



**HAL**  
open science

## Cumulative and Averaging Fusion of Beliefs

Audun Jøsang, Javier Diaz, Maria Rifqi

► **To cite this version:**

Audun Jøsang, Javier Diaz, Maria Rifqi. Cumulative and Averaging Fusion of Beliefs. *Information Fusion*, 2010, 11 (2), pp.192-200. hal-01078654

**HAL Id: hal-01078654**

**<https://hal.science/hal-01078654>**

Submitted on 29 Oct 2014

**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.

# Cumulative and Averaging Fusion of Beliefs

Audun Jøsang<sup>a</sup> Javier Diaz<sup>b</sup> Maria Rifqi<sup>b</sup>

<sup>a</sup>*UNIK Graduate Center, University of Oslo, Norway  
josang @ unik.no*

<sup>b</sup>*CNRS - UMR 7606 LIP6, Université Paris VI, France  
{Javier.Diaz, Maria.Rifqi} @ lip6.fr*

---

## Abstract

The problem of fusing beliefs in the Dempster-Shafer belief theory has attracted considerable attention over the last two decades. The classical Dempster's Rule has often been criticised, and many alternative rules for belief combination have been proposed in the literature. We show that it is crucial to consider the nature of the situation at hand and to select the appropriate fusion operator as a function thereof. In this paper we present the cumulative rule and the averaging rule of belief fusion, which represent generalisations of the subjective logic consensus operator for independent and dependent opinions respectively. The generalised operators are applicable to the combination of general basic belief assignments (bbas). These rules, which can be directly derived from classical statistical theory, produce results that correspond well with human intuition.

*Key words:* Belief Theory, Subjective Logic, Fusion, Dempster's Rule

---

## 1 Introduction

Belief theory has its origin in a model for upper and lower probabilities proposed by Dempster in 1960. Shafer later used the same fundamental framework as a model for expressing beliefs [1]. The main idea behind the theory of belief functions is to abandon the additivity principle of probability theory, i.e. that the sum of probabilities on all pairwise disjoint possibilities always equals one. Instead belief theory gives observers the ability to assign belief masses to any subset of a state space, i.e. to non-disjoint possibilities including the whole state space itself. The advantage of this approach over classical probabilistic modelling is that uncertainty about subset probabilities, e.g. due to missing evidence, can be explicitly expressed. Uncertainty can for example be expressed by assigning belief mass to the union of singletons, or to the whole state space itself. Consistency is preserved by requiring that the sum of all belief masses always is one. The difference between probability additivity

and belief mass additivity is that for probabilities the states must all be mutually disjoint, whereas for belief masses the states can be overlapping. Shafer's book [1] describes many characteristics of belief functions, but the two main elements are 1) a flexible way of expressing beliefs, and 2) a method for fusing beliefs, commonly known as Dempster's Rule.

There are well known examples where Dempster's rule produces counter-intuitive results, especially in case of strong conflict between the two argument beliefs. Motivated by this observation, numerous authors have proposed alternative methods for fusing beliefs, e.g. [2,3,4,5,6,7,8,9]. An overview of belief fusion rules that have been proposed in the literature is provided in [10]. These rules express different behaviours with respect to the results of fusing beliefs, but have in general been proposed with the same basic purpose in mind: to combine two beliefs into a single belief that reflects the two possibly conflicting beliefs in a fair and equal way.

However, situations that may seem similar at first glance can be very different when examined more closely, and will therefore require different operators. For example, the right operator for modelling the strength of a chain is the principle of the weakest link. The right operator for modelling the strength of a relay swimming team is the average strength of each member. Applying the weakest swimmer principle to assess the overall strength of the relay team might represent an approximation, but it is incorrect in general, and would give very unreliable predictions. Similarly, applying the principle of average strength of the links in a chain to assess the overall strength of the chain might represent an approximation, but it is incorrect in general and could be fatal if life depended on it. The observation of these simple examples tells us that it is crucial to properly understand the situation at hand in order to find the correct model for analysing it.

In our view researchers in the belief theory community have not paid sufficient attention to analysing the actual situation to be modelled in order to determine whether Dempster's rule or any other rule can be correctly applied. Instead, researchers have often tried to assess the merits of fusion operators based solely on algebraic properties, such as commutativity and associativity, which do not represent sufficient criteria for judging an operator's applicability to a particular situation. For example, no matter how solid the theoretic basis for the average operator is, it will never represent a correct model for the strength of a chain.

In this article we present two belief fusion operators called the cumulative and averaging rules of combining beliefs. These rules do not represent an alternative or competitor to Dempster's rule because they are applicable in different types of situations than Dempster's rule.

The terms *cumulative rule* and *averaging rule* have explicitly been chosen in order to have descriptive names for the types of situations to which they apply. The cumulative rule of combination is applicable to situations where independent be-

belief functions are combined as a function of accumulation of the evidence. The averaging rule of combination is applicable to situations where dependent belief functions are combined as a function of the average of the evidence. This is different from Dempster's rule, where beliefs are combined by normalised conjunction. It can be shown that non-normalised conjunction corresponds to multiplication of belief functions, and normalised conjunction corresponds to a series of stochastic constraints [11] represented by the argument belief functions. This difference will be illustrated by examples below.

The cumulative and averaging rules can be applied to the combination of general belief functions, and represents a generalisation of cumulative and averaging fusion of opinions in subjective logic[5,12,13,14]. The cumulative rule itself is then simply equivalent to the additive combination of Dirichlet distributions, and the averaging rule is simply equivalent to the average of Dirichlet distributions. This also provides an equivalence mapping between Dirichlet distributions and belief functions [15]. In this way, belief fusion in the form of the cumulative and averaging rules is firmly based on classical statistical theory.

