict minimization

$$m^* \in \underset{m_{12}\in\mathcal{J}_{12}^{\Theta\times\Theta}}{\arg\min} K(m_{12}) \tag{5.33}$$

This optimization problem has in general multiple solutions. He thus argues that among the minimally conflictual joint mass functions, the least specific (3.7) among the Conj $[m^*]$ should be retained as final result. Unfortunately, there may still be multiple mass functions minimizing both conflict and specificity. When the induced rule is well defined, the obtained combination is commutative and idempotent.

Destercke and Dubois 2011 advocate for least committed solutions in order to select joint mass functions in $\mathcal{J}_{12}^{\Theta\times\Theta}$ but again such solutions are not unique. They also prove the existence of idempotent conjunctive rules that can be built using the commensuration method [22] so that the solution is in $\mathcal{J}_{12}^{\Theta\times\Theta}$ and has maximal expected cardinality (3.8). However, they fail to obtain an operational way to compute the combined mass function in the general case.

Cattaneo 2011 made a second contribution in the wake of the aforementioned one. He argues that the combined joint mass should be

$$m \in \text{Conj}\left[\underset{m_{12}\in\mathcal{J}_{12}^{\Theta\times\Theta}}{\arg\max}\, pl_{12}\right] = \underset{m_0\in\mathcal{M}_s(m_1)\cap\mathcal{M}_s(m_2)}{\arg\max} pl_0, \tag{5.34}$$

or the equivalent process with commonality functions. Since there is again, in general, multiple solutions, the author recommends to use an informative content measure to select the least informative function among candidate ones. Actually the approach of Destercke and Dubois 2011 is in line with the above where the measure in question is the expected cardinality. Any rule of this kind is ensured to be conjunctive, commutative, idempotent, quasi-associative and to generalize Dempster's conditioning.

However, these rules are not fully specified since the information measure will not mandatorily yield a unique solution either and the maximization problem is not trivial to solve. Cattaneo thus suggests to use another rule as a proxy for the above one. This rule is denoted by $\circledwedge$ and specified by the following system of equations

$$\tilde{m}(A) = \max\left\{0; 1 - pl_1 \curlywedge pl_2(A^c) - \sum_{B\subsetneq A}\tilde{m}(B)\right\}, \forall A\subseteq\Theta, \tag{5.35}$$

where

$$pl_1 \curlywedge pl_2(A) = \min_{B\subseteq A} pl_1(B) + pl_2(A\setminus B). \tag{5.36}$$

The system can be solved incrementally by starting to compute $\tilde{m}(\emptyset)$. The function $\tilde{m}$ needs to be renormalized to obtain a bona fide mass function so the combination result is

$$m_1 \circledwedge m_2 = \begin{cases} \frac{\tilde{m}}{\sum\limits_{\substack{B\subseteq\Theta\\B\neq\emptyset}}\tilde{m}(B)} & \text{if } \sum\limits_{\substack{B\subseteq\Theta\\B\neq\emptyset}}\tilde{m}(B) > 0 \\ m_\Theta & \text{otherwise}\end{cases}. \tag{5.37}$$

This rule is also conjunctive, commutative, idempotent, quasi-associative and a generalization of Dempster's conditioning. On top of that, it is invariant w.r.t. refinement. The author actually introduces two general class of rules with equivalent properties (one class for plausibilities and another one for commonalities). This rule is an approximation of the objective one as it coincides in a number of situations (when $n = 2$ for instance) but when it does not, we cannot really quantify how close the

22. A pair of mass functions such that they have the same number of focal elements and the same mass values are said to be commensurate. For instance, $m_1 = 0.25m_{\{a\}} + 0.75m_{\{b,c\}}$ and $m_2 = 0.25m_{\{a,b\}} + 0.75m_{\{a,c\}}$ are commensurate.

The commensuration method (Dubois and Yager 1992) consists in subdividing (in an additive way) each mass value into a set of smaller values so that subdivisions of mass functions are commensurate. They can be then combined by assigning mass subdivisions to the intersection of the focal elements to which they belong in each of the input mass functions. The result is dependent on the way masses are split and focal element are indexed.

two rules are. As final remark, the vacuous mass function is the neutral element of both the objective and proxy rules. From (5.34), this is obvious as $\mathcal{M}_s(m_\Theta) = \mathcal{M}$. For the proxy rule, we can write

$$pl_1 \curlywedge pl_\Theta(A) = \min_{B\subseteq A} pl_1(B) + pl_\Theta(A \setminus B), \tag{5.38}$$

$$= \min_{B\subseteq A} pl_1(B) + 1, \tag{5.39}$$

$$= \min_{B\subseteq A} pl_1(B), \tag{5.40}$$

$$= pl_1(A). \tag{5.41}$$

In this case, the computation of $\tilde{m}$ is just a straightforward computation of m_1 from pl_1.

5.4.2 *Meta-information integration*

We have seen in 2.6 that belief functions are instrumental to tackle the problem of information fusion under meta-knowledge regarding the truthfulness of the sources. In particular, α-junctions allow to rectify special forms of untruthfulnesses inside the combination process. Pichon et al. 2014 proposed a more general combination framework that generalizes α-junctions and allow to perform combination under various forms of untruthfulness (partially or completely deceptive information) or irrelevance (lack of useful information).

When a piece of information is irrelevant, we have to throw it away. For instance, a broken sensor may deliver a constant value that is just meaningless. When a piece of information is untruthful, there may be a way to rectify it and integrate it. For instance, a biased sensor delivers useful information once the bias is subtracted to the reported value.

Example 27. Let T and R denote two binary variables representing respectively the truthfulness and the relevance of a piece of information. Let $\mathcal{S}$ denote the space of pairs (R, T). Define the following multi-valued mapping $\Gamma_A : \mathcal{S} \to 2^\Theta$:

$$\Gamma_A(R,T) = \begin{cases} A & \text{if } R = T = 1 \\ A^c & \text{if } R = 1 \text{ and } T = 0 \\ \Theta & \text{otherwise} \end{cases}.$$

The source $(\mathcal{S}, \sigma_\mathcal{S}, \mu_s, \Gamma_A)$ induces a belief function on Θ. The corresponding mass function m is specified by two probability values: $q = \mu_s(T = 1|R = 1)$ and $p = \mu_s(R = 1)$. We obtain the following separable mass function

$$m(B) = \begin{cases} p.q & \text{if } B = A \\ p.(1-q) & \text{if } B = A^c \\ 1-p & \text{if } B = \Theta \\ 0 & \text{otherwise} \end{cases}.$$

This mass function is interpreted as the rectified version of the testimony $\theta \in A$ under uncertain truthfulness and relevance. When information is surely truthful as long as it is relevant ($q = 1$), the discounting operation (2.10) is retrieved and the rectified mass function is m_A^{1-p}.

The above example can be generalized in several ways. If the basic testimony is not a categorical mass function m_A but a general one m induced by the source $(\Omega, \sigma_\Omega, \mu, \Gamma)$, then we can define a rectified source $(\Omega \times \mathcal{S}, \sigma_{\Omega\times\mathcal{S}}, \mu \otimes \mu_s, \tilde{\Gamma})$ and

$$\tilde{\Gamma}(\omega, R, T) = \begin{cases} \Gamma(\omega) & \text{if } R = T = 1 \\ \Gamma(\omega)^c & \text{if } R = 1 \text{ and } T = 0\,, \\ \Theta & \text{otherwise} \end{cases} \tag{5.42}$$

$$= \Gamma_{\Gamma(\omega)}(R, T)\,. \tag{5.43}$$

In a more general case, meta-information also comes in the form of a belief function on $\mathcal{S}$. So here, we have two sources:

- a source containing information on θ: $(\Omega, \sigma_\Omega, \mu, \Gamma)$ which induces a belief function on Θ. The corresponding unnormalized mass function is denoted by $m = \mu \circ \Gamma^{-1}$.
- a source containing meta-information on (R, T): $(\Omega', \sigma_{\Omega'}, \mu_{\text{meta}}, \Gamma_{\text{meta}})$ which induces a belief function on $\mathcal{S}$. The corresponding unnormalized mass function is denoted by $m_{\text{meta}} = \mu_{\text{meta}} \circ \Gamma_{\text{meta}}^{-1}$.

The **behavior based correction** (BBC) scheme consists in the following combination of these two sources so as to obtain the rectified source $(\Omega \times \Omega', \sigma_{\Omega\times\Omega'}, \mu \otimes \mu_{\text{meta}}, \Gamma_{\text{rec}})$ and $\forall \omega \in \Omega, \forall \omega' \in \Omega'$

$$\Gamma_{\text{rec}}(\omega, \omega') = \bigcup_{s \in \Gamma_{\text{meta}}(\omega')} \tilde{\Gamma}(\omega, s)\,. \tag{5.44}$$

So if m_{rec} denotes the rectified version of mass function m that is induced by the rectified source, we obtain for all $B \subseteq \Theta$

$$m_{\text{rec}}(B) = \mu \otimes \mu_{\text{meta}} \circ \Gamma_{\text{rec}}^{-1}(B)\,, \tag{5.45}$$

$$= \sum_{\substack{\omega\in\Omega, \omega'\in\Omega' \\ \Gamma_{\text{rec}}(\omega,\omega')=B}} \mu \otimes \mu_{\text{meta}}(\omega, \omega')\,, \tag{5.46}$$

$$= \sum_{\substack{\omega\in\Omega, \omega'\in\Omega' \\ \Gamma_{\text{rec}}(\omega,\omega')=B}} \mu(\omega) \times \mu_{\text{meta}}(\omega')\,, \tag{5.47}$$

$$= \sum_{\substack{A\subseteq\Theta, H\subseteq\mathcal{S} \\ \bigcup_{s\in H} \Gamma_A(s)=B}} m(A) \times m_{\text{meta}}(H)\,. \tag{5.48}$$

To better understand how (5.48) is obtained from (5.47): let H denote a subset of $\mathcal{S}$; let A denote a subset of Θ such that $\bigcup_{s\in H} \Gamma_A(s) = B$.

