Defining new approximations of belief functions by means of Dempster's combination

Thomas Burger
Université Européenne de Bretagne, Université de Bretagne-Sud, CNRS, Lab-STICC
Centre de Recherche Yves Coppens, BP 573, F-56017 Vannes cedex, France
Email: thomas.burger@univ-ubs.fr

***Abstract*—Approximating a belief function (with a probability distribution or with another belief function with a restricted number of focal elements) is an important issue in Dempster-Shafer Theory. The reason is that such approximations are really useful in two different situations: (1) decision making and (2) computational saving. In this paper, we propose to consider the definition of a proxy for a belief function as the result of the Dempster's combination of two belief functions: The first one is the belief function to approximate and the second one is a Bayesian belief function which encodes a meta-information describing the support of the approximation (i.e. the set of the potential focal elements of the proxy).**

Keywords: belief function theory, Dempster-Shafer Theory, Bayesian approximation, k-additive belief functions, hidden Dempster's combination

I. Introduction

Dempster-Shafer Theory [1] (as well as several of its extensions, such as [2]) is a very popular tool to handle imprecision and uncertainty in knowledge discovery. Nonetheless, two major drawbacks have prevented its early and complete spreading into other communities with similar scientific goals, such as statistic, data mining, Bayesian machine learning, etc.

The first one is the computational burden [3]: In the formalism of belief functions, it is necessary to deal with a distribution defined on a powerset rather than a set. Then, the computational cost grows in an exponential manner with respect to the size of the state-space. Hence, more compact representations are sought, i.e. by means of belief functions with a restricted number of focal elements.

The second drawback is the lack of intuitive significance for a belief function with several focal elements of different cardinality. As explained in [4], [5], it is not trivial for an expert to capture the very meaning of a each of these focal elements. Consequently, decision making in such a context is not intuitive. On the other hand, there are many representations which are easier to handle (such as probabilities or necessity measures, etc.), and a popular solution is to convert the belief function into another belief function with a more intuitive (and often more compact) set of definition.

Finally, in spite of their differences, these two problems have led to the same question: How can a belief function be approximated with another more compact belief function? For obvious reasons, these approximations are often Bayesian belief functions, but other kinds of proxies are also interesting.

In this paper, the investigation of such proxies is pushed further in a particular direction, inspired by an article of R. Haenni [6]. He shows that several popular operators of Dempster-Shafer Theory can be expressed in terms of a Dempster's combination with a special belief function. He considers several operators: Discounting, disjunctive combination, refinement, etc. but there is no study on approximating operators. They are considered in this paper.

Section 2 gives a the state-of-the-art containing a brief inventory of the Bayesian/non-Bayesian proxies in the literature and a summary of the framework of [6]. Then, this framework is used in Section 3 to redefine an already well-known Bayesian proxy (the relative plausibility). It is explained why the contextualization of this result into Haenni's works opens new issues: it allows for the generalization of this proxy in several non-Bayesian ways. Finally, Section 4 is an outlook on future works.

II. State of the art

Let X be a variable which takes its value on $\Omega_X = (x_1, x_2, \ldots, x_N)$. The piece of information $[\mathcal{I}]$ available[1] on the outcome of X is equivalently encoded in the four following functions defined on $\mathcal{P}(\Omega_X)$: The mass function $m_X^{[\mathcal{I}]}$, the belief function $Bel_X^{[\mathcal{I}]}$, the plausibility function $Pl_X^{[\mathcal{I}]}$ and the commonality function $q_X^{[\mathcal{I}]}$. It is possible to switch from one form to another by means of sum functions or with Möbius inversions [7]. The core of a mass function is the union of its focal elements. Here, we call *support* the set of potential focal elements according to the membership to a dedicated familly of mass functions[2]. The core and the support are noted $\mathcal{F}\left(m_X^{[\mathcal{I}]}\right) = \{F_1, \ldots, F_c\}\,,\ c \leq 2^{|\Omega_X|}$, and $\mathcal{S}\left(m_X^{[\mathcal{I}]}\right) = \{S_1, \ldots, S_C\}$. Of course, we have

$$\mathcal{F}\left(m_X^{[\mathcal{I}]}\right) \subseteq \mathcal{S}\left(m_X^{[\mathcal{I}]}\right).$$

A proxy is defined by an operator $\mathcal{O}$ that maps the set of mass functions on $\mathcal{P}(\Omega_X)$ onto the same set, so that the image

[1] Depending on the context, the piece of information $[\mathcal{I}]$ may correspond to (1) a partial knowledge derived from sample data, (2) the subjective opinion of an agent or (3) the fusion of several other pieces of information.

[2] Hence, this definition is not unique, and it is mainly used to simplify the notion of proxy. For instance, the support of any Bayesian belief function is the set of all the singletons, eventhough some of these singletons are not focal elements : no mass is associated to non-singleton, and a mass may potentially be associated to any of the singletons.

of the mapping (the proxy) has a smaller support:

$$\left|\mathcal{S}\left(\mathcal{O}\left[m_X^{[\mathcal{I}]}\right]\right)\right| \leq \left|\mathcal{S}\left(m_X^{[\mathcal{I}]}\right)\right|$$

A "good" proxy is defined so that the piece of information $[\mathcal{I}]$ expressed in the original mass function is "well preserved" in the proxy. Depending on the authors, this "preservation" is impersonated by different *consistency properties* that the proxy must verify: Commutativity with respect to Dempster's combination [8], convex linearity [9], dominance property [10], consistency with upper and lower bounds (for Bayesian proxies) [11], etc. In this paper, the commutativity with Dempster's combination plays a major role, but it does not diminish the interest of other consistency properties.

