Complementary Lipschitz continuity results for the distribution of intersections or unions of independent random sets in finite discrete spaces

John Klein

HAL Id: hal-02108324

https://hal.science/hal-02108324

Submitted on 9 May 2019

Complementary Lipschitz continuity results for the distribution of intersections or unions of independent random sets in finite discrete spaces

John Klein

Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille

Abstract

We prove that intersections and unions of independent random sets in finite spaces achieve a form of Lipschitz continuity. More precisely, given the distribution of a random set Ξ, the function mapping any random set distribution to the distribution of its intersection (under independence assumption) with Ξ is Lipschitz continuous with unit Lipschitz constant if the space of random set distributions is endowed with a metric defined as the L_k norm distance between inclusion functionals also known as commonalities. Moreover, the function mapping any random set distribution to the distribution of its union (under independence assumption) with Ξ is Lipschitz continuous with unit Lipschitz constant if the space of random set distributions is endowed with a metric defined as the L_k norm distance between hitting functionals also known as plausibilities.

Using the epistemic random set interpretation of belief functions, we also discuss the ability of these distances to yield conflict measures. All the proofs in this paper are derived in the framework of Dempster-Shafer belief functions. Let alone the discussion on conflict measures, it is straightforward to transcribe the proofs into the general (non necessarily epistemic) random set terminology.

1

1 Introduction

When one is interested in a point-valued random variable but has access to set-valued (imprecise) observations of this latter, Dempster [8] proposed to use a probabilistic model relying on multi-valued mappings. This model was further developed by Shafer [40] in a self-contained framework known today as Dempster-Shafer theory, evidence theory or belief function theory. In this framework, the uncertainty on the value of the random variable is equivalently[1] captured (among others) by set functions known as the belief, plausibility and commonality functions. These functions evaluate respectively how likely it is that the imprecise observations imply / are consistent / are implied by some event. In this paper, we focus on the case where these events are disjunctions of elements of a finite and discrete space.

Besides, belief functions are also known to be formally equivalent to random sets [32] and are interpretable as epistemic ones [3]. A random set is a random element whose realizations are set-valued. The probability masses governing the random set can also be uniquely characterized by set functions that are capacities [28] (non additive measures). Some of these set functions are:

- the containment functional that captures the probabilities that a given set contains the random set,

- the hitting functional (or capacity functional) that captures the probabilities that the random set intersects a given set,

- the inclusion functional that captures the probabilities that a given set is included in the random set.

The above functions are the respective random set terminology for the belief, plausibility and commonality functions.

In the deterministic setting, one can observe a form of consistency between operations like union or intersection with set-distances in the sense that, for some of these distances, if one intersects (resp. unites) the same set with two other ones, say X_1 and X_2, then the obtained intersections (resp. unions) are at least as close as X_1 and X_2 were before.

In this paper, we investigate if this observed consistency propagates to some extent to the random setting. We prove that if one intersects (resp.

---

[1]These functions are in bijective correspondence [40].

2

unites) the same random set with two other ones, say Ξ_1 and Ξ_2, then the obtained random intersections (resp. unions) have distributions that are at least as close as the distributions of Ξ_1 and Ξ_2 were before. These results are dependent on the chosen metric for random set distributions. We examine metrics that consist in L_k norm based distances between either hitting or inclusion functionals. In addition, the results can be rephrased as Lipschitz continuity (with a unit Lipschitz constant) for functions that map any random set distribution to the distribution of its intersection (resp. union) with a given fixed random set.

In the belief function framework, the closest related works are those of Loudahi et al. [25, 26]. The authors proved that the consistency under study holds between some belief function distances and combination operators that yield the distribution of the intersection or union of independent random sets. The results that we introduce in this paper involves distances that are computationally more tractable than those introduced in [25, 26] but also rely on independence assumptions. In the random set literature, many results regarding unions of i.i.d. random sets and random set metrics are available [30] but they do not address Lipschitz continuity. Also, we do not require the examined random sets to be identically distributed.

Furthermore, building upon recent work from Pichon and Jousselme [35], we also investigate if our results can be instrumental to span new degrees of conflict. We prove that the consistency of a distance with the conjunctive rule makes the corresponding conflict degree compliant with at least one requirement discussed in Destercke and Burger [11]. However, we also show that several distances relying on an L_k norm are not appropriate to yield a degree of conflict as suggested in [35] when k is finite. These last developments focus on information fusion aspects of the belief function framework only and do not generalize to non-epistemic random sets.

This article is organized as follows: section 2 gives necessary background on the theory of belief functions and random sets. Section 3 is an overview of distances between random set distributions and the sought Lipschitz continuity property is stated. Section 4 contains the main results of the paper, i.e. Lipschitz continuity for the distribution of intersection and union of independent random sets. Finally, in section 5, we make use of the aforementioned results to investigate if the examined distances can yield relevant degrees of conflict in the belief function framework. All the proofs of the newly introduced results are given in the appendices.

Most of the paper is written using belief function terminology and usual

3

notations in this framework but it can be transcribed to the random set framework by merely switching the set function names as explained in the first paragraphs of this introduction. Special care was paid to allow easy readability for readers familiar with any of these two frameworks.

2 Belief functions and random sets

In this section, some mathematical notations for baseline belief function and random set concepts are given. The reader is expected to be familiar with one of these frameworks. More material on belief functions basics is found for instance in [40, 7] and on random sets in [30, 31, 4].

Belief functions can be applied in the context of uncountable spaces [32, 43, 31, 10] but a majority of results were derived in the finite case and we also make this assumption in this article.

2.1 Baseline definitions

A random set in a finite and discrete space $\Omega = \{\omega_t\}_{t=1}^n$ is a random element whose realizations are subsets of Ω. When one is interested in a point valued variable but has access to set valued (imprecise) observations, one can try to infer the distribution of an epistemic random set [29]. Belief functions are in line with this epistemic interpretation. When one is interested in a set valued variable and has access to corresponding samples, one can try to infer the distribution of an ontic random set [28].

Both types of uncertainty lead to formally equivalent objects although these objects need occasionally to be processed and understood in different ways [3]. In the finite (and consequently countable) setting, the distribution of a random set Ξ_i is a set function called **mass function** and is denoted by m_i. The power set 2^Ω is the set of all subsets of Ω and it is the domain of mass functions. For any $A \in 2^\Omega$, the **cardinality** of this set is denoted by $|A|$ and we thus have $|\Omega| = n$. The cardinality of 2^Ω is denoted by $N = 2^n$. Mass functions have $[0, 1]$ as co-domain and they sum to one: $\sum_{A \in 2^\Omega} m_i(A) = 1$. A **focal element** of a mass function m_i is a set $A \subseteq \Omega$ such that $m_i(A) > 0$. A mass function having only one focal element A is called a **categorical mass function** and is denoted by m_A. A **simple mass function** is the convex combination of m_Ω with some categorical mass function m_A.

Several alternative set functions are commonly used as equivalent characterizations of Ξ_i. The **belief**, **plausibility** and **commonality** functions of a set A are defined as

$$bel_i(A) = \sum_{E \subseteq A, E \neq \emptyset} m_i(E), \tag{1}$$

$$pl_i(A) = \sum_{E \cap A \neq \emptyset} m_i(E), \tag{2}$$

$$q_i(A) = \sum_{E \supseteq A} m_i(E) \tag{3}$$

and respectively represent how much likely it is that A contains / intersects / is included in the underlying random set. In the random set literature, these set functions are respectively referred to as the **containment**, **hitting** and **inclusion functionals**. When the empty set has a positive mass, another representation is provided by **implicability** functions b_i. These functions are closely related to belief and plausibility functions through the following relations: $\forall A \in 2^\Omega$,

$$b_i(A) = bel_i(A) + m_i(\emptyset), \tag{4}$$
$$b_i(A) = 1 - pl_i(A^c). \tag{5}$$

Another useful concept is the **negation** (or complement) $\overline{m}_i$ of a mass function m_i introduced by Dubois and Prade [14]. The function $\overline{m}_i$ is such that $\forall A \subseteq \Omega$, $\overline{m}_i(A) = m_i(A^c)$ with $A^c = \Omega \setminus A$. The authors also provide a result that will be instrumental in the proof of proposition 4. This result reads

$$\overline{b}_i(A^c) = q_i(A), \forall A \subseteq \Omega, \tag{6}$$

where $\overline{b}_i$ denotes the implicability function in correspondence with $\overline{m}_i$.

2.2 Intersection, union and information fusion

Information fusion in the framework of belief functions is performed using an operator mapping an arbitrary large set of input mass functions to a single output mass function which summarizes all information contained in the input ones. On top of this minimal requirement, the operator must also follow a certain policy in the way that the information encoded in the input

mass functions is processed to build the output one. There are two canonical and dual such policies: conjunctivity and disjunctivity.

Suppose $\sqsubseteq$ denotes an informational partial order [46, 14, 9] for mass functions in the sense that one writes $m_1 \sqsubseteq m_2$ if m_1 contains at least as much (epistemic) information as m_2. Following [13], a fusion operator is conjunctive if its output is more informative than any input. The conjunctive rule operator [44] denoted by $\ocircle$ is defined as follows

$$m_1 \ocircle m_2 (E) = \sum_{\substack{A,B \subseteq \Omega \\ A \cap B = E}} m_1(A) m_2(B), \forall E \subseteq \Omega. \tag{7}$$

The conjunctive rule is associative and commutative and the generalization of the above expression to more than two input mass functions is immediate. This rule is the unnormalized version of Dempster's rule [8] and on the random set side, it can be understood as the distribution of the intersection of two independent[2] random sets. Obviously, this operator is a generalization of the set intersection as we have $m_A \ocircle m_B = m_{A \cap B}$ for any two subsets A and B of Ω. For the sake of equation concision we adopt the following notation $m_{1 \cap 2} = m_1 \ocircle m_2$. This combination is very simple to compute when dealing with commonality functions:

$$q_{1 \cap 2}(A) = q_1(A) q_2(A), \forall A \subseteq \Omega. \tag{8}$$

It can be easily proved that the conjunctive rule is conjunctive. Consider the informational partial order based on commonalities $\sqsubseteq_q$ which is defined as

$$m_1 \sqsubseteq_q m_2 \Leftrightarrow q_1(E) \leq q_2(E), \forall E \subseteq \Omega. \tag{9}$$

Since $q_{1 \cap 2}$ is the elementwise multiplication of q_1 and q_2, we obtain $m_{1 \cap 2} \sqsubseteq_q m_1$ and $m_{1 \cap 2} \sqsubseteq_q m_2$.