## 2 Theory of Belief Functions

In this section several concepts of the Dempster-Shafer theory of evidence [1] are recalled in order to introduce notations used throughout the article. The term *frame of discernment* is used in belief theory with the equivalent meaning of state space from probability theory. A frame denoted by  $\Theta = \{\theta_i; i = 1, \dots, k\}$  represents a finite set of exhaustive and exclusive possible values for a state variable of interest. The terms *states*, *elements* or *outcomes* will be used to denote the state variable. The frame can for example be the set of six possible outcomes of throwing a dice, so that the (unknown) outcome of a particular instance of throwing the dice becomes the state variable. A bba (basic belief assignment<sup>1</sup>), denoted by  $m$ , is defined as a belief distribution function from the powerset  $2^\Theta$  to  $[0, 1]$  satisfying:

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{x \subseteq \Theta} m(x) = 1. \quad (1)$$

Values of a bba are called *belief masses*. Each subset  $x \subseteq \Theta$  such that  $m(x) > 0$  is called a focal element of  $\Theta$ .

From a bba  $m$  can be derived a set of non-additive belief functions  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$ , defined as:

$$\text{Bel}(x) \triangleq \sum_{\emptyset \neq y \subseteq x} m(y) \quad \forall x \subseteq \Theta. \quad (2)$$

---

<sup>1</sup> Called *basic probability assignment* in [1], and *Belief Mass Assignment* (BMA) in [16,5].

The quantity  $\text{Bel}(x)$  can be interpreted as a measure of one's total belief committed to the hypothesis that  $x$  is true. Note that functions  $m$  and  $\text{Bel}$  are in one-to-one correspondence [1] and can be seen as two facets of the same piece of information.

A few special classes of bba can be mentioned. A vacuous bba has  $m(\Theta) = 1$ , i.e. no belief mass committed to any proper subset of  $\Theta$ . A *Bayesian* bba is when all the focal elements are singletons, i.e. one-element subsets of  $\Theta$ . If all the focal elements are nestable (i.e. linearly ordered by inclusion) then we speak about a *consonant* bba. A *dogmatic* bba is defined by Smets as a bba for which  $m(\Theta) = 0$ . Let us note, that trivially, every Bayesian bba is dogmatic.

The powerset of the frame  $\Theta$  is defined as  $2^\Theta = \{x_i; x_i \subseteq \Theta\}$ . We will define the state space  $X$  as a special representation of the powerset of  $\Theta$ . More precisely, the set  $X$  is defined as:

$$X = 2^\Theta \setminus \Theta \quad (3)$$

meaning that all proper subsets of  $\Theta$  are elements of  $X$ . By considering  $X$  as a frame in itself, a general bba on  $\Theta$  becomes a particular bba on  $X$  called a *Dirichlet bba*. A belief mass on a proper subset of  $\Theta$  then becomes a belief mass on a singleton of  $X$ . In addition we define  $m(X) = m(\Theta)$ . In this way, a Dirichlet bba on  $X$  derived from a general bba on  $\Theta$ , is characterised by having mutually disjoint focal elements, except the whole frame  $X$  itself. This is formally defined as follows.

**Definition 1 (Dirichlet bba)** *Let  $X$  be a frame of discernment. A bba where the only focal elements are  $X$  and/or singletons of  $X$ , is called a Dirichlet belief mass distribution function, or Dirichlet bba for short.*

Fig.1 below illustrates a possible Dirichlet bba on  $X$ , where the shaded circles around singletons and the shaded ellipse around  $X$  represent belief masses on those subsets. The focal elements in this example are  $X$ ,  $x_1$ ,  $x_2$  and  $x_4$ .

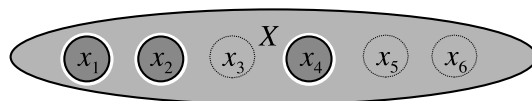


Figure 1. Example Dirichlet bba, characterised by belief masses on singletons and  $X$

The number of elements in  $X$  is  $|X| = 2^{|\Theta|} - 2$  when excluding  $\emptyset$ . For example, Fig.1 illustrates  $X$  as having cardinality 6, meaning that it is the powerset of a ternary frame of discernment. The subsets of  $\Theta$  and the elements of  $X$  carry the same belief masses, so is natural to make the correspondence as simple as possible. The following example defines a possible correspondence between subsets of  $\Theta$

and elements of  $X$ :

$$\begin{aligned}
 x_1 &\longleftrightarrow \theta_1, & x_4 &\longleftrightarrow \theta_1 \cup \theta_2, & X &\longleftrightarrow \Theta . \\
 x_2 &\longleftrightarrow \theta_2, & x_5 &\longleftrightarrow \theta_1 \cup \theta_3, & & \\
 x_3 &\longleftrightarrow \theta_3, & x_6 &\longleftrightarrow \theta_2 \cup \theta_3, & & 
 \end{aligned} \tag{4}$$

The number of focal elements of a Dirichlet bba on  $X$  can be at most  $|X| + 1$ , which happens when every element as well as  $X$  is a focal element.

The name ‘‘Dirichlet’’ bba is used because bbas of this type are equivalent to Dirichlet probability density functions under a specific mapping. A bijective mapping between Dirichlet bbas and Dirichlet probability density functions is defined in [15], and is also described in Sec.4 below. Our approach is different from that of Walley’s Imprecise Dirichlet Model [17] which interprets the situations of frame ignorance and frame certainty in the Dirichlet model as lower and upper probability in the belief model.