First, the most popular Bayesian approximations of a belief function are recalled[3].Then, non-Bayesian proxies are considered. Finally, we summarize the procedure described in [6] by Haenni to uncover Dempster's combinations in several operators of belief function theory.

A. Bayesian proxies

Given a mass function $m_X^{[\mathcal{I}]}$, the *pignistic probability* distribution is defined as :

$$m_X^{[BetP]}(x_i) = \frac{1}{1 - m_X^{[\mathcal{I}]}(\emptyset)} \sum_{\substack{x_i \in A \\ A \subseteq \Omega_X}} \frac{m_X^{[\mathcal{I}]}(A)}{|A|} \qquad \forall x_i \in \Omega_X$$

The corresponding operator is called the *pignistic transform*. Because it does not commute with Dempster's combination, the pignistic probability is not used as a computationally efficient proxy. Nonetheless, its convex linearity [9], its superposition with the barycenter of the dominating probabilities [12] and its obvious interpretation throughout the insufficient reason principle[4] makes it a major tool for decision making in the Transferable Belief Model [2]. Finally, Daniel [16] proposed using a mass function $m_X^{[\mathcal{W}]}$ to weight the pignistic transform according to prior knowledge. $\forall x_i \in \Omega_X$:

$$m_X^{[WBetP]}(x_i) = \frac{1}{1 - m_X^{[\mathcal{I}]}(\emptyset)} \sum_{\substack{x_i \in A \\ A \subseteq \Omega_X}} \frac{m_X^{[\mathcal{W}]}(x_i)}{\sum_{B \in A} m_X^{[\mathcal{W}]}(B)} \frac{m_X^{[\mathcal{I}]}(A)}{|A|}$$

The only proxy which is as popular as the pignistic probability is simply defined by normalizing the values of $Pl_X^{[\mathcal{I}]}$ of singletons so that they sum over Ω_X up to 1. $\forall x_i \in \Omega_X$:

$$m_X^{[RelPl]}(x_i) = \frac{Pl_X^{[\mathcal{I}]}(x_i)}{\sum\limits_{x_j \in \Omega_X} Pl_X^{[\mathcal{I}]}(x_j)} = \frac{Pl_X^{[\mathcal{I}]}(x_i)}{\sum\limits_{A \subseteq \Omega_X} m_X^{[\mathcal{I}]}(A) \cdot |A|}$$

This proxy is really interesting as it is both computationally efficient and useful for decision making: It was first introduced in a formal way in 1989 [3] by Voorbraak as a normalization up to 1 of the values of the commonality of the singletons[5], but its use for decision making came earlier, as the problem of finding *the most plausible configuration* was previously addressed (e.g. [13], [14]). In spite of its lack of dominating properties, it has been widely studied by various authors and widely justified. Hence, it has numerous names in the literature, such as *Bayesian approximation* [3] (1989), *proportional plausibility probability* [15] (2001), *plausibility transform* [8] (2003), *Cautious probabilistic transform* [16] (2006), *relative plausibility of singleton* [10], etc.

In a similar logic, it is possible to derive a proxy by normalizing the belief of singletons. As long as $\exists F \in \mathcal{F}(m_X^{[\mathcal{I}]})$ such that $|F| = 1$, it is defined by:

$$m_X^{[RelBel]}(x_i) = \frac{Bel_X^{[\mathcal{I}]}(x_i)}{\sum\limits_{x_j \in \Omega_X} Bel_X^{[\mathcal{I}]}(x_j)} \qquad \forall x_i \in \Omega$$

Basically, using this proxy means dropping the focal elements with cardinality greater than or equal to 2. This proxy has several interesting properties: In [15] (2001), Sudano briefly introduced it as the *proportional belief probability*. Then, it was introduced [19] (2003) and extensively studied [16] (2006) by Daniel as the *disjunctive probabilistic transform*. In the latter, Daniel also briefly discussed the interactions of several transforms with belief and plausibility functions from a geometric point of view. Meanwhile, Cuzzolin invested in a deeper way the geometric properties of the space of belief functions [10] and used them a few years later to derive the existence of this proxy, that he called the *relative belief of singletons* [17], [18] (2008). Interestingly enough, this proxy has alternatively been proposed for decision making and for computational saving.

In a similar way, Cuzzolin derived from geometric considerations the *orthogonal projection* and the *intersection probability* [10]. The *probability deficiency proportional plausibilities*, was very briefly introduced by Sudano [15] and shares important similarities with the intersection probability. Both of them have the following structure:

$$Bel_X^{[\mathcal{I}]}(x_i) + \left[1 - \sum_{x_j \in \Omega_X} Bel_X^{[\mathcal{I}]}(x_j)\right] \cdot \frac{E_X^{[\mathcal{I}]}(x_i)}{\sum\limits_{x_j \in \Omega_X} E_X^{[\mathcal{I}]}(x_j)}$$

$\forall x_i \in \Omega_X$, with $E_X^{[\mathcal{I}]} = Pl_X^{[\mathcal{I}]}$ for Sudano's proxy, and with $E_X^{[\mathcal{I}]} = \left(Pl_X^{[\mathcal{I}]} - m_X^{[\mathcal{I}]}\right)$ for Cuzzolin's.