When $m_2 = m_E$, i.e. m_2 is categorical, the result of the conjunctive combination between m_1 and m_2 is referred to as the conditioning of m_1 given E because this operation is a generalization of probabilistic conditioning[3]

---

[2]There are several notions of independence for random sets [4, chapter 2]. In this paper, we only consider the usual probabilistic notion, i.e. joint distributions factorizing as the product of their marginals.

[3]If focal elements of m_1 are singletons, i.e. the random set is point valued, then Dempster's conditioning coincides with Bayes rule: $m_{1|E}(A) = \frac{m_1(A)}{m_1(E)}$.

6

[40]. The mass function $m_1 \bigcirc m_E$ is also denoted by $m_{1|E}$. The following property of the implicability function w.r.t. conditioning will be instrumental in some proofs:

Lemma 1. *For any mass function m_1, any categorical mass function m_E and any subset $A \subseteq \Omega$, we have*

$$b_{1|E}(A) = b_1\left(\left(E \setminus A\right)^c\right) = b_1\left(E^c \cup A\right). \tag{10}$$

To the best of our knowledge, this property is not reported in the belief function literature, we thus provide a proof in B.

As for disjunction, the output is required to be less informative than any input and it is thus considered as an extremely conservative fusion policy. The disjunctive rule operator [44] denoted by $\bigcirc\!\!\!\!\!\bigcirc$ is defined as follows

$$m_1 \bigcirc\!\!\!\!\!\bigcirc m_2(E) = \sum_{\substack{A,B \subseteq \Omega \\ A \cup B = E}} m_1(A)\, m_2(B),\, \forall E \subseteq \Omega. \tag{11}$$

The disjunctive rule is also associative and commutative and it is a generalization of set union as we have $m_A \bigcirc\!\!\!\!\!\bigcirc m_B = m_{A \cup B}$ for any two subsets A and B of Ω. We denote by $m_{1 \cup 2}$ the result of the following combination: $m_1 \bigcirc\!\!\!\!\!\bigcirc m_2$. On the random set side, $m_{1 \cup 2}$ is understood as the distribution of the union of two independent random sets. The disjunctive combination is very simple to compute when dealing with implicability functions:

$$b_{1 \cup 2}(A) = b_1(A)\, b_2(A),\, \forall A \subseteq \Omega. \tag{12}$$

The disjunctivity of this rule can be proved using the partial order based on implicabilities $\sqsubseteq_b$. This latter reads

$$m_1 \sqsubseteq_b m_2 \Leftrightarrow b_1(E) \geq b_2(E),\, \forall E \subseteq \Omega. \tag{13}$$

Since $b_{1 \cup 2}$ is the elementwise multiplication of b_1 and b_2, we obtain the desired conclusion.

The disjunctive rule is related to the conjunctive rule by the following De Morgan relation [14]: for any mass functions m_1 and m_2

$$\overline{m_1 \bigcirc m_2} = \overline{m}_1 \bigcirc\!\!\!\!\!\bigcirc \overline{m}_2. \tag{14}$$

7

In K, we mention a more general family of combination rules for belief functions which encompasses the conjunctive and disjunctive rules. These rules are known as α-junctions [41]. Since α-junctions have a more limited impact in the belief function literature than the conjunctive and disjunctive rules, we chose not to mention them in the main body of this article. However, the results that we prove in the next sections do propagate to α-junctions. See K for the corresponding proofs.

2.3   The space of mass 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 $\{m_A\}_{A \subseteq \Omega}$. This simplex is also called **mass space** [5] and has finite Lebesgue measure but contains uncountably many mass functions.

Embedding mass functions in a vector space is particularly useful when computing either $m_{1 \cap 2}$ or $m_{1 \cup 2}$ because they can be obtained as the dot product of some matrix with one of the input mass functions (seen as a column vector) [42]. Each such matrix is in one-to-one correspondence with the other mass function. The vector form of any set function will be denoted using bold characters, for instance, the vector form of a mass function m_i is denoted by $\mathbf{m}_i$.

Let $\mathbf{S}_1$ denote the **specialization** matrix [14] in bijective correspondence with m_1. Each entry of $\mathbf{S}_1$ is given by $S_1(A, B) = m_{1|B}(A)$. From a geometric point of view [6], each column of $\mathbf{S}_1$ corresponds to the vertex of a polytope $\mathcal{P}_1$, called the **conditional subspace** of m_1. Any mass function $m \in \mathcal{P}_1$ is the result of the combination of m_1 with another mass function using $\bigcirc\!\!\!\!\cap$. Most importantly, for any mass functions m_1 and m_2, we have

$$\mathbf{m}_{1 \cap 2} = \mathbf{S}_1 \cdot \mathbf{m}_2. \tag{15}$$

Let $\mathbf{G}_1$ denote the **generalization matrix** in bijective correspondence with m_1. Each entry of $\mathbf{G}_1$ is given by $G_1(A, B) = m_{1 \cup B}(A)$. For any mass functions m_1 and m_2, we have

$$\mathbf{m}_{1 \cup 2} = \mathbf{G}_1 \cdot \mathbf{m}_2. \tag{16}$$

There are also transfer matrices allowing to turn mass functions in commonality or implicability functions using a right-handed dot product. They are presented in more details in K.

3 From mass function metrics to Lipschitz continuity

In this section, we will first recall the definitions of some existing distances between mass functions. We focus on (full) metrics and do not discuss dissimilarities [45, 47] which have fewer baseline properties as compared to metrics. The Lipschitz continuity property that we seek will then be stated and its desirability will be justified by analyzing set-distances.

3.1 Vector-based distances

A distance, or metric, provides a positive real value assessing the discrepancies between two elements. Let us first give a general definition of such an application when the compared vectors are mass functions:

Definition 1. Given a domain Ω and its related mass space $\mathcal{M}$, a mapping $d : \mathcal{M} \times \mathcal{M} \longrightarrow [0, a]$ with $a \in \mathbb{R}^+$ is a **distance** between two mass functions m_1 and m_2 defined on Ω if the following properties hold:

- Symmetry : $d(m_1, m_2) = d(m_2, m_1)$,

- Definiteness : $d(m_1, m_2) = 0 \Leftrightarrow m_1 = m_2$,

- Triangle inequality : $d(m_1, m_2) \leq d(m_1, m_3) + d(m_3, m_2)$.

If the mapping fails to possess some of the above properties, then it degrades into unnormalized distance, dissimilarity or pseudo-distance. Only full metrics are able to provide a positive finite value that matches the intuitive notion of gap^4 between elements of a given space.
If $a \neq +\infty$, then the distance is bounded and if in addition $a = 1$, the distance is **normalized**. Provided that a mass function distance d is bounded, this

---

[4]This term was used by Frechet [15] in his early works on metric spaces, *i.e.* spaces endowed with a distance.

distance can be normalized by dividing it with $\rho = \max_{A,B \in 2^{\Omega}} d\left(m_A, m_B\right)$ which is the **diameter** of $\mathcal{M}$ [26].

The most popular metric in the belief function literature is Jousselme distance [17]. It is based on an inner product relying on a similarity matrix. This distance is given by:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2} (\mathbf{m_1} - \mathbf{m_2})^T \cdot \mathbf{D} \cdot (\mathbf{m_1} - \mathbf{m_2})}, \qquad (17)$$

where $\mathbf{m}_i$ denotes the column vector version of mass function m_i and $\mathbf{D}$ is the Jaccard similarity matrix [16] between focal elements. Its components are:

$$D(A, B) = \begin{cases} 1 & \text{if } A = B = \emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \text{otherwise} \end{cases}. \qquad (18)$$

Thanks to the matrix $\mathbf{D}$, Jousselme distance takes into account the dependencies between the base vectors of $\mathcal{M}$. Consequently, the poset structure of $\left(2^{\Omega}, \sqsubseteq\right)$ has an impact on distance values, allowing a better match with the user's expectations.

Many other mass function distances are defined similarly by substituting matrix $\mathbf{D}$ with another matrix evaluating the similarity between base vectors in different ways [12, 5]. Experimental material in [18] shows that these distances are highly correlated to d_J.

Observe that the aforementioned distances are the L_2 norm of the difference of two vectors which are obtained by applying the same linear mapping to each mass function under comparison. We can thus build other distances by resorting to other norms. In particular, when the linear mapping maps a mass function to its corresponding plausibility, commonality or implicability function[5], we obtain distances that will be instrumental in the sequel of this paper. The formal definition of these distances follows.

Definition 2. For some family $f \in \{q, bel, pl, b\}$ of set functions in bijective correspondence with mass functions, an L_k norm based **f-distance** $d_{f,k}$ is the following mapping:

$$d_{f,k} : \begin{array}{rcl} \mathcal{M} \times \mathcal{M} & \to & [0, 1], \\ (m_1, m_2) & \to & \frac{1}{\rho} \left\| \mathbf{f_1} - \mathbf{f_2} \right\|_k. \end{array}$$

---

[5]See [42] for the definition of this linear mappings.

$\mathbf{f}_i$ is the vector representation of the set function f_i (in correspondence with m_i) and ρ is a normalization factor given by

$$\rho = \max_{A,B \in 2^\Omega} \|\mathbf{f}_A - \mathbf{f}_B\|_k.$$

For any vector $\mathbf{f} \in \mathbb{R}^N$, its L^k norm are given by:

$$\|\mathbf{f}\|_k = \left(\sum_{A \subseteq \Omega} |f(A)|^k \right)^{\frac{1}{k}}. \tag{19}$$

Given relation (5), we see that $d_{pl,k} = d_{b,k}$ for any k. Consequently, we do not further mention distances between implicability functions in the sequel of this article. We end this subsection with a small result giving closed form expressions for constant ρ for distances between plausibilities and commonalities.

Lemma 2. *For L_k norm based distance between commonality or plausibility functions, we have*

$$\rho = \begin{cases} (N-1)^{1/k} & \textit{if } k < \infty \\ 1 & \textit{if } k = \infty \end{cases}. \tag{20}$$

Proof. (sketch) Given proposition 2 in [20], for any of the distances evoked in the lemma, we have

$$\max_{A,B \in 2^\Omega} d(m_A, m_B) = d(m_\Omega, m_\emptyset). \tag{21}$$

Finally, for $f \in \{q, pl\}$, we always have $|f_\Omega(A) - f_\emptyset(A)| = 1$ if $A \neq \emptyset$ and $|f_\Omega(\emptyset) - f_\emptyset(\emptyset)| = 0$. $\qquad \square$

3.2 Matrix-based distances

Since specialization and generalization matrices are also in bijective correspondence with mass functions, we can use the same recipe as in definition 2 to build new mass function distances. The only difference is that mass functions are mapped to matrices and one must thus resort to matrix norms instead of vector norms. Such distances were first introduced in [25, 26]. A subset of these distances are defined as follows:

Definition 3. The L_k norm based **specialization distance** $d_{\mathrm{spe},k}$ is the following mapping:

$$d_{\mathrm{spe},k}: \begin{array}{ccc} \mathcal{M} \times \mathcal{M} & \to & [0,1], \\ (m_1, m_2) & \to & \frac{1}{\rho} \left\| \mathbf{S}_1 - \mathbf{S}_2 \right\|_k. \end{array}$$

$\mathbf{S}_i$ is the specialization matrix in correspondence with m_i and ρ is a normalization factor given by

$$\rho = \begin{cases} (2(N-1))^{1/k} & \text{if } k < \infty \\ 1 & \text{if } k = \infty \end{cases}$$

For any matrix $\mathbf{F} \in \mathbb{R}^N$, its L^k norm is given by:

$$\left\| \mathbf{F} \right\|_k = \left(\sum_{A,B \subseteq \Omega} |F(A,B)|^k \right)^{\frac{1}{k}}. \tag{22}$$

It was proved in [26] that if we use generalization matrices in the same way as the above definition, we obtain a distance that coincides with the specialization distance.

