A New Combination Function in Evidence Theory

Vicenç Torra
Dpt. d'enginyeria informática
Escola Tècnica Superior d'Enginyeria (ETSE)
Universitat Rovira i Virgili
Carretera de Salou, s/n
E-43006 Tarragona (Spain)
E-mail: vtorra@etse.urv.es

Abstract

This paper studies the combination of basic probability assignment (bpa) in evidence theory. After introducing an interpretation of the mass function we show that given two *bpa* Dempster's rule of combination does not build a coherent *bpa* with respect to the interpretation. Next we give a new combination function that overcomes this problems and study some of its properties.

0. Introduction

One of the models of approximate reasoning is evidence theory[8,9] where the knowledge over a frame of discernment θ is expressed by means of belief and plausibility measures. They are functions from (θ) (i.e., parts of θ) into [0,1]. In particular, a belief measure, Bel, should satisfy the following axioms[7]:

(i) $Bel(\emptyset)=0$ and $Bel(\theta)=1$
(ii) For every A, B (θ), if $A \prod B$ then $Bel(A)=Bel(B)$
(iii) $Bel(A_1 \approx A_2 \approx ... \approx A_n) =$

$$? _i Bel(A_i) - ? _{i<j} Bel(A_i \leftrightarrow A_j) +...+ (-1)^{n+1} Bel(A_1 \leftrightarrow A_2 \leftrightarrow ... \leftrightarrow A_n),$$

for all n M.

On the other hand, a plausibility measure, Pl, should satisfy (i), (ii) and also (iii)':

(iii)' $Pl(A_1 \leftrightarrow A_2 \leftrightarrow ... \leftrightarrow A_n) =$

$$? _i Pl(A_i) - ? _{i<j} Pl(A_i \approx A_j) +...+ (-1)^{n+1} Pl(A_1 \approx A_2 \approx ... \approx A_n),$$

for all n M.

In fact, belief and plausibility measures are related to each other. For each belief measure Bel, it can be defined a plausibility measure Pl in the following way: $Pl(A)=1 - Bel(A^c)$ for all A in (A), where A^c is the complement of A.

Similarly, we can define a belief measure, Bel, as Bel(A)=1 - Pl(A^C) for all A in (A). Belief and plausibility measures are, therefore, reciprocally dual.

Every belief measure and its dual plausibility measure can equivalently be expressed in terms of a function m from parts of θ onto [0,1] such that

$$m(\emptyset)=0$$
$$? \; _{A \prod \theta} \; m(A) = 1$$

where m(A) is interpreted as the degree of evidence supporting the claim that a specific element of X belongs to the set A but not to any special subset of A. This function *m* is called a mass function or, also, a basic probability assignment (bpa).

Given a basic probability assignment *m*, their corresponding belief and plausibility measures are uniquely determined by the formulae:

$$Bel(A) = ? \; _{B \prod A} \; m(B)$$
$$Pl(A) = ? \; _{B \leftrightarrow A ? \emptyset} \; m(B)$$

Conversely, given a belief function Bel, the corresponding bpa is given by:

$$m(A)=? \; _{B \wp A} \; (-1)^{|A-B|} \; Bel(B)$$

A mass function m generalises both probability distributions (where m(x) ? 0 if x?{e} e θ) and fuzzy sets (when m(x) = 0 if x $\{x_n=\theta, x_{n-1}, \ldots, x_1, x_0=\emptyset\}$ with x_{i+1} x_i}. In the former case, there are two main ways of relating belief measures to probability distributions[5]:

i) *Belief and plausibility as inner and outer measures.* In a probability space (θ,*X*,Pr), the probability measure Pr is not necessarily defined on (θ), but only on *X* In this case, we can extend Pr to (θ) in the following ways:

$Pr_*(A) = \sup \{Pr(x) |$ x∏A and x *X*$\}$ for all A (θ)

$Pr^*(A) = \inf \{Pr(x) |$ x∈A and x *X*$\}$ for all A (θ)

The function Pr_* thus defined satisfies axioms (i)-(iii) given above, and therefore it is a belief function. The function Pr^*, instead, satisfies axioms (i), (ii) and (iii)', so, it is a plausibility measure.

ii) *Probability interval induced by a belief measure.* Each belief measure, with its dual plausibility, induces a probability interval P_{Bel}:

$$P_{Bel} = \{p \mid Bel(A) = p(A) = Pl(A) \text{ for all } A \quad (\theta)\}$$

As a probability p such that Bel(A)=p(A)=Pl(A) for all A (θ) is called a probability consistent with Bel, P_{Bel} is the set of all consistent probabilities with a given Bel.