### 3 The Dirichlet Multinomial Model

The cumulative and averaging rules of combination, to be described in detail in the following sections, are firmly rooted in the classical Bayesian inference theory, and are equivalent to addition and averaging of multinomial observations respectively. For self-containment, we briefly outline the Dirichlet multinomial model below, and refer to [18] for more details.

#### 3.1 The Dirichlet Distribution

We are interested in knowing the probability distribution over the disjoint elements of a frame based on observed instances of these elements. In case of binary frames, it is determined by the Beta distribution. In the general case it is determined by the Dirichlet distribution, which describes the probability distribution over a  $k$ -component random variable  $p(x_i)$ ,  $i = 1, \dots, k$  with sample space  $[0, 1]^k$ , subject to the additivity criterion:

$$\sum_{i=1}^k p(x_i) = 1 . \tag{5}$$

Note that for any sample from a Dirichlet random variable, it is sufficient to determine values for  $p(x_i)$  for any  $k - 1$  elements  $i$  of  $\{1, \dots, k\}$ , as this uniquely determines the value of the remaining variable.

The Dirichlet distribution with prior captures a sequence of observations of the  $k$  possible outcomes with  $k$  positive real observation variables  $r(x_i)$ ,  $i = 1 \dots k$ , each corresponding to one of the possible outcomes. In order to have a compact notation we define a vector  $\vec{p} = \{p(x_i) \mid 1 \leq i \leq k\}$  to denote the  $k$ -component random probability variable, and a vector  $\vec{r} = \{r_i \mid 1 \leq i \leq k\}$  to denote the  $k$ -component random observation variable  $[r(x_i)]_{i=1}^k$ .

In order to distinguish between the *a priori* information and the *a posteriori* evidence, the Dirichlet distribution must be expressed with prior information represented as a base rate vector  $\vec{a}$  over the frame as well as the non-informative prior weight  $C$ . Eq.(6) represents this Dirichlet Distribution with Prior.

$$f(\vec{p} \mid \vec{r}, \vec{a}) = \frac{\Gamma\left(\sum_{i=1}^k (r(x_i) + a(x_i)C)\right)}{\prod_{i=1}^k \Gamma(r(x_i) + a(x_i)C)} \prod_{i=1}^k p(x_i)^{r(x_i) + a(x_i)C - 1} \quad (6)$$

The vector  $\vec{p}$  represents first order probability variables over the elements of  $X$  satisfying Eq.(5), whereas the density  $f(\vec{p} \mid \vec{r}, \vec{a})$  represents the probability of specific sets of values for the first-order variables. Since the first-order variables  $\vec{p}$  are continuous, the second-order probability  $f(\vec{p} \mid \vec{r}, \vec{a})$  for any given value of  $p(x_i) \in [0, 1]$  is vanishingly small and therefore meaningless as such. It is only meaningful to compute  $\int_{p_1}^{p_2} f(p(x_i) \mid \vec{r}, \vec{a})$  for a given interval  $[p_1, p_2]$  and level  $x_i$ , or simply to compute the expectation value of  $p(x_i)$ . As will be shown below, this provides a sound mathematical basis for accumulating and averaging evidence.

Given the Dirichlet distribution of Eq.(6), the probability expectation of any of the  $k$  random variables can now be written as:

$$E(p(x_i) \mid \vec{r}, \vec{a}) = \frac{r(x_i) + a(x_i)C}{C + \sum_{t=1}^k r(x_t)}. \quad (7)$$

Eq.(6) is useful, because it allows the determination of the probability distribution with arbitrary amounts of observation evidence, even without any observations.

The *non-informative* prior weight  $C$  is set to  $C = 2$  when a uniform distribution over binary frames is assumed. Selecting a larger value for  $C$  will result in new observations having less influence over the Dirichlet distribution. A distribution is non-informative when it only reflects knowledge of the frame, and does not reflect any observation evidence.

It can be noted that it would be unnatural to require a uniform distribution over arbitrary large frames because it would make the sensitivity to new evidence arbitrarily small. For example, requiring a uniform *a priori* distribution over a frame

of cardinality 100, would force  $C = 100$ . In case an event of interest has been observed 100 times, and no other event has been observed, the derived probability expectation of the event of interest will still only be about  $\frac{1}{2}$ , which would seem totally counter-intuitive. In contrast, when a uniform distribution is assumed in the binary case, and the same 100 observations are analysed, the derived probability expectation of the event of interest would be close to 1, as intuition would dictate. The value of  $C$  determines the approximate number of observations of a particular element in the frame needed to influence the probability expectation value of that element from 0 to 0.5

