



HAL
open science

Cross-Entropy and Relative Entropy of Basic Belief Assignments

Jean Dezert, Frédéric Dambreville

► **To cite this version:**

Jean Dezert, Frédéric Dambreville. Cross-Entropy and Relative Entropy of Basic Belief Assignments. International Conference on Information Fusion (FUSION 2023), Jun 2023, Charleston, United States. pp.1-8, 10.23919/FUSION52260.2023.10224207 . hal-04628203

HAL Id: hal-04628203

<https://hal.science/hal-04628203v1>

Submitted on 28 Jun 2024

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

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

Cross-Entropy and Relative Entropy of Basic Belief Assignments

Jean Dezert, Frédéric Dambreville
Department of Information Processing and Systems
The French Aerospace Lab - ONERA
Palaiseau, France.
jean.dezert@onera.fr, frederic.dambreville@onera.fr

Abstract—This paper introduces the concept of cross-entropy and relative entropy of two basic belief assignments. It is based on the new entropy measure presented recently. We prove that the cross-entropy satisfies a generalized Gibbs-alike inequality from which a generalized Kullback-Leibler divergence measure can be established in the framework of belief functions. We show on a simple illustrating example how these concepts can be used for decision-making under uncertainty.

Keywords: generalized entropy, cross-entropy, relative entropy, Kullback-Leibler divergence, belief functions

I. INTRODUCTION

In Shannon's theory of communication developed in 1948 [1], [2], the measure of uncertainty (MoU), also called entropy, for characterizing a source of information (from signal transmission standpoint) is defined by Shannon entropy. This entropy measures the randomness of a probability distribution P and is usually noted by $H(P)$. Shannon entropy does not concern the semantic aspects of the content of a message but only its transmission [3]–[5]. $H(P)$ has played a very important role in the development of modern communication systems and cryptography [6] until today. According to Cover and Thomas [7], the cross-entropy denoted by $H(P, Q)$ is the average number of bits needed to encode data coming from a source with a probability distribution P when we use a distribution model Q to define our codebook. Cross-entropy is commonly used in machine learning as a loss function [8], and the cross-entropy method is often used in practice to estimate an unknown true pmf (probability mass function) based on a test set where Q is the assumed (or eventually empirical) pmf model. The minimization of the cross-entropy is related with the principle of the maximization of the likelihood. That is why cross-entropy plays a major role in many statistical applications. The relative entropy, often referred as Kullback-Leibler divergence [9], is the difference between the cross-entropy and Shannon entropy, and so it is $H(P, Q) - H(P)$. All these aforementioned basic concepts have been well established (and strongly justified) from the mid of 20th century, and all use the theory of probability as the fundamental underlying mathematical framework.

In this paper we go beyond the classical probabilistic framework because we want to work possibly with epistemic uncertainty represented by non-probabilistic models thanks to the mathematical framework of belief functions introduced by

Shafer [10], and in this context the legitimate and important question is to know if it is possible, or not, to extend the concepts of entropy, cross-entropy and relative entropy for the belief functions. Concerning the concept of entropy, the answer is affirmative and very recently a new generalized entropy measure has been proposed in [11] in the framework of the theory of belief functions. Concerning the second and third theoretical questions about cross-entropy and relative entropy concepts, we give new comprehensive and better answers to these questions in this paper. This is our new theoretical contribution in the field. The concrete meaning of relative entropy and cross-entropy measures in the belief functions framework is a challenging question because the entropy of belief function is merely related to the uncertainty of epistemic knowledge rather than of statistical knowledge. No concrete meaning of these notions has been firmly established so far. This interesting open question is left for future research works.

To make the material of this paper quite self-contained, we recall the basic classical concepts related to entropy (Shannon entropy, cross-entropy, and relative entropy) in the section II, and we present the basics of belief functions [10] in Section III with the new concept of entropy measure of basic belief mass assignment (BBA) [11] in the section IV. After recalling a very recent definition of cross-entropy of BBAs [12] based on the non effective Deng's entropy definition [13], we present in the section V a new cross-entropy definition based on our new effective entropy definition. The section VI presents the concept of relative entropy of BBAs which can be interpreted as a generalization of the Kullback-Leibler divergence measure for belief functions. An example of the use of these concepts for decision-making under uncertainty is given in the section VII. Concluding remarks and perspectives are given in the section VIII.

II. CLASSICAL NOTIONS RELATED TO ENTROPY

A. Shannon entropy

Consider a discrete random variable θ represented by a probability mass function (pmf) $P_N = (p_1, p_2, \dots, p_N)$, where $p_i = P(\theta_i)$ is the probability of the i -th state θ_i (i.e. outcome) of $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$. Shannon was interested in communication systems where the various events were the carriers of coded messages, and he did justify his entropy measure as appropriate measure of average uncertainty (or

measure of randomness) of a random variable [1], [2], [6], [7]. The entropy of a random variable is the average level of *surprisal*, or *uncertainty* inherent in the variable's possible outcomes [14]. Shannon entropy is defined by¹

$$H(P_N) \triangleq - \sum_{i=1}^{|\Theta|} P(\theta_i) \log(P(\theta_i)). \quad (1)$$

By convention, $P(\theta_i) \log(P(\theta_i)) = 0$ if $P(\theta_i) = 0$. This is easily justified by continuity because $\lim_{x \rightarrow 0^+} x \log x = 0$, which can be proved using L'Hôpital rule [15]. Adding terms of zero probability does not change the entropy value. In (1) we use the natural logarithm (i.e. base e logarithm) and in this case the Shannon entropy value is expressed in *nats* unity. We can also use the base 2 logarithm (\log_2) function instead of the natural logarithm, and if so the Shannon entropy value will be expressed in *bits*. Shannon entropy can be interpreted as a generalization of Hartley entropy (1928) [16] when presuming the pmf of equally probable states (i.e. uniform² pmf P_N^{unif}), hence getting $H(P_N^{\text{unif}}) = \log(|\Theta|) = \log(N)$. Note that if we have a uniform pmf P_N^{unif} defined on Θ with $|\Theta| = N$ and another uniform pmf $P_{N'}^{\text{unif}}$ defined on Θ' with $|\Theta'| = N'$, and if $|\Theta| < |\Theta'|$ then $H(P_N^{\text{unif}}) < H(P_{N'}^{\text{unif}})$ because $\log(|\Theta|) < \log(|\Theta'|)$ since $\log(x)$ is an increasing function. The minimum value of Shannon entropy is zero, which characterizes a *non-random* (or *sure*) event θ_j for which $P(\theta_j) = 1$.