Some of the terms in the sum of (5.47) are the following

$$\sum_{\omega\in\Gamma^{-1}(A), \omega'\in\Gamma_{\text{meta}}^{-1}(H)} \mu(\omega) \times \mu_{\text{meta}}(\omega')$$

and the above sum is equal to $m(A) \times m_{\text{meta}}(H)$. So (5.47) can be rewritten as a sum of such partial sums for any H and A that are compliant with the constraint.

BBC is a generalization of the ballooning extension (2.19). Besides, when the meta-source induces a vacuous mass function on $\mathcal{S}$ and $\mathcal{S}$ is the meta-frame from example 27, we obtain

$$m_{\text{rec}}(B) = \sum_{\substack{A \subseteq \Theta, H \subseteq \mathcal{S} \\ \bigcup_{s \in H} \Gamma_A(s) = B}} m(A)\, m_{\text{meta}}(H), \tag{5.49}$$

$$= \sum_{\substack{A \subseteq \Theta \\ \bigcup_{s \in \mathcal{S}} \Gamma_A(s) = B}} m(A), \tag{5.50}$$

$$= \sum_{\substack{A \subseteq \Theta \\ \Theta = B}} m(A), \tag{5.51}$$

$$= \begin{cases} 1 & \text{if } B = \Theta \\ 0 & \text{otherwise} \end{cases}. \tag{5.52}$$

So when the meta-knowledge is not available, BBC follows an extremely cautious strategy and suggest to throw away all information.

Finally, if we want to combine two sources and their respective meta sources, we can just rectify each source individually and then combine the rectified sources using the conjunctive rule. The frame $\mathcal{S}$ can take more general forms than just (R, T) pairs, see Pichon et al. 2014, example 1 and 2.

5.5 Conclusions

By viewing combination rules as binary operations, we can examine what algebraic properties hold for the mass space, depending on the chosen rule it is endowed with. Like for other kinds of structures previously discussed in this monograph, our main conclusion is that the mass space cannot attain very rich structures. Even the group structure does not seem to be achievable for belief functions because combination operations are difficult to reverse, albeit not totally impossible. The absence of group structure is not a strong impediment to the theory of belief functions because there are more general or complementary operations that can help to recover a previous state of belief if necessary.

In the hierarchy of algebraic structures, the entire mass space can be granted (at most) the status of commutative monoid. Some subsets of it can be granted the status of semilattice, as already seen in chapter 3. We review the structure status of the mass space w.r.t. the rules presented in this chapter in table 5.4.

Rule	Algebraic structure	Space
α-junction	Commutative Monoid	Mass space
$\sqcap$ and $\sqcup$	Idempotent Commutative Magma	Mass space
$\owedge$	Semilattice	Set of non dogmatic mass functions
$\ovee$	Semilattice	Set of subnormal mass functions
Yager	Commutative Magma	mass space
Dubois and Prade	Commutative Magma	mass space
Lefervre	Commutative Magma	mass space
PCR	Commutative Magma	mass space
RCR	Commutative Magma	mass space
Mix	Commutative Magma	mass space
$\circledwedge$	Idempotent Commutative Magma	Mass space

Table 5.4: Algebraic structures induced by combination rules.

General Conclusion

In this general conclusion, we start with a bundle of the most important general comments we made on the theory of belief functions throughout this monograph. Finally, we examine several research directions that we plan to follow in short, middle and long term.

Take home messages on belief functions and associated structures

In the quest for a mathematical theory of uncertainty, several frameworks have been developed through the past centuries. All of them are somehow connected and share a vision of quantifiable chances of events through numbers living in the unit interval $[0;1]$. These models allow us to evaluate what outcome is more likely than the others with respect to an ill known variable of interest.

A major difficulty in deriving a unified and comprehensive theory of uncertainty is that uncertainty has several facets. We have evoked aleatory and epistemic uncertainties but there are other taxonomies. The diversity in the nature of uncertainty makes it difficult for a mathematical tool to meet the requirements induced by each uncertainty type.

One aspect of uncertainty, known as Knightian uncertainty, is that there are usually several levels of uncertainty especially in a statistical context where we want to infer the variable of interest from data. If a model allows us to compute a real valued estimate of the chances of an event (first level), this value itself may not be certain. We can also evaluate how likely this value is (second level) but again this evaluation is subject to uncertainty. This phenomenon is reminiscent of a fractal behavior. Obviously, the mathematical complexity grows with the uncertainty level one touches and we most often build first order models (with credible or confidence intervals) and sometimes second order ones. In this monograph, the presented frameworks address Knightian uncertainty and are attempts to provide intermediate solutions between the first and the second level. For the first level, probability theory is undisputed and has proved to provide excellent results in almost every field of science. The full second order consists in probabilities of probabilities but is very abstract and complex to use in practice. Consequently, it appears justified to look for trade-off frameworks with higher representation power than probabilities but less computation burden than second order probabilities.

This monograph offers insights into the theory of belief functions which is a simplified second order model that generalizes probability theory. The axioms in this theory allow to perform calculus in a computationally feasible way in many situations and applicative contexts. A belief function can be interpreted as probability bounds, but not all set of probability bounds can be represented by a belief function. Belief functions can also be interpreted as the probability that evidence implies that the variable of interest belongs to a given set. The theory of belief functions comprises mathematical tools allowing to build belief functions from data, update them when necessary, perform inference and draw conclusions in a principled way.

It is nonetheless, in general, more difficult to use belief functions than standard probabilities as some of these tools have subtle pre-requisites. In the probability framework, independence assumption are frequently assumed and even if it is not verified, the model remains consistent. For example the naive Bayes classifier makes such an assumption which is actually instrumental to prevent overfitting. In contrast, alleging illegitimate independence when using belief functions and for instance combine them using Dempster's rule can lead to misleading conclusions. Let alone these pitfalls, belief function calculus has most of the time a higher time and memory complexity than probabilistic calculus.

Since belief functions are rather general objects as compared to probability distributions or sets, a very rich structure for belief spaces cannot be achieved. By belief spaces, we mean the space where belief functions live and those isomorphic to this one. Three types of structures have been reviewed in this manuscript:

- order theoretic structure,
- metric structure,
- algebraic structure.

By combining all these structures, a belief space is (at most) a partially ordered metric commutative monoid. Some subspaces can be granted the status of semi-lattice and one of them is a (full) lattice.

A linear presentation of these structures is nearly impossible as they are quite intricate notions. Metrics can be used to derive pre-orders or combination rules. Some combination rules induce partial orders (for subclasses of belief functions). Partial orders and combination rules should be consistent with metrics. In conclusion, the global picture is rather complex. We hope that this monograph is nonetheless pleasant to read and does justice to the logic behind belief functions and the structures of their space.

Future research directions

In this section, we give research perspectives regarding open problems and challenging tasks in connection with the theory of belief functions. These perspectives will be addressed in the coming year (short term), in the next two years (middle term) or at a horizon of five years (long term)

Short term perspectives

Most of the personal contributions reported in this document are related to distances for belief functions. Since some expertise in this field has been acquired, we plan to pursue some efforts in this direction. In particular, an open problem evoked in the conclusions of chapter 4 is that no distance has been proved to be consistent with both a combination rule and an informative partial order. We will investigate this question in the next months. The consistency of evidential distance with other aspects (like uncertainty measures) is also a complementary question to be addressed.

In addition, the collaboration with Sebastien Destercke will be continued and his expertise in imprecise probabilities will allow us to investigate the possibility to carry over these concepts to this more general framework. To that end, we will need to identify candidate distances and informative partial orders. Obviously, the set inclusion relation between closed convex subsets of probability distributions induces an informative partial order for lower previsions. The partial order obtained by comparing plausibility functions can be generalized to a comparison between upper probabilities. Some dissimilarities for lower previsions are introduced in Abellán and Gómez 2006. One of these dissimilarities possess properties that are related to our definition of consistency between distances and partial orders in the belief function theory.

An important topic which is outside of the scope of this monograph (but no less important) is the application of the theory of belief functions in various fields of engineering sciences and econometrics. A field of application in which we have made a contribution is signal and image processing. In this field, the aim is to process functional data (i.e signals) and extract relevant information and patterns from them. In Klein and Colot 2012, we introduced an image contour detection method based on an evidential model of pixel values. A grayscale image is a 2D finite grid. Each grid point is called pixel and is assigned a quantized value. Most of the time this value is encoded on one byte, hence 256 pixels values are possible. We proposed a belief function model for pixel values. The corresponding belief function is consonant and the focal elements are nested supersets of the sensor value. We showed that Dempster's degree of conflict obtained when combining belief functions induced by a pixel neighborhood provides remarkable contour detection performances as compared to standard approaches.