Finally, Daniel [16] and Sudano [15] proposed several other proxies. A part of them share a structure similar to the previous ones, but the probability masses are weighted with the mass values (in order to be consistent with upper and lower bounds [11]). The proliferation of various Bayesian proxies shows that the search for new proxies and a better understanding of their definition are real issues. In a larger perspective, some

[3]In [3], "Bayesian approximation" refers to a particular proxy (the relative plausibility). In this paper, a Bayesian approximation is the result of an approximation operator that maps the set of belief functions onto the set of Bayesian belief functions (i.e. the support of the Bayesian proxy is Ω). Hence, Voorkraak's proxy is one among the various Bayesian approximations we consider.

[4]The insufficient reason principle helps to understand the intuition of the pignistic transform, but it does not justify it, as stressed by Smets [9].

[5]$Pl_X^{[\mathcal{I}]}$ and $q_X^{[\mathcal{I}]}$ are equivalent for singletons.

authors consider the definition of Bayesian proxies for more general distributions than belief functions, such as imprecise probabilities [20], or even fuzzy measures [21].

B. Non-Bayesian proxies

Up to now, we have only considered Bayesian proxies, as (1) their natural understanding, (2) their compactness and (3) their obvious relation to decision making make them rather popular. On the other hand, non-Bayesian proxies are also interesting and have been investigated.

The sets of the strong inner (resp. outer) approximations have extensively been studied. In [22], Dubois and Prade proposed to consider these two sets and to seek for two mass functions which:

- are consonant (the focal elements are such that $F_i \subseteq F_j$ $\forall 0 < i < j \leq c$ with $c \leq |\Omega_X|$),
- minimize some distance criterion with the original belief function.

This leads to the definition of the *minimal consonant outer approximation* and the *maximal consonant strong inner approximation*. These proxies are particularly interesting as they can be seen as fuzzy sets or as necessity measures. Unfortunately, these approximations do not commute with Dempster's combination.

In [25], Denœux described the construction of an inner (resp. outer) strong approximation $\underline{m}_X^{[\mathcal{I}]}$ (resp. $\overline{m}_X^{[\mathcal{I}]}$) as the result of an inner (resp. outer) reduction (a reduction is a particular mapping from Ω_X onto a coarser frame $\Theta(\Omega_X)$, where Θ is a partition of Ω_X). To lose as little information as possible, $\Theta(\Omega_X)$ is defined as the result of a hierarchical clustering of Ω_X. Then, in [26], it is shown that $\underline{m}_X^{[\mathcal{I}]}$ (resp. $\overline{m}_X^{[\mathcal{I}]}$) commutes with disjunctive (resp. conjunctive) combinations, and that they form an efficient basis with the Fast Möbius Transform [27] to approximate the result of a combination.

Then, there is the set k-additive belief functions [5]. A k-additive belief function has no focal element of cardinality $> k$, and at least one focal element of cardinality k. The set of k-additive belief functions dominating a dedicated fuzzy measure is described in [21]. In [23], [24], we proposed to generalize the pignistic transform so that, its result is a k-additive belief function instead of a Bayesian (or 1-additive) belief function, $\forall k \leq N$ chosen by the user. As the pignistic transform, this generalization does not commute with Dempster's combination. Thus, its main interest is to provide a decision making framework where it is possible to compare imprecise and precise decisions.

Finally, in [4], Tessem described the following algorithm to approximate a mass function: *(1) Sort the focal elements by decreasing mass. (2) Remove iteratively the focal element with the smallest mass until a criterion on the number of focal elements or on the mass discarded in the process is reached. (3) Normalize the remaining masses to 1.* The mass function obtained is the proxy. It is justified by the following concerns: As the focal elements with small mass become immaterial during a succession of Dempster's combinations, deleting them at the very beginning is not problematic, but computation saving.

C. Hidden Dempster's combinations

In [6], Haenni proposes to uncover Dempster's combinations in the process of applying an operator[6] $\mathcal{O}$ to a mass function $m_X^{[\mathcal{I}]}$. The application of the operator is seen as a combination of two mass functions: the first one is the belief function to process $m_X^{[\mathcal{I}]}$. The second one is a mass function that encodes a meta-information $[\mathcal{M}]$ corresponding to the semantics of the operator $\mathcal{O}$. Let us call Y the variable on which $[\mathcal{M}]$ is informative. Here is the description of the process from [6]:

1) Define $m_Y^{[\mathcal{M}]}$ and extend it to (X, Y). Depending on the relationship between X and Y this extension may be a vacuous one $m_Y^{[\mathcal{M}]\uparrow(X,Y)}$ or a ballooning one: $m_Y^{[\mathcal{M}]\Uparrow(X,Y)}$. Let us use the generic notation $m_{(X,Y)}^{[\mathcal{M}]}$.
2) Express $[\mathcal{I}]$ as a piece of information on the outcome of (X, Y) rather than on X, so that it is possible to define $m_{(X,Y)}^{[\mathcal{I}]}$. Once again, this extension may be a vacuous or a ballooning one[7].
3) Compute the Dempster's combination of these two masses and marginalize the result on the appropriate[8] space Z.