The main algebraic properties of Shannon entropy are, see [17] (p. 30) for details: the symmetry, the normality³, expansibility, decisivity, sub-additivity and recursivity. We recall that Shannon entropy value $H(P_N)$ is always smaller than $H(P_N^{\text{unif}})$ if $P_N \neq P_N^{\text{unif}}$, expressing the fact that the uniform pmf is the only pmf giving the maximal Shannon entropy value, and characterizing the maximum of uncertainty (or randomness), which is called the maximality property.

B. The cross-entropy

Consider a finite set of exhaustive events $\Theta = \{\theta_1, \dots, \theta_n\}$ where θ_i are mutually exclusive (i.e. $\theta_i \cap \theta_j = \emptyset$ if $i \neq j$). Suppose that $P = \{P(\theta_1) = p_1, \dots, P(\theta_n) = p_n\}$ is a probability distribution over the set Θ . Then for any other probability distribution $Q = \{Q(\theta_1) = q_1, \dots, Q(\theta_n) = q_n\}$ the Gibbs inequality holds [18]

$$- \sum_{i=1}^n p_i \log(q_i) \geq - \sum_{i=1}^n p_i \log(p_i). \quad (2)$$

The cross-entropy between probability distributions P and Q over the same underlying set of events Θ is defined by

$$H(P, Q) = - \sum_{i=1}^n p_i \log(q_i) = - \sum_{X \in \Theta} P(X) \log(Q(X)). \quad (3)$$

One can easily verify that $H(P, Q) = H(P)$ when $Q = P$, i.e. when the probability distribution Q coincides with the

true probability distribution P the cross-entropy value equals Shannon entropy of P . Gibbs inequality is $H(P, Q) \geq H(P)$.

C. The relative entropy

The difference between the cross-entropy $H(P, Q)$ and Shannon entropy $H(P)$ is named the relative entropy or the Kullback-Leibler (KL) divergence [9]), and is often denoted by⁴

$$D_{\text{KL}}(P \parallel Q) \triangleq H(P, Q) - H(P) = \sum_{i=1}^n p_i \log(p_i/q_i). \quad (4)$$

$D_{\text{KL}}(P \parallel Q)$ measures how the probability distribution P is different from a second, reference probability distribution Q . It corresponds to the expectation of the logarithmic difference between the probability distributions P and Q , where the expectation is taken using the distribution P . In general the relative entropy $D_{\text{KL}}(P \parallel Q)$ is not symmetric under interchange of the distributions P and Q and we have $D_{\text{KL}}(P \parallel Q) \neq D_{\text{KL}}(Q \parallel P)$. Therefore, D_{KL} is not strictly a distance even if it is often abusively called a *distance* in the literature, even by Cover in [7]. This relative entropy (i.e. divergence measure) is important in pattern recognition and neural networks for making classification, as well as in information theory. Kullback and Leibler also proposed a symmetrized measure in [9] defined as $D_{\text{KL}}(P \parallel Q) + D_{\text{KL}}(Q \parallel P)$. Another renown symmetric version of the KL divergence is the Jensen-Shannon (JS) divergence defined by Lin in [19]

$$D_{\text{JS}}(P \parallel Q) \triangleq \frac{1}{2} D_{\text{KL}} \left(P \parallel \frac{P+Q}{2} \right) + \frac{1}{2} D_{\text{KL}} \left(Q \parallel \frac{P+Q}{2} \right). \quad (5)$$

The Jensen-Shannon divergence can be interpreted as the total Kullback-Leibler divergence to the average probability distribution $(P+Q)/2$. This JS divergence is often used in practice because its square root is a metric often referred to as Jensen-Shannon distance [20], that is

$$d_{\text{JS}}(P, Q) \triangleq \sqrt{D_{\text{JS}}(P \parallel Q)}. \quad (6)$$

Jensen-Shannon divergence has been applied in different fields of applications (e.g. bioinformatics, social sciences, fire experiments, machine learning, in deep learning for studying generative adversarial networks, etc), see [21].

III. BELIEF FUNCTIONS

The belief functions (BF) were introduced by Shafer [10] for modeling epistemic uncertainty, reasoning about uncertainty and combining distinct sources of evidence. The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD) $\Theta = \{\theta_1, \dots, \theta_N\}$ where all elements (i.e. members) of Θ are exhaustive and exclusive. The set of all subsets of Θ (including empty set

¹The symbol \triangleq means *equal by definition*.

²for which $P(\theta_i) = 1/N$ for $i = 1, 2, \dots, N$.

³This stipulates that $H(P_2^{\text{unif}}) = 1$ using base 2 logarithm function in (1).

⁴As in [7] (p. 19), in the formula (4) we use the conventions that $0 \log(0/0) = 0$, $0 \log(0/q) = 0$, and $p \log(p/0) = \infty$. So, if there is any $X \in \Theta$ such that $P(X) > 0$ and $Q(X) = 0$, $D_{\text{KL}}(P \parallel Q) = \infty$.