It can be noted that according to Walley's IDM (Imprecise Dirichlet Model) [17] the upper and lower probability for a state  $x_i$  are obtained by setting  $a(x_i) = 1$  and  $a(x_i) = 0$  respectively. The lower probability is thus based on a zero base rate, and the upper probability is based on a base rate equal to one. The upper and lower probabilities are then interpreted as the upper and lower bounds for the relative frequency of the outcome. While this is an interesting interpretation of the Dirichlet distribution, it can not be taken literally. According to this model, the upper and lower probability values for an outcome  $x_i$  are defined as:

$$\text{IDM Upper probability: } \bar{P}(x_i) = \frac{r(x_i) + C}{C + \sum_{i=1}^k r(x_i)} \quad (8)$$

$$\text{IDM Lower probability: } \underline{P}(x_i) = \frac{r(x_i)}{C + \sum_{i=1}^k r(x_i)} \quad (9)$$

where  $r(x_i)$  is the number of observations of outcome  $x_i$ , and  $C$  is the weight of the non-informative prior probability distribution. It can easily be shown that these values can be misleading. For example, assume an urn containing 9 red balls and 1 black ball, meaning that the relative frequencies of red and black balls are  $p(\text{red}) = 0.9$  and  $p(\text{black}) = 0.1$ . The *a priori* weight is set to  $C = 2$ . Assume further that an observer picks one ball which turns out to be black. According to Eq.(9) the lower probability is then  $\underline{P}(\text{black}) = \frac{1}{3}$ . It would be incorrect to literally interpret this value as the lower bound for the relative frequency because it obviously is greater than the actual relative frequency of black balls. This example shows that there is no guarantee that the actual probability of an event is inside the interval defined by the upper and lower probabilities as described by the IDM. The terms "upper" and "lower" must therefore be interpreted as rough terms for imprecision, and not as absolute bounds.

The traditional approach in Bayesian analysis is to interpret the combination of the base rate vector  $\vec{a}$  and the *a priori* weight  $C$  as representing specific *a priori* information such as provided by a domain expert.



### 3.2 Visualising Dirichlet Distributions

Visualising Dirichlet distributions is challenging because it is a density function over  $k - 1$  dimensions, where  $k$  is the frame cardinality. For this reason, Dirichlet distributions over ternary frames are the largest that can be easily visualised on paper.

With  $k = 3$ , the probability distribution has 2 degrees of freedom, and the additivity equation  $p(x_1) + p(x_2) + p(x_3) = 1$ , which is an instantiation of Eq.(5), defines a triangular plane as illustrated in Fig.2.

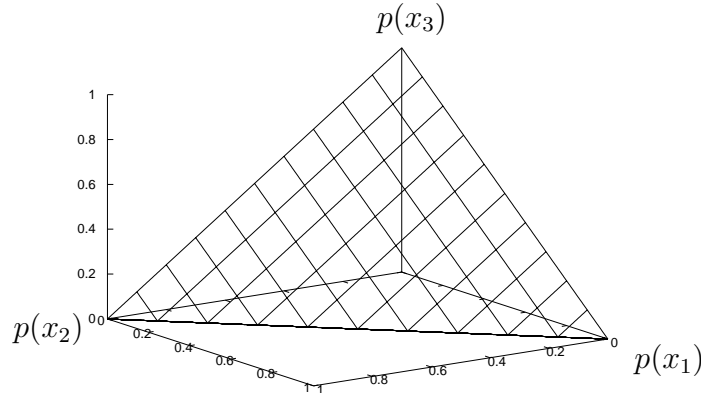


Figure 2. Triangular plane

In order to visualise probability density over the triangular plane, it is convenient to lay the triangular plane horizontally in the X:Y-plane, and visualise the density dimension along the Z-axis.

Let us consider the example of an urn containing balls that can have the inscription  $x_1$ ,  $x_2$  or  $x_3$  (i.e.  $k = 3$ ). Let us first assume that no other information than the cardinality of the frame is available, meaning that  $r(x_1) = r(x_2) = r(x_3) = 0$ , and  $a(x_1) = a(x_2) = a(x_3) = 1/3$ . Then Eq.(7) dictates that the expected probability of picking a ball of any type is equal to the base rate probability, which is  $\frac{1}{3}$ . The *a priori* Dirichlet density function is illustrated in Fig.3.a.

Let us now assume that an observer has picked  $x_1$  6 times,  $x_2$  once and  $x_3$  once, i.e.  $r(x_1) = 6$ ,  $r(x_2) = 1$ ,  $r(x_3) = 1$ . By assuming a prior weight  $C = 2$  as before, the *a posteriori* expected probability of picking a ball with  $x_1$  can be computed as  $E(p(x_1)) = \frac{2}{3}$ . The *a posteriori* Dirichlet density function is illustrated in Fig.3.b.

We reuse the example of the urn containing balls with the inscriptions  $x_1$ ,  $x_2$ , and  $x_3$ , but this time we assume a binary partition of  $X$  into  $\{x_1, \bar{x}_1\}$ , i.e. where  $\bar{x}_1 = x_2 \cup x_3$ . The base rate of picking a ball with inscription  $x_1$  is set to  $a(x_1) = \frac{1}{3}$  as before because the urn still contains balls of three different types, and the grouping of states  $x_2$  and  $x_3$  is purely technical.

Let us again assume that an observer has picked (with return)  $x_1$  6 times, and