In this paper we introduce a new interpretation that also relates evidence theory with probability theory. In our case, every mass function is understood as a probability that has suffered from a process of information losing.

Due to the dual nature of belief functions, there are operations defined over mass functions (and also over beliefs) that generalise operations over sets[2] (e.g., union, intersection) and over probability[5]. In this latter case there are functions to update belief measures in the light of new evidence (generalisations of Bayes' and Jeffrey's rule in probability) and functions to combine two belief measures. Dubois and Prade difference two types of conditioning, named *focusing* and *revision* (updating in preliminary papers[3]).

Focusing: When no new piece of information is introduced, we just consider another reference class by focusing our attention on a given subset of the original set.

Revision: The probability is modified to take into account a new piece of evidence.

In the revision process, the piece of introduced evidence can take several forms. It could be crisp information, e.g., "an event B has occurred", so the new belief function would be of the form: Bel(·|B). See references [6, 10, 13] for some definitions of conditional belief as generalisations of Bayes' and Jeffrey's rules. It

could be, instead, information represented by means of a belief measure. In this later case, the combination function is either defined according to a set of needed properties, i.e., axiomatically (see reference[9] for the properties of Dempster's rule of combination) or defined according to the interpretation given to belief functions. For instance, when we consider probability intervals induced by a belief function, the combination function should consider the intersection of the intervals induced by initial belief measures. Chateauneuf[1] studies the compatibility of two belief functions in this later setting.

In this paper, after introducing a new interpretation of mass functions based on probabilities and on an information losing function (section 1), we show that when two masses are defined from a single probability and they are combined through Dempster's rule the new function can be non-definable from the same probability (section 2). We introduce then a combination function that satisfies this condition (section 3) and we study some of its properties (section 4).

1. Preliminaries

In this section we define what an information losing function is, and when a mass m_2 is more general than another mass m_1. We also show how to define a mass from a probability and a mass attached to a partition **Y** of the basis of a σ-algebra X.

Definition 1. Let m_1, m_2 be two masses and let x_0, x_1 be subsets of θ that satisfy $x_0 \wp x_1$. In these conditions it is said that *the set x_0 in mass m_1 has lost information in favour of x_1 in mass m_2* if and only if:

$m_2(x)=m_1(x)$ for all $x \prod \theta$, such that $x?x_0$ and $x?x_1$
$m_2(x_0)=m_1(x_0) - k$ with $m_1(x_0)=k$
$m_2(x_1)=m_1(x_1) + k$

Definition 2. Let m_1, m_2 be two masses. Then, it is said that *the mass m_1 has lost information in favour of mass m_2* if exists $x_0 \wp x_1 \prod \theta$ such that x_0 in mass m_1 has lost information in favour of x_1 in mass m_2.

We say that m_1 has lost information because some of the mass that before supported x_0 is now supporting the whole set x_1, so not only x_0 but also the elements in x_1-x_0.

Definition 3. Let m_1, m_2 be two masses. Then *mass m_2 is said to be more general than the mass m_1* if and only if we can build m_2 from m_1 through a sequence m_i, m_{i+1}, ..., m_j where $m_1=m_i$, $m_2=m_j$ and where each m_k with k {i,...,j-1} has lost information in relation to mass m_{k+1}.

Next we define the concept of information losing function (called *pdi*), as a function that given a mass m_1 builds another more general m_2.

Definition 4. Let *m* be a mass function defined over a frame of discernment θ. In these conditions a function *pdi*: (θ)* (θ) -> [0,1] is called an *information losing function* over m if it satisfies

$$m(A) + ?_{x_0:x_0 \Pi A}\, pdi_A(x_0) - ?_{x_1:A \wp x_1}\, pdi_{x_1}(A) = 1 \qquad \text{for all } A \quad (\theta)$$

where $pdi_x(y)$ stands for pdi(x,y), and $pdi_{x_1}(x_0)=0$ for all x_0, x_1 (θ) that do not satisfy $x_0 \wp x_1 \prod \theta$.

The expression $pdi_{x_1}(x_0)$ is read as the quantity of information that x_0 loses in favour of x_1. Note that the function *pdi* is a transference function that only allows transference between two sets x_0 and x_1 if $x_0 \wp x_1$.