More generally, we have the feeling that Dempster's idea of a random variable which cannot be precisely observed is well adapted to appraise the uncertainty in digital signals. A similar idea is also developed by Cooman and Zaffalon 2004 using lower previsions. A manuscript is under preparation in which we elaborate on this idea to derive pixel value upper probabilities. We customize the general digital signal modelization problem to the case of photon counting sensors and apply it image denoising. The preliminary results proves that the model produces competitive performances as compared to the famous Yaroslavsky filter.

Middle term perspectives

As middle term perspective, we plan to focus on another application field of belief functions or ill-known probabilities at large: machine learning. A first track to follow is the adaptation of the probabilistic classifier combinations that have been studied as part of Mahmoud Albardan's PhD. The bulk of this seminal work is an attempt to circumvent the classifier output dependencies by applying idempotent probability distribution aggregation using t-norms followed by a renormalization. The aggregated distributions are empirical performance conditional probabilities (actual class given predicted class). This approach allows to define a continuum of models ranging form independence to total dependence using only one hyperparameter. By using a model to build belief functions for the same variable, idempotent combination rules may be tried as well.

Another idea is to apply Pichon et al. 2014's behavior based correction to derive robust classifiers. The relevance of a classifier may be evaluated empirically at training time through cross-validation. If the output of the classifier is a belief function, we can use this information to compute rectified belief functions so that poorly relevant classifiers have less influence in the final class prediction. As for truthfulness, the goal may not be to estimate it from data but to obtain a classifier that can be updated at limited cost, i.e. we do not have to retrain the classifier to integrate this piece of meta-information.

Long term perspectives

As remarked in chapter 5, deriving a reversible combination rule is extremely challenging. A more realistic goal is to introduce a methodology that relies on several rules or other mechanisms and achieves evidence deletion and insertion. The behavior based correction mechanism is one such approach but it requires meta-information. When this meta-information is missing, a spare technique would be useful. Intuitively, the solution lies in a subtle mixture of conjunctive and disjunctive rules. For instance, the rule introduced Dubois and Prade 1992 is a partial answer. Indeed for any mass function whose focal core is A, its combination with m_{A^c} using this rule returns the vacuous mass function m_Θ, so we could conclude that any combination with a simple mass function can be "undone". As shown in the following example, this is not true because the rule is not associative.

Example 28. Let $\Theta = \{a, b, c\}$ denote a space. Suppose we have combined $m_1 = \frac{1}{2}m_{\{a\}} + \frac{1}{2}m_{\{b\}}$ and $m_2 = \frac{1}{2}m_{\{a\}} + \frac{1}{2}m_{\{a,b\}}$ using Dubois and Prade's rule. In this case, we obtain $m_{\text{dp}} = \frac{1}{2}m_{\{a\}} + \frac{1}{4}m_{\{b\}} + \frac{1}{4}m_{\{a,b\}}$. The question is: can we find a mass function m such that by combination under Dubois and Prade's rule with m_{dp} yields m_1?

First, as suggested above, we may want to try $m = m_{\{c\}}$ because in this case the combination of m and m_2 is the vacuous mass function. This does not work as m_{dp} combined with $m_{\{c\}}$ using Dubois and Prade's rule yields also the vacuous mass function, so we do not recover m_1.

Second, we remark that a focal element of m that does not contain a generates conflict with the mass assigned to $\{a\}$ in m_{dp} so some positive mass will be assigned to their disjunction which is not a focal element of m_1. Likewise, a focal element of m must contain b as well so it must be a superset of $\{a, b\}$. But since $\{a, b\}$ is a focal element of m_{dp}, a positive mass will be assigned to $\{a, b\}$ by the conjunctive component of Dubois and Prades rule, ergo m_1 cannot be recovered.

Although reversibility is not achieved in the above toy example, there are many situations in which Dubois and Prade's rule achieves decombination because its conjunctive component is compensated by the disjunctive one and conversely.

Example 29. Let $m_1 = m_A$ and $m_2 = m_B$ denote two categorical mass functions on Θ such that $A \cap B \neq \emptyset$. In this case, their combination by Dubois and Prade's rule is $m_{\text{dp}} = m_{A\cap B}$. The question is: can we find a mass function m such that by combination under Dubois and Prade's rule with m_{dp} yields m_1?

This time the answer is yes. Take $m = m_{A\setminus(A\cap B)}$. The combination of m_{dp} and m using Dubois and Prade's rule amounts to a the disjunctive rule which yields $m_1 = m_A$.

A first stepping stone would be to prove that such compensations are possible at least using a pair of rules.

Career review

In this very last section, I give some comments on my eight-year long experience as assistant professor in the university of Lille1. Since these comments are personal, I will not use the conventional "we" but the subjective "I" instead.

Teaching

Teaching is much more challenging than what general public pictures. From my experience, a key point in the success of a teaching unit is the ability of the lecturer to adapt the message to the audience. For instance, I teach machine learning related units to three different groups of students (students from the computer sciences department, students from the electrical engineering and automation department and students from Ecole Centrale de Lille who have a strong mathematical background). For each group, I had to find different justifications as to why this course unit was important in their future careers. I also had to design different practicals because they do not have the same expectations. Mathematicians want to see equations in actions. Computer scientists want to code and use libraries. People from engineering sciences want to apply it in their field of expertise.

Of course teaching is not just of question of pleasing students and meet their expectations. Sometimes, complex concepts and technicalities need to be detailed. Students can accept them if they are rapidly proved to be relevant and not just piled for the whole semester.

I believe I have improved my abilities to pass on knowledge but there is also room for further improvements. I had the opportunity to benefit from colleagues to exchange ideas and good practices. In Lille1 and in Centrale, colleagues are strongly committing themselves to improve teaching materials year after year. On the downside, the volume of hours we need to teach is, in my opinion, too big to really be efficient in this task. Another problem is that the profile of students is changing rapidly and we are not really aware of what they really know or not. Also, at the university, the audience is very heterogeneous and it is almost impossible to deliver teachings that satisfy the best students while not letting others drowning.

Research

I have achieved several goals that I had when I took my assistant professor position. I have pursued my interest in belief functions and have now the feeling to have a much deeper understanding of the theory than at the end of my PhD. Perhaps, the greatest aspect of my job is that I keep learning day after day and I now see that this is essential.

Another satisfaction is that I have enlarged my field of expertise to medical image processing and machine learning. For each of these topics I have supervised (or am supervising) a PhD. Ideally, I would like to take advantage of these skills by mixing them (use evidential models to derive classifiers, use machine learning for medical image segmentation and so on).

Sometimes, I have a few regrets not to have worked as postdoc for a couple of years because this allows to boost the publication list since postdoc people have acquired scientific writing abilities and do not have to spend time on teaching. Postdoc positions are also very instrumental to build an international network and increase the visibility of research activities. I have national collaborations but an objective for the next years is to start also international ones.

The pace of my publications has increased and I am confident that I can maintain it. In spite of the pressure of institutions, I have tried to focus on publication quality not quantity. I select journals whose review policy is trustworthy and there is limited overlap between my papers (except when a journal article is an extended version of a conference one).

Another point that I need to work on is research project management. In France, as in almost every country, there are national institutions (CNRS and ANR) that launch project campaigns every year. The corresponding project are endowed with substantial fundings and increase a researcher visibility. Perhaps, I had the feeling that I needed to accomplish more before applying to these calls for projects but I think the time has come for stepping up. My laboratory (CRIStAL) and my research team (SigMA) provide a nice research and motivating environment that is helpful to apply for such projects.

Team leading

Team leading is really not something I expected to do so early and it was not a motivation when I took my assistant professor position. Yet, after three years the teaching team leader got retired and the dean of the computer sciences, electrical engineering and automation faculty asked me to be his successor. I accepted and started to involve myself in the management of my university.

The first year was difficult because I took part in the organization revision of the electrical engineering and automation bachelor which lead to endless negotiations as to what teaching units we keep, modify or leave untouched. After two years, I got a better understanding of the university system and started to enjoy the responsibilities. I have the feeling that I succeeded to defend the interests of my team. I also learned that team leading is a very daily responsibility and I am often interrupted in the middle of a research effort because there are urgent matter to discuss. An essential point in team leading is being a good listener and pay attention to colleagues.

Appendices

A

a.1 Rationality requirement axioms

Savage 1954 defined seven axioms which he argues are necessary conditions to make rational decisions. For some preorder $\preceq$ endowing the set of acts $\mathbb{Y}$, the seven axioms are

(i) the preorder $\preceq$ is a total preorder.

(ii) **sure thing principle**: for any pair of acts (y, y') let yAy' denote the following compound act:

$$\left(yAy'\right)(a) = \begin{cases} y(a) & \text{if } a \in A \\ y'(a) & \text{if } a \in A^c \end{cases}. \tag{A.1}$$

If $y_1Ay' \preceq y_2Ay'$, then we say that $y_1 \preceq y_2$ given A w.r.t. y'. The sure thing principle states that if $y_1 \preceq y_2$ given A holds w.r.t. y' then it holds as well w.r.t. any other act.