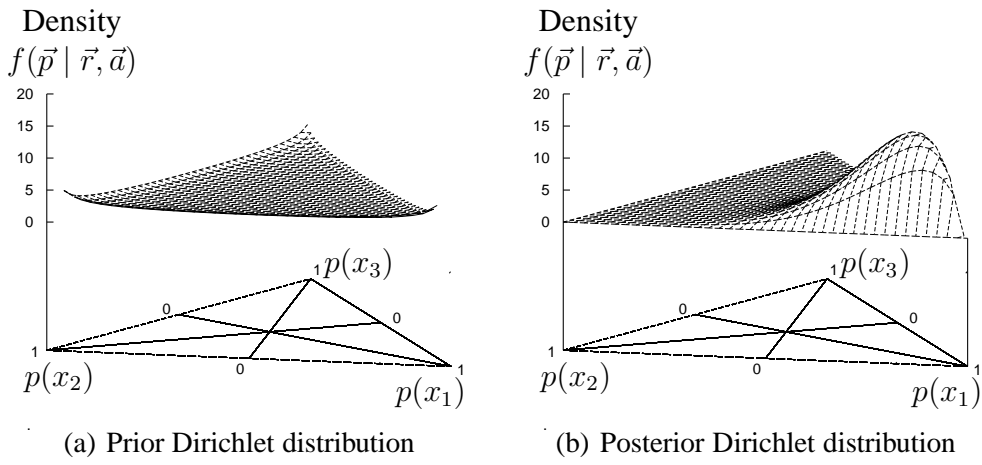


Figure 3. Visualising *a priori* and *a posteriori* Dirichlet distributions

$\{x_2 \text{ or } x_3\}$  twice, i.e.  $r(x_1) = 6$  and  $r(\bar{x}_1) = 2$ .

Since the frame has been reduced to binary, the Dirichlet distribution is reduced to a Beta distribution which is simple to visualise. The *a priori* and *a posteriori* density functions are illustrated in Fig.3.2.

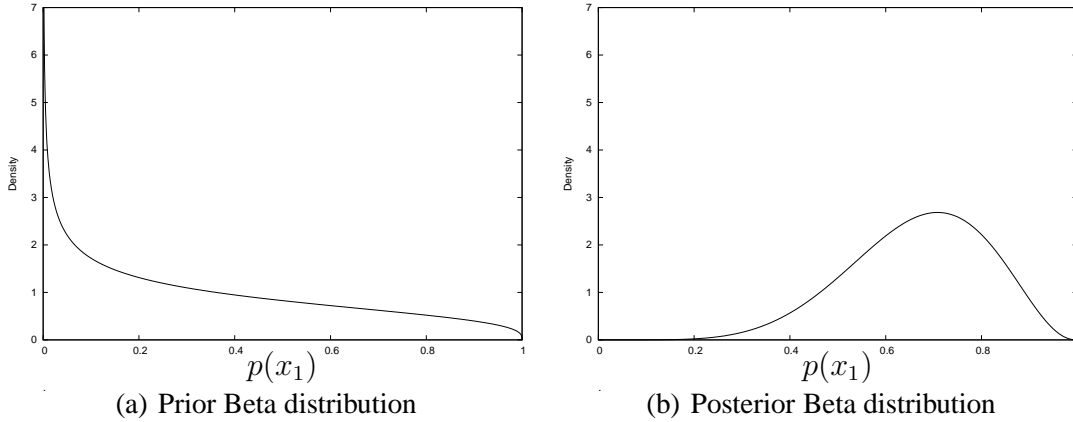


Figure 4. Visualising prior and posterior Beta distributions

The *a posteriori* expected probability of picking a ball with inscription  $x_1$  can be computed with Eq.(7) as  $E(p(x_1)) = \frac{2}{3}$ , which is the same as before the coarsening, as illustrated in Fig.3.b This shows that the coarsening does not influence the probability expectation value of specific events.

#### 4 Mapping Between Dirichlet Distribution and Belief Distribution Functions

In this section we will define a bijective mapping between Dirichlet probability distributions described in Sec.3, and Dirichlet bbas described in Sec.2.

Let  $X = \{x_i; i = 1, \dots, k\}$  be a frame where each singleton represents a possible outcome of a state variable. It is assumed that  $X$  represents the powerset of a frame of discernment  $\Theta$  according to Eq.(eq:powerset). Let  $m$  be a general bba on  $\Theta$  and therefore a Dirichlet bba on  $X$ , and let  $f(\vec{p} \mid \vec{r}, \vec{a})$  be a Dirichlet distribution over  $X$ .

The bijective mapping between  $m$  and  $f(\vec{p} \mid \vec{r}, \vec{a})$  is based on simple intuitive criteria specified below. The mathematical expressions for the bijective mapping can be directly derived from the criteria.

As first criterion we require equality between the pignistic probability values  $\wp(x_i)$  derived from  $m$ , and the probability expectation values  $E(p(x_i))$  of  $f(\vec{p} \mid \vec{r}, \vec{a})$ . For all  $x_i \in X$ , this constraint is expressed as:

$$\wp(x_i) = E(p(x_i) \mid \vec{r}, \vec{a}) \iff m(x_i) + a(x_i)m(X) = \frac{r(x_i) + a(x_i)C}{C + \sum_{t=1}^k r(x_t)} \quad (10)$$

We also require  $m(x_i)$  to be an increasing function of  $r(x_i)$ , and  $m(X)$  to be a decreasing function of  $\sum_{t=1}^k r(x_t)$ . In other words, the more evidence in favour of a particular outcome, the greater its belief mass. Furthermore, the less evidence available in general, the more vacuous the bba (i.e. the greater  $m(X)$ ). These intuitive requirements together with Eq.(10) imply the bijective mapping defined by Eq.(11).

For  $m(X) \neq 0$  :

$$\left\{ \begin{array}{l} m(x_i) = \frac{r(x_i)}{C + \sum_{t=1}^k r(x_t)} \\ m(X) = \frac{C}{C + \sum_{t=1}^k r(x_t)} \end{array} \right\} \iff \left\{ \begin{array}{l} r(x_i) = \frac{Cm(x_i)}{m(X)} \\ 1 = m(X) + \sum_{i=1}^k m(x_i) \end{array} \right. \quad (11)$$

Next we consider the case of zero uncertainty. In case  $m(X) \rightarrow 0$ , then necessarily  $\sum_{i=1}^k m(x_i) \rightarrow 1$ , and  $\sum_{i=1}^k r(x_i) \rightarrow \infty$ , meaning that at least some, but not necessarily all, of the evidence parameters  $r(x_i)$  are infinite.