Definition 5. Given a mass *m* and an information losing function *pdi* over *m*, we define the function m_{pdi}: (θ) -> [0,1] by:

$$m_{pdi,m}(A) = m(A) + ?_{x_0 \wp A}\, pdi_A(x_0) - ?_{A \wp x_1}\, pdi_{x_1}(A)$$

Next proposition settles that the function defined above is a mass function.

Proposition 1. Let *m* be a mass function defined over a frame of discernment θ, let *pdi* be an information losing function over m. In these conditions, m_{pdi} is a mass function.

Example 1. Let m_1, m_2 be two masses defined over a single frame of discernment θ; let the set x_0 in mass m_1 be a set that has lost information in

favour of x_1 ($x_0 \wp x_1 \prod \theta$) in mass m_2 according to definition 1. In these conditions, according to definition 4, if we define the following *pdi*:

$pdi_B(A)=k$ if $A=x_0$ and $B=x_1$
$pdi_B(A)=0$ otherwise

then, according to definition 5, m_2 can be expressed of the form:

$$m_2=m_{pdi,m_1}$$

Next we define a basis of a σ-algebra X of subsets of θ. X is a σ-algebra of θ if X is a set of subsets of θ that contains θ and is closed with respect to union and intersection.

Definition 6. A subset y of X is a basis if[4]

1) the members of y are disjoint and not empty, and
2) X exactly consists on the numerable union of y members

In our case, as θ is supposed to be finite, X is finite. So, the basis y of X is unique (see reference[4]) and consists precisely of the minimal elements of X.

Next we see that (i) any probability distribution defined over a σ-algebra X generates a mass function, and that (2) every mass function can be seen as the result of applying an information losing process to a probability function.

Proposition 2[5]**.** Given a probability distribution P defined over the σ-algebra X, the function m_P: (θ)->[0,1] defined as:

$m_P(A) =$ P(A) if A y, where y is the basis of X
0 otherwise

is a mass function.

Proposition 3. For each mass function there is at least one probability distribution that can generate it by means of an information losing function.

Proof. Let *m* be a mass over the power set of θ, then the probability distribution P defined also over (θ) as:

$$P(\{e\}) = ?_{A:e\ A}\ m(A) / |A| \qquad \text{for all } e\ \theta$$

can generate *m* through the following information losing function *pdi*:

$$pdi_{x_1}(x_0) = \begin{cases} m(x_1)/|x_1| & \text{if } x_0=\{e\} \text{ and } e\ x_1 \text{ for some } e\ \theta \\ 0 & \text{otherwise} \end{cases}$$

o

As stated by Dubois and Prade[2], the probability distribution defined in this way is in according with Laplace's principle of modelling a lack of information by uniformly distributed probability allocations.

Definition 7. Given a probability distribution P defined over a σ-algebra X, and a partition **Y** of the basis of X, we define *the mass attached to the partition* **Y** as:

$$m_{P,\mathbf{Y}}(A) = \begin{cases} ?_{x_i\ x}\ P(x_i) & \text{if } A=\approx_{x_i\ x}\ x_i \text{ for a certain } x\ \mathbf{Y} \\ 0 & \text{otherwise} \end{cases}$$

That is, the mass of a set A is defined as the sum of the probability of all those sets x_i of the basis of X that define A, i.e., the sets x_i such that $A=\approx_{x_i\ x}\ x_i$. Due to the fact that **Y** is a partition over a basis, the function defined above is a mass function: each x_i of the basis of X is only included in a single set A (the set A where x_i) belongs.

Next we prove that, given a probability distribution P defined over a σ-algebra X, every mass attached to a partition **Y** of the basis of X is more general than the mass defined according to proposition 2.

Proposition 4. Let (θ,X,P) be a probability space. Let **Y** be a partition of the basis of X. In these conditions, the mass $m_{P,\mathbf{Y}}$ attached to the partition **Y** defined according to definition 7 is more general than the mass m_P defined according to proposition 2.

Proof. It is proved by defining

$pdi_A(x_i) = m_P(x_i)$ if $A=\approx_{x_i}\ x\ x_i$ for a certain x **Y** and x_i?A
$pdi_A(B) = 0$ otherwise o

2. Dempster's rule of combination

In this section we study the behaviour of Dempster's rule of combination in the framework of information losing functions. We give an example in order to prove that given two masses defined from a single mass *m* with two information losing functions, then the combined mass according to Dempster's rule of combination cannot be always defined from the same mass *m* with another information losing function.