(iii) If for any pair of acts (y, y'), $y \preceq y'$ given A, then A is said to be null. If A is not null and if (y, y') are constant acts, then for any act y_0 $y_0Ay \preceq y_0Ay' \Leftrightarrow y \preceq y'$.

(iv) Let y_1, y_1', y_2, y_2' denote four constant acts with the following strict preference relations: $y_1 \prec y_1'$ and $y_2 \prec y_2'$. For any $A, B \subseteq \Theta$,

$$y_1Ay_1' \preceq y_1By_1' \Leftrightarrow y_2Ay_2' \preceq y_2By_2'. \tag{A.2}$$

(v) There are two constant acts $y_\perp$ and $y_\top$ with a strict preference: $y_\perp \prec y_\top$.

(vi) For any y_1, y_2, y_3 in $\mathbb{Y}$ such that $y_1 \preceq y_2$, there is a finite partition $(B_i)_{i=1}^k$ of Θ such that

$$y_1B_iy_3 \preceq y_2 \text{ and } y_1 \preceq y_2B_iy_3, \forall i. \tag{A.3}$$

(vii) For two acts y_1 and y_2, if y_1 is preferred (resp. less preferred) to the constant act $y_2(a)$ for each $a \in A$, then $y_1 \succeq y_2$ (resp. $y_1 \preceq y_2$) given A.

a.2 Proofs of results on the consistency of distances with α-junctions

Proof. (of prop. 3) - Suppose m_1, m_2 and m_3 are three mass functions defined on Θ. $\mathbf{K}_1$, $\mathbf{K}_2$ and $\mathbf{K}_3$ are their respective evidential matrices with respect to a given α-junction denoted by $\odot^\alpha$. The 1-operator-norm has the sub-multiplicative property, i.e. for all matrices $\mathbf{A}$ and $\mathbf{B}$, one has:

$$\|\mathbf{AB}\|_{op1} \leq \|\mathbf{A}\|_{op1} . \|\mathbf{B}\|_{op1}$$

One can thus write:

$$\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_{op1} \quad \leq \quad \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} . \|\mathbf{K}_3\|_{op1},$$

Given that $\|\mathbf{K}_3\|_{op1} = 1$, we have:

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_{op1} &\leq \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1}, \\ \Leftrightarrow d_{op1}\left(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3\right) &\leq d_{op1}\left(m_1, m_2\right). \end{aligned}$$

By definition, this latter inequality means that distance d_{op1} is consistent with rule $\odot^\alpha$. □

Proof. (of prop. 4) - Suppose m_1, m_2 and m_3 are three mass functions defined on Θ. Suppose $\mathbf{K}_1$, $\mathbf{K}_2$ and $\mathbf{K}_3$ are their respective evidential matrices with respect to an α-junction denoted by $\odot^\alpha$. The L_1 norm of a matrix is the sum over all columns of the L_1 norms of its column vectors:

$$\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 = \sum_{A\subseteq\Theta} \left\|\mathbf{m}_{13|^\alpha A} - \mathbf{m}_{23|^\alpha A}\right\|_1, \tag{A.4}$$

with $m_{i3|^\alpha A} = m_i \odot^\alpha m_A \odot^\alpha m_3$. Besides, according to proposition 3, d_{op1} is consistent with $\odot^\alpha$. Inequality (4.15) thus applies, and after multiplying both sides of this inequality by the normalizing constant of the distance, we have that for any $A \subseteq \Theta$:

$$\left\|\left(\mathbf{K}_{1|^\alpha A} - \mathbf{K}_{2|^\alpha A}\right)\mathbf{K}_3\right\|_{op1} \leq \left\|\mathbf{K}_{1|^\alpha A} - \mathbf{K}_{2|^\alpha A}\right\|_{op1},$$

with $\mathbf{K}_{i|^\alpha A}$ the evidential matrix corresponding to $m_{i|^\alpha A}$ with respect to $\odot^\alpha$. Now applying corollary 1 on each side of the above inequality gives:

$$\left\|\mathbf{m}_{13|^\alpha A} - \mathbf{m}_{23|^\alpha A}\right\|_1 \leq \left\|\mathbf{m}_{1|^\alpha A} - \mathbf{m}_{2|^\alpha A}\right\|_1. \tag{A.5}$$

Let us now use this inequality in equation (A.4):

$$\begin{aligned}
\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 &\leq \sum_{A\subseteq\Theta} \left\|\mathbf{m}_{1|^\alpha A} - \mathbf{m}_{2|^\alpha A}\right\|_1, \\
\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 &\leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_1.
\end{aligned}$$

Finally, dividing both sides of the above inequality by the normalizing constant of the distance gives:

$$d_{mat,1}\left(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3\right) \leq d_{mat,1}\left(m_1, m_2\right).$$

By definition, this latter inequality means that distance $d_{mat,1}$ is consistent with rule $\odot^\alpha$. □

Proof. (of prop. 5) - Suppose m_1, m_2 and m_3 are three mass functions defined on Θ. $\mathbf{K}_1$, $\mathbf{K}_2$ and $\mathbf{K}_3$ are their respective evidential matrices with respect to an α-junction $\odot^\alpha$. The L_∞ norm of a matrix is the max of the L_∞ norms of its column vectors. Since a column vector of $\mathbf{K}_i$ writes as $\mathbf{m}_{i|^\alpha B}$ with $B \subseteq \Theta$, there exists a subset E such that :

$$\begin{aligned}
\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_\infty &= \left\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{m}_{3|^\alpha E}\right\|_\infty, \\
&= \left\|(\mathbf{K}_1 - \mathbf{K}_2)\sum_{Y\subseteq\Theta} m_{3|^\alpha E}(Y)\,\mathbf{m}_Y\right\|_\infty, \\
&\leq \sum_{Y\subseteq\Theta} m_{3|^\alpha E}(Y)\,\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{m}_Y\|_\infty.
\end{aligned}$$

Again, the L_∞ norm of $\mathbf{K}_1 - \mathbf{K}_2$ is the max of the L_∞ norms of its colums vectors : $\max_{Y\subseteq\Theta} \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{m}_Y\|_\infty = \|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty$. Each term in the previous inequation is maximized by $\|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty$ which gives:

$$\begin{aligned}
\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_\infty &\leq \sum_{Y\subseteq\Theta} m_{3|^\alpha X}(Y)\,\|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty, \\
&\leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty.
\end{aligned}$$

After normalization, the above inequation gives

$$d_{mat,\infty}\left(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3\right) \leq d_{mat,\infty}\left(m_1, m_2\right)$$

. □

a.3 Proofs of results on the consistency of distances with partial orders

Proof. (of prop. 6) - The second implication follows from the first one by contraposition, hence we will only show the first one. For this, take any triplet m_1, m_2, m_3 such that $m_1 \sqsubseteq_x m_2 \sqsubseteq_x m_3$. We then have

$$\begin{aligned} m_1 \sqsubseteq_x m_2 \sqsubseteq_x m_3 &\Rightarrow m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3 \\ &\Rightarrow \max\{d_{12}; d_{23}\} \leq d_{13} \end{aligned}$$

where $d_{ij} = d(m_i, m_j)$. The first implication following from $\sqsubseteq_x \Rightarrow \sqsubseteq_y$, and the second one from the $\sqsubseteq_y$-compatibility of d. □

Proof. (of prop. 7) - We will start with distance $d_{pl,k}$ with $k < \infty$, and will then proceed with the others:

- for distances $d_{pl,k}$: let pl_1, pl_2 and pl_3 denote three plausibility functions induced by m_1, m_2, m_3. Let us suppose that $m_1 \sqsubset_{pl} m_2 \sqsubset_{pl} m_3$. We can write $\left(d_{pl,k}(m_1, m_3)\right)^k$

$$\begin{aligned} &= \left(\|\mathbf{pl}_1 - \mathbf{pl}_3\|_k\right)^k, \\ &= \sum_{A \subseteq \Theta} |pl_1(A) - pl_3(A)|^k. \end{aligned}$$

  Since $m_1 \sqsubset_{pl} m_2 \sqsubset_{pl} m_3$, we know that for any $A \subseteq \Theta$:

$$|pl_1(A) - pl_3(A)| \geq \max\{pl_3(A) - pl_2(A); pl_2(A) - pl_1(A)\} \geq 0$$

  and that the inequality is strict for at least one subset. This gives :

$$\begin{aligned} \left(d_{pl,k}(m_1, m_3)\right)^k > \sum_{A \subseteq \Theta} \max \Big\{ &(pl_3(A) - pl_2(A))^k, \\ &(pl_2(A) - pl_1(A))^k \Big\} \end{aligned}$$

  As the sum of maxima is always higher than the maximum of the sums, this gives

$$\begin{aligned} \left(d_{pl,k}(m_1, m_3)\right)^k &> \max \Big\{ \sum_{A \subseteq \Theta} (pl_3(A) - pl_2(A))^k, \\ &\qquad \sum_{A \subseteq \Theta} (pl_2(A) - pl_1(A))^k \Big\} \\ &> \max \Big\{ \left(\|\mathbf{pl}_1 - \mathbf{pl}_2\|_k\right)^k, \\ &\qquad \left(\|\mathbf{pl}_2 - \mathbf{pl}_3\|_k\right)^k \Big\}. \end{aligned}$$