We define  $\eta(x_i)$  as the the relative degree of infinity between the corresponding infinite evidence parameters  $r(x_i)$  such that  $\sum_{i=1}^k \eta(x_i) = 1$ . When infinite evidence parameters exist, any finite evidence parameter  $r(x_i)$  can be assumed to be zero in any practical situation because it will have  $\eta(x_i) = 0$ , i.e. it will carry zero weight relative to the infinite evidence parameters. This leads to the bijective mapping defined by Eq.(12).

For  $m(X) = 0$  :

$$\left\{ \begin{array}{l} m(x_i) = \eta(x_i) \\ m(X) = 0 \end{array} \right. \iff \left\{ \begin{array}{l} r(x_i) = \eta(x_i) \sum_{t=1}^k r(x_t) = \eta(x_i) \infty \\ 1 = \sum_{t=1}^k m(x_t) \end{array} \right. \quad (12)$$

In case  $\eta(x_i) = 1$  for a particular evidence parameter  $r(x_i)$ , then  $r(x_j) = \infty$  and all the other evidence parameters are finite. In case  $\eta(x_j) = 1/l$  for all  $j = 1 \dots l$ , then all the evidence parameters are all equally infinite.

## 5 Deriving the Cumulative Rule of Belief Fusion

The cumulative rule is equivalent to *a posteriori* updating of Dirichlet distributions, and is based on the bijective mapping described in the previous section.

Assume a process with possible outcomes defined by the frame of discernment  $\Theta$ . Let  $X = \{x_i; i = 1, \dots, k\}$  represent the powerset of  $\Theta$  according to Eq.(3). Let agents  $A$  and  $B$  observe the outcomes of the process over two separate time periods, assuming that they apply the same base rate vector  $\vec{a}$  to  $\Theta$ . Let the two observers' respective observations be expressed as  $\vec{r}_A$  and  $\vec{r}_B$ . The Dirichlet distributions resulting from these separate bodies of evidence can be expressed as  $f(\vec{p} | \vec{r}_A, \vec{a})$  and  $f(\vec{p} | \vec{r}_B, \vec{a})$

The fusion of these two bodies of evidence simply consists of vector addition of  $\vec{r}_A$  and  $\vec{r}_B$ . In terms of Dirichlet distributions, this can be expressed as:

$$\begin{aligned} f(\vec{p} | \vec{r}_{A \diamond B}, \vec{a}) &= f(\vec{p} | \vec{r}_A, \vec{a}) \oplus f(\vec{p} | \vec{r}_B, \vec{a}) \\ &= f(\vec{p} | (\vec{r}_A + \vec{r}_B), \vec{a}) . \end{aligned} \quad (13)$$

The symbol “ $\diamond$ ” denotes the cumulative fusion of two observers  $A$  and  $B$  into a single imaginary observer denoted as  $A \diamond B$ . All the necessary elements are now in place for presenting the cumulative rule for belief fusion.

### Theorem 1 (Cumulative Fusion Rule)

Let  $m_A$  and  $m_B$  be bbas respectively held by agents  $A$  and  $B$  over the same frame of discernment  $\Theta$ . Let  $X = \{x_i; i = 1, \dots, k\}$  represent the powerset of  $\Theta$  according to Eq.(3). Let  $m_{A \diamond B}$  be the bba such that:

Case I: For  $m_A(\Theta) \neq 0 \vee m_B(\Theta) \neq 0$ :

$$\begin{cases} m_{A \diamond B}(x_i) = \frac{m_A(x_i)m_B(\Theta) + m_B(x_i)m_A(\Theta)}{m_A(\Theta) + m_B(\Theta) - m_A(\Theta)m_B(\Theta)} \\ m_{A \diamond B}(\Theta) = \frac{m_A(\Theta)m_B(\Theta)}{m_A(\Theta) + m_B(\Theta) - m_A(\Theta)m_B(\Theta)} \end{cases} \quad (14)$$

Case II: For  $m_A(\Theta) = 0 \wedge m_B(\Theta) = 0$ :

$$\begin{cases} m_{A \diamond B}(x_i) = \gamma^A m_A(x_i) + \gamma^B m_B(x_i) \\ m_{A \diamond B}(\Theta) = 0 \end{cases} \quad \text{where} \quad \begin{cases} \gamma^A = \lim_{\substack{m_A(\Theta) \rightarrow 0 \\ m_B(\Theta) \rightarrow 0}} \frac{m_B(\Theta)}{m_A(\Theta) + m_B(\Theta)} \\ \gamma^B = \lim_{\substack{m_A(\Theta) \rightarrow 0 \\ m_B(\Theta) \rightarrow 0}} \frac{m_A(\Theta)}{m_A(\Theta) + m_B(\Theta)} \end{cases} \quad (15)$$

Then  $m_{A \diamond B}$  is called the cumulatively fused bba of  $m_A$  and  $m_B$ , representing an imaginary agent  $[A \diamond B]$ 's bba, as if that agent represented both  $A$  and  $B$ . By using the symbol ' $\oplus$ ' to designate this belief operator, we define  $m_{A \diamond B} \equiv m_A \oplus m_B$ .