In this example (example 2), after defining a σ-algebra X, a probability distribution P over X, and m_P according to proposition 2, we introduce two partitions $\mathbf{Y_1}$ and $\mathbf{Y_2}$ of the basis of X together with the masses m_1 and m_2 attached, respectively, to $\mathbf{Y_1}$ and $\mathbf{Y_2}$. We compute then m_c, the combined mass of m_1 and m_2 according to Dempster's rule of combination.

Proposition 5. Let m_1 and m_2 be two mass functions defined from a single probability distribution P through two different information losing functions (pdi_1, pdi_2), let m_c be the mass function obtained by combination of m_1 and m_2 through Dempster's rule of combination. In these conditions, it is not always possible to obtain m_c from P through an information losing function.

To prove the above proposition we give the following example:

Example 2 Given a frame of discernment θ, we consider three subsets $[P_1]$, $[P_2]$ and $[C_1]$ that correspond to those elements of θ that satisfy the properties P_1, P_2 and C_1 respectively. These subsets of θ point out the σ-algebra X with the following set y as its basis:

$$y=\{[\neg P_1 \wedge \neg P_2 \wedge \neg C_1], [\neg P_1 \wedge P_2 \wedge \neg C_1], [\neg P_1 \wedge P_2 \wedge C_1], [\neg P_1 \wedge \neg P_2 \wedge C_1],$$
$$[P_1 \wedge P_2 \wedge C_1], [P_1 \wedge \neg P_2 \wedge C_1], [P_1 \wedge P_2 \wedge \neg C_1], [P_1 \wedge \neg P_2 \wedge \neg C_1]\}$$

where [x^y^z] stands for the set of objects that satisfy properties x, y and z (note that [x^y^z]=[x]↔[y]↔[z]).

Let P, the probability distribution, be defined as

$P([\neg P_1 \wedge \neg P_2 \wedge \neg C_1]) = .1111$	$P([\neg P_1 \wedge P_2 \wedge \neg C_1]) = .0556$
$P([\neg P_1 \wedge P_2 \wedge C_1]) = .0833$	$P([\neg P_1 \wedge \neg P_2 \wedge C_1]) = .0278$
$P([P_1 \wedge P_2 \wedge C_1]) = .1389$	$P([P_1 \wedge \neg P_2 \wedge C_1]) = .1667$
$P([P_1 \wedge P_2 \wedge \neg C_1]) = .1944$	$P([P_1 \wedge \neg P_2 \wedge \neg C_1]) = .2222$

and therefore let m_P according to proposition 2 be:

$m_p(A) = P(A)$	if A y
$m_p(A) = 0$	otherwise

Let $\mathbf{Y_1}$ and $\mathbf{Y_2}$ be two partitions defined over the basis y:

$\mathbf{Y_1} = \{ \{[\neg P_1 \wedge \neg P_2 \wedge \neg C_1], [\neg P_1 \wedge P_2 \wedge \neg C_1]\}, \{[\neg P_1 \wedge P_2 \wedge C_1], [\neg P_1 \wedge \neg P_2 \wedge C_1]\}, \{[P_1 \wedge P_2 \wedge C_1], [P_1 \wedge \neg P_2 \wedge C_1]\}, \{[P_1 \wedge P_2 \wedge \neg C_1], [P_1 \wedge \neg P_2 \wedge \neg C_1]\}\}$

$\mathbf{Y_2} = \{\{[\neg P_1 \wedge P_2 \wedge C_1], [P_1 \wedge P_2 \wedge C_1]\}, \{[\neg P_1 \wedge \neg P_2 \wedge C_1], [P_1 \wedge \neg P_2 \wedge C_1]\}, \{[\neg P_1 \wedge P_2 \wedge \neg C_1], [P_1 \wedge P_2 \wedge \neg C_1]\}, \{[\neg P_1 \wedge \neg P_2 \wedge \neg C_1], [P_1 \wedge \neg P_2 \wedge \neg C_1]\}\}$

that correspond, respectively, to the sets:

$\mathbf{S_1} = \{ [P_1 \wedge C_1], [\neg P_1 \wedge C_1], [P_1 \wedge \neg C_1], [\neg P_1 \wedge \neg C_1] \}$

$\mathbf{S_2} = \{ [P_2 \wedge C_1], [\neg P_2 \wedge C_1], [P_2 \wedge \neg C_1], [\neg P_2 \wedge \neg C_1] \}$

Figure 1 shows the partition $\mathbf{Y_1}$ (on the left hand side it is underlined the subset $[P_1]$ and on the right hand side the subset $[C_1]$) and figure 2 corresponds to the partitions $\mathbf{Y_1}$, $\mathbf{Y_2}$ and the basis y.