The result of this procedure corresponds to the result of the application of the operator $\mathcal{O}^{[\mathcal{M}]}$. Thus, if we call $[\mathcal{R}]$ the resulting piece of information, one has:

$$\mathcal{O}^{[\mathcal{M}]}\left(m_X^{[\mathcal{I}]}\right) = m_X^{[\mathcal{R}]} = \left[m_{(X,Y)}^{[\mathcal{I}]} \ \oplus \ m_{(X,Y)}^{[\mathcal{M}]}\right]^{\downarrow Z}$$

Let us note that a similar method is used in [28] to provide a meaning to α-junctions [29].

III. Defining proxies by means of Dempster's combination

In this work, $\mathcal{O}$ is supposed to be an approximating operator. If $\mathcal{O}$ can be expressed in terms of a Dempster's combination, then, it obviously commutes with it. Therefore, the approximating operators which do not commute with Dempster's combination are obviously not in the scope of this paper (such as the consonant proxies, as the consonant structure is broken by Dempster's combination [22]).

In this section, we apply Haenni's procedure to provide a definition of several proxies. First, some Dempster's combinations are uncovered in several well-known proxies. Then, we consider several generalization of the relative plausibility in the framework of Haenni.

[6]Several operators are considered: Discounting, disjunctive combination, refinement, coarsening (among which inner and outer reductions) and enlarging the frame to the open world assumption.

[7]In [6], it is a ballooning extension for the discounting operator and combinations, and a vacuous extension for coarsening/refinement operators.

[8]Most of the time, $Z = X$.

A. Hidden combination in approximating operators

In [3], it is established that the relative plausibility commutes with Dempster's combination. In [11], the relative plausibility in the case of frames of two elements is introduced as the result of a given homomorphism on a Dempster semi-group. On frames of more than 2 elements, the Dempster semi-group structure no longer holds, but it is demonstrated that its main properties still holds: The relative plausibility is the result of a Dempster's combination with the following uniform Bayesian mass function:

$$\begin{array}{rcll} m_X^{[\mathcal{U}ni]}(x_i) & = & \frac{1}{N}, & \forall i \leq N \\ m_X^{[\mathcal{U}ni]}(.) & = & 0 & \text{otherwise.} \end{array}$$

It is almost trivial to rewrite the proof of [16] in the framework of Haenni: Obviously, one has, $\Omega_Y = \Omega_X$ and $m_Y^{[\mathcal{M}]} = m_X^{[\mathcal{U}ni]}$. Thus, the combination, as well as the following marginalization described in Haenni's procedure are completely straightforward. In this setting, here is an interpretation for the meta-information $[\mathcal{M}]$: It is a list of the focal elements which are expected for the proxy $m_X^{[RelPl]}$, weighted according to their respective importance (the uniform distribution means they are all as much important).

It is possible to generalize this proxy by using a different meta-information $[\mathcal{M}]$ which is not necessarily $[\mathcal{U}ni]$: The same focal elements are considered, but some of them are promoted with respect to the others. Then, the corresponding Bayesian mass function $m_X^{[\mathcal{M}]}$ does not read uniform distribution. It is straightforward to establish the following proposition:

Proposition 1: *Let $m_X^{[\mathcal{M}]}$ be a Bayesian mass function and $m_X^{[\mathcal{I}]}$ be a mass function. We define the mass function*

$$m_X^{[\mathcal{R}_{Pl}]} = \mathcal{O}^{[\mathcal{M}]}\left(m_X^{[\mathcal{I}]}\right) = m_X^{[\mathcal{M}]} \oplus m_X^{[\mathcal{I}]}$$

$m_X^{[\mathcal{R}_{Pl}]}(A) = 0,\ \forall A\ /\ |A| \neq 1$, *and* $\forall x_i \in \Omega_X$, *we have:*

$$\begin{array}{rcl} m_X^{[\mathcal{R}_{Pl}]}(x_i) & = & \mathcal{K}[\mathcal{R}_{Pl}] \cdot q_X^{[\mathcal{M}]}(x_i) \cdot q_X^{[\mathcal{I}]}(x_i) \\ & = & \mathcal{K}[\mathcal{R}_{Pl}] \cdot m_X^{[\mathcal{M}]}(x_i) \cdot Pl_X^{[\mathcal{I}]}(x_i) \\ \textit{where } \mathcal{K}[\mathcal{R}_{Pl}] & = & \left[\displaystyle\sum_{x_j \in \Omega_X} m_X^{[\mathcal{M}]}(x_j) \cdot Pl_X^{[\mathcal{I}]}(x_j)\right]^{-1} \end{array}$$

$\mathcal{O}^{[\mathcal{M}]}$ is an Bayesian approximating operator. Hence $m_X^{[\mathcal{R}_{Pl}]}$ is a Bayesian approximation of $m_X^{[\mathcal{I}]}$. Moreover, in the particular case where $m_X^{[\mathcal{M}]}$ is uniform, $\mathcal{O}^{[\mathcal{M}]}$ corresponds to the plausibility transform ($m_X^{[\mathcal{R}_{Pl}]}$ is the relative plausibility of $m_X^{[\mathcal{I}]}$).