In Case II,  $\gamma^A$  and  $\gamma^B$  are relative weights satisfying  $\gamma^A + \gamma^B = 1$ . The default values are  $\gamma^A = \gamma^B = 0.5$ .

The proof below provides details about how the expression for the cumulative rule is derived.

**Proof 1** Let  $m_A$  and  $m_B$  be Dirichlet bbas. The mapping from Dirichlet bbas to Dirichlet distributions is done according to the right hand sides of Eq.(11) and Eq.(12), expressed as:

$$\begin{aligned} m_A &\longmapsto f(\vec{p} \mid \vec{r}_A, \vec{a}) \\ m_B &\longmapsto f(\vec{p} \mid \vec{r}_B, \vec{a}) \end{aligned} \quad (16)$$

These Dirichlet distributions can now be fused according to Eq.(13), expressed as:

$$f(\vec{p} \mid \vec{r}_A, \vec{a}) \oplus f(\vec{p} \mid \vec{r}_B, \vec{a}) = f(\vec{p} \mid (\vec{r}_A + \vec{r}_B), \vec{a}) \quad (17)$$

Finally, the result of Eq.(17) is mapped back to a (cluster) Dirichlet bba again using Eq.(11). This can be written as:

$$f(\vec{p} \mid (\vec{r}_A + \vec{r}_B), \vec{a}) \longmapsto m_{A \diamond B} \quad (18)$$

By inserting the full expressions for the parameters in Eqs.(16), (17) and (18), the

expressions of Eqs.(14) and (15) in Theorem 1 emerge.

It can be verified that the cumulative rule is commutative, associative and non-idempotent. The non-idempotence means that cumulative fusion of two equal argument bbas will result in a different bba. In Case II of Theorem 1 (Bayesian bbas, which can also be described as dogmatic Dirichlet bbas), the associativity depends on the preservation of relative weights of intermediate results, which requires the additional weight variable  $\gamma$ . In this case, the cumulative rule is equivalent to the weighted average of probabilities.

It is interesting to notice that the expression for the cumulative rule is independent of the *a priori* weight  $C$ . That means that the choice of a uniform Dirichlet distribution in the binary case in fact only influences the mapping between Dirichlet distributions and Dirichlet bbas, not the cumulative rule itself. This shows that the cumulative rule is firmly based on classical statistical analysis, and not dependent on arbitrary choices of prior.

The consensus operator [5,12,16] is a special case of the cumulative rule, and emerges directly from Theorem 1 by assuming a binary frame.

## 6 Deriving the Averaging Rule of Belief Fusion

The averaging rule is equivalent to averaging the evidence of Dirichlet distributions, and is based on the bijective mapping between the belief and evidence notations described in Sec.4.

Assume a process with possible outcomes defined by the frame of discernment  $\Theta$ . Let  $X = \{x_i; i = 1, \dots, k\}$  represent the powerset of  $\Theta$  according to Eq.(3). Let two sensors  $A$  and  $B$  observe the same outcomes of the process, i.e. over the same time period, assuming that they apply the same base rate vector  $\vec{a}$  to  $\Theta$ . Let the two sensors' respective observations be expressed as  $\vec{r}_A$  and  $\vec{r}_B$ . The Dirichlet distributions resulting from these separate bodies of evidence can be expressed as  $f(\vec{p} | \vec{r}_A, \vec{a})$  and  $f(\vec{p} | \vec{r}_B, \vec{a})$

The averaging fusion of these two bodies of evidence simply consists of the average vector value of  $\vec{r}_A$  and  $\vec{r}_B$ . In terms of Dirichlet distributions, this is expressed as:

$$\begin{aligned} f(\vec{p} | \vec{r}_{A \diamond B}, \vec{a}) &= f(\vec{p} | \vec{r}_A, \vec{a}) \oplus f(\vec{p} | \vec{r}_B, \vec{a}) \\ &= f(\vec{p} | (\frac{\vec{r}_A + \vec{r}_B}{2}), \vec{a}) . \end{aligned} \tag{19}$$

The symbol " $\diamond$ " denotes the averaging fusion of two observers  $A$  and  $B$  into a single imaginary observer denoted as  $A \diamond B$ .

## Theorem 2 Averaging Fusion Rule

Let  $m_A$  and  $m_B$  be bbas respectively held by agents  $A$  and  $B$  over the same frame  $\Theta$ . Let  $X = \{x_i; i = 1, \dots, k\}$  represent the powerset of  $\Theta$  according to Eq.(3). Let  $m_{A \diamond B}$  be the bba such that:

Case I: For  $m_A(\Theta) \neq 0 \vee m_B(\Theta) \neq 0$ :

$$\begin{cases} m_{A \diamond B}(x_i) = \frac{m_A(x_i)m_B(\Theta) + m_B(x_i)m_A(\Theta)}{m_A(\Theta) + m_B(\Theta)} \\ m_{A \diamond B}(\Theta) = \frac{2m_A(\Theta)m_B(\Theta)}{m_A(\Theta) + m_B(\Theta)} \end{cases}$$

Case II: For  $m_A(\Theta) = 0 \wedge m_B(\Theta) = 0$ :

$$\begin{cases} m_{A \diamond B}(x_i) = \gamma^A m_A(x_i) + \gamma^B m_B(x_i) \\ m_{A \diamond B}(\Theta) = 0 \end{cases} \quad \text{where} \quad \begin{cases} \gamma^A = \lim_{\substack{m_A(\Theta) \rightarrow 0 \\ m_B(\Theta) \rightarrow 0}} \frac{m_B(\Theta)}{m_A(\Theta) + m_B(\Theta)} \\ \gamma^B = \lim_{\substack{m_A(\Theta) \rightarrow 0 \\ m_B(\Theta) \rightarrow 0}} \frac{m_A(\Theta)}{m_A(\Theta) + m_B(\Theta)} \end{cases} \quad (20)$$

Then  $m_{A \diamond B}$  is called the averaged bba of  $m_A$  and  $m_B$ , representing the averaging fusion of the bbas of  $A$  and  $B$ . By using the symbol ' $\oplus$ ' to designate this belief operator, we define  $m_{A \diamond B} \equiv m_A \oplus m_B$ .