Other matrix norms were investigated in [25, 26], i.e. operator norms. These norms lead to mass function distances that have fewer desirable properties[6] as compared to L_k matrix norm based ones. They are thus not mentioned in this article.

3.3 Union, intersection and set-distances

There are two main types of metrics between sets [22]: those accounting for how many elements are shared by the subsets and those that also account for the number of elements that they do not share. Examples of each category are the following:

- the Jaccard distance

$$d_{\mathrm{jac}}(A,B) = \begin{cases} 0 & \text{if } A = B = \emptyset \\ 1 - D(A,B) = \frac{|A \triangle B|}{|A \cup B|} & \text{otherwise} \end{cases},$$

---

[6]These distances are not consistent with informational partial orders that generalize set inclusion. See [20] for a definition of the consistency of mass function distances with partial orders.

- the (normalized) Hamming set distance $d_{\text{ham}}(A, B) = \frac{|A \Delta B|}{n}$,

where Δ denotes the set symmetric difference.

The Jaccard distance belongs to the first type of metric while the Hamming distance belongs to the second one. The following example illustrates their main difference.

Example 1. In this example, we replace subsets by their binary representations, i.e. $A = 0011$ means that $n = 4$ and the elements of A are the third and fourth elements of Ω. We have

$$d_{\text{jac}}(0011, 0110) = \frac{2}{3} \text{ and } d_{\text{ham}}(0011, 0110) = \frac{2}{4}, \tag{23}$$

$$\text{while } d_{\text{jac}}(00011, 00110) = \frac{2}{3} \text{ and } d_{\text{ham}}(00011, 00110) = \frac{2}{5}. \tag{24}$$

The hamming distance decreases as a fifth element is contained in Ω and thus this distance depends on elements that the subset do not share.

We can wonder how these distances interact with set operations like intersection and union. Actually, the nature of these interactions are highly dependent on how one wishes to perform information fusion using either intersections or unions:

- Suppose subsets represent a collection of candidate contents that a classifier must assign to an input image. If we want to evaluate if two images have similar contents, we can use a set distance between their imprecise tags. Suppose image_a is tagged as {cat or dog} and image_b is tagged as {dog or bike}. If we learn from a second classifier that both images contain pictures of an animal, then we deduce that the image contents are more likely to be closer after inserting this information. More formally, if one intersects both A and B with a third party subset C, the result of these intersections cannot be more distant than A and B were initially, which reads

$$\text{(a)} \quad d(A \cap C, B \cap C) \leq d(A, B).$$

- Suppose subsets are lists of attributes of some streaming video service users. Suppose that it is known that user_a likes action movies, is a male and lives in the US. Suppose user_b likes comedies, is a male and

lives in the UK. Suppose we learn that both of them also like science fiction movies, then we deduce that these users have closer profiles than previously thought. More formally, if one unites both A and B with a third party subset C, the result of these unions cannot be more distant than A and B were initially, which reads

$$\text{(b)} \quad d\left(A \cup C, B \cup C\right) \leq d\left(A, B\right).$$

Observe that in both of these examples, one adopts a conjunctive information fusion policy in the sense that aggregation results are more informative than each input. The conjunctive/disjunctive nature of an operator (like intersection or union) depends on the type of underlying uncertainty. What these examples are meant to highlight is that, should you intend to combine the informative content of subsets using either intersections or unions, then a set distance should comply to either (a) or (b) in order to translate in the numerical distance values that informative contents are more similar after fusion.

It can be proved that the Hamming set distance verifies (a) and (b) while the Jaccard distance verifies (b) only, c.f. A for more details. From the applicative contexts of the above examples, we see that the desirability of property (a) or (b) depends on the information fusion operator. Outside the scope of information fusion, we may not require any of these properties for a set metric. When one intends to perform information fusion with random sets, it makes sense to wonder if some mass function distances can generalize these properties with respect to information fusion operators defined for them.

3.4   Union, intersection and random set distribution distances

One way to generalize the properties (a) or (b) to random sets is stated by the following property:

Definition 4. Let $*$ be a combination operator and d a mass function distance. d is said to be **consistent** with respect to $*$ if any of the following conditions is verified:

(i) for any mass functions m_1, m_2 and m on Ω:

$$d\left(m_1 * m, m_2 * m\right) \leq d\left(m_1, m_2\right). \tag{25}$$

(ii) for any mass function m, the mapping $F_m : \mathcal{M} \longrightarrow \mathcal{M}$ of the form

$$F_m(m_0) = m_0 * m \tag{26}$$

is Lipschitz continuous with 1 as Lipschitz constant.

The equivalence between the two conditions follows from the very definition of Lipschitz continuity with 1 as Lipschitz constant. Indeed, for mapping F_m to qualify as such, it means that we have

$$d(F_m(m_1), F_m(m_2)) \leq d(m_1, m_2)$$

for any pair of mass functions (m_1, m_2). In the remainder of this article, we will refer to this property either as consistency property between an operator and distance or, by small abuse of language, as **1-Lipschitz continuity** of an operator w.r.t. to a distance.

Under this property, repeated combinations with a given mass function m cannot pull away any pair of mass functions. Such mappings are also called non expansive or short maps. Lipschitz continuity is stronger than uniform continuity. In particular, it implies a form of regularity for the corresponding combination mechanism in the sense that the norm of its gradient is bounded by 1 meaning that the combined mass function does not change very fast or wiggle in the vicinity of functions m_0 or m.

From an informative content standpoint, this property also has an impact. Suppose a mass function m is separable[7], i.e. the combination under rule $*$ of elementary pieces of information embodied by simple mass functions yields function m. Using a consistent distance w.r.t. $*$, mass functions are all the closer as their decompositions involve identical elementary components.

Proving that a fusion operator achieves Lipschitz continuity is not trivial because $\mathcal{M}$ is not finite but instead a compact subset of an uncountable space. In [25], Loudahi et al. established the consistency of the L_1 and L_∞ based specialization distances w.r.t. the conjunctive and disjunctive rules. Numerical experiments also show that the L_2 based specialization distance is not consistent w.r.t. the conjunctive or disjunctive rule in the sense of definition 4. The experiments also show that Jousselme distance is not consistent

---

[7]Shafer [40] introduced this terminology for decompositions w.r.t. Dempster's rule but we understand it in a more general perspective here by considering decompositions w.r.t. some arbitrary rule $*$.

w.r.t. the conjunctive rule. Its consistency with the disjunctive rule remains undetermined.

Observe that we compare random sets only in distribution here. One could think of other generalizations of properties (a) or (b) in probability or in an almost sure fashion.

4 Lipschitz continuity results

In this section, we provide new Lipschitz continuity results of L_k norm based distances between commonality or plausibility functions with the conjunctive and disjunctive rules.

4.1 Main results on Lipschitz continuity for the conjunctive rule

Proposition 1. *For $1 \leq k \leq \infty$, $\text{\textcircled{$\cap$}}$ is 1-Lipschitz continuous w.r.t. the L_k norm based q-distance $d_{q,k}$.*

See C for proof.

Proposition 2. *$\text{\textcircled{$\cap$}}$ is 1-Lipschitz continuous w.r.t. the L_∞ norm based pl-distance $d_{pl,\infty}$.*

See D for proof.

As compared to previous Lipschitz continuity results [25, 26], specialization distances have a greater time complexity as compared to commonality ones. Indeed, although the construction of specialization distances can be sped up [24], the time complexity for the specialization distance is quadratic in N. More precisely, the time complexity to build a specialization matrix is $O\left(N^{\frac{\log(3)}{\log(2)}}\right) \approx O\left(N^{1.58}\right)$. However, computing the norm of such a matrix has time complexity $O\left(N^2\right)$. The time complexity to compute a commonality function [19] is $O\left(N \log\left(N\right)\right)$ while that of computing the norm of commonality is $O\left(N\right)$. Given relations (5) and (6), the time complexity to compute a plausibility function is identical to that of commonality ones. Moreover, the memory complexity is obviously reduced as well.

4.2 Lipschitz continuity for the conjunctive rule: numerical experiments and counter-examples

This subsection contains experiments illustrating the (in)consistency of several mass function distances with respect to the conjunctive rule. We generate randomly [2] $1e4$ triplets of simple mass functions and check how many times inequality (25) is verified (when $* = \cap\!\!\!\!\bigcirc$) for several distances. The corresponding success rates are reported in Table 1.

Table 1: Consistency rates for several mass function distances w.r.t. $\cap\!\!\!\!\bigcirc$

Distance	d_J	$d_{q,1}$	$d_{q,2}$	$d_{q,\infty}$	$d_{pl,1}$	$d_{pl,2}$	$d_{pl,\infty}$	d_{spe}
Consistency rate	86.42%	100%	100%	100%	38.22%	63.60%	100%	100%

The results are compliant with proposition 1 and 2 as all commonality distances and $d_{pl,\infty}$ achieve 100% of success. The results also show that Jousselme distance and L_1 or L_2 norm based distances between plausibilities are not consistent with $\cap\!\!\!\!\bigcirc$. The rates also show that the circumstances in which Lipschitz continuity does not hold for these distances are not rare events.