  The last inequality is equivalent to :

$$d_{pl,k}(m_1, m_3) > \max\Big\{d_{pl,k}(m_1, m_2); d_{pl,k}(m_2, m_3)\Big\}. \quad (A.6)$$

- for distances $d_{b,k}$: given equation (A.6) and $d_{b,k} = d_{pl,k}$ (Lemma 1 in Klein et al. 2016c), the proof is immediate.
- for distances $d_{bel,k}$: the proof is actually similar to the one for $d_{pl,k}$. Let bel_1, bel_2 and bel_3 denote three belief functions induced by m_1, m_2, m_3 and suppose that $m_1 \sqsubset_{bel} m_2 \sqsubset_{bel} m_3$. We then have:

$$\begin{aligned} d_{bel,k}(m_1, m_3)^k &= \left(\|\mathbf{bel}_1 - \mathbf{bel}_3\|_k\right)^k, \\ &= \sum_{A \subseteq \Theta} |bel_1(A) - bel_3(A)|^k. \end{aligned}$$

Since $m_1 \sqsubset_{bel} m_2 \sqsubset_{bel} m_3$, we know that for any $A \subseteq \Theta$:

$$|bel_1(A) - bel_3(A)| \geq \max\{bel_2(A) - bel_3(A); bel_1(A) - bel_2(A)\} \geq 0$$

and that the inequality is strict for at least one subset. The proof then follows by an analogous reasoning to the one used for plausibilities.

- for distances $d_{q,k}$, the proof follows the same pattern. Simply consider q_1, q_2, q_3 induced by m_1, m_2, m_3 such that $m_1 \sqsubset_q m_2 \sqsubset_q m_3$, then

$$\begin{aligned} d_{q,k}(m_1, m_3)^k &= (\|\mathbf{q}_1 - \mathbf{q}_3\|_k)^k, \\ &= \sum_{A \subseteq \Theta} |q_1(A) - q_3(A)|^k. \end{aligned}$$

and the proof follows similarly to the previous cases.

- for pseudo-distances $d_{\pi,k}$, the proof follows again the same pattern (with a sum over $a \in \Theta$).

The proof for the $k = \infty$ case is omitted. It is given in Klein et al. 2016c, Appendix B. □

a.4 Proofs of results on distance based combination rules

Proof. (of lemma 5) - We give a proof for $\sqcap_q$. The one for $\sqcap_{pl}$ follows a similar scheme.

Let us denote by m^* the mass function yielded by $m_1 \sqcap_q m_2$. For any mass function $m \in \mathcal{M}_q(m_1) \cap .. \cap \mathcal{M}_q(m_\ell)$, we thus have

$$\begin{aligned} \|q - q_\Theta\|_k &\geq \|q^* - q_\Theta\|_k, \\ \Leftrightarrow \|1 - q\|_k &\geq \|1 - q^*\|_k. \end{aligned}$$

The above inequality comes from the fact that commonalities for the vacuous mass function are constant with value one. Observing that there is a symmetry relating function $g(\mathbf{x}) = \|\mathbf{1} - \mathbf{x}\|$ with function $h(\mathbf{x}) = \|\mathbf{x}\|$ for any vector $\mathbf{x}$ in the unit hypercube, we deduce

$$\begin{aligned} \|q\|_k &\leq \|q^*\|_k, \\ \Leftrightarrow \|q - q_\emptyset\|_k &\leq \|q^* - q_\emptyset\|_k. \end{aligned}$$

The above inequality is obtained by remembering that $q_\emptyset$ has null value for all non-empty set. It has value one for $\emptyset$ but this is tantamount to add the same constant term to both sides of the inequality. □

Proof. (of prop. 10) - The commonality function corresponding to the categorical mass function m_A is given by

$$q_A(B) = \begin{cases} 1 \text{ if } B \subseteq A \\ 0 \text{ otherwise} \end{cases}. \tag{A.7}$$

From this, one obviously has $q_0(B)\, q_A(B) = q_0(B) \wedge q_A(B)$. Remembering that entrywise product of two commonality functions is the commonality function of their conjunctive combination, we have

$$\mathcal{M}_q(m_0) \cap \mathcal{M}_q(m_A) = \mathcal{M}_q\left(m_{0|A}\right).$$

By definition of $\mathcal{M}_q\left(m_{0|A}\right)$, its unique maximal element is $m_{0|A}$, meaning that $\forall m \in \mathcal{M}_q\left(m_{0|A}\right)$, one has $m \sqsubseteq_q m_{0|A}$. Now since we also have that $m_{0|A} \sqsubseteq_q m_\Theta$ and $d_{q,k}$ is $\sqsubseteq_q$-compatible, then

$$\underset{m \in \mathcal{M}_q(m_0) \cap \mathcal{M}_q(m_A)}{\arg\min} d_{q,k}(m, m_\Theta) = m_{0|A}. \tag{A.8}$$

□

Proof. (of prop. 11) - The vacuous mass function m_Θ is the maximum of $\left(\mathcal{M}, \sqsubseteq_f\right)$ for $f \in \{d, s, q, pl\}$ which implies that $\mathcal{M}_f\left(m_\Theta\right) = \mathcal{M}$. Consequently, the feasible set of $m \sqcap_{f,k} m_\Theta$ is $\mathcal{M}_f\left(m\right)$. By defintion of $\mathcal{M}_f\left(m\right)$, we have $m \sqcap_{f,k} m_\Theta \sqsubseteq_f m$. Since distance d is consistent with $\sqsubseteq_f$, $m \sqcap_{f,k} m_\Theta \sqsubseteq_f m \sqsubseteq_f m_\Theta$ implies $d\left(m \sqcap_{f,k} m_\Theta, m_\Theta\right) \leq d\left(m, m_\Theta\right)$. But $m \sqcap_{f,k} m_\Theta$ has by definition minimal distance to m_Θ therefore we also have $d\left(m \sqcap_{f,k} m_\Theta, m_\Theta\right) \geq d\left(m, m_\Theta\right)$, hence $m \sqcap_{f,k} m_\Theta = m$.

Furthermore, suppose $m_{\mathrm{e}} \neq m_\Theta$ is a neutral element. Since m_Θ is neutral, $m_{\mathrm{e}} \sqcap_{f,k} m_\Theta = m_{\mathrm{e}}$ but since m_{e} is neutral as well then $m_{\mathrm{e}} \sqcap_{f,k} m_\Theta = m_\Theta$ hence a contradiction. $\square$

Bibliography

Abellán, Joaquín, and Manuel Gómez. 2006. "Measures of divergence on credal sets." *Fuzzy Sets and Systems* 157 (11): 1514 –1531. doi:http://dx.doi.org/10.1016/j.fss.2005.11.021. (Cited on page 114).

Autchariyapanitkul, Kittawit, Somsak Chanaim, Songsak Sriboonchitta, and Thierry Denoeux. 2014. "Predicting Stock Returns in the Capital Asset Pricing Model Using Quantile Regression and Belief Functions" [in English]. In *Belief Functions: Theory and Applications,* edited by Fabio Cuzzolin, 8764:219–226. Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-319-11191-9_24. (Cited on page 12).

Burger, Thomas. 2014. "Geometric Interpretations of Conflict: A Viewpoint." In *Belief Functions: Theory and Applications: Third International Conference, BELIEF 2014, Oxford, UK, September 26-28, 2014. Proceedings,* edited by Fabio Cuzzolin, 412–421. Cham: Springer International Publishing. doi:10.1007/978-3-319-11191-9_45. (Cited on page 80).

Burger, Thomas, Sebastien Destercke, et al. 2013. "How to randomly generate mass functions." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems:* 645–673. (Cited on page 82).

Cattaneo, Marco E. G. V. 2011. "Belief functions combination without the assumption of independence of the information sources." Dependence Issues in Knowledge-Based Systems, *International Journal of Approximate Reasoning* 52 (3): 299 –315. doi:http://dx.doi.org/10.1016/j.ijar.2010.10.006. (Cited on pages 102, 107).

———. 2003. "Combining belief functions issued from dependent sources." In *Third Int. Symposium on Imprecise Probabilities and Their Applications,* edited by J.M. Bernad, T. Seidenfeld, and M. Zaffalon (Eds.), 133–147. Lugano (Switzerland): Carleton Scientific. (Cited on pages 102, 106).

Clemens, Joachim, Thomas Reineking, and Tobias Kluth. 2016. "An evidential approach to SLAM, path planning, and active exploration." *International Journal of Approximate Reasoning* 73:1 –26. doi:http://dx.doi.org/10.1016/j.ijar.2016.02.003. (Cited on page 12).

Cooman, Gert de, and Marco Zaffalon. 2004. "Updating beliefs with incomplete observations." *Artificial Intelligence* 159 (1): 75 –125. doi:https://doi.org/10.1016/j.artint.2004.05.006. (Cited on page 114).

Couso, Inés, and Didier Dubois. 2014. "Statistical reasoning with set-valued information: Ontic vs. epistemic views." Special issue: Harnessing the information contained in low-quality data sources, *International Journal of Approximate Reasoning* 55 (7): 1502 –1518. doi:http://dx.doi.org/10.1016/j.ijar.2013.07.002. (Cited on page 33).