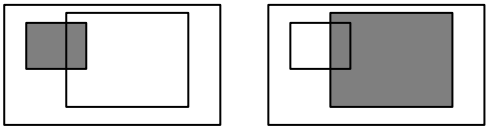
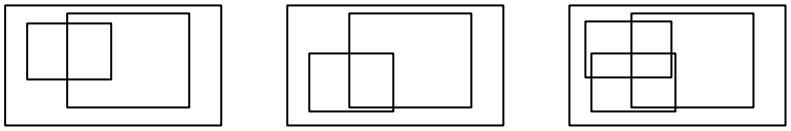


Figure 1. $[P_1]$ and $[C_1]$



Partition $\mathbf{Y_1}$ Partition $\mathbf{Y_2}$ Basis y

Figure 2

Given the probability distribution p defined above over the σ-algebra X; let m_1 and m_2 be the basic probability assignments attached to the partitions $\mathbf{Y_1}$ and $\mathbf{Y_2}$ according to definition 7. In these conditions m_1 and m_2 are of the form:

$m_1([P_1 \wedge C_1]) = .3056$ $\quad$ $m_1([\neg P_1 \wedge C_1]) = .1111$

$m_1([P_1 \wedge \neg C_1]) = .4166$ $\quad$ $m_1([\neg P_1 \wedge \neg C_1]) = .1667$

$m_2([P_2 \wedge C_1]) = .2222$ $\quad$ $m_2([\neg P_2 \wedge C_1]) = .1945$

$m_2([P_2 \wedge \neg C_1]) = .25$ $\quad$ $m_2([\neg P_2 \wedge \neg C_1]) = .3333$

Next we combine this two masses according to Dempster's rule of combination[8,9]:

$$m_c(Z) = \frac{\bullet_{X \cap Y = Z} \; m_1(X)\, m_2(Y)}{1 - \bullet_{X \cap Y = \emptyset} \; m_1(X)\, m_2(Y)} \qquad \text{for all } Z \in P(\Theta)$$

In our case, the mass obtained, m_c, is also defined over a partition of the basis X due to the fact that given two partitions $\mathbf{Y_a}$ and $\mathbf{Y_b}$ of y, the set $\mathbf{Y_c}$ defined as:

$$\mathbf{Y_c} = \{A \leftrightarrow B: A \in \mathbf{Y_a}, B \in \mathbf{Y_b}\} - \{\emptyset\}$$

is a also a partition.

In table 1 are shown all the non-zero sets (they are, in fact, the elements of the partition Y_c) obtained through Dempster's rule of combination. We considered empty only those sets where the intersection of $[C_1]$ and $[\neg C_1]$ appears. It is clear that $[C_1 \wedge \neg C_1]=\emptyset$. This partition coincides with the basis of the σ-algebra X.

	$P_1 \wedge C_1$	$P_1 \wedge \neg C_1$	$\neg P_1 \wedge C_1$	$\neg P_1 \wedge \neg C_1$
$P_2 \wedge C_1$	$P_1 \wedge P_2 \wedge C_1$	ø	$\neg P_1 \wedge P_2 \wedge C_1$	ø
$P_2 \wedge \neg C_1$	ø	$P_1 \wedge P_2 \wedge \neg C_1$	ø	$\neg P_1 \wedge P_2 \wedge \neg C_1$
$\neg P_2 \wedge C_1$	$P_1 \wedge \neg P_2 \wedge C_1$	ø	$\neg P_1 \wedge \neg P_2 \wedge C_1$	ø
$\neg P_2 \wedge \neg C_1$	ø	$P_1 \wedge \neg P_2 \wedge \neg C_1$	ø	$\neg P_1 \wedge \neg P_2 \wedge \neg C_1$

Table 1

The combined mass results:

$m_c([\neg P_1 \wedge \neg P_2 \wedge \neg C_1]) = .1081$ $m_c([\neg P_1 \wedge P_2 \wedge \neg C_1]) = .0811$

$m_c([\neg P_1 \wedge P_2 \wedge C_1]) = .0480$ $m_c([\neg P_1 \wedge \neg P_2 \wedge C_1]) = .0421$

$m_c([P_1 \wedge P_2 \wedge C_1]) = .1321$ $m_c([P_1 \wedge \neg P_2 \wedge C_1]) = .1157$

$m_c([P_1 \wedge P_2 \wedge \neg C_1]) = .2027$ $m_c([P_1 \wedge \neg P_2 \wedge \neg C_1]) = .2702$

It can be shown easily that the result obtained is a mass that can not be obtained from the initial probability. In fact, this mass is also a probability on X. But this probability is not the original one because there exists x X such that $m_c(x)?P(x)$.