\emptyset , and Θ) is the power-set of Θ denoted by 2^Θ . The number of elements (i.e. the cardinality) of the power-set is $2^{|\Theta|}$. A (normalized) basic belief assignment (BBA) associated with a given source of evidence is a mapping $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$ such that⁵ $m^\Theta(\emptyset) = 0$ and $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$. A BBA $m^\Theta(\cdot)$ characterizes a source of evidence related with a FoD Θ . For notation shorthand, we can omit the superscript Θ in $m^\Theta(\cdot)$ notation if there is no ambiguity on the FoD we work with. The quantity $m(X)$ is called the mass of belief of X . $X \in 2^\Theta$ is called a focal element (FE) of $m(\cdot)$ if $m(X) > 0$. The set of all focal elements of $m(\cdot)$ is denoted by $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$. The belief and the plausibility of X are respectively defined for any $X \in 2^\Theta$ by [10]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y), \quad (7)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}), \quad (8)$$

where $\bar{X} \triangleq \Theta \setminus \{X\}$ is the complement of X in Θ .

One has always $0 \leq Bel(X) \leq Pl(X) \leq 1$, see [10]. For $X = \emptyset$, $Bel(\emptyset) = Pl(\emptyset) = 0$, and for $X = \Theta$ one has $Bel(\Theta) = Pl(\Theta) = 1$. $Bel(X)$ and $Pl(X)$ are often interpreted as the lower and upper bounds of unknown probability $P(X)$ of X , that is $Bel(X) \leq P(X) \leq Pl(X)$. To quantify the uncertainty (i.e. the imprecision) of $P(X) \in [Bel(X), Pl(X)]$, we use $u(X) \in [0, 1]$ defined by

$$u(X) \triangleq Pl(X) - Bel(X). \quad (9)$$

The quantity $u(X) = 0$ if $Bel(X) = Pl(X)$ which means that $P(X)$ is known precisely, and one has $P(X) = Bel(X) = Pl(X)$. One has $u(\emptyset) = 0$ because $Bel(\emptyset) = Pl(\emptyset) = 0$, and one has $u(\Theta) = 0$ because $Bel(\Theta) = Pl(\Theta) = 1$. If all focal elements of $m(\cdot)$ are singletons of 2^Θ the BBA $m(\cdot)$ is a Bayesian BBA because $\forall X \in 2^\Theta$ one has $Bel(X) = Pl(X) = P(X)$ and $u(X) = 0$. Hence the belief and plausibility of X coincide with a probability measure $P(X)$ defined on the FoD Θ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by $m_v(X) = 1$ for $X = \Theta$, and $m_v(X) = 0$ for all $X \in 2^\Theta$ different of Θ . This very particular BBA plays a major role in the establishment of a new effective measure of uncertainty for BBA.

IV. ENTROPY OF BASIC BELIEF ASSIGNMENTS

In [22] we did analyze in details forty-five measures of uncertainty (MoU) of BBAs by covering 40 years of research works on this topic. Some of these MoUs capture only a particular aspect of the uncertainty inherent to a BBA (typically, the non-specificity and the conflict). Other MoUs propose a total uncertainty measure to capture jointly several aspects of the uncertainty. Unfortunately, most of these MoUs fail to satisfy four very simple reasonable and essential desiderata, and so they cannot be considered as really effective and

useful. Actually only five MoUs can be considered as effective from the mathematical sense presented next, but unfortunately they appear as conceptually defective and disputable, see discussions in [22]. That is why, a better effective measure of uncertainty (MoU), i.e. generalized entropy of BBAs has been developed and presented in [11]. The mathematical definition of this new effective entropy is given by

$$U(m) = \sum_{X \in 2^\Theta} s(X) \quad (10)$$

with

$$s(X) \triangleq -m(X)(1 - u(X)) \log(m(X)) + u(X)(1 - m(X)). \quad (11)$$

$s(X)$ is the uncertainty contribution related to X named the *entropiece* of X . This entropiece $s(X)$ involves $m(X)$ and the imprecision $u(X) = Pl(X) - Bel(X)$ about the unknown probability of X in a subtle interwoven manner. Because $u(X) \in [0, 1]$ and $m(X) \in [0, 1]$ one has $s(X) \geq 0$, and $U(m) \geq 0$. The quantity $U(m)$ is expressed in *nats* because we use the natural logarithm. $U(m)$ can be expressed in *bits* by dividing the $U(m)$ value in *nats* by $\log(2) = 0.69314718\dots$. This measure of uncertainty $U(m)$ is a continuous function in its basic belief mass arguments because it is a summation of continuous functions. In formula (11), we always take $m(X) \log(m(X)) = 0$ when $m(X) = 0$ because $\lim_{m(X) \rightarrow 0^+} m(X) \log(m(X)) = 0$. Note that for any BBA m , one has $s(\emptyset) = 0$ because $m(\emptyset) = 0$ and $u(\emptyset) = 0$. For the vacuous BBA, one has $s(\Theta) = 0$ because $m_v(\Theta) = 1$ and $u(\Theta) = 0$.

This measure of uncertainty $U(m)$ is effective because it can be proved (see proofs in [11]) that it satisfies the following four essential properties:

- 1) $U(m) = 0$ for any BBA $m(\cdot)$ focused on a singleton X of 2^Θ .
- 2) $U(m_v^\Theta) < U(m_v^{\Theta'})$ if $|\Theta| < |\Theta'|$.
- 3) $U(m) = -\sum_{X \in \Theta} m(X) \log(m(X))$ if the BBA $m(\cdot)$ is a Bayesian BBA. Hence, $U(m)$ reduces to Shannon entropy [1] in this case.
- 4) $U(m) < U(m_v)$ for any non-vacuous BBA $m(\cdot)$ and for the vacuous BBA $m_v(\cdot)$ defined with respect to the same FoD.