Cuzzolin, Fabio. 2008. "A Geometric Approach to the Theory of Evidence." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 38 (4): 522–534. doi:10.1109/TSMCC.2008.919174. (Cited on pages 55, 75).

———. 2010a. "Geometric conditioning of belief functions." In *Proceedings of BELIEF 2010, International workshop on the theory of belief functions,* 1–6. Brest, France. (Cited on pages 75, 86).

———. 2004. "Geometry of Dempster's rule of combination." *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 34 (2): 961 –977. doi:10.1109/TSMCB.2003.818431. (Cited on pages 56, 75, 99).

Cuzzolin, Fabio. 2011. "On Consistent Approximations of Belief Functions in the Mass Space." In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty,* edited by Weiru Liu, 6717:287–298. Lecture Notes in Computer Science. Springer Berlin Heidelberg. (Cited on page 76).

———. 2010b. "The geometry of consonant belief functions: Simplicial complexes of necessity measures." Theme: Aggregation functions, *Fuzzy Sets and Systems* 161 (10): 1459 –1479. doi:http://dx.doi.org/10.1016/j.fss.2009.09.024. (Cited on page 75).

Daniel, Milan. 2004. "Algebraic structures related to the combination of belief functions." *Scientiae Mathematicae Japonicae* 60 (2): 245–256. (Cited on page 88).

Delmotte, François, Laurent Dubois, and Pierre Borne. 1996. "Context-dependent trust in data fusion within the possibility theory." In *IEEE Int.Conf. on Systems, Man, and Cybernetics (ICSMC'96),* 538–543. Beijing (China). (Cited on page 95).

Delmotte, François, Laurent Dubois, Anne-Marie Desodt, and Pierre Borne. 1995. "Using trust in uncertainty theories." *Information and Systems Engineering* 1:303–314. (Cited on page 94).

Démotier, Sabrina, Walter Schon, and Thierry Denœux. 2006. "Risk assessment based on weak information using belief functions: a case study in water treatment." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36 (3): 382–396. (Cited on page 12).

Dempster, Arthur P. 1966. "New Methods for Reasoning Towards Posterior Distributions Based on Sample Data." *Annals of Mathematical Satistics* 37:355–374. (Cited on page 36).

———. 2008. "The Dempster–Shafer calculus for statisticians." In Memory of Philippe Smets (1938–2005), *International Journal of Approximate Reasoning* 48 (2): 365 –377. doi:http://dx.doi.org/10.1016/j.ijar.2007.03.004. (Cited on page 32).

———. 1967. "Upper and lower probabilities induced by a multiple valued mapping." *Annals of Mathematical Satistics* 38 (2): 325–339. (Cited on pages 3, 5, 27, 36, 37, 92).

Denœux, Thierry. 1995. "A k-nearest neighbour classification rule based on Dempster-Shafer theory." *IEEE trans. on Systems Man and Cybernetics* 25 (5): 804–813. (Cited on page 12).

———. 2008. "Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence." *Artificial Intelligence* 172 (2-3): 234–264. (Cited on pages 54, 62, 71, 72, 92, 97, 98, 101, 102, 104).

———. 2001. "Inner and outer approximation of belief structures using a hierarchical clustering approach." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 9 (04): 437–460. (Cited on page 76).

———. 2014. "Likelihood-based belief function: Justification and some extensions to low-quality data." Special issue: Harnessing the information contained in low-quality data sources, *International Journal of Approximate Reasoning* 55 (7): 1535 –1547. doi:http://dx.doi.org/10.1016/j.ijar.2013.06.007. (Cited on pages 34, 58).

Denœux, Thierry, Orakanya Kanjanatarakul, and Songsak Sriboonchitta. 2015. "EK-NNclus: a clustering procedure based on the evidential K-nearest neighbor rule." *Knowledge-Based Systems* 88:57–69. (Cited on page 12).

Destercke, Sébastien, and Thomas Burger. 2013. "Toward an axiomatic definition of conflict between belief functions." *IEEE transactions on cybernetics* 43 (2): 585–596. (Cited on pages 64, 80, 99).

Destercke, Sebastien, and Didier Dubois. 2011. "Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory." *Information Sciences* 181 (18): 3925 –3945. doi:http://dx.doi.org/10.1016/j.ins.2011.05.007. (Cited on pages 102, 106, 107).

Dhaou, Salma Ben, Kuang Zhou, Mouloud Kharoune, Arnaud Martin, and Boutheina Ben Yaghlane. 2017. "The advantage of evidential attributes in social networks." In *2017 20th International Conference on Information Fusion (Fusion),* 1–8. doi:10.23919/ICIF.2017.8009758. (Cited on page 12).

Dubois, Didier, and Thierry Denœux. 2010. "Statistical inference with belief functions and possibility measures: a discussion of basic assumptions." *Combining Soft Computing and Statistical Methods in Data Analysis:* 217–225. (Cited on page 58).

Dubois, Didier, and Henri Prade. 1985. "A note on measures of specificity for fuzzy sets." *International Journal of General Systems* 10:279–283. (Cited on page 64).

Dubois, Didier, and Henri Prade. 1986. "A set-theoretic view of belief functions: logical operations and approximatons by fuzzy sets." *Int. Journal of General Systems* 12 (3): 193–226. (Cited on pages 62, 64).

———. 1988. *Possibility theory: An approach to the computerized processing of information.* Plenum Press, New York. (Cited on page 39).

Dubois, Didier, and Henri Prade. 1992. "Reliability Data Collection and Analysis." Chap. On the combination of evidence in various mathematical frameworks, edited by J; Flamm and Brussels T. Luisi, 213–241. (Cited on pages 93, 115).

Dubois, Didier, and Ronald R. Yager. 1992. "Fuzzy set connectives as combinations of belief structures." *Information Sciences* 66 (3): 245 –276. doi:http://dx.doi.org/10.1016/0020-0255(92)90096-Q. (Cited on page 107).

Dubois, Didier, Henri Prade, and Philippe Smets. 2008. "A definition of subjective possibility." In Memory of Philippe Smets (1938–2005), *International Journal of Approximate Reasoning* 48 (2): 352 –364. doi:http://dx.doi.org/10.1016/j.ijar.2007.01.005. (Cited on page 97).

———. 2001. "New Semantics for Quantitative Possibility Theory" [in English]. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty,* edited by Salem Benferhat and Philippe Besnard, 2143:410–421. Lecture Notes in Computer Science. Springer Berlin Heidelberg. doi:10.1007/3-540-44652-4_36. (Cited on page 68).

———. 1996. "Representing partial ignorance." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 26 (3): 361–377. doi:10.1109/3468.487961. (Cited on pages 32, 37, 39).

Dubois, Didier, Weiru Liu, Jianbing Ma, and Henri Prade. 2016. "The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks." *Information Fusion* 32:12–39. (Cited on pages 93, 95, 102).

El Zoghby, Nicole, Véronique Cherfaoui, and Thierry Denœux. 2014. "Evidential distributed dynamic map for cooperative perception in VANets." In *2014 IEEE Intelligent Vehicles Symposium Proceedings,* 1421–1426. doi:10.1109/IVS.2014.6856550. (Cited on page 12).

Fisher, Ronald A. 1930. "Inverse probability." In *Mathematical Proceedings of the Cambridge Philosophical Society,* 26:528–535. 4. Cambridge University Press. (Cited on page 35).

Fixsen, Dale, and Ronald P.S. Mahler. 1997. "The modified Dempster-Shafer approach to classification." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 27 (1): 96–104. doi:10.1109/3468.553228. (Cited on page 86).

Florea, Mihai C., Anne-Laure Jousselme, Eloi Boissé, and Dominic Grenier. 2009. "Robust combination rules for evidence theory." *Information Fusion* 10:183–197. (Cited on page 95).

Fréchet, Maurice. 1906. "Sur quelques points du calcul fonctionnel." PhD diss., Faculté des Sciences de Paris. (Cited on page 75).

Gilboa, Itzhak. 1987. "Expected utility with purely subjective non-additive probabilities." *Journal of Mathematical Economics* 16 (1): 65 –88. doi:http://dx.doi.org/10.1016/0304-4068(87)90022-X. (Cited on page 60).

Gong, Wenjuan, and Fabio Cuzzolin. 2017. "A belief-theoretical approach to example-based pose estimation." *IEEE Transactions on Fuzzy Systems* PP (99): 1–1. doi:10.1109/TFUZZ.2017.2686803. (Cited on page 12).

Hájek, Petr, and Julio V Valdes. 1991. "A generalized algebraic approach to uncertainty processing in rule-based expert systems (dempsteroids)." *Computers and Artificial Intelligence* 10 (1): 29–42. (Cited on page 88).

Halpern, Joseph Y, and Ronald Fagin. 1992. "Two views of belief: belief as generalized probability and belief as evidence." *Artificial intelligence* 54 (3): 275–317. (Cited on page 37).

Han, Deqiang, Jean Dezert, Chongzhao Han, and Yi Yang. 2011. "New dissimilarity measures in evidence theory." In *14th International Conference on Information Fusion,* 1–7. (Cited on page 80).