To get a better insight as to why $d_{pl,k}$ is not consistent with $\cap\!\!\!\!\bigcirc$ when k is finite, we provide the following counter-example:

Example 2. Let $\Omega = \{a, b, c\}$. Suppose $m_1 = \frac{1}{2}m_{\{a,b\}} + \frac{1}{2}m_\Omega$, $m_2 = \frac{1}{2}m_{\{a,c\}} + \frac{1}{2}m_\Omega$ and $m_3 = m_{\{b\}}$. By conjunctive combination, we obtain

$$m_1 \cap\!\!\!\!\bigcirc m_3 = m_{\{b\}}, \tag{27}$$

$$\text{and } m_2 \cap\!\!\!\!\bigcirc m_3 = \frac{1}{2}m_{\{b\}} + \frac{1}{2}m_\emptyset. \tag{28}$$

The plausibilities are

	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Ω
pl_1	0	1	1	1	$\frac{1}{2}$	1	1	1
pl_2	0	1	$\frac{1}{2}$	1	1	1	1	1
$pl_{1\cap3}$	0	0	1	1	0	0	1	1
$pl_{2\cap3}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$

We see that

$$d_{pl,k}\left(m_1, m_2\right) = \frac{1}{\rho}\left(2\left(\frac{1}{2}\right)^k\right)^{1/k} \tag{29}$$

$$\text{while } d_{pl,k}\left(m_1 \text{\textcircled{\tiny$\cap$}} m_3, m_2 \text{\textcircled{\tiny$\cap$}} m_3\right) = \frac{1}{\rho}\left(4\left(\frac{1}{2}\right)^k\right)^{1/k}. \tag{30}$$

This counter-example tends to show that the inconsistency of these distances lies (at least partially) in the way that the mass of the empty set is assigned by the conjunctive rule.

4.3 Main results on Lipschitz continuity for the disjunctive rule

Proposition 3. *For $1 \le k \le \infty$, $\text{\textcircled{\tiny$\cup$}}$ is 1-Lipschitz continuous w.r.t. the L_k norm based pl-distance $d_{pl,k}$.*

See E for proof.

Proposition 4. *$\text{\textcircled{\tiny$\cup$}}$ is 1-Lipschitz continuous w.r.t the L_∞ norm based q-distance $d_{q,\infty}$.*

See F for proof.

The same type of arguments outlining the added value of our new Lipschitz continuity results in the conjunctive case also hold in the disjunctive one. The distances between plausibilities or commonalities have a smaller time and memory complexities as compared to the specialization distances. It must be noted that $d_{\text{spe},1}$, $d_{\text{spe},\infty}$, $d_{q,\infty}$ and $d_{pl,\infty}$ are the only distances that are reported to be consistent with both the conjunctive and disjunctive rules. As the numerical experiments presented in the next paragraph will show, L_k norm based distances between commonalities are not consistent with $\text{\textcircled{\tiny$\cup$}}$ when k is finite. The counter-example presented in 4.2 proves that L_k norm based distances between plausibilities are not consistent with $\text{\textcircled{\tiny$\cap$}}$ when k is finite.

4.4 Lipschitz continuity for the disjunctive rule: numerical experiments and counter-examples

This subsection contains experiments illustrating the (in)consistency of several mass function distances with respect to the disjunctive rule. We gener-

ate randomly [2] $1e4$ triplets of simple mass functions and check how many times inequality (25) is verified (when $* = \bigcirc$) for several distances. The corresponding success rates are reported in Table 2.

Table 2: Consistency rates for several mass function distances w.r.t. $\bigcirc$

Distance	d_J	$d_{q,1}$	$d_{q,2}$	$d_{q,\infty}$	$d_{pl,1}$	$d_{pl,2}$	$d_{pl,\infty}$	d_{spe}
Consistency rate	100%	94.76%	94.09%	100%	100%	100%	100%	100%

The results are compliant with proposition 3 and 4 as all plausibility distances and $d_{q,\infty}$ achieve 100% of success. The results also show that L_1 or L_2 norm based distances between commonalities are not consistent with $\bigcirc$. The consistency of Jousselme distance can be conjectured. This distance also achieves 100% of success if one draws random mass functions and not just random simple mass functions.

To get a better insight as to why $d_{q,k}$ is not consistent with $\bigcirc$ when k is finite, we provide the following counter-example:

Example 3. Let $\Omega = \{a, b, c\}$. Suppose $m_1 = m_{\{a\}}$, $m_2 = m_{\{a,c\}}$ and $m_3 = m_{\{b\}}$. By disjunctive combination, we obtain

$$m_1 \bigcirc m_3 = m_{\{a,b\}}, \tag{31}$$

$$\text{and } m_2 \bigcirc m_3 = m_\Omega. \tag{32}$$

The commonalities are

	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	Ω
q_1	1	1	0	0	0	0	0	0
q_2	1	1	0	0	1	1	0	0
$q_{1\cup3}$	1	1	1	1	0	0	0	0
$q_{2\cup3}$	1	1	1	1	1	1	1	1

We see that

$$d_{q,k}(m_1, m_2) = \frac{1}{\rho}(2)^{1/k} \tag{33}$$

$$\text{while } d_{q,k}(m_1 \bigcirc m_3, m_2 \bigcirc m_3) = \frac{1}{\rho}(4)^{1/k}. \tag{34}$$

19

5 Conflict degrees spanned by consistent distances with the conjunctive rule

When information sources support antagonistic assumptions, it is important to provide a way to numerically assess the level of inconsistency in their respective messages. This is the purpose of degrees of conflict defined in the framework of belief functions. If such a degree of conflict is bounded, we can use different information fusion strategies in order to make more robust decisions.

In the theory of belief functions, such a situation typically occurs when there is a pair of subsets (A, B) such that $A \cap B = \emptyset$ and $m_1(A) > 0$ and $m_2(B) > 0$. In the following paragraphs, we give a brief reminder of existing conflict degrees in the belief function literature as well as desirable properties for such degrees. Next, we also comment on the advisability of building new degrees using distances that are consistent with $\cap$.

5.1 Assessing the degree of conflict between belief functions

In his pioneering article, Dempster [8] already provides a way to assess the degree of conflict between two mass functions. Let κ denote this criterion which is known as **Dempster's degree of conflict** and reads

$$\kappa(m_1, m_2) = m_{1 \cap 2}(\emptyset). \tag{35}$$

More recently, Destercke and Burger [11] outline that this degree can be built upon a consistency measure ϕ which evaluates to what extent a single mass function is not self-contradictory. In the case of Dempster's degree of conflict, this measure is simply given by

$$\phi(m) = 1 - m(\emptyset). \tag{36}$$

They also introduce the following strong consistency measure Φ which is such that

$$\Phi(m) = \max_{a \in \Omega} pl(\{a\}). \tag{37}$$

This second measure is the L_∞ norm of the contour function[8]. It is stronger in the sense that $\phi(m) < 1 \Rightarrow \Phi(m) < 1$ while the opposite implication is

---

[8]The contour function is the restriction of the plausibility function to singletons.

not true. This means that Φ can detect a finer level of self-(in)consistencies in the information encoded by some mass function.

Destercke and Burger [11] also introduce the following list of desirable properties for some degree of conflict $\mathcal{C}$:

(i) (extreme conflict values) $\mathcal{C}(m_1, m_2) = 0$ iff m_1 and m_2 are non con-flicting and $\mathcal{C}(m_1, m_2) = 1$ iff $m_1 \textcircled{\cap} m_2 = m_\emptyset$,

(ii) (symmetry) $\mathcal{C}(m_1, m_2) = \mathcal{C}(m_2, m_1)$,

(iii) (imprecision monotonicity) if $m_1 \sqsubseteq m'_1$ then $\mathcal{C}(m'_1, m_2) \leq \mathcal{C}(m_1, m_2)$,

(iv) (ignorance is bliss) $\mathcal{C}(m_1, m_\Omega) = 1 - \mathfrak{I}(m_1)$ where $\mathfrak{I}$ is a consistency measure such as the aforementioned ones,

(v) (invariance to refinement) for some multi-valued mapping $\rho : \Omega \to 2^\Theta$ with $|\Omega| < |\Theta| < \infty$ and a mass function m'_1 such that $m'_1 \left(\underset{a \in A}{\cup} \rho(a) \right) = m_1(A)$ for any $A \subseteq \Omega$, we have $\mathcal{C}(m_1, m_2) = \mathcal{C}(m'_1, m'_2)$.

The definition of non-conflicting mass functions is not specified in property (i) because several such notions can be considered. The authors explain that if non-conflict means that the intersection of any focal element of m_1 with any focal element of m_2 is not empty then κ satisfies each property with $\mathfrak{I} = \phi$. Moreover, if non-conflict means that the intersection of all the focal elements of both mass functions is not empty then $K(m_1, m_2) = 1 - \underset{a \in \Omega}{\max} \, pl_{1 \cap 2}(\{a\})$ satisfies each property with $\mathfrak{I} = \Phi$. We will refer to K as the **degree of strong conflict**. The informational partial order in property (iii) is the specialization partial order [14] for both degrees κ and K.

5.2 Deriving new degrees of conflict

Prior to Destercke and Burger [11], several authors [27, 23] proposed to derive new degrees of conflict to overcome the limitations of κ. Indeed, Dempster's degree of conflict evaluates two pairs of mass functions as equally conflicting as long as they assign the same mass to $\emptyset$ (after their respective conjunctive combinations). Let $m_\cap$ denote the conjunctive combination of the first pair and $m'_\cap$ the combination of the second one. Suppose the focal elements of

$m_\cap$ are $\{\emptyset, A\}$ and those of $m'_\cap$ are $\{\emptyset, A, B\}$. If $A \cap B = \emptyset$, then $m'_\cap$ carries intuitively a higher level of inconsistency which κ fails to grasp.

The degrees of conflict introduced in [27, 23] are built using pairwise distances $d(m_1, m_2)$. There are however several arguments [11, 1] outlining that this practice is ill advised. However, in a recent work, Pichon and Jousselme [35] highlighted that non-pairwise distances can be instrumental in the construction of degrees of conflict. The authors examine the distance between the conjunctive combination $m_{1\cap2}$ and some reference mass function, i.e. the total conflict mass function $m_\emptyset$. Indeed we have

$$\kappa(m_1, m_2) = 1 - d_{pl,\infty}(m_{1\cap2}, m_\emptyset). \tag{38}$$

Similarly, K can be retrieved as the L_∞ norm based distances between the contour functions of m_1 and m_2. This observation raises the following question: can we build other relevant degrees of conflict in the same fashion as in (38) but using other distances than $d_{pl,\infty}$? We try to provide some answers to this question in the next paragraphs when the examined distances are consistent with $\bigcirc\!\!\!\!\cap$.

Proposition 5. *Let d denote a mass function distance which is either an L_k norm based distance between commonalities, plausibilities, or specialization matrices. Let $\mathcal{C} : \mathcal{M} \times \mathcal{M} \to [0; 1]$ denote the following mapping*

$$\mathcal{C}(m_1, m_2) = 1 - d(m_{1\cap2}, m_\emptyset). \tag{39}$$

Then $\mathcal{C}$ does not satisfy property (i) if k is finite.

See G for a proof.