For instance, with $x=[P_1 \wedge \neg P_2 \wedge \neg C_1]$ we have

$$m_c(x)=.2702?P(x)=.2222$$

3. An alternative combination function

In example 2 we have shown that although the masses m_1 and m_2 are both more general than a mass m, the combined mass is not always more general than m. Due to the fact that if m_1 and m_2 have been built from a single P suffering from a process of information losing, it is reasonable that their combined mass can also be built from that P. We introduce a combination function that satisfies this condition. Some of the properties of this function are outlined in section 4.

The function MM_c, defined below, combines m_1 and m_2 and returns a new mass m_c.

```
L0    Function MM_c (m_1, m_2: mass): m_c:mass
L1    var card, k: integer; X,Y,Z: set
L2    for card:=1 to card(θ) do
L3          for all X s.t. XΠθ & (m_1(X)?0 o m_2(X)?0) & Cardinality(X)=card
fer
L4           m_c(X) = max (m_1(X), m_2(X)): k:=0
L5           if m_c(X)=m_1(X) then k:=2; m_1(X):=0
L6           if m_c(X)=m_2(X) then k:=1; m_2(X):=0
L7           α:=m_c(X)-m_k(X)
L8           if α?0 then
L9              acu := 0; Z := ø
L10             for all Y s.t. (X ↔ Y ? ø) & m_k(Y)?0 do
L11                m_k'(Y) = m_k(Y) - α
L12                m_k''(Y)= max(m_k'(Y),0)
L13                Z:=Z ≈ Y
L14                acu := acu + m_k(Y) - m_k''(Y)
L15             end for all
L16             if acu=α then
L17                m_k'(Z)=m_k(Z) + acu - α
L18                m'_k(X) :=m_k(X) + α
L19             else ERROR: BOTH MASSES FROM DIFFERENT PROBABILITIES
L20             end if
L21          end if
L22       end for all
L23    end for
L24    return m_c
L25   end MM_c.
```

This function determines the mass corresponding to the combined function (m_c) incrementally. In the first iteration of loop L2-L23, we get the mass for all X (θ) such that their cardinality is one. Then, we obtain the mass for all sets with cardinality two, and so on... In general, in the i-th iteration of loop L2-L23, the function settles the mass corresponding to those X (θ) with cardinality i.

When a subset X of θ is chosen in L3, the function defines $m_c(X)=\max(m_1(X),m_2(X))$ (L4). We define $m_c(X)$ in this way due to the fact that m_1 and m_2 come from the same probability, so they satisfy:

$$P(X) = \max(Bel_1(X),Bel_2(X))$$

Besides, when $m_c(X)$ is defined, $m_c(x_0)$ is already set for all $x_0 \wp X$ and m_1 and m_2 have been modified to satisfy $m_1(x_0)=m_2(x_0)=m_c(x_0)$. Therefore, we have

$$Bel_1(X) = ?_{x_0 \wp X}\, m_c(x_0) + m_1(X)$$
$$Bel_2(X) = ?_{x_0 \wp X}\, m_c(x_0) + m_2(X)$$

So, in order to get

$$Bel_c(X)=\max(Bel_1(X),Bel_2(X))$$

we set $m_c(X) = \max(m_1(X),m_2(X))$.

Once $m_c(X)=\max(m_1(X),m_2(X))$, one of the mass, say m_1, satisfies the condition $m_c(X)=m_1(X)$ but the other, m_2, does not. In order that m_2 satisfies the equality $m_c(X)=m_2(X)$, $m_2(X)$ has to be increased $\alpha=m_c(X)-m_2(X)$ and some $m_2(Y)$ decreased. Because for all $x_0 \wp X$ $m_1(x_0)=m_2(x_0)$ and there is a probability distribution from which m_1 and m_2 have been built through information losing function, then the mass α has to be in sets with a non empty intersection and with cardinality greater or equal than the cardinality of X.