The maximum of entropy value is obtained for the vacuous BBA m_v over a FoD Θ , because m_v characterizes a source of evidence with a full lack of information. This maximum entropy value is $U(m_v^\Theta) = 2^{|\Theta|} - 2$ (see derivation in [11]) and it represents the sum of all imprecisions of $P(X)$ for all $X \in 2^\Theta$. Because for all $X \in 2^\Theta \setminus \{\emptyset, \Theta\}$ one has $u(X) = 1$ because $[Bel(X), Pl(X)] = [0, 1]$, and one has $u(\emptyset) = 0$ and $u(\Theta) = 0$ when considering the vacuous BBA then the sum of all imprecisions $u(X)$ about $P(X)$ is equal to $2^{|\Theta|} - 2$. It is worth mentioning that one has always $U(m_v^\Theta) > \log(|\Theta|)$ which means that the vacuous BBA has always an entropy greater than the maximum of Shannon entropy $\log(|\Theta|)$ obtained with the uniform pmf on Θ .

⁵In Shafer's theory of BFs we work with a closed FoD and the mass of the empty set must always be equal to zero.

V. CROSS-ENTROPY OF TWO BBAS

A. Cross-entropy derived from Deng's entropy

Very recently in [12], Gao et al. proposed a definition of the cross-entropy of two BBAs inspired by the non-effective Deng's entropy $E_d(m)$ proposed earlier by Deng in [13] and defined as follows:

$$E_d(m) = - \sum_{X \subseteq \Theta} m(X) \log\left(\frac{m(X)}{2^{|X|} - 1}\right) \quad (12)$$

where $m(X)$ is the mass of belief of any subset X of the frame of discernment Θ , and where $|X|$ is the cardinality of X . If $m(X) = 0$, the term $m(X) \log\left(\frac{m(X)}{2^{|X|} - 1}\right)$ is set to zero.

Deng's entropy definition is unfortunately not recommended because it is non-effective. Indeed, we can have $E_d(m) > E_d(m_v)$ indicating that a non-vacuous BBA $m(\cdot)$ can be more uncertain than the vacuous BBA $m_v(\cdot)$, which obviously is not appropriate because the vacuous BBA characterizes the state of total ignorance. As a simple counterexample of Deng's entropy consider $\Theta = \{A, B, C\}$ the vacuous BBA $m_v(\cdot)$ with $m_v(A \cup B \cup C) = 1$, and the non-vacuous BBA $m(\cdot)$ with $m(A \cup B) = m(A \cup C) = m(B \cup C) = 1/3$. Clearly, one gets $E_d(m) > E_d(m_v)$. See the paper [22] for more discussions about other non-effective entropy proposals. For this counterexample, the values of Deng's entropies are

$$\begin{aligned} E_d(m_v) &= -m_v(A \cup B \cup C) \log\left(\frac{m_v(A \cup B \cup C)}{2^{|A \cup B \cup C|} - 1}\right) \\ &= -1 \cdot \log\left(\frac{1}{2^3 - 1}\right) = -\log\left(\frac{1}{7}\right) \approx 1.9459, \end{aligned}$$

$$\begin{aligned} E_d(m) &= -m(A \cup B) \log\left(\frac{m(A \cup B)}{2^{|A \cup B|} - 1}\right) \\ &\quad - m(A \cup C) \log\left(\frac{m(A \cup C)}{2^{|A \cup C|} - 1}\right) \\ &\quad - m(B \cup C) \log\left(\frac{m(B \cup C)}{2^{|B \cup C|} - 1}\right) \\ &= -3 \cdot \frac{1}{3} \cdot \log\left(\frac{1}{3} \cdot \frac{1}{2^2 - 1}\right) = -\log(1/9) \approx 2.1972. \end{aligned}$$

Based on this non-effective entropy measure, the cross-entropy defined by Gao et al. [12] between BBAs m_1 and m_2 is based on a mimicry of the classical cross-entropy definition using Deng's entropy, that is

$$C(m_1, m_2) = - \sum_{X \subseteq \Theta} m_1(X) \log\left(\frac{m_2(X)}{2^{|X|} - 1}\right).$$

Similarly, the cross-entropy between m_2 and m_1 is

$$C(m_2, m_1) = - \sum_{X \subseteq \Theta} m_2(X) \log\left(\frac{m_1(X)}{2^{|X|} - 1}\right).$$

Because Deng's entropy is non effective, we have serious doubt on the validity of the cross-entropy concept defined by $C(m_1, m_2)$ and $C(m_2, m_1)$ formulas. This matter of fact justifies the necessity of using a better entropy measure [11] defined by (10)–(11), and the development of a better cross-entropy measure. This is what we present in the next section.

B. A new definition of cross-entropy

Based on the definition (3) of cross-entropy in the probabilistic framework, and the definition of the effective generalized entropy $U(m)$ given in (10), it seems quite natural to try to extend directly the concept of cross-entropy of two pdfs p and q to the cross-entropy of two BBAs m_1 and m_2 defined over the same FoD Θ . The extension of the classical cross-entropy formula (3) applied with generalized entropy $U(m)$ given in (10) suggests directly the following generic formula of the cross-entropy between two BBAs

$$U(m_1, m_2) = \sum_{X \subseteq \Theta} s_{1,2}(X) \quad (13)$$

with

$$\begin{aligned} s_{1,2}(X) \triangleq & -m_1(X)(1 - u_i(X)) \log(m_2(X)) \\ & + u_j(X)(1 - m_k(X)) \end{aligned} \quad (14)$$

where indexes i, j and k have to belong to the set $\{1, 2\}$.

From this generic formulation, one sees that we could a priori define eight different cross-entropies between two BBAs depending on the choice of indexes (i, j, k) listed in Table I.

Table I
POSSIBLE TRIPLETS (i, j, k) .

Triplet $T = (i, j, k)$	Value
T_1	(1,1,1)
T_2	(1,1,2)
T_3	(1,2,1)
T_4	(1,2,2)
T_5	(2,1,1)
T_6	(2,1,2)
T_7	(2,2,1)
T_8	(2,2,2)

It is worth mentioning that if $m_2 = m_1$ the cross-entropy measure coincides with the entropy measure, that is $U(m_1, m_2) = U(m_1, m_1) = U(m_1)$.