From the above result, building a conflict degree using (39) using $d_{q,k}$, $d_{pl,k}$ or $d_{\mathrm{spe},k}$ is ill-advised whenever $k \neq \infty$. Intuitively, degrees of conflict relying on L_∞ are better candidates to verify (i) because the maximal norm value is not uniquely achieved for $m_{1\cap2} = m_\Omega$.

Proposition 6. *Let $\mathcal{C}_{q,\infty} : \mathcal{M} \times \mathcal{M} \to [0; 1]$ denote the following mapping*

$$\mathcal{C}_{q,\infty}(m_1, m_2) = 1 - d_{q,\infty}(m_{1\cap2}, m_\emptyset). \tag{40}$$

Then $\mathcal{C}_{q,\infty}$ coincides with the degree of strong conflict K.

Proposition 7. *Let $\mathcal{C}_{spe,\infty} : \mathcal{M} \times \mathcal{M} \to [0;1]$ denote the following mapping*

$$\mathcal{C}_{spe,\infty}(m_1, m_2) = 1 - d_{spe,\infty}(m_{1 \cap 2}, m_\emptyset). \tag{41}$$

Then $\mathcal{C}_{spe,\infty}$ coincides with Dempster's degree of conflict κ.

See H and I for proofs.

Proposition 6 shows that the degree of strong conflict is retrieved through the L_∞ norm based commonality distance while remark 7 shows that Dempster's degree of conflict is retrieved through the L_∞ norm based specialization distance which are complementary observations in line with [35]. We continue with another more general remark, i.e. outside the sole scope of a given family of mass function distances.

Proposition 8. *Let d denote a mass function distance which is consistent w.r.t. $\odot$. Let $\mathcal{C} : \mathcal{M} \times \mathcal{M} \to [0;1]$ denote the mapping defined from (39). Then $\mathcal{C}$ satisfies property (iii) w.r.t. the Dempsterian partial order $\sqsubseteq_d$.*

See J for a proof. Following proposition 8, it seems that in general, distances consistent w.r.t. $\odot$ are good candidates to possibly yield a relevant conflict degree.

6   Conclusion

In the scope of the theory of belief functions, this paper provides new results on the consistency of L_k norm based distances between commonalities and the L_∞ norm based distance between plausibilities with the conjunctive rule of combination. We also prove the consistency of L_k norm based distances between plausibilities and the L_∞ norm based distance between commonalities with the disjunctive rule of combination. The investigated form of consistency is equivalent to Lipschitz continuity of the mapping obtained by fixing one of the operand of pairwise combinations under these rules. Since the corresponding Lipschitz constant is 1, this property means that combining any pair of belief functions with any third party belief function is a non-expansive operation.

Outside the scope of belief functions, the results apply to random set distributions as belief functions can be interpreted as epistemic random sets. In this more general context, the conjunctive rule yields the distribution of the intersection of two independent random sets while the disjunctive

rule yields the distribution of the union of two independent random sets. Commonalities map any subset A to the probability that the random set is a superset of A (inclusion functional). Plausibilities map any subset A to the probability that the random set intersects A (hitting functional). Our results prove that if F maps the distribution of a random set to the distribution of the intersection of this random set with a given (fixed) independent one, then F is Lipschitz continuous with Lipschitz constant 1 w.r.t. L_k norm based distances between inclusion functionals or the L_∞ norm based distance between hitting functionals. Similarly, if F maps the distribution of a random set to the distribution of the union of this random set with a given (fixed) independent one, then F is Lipschitz continuous with Lipschitz constant 1 w.r.t. L_k norm based distances between hitting functionals or the L_∞ norm based distance between inclusion functionals.

We only investigate belief functions and random sets on finite spaces. Extending these results to uncountable spaces is an important perspective for future works. In the uncountable setting, random closed sets are defined as measurable mappings with respect to the Effros σ-algebra on the family of closed subsets of some locally compact Haussdorf completely separable topological space. The main results obtained in the finite case essentially rely on two aspects:

(i) intersection (resp. union) of independent random sets can be characterized by the elementwise multiplication of their inclusion (resp. containment) functionals,

(ii) a closed form expression of L_k norms for inclusion or hitting functionals.

Concerning the first aspect, the fact that $(A \supseteq C \text{ and } B \supseteq C) \Leftrightarrow A \cap B \supseteq C$ and $(A \subseteq C \text{ and } B \subseteq C) \Leftrightarrow A \cup B \subseteq C$ is intuitively sufficient to obtain equivalent relations in uncountable spaces. The relation for unions of independent random closed sets is indirectly evoked for containment functionals in [30, p.82]. The second aspects seems more challenging to generalize because one needs to introduce a norm for capacity functionals that is not just a vector norm. It may be possible to build such norms using Choquet integrals.

Another relevant research track for future works consists in investigating if the proposed Liptschitz continuity results hold as well when independence assumptions are not verified. Intuitively intersecting or uniting some pair of random sets with a third one that may or may not be dependent on either of them should still make their corresponding distributions closer (w.r.t. the

appropriate distance). The difficulty in this regard is that one can no longer just work with marginal inclusion or containment functionals but with multivariate ones [36] that do not factorize as elementwise products of marginal functionals. Recent work [37, 38, 39] generalizing Sklar's theorem on copulas to joint or multivariate capacity functionals of random sets may be useful in this quest because it gives an explicit connection between marginal functionals and multivariate ones. However, in contrast to point-valued random variables, a family of copulas is necessary to characterize this link.

Finally, going back to belief functions, we also discuss the advisability of building new degrees of conflict using distances that are consistent with the conjunctive rule. Such degrees are mainly interesting in information fusion applications of belief functions. We show that distances consistent with the conjunctive rule can be deemed to be relevant candidates for this purpose as they will achieve a desirable property for degrees of conflict. As for the distances for which we provide new consistency results (distance between commonalities or plausibilities), it turns out that they either violate another desirable property or coincide with an already known degree of conflict.

Acknowledgments

A   Set metrics and their consistency with set operations

This appendix is meant to show that some distances between sets verify either property (a) or (b), see subsection 3.3 for the definitions of these latter. We examine, the Hamming distance and the Jaccard distance:

- Hamming distance: for any subsets A, B and C, we have

$$d_{\mathrm{ham}}\left(A \cap C, B \cap C\right) = \frac{\left|\left(A \cap C\right) \Delta \left(B \cap C\right)\right|}{n} \tag{42}$$

$$= \frac{\left|\left(A \Delta B\right) \cap C\right|}{n} \tag{43}$$

$$\leq \frac{\left|A \Delta B\right|}{n}. \tag{44}$$

So we see that d_{ham} verifies (a).

We can also write

$$d_{\mathrm{ham}}\left(A \cup C, B \cup C\right) = \frac{\left|\left(A \cup C\right) \Delta \left(B \cup C\right)\right|}{n} \tag{45}$$

$$= \frac{\left|\left(A \cup B \cup C\right) \setminus \left(\left(A \cup C\right) \cap \left(B \cup C\right)\right)\right|}{n} \tag{46}$$

$$= \frac{\left|\left(A \cup B \cup C\right) \setminus \left(\left(A \cap B\right) \cup C\right)\right|}{n} \tag{47}$$

$$= \frac{\left|\left(\left(A \setminus C\right) \cup \left(B \setminus C\right)\right) \setminus \left(A \cap B\right)\right|}{n} \tag{48}$$

$$\leq \frac{\left|\left(A \cup B\right) \setminus \left(A \cap B\right)\right|}{n} \tag{49}$$

$$\leq \frac{\left|A \Delta B\right|}{n}, \tag{50}$$

and d_{ham} verifies (b) as well.

- Jaccard distance: for any subsets A, B and C, we have

$$d_{\mathrm{jac}}\left(A \cup C, B \cup C\right) = \frac{\left|\left(A \cup C\right) \Delta \left(B \cup C\right)\right|}{\left|\left(A \cup C\right) \cup \left(B \cup C\right)\right|} \tag{51}$$

$$= \frac{\left|A \Delta B\right| - \left|\left(A \Delta B\right) \setminus C\right|}{\left|A \cup B\right| + \left|C \setminus \left(A \cup B\right)\right|} \tag{52}$$

$$\leq \frac{\left|A \Delta B\right|}{\left|A \cup B\right|} \tag{53}$$

So we see that d_{jac} verifies (b).

To see that d_{jac} does not verify property (a), we provide the following counter-example.

26

Example 4. Let A and B denote two subsets that are not disjoint, i.e. $A \cap B \neq \emptyset$ and therefore $d_{\text{jac}}(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} < 1$.

If $C = A \Delta B$, then $A \cap C = A \setminus B$ and $B \cap C = B \setminus A$. This also implies that $(A \cap C) \Delta (B \cap C) = A \Delta B$ because $A \cap C$ and $B \cap C$ are disjoint. We also have $(A \cap C) \cup (B \cap C) = A \Delta B$ and consequently $d_{\text{jac}}(A \cap C, B \cap C) = 1$.

B   Proof of lemma 1

In this appendix, we give a proof that $b_{1|E}(A) = b_1((E \setminus A)^c)$ for any implicability function b_1 and any subset A and E.

Proof. By definition of the implicability function and conditioning, we have for any $A \subseteq \Omega$

$$b_{1|E}(A) = \sum_{B \subseteq A} \sum_{\substack{C \subseteq \Omega \\ \text{s.t.} \\ C \cap E = B}} m_1(C). \tag{54}$$

The second sum is empty if E is not a superset of B. If this condition is verified, we remark that subsets C are necessarily the union of B and some subset of E^c. This gives

$$b_{1|E}(A) = \sum_{B \subseteq A \cap E} \sum_{D \subseteq E^c} m_1(B \cup D). \tag{55}$$

Finally, any subset X of $(E \setminus A)^c$ can be partitioned w.r.t. $A \cap E$ and E^c, meaning that $\exists! Y \subseteq A \cap E$ and $\exists! Y' \subseteq E^c$ such that $X = Y \cup Y'$. Consequently, we have

$$b_{1|E}(A) = \sum_{X \subseteq (E \setminus A)^c} m_1(X) \tag{56}$$

$$= b_1((E \setminus A)^c) \tag{57}$$

$\square$

C Proof of proposition 1

In this appendix, we give a proof that for $1 \leq k \leq \infty$, $\textcircled{\cap}$ is 1-Lipschitz continuous w.r.t. the L_k norm based q-distance $d_{q,k}$.