A proof of a more general theorem is given in section *3.C*, but the idea is to compute the Dempster's combination by the use of the communalities, and by using the fact that q and m are equal for Bayesian mass functions. Moreover, it is easy to see that, in case $m_X^{[\mathcal{M}]}(x_i) = \frac{1}{|\mathcal{F}|} = \frac{1}{N}, \forall x_i \in \Omega_X$, we have,

$$\frac{|\mathcal{F}|}{\mathcal{K}[\mathcal{R}_{Pl}]} = \sum_{x_j \in \Omega_X} Pl_X^{[\mathcal{I}]}(x_j)$$

which leads to the particular case of the relative plausibility.

By now, let us note that the role of $m_X^{[\mathcal{M}]}$ is also really close to the one of $m_X^{[\mathcal{W}]}$ in the $[\mathcal{W}]$-weighted pignistic transform: The piece of meta-information is used both as a manner to encode the approximating operator (the plausibility transform) and as a prior knowledge to weight the various masses of the proxy.

Let us now consider the outer/inner proxies which commute with Dempster's combination. In [6], it is also shown that inner and outer reductions are operators which correspond to hidden Dempster's combinations. Then, $\underline{m}_X^{[\mathcal{I}]}$ and $\overline{m}_X^{[\mathcal{I}]}$ can obviously be described in terms of hidden Dempster's combinations. In any of these cases, the meta-information $[\mathcal{M}]$ depicts the structure of a coarsening Θ (see [6]). Thus, Y corresponds to $\Theta(\Omega_X)$. One has:

$$m_{(X,Y)}^{[\mathcal{M}]}\left(\bigcup_{\theta_i \in \Theta(\Omega_X)} \theta_i \times \{f_\Theta(\theta_i)\}\right) = 1$$

where f_Θ is a function that maps each element θ_i of the partition $\Theta(\Omega_X)$ to the corresponding set of elements $\{x_{i,1}, x_{i,2}, \ldots, x_{i,j}\}$ of Ω_X.

B. Interesting structures

A major goal of this paper is to generalize the relative plausibility. In the previous section, the plausibility transform is interpreted in Haenni's framework, and it leads to a first generalization, $\mathcal{O}^{[\mathcal{M}]}$, which remains Bayesian. We aim at being even more general, so that non-Bayesian generalizations of the relative plausibility are defined. So far, we have intuitively understood that $[\mathcal{M}]$ describes which elements of $\mathcal{P}(\Omega_X)$ potentially belong to the core of the proxy. In fact, the support of the proxy depends on the core of $m_X^{[\mathcal{M}]}$:

$$\mathcal{S}\left(m_X^{[\mathcal{R}_{Pl}]}\right) = \mathcal{F}\left(m_X^{[\mathcal{M}]}\right)$$

In this subsection, we discuss the alternative structure for the core of the proxy: $[\mathcal{M}]$ is replaced by another piece of meta-information $[\mathcal{M}']$ and we discuss the structure of the core of the mass function which encodes $[\mathcal{M}']$.

From the previous section, $m_Y^{[\mathcal{M}']}$ is conveniently encoded in a Bayesian structure. On the other hand, we need to specify some focal elements which are not singletons, as the proxy is not bound to be Bayesian anymore. Then, it is natural to consider that $m_Y^{[\mathcal{M}']}$ is defined on Ω_Y, where Ω_Y is a set of atoms y_i, each of them representing a particular subset of $\mathcal{P}(\Omega_X)$. As a consequence, Ω_Y is made of $2^{|\Omega_X|}$ elements. Let us note as $\diamondsuit_{A_i} = \diamondsuit_{\{x_{i_1}, x_{i_2}, \ldots x_{i_j}\}}$ an atomic element y_i of Ω_Y which corresponds to the focal element $A_i = \{x_{i_1}, x_{i_2}, \ldots x_{i_j}\}$ with $j \leq N$. As the y_i's are seen as atoms, there is no inclusion relation amongst them, as with classical focal elements. Thus,

$m_Y^{[\mathcal{M}']}$ remains Bayesian, but any type of support for the proxy can be described. Moreover, it is possible to specify in $[\mathcal{M}']$ the presence of a potential focal element S_i of great cardinality, without implying the presence of the potential focal elements which are included in S_i (on the contrary to what would occur by using Dempster's combination with a non Bayesian mass function).

The most straightforward generalization of the relative plausibility of singletons is to consider focal elements which are k-uples (with $2 \leq k \leq N$) rather than singletons. In such a case, $[\mathcal{M}']$ reads "All and only the k-uples in Ω_X are potential focal elements for the proxies". To remain pedagogical, let us only consider uniformly distributed mass functions[9]. In such a case, the support of the proxy is made of $\binom{N}{k} = \frac{N!}{k!(N-k)!}$ potential focal elements of cardinality k (even if $m_Y^{[\mathcal{M}']}$ is Bayesian). We have:

$$\begin{aligned} m_Y^{[\mathcal{M}']}(\diamondsuit_{A_i}) &= \frac{k!(N-k)!}{N!} \\ &\quad \forall \diamondsuit_{A_i} \in \Omega_Y, \ / \ A_i \subseteq \Omega_X, |A_i| = k \\ m_Y^{[\mathcal{M}']}(.) &= 0 \qquad \text{otherwise} \end{aligned}$$