What is the best definition of the cross-entropy of two BBAs among the eight possible definitions? Or equivalently, what is the most suitable triplet of indexes (i, j, k) to plug in the generic cross-entropy formula (14)? To answer to this important question, we propose to consider as the effective choice of triplet (i, j, k) the one which allows the information entropy of a BBA m_1 to be less than or equal to its cross-entropy with any other BBA m_2 . More precisely, select the triplet (i, j, k) such that for any BBAs m_1 and m_2 defined on the same FoD, the following inequality holds

$$U(m_1, m_2) \geq U(m_1). \quad (15)$$

Actually for the eight a priori possible definitions of cross-entropy drawn from (13)–(14), one can easily find by Monte-Carlo simulations of random pairs (m_1, m_2) of BBAs that the choices of triplets $(1, 1, 1)$, $(1, 2, 1)$, $(1, 2, 2)$, $(2, 1, 1)$, $(2, 1, 2)$, $(2, 2, 1)$, and $(2, 2, 2)$ are not judicious because the inequality (15) can be violated, see some examples in the appendix. Because our Monte-Carlo analysis based on 100000

random pairs (m_1, m_2) of BBAs revealed that the inequality (15) was satisfied only for the triplet $(i, j, k) = (1, 1, 2)$ for different cardinalities of frames of discernment tested up to $|\Theta| = 10$, we did conjecture that the satisfactory definition of a cross-entropy of two BBAs satisfying inequality (15) is mathematically defined by (13) with

$$s_{1,2}(X) \triangleq -m_1(X)(1 - u_1(X)) \log(m_2(X)) + u_1(X)(1 - m_2(X)). \quad (16)$$

The term $s_{1,2}(X)$ defined in (16) is called the *cross-entropiece* of X .

Theorem 1: Let m_1 and m_2 be BBAs defined on the same frame of discernment. The cross-entropy $U(m_1, m_2)$ defined by (13) and (16) always satisfies the inequality $U(m_1, m_2) \geq U(m_1)$, with equality only if $m_1 = m_2$.

Proof: see appendix.

Proposition: If the BBAs m_1 and m_2 are Bayesian the cross-entropy defined by (13) and (16) coincides with the classical cross-entropy given by (3).

Proof: Since $u_1(X) = 0$ for all $X \in 2^\Theta$ for any Bayesian BBA $m_1(\cdot)$, the proposition is immediate.

VI. RELATIVE ENTROPY OF TWO BBAS

It is worth mentioning that the inequality (15) is a generalization of the well-known Gibbs inequality (2), and it coincides with Gibbs inequality when the BBAs m_1 and m_2 are Bayesian BBAs. The generalized relative entropy (GRE) of two BBAs m_1 and m_2 that are defined over the same frame of discernment Θ is naturally defined by

$$U(m_1 \parallel m_2) \triangleq U(m_1, m_2) - U(m_1). \quad (17)$$

Because Theorem 1 holds, one has always $U(m_1 \parallel m_2) \geq 0$, with equality if m_1 equals m_2 . As for the classical relative entropy defined by (4), the GRE is not symmetric under the interchange of the BBAs m_1 and m_2 , so that in general $U(m_1 \parallel m_2) \neq U(m_2 \parallel m_1)$. Therefore GRE must also not be considered as a distance. This GRE is a direct generalization of Kullback-Leibler (KL) divergence measure in the framework of belief functions. Using expressions (16) and⁶ (11) the mathematical definition of $U(m_1 \parallel m_2)$ is

$$U(m_1 \parallel m_2) = \sum_{X \subseteq \Theta} [m_1(X)(1 - u_1(X)) \cdot (\log(m_1(X)) - \log(m_2(X))) + u_1(X)(m_1(X) - m_2(X))]. \quad (18)$$

GRE coincides with *KL*-divergence formula (4) when the BBAs m_1 and m_2 are Bayesian because if focal elements of m_1 and m_2 are singletons of 2^Θ then $u_1(X) = 0$ and

⁶with m replaced by m_1 .

$$\begin{aligned} U(m_1 \parallel m_2) &= \sum_{X \subseteq \Theta} m_1(X)(\log(m_1(X)) - \log(m_2(X))) \\ &= \sum_{X \subseteq \Theta} m_1(X) \log\left(\frac{m_1(X)}{m_2(X)}\right) \end{aligned} \quad (19)$$

which is equivalent to formula (4) when interpreting the Bayesian BBA m_1 as a probability measure p , and the Bayesian BBA m_2 as a probability measure q over the set Θ .

VII. EXAMPLE OF APPLICATION

In this section we present an example of the use of the entropy, cross-entropy and relative entropy concepts defined in this paper for the purpose of decision-making under uncertainty. More precisely, given a BBA $m(\cdot)$ defined over a FoD Θ , how to make a decision based on $m(\cdot)$ and how to select the most pertinent element θ_i of Θ ?

A. Decision using relative entropy

Classically the decision-making from a BBA is based on the max of $Pl(\cdot)$, on the max of $Bel(\cdot)$, or on the max of pignistic probability depending on the attitude chosen by the decision-maker (resp. optimistic, pessimistic or in-between attitudes). Here we propose to make the decision based on the relative entropy measure. More precisely, from any BBA m defined over a FoD $\Theta = \{\theta_i, i = 1, \dots, n\}$, we calculate the divergences $U(m_i \parallel m)$ for $i = 1, 2, \dots, n$, where m_i is the BBA focused on the element $\theta_i \in \Theta$ such that $m_i(\theta_i) = 1$. We will take as decision $\hat{\theta}$ the element θ_i for which the divergence between m and m_i is minimal, that is $\hat{\theta} = \theta_{i^*}$ with $i^* = \arg \min_{i \in \{1, 2, \dots, n\}} U(m_i \parallel m)$.