Proof. Suppose m_1, m_2 and m_3 are three mass functions on Ω and $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ are their respective commonality vectors. For any positive finite integer k, we have:

$$
\begin{aligned}
[d_{q,k}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3)]^k &= [\|\mathbf{q}_{1 \cap 3} - \mathbf{q}_{2 \cap 3}\|_k]^k \\
&= \sum_{A \subseteq \Omega} |q_{1 \cap 3}(A) - q_{2 \cap 3}(A)|^k \\
&= \sum_{A \subseteq \Omega} |q_1(A) q_3(A) - q_2(A) q_3(A)|^k \\
&= \sum_{A \subseteq \Omega} |q_3(A)|^k |q_1(A) - q_2(A)|^k,
\end{aligned}
$$

For any subset $A \subseteq \Omega$, we have that $0 \leq q_3(A) \leq 1$ thus we obtain

$$
\begin{aligned}
[d_{q,k}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3)]^k &\leq \sum_{A \subseteq \Omega} |q_1(A) - q_2(A)|^k \\
&\leq [d_{q,k}(m_1, m_2)]^k.
\end{aligned}
$$

By definition, this latter inequality means that distance $d_{q,k}$ is consistent with rule $\textcircled{\cap}$.

If $k = \infty$, we have:

$$
\begin{aligned}
d_{q,\infty}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3) &= \|\mathbf{q}_{1 \cap 3} - \mathbf{q}_{2 \cap 3}\|_\infty \\
&= \max_{A \subseteq \Omega} |q_{1 \cap 3}(A) - q_{2 \cap 3}(A)| \\
&= \max_{A \subseteq \Omega} |q_1(A) q_3(A) - q_2(A) q_3(A)| \\
&= q_3(B) |q_1(B) - q_2(B)|,
\end{aligned}
$$

with $B = \arg\max_{A \subseteq \Omega} \{q_3(A) |q_1(A) - q_2(A)|\}$. It follows that

$$
\begin{aligned}
d_{q,\infty}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3) &\leq |q_1(B) - q_2(B)| \\
&\leq \max_{A \subseteq \Omega} |q_1(A) - q_2(A)| \\
&\leq d_{q,\infty}(m_1, m_2).
\end{aligned}
$$

By definition, this latter inequality means that $\textcircled{\cap}$ is 1-Lipschitz continuous w.r.t. $d_{q,\infty}$. $\qquad\square$

D   Proof of proposition 2

In this appendix, we give a proof that $\odot$ is 1-Lipschitz continuous w.r.t. the L_∞ norm based pl-distance $d_{pl,\infty}$.

Proof. Suppose m_1, m_2 and m_3 are three mass functions on Ω and $\mathbf{pl}_1$, $\mathbf{pl}_2$ and $\mathbf{pl}_3$ are their respective plausibility vectors. Let us first prove an intermediate result in case $m_3 = m_E$ is categorical. We can write

$$d_{pl,\infty}\left(m_{1|E}, m_{2|E}\right) = \max_{A \subseteq \Omega} |pl_{1|E}(A) - pl_{2|E}(A)|. \tag{58}$$

Using the fact that pl and b-distances coincide and lemma 1, we obtain

$$d_{pl,\infty}\left(m_{1|E}, m_{2|E}\right) = \max_{A \subseteq \Omega} |b_{1|E}(A) - b_{2|E}(A)| \tag{59}$$

$$= \max_{A \subseteq \Omega} |b_1\left((E \setminus A)^c\right) - b_2\left((E \setminus A)^c\right)| \tag{60}$$

$$\leq \max_{A \subseteq \Omega} |b_1(A) - b_2(A)|. \tag{61}$$

So the consistency condition is verified when m_3 is categorical. Now, let us examine the general case where m_3 is not necessarily categorical. Let $\mathbf{B}$ denote the transfer matrix [42] allowing to obtain vector forms of implicability functions by right-handed dot product with the vector form of their corresponding mass functions. We can write

$$d_{pl,\infty}(m_1 \odot m_3, m_2 \odot m_3) = \|\mathbf{pl}_{1 \cap 3} - \mathbf{pl}_{2 \cap 3}\|_\infty \tag{62}$$

$$= \|\mathbf{b}_{1 \cap 3} - \mathbf{b}_{2 \cap 3}\|_\infty \tag{63}$$

$$= \|\mathbf{B} \cdot (\mathbf{m}_{1 \cap 3} - \mathbf{m}_{2 \cap 3})\|_\infty \tag{64}$$

$$= \|\mathbf{B} \cdot (\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{m}_3\|_\infty. \tag{65}$$

One can always decompose a mass function as a convex combination of categorical ones: $\mathbf{m}_3 = \sum_{E \subseteq \Omega} m_3(E) \mathbf{m}_E$. We obtain

$$d_{pl,\infty}(m_1 \odot m_3, m_2 \odot m_3) = \left\| \sum_{E \subseteq \Omega} m_3(E) \mathbf{B} \cdot (\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{m}_E \right\|_\infty \tag{66}$$

$$\leq \sum_{E \subseteq \Omega} m_3(E) \|\mathbf{B} \cdot (\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{m}_E\|_\infty, \tag{67}$$

with the last inequality following from the triangle inequality and absolute homogeneity properties of the L_∞ norm.

From our intermediate result, we know that for any E, $\|\mathbf{B} \cdot (\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{m}_E\|_\infty = d_{pl,\infty}\left(m_{1|E}, m_{2|E}\right) \leq d_{pl,\infty}\left(m_1, m_2\right)$. Consequently, we have

$$d_{pl,\infty}\left(m_1 \textcircled{$\cap$} m_3, m_2 \textcircled{$\cap$} m_3\right) \leq d_{pl,\infty}\left(m_1, m_2\right) \sum_{E \subseteq \Omega} m_3\left(E\right) \tag{68}$$

$$\leq d_{pl,\infty}\left(m_1, m_2\right). \tag{69}$$

By definition, this latter inequality means that $\textcircled{$\cap$}$ is 1-Lipschitz continuous w.r.t. $d_{pl,\infty}$. $\qquad\square$

E   Proof of proposition 3

In this appendix, we give a proof that for $1 \leq k \leq \infty$, $\textcircled{$\circ$}$ is 1-Lipschitz continuous w.r.t. the L_k norm based pl-distance $d_{pl,k}$.

Proof. Suppose m_1, m_2 and m_3 are three mass functions on Ω and $\mathbf{pl}_1$, $\mathbf{pl}_2$ and $\mathbf{pl}_3$ are their respective plausibility vectors. For any positive finite integer k, we have:

$$
\begin{aligned}
\left[d_{pl,k}\left(m_1 \textcircled{$\circ$} m_3, m_2 \textcircled{$\circ$} m_3\right)\right]^k &= \left[\|\mathbf{pl}_{1 \cup 3} - \mathbf{pl}_{2 \cup 3}\|_k\right]^k \\
&= \sum_{A \subseteq \Omega} \left|pl_{1 \cup 3}\left(A\right) - pl_{2 \cup 3}\left(A\right)\right|^k \\
&= \sum_{A \subseteq \Omega} \left|b_{1 \cup 3}\left(A^c\right) - b_{2 \cup 3}\left(A^c\right)\right|^k \\
&= \sum_{A \subseteq \Omega} \left|b_1\left(A^c\right) b_3\left(A^c\right) - b_2\left(A^c\right) b_3\left(A^c\right)\right|^k \\
&= \sum_{A \subseteq \Omega} \left|b_3\left(A^c\right)\right|^k \left|b_1\left(A^c\right) - b_2\left(A^c\right)\right|^k.
\end{aligned}
$$

For any subset $A \subseteq \Omega$, we have that $0 \leq b_3\left(A^c\right) \leq 1$ and we obtain

$$
\begin{aligned}
\left[d_{pl,k}\left(m_1 \textcircled{$\circ$} m_3, m_2 \textcircled{$\circ$} m_3\right)\right]^k &\leq \sum_{A \subseteq \Omega} \left|b_1\left(A^c\right) - b_2\left(A^c\right)\right|^k \\
&\leq \sum_{A \subseteq \Omega} \left|pl_1\left(A\right) - pl_2\left(A\right)\right|^k \tag{70} \\
&\leq \left[d_{pl,k}\left(m_1, m_2\right)\right]^k.
\end{aligned}
$$

30

By definition, this latter inequality means that distance $d_{pl,k}$ is consistent with rule $\copyright$.

If $k = \infty$, we have:

$$
\begin{aligned}
d_{pl,\infty}\left(m_1 \copyright m_3, m_2 \copyright m_3\right) &= \left\| \mathbf{pl}_{1\cup3} - \mathbf{pl}_{2\cup3} \right\|_\infty, \\
&= \max_{A \subseteq \Omega} \left| pl_{1\cup3}\left(A\right) - pl_{2\cup3}\left(A\right) \right| \\
&= \max_{A \subseteq \Omega} \left| b_{1\cup3}\left(A^c\right) - b_{2\cup3}\left(A^c\right) \right| \\
&= \max_{A \subseteq \Omega} \left| b_1\left(A^c\right) b_3\left(A^c\right) - b_2\left(A^c\right) b_3\left(A^c\right) \right| \\
&= b_3\left(B^c\right) \left| b_1\left(B^c\right) - b_2\left(B^c\right) \right|,
\end{aligned}
$$

with $B = \arg\max_{A \subseteq \Omega} \left\{ b_3\left(A^c\right) \left| b_1\left(A^c\right) - b_2\left(A^c\right) \right| \right\}$. It follows that

$$
\begin{aligned}
d_{pl,\infty}\left(m_1 \copyright m_3, m_2 \copyright m_3\right) &\leq \left| b_1\left(B^c\right) - b_2\left(B^c\right) \right| \\
&\leq \left| pl_1\left(B\right) - pl_2\left(B\right) \right| \\
&\leq \max_{A \subseteq \Omega} \left| pl_1\left(A\right) - pl_2\left(A\right) \right| \\
&\leq d_{pl,\infty}\left(m_1, m_2\right).
\end{aligned}
$$

By definition, this latter inequality means that $\copyright$ is 1-Lipschitz continuous w.r.t. $d_{pl,\infty}$.

$\square$

F   Proof of proposition 4

In this appendix, we give a proof that $\copyright$ is 1-Lipschitz continuous w.r.t the L_∞ norm based q-distance $d_{q,\infty}$.