Because the lost mass α is in some of the Y such that $X \leftrightarrow Y ?0$ but we do not know which, we extract the mass α to each of these sets ($m'_2(Y)=m_2(Y)-\alpha$, L11). The variable acu accumulates the extracted mass. Once the mass is in acu $m_2(X)$ can be correctly defined ($m'_2(X)=m_2(X)+\alpha$, L18). Then the remaining mass, acu-α, has to be redistributed, but because as before we do not know to which one of the sets Y have to be returned, we return it to the union set Z ($m'_2(Z)=m_2(Z)$ + (acu-α) where $Z=\approx_{Y \leftrightarrow X ?\emptyset} Y$, L17). The set Z is defined incrementally in line L13. In each step of the loop L10-L15 we add to the set Z a set Y such that $X \leftrightarrow Y ?\emptyset$ and $m(Y)?0$. To sum it up, in this way, we have transferred acu-α mass that was associated to the sets Y such that $X \leftrightarrow Y ?\emptyset$ to the set Z.

We have also to underline the following considerations that have not been considered yet to simplify the description of the algorithm:

• When mass is subtracted to each Y such that $Y \leftrightarrow X ? \emptyset$ it is possible that $m_2(Y)<\alpha$ in this case instead of subtracting α we subtract only the quantity $m_2(Y)$ and therefore $m_2"(Y)$ would be 0. (L12).

• If `acu` does not accumulate enough mass (that is, if `acu`$<\alpha$, L16) when m_2 has to be increased, there is an inconsistency due to the fact that both masses can not be defined from the same probability (L19).

• The outer loop (L2-L23) forces the chosen sets to have an increasing cardinality. In this way, in m_c, mass is transferred from sets with a lower cardinality to sets with a greater cardinality.

4. Properties of the combination function

In this section some of the conditions given by Smets[9] about Dempster's rule of combination are studied for the function introduced above.

The function satisfies unanimity, that is, given two masses m_1, m_2 such that $m_1=m_2$ then $m_c=m_1=m_2$, compositionality ($bel_c(A)$ is a function of A, bel_1 and bel_2) and, depending on the selection function of X in L3, symmetry (for instance it is satisfied if at each step we choose the set X with a minimum m(X), or the smaller X in lexicographic order or any other predefined order, ...). It has to be said that with different selection functions in L3 different combined mass results. Neither associativity nor autofunctionality ($m_c(A)$ is independent of the masses given by m_1 and m_2 to propositions $B \wp \neg A$) are satisfied. Conditioning[9], that is, if m_2 is such that $m_2(B)=1$ then

$$m_c(A) = \begin{cases} ?_{C \to \neg B}\, m_1(A \vee C) & \text{for all } A \to B \\ 0 & \text{otherwise} \end{cases}$$

is neither satisfied, although it does if lines L1.1-L1.2 given below are included.

```
L1.1 O1=≈A∏θ:m1(A)?0 A
L1.2 O2=≈A∏θ:m2(A)?0 A
L1.3 m'1= 0: for all A℘θ do m'1(A↔O2) :=m'1(A↔O2) + m1(A) end for all
L1.4 m'2= 0: for all A℘θ do m'2(A↔O1) :=m'2(A↔O1) + m2(A) end for all
```

The MM_c function satisfies the property of building a mass obtainable from a probability when mass m_1 and m_2 have been defined from that probability. However, in general, belief and plausibility functions associated to the combined function do not satisfy

$$Bel_c(X) = \max(Bel_1(X),Bel_2(X))$$
$$Pl_c(X) = \min(Pl_1(X),P_2(X))$$

This is due to the fact that L17 can reduce the Bel of some set Y transferring mass assigned to Y to a bigger one Z (i.e., Y $\wp$ Z). In the same way plausibility of a set X=Z$\leftrightarrow\neg$Y increases when mass attached to Y is transferred to Z.

We finish this section giving the results of applying this combination function to masses m_1 and m_2 in example 2 (the selection function of L3 was to choose at each step the set X that has a minimum m(X)):

$m_C([P_1 \wedge C_1]) = .1103$ $\quad$ $m_C([\neg P_1 \wedge C_1]) = .1103$

$m_C([P_1 \wedge \neg C_1]) = .1672$ $\quad$ $m_C([\neg P_1 \wedge \neg C_1]) = .1666$

$m_C([P_2 \wedge C_1]) = .1118$ $\quad$ $m_C([\neg P_2 \wedge C_1]) = .0833$

$m_C([P_2 \wedge \neg C_1]) = .0828$ $\quad$ $m_C([\neg P_2 \wedge \neg C_1]) = .1664$