Example: Consider the FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$, and after some fusion processing suppose we obtain the following BBA $m(\cdot)$ defined by $m(\theta_1) = 0.1$, $m(\theta_2) = 0.2$, $m(\theta_3) = 0.3$, $m(\theta_1 \cup \theta_2) = 0.01$, $m(\theta_1 \cup \theta_3) = 0.02$, $m(\theta_2 \cup \theta_3) = 0.07$ and $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.3$. Then we get $U(m_1 \parallel m) \approx 2.30$, $U(m_2 \parallel m) \approx 1.60$ and $U(m_3 \parallel m) \approx 1.20$. Based on this result the decision will be $\hat{\theta} = \theta_3$ because the divergence $U(m_3 \parallel m) = 1.20$ is the least value among the values 2.30, 1.60 and 1.20. This decision is consistent with what we intuitively expect because $[Bel(\theta_1), Pl(\theta_1)] = [0.10, 0.43]$, $[Bel(\theta_2), Pl(\theta_2)] = [0.20, 0.58]$ and $[Bel(\theta_3), Pl(\theta_3)] = [0.30, 0.69]$ showing that θ_3 is the element of Θ that has the maximum of belief and also the maximum of plausibility.

Remark 1: We could not use $U(m \parallel m_i)$ instead of $U(m_i \parallel m)$. Indeed, we get $U(m \parallel m_i) = +\infty$, and thus cannot decide. But the use of $U(m_i \parallel m)$ is however not completely satisfactory because for m_i we have $u_i(X) = 0$ for all $X \in 2^\Theta$ and $U(m_i) = 0$, so that $U(m_i \parallel m) = U(m_i, m) - U(m_i) = -\log(m(\theta_i))$. Thus, the decision is made with only part of the information of m about θ_i and not with the other mass values of non-singleton focal elements of m (if any). Subsequently, a pseudo-distance inspired by Jensen-Shannon is proposed which uses the whole BBA information.

B. Decision using Jensen-Shannon pseudo-distance

In [23] we did propose a decision-making method based on the minimum of belief-interval distance that used Wasserstein distance. We take for decision $\hat{\theta}$ the element θ_i for which the distance $d(m, m_i)$ between m and m_i is minimal, that is $\hat{\theta} = \theta_{i^*}$ with $i^* = \arg \min_{i \in \{1, 2, \dots, n\}} d(m, m_i)$. This method implicitly assumes the uniform distribution of the probability $P(X)$ in $[Bel(X), Pl(X)]$ which is disputable because we cannot check in practice if this assumption is true, or not. To circumvent this problem, we propose to replace the belief-interval distance between BBAs by the Jensen-Shannon-like pseudo-distance derived from our relative entropy concept, which would be defined by

$$d(m, m') \triangleq \sqrt{\frac{1}{2} [U(m \parallel \frac{m+m'}{2}) + U(m' \parallel \frac{m+m'}{2})]}. \quad (20)$$

Note that $d(m, m')$ coincides with Jensen-Shannon distance (6) when the BBAs m and m' are bayesian BBAs. One has also $d(m, m') = d(m', m)$, $d(m, m') \geq 0$ and $d(m, m') = 0$ when $m = m'$ because $U(m \parallel \frac{m+m}{2}) = U(m \parallel m)$, and $U(m \parallel m) = U(m, m) - U(m) = 0$.

In our example, we obtain the following pseudo-distances: $d(m, m_1) \approx 0.67$, $d(m, m_2) \approx 0.60$, and $d(m, m_3) \approx 0.55$. Based on these values we will take the decision $\hat{\theta} = \theta_3$.

Note also that if m is the vacuous BBA (i.e. $m = m_v$), then in this particular case we will obtain $d(m_v, m_1) = d(m_v, m_2) = d(m_v, m_3) = 0.6656$ so that no clear decision can be drawn from the vacuous BBA since it does not contain useful information, which makes perfect sense. Note that the inequality $(d(m, m_1) = 0.6763) > (d(m_v, m_1) = 0.6656)$ is not surprising because the BBA m is more unfavorable to θ_1 than the vacuous BBA m_v is.

Remark 2: We tested (20) against the triangular inequality $d(m, m') + d(m', m'') \geq d(m, m'')$. A crude Monte Carlo analysis based on millions of random BBAs generated uniformly over different frames of discernment up to cardinality $|\Theta| = 13$ revealed no counterexample. This indicates that such counterexamples are rare events. However, we tried a refined Monte Carlo analysis, where the set of focal elements were generated prior to the BBA. On the basis of 10000 different generated BBAs and near $500 \cdot 10^9$ combination cases, we have found a rate of $2 \cdot 10^{-5}$ counterexamples to the triangular inequality. This is quite small. More interestingly, the degree of violation of the triangular inequality was small, since we found 1.17 as the maximum value for $\frac{d(m, m'')}{d(m, m') + d(m', m'')}$. That is why we consider $d(m, m')$ only as a pseudo-distance, i.e. a semimetric. But our simulations suggest that this semimetric satisfies a sharp ρ -relaxed triangle inequality:

$$d(m, m'') \leq \rho(d(m, m') + d(m', m'')) \quad \text{with } \rho \geq 1.2.$$

In conclusion, the topology induced by this semimetric is certainly very close to a true metric topology.

Counterexample of triangular inequality:

Consider $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and the three BBAs m , m' and m'' as follows:

$$\begin{aligned} m(\theta_3) &= 0.25, m(\theta_1 \cup \theta_3) = 0.19, m(\theta_2 \cup \theta_3) = 0.21, m(\Theta) = 0.35, \\ m'(\theta_1 \cup \theta_3) &= 0.25, m'(\theta_2 \cup \theta_3) = 0.26, m'(\Theta) = 0.49, \\ m''(\theta_1 \cup \theta_2) &= 0.44, m''(\Theta) = 0.56. \end{aligned}$$

We get $d(m, m') \approx 0.1144$, $d(m', m'') \approx 0.1800$ and $d(m, m'') \approx 0.3306$. Hence $d(m, m') + d(m', m'') = 0.2945$ which is smaller than $d(m, m'') = 0.3306$. So there, the triangular inequality $d(m, m') + d(m', m'') \geq d(m, m'')$ is violated.