Proof. Suppose m_1, m_2 and m_3 are three mass functions on Ω and q_1, q_2 and q_3 are their respective commonality functions. Using relations (6) and (14), we can write

$$
d_{q,\infty}\left(m_{1\cup3}, m_{2\cup3}\right) = \max_{A \subseteq \Omega} \left| q_{1\cup3}\left(A\right) - q_{2\cup3}\left(A\right) \right| \tag{71}
$$

$$
= \max_{A \subseteq \Omega} \left| \overline{b}_{1\cup3}\left(A^c\right) - \overline{b}_{2\cup3}\left(A^c\right) \right| \tag{72}
$$

$$
= d_{b,\infty}\left(\overline{m_1 \copyright m_3}, \overline{m_2 \copyright m_3}\right) \tag{73}
$$

$$
= d_{pl,\infty}\left(\overline{m_1 \copyright m_3}, \overline{m_2 \copyright m_3}\right) \tag{74}
$$

$$
= d_{pl,\infty}\left(\overline{m}_1 \copyright \overline{m}_3, \overline{m}_2 \copyright \overline{m}_3\right). \tag{75}
$$

31

Since $d_{pl,\infty}$ is consistent w.r.t. $\cap$, we obtain

$$d_{q,\infty}\left(m_{1\cup3}, m_{2\cup3}\right) \leq d_{pl,\infty}\left(\overline{m}_1, \overline{m}_2\right) \tag{76}$$

$$\text{and } d_{pl,\infty}\left(\overline{m}_1, \overline{m}_2\right) = d_{b,\infty}\left(\overline{m}_1, \overline{m}_2\right) \tag{77}$$

$$= \max_{A \subseteq \Omega} |\overline{b}_1\left(A\right) - \overline{b}_2\left(A\right)| \tag{78}$$

$$= \max_{A \subseteq \Omega} |q_1\left(A^c\right) - q_2\left(A^c\right)| \tag{79}$$

$$= d_{q,\infty}\left(m_1, m_2\right). \tag{80}$$

By definition, this latter inequality means that $\copyright$ is 1-Lipschitz continuous w.r.t. $d_{q,\infty}$. $\qquad\blacksquare$

G   Proof of proposition 5

In this appendix, we provide a proof that the extreme conflict values property cannot be verified when a degree of conflict $\mathcal{C}$ is defined as

$$\mathcal{C}\left(m_1, m_2\right) = 1 - d\left(m_{1\cap2}, m_\emptyset\right)$$

where d is an L_k norm based distance between either commonality functions, plausibility functions or specialization matrices when k is finite.

Proof. We obviously have $\mathcal{C}\left(m_1, m_2\right) = 1$ iff $m_{1\cap2} = m_\emptyset$ and mass functions m_1 and m_2 are maximally conflicting so the problem does not come from this side of property (i).

Now, suppose $d = d_{q,k}$ and k is finite. Since $\mathcal{C}\left(m_1, m_2\right) = 0 \Leftrightarrow d_{q,k}\left(m_{1\cap2}, m_\emptyset\right) = 1$, we can write

$$d_{q,k}\left(m_{1\cap2}, m_\emptyset\right) = 1 \tag{81}$$

$$\Leftrightarrow \frac{1}{\rho}\left\|\mathbf{q}_{1\cap2} - \mathbf{q}_\emptyset\right\|_k = 1 \tag{82}$$

$$\Leftrightarrow \frac{1}{\rho^k}\sum_{A \subseteq \Omega}|q_{1\cap2}\left(A\right) - q_\emptyset\left(A\right)|^k = 1. \tag{83}$$

Remember that $\rho^k = N - 1$ for L_k norm based commonality distances. Moreover, $q_\emptyset\left(A\right) = 0$ when $A \neq \emptyset$ and that $q\left(\emptyset\right) = 1$ for any commonality function

therefore we obtain

$$d_{q,k}(m_{1\cap2}, m_\emptyset) = 1 \tag{84}$$

$$\Leftrightarrow \sum_{\substack{A\subseteq\Omega \\ A\neq\emptyset}} |q_{1\cap2}(A)|^k = N - 1. \tag{85}$$

Finally, the sum of commonalities (power k) cannot be equal to $N-1$ unless $q_{1\cap2}(A) = 1$ for each $A \neq \emptyset$. This means that $q_{1\cap2} = q_\Omega$ which implies that $q_1 = q_2 = q_\Omega$. This latter condition is not an admissible definition of non-conflict. The proof for distances between plausibilities is extremely similar and is thus omitted.

Concerning, distances between specialization matrices, the philosophy is also similar but we provide a sketch of the proof. When dealing with an L_k norm based specialization distance, we have $\rho^k = 2(N-1)$ and this distance is achieved for the pair $(m_\Omega, m_\emptyset)$. To see that m_Ω is the only mass function achieving maximal distance with $m_\emptyset$, one just needs to observe that the matrix entry $S_\emptyset(A, B) = 1$ if $A = \emptyset$ and $S_\emptyset(A, B) = 0$ otherwise. We can write

$$[d_{\text{spe},k}(m_{1\cap2}, m_\emptyset)]^k = \sum_{E\subseteq\Omega} \left(\left\|\mathbf{m}_{1\cap2|E} - \mathbf{m}_\emptyset\right\|_k\right)^k. \tag{86}$$

The only way to maximize the above expression is to maximize each $\left\|\mathbf{m}_{1\cap2|E} - \mathbf{m}_\emptyset\right\|_k$ individually for $E \neq \emptyset$. We know that $\left\|\mathbf{m}_{1\cap2|E} - \mathbf{m}_\emptyset\right\|_k \leq 2$ and we need $m_{1\cap2|E}(\emptyset) = 0$ to achieve this maximal value. This is not possible unless $m_{1\cap2} = m_\Omega$. Again, $m_{1\cap2} = m_\Omega$ implies that $m_1 = m_2 = m_\Omega$. $\square$

H   Proof of proposition 6

In this appendix, we provide a proof that the degree of conflict $\mathcal{C}_{q,\infty}$ defined as

$$\mathcal{C}_{q,\infty}(m_1, m_2) = 1 - d_{q,\infty}(m_{1\cap2}, m_\emptyset)$$

coincides with the degree of strong conflict K.

Proof. By definition of $\mathcal{C}_{q,\infty}$, we can write

$$\mathcal{C}_{q,\infty}(m_1, m_2) = 1 - \max_{A\subseteq\Omega} |q_{1\cap2}(A) - q_\emptyset(A)|. \tag{87}$$

Since $q_\emptyset (A) = 0$ when $A \neq \emptyset$ and that $q (\emptyset) = 1$ for any commonality function, we obtain

$$\mathcal{C}_{q,\infty} (m_1, m_2) = 1 - \max_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} |q_{1 \cap 2} (A)|. \tag{88}$$

Any commonality function is such that $q (A) \geq q (A')$ if $A \subseteq A'$, which gives

$$\mathcal{C}_{q,\infty} (m_1, m_2) = 1 - \max_{\substack{a \in \Omega \\ A \neq \emptyset}} |q_{1 \cap 2} (\{a\})|. \tag{89}$$

Finally commonality and plausibility functions coincide on singletons, hence $\mathcal{C}_{q,\infty} (m_1, m_2) = K (m_1, m_2)$. □

I Proof of proposition 7

In this appendix, we provide a proof that the degree of conflict $\mathcal{C}_{\mathrm{spe},\infty}$ defined as

$$\mathcal{C}_{\mathrm{spe},\infty} (m_1, m_2) = 1 - d_{\mathrm{spe},\infty} (m_{1 \cap 2}, m_\emptyset)$$

coincides with Dempster's degree of conflict κ.

Proof. By definition, we have

$$d_{\mathrm{spe},\infty} (m_{1 \cap 2}, m_\emptyset) = \max_{A,B \subseteq \Omega} |S_{1 \cap 2} (A, B) - S_\emptyset (A, B)| \tag{90}$$

$$= \max_{A \subseteq \Omega} \left\| \mathbf{m}_{1 \cap 2 | A} - \mathbf{m}_\emptyset \right\|_\infty. \tag{91}$$

For any $A \subseteq \Omega$, we have

$$\left\| \mathbf{m}_{1 \cap 2 | A} - \mathbf{m}_\emptyset \right\|_\infty = \max \left\{ 1 - m_{1 \cap 2 | A} (\emptyset) ; \max_{\substack{E \subseteq \Omega \\ E \neq \emptyset}} m_{1 \cap 2 | A} (E) \right\}. \tag{92}$$

For any $E \neq \emptyset$, we have

$$m_{1 \cap 2 | A} (E) \leq \sum_{\substack{E' \subseteq \Omega \\ E' \neq \emptyset}} m_{1 \cap 2 | A} (E') \tag{93}$$

$$\leq 1 - m_{1 \cap 2 | A} (\emptyset). \tag{94}$$

We deduce that $\left\| \mathbf{m}_{1 \cap 2 | A} - \mathbf{m}_\emptyset \right\|_\infty = 1 - m_{1 \cap 2 | A} (\emptyset)$. Finally, Dempster's degree of conflict can only grow as one performs a conjunctive combination therefore $\max_{A \subseteq \Omega} 1 - m_{1 \cap 2 | A} (\emptyset) = 1 - m_{1 \cap 2} (\emptyset)$, hence $\mathcal{C}_{\mathrm{spe},\infty} (m_1, m_2) = \kappa (m_1, m_2)$. □

34

J Proof of proposition 8

In this appendix, we give a proof that for some degree of conflict $\mathcal{C}$ defined as

$$\mathcal{C}(m_1, m_2) = 1 - d(m_{1 \cap 2}, m_\emptyset),$$

where d is a mass function distance consistent with $\cap$, then $\mathcal{C}$ satisfies property (iii) (from [11]) w.r.t. the Dempsterian partial order $\sqsubseteq_d$.

Proof. By definition of the Dempsterian partial order, $m_1 \sqsubseteq_d m_1'$ means that there exists a mass function m such that $m_1 = m_1' \cap m$. If d is consistent w.r.t. $\cap$, then for any mass function m_2, we have

$$d(m_1' \cap m_2 \cap m, m_\emptyset \cap m) \leq d(m_1' \cap m_2, m_\emptyset) \tag{95}$$

$$\Leftrightarrow d(m_1 \cap m_2, m_\emptyset) \leq d(m_1' \cap m_2, m_\emptyset) \tag{96}$$

$$\Leftrightarrow \mathcal{C}(m_1, m_2) \geq \mathcal{C}(m_1', m_2). \tag{97}$$

$\square$

K Consistency of distances with α-junctions

In [26], two families of mass function distances are introduced. Each of them relies on a given type of evidential matrix and a matrix norm. The evidential matrices in question are either an α-specialization matrix or an α-generalization matrix. These matrices are a generalization of the specialization matrix and generalization matrix, in the sense that these two matrices are retrieved by setting $\alpha = 1$. The definition of these more general matrices stems from a class of combination rules known as α-junctions [41].

In short, α-junctions are linear combination rules that do not depend on the order in which mass functions are combined. This axiomatic justification of these properties is detailed in [41]. These rules also have a meta-data dependent interpretation. These meta-data characterize the truthfulness of the sources that induced the mass functions. A source is truthful if it conveys the pieces of information it possesses and it is untruthful if it conveys inconsistent pieces of information as compared to the ones it possesses. For example, suppose a source has inferred that $\{\theta \in A\}$. If it is truthful it conveys the mass function m_A while it conveys m_{A^c} if it is untruthful. The α-junctions allow to combine mass functions in several situations ranging between these two extreme cases. This interpretation is documented in [34, 33, 26, 21].