It can be seen that belief and plausibility function define in this case an interval in which probability is enclosed. For instance, for the set $x_8=[P_1 \wedge \neg P_2 \wedge \neg C_1]$ we have:

$$Bel_c([P_1 \wedge \neg P_2 \wedge \neg C_1])=0$$
$$Pl_c([P_1 \wedge \neg P_2 \wedge \neg C_1])=.3336$$

and that,

$$P([P_1 \wedge \neg P_2 \wedge \neg C_1]) \in [Bel_c([P_1 \wedge \neg P_2 \wedge \neg C_1]),Pl_c([P_1 \wedge \neg P_2 \wedge \neg C_1])]$$

5. Conclusions

In this paper we have introduced an interpretation of mass functions in evidence theory based on probability distributions. We have analysed Dempster's rule of combination in this framework and proved that given two masses in this framework it returns an inconsistent mass. We have introduced a new combination function that builds a new function coherent with the interpretation and studied some of the properties of the combination function. We give an example of the application of the function introduced.

Acknowledgement

The help of Professors Claudi Alsina and Lluís Godo is gratefully acknowledged. I would thank the referees for their suggestions and comments. I would also thank the support of the European Community, under the contract VIM: CHRX-CT93-0401 that allowed me to discuss this work with some European partners. Part of this work was completed at the Institut d'Investigació en Intel.ligència Artificial (IIIA, CSIC).

References

1. A. Chateauneuf, "Combination of compatible belief functions and relation of specificity" in *Advances in Dempster-Shafer Theory of evidence*, Wiley, 97-114, 1994.

2. D. Dubois, H. Prade, "A set-theoretic view of belief functions", *Int. J. of General Systems* **12**, 193-226 (1986).

3. D. Dubois, H. Prade, "Evidence, Knowledge, and Belief Functions", *Int. J. of Approx. Reasoning*, **6**, 295-319 (1992).

4. R. Fagin, J. Y. Halpern, "A new approach to updating beliefs", in P.Bonissone, M.Henrion, L.N. Kanal, J.F.Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, Elsevier Science Publishers, pp.347-374, 1991.

5. J. Y. Halpern, R. Fagin, "Two views of belief: belief as generalized probability and belief as evidence", *Artificial Intelligence*, **54**, 275-317 (1992).

6. H. Ichihashi, H. Tanaka, "Jeffrey-like rules of conditioning for Dempster-Shafer theory of evidence", *Int. J. of Approx. Reasoning*, **3**, 143-156 (1989).

7. G. J. Klir, T. A. Folger, *Fuzzy sets, uncertainty, and information*, Prentice-Hall, 1988.

8. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

9. P. Smets, "Belief Functions", pp. 153-286, in P. Smets et al., 1988.

10. P. Smets, "Jeffrey's rule of conditioning generalized to belief functions", in D. Heckerman, A. Mamdani, (Eds), *Proceedings of the 9th conference in Uncertainty in AI'93*, 1993.

11. P. Smets, E. H. Mamdami, D. Dubois, H. Prade, *Non-Standard Logics for Automated Reasoning*, Academic Press, 1988.

12. V. Torra, *Contribució a l'estudi de funcions de síntesi per a la intel.ligencia artificial*, PhD Dissertation, Departament de Llenguatges i Sistemes Informátics, Universitat Politécnica de Catalunya, Barcelona, Maig, 1994.

13. C.G. Wagner, "Generalizing Jeffrey Conditionalization", in Dubois, Wellman, d'Ambrosio, Smets, (Eds), Proceedings of the 8th conference in Uncertainty in AI'92, Morgan Kaufmann, San Mateo, CA, USA, 331-335, 1992.

Figures

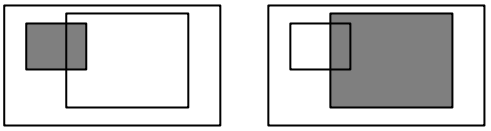
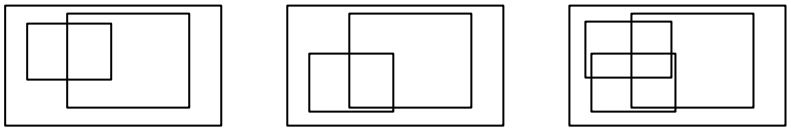


Figure 1. [P_1] and [C_1]



Partition $\mathbf{Y_1}$ Partition $\mathbf{Y_2}$ Basis y

Figure 2