A more interesting generalization is to consider k-additive belief functions, such as those recalled in Section 2. Of course, $m_Y^{[\mathcal{M}']}$ is still Bayesian, and even if we only consider the case of uniform distribution, we have:

$$\begin{aligned} m_Y^{[\mathcal{M}']}(\diamondsuit_{A_i}) &= \frac{1}{|\mathcal{F}|} \qquad \forall A_i \in \Omega_X, |A_i| \leq k \\ m_Y^{[\mathcal{M}']}(\diamondsuit_{A_i}) &= 0 \qquad \text{otherwise} \end{aligned}$$

Another generalization proposed by Cuzzolin [10], is to consider that the cardinality of focal elements must be either 1 or k. For the corresponding class of belief functions the pignistic probability and the intersection probability are equal. Finally, the dual case of k-additive belief function is called k-intolerant belief functions [30]. In this case, there is no focal element with a cardinality lower than $k-1$. All these families have interesting supports, and in the sequel, we aim at defining a generalization of the plausibility transform, which can be parametrized to provide a proxy belonging to any of them.

C. The generalized relative plausibility

We do not consider each family (k-intolerant, k-additive, etc.) of belief functions separately. We directly consider the case of a general Bayesian mass function $m_Y^{[\mathcal{M}']}$, the core of which reads any type of structure, and its mass not necessarily being uniformly distributed among the focal elements.

So far, $m_X^{[\mathcal{I}]}$ (the mass function to approximate) and Y are known. Moreover, the meaning of $[\mathcal{M}']$ was preciously fully detailed. Hence, the definition of $m_Y^{[\mathcal{M}']}$ is rather simple. The only point is to remind that in Ω_Y, $m_Y^{[\mathcal{M}']}$ is Bayesian. Then, $\forall \diamondsuit_{A_i} \in \Omega_Y$, one has:

$$\begin{aligned} m_Y^{[\mathcal{M}']}(\diamondsuit_{A_i}) &= a_i \quad \text{such that} \quad \textstyle\sum_i a_i = 1 \\ m_Y^{[\mathcal{M}']}(.) &= 0 \qquad \text{otherwise} \end{aligned}$$

where A_i are subsets of Ω_X and $\diamondsuit_{A_i}$ are elements of Ω_Y.

The extension of $m_Y^{[\mathcal{M}']}$ to $m_{(X,Y)}^{[\mathcal{M}']}$ has not been discussed yet. As previously explained, Y is a set of elements representing the sets of $\mathcal{P}(\Omega_X)$. Then, Ω_Y is richer than Ω_X, in the meaning that, it is always possible to express a function on $\mathcal{P}(\Omega_X)$ as a function on $\mathcal{P}(\Omega_Y)$, which is a kind of powerset of the powerset of Ω_X. Thus, working on the vector (X,Y) may appear as a bit cumbersome, and working on Y may be sufficient. We adopt this latter strategy, as it simplifies the notation.

Then, it is necessary to express $[\mathcal{I}]$ as a piece of information on Y rather than on X(Later, this operation is noted $\uparrow\uparrow$). To do so, we propose to using the following mass function:

$$m_Y^{[\mathcal{I}]}(\nabla_A) = m_X^{[\mathcal{I}]}(A) \qquad \forall A \subseteq \Omega_X$$

where ∇_A is the set of all the $y_k = \diamondsuit_{\{k_1,k_2,\ldots,k_j\}}$ such that $\left\{x_{k_1}, x_{k_2}, \ldots x_{k_j}\right\} \subseteq A$. This cumbersome notation is necessary to make sure that $\forall A \subseteq \Omega_X$:

$$\left[m_Y^{[\mathcal{I}]} \oplus m_Y^{[\mathcal{J}]}\right](\nabla_A) = \left[m_X^{[\mathcal{I}]} \oplus m_X^{[\mathcal{J}]}\right](A)$$

Let us compute the Dempster's combination

$$m_Y^{[\mathcal{G}_{Pl}]} = m_Y^{[\mathcal{I}]} \ \oplus \ m_Y^{[\mathcal{M}']}$$

with the communalities: $q_Y^{[\mathcal{G}_{Pl}]} = \mathcal{K}' \cdot q_Y^{[\mathcal{I}]} \cdot q_Y^{[\mathcal{M}']}$, with,

$$\mathcal{K}' = \left[\sum_{A \neq \emptyset} (-1)^{|A|+1} \cdot q_Y^{[\mathcal{M}']}(A) \cdot q_Y^{[\mathcal{I}]}(A)\right]^{-1}$$

Moreover, $m_Y^{[\mathcal{M}']}$ is Bayesian. Thus, $m_Y^{[\mathcal{G}_{Pl}]}$ is Bayesian, and we have $m_Y^{[\mathcal{M}']} = q_Y^{[\mathcal{M}']}$ and $m_Y^{[\mathcal{G}_{Pl}]} = q_Y^{[\mathcal{G}_{Pl}]}$. Thus $\forall A \subseteq \Omega_X$:

$$\begin{aligned} m_Y^{[\mathcal{G}_{Pl}]}(\diamondsuit_A) &= \mathcal{K}' \cdot m_Y^{[\mathcal{M}']}(\diamondsuit_A) \cdot q_Y^{[\mathcal{I}]}(\diamondsuit_A) \\ \text{and} \quad q_Y^{[\mathcal{G}_{Pl}]}(.) &= 0 \qquad \text{otherwise} \end{aligned}$$