Concerning evidential matrices, the most important point for the present discussion is that for each $\alpha \in [0; 1]$, there is a bijective correspondence between a given mass function m_i and an α-specialization matrix $\mathbf{D}_i^{(\alpha)}$. Another bijective correspondence also exists between a given mass function m_i and an α-generalization matrix $\mathbf{G}_i^{(\alpha)}$. The main result in [26] states that, for any $\alpha \in [0; 1]$, the distance induced by the L_1 matrix norm of the difference between a pair of α-specialization matrices is consistent with the α-conjunctive rule. Likewise, for any $\alpha \in [0; 1]$, the distance induced by the L_1 matrix norm of the difference between a pair of α-generalization matrices is consistent with the α-disjunctive rule.

Evidential matrices are not the only representation of states of beliefs induced by α-junctions. One can also define α-commonality functions [41] $q_i^{(\alpha)}$. Let $\mathbf{Q}^{(\alpha)}$ denote the matrix obtained by n Kronecker product as

$$\mathbf{Q}^{(\alpha)} = \mathrm{Kron}\left(\begin{bmatrix} 1 & 1 \\ \alpha - 1 & 1 \end{bmatrix}, \ldots, \mathrm{Kron}\left(\begin{bmatrix} 1 & 1 \\ \alpha - 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ \alpha - 1 & 1 \end{bmatrix} \right) \right). \quad (98)$$

The vector form of function $q_i^{(\alpha)}$ is obtained as

$$\mathbf{q}_i^{(\alpha)} = \mathbf{Q}^{(\alpha)} \cdot \mathbf{m}_i, \quad (99)$$

with $\mathbf{m}_i$ the vector form of some mass function m_i. There is a bijective correspondence between α-commonality functions and mass functions and the α-commonality function in correspondence with the result of an α-conjunctive combination is equal to the entrywise product of the α-commonality functions in correspondence with the combined mass functions. Using this property, the same reasoning as in the proof of proposition 1 applies and any L_k norm based distance between α-commonality functions is consistent with the corresponding α-conjunctive rule. For the proof to hold, one also needs that $|q_i^{(\alpha)}(B)| \leq 1$ for any $B \subseteq \Omega$. Looking at equation (98), we actually have

$$\alpha - 1 \leq q_i^{(\alpha)}(B) \leq 1, \forall B \subseteq \Omega. \quad (100)$$

Similarly, Smets [41] also introduces α-implicability functions $b_i^{(\alpha)}$. Let $\mathbf{B}^{(\alpha)}$ denote the matrix obtained by n Kronecker product as

$$\mathbf{B}^{(\alpha)} = \mathrm{Kron}\left(\begin{bmatrix} 1 & 1 \\ 1 & \alpha - 1 \end{bmatrix}, \ldots, \mathrm{Kron}\left(\begin{bmatrix} 1 & 1 \\ 1 & \alpha - 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & \alpha - 1 \end{bmatrix} \right) \right). \quad (101)$$

The vector form of function $b_i^{(\alpha)}$ is obtained as

$$\mathbf{b}_i^{(\alpha)} = \mathbf{B}^{(\alpha)} \cdot \mathbf{m}_i, \tag{102}$$

with $\mathbf{m}_i$ the vector form of some mass function m_i. There is a bijective correspondence between α-implicability functions and mass functions and the α-implicability function in correspondence with the result of an α-disjunctive combination is equal to the entrywise product of the α-implicability functions in correspondence with the combined mass functions. Using this property, the same reasoning as in the proof of proposition 3 applies and any L_k norm based distance between α-implicability functions is consistent with the corresponding α-disjunctive rule. For the proof to hold, one also needs that $|b_i^{(\alpha)}(B)| \leq 1$ for any $B \subseteq \Omega$. Again, from equation (101), we actually have

$$\alpha - 1 \leq b_i^{(\alpha)}(B) \leq 1, \forall B \subseteq \Omega. \tag{103}$$

Computational difficulties are found when n increases for computing α-commonality and α-implicability functions but this is also true for computing α-specialization or α-generalization matrices, therefore all the distances evoked in this appendix section are on an equal footing from both theoretical and practical considerations.

References

[1] T. Burger. Geometric views on conflicting mass functions: From distances to angles. *International Journal of Approximate Reasoning*, 70:36–50, 2016.

[2] T. Burger, S. Destercke, et al. How to randomly generate mass functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pages 645–673, 2013.

[3] I. Couso and D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502 – 1518, 2014. Special issue: Harnessing the information contained in low-quality data sources.

[4] I. Couso, D. Dubois, and L. Sánchez. Random sets and random fuzzy sets as ill-perceived random variables. In *SpringerBriefs in Computational Intelligence*. Springer, 2014.

[5] F. Cuzzolin. A geometric approach to the theory of evidence. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 38(4):522–534, July 2008.

[6] F. Cuzzolin. Geometric conditioning of belief functions. In *Proceedings of BELIEF 2010, International workshop on the theory of belief functions*, pages 1–6, Brest, France, 2010.

[7] F. Cuzzolin. *Visions of a generalized probability theory*. Lambert Academic Publishing, London, UK, 2014.

[8] A. P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Satistics*, 38(2):325–339, 1967.

[9] T. Denux. Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence*, 172(2-3):234–264, february 2008.

[10] T. Denux. Likelihood-based belief function: Justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535 – 1547, 2014. Special issue: Harnessing the information contained in low-quality data sources.

[11] S. Destercke and T. Burger. Toward an axiomatic definition of conflict between belief functions. *IEEE transactions on cybernetics*, 43(2):585–596, 2013.

[12] J. Diaz, M. Rifqi, and B. Bouchon-Meunier. A similarity measure between basic belief assignments. In *Int. Conf. on Information Fusion (FUSION'06)*, pages 1–6, Florence (Italy), july 2006.

[13] D. Dubois, W. Liu, J. Ma, and H. Prade. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. *Information Fusion*, 32:12–39, 2016.

[14] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximatons by fuzzy sets. *Int. Journal of General Systems*, 12(3):193–226, 1986.

[15] M. Fréchet. *Sur quelques points du calcul fonctionnel*. PhD thesis, Faculté des Sciences de Paris, 1906.

[16] P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272, 1901.

[17] A.-L. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001.

[18] A.-L. Jousselme and P. Maupin. Distances in evidence theory: Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53(2):118 – 145, 2012. Theory of Belief Functions (BELIEF 2010).

[19] R. Kennes and P. Smets. Computational aspects of the mobius transform. *arXiv preprint arXiv:1304.1122*, 2013.

[20] J. Klein, S. Destercke, and O. Colot. Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation. *International Journal of Approximate Reasoning*, 71:15 – 33, 2016.

[21] J. Klein, M. Loudahi, J.-M. Vannobel, and O. Colot. α-junctions of categorical mass functions. In F. Cuzzolin, editor, *Belief Functions: Theory and Applications*, volume 8764 of *Lecture Notes in Computer Science*, pages 1–10. Springer International Publishing, 2014.

[22] M.-J. Lesot, M. Rifqi, and H. Benhadda. Similarity measures for binary and numerical data: a survey. *International Journal of Knowledge Engineering and Soft Data Paradigms*, 1(1):63–84, 2008.

[23] W. Liu. Analyzing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909 – 924, 2006.

[24] M. Loudahi, J. Klein, J.-M. Vannobel, and O. Colot. Fast computation of l_p norm-based specialization distances between bodies of evidence. In F. Cuzzolin, editor, *Belief Functions: Theory and Applications*, volume 8764 of *Lecture Notes in Computer Science*, pages 422–431. Springer International Publishing, 2014.

[25] M. Loudahi, J. Klein, J.-M. Vannobel, and O. Colot. 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, 2014.

[26] M. Loudahi, J. Klein, J.-M. Vannobel, and O. Colot. Evidential matrix metrics as distances between meta-data dependent bodies of evidence. *Cybernetics, IEEE Transactions on*, 46(1):109–122, Jan 2016.

[27] A. Martin, A.-L. Jouselme, and C. Osswald. Conflict measure for the discounting operation on belief functions. In *Proceedings of the Eleventh International Conference on Information Fusion*, pages 1–8, 2008.

[28] G. Mathéron. *Random sets and integral geometry*. Wiley New York, 1975.

[29] E. Miranda, I. Couso, and P. Gil. Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications*, 307(1):32–47, 2005.

[30] I. S. Molchanov. *Theory of random sets*, volume 19. Springer, 2005.

[31] H. T. Nguyen. An introduction to random sets. 2006. *Chapman&Hall/CRC, Boca Raton, FL*.

[32] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65(3):531 – 542, 1978.

[33] F. Pichon. On the α-conjunctions for combining belief functions. In T. Denoeux and M.-H. Masson, editors, *Belief Functions: Theory and Applications*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 285–292. Springer Berlin Heidelberg, 2012.

[34] F. Pichon and T. Denoeux. 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., 2009.

[35] F. Pichon, A.-L. Jousselme, and N. B. Abdallah. Several shades of conflict. *Fuzzy Sets and Systems*, 366:63 – 84, 2019.

[36] B. Schmelzer. Characterizing joint distributions of random sets by multivariate capacities. *International Journal of Approximate Reasoning*, 53(8):1228–1247, 2012.

[37] B. Schmelzer. Joint distributions of random sets and their relation to copulas. *International Journal of Approximate Reasoning*, 65:59–69, 2015.

[38] B. Schmelzer. Sklar's theorem for minitive belief functions. *International Journal of Approximate Reasoning*, 63:48–61, 2015.

[39] B. Schmelzer. Multivariate capacity functionals vs. capacity functionals on product spaces. *Fuzzy Sets and Systems*, 364:1–35, 2019.

[40] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University press, Princeton (NJ), USA, 1976.

[41] P. Smets. The alpha-junctions: Combination operators applicable to belief functions. In *First Int. Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQUARU-FAPR'97)*, volume 1244 of *Lecture Notes in Computer Sciences*, pages 131–153. Springer International Publishing, Bad Honef, Germany, 1997.

[42] P. Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(12):1 – 30, 2002.

[43] P. Smets. Belief functions on real numbers. *International Journal of Approximate Reasoning*, 40:181–223, 2005.

[44] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.

[45] B. Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315 – 329, 1993.

[46] R. R. Yager. The entailment principle for dempstershafer granules. *International Journal of Intelligent Systems*, 1(4):247–262, 1986.

[47] L. M. Zouhal and T. Denœux. 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, 1998.