VIII. CONCLUSION

In this paper we have proposed new measures of cross-entropy and relative entropy of two basic belief assignments based on the new effective measure of entropy of belief function presented in 2022. These new concepts are mathematically well-defined and are direct generalizations of their classical formulations drawn of the probabilistic framework. It is expected that these new theoretical concepts will become useful in some applications for decision-making under uncertainty. As research perspectives, we hope to improve them a bit more in order to provide a true Jensen-Shannon metric for belief functions in a near future. Also, applications of these new concepts are under development and they will be reported in future publications.

APPENDIX

A. Counterexamples of inequality (15)

We consider the FoD $\Theta = \{A, B, C\}$ and we give BBAs⁷ $m_1(\cdot)$ and $m_2(\cdot)$ such that inequality (15) is violated for the different choices of triplet (i, j, k) used in the formula (14).

- Consider $(i, j, k) = (1, 1, 1)$ and the BBAs of Table II. We get $U(m_1) = 3.9742$ and $U(m_1, m_2) = 3.9432$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.
- Consider $(i, j, k) = (1, 2, 1)$ and the BBAs of Table III. We get $U(m_1) = 3.7447$ and $U(m_1, m_2) = 2.4995$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.
- Consider $(i, j, k) = (1, 2, 2)$ and the BBAs of Table IV. We get $U(m_1) = 3.9568$ and $U(m_1, m_2) = 2.5086$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.
- Consider $(i, j, k) = (2, 1, 1)$ and the BBAs of Table V. We get $U(m_1) = 3.2115$ and $U(m_1, m_2) = 2.8616$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.
- Consider $(i, j, k) = (2, 1, 2)$ and the BBAs of Table VI. We get $U(m_1) = 2.5542$ and $U(m_1, m_2) = 2.2147$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.
- Consider $(i, j, k) = (2, 2, 1)$ and the BBAs of Table VII. We get $U(m_1) = 4.5243$ and $U(m_1, m_2) = 3.8714$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.

⁷The numerical values entering in the tables have been approximated to their fourth decimal for convenience.

- Consider $(i, j, k) = (2, 2, 2)$ and the BBAs of Table VIII. We get $U(m_1) = 3.8858$ and $U(m_1, m_2) = 3.0406$. The inequality (15) is violated because $U(m_1) > U(m_1, m_2)$.

Table II
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.2094	0.1199
B	0.0537	0.0885
$A \cup B$	0.3016	0.2833
C	0.0054	0.0112
$A \cup C$	0.0713	0.0539
$B \cup C$	0.0712	0.0646
$A \cup B \cup C$	0.2874	0.3786

Table III
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1443	0.4612
B	0.0695	0.0657
$A \cup B$	0.0128	0.0291
C	0.0903	0.2119
$A \cup C$	0.2922	0.1056
$B \cup C$	0.2324	0.0305
$A \cup B \cup C$	0.1585	0.0960

Table IV
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1585	0.2677
B	0.0180	0.2017
$A \cup B$	0.2202	0.0566
C	0.0758	0.2432
$A \cup C$	0.1396	0.0120
$B \cup C$	0.1681	0.1051
$A \cup B \cup C$	0.2198	0.1137

Table V
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1718	0.0852
B	0.1045	0.0246
$A \cup B$	0.1721	0.0551
C	0.1329	0.0598
$A \cup C$	0.1721	0.2664
$B \cup C$	0.2078	0.2391
$A \cup B \cup C$	0.0388	0.2698

B. Proof of the Theorem 1

Subsequently, log is the natural logarithm function to the base of the mathematical Euler constant e . To prove the Theorem 1, we first prove the theorem 2 below.

Theorem 2: Let $\mathcal{M}(\Theta)$ be the set of basic belief assignments over Θ . Then:

$$\arg \min_{m \in \mathcal{M}(\Theta)} U(m_1, m) = \{m_1\}.$$

Table VI
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.2228	0.0284
B	0.2112	0.0617
$A \cup B$	0.0767	0.3397
C	0.2523	0.0428
$A \cup C$	0.0726	0.1748
$B \cup C$	0.1196	0.2164
$A \cup B \cup C$	0.0448	0.1362

Table VII
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.0443	0.1234
B	0.0892	0.2140
$A \cup B$	0.2486	0.1321
C	0.0473	0.0962
$A \cup C$	0.0562	0.2928
$B \cup C$	0.0803	0.0245
$A \cup B \cup C$	0.4341	0.1170

Table VIII
BBAS $m_1(\cdot)$ AND $m_2(\cdot)$.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.0164	0.2007
B	0.0196	0.0967
$A \cup B$	0.2219	0.1484
C	0.0746	0.2518
$A \cup C$	0.1052	0.0996
$B \cup C$	0.4906	0.1538
$A \cup B \cup C$	0.0717	0.0490

Proof: Let $F_1 \subset 2^\Theta \setminus \{\emptyset\}$ be the set of focal elements of m_1 . First at all, it is noticed that $u_1(X) < 1$ for all $X \in F_1$. Moreover, if there are $X \in F_1$ and $m \in \mathcal{M}(\Theta)$ such that $m(X) = 0$, then $U(m_1, m) = +\infty$. As a consequence, if m minimizes $U(m_1, m)$, then its set of focal elements contains the set of focal elements of m_1 .

Optimizations. Let F be such that $F_1 \subset F \subset 2^\Theta \setminus \{\emptyset\}$. The proof is done by solving:

$$\min_{m: F \rightarrow \mathbb{R}_+^*} f(m) \quad (21)$$

under constraint

$$\sum_{X \in F} m(X) = 1 \quad (22)$$

where

$$f(m) = U(m_1, m) = \sum_{\emptyset \neq X \neq \Theta} u_1(X)(1 - m(X)) + \sum_{X \in F_1} -m_1(X)(1 - u_1(X)) \log(m(X)). \quad (23)$$

It is worth noting that $m(X)$ is nothing but the component of index X of the unknown map vector $m : F \rightarrow \mathbb{R}_+^*$. The optimization (21) with equality constraint (22) could typically be solved by means of Lagrangian multiplier method.