The last step is to "marginalize back" the result on X (noted $\downarrow\downarrow$). As each focal element is of the form $\diamondsuit_A$, there is a one-to-one correspondence between the focal elements of $m_Y^{[\mathcal{G}_{Pl}]}$ and the elements of Ω_X. Thus, we define the following new masses:

$$\begin{aligned} m_X^{[\mathcal{G}_{Pl}]}(A) &= m_Y^{[\mathcal{G}_{Pl}]}(\diamondsuit_A) \\ m_X^{[\mathcal{M}']}(A) &= m_Y^{[\mathcal{M}']}(\diamondsuit_A) \end{aligned}$$

Moreover, we have, by construction, $q_Y^{[\mathcal{I}]}(\diamondsuit_A) = q_Y^{[\mathcal{I}]}(\nabla_A)$, as the $\nabla_{\{.\}}$ are the only focal elements of $m_Y^{[\mathcal{I}]}$, and as

$$\nabla_A = \diamondsuit_A \cup \left[\bigcup_{B \subset A} \nabla_B\right]$$

Moreover, as by definition, $q_Y^{[\mathcal{I}]}(\nabla_A) = q_X^{[\mathcal{I}]}(A)$, one has the following result:

[9]Of course, this is not mandatory, but (1) it helps to understand the global procedure, (2) it corresponds to a "fair" situation where no prior information is available, nor prior weight $[\mathcal{W}]$ is used.

Theorem 1: *Let $m_X^{[\mathcal{M}']}$ and $m_X^{[\mathcal{I}]}$ be two mass functions. $m_Y^{[\mathcal{M}']}$ is derived from $m_X^{[\mathcal{M}']}$ according to the process described above. We define the following mass function:*

$$m_X^{[\mathcal{G}_{Pl}]} = \mathcal{O}^{[\mathcal{M}']}\left(m_X^{[\mathcal{I}]}\right) = \left[\left[m_Y^{[\mathcal{M}']}\right] \oplus \left[m_X^{[\mathcal{I}]}\right]^{\uparrow\uparrow Y}\right]^{\downarrow\downarrow X}$$

1) *$\forall A \subseteq \Omega_X$, we have:*
$$m_X^{[\mathcal{G}_{Pl}]}(A) = \mathcal{K}' \cdot m_X^{[\mathcal{M}']}(A) \cdot q_X^{[\mathcal{I}]}(A)$$
where $\mathcal{K}'$ is a normalizing factor.
2) *$\mathcal{O}^{[\mathcal{M}']}$ is an approximating operator thant maps the set of belief functions onto the set of belief functions the support of which is $\mathcal{F}\left(m_X^{[\mathcal{M}']}\right)$. In other words, $m_X^{[\mathcal{G}_{Pl}]}$ is a proxy for $m_X^{[\mathcal{I}]}$.*
3) *If $m_X^{[\mathcal{M}']}$ is Bayesian, then $\mathcal{O}^{[\mathcal{M}']}$ is a Bayesian approximating operator. Moreover, its result $m_X^{[\mathcal{G}_{Pl}]}$ corresponds to $m_X^{[\mathcal{R}_{Pl}]}$ defined in Proposition 1.*
4) *If $m_X^{[\mathcal{M}']}$ is Bayesian and uniform, then $\mathcal{O}^{[\mathcal{M}']}$ corresponds to the plausibility transform, (or equivalently, $m_X^{[\mathcal{G}_{Pl}]} = m_X^{[RelPl]}$ is the relative plausibility of $m_X^{[\mathcal{I}]}$).*
5) *$\mathcal{O}^{[\mathcal{M}']}$ is a generalization of the plausibility transform, parametred by $m_X^{[\mathcal{M}']}$.*
6) *$\mathcal{O}^{[\mathcal{M}']}$ commutes with Dempster's combination.*

proofs: 1) and 2) see above. 3) and 4) consequences of 2) and see the explanations after Proposition 1. 5) Consequences of 1)-4). 6) Obvious. □

IV. Conclusion & outlook

A parameterized family of proxies is defined, by the use of Dempster's combination and a meta-information encoding a particular structure of support. This family of proxy generalizes the relative plausibility in many ways, such as, for instance, the k-additive relative plausibility, or the weighted relative plausibility. All these non-Bayesian proxies are based on the commonality function, such as initially defined by Voorbraak [3].

In parallel to this work on relative plausibility, we have seen that other well-known proxies can be defined in Haenni's framework, such as Denœux' strong inner approximation, and Denœux' strong outer approximation. A similar work on the relative belief woud be interesting, for two reasons: Firstly, it is shown in [6], that the disjunctive combination is a hidden Dempster's combination (*theorem 3.2*). Then, the relative belief can be expressed in Haenni's framework, even if the relative belief commutes with the disjunctive combination and does not commute with Dempster's combination. Secondly, it is shown in [17], [18], that the relative belief of a plausibility function (then, the plausibility function is seen as a belief function) commutes with Dempster's combination.