Harmanec, David, George J. Klir, and Zhenyuan Wang. 1996. "Modal logic interpretation of Dempster-Shafer theory: An infinite case." *International Journal of Approximate Reasoning* 14 (2): 81 –93. doi:http://dx.doi.org/10.1016/0888-613X(95)00111-S. (Cited on page 73).

Jirousek, Radim, and Prakash P. Shenoy. 2018. "A new definition of entropy of belief functions in the Dempster-Shafer theory." *International Journal of Approximate Reasoning* 92 (Supplement C): 49 –65. doi:https://doi.org/10.1016/j.ijar.2017.10.010. (Cited on page 86).

Jousselme, Anne-Laure, and Patrick Maupin. 2012. "Distances in evidence theory: Comprehensive survey and generalizations." Theory of Belief Functions (BELIEF 2010), *International Journal of Approximate Reasoning* 53 (2): 118 –145. doi:http://dx.doi.org/10.1016/j.ijar.2011.07.006. (Cited on page 77).

Jousselme, Anne-Laure, Dominic Grenier, and Eloi Bossé. 2001. "A new distance between two bodies of evidence." *Information Fusion* 2:91–101. (Cited on pages 66, 77).

Kallel, Abdelaziz, and Sylvie Le Hégarat-Mascle. 2009. "Combination of partially non-distinct beliefs: the cautious-adaptative rule." *International Journal of Approximate Reasoning* 50:1000–1021. (Cited on page 49).

Kanjanatarakul, Orakanya, Songsak Sriboonchitta, and Thierry Denœux. 2014. "Forecasting using belief functions: An application to marketing econometrics." *International Journal of Approximate Reasoning* 55 (5): 1113 –1128. doi:http://dx.doi.org/10.1016/j.ijar.2014.01.005. (Cited on page 34).

Kanjanatarakul, Orakanya, Thierry Denœux, and Songsak Sriboonchitta. 2016. "Prediction of future observations using belief functions: A likelihood-based approach." {BELIEF} 2014 – Third International Conference on Belief Functions, *International Journal of Approximate Reasoning* 72:71 –94. doi:http://dx.doi.org/10.1016/j.ijar.2015.12.004. (Cited on page 38).

Ke, Xiaolu, Liyao Ma, and Yong Wang. 2013. "A dissimilarity measure based on singular value and its application in incremental discounting." In *Proceedings of the 16th International Conference on Information Fusion,* 1391–1397. (Cited on page 80).

Kendall, David G. 1974. "Foundations of a theory of random sets." *Stochastic geometry* 3 (9). (Cited on page 32).

Kennes, Robert. 1991. "Evidential reasoning in a categorial perspective: conjunction and disjunction of belief functions." In *Proceedings of the Seventh conference on Uncertainty in Artificial Intelligence,* 174–181. Morgan Kaufmann Publishers Inc. (Cited on page 70).

Keynes, John Maynard. 1921. "Chapter iv: The principle of indifference." *A treatise on probability* 4:41–64. (Cited on page 26).

Klawonn, F., and E. Schwecke. 1992. "On the axiomatic justification of Dempster's rule of combination." *International Journal of Intelligent Systems* 7 (5): 469–478. doi:10.1002/int.4550070506. (Cited on page 34).

Klawonn, Frank, and Philippe Smets. 1992. "The dynamic of belief in the transferable belief model and specialization-generalization matrices." In *Proceedings of the eighth conference on Uncertainty in Artificial Intelligence,* 130–137. UAI '92. Stanford, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. (Cited on page 34).

Klein, John, and Olivier Colot. 2012. "A belief function model for pixel data." *Belief Functions: Theory and Applications:* 189–196. (Cited on page 114).

———. 2010. "Automatic discounting rate computation using a dissent criterion." In *Workshop on the theory of belief functions,* 1–6. Brest, France. (Cited on page 49).

———. 2011. "Singular sources mining using evidential conflict analysis." *International Journal of Approximate Reasoning* 52 (0): 1433–1451. doi:10.1016/j.ijar.2011.08.005. (Cited on pages 66, 68).

Klein, John, Mehena Loudahi, Jean-Marc Vannobel, and Olivier Colot. 2014. "α-Junctions of Categorical Mass Functions" [in English]. In *Belief Functions: Theory and Applications,* edited by Fabio Cuzzolin, 8764:1–10. Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-319-11191-9_1. (Cited on page 52).

Klein, John, Chritèle Lecomte, and Pierre Miché. 2009. "Hierarchical and conditional combination of belief functions induced by visual tracking." *International Journal of Approximate Reasoning* 51:410–428. (Cited on pages 12, 92).

Klein, John, Sebastien Destercke, and Olivier Colot. 2016a. "Idempotent Conjunctive and Disjunctive Combination of Belief Functions by Distance Minimization." Edited by Jiřina Vejnarová and Václav Kratochvíl, *International Journal of Approximate Reasoning* (Cham) submitted to:XX–XX. (Cited on pages 96, 102).

———. 2016b. "Idempotent Conjunctive Combination of Belief Functions by Distance Minimization." In *Belief Functions: Theory and Applications: 4th International Conference, BELIEF 2016, Prague, Czech Republic, September 21-23, 2016, Proceedings,* edited by Jiřina Vejnarová and Václav Kratochvíl, 156–163. Cham: Springer International Publishing. doi:10.1007/978-3-319-45559-4_16. (Cited on pages 76, 86, 96, 97).

———. 2016c. "Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation." *International Journal of Approximate Reasoning* 71:15 –33. doi:http://dx.doi.org/10.1016/j.ijar.2016.01.001. (Cited on pages 84, 86, 97, 100, 123, 124).

Kohlas, Jürg, and Paul-André Monney. 1995. *A mathematical theory of hints: An approach to the Dempster-Shafer theory of evidence.* Vol. 425. Lecture Notes in Economics and Mathematical Systems. Springer Science & Business Media. (Cited on page 36).

———. 2008. "An algebraic theory for statistical information based on the theory of hints." In Memory of Philippe Smets (1938-2005), *International Journal of Approximate Reasoning* 48 (2): 378 –398. doi:https://doi.org/10.1016/j.ijar.2007.05.003. (Cited on page 36).

Labourey, Quentin, Denis Pellerin, Michele Rombaut, Olivier Aycard, and Catherine Garbay. 2015. "Sound classification in indoor environment thanks to belief functions." In *Signal Processing Conference (EUSIPCO), 2015 23rd European,* 2286–2290. IEEE. (Cited on page 12).

Lefevre, Eric, Olivier Colot, and Patrick Vannoorenberghe. 2002. "Belief function combination and conflict management." *Information Fusion* 3:149–162. (Cited on page 93).

Lian, Chunfeng, Su Ruan, Thierry Denœux, Hua Li, and Pierre Vera. 2017. "Spatial Evidential Clustering with Adaptive Distance Metric for Tumor Segmentation in FDG-PET Images." *IEEE Transactions on Biomedical Engineering* PP (99): 1–1. doi:10.1109/TBME.2017.2688453. (Cited on page 12).

Liu, Weiru. 2006. "Analyzing the degree of conflict among belief functions." *Artificial Intelligence* 170 (11): 909 – 924. doi:http://dx.doi.org/10.1016/j.artint.2006.05.002. (Cited on page 80).

Liu, Zhunga, Jean Dezert, Quan Pan, and Grégoire Mercier. 2011. "Combination of sources of evidence with different discounting factors based on a new dissimilarity measure." *Decision Support Systems* 52 (1): 133 –141. doi:http://dx.doi.org/10.1016/j.dss.2011.06.002. (Cited on page 80).

Loudahi, Mehena, John Klein, Jean-Marc Vannobel, and Olivier Colot. 2016. "Evidential Matrix Metrics as Distances Between Meta-Data Dependent Bodies of Evidence." *Cybernetics, IEEE Transactions on* 46 (1): 109–122. doi:10.1109/TCYB.2015.2395877. (Cited on pages 77, 78, 81).

———. 2014a. "Fast Computation of L_p Norm-Based Specialization Distances between Bodies of Evidence" [in English]. In *Belief Functions: Theory and Applications,* edited by Fabio Cuzzolin, 8764:422–431. Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-319-11191-9_46. (Cited on page 80).

———. 2014b. "New distances between bodies of evidence based on Dempsterian specialization matrices and their consistency with the conjunctive combination rule." *International Journal of Approximate Reasoning* 55 (5): 1093 –1112. doi:http://dx.doi.org/10.1016/j.ijar.2014.02.007. (Cited on pages 78, 81).

Martin, Arnaud, and Christophe Osswald. 2007. "Toward a combination rule to deal with partial conflict and specificity in belief functions theory." In *Int. Conf. on Information Fusion (FUSION'07),* 1–8. Quebec (Canada). (Cited on pages 94, 95).

Martin, Arnaud, Anne-Laure Jouselme, and Christophe Osswald. 2008. "Conflict measure for the discounting operation on belief functions." In *Proceedings of the Eleventh International Conference on Information Fusion,* 1–8. (Cited on pages 66, 86).

Martin, Ryan, Jianchun Zhang, and Chuanhai Liu. 2010. "Dempster–Shafer Theory and Statistical Inference with Weak Beliefs." *Statist. Sci.* 25, no. 1 (February): 72–87. doi:10.1214/10-STS322. (Cited on pages 12, 34, 36, 38).