Because $\log(m(X))$ is a concave function of $m(X)$, the term $-m_1(X)(1 - u_1(X)) \log(m(X))$ is proportional to $-\log(m(X))$ and is a convex function of $m(X)$. And because $u_1(X)(1 - m(X))$ is a linear function of $m(X)$, the term $-m_1(X)(1 - u_1(X)) \log(m(X)) + u_1(X)(1 - m(X))$ is a convex function of $m(X)$. Therefore, the function $f(m)$ is a convex function. We are then ensured that Lagrangian multiplier condition will point, if it is fulfilled, to the minima of the function.

Lagrangian multiplier is defined for this problem by:

$$L(m, \lambda) \triangleq \sum_{X \in F_1} -m_1(X)(1 - u_1(X)) \log(m(X)) + \sum_{\emptyset \neq X \neq \Theta} u_1(X)(1 - m(X)) + \lambda[1 - \sum_{X \in F} m(X)]. \quad (24)$$

The optimality conditions are:

$$D_{m(X)}L(m, \lambda) = 0 \text{ for all } X \in F.$$

Where $D_{m(X)}L(m, \lambda)$ is the differential of $L(m, \lambda)$ with respect to $m(X)$ given by

$$D_{m(X)}L(m, \lambda) = \frac{-m_1(X)(1 - u_1(X))}{m(X)} - u_1(X) - \lambda,$$

for $X \in F_1$, and:

$$D_{m(X)}L(m, \lambda) = -u_1(X) - \lambda, \text{ for } X \in F \setminus F_1.$$

Then, the optimal solution for (21) is m^{opt} such that:

$$m^{\text{opt}}(X) = \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))}, \text{ for all } X \in F_1, \quad (25)$$

with λ chosen such that:

$$-\lambda = u_1(X) \text{ for all } X \in F \setminus F_1, \quad (26)$$

$$\sum_{X \in F \setminus F_1} m^{\text{opt}}(X) + \sum_{X \in F_1} \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))} = 1. \quad (27)$$

Noticed that (25) implies $-\lambda - u_1(X) > 0$ for all $X \in F_1$.

Case $F \neq F_1$: Condition (26) implies $-\lambda \leq 1$ and then:

$$\sum_{X \in F_1} \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))} \geq \sum_{X \in F_1} m_1(X) = 1.$$

Then by (27), it comes $m^{\text{opt}}(X) = 0$ for $X \in F \setminus F_1$, which contradicts hypothesis that F is the set of focal elements of m . *There is no solution with more focal elements than m_1 .*

Case $F = F_1$: Choice $\lambda = -1$ is obvious. Therefore, *the*

unique minimizer $m^{\text{opt}} = m_1$ is obtained.

Conclusion. It has been shown that minimizer of $U(m_1, m)$ only exists if it has the same set of focal elements than m_1 . Moreover, it is shown in that case that the only minimizer is m_1 . As a consequence:

$$\arg \min_{m \in \mathcal{M}(\Theta)} U(m_1, m) = \{m_1\} \text{ for all } m \in \mathcal{M}(\Theta).$$

Because Theorem 2 holds, we have $U(m_1, m) > U(m_1)$ when $m \neq m_1$, and $U(m_1, m_1) = U(m_1)$ when $m = m_1$. Therefore, $U(m_1, m) \geq U(m_1)$ for any BBA $m \in \mathcal{M}(\Theta)$. Thus the inequality (15) holds, with equality only if $m_1 = m_2$, which completes the proof of the Theorem 1.

REFERENCES

- [1] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423 & 623-656, July & October 1948 (reprinted in [6]).
- [2] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1-117, 1949.
- [3] M. Higashi, G.J. Klir, *Measures of uncertainty and information based on possibility distributions*, Int. J. Gen. Systems, Vol. 9, pp. 43-58, 1983.
- [4] O. Lombardi, F. Holik, L. Vanni, *What is Shannon information?*, Synthese 193, pp. 1983-2012, 2016.
- [5] O. Rioul, *This is IT: A Primer on Shannon's Entropy and Information*, L'Information, Séminaire Poincaré XXIII, 43-77, 2018.
- [6] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [7] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, New Jersey, Second Edition, 2006.
- [8] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- [9] S. Kullback, R.A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, Vol. 22, No. 1, pp. 79-86, 1951.
- [10] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [11] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, Fusion 2022 Conf., Linköping, Sweden, pp. 1-10, July, 2022.
- [12] X. Gao, L. Pan, Y. Deng, *Cross entropy of mass function and its application in similarity measure*, Applied Intelligence, Vol. 52, pp. 8337-8350, 2022.
- [13] Y. Deng, *Deng entropy*, Chaos, Solitons & Fractals, Vol. 91, pp. 549-553, Oct. 2016.
- [14] [https://en.wikipedia.org/wiki/Entropy_\(information_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))
- [15] R.E. Bradley, S.J. Petrilli, C.E. Sandifer, *L'Hôpital's analyse des infiniments petits (An annotated translation with source material by Johann Bernoulli)*, Birkhäuser, 311 pages, 2015.
- [16] R. Hartley, *Transmission of information*, Bell System Tech. J., Vol. 7, pp. 535-563, 1928.
- [17] J. Aczél, Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, 1975.
- [18] D.J.C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.
- [19] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. on Information Theory, Vol. 37, No. 1, pp. 145-151, 1991.
- [20] D.M. Endres, J.E. Schindelin, *A new metric for probability distributions*, IEEE Trans. Inf. Theory, Vol. 49, No. 7, pp. 1858-1860, July 2003.
- [21] https://en.wikipedia.org/wiki/Jensen-Shannon_divergence
- [22] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, Feb. 2022.
- [23] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21-23, 2016.