In the future, a discussion on the distribution of the mass function $m_Y^{[\mathcal{M}']}$ which parameterizes the transform would be interesting. It would also be interesting to look for measures to quantify the distances between a mass function and its various proxies, in order to choose the closest one, or to choose the one which provides the best "distance/size of the support" ratio.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton U. P., 1976.
[2] P. Smets and R. Kennes, "The transferable belief model", *Artificial Intelligence*, vol. 66 no.2, pp. 191–234, 1994.
[3] F. Voorbraak, "A computationally effcient approximation of Dempster-Shafer theory," *Int. J. Man-Machine Studies*, vol. 30, pp. 525–536, 1989.
[4] B. Tessem, "Approximations for Efficient Computation in the Theory of Evidence," *Artif. Intell.* Vol. 61, No. 2, pp. 315–329, 1993.
[5] M. Grabisch, "K-order additive discrete fuzzy measures and their representation," *Fuzzy sets and systems*, vol. 92, pp. 167–189, 1997.
[6] R. Haenni, "Uncover Dempster's Rule Where It Is Hidden," *Proceedings of the Inter. Conf. on Information Fusion*, Florence, Italy, 2006.
[7] G. C. Rota, "On the foundations of combinatorial theory I. Theory of Möbius functions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol.2, pp.340–368, 1964.
[8] B. R. Cobb and P. P. Shenoy, "A Comparison of Methods for Transforming Belief Functions Models to Probability Models," *Lecture Notes in Artificial Intelligence* vol. 2711, pp. 255–266, 2003.
[9] P. Smets, "Decision Making in the TBM: the Necessity of the Pignistic Transformation," *Int. J. Approximate Reasoning*, vol. 38, 2005.
[10] F. Cuzzolin, "Two new Bayesian approximations of belief functions based on convex geometry," *IEEE Trans. on Systems, Man, and Cybernetics*, part. B, vol.37, no. 4, pp. 993–1008, 2007.
[11] M. Daniel, "Probabilistic Transformations of Belief Functions," *L. Godo (Ed.): Proceedings of ECSQARU'05*, LNAI 3571, pp. 539–551, 2005.
[12] D. Dubois, H. Prade and P. Smets, "New semantics for quantitative possibility theory," *in the proceedings of ISIPTA*, pp. 152–161, 2001.
[13] G. Shafer and P. P. Shenoy, *Local Computation on hypertrees*. Working paper #201, School of Business, University of Kansas, 1988.
[14] R. Giles, "Foundations for a possibility theory". *M. M. Gupta and E. Sanchez (eds.), Fuzzy Information and Decision Processes*, North-Holland, Amsterdam, pp.83–195, 1982.
[15] J. J. Sudano, "Pignistic Probability Transforms for Mixes of Low- and High- Probability Events", *Proceedings of the International Conference on Information Fusion*, Montreal, Canada, August 7-10, 2001.
[16] M. Daniel, "On Transformations of Belief Functions to Probabilities," *Int. J. of Intelligent systems*, vol. 21, pp. 261–282, 2006.
[17] F. Cuzzolin, "Semantics of the relative belief of singletons," *Workshop on Uncertainty and Logic*, 2008.
[18] F. Cuzzolin, "Dual properties of the relative belief of singletons," *In the proceedings of PRICAI'08*, 2008.
[19] M. Daniel, "On Transformations of Belief Functions to Probabilities," J. Vejnarová, ed., *Proceedings of the 6th Workshop on Uncertainty Processing in Expert Systems*, pp. 77–90, 2003.
[20] P. Baroni and P. Vicig,"Transformations from Imprecise to Precise Probabilities," *Nielsen, T. D., Zhang, N. L. (eds.): Proceedings of ECSQARU 2003*, LNAI 2711, pp. 37–49, 2003.
[21] P. Miranda, M. Grabisch and P. Gil, "Dominance of capacities by k-additive belief functions," *EJOR*, vol. 175, pp. 912–930, 2006.
[22] D. Dubois and H. Prade, "Consonant approximations of belief measures," *Int. J. of Approximate Reasoning*, Vol. 4, pp. 419–449, 1990.
[23] T. Burger and A. Caplier, "A generalization of the Pignistic Transform for partial bet," *C. Sossai and G. Chemello (eds.): Proceedings of ECSQARU 2009*, LNAI 5590, pp. 252–263, 2009.
[24] O. Aran, T. Burger, A. Caplier and L. Akarun, "A Belief-Based Sequential Fusion Approach for Fusing Manual and Non-Manual Signs," *Pattern Recognition*, vol. 42, No. 5, pp. 812–822, May 2009.
[25] T. Denoeux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, No. 4, pp. 437–460, 2001.
[26] T. Denoeux and A. Ben Yaghlane, "Approximating the Combination of Belief Functions using the Fast Moebius Transform in a coarsened frame," *Int. J. of Approximate Reasoning*, Vol. 31, No. 1-2, pp. 77–101, 2002.
[27] R. Kennes, "Computational aspects of the Möbius transform of graphs", IEEE Tran. on Systems, Man, & Cybernetics Vol. 22 pp. 201–223, 1992.
[28] F. Pichon and T. Denoeux, "Interpretation and computation of α-junctions for combining belief functions," *in the proceedings of the sixth int. symp. on Imprecise Probability: Theories and Applications*, July 2009.
[29] P. Smets, "The α-junctions: combination operators applicable to belief functions," *Proceedings of the first int. joint conf. on Qualitative and Quantitative Practical Reasoning*, LNCS 1244, pp. 131–153, 1997.
[30] J.-L. Marichal, "k-intolerant capacities and Choquet integrals", *EJOR*, Vol. 177, No. 3, pp. 1453-1468, 2007.