Mathéron, Georges. 1975. *Random sets and integral geometry.* Wiley New York. (Cited on page 32).

Mercier, David, Eric Lefevre, and François Delmotte. 2010. "Belief Functions Contextual Discounting and Canonical Decompositions." *International Journal of Approximate Reasoning:* in press. (Cited on page 49).

Mercier, David, Benjmain Quost, and Thierry Denœux. 2008. "Refined modeling of sensor reliability in the belief function framework using contextual discounting." *Information Fusion* 9:246–258. (Cited on page 49).

Mo, Hongming, Xi Lu, and Yong Deng. 2016. "A generalized evidence distance." *Journal of Systems Engineering and Electronics* 27 (2): 470–476. doi:10.1109/JSEE.2016.00049. (Cited on page 80).

Nguyen, Hung T. 1978. "On random sets and belief functions." *Journal of Mathematical Analysis and Applications* 65 (3): 531 –542. doi:http://dx.doi.org/10.1016/0022-247X(78)90161-0. (Cited on pages 31, 32).

Pearl, Judea. 1990. "Reasoning with belief functions: An analysis of compatibility." *International Journal of Approximate Reasoning* 4 (5): 363 –389. doi:http://dx.doi.org/10.1016/0888-613X(90)90013-R. (Cited on pages 36, 40).

———. 1992. "Rejoinder to comments on "reasoning with belief functions: An analysis of compatibility"." *International Journal of Approximate Reasoning* 6 (3): 425 –443. doi:http://dx.doi.org/10.1016/0888-613X(92)90034-W. (Cited on page 37).

Pichon, Frédéric. 2012. "On the α-Conjunctions for Combining Belief Functions." In *Belief Functions: Theory and Applications,* edited by Thierry Denoeux and Marie-Hélène Masson, 164:285–292. Advances in Intelligent and Soft Computing. Springer Berlin Heidelberg. doi:10.1007/978-3-642-29461-7_33. (Cited on page 53).

Pichon, Frédéric, and Thierry Denoeux. 2009. "Interpretation and Computation of α-Junctions for Combining Belief Functions." In *6th Int. Symposium on Imprecise Probability: Theories and Applications (ISIPTA '09).* Durham, U.K. (Cited on page 52).

Pichon, Frédéric, Didier Dubois, and Thierry Denœux. 2012. "Relevance and truthfulness in information correction and fusion." Theory of Belief Functions (BELIEF 2010), *International Journal of Approximate Reasoning* 53 (2): 159 –175. doi:http://dx.doi.org/10.1016/j.ijar.2011.02.006. (Cited on pages 54, 89).

Pichon, Frédéric, David Mercier, François Delmotte, and Éric Lefèvre. 2014. "Truthfulness in Contextual Information Correction" [in English]. In *Belief Functions: Theory and Applications,* edited by Fabio Cuzzolin, 8764:11–20. Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-319-11191-9_2. (Cited on pages 108, 110, 115).

Pichon, Frédéric, and Anne-Laure Jousselme. 2016. "A norm-based view on conflict." In *Proceedings of the Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2016),* 153–159. La Rochelle, France: Cépaduès. (Cited on page 80).

Savage, Leonard J. 1954. "The foundations of statistics." *NY, John Wiley:* 188–190. (Cited on page 121).

Schubert, Johan. 2010. "Conflict management in Dempster–Shafer theory using the degree of falsity." *International Journal of Approximate Reasoning* 52:449–460. (Cited on pages 49, 65).

———. 2016. "Entropy-Based Counter-Deception in Information Fusion." In *Belief Functions: Theory and Applications: 4th International Conference, BELIEF 2016, Prague, Czech Republic, September 21-23, 2016, Proceedings,* edited by Jiřina Vejnarová and Václav Kratochvíl, 174–181. Cham: Springer International Publishing. doi:10.1007/978-3-319-45559-4_18. (Cited on page 68).

———. 1996. "Specifying Nonspecific Evidence." *International Journal of Intelligent Systems* 11:525–563. (Cited on page 65).

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence.* Princeton (NJ), USA: Princeton University press. (Cited on pages 3, 30, 47, 49, 58, 67, 80).

———. 1973. "Allocation of probability: a theory of partial belief." PhD diss., Princeton University, department of statistics. (Cited on page 31).

———. 1981. "Constructive probability." *Synthese* 48 (1): 1–60. (Cited on page 33).

———. 1992. "Rejoinders to comments on "perspectives on the theory and practice of belief functions"." *International Journal of Approximate Reasoning* 6 (3): 445 –480. doi:http://dx.doi.org/10.1016/0888-613X(92)90035-X. (Cited on page 37).

Shenoy, Prakash P, and Glenn Shafer. 2008. "Axioms for Probability and Belief-Function Propagation." *Classic Works of the Dempster-Shafer Theory of Belief Functions* 219:499–528. (Cited on page 39).

Smarandache, Florentin, and Jean Dezert. 2005. "Information fusion based on new proportional conflict redistribution rules." In *Int. Conf. on Information Fusion (FUSION'05),* 1–8. Philadelphia (USA). (Cited on page 94).

Smets, Philippe. 2007. "Analyzing the combination of conflicting belief functions." *Information Fusion* 8:387–412. (Cited on page 49).

———. 2005. "Belief functions on real numbers." *International Journal of Approximate Reasoning* 40:181–223. (Cited on page 39).

———. 1993. "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem." *International Journal of Approximate Reasoning* 9 (1): 1 –35. doi:http://dx.doi.org/10.1016/0888-613X(93)90005-X. (Cited on page 58).

Smets, Philippe. 1997. "The alpha-junctions: Combination operators applicable to belief functions." In *First Int. Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQUARU-FAPR'97),* 1244:131–153. Lecture Notes in Computer Sciences. Bad Honef, Germany: Springer International Publishing. (Cited on pages 52, 56).

———. 2002. "The application of the matrix calculus to belief functions." *International Journal of Approximate Reasoning* 31 (1–2): 1 –30. doi:http://dx.doi.org/10.1016/S0888-613X(02)00066-X. (Cited on pages 55, 56, 73, 77).

———. 1995. "The canonical decomposition of weighted belief." In *Int. Joint. Conf. on Artificial Intelligence,* 1896–1901. (Cited on pages 54, 92).

———. 1990. "The combination of evidence in the transferable belief model." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (5): 447–458. (Cited on page 34).

Smets, Philippe, and Robert Kennes. 1994. "The transferable Belief Model." *Artificial Intelligence* 66 (2): 191–234. (Cited on pages 33, 58).

Sunberg, Zachary, and Jonathan Rogers. 2013. "A belief function distance metric for orderable sets." *Information Fusion* 14 (4): 361 –373. doi:http://dx.doi.org/10.1016/j.inffus.2013.03.003. (Cited on page 80).

Tessem, Bjørnar. 1993. "Approximations for efficient computation in the theory of evidence." *Artificial Intelligence* 61 (2): 315 –329. doi:http://dx.doi.org/10.1016/0004-3702(93)90072-J. (Cited on page 80).

Walley, Peter. 1991. *Statistical reasoning with imprecise probabilities.* Chapman / Hall/CRC. (Cited on page 41).

———. 2000. "Towards a unified theory of imprecise probability." *International Journal of Approximate Reasoning* 24 (2): 125 –148. doi:http://dx.doi.org/10.1016/S0888-613X(00)00031-1. (Cited on page 41).

Wen, Chenglin, Yingchang Wang, and Xiaobin Xu. 2008. "Fuzzy Information Fusion Algorithm of Fault Diagnosis Based on Similarity Measure of Evidence." In *Advances in Neural Networks - ISNN 2008: 5th International Symposium on Neural Networks, ISNN 2008, Beijing, China, September 24-28, 2008, Proceedings, Part II,* edited by Fuchun Sun, Jianwei Zhang, Ying Tan, Jinde Cao, and Wen Yu, 506–515. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-87734-9_58. (Cited on page 80).

Xu, Philippe, Franck Davoine, Hongbin Zha, and Thierry Denœux. 2015. "Evidential calibration of binary {SVM} classifiers." *International Journal of Approximate Reasoning:* –. doi:http://dx.doi.org/10.1016/j.ijar.2015.05.002. (Cited on pages 12, 34).

Yager, Ronald R. 1987. "On the Dempster-Shafer framework and new combination rules." *Information Sciences* 41:93–138. (Cited on page 92).

———. 1986. "The entailment principle for dempster—shafer granules." *International Journal of Intelligent Systems* 1 (4): 247–262. (Cited on page 62).

Zadeh, Lotfi A. 1986. "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination." *the Artificial Intelligence Magazine* 7:85–90. (Cited on pages 93, 102).

———. 1978. "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems* 1:3–28. (Cited on page 39).

Zhang, Peng, Isabelle Gardin, and Patrick Vannoorenberghe. 2007. "Information fusion using evidence theory for segmentation of medical images." In *International Colloquium on Information Fusion,* 265–272. Xi'An, P.R.China. (Cited on page 49).

Zouhal, Lalla Mariem, and Thierry Denœux. 1998. "An evidence-theoretic k-NN rule with parameter optimisation." *IEEE Transactions on Systems, Man and Cybernetics. Part C: Application and reviews* 28 (2): 263 –271. (Cited on pages 80, 86).

This bibliography has 125 entries.