It can be verified that the averaging fusion rule is commutative, and idempotent, but not associative. The non-associativity means that averaging fusion of three argument bbas will produce different results depending on which two bbas are fused first. The cumulative rule represents a generalisation of the consensus rule for dependent opinions [12].

In Case II,  $\gamma^A$  and  $\gamma^B$  are relative weights satisfying  $\gamma^A + \gamma^B = 1$ . The default values are  $\gamma^A = \gamma^B = 0.5$ .

The proof below provides details about how the expression for the averaging rule is derived.

**Proof 2** Let  $m_A$  and  $m_B$  be Dirichlet bbas. The mapping from Dirichlet bbas to Dirichlet distributions is done according to the right hand sides of Eq.(11) and Eq.(12), expressed as:

$$\begin{aligned} m_A &\longmapsto f(\vec{p} \mid \vec{r}_A, \vec{a}) \\ m_B &\longmapsto f(\vec{p} \mid \vec{r}_B, \vec{a}) \end{aligned} \quad (21)$$

The average of these Dirichlet distributions can now be obtained according to

Eq.(19), expressed as:

$$f(\vec{p} \mid \vec{r}_A, \vec{a}) \oplus f(\vec{p} \mid \vec{r}_B, \vec{a}) = f(\vec{p} \mid (\frac{\vec{r}_A + \vec{r}_B}{2}), \vec{a}) \quad (22)$$

Finally, the result of Eq.(22) is mapped back to a Dirichlet bba again using Eq.(11). This can be written as:

$$f(\vec{p} \mid (\frac{\vec{r}_A + \vec{r}_B}{2}), \vec{a}) \mapsto m_{A \circ B} \quad (23)$$

By inserting the full expressions for the parameters in Eqs.(21), (22) and (23), the expressions of Eqs.(20) and (20) in Theorem 2 emerge.

## 7 Examples

In this section, we will illustrate by examples the results of applying the cumulative and averaging rules, as well as Dempster's rule of fusing beliefs. Each example is chosen to illustrate that it is crucial to select the appropriate rule for modelling a specific situation.

To make the presentation self contained, we also include the definitions of Conjunctive Rule and of Dempster's Rule below:

**Definition 2 (The Conjunctive Rule)** .

$$[m_A \otimes m_B](x) = \sum_{y \cap z = x} m_A(y) m_B(z) \quad \forall x \subseteq X. \quad (24)$$

This rule is referred to as the conjunctive rule of combination, or the non-normalised Dempster's rule. If necessary, the normality assumption  $m(\emptyset) = 0$  can be recovered by dividing each mass by a normalisation coefficient. The resulting operator known as Dempster's rule is defined as:

**Definition 3 (Dempster's Rule)** .

$$[m_A \odot m_B](x) = \frac{[m_A \otimes m_B](x)}{1 - [m_A \otimes m_B](\emptyset)} \quad \forall x \subseteq X, x \neq \emptyset \quad (25)$$

The use of Dempster's rule is possible only if  $m_A$  and  $m_B$  are not totally conflicting, i.e., if there exist two focal elements  $y$  and  $z$  of  $m_A$  and  $m_B$  satisfying  $y \cap z \neq \emptyset$ .



## 7.1 Zadeh's Example

This well known example was put forward by Zadeh [19] to show that Dempster's rule can produce counter-intuitive results when applied to particular situations.

The averaging rule produces results well in line with intuition when applied to this example, as will be shown below.

Suppose that we have a murder case with three suspects; Peter, Paul and Mary, and two witnesses  $W_A$  and  $W_B$  who give highly conflicting testimonies. The example assumes that the most reasonable conclusion about the likely murderer can be obtained by fusing the beliefs expressed by the two witnesses. Table 1 gives the witnesses' belief masses in the case of Zadeh's example and the resulting belief masses after applying Dempster's rule. The abbreviations "CR", "AR" and "DR" stand for Cumulative Rule, Averaging Rule, and Dempster's Rule respectively.

	$W_A$	$W_B$	CR	AR	DR
$m(\text{Peter}) =$	0.99	0.00	0.495	0.495	0.00
$m(\text{Paul}) =$	0.01	0.01	0.010	0.010	1.00
$m(\text{Mary}) =$	0.00	0.99	0.495	0.495	0.00
$m(\Theta) =$	0.00	0.00	0.000	0.000	0.00

Table 1  
Zadeh's example (averaging situation)

In case of Bayesian bbas such as in Zadeh's example, the cumulative rule and the averaging rules are equivalent, and represent the weighted average of probabilities.

The question now arises whether Zadeh's example represents a cumulative, an averaging or a conjunctive situation. Said differently, in case the testimonies can be considered as statistical evidence which is accumulated during the trial, then the cumulative rule should be applied. In case the testimonies should be weighted against each other to produce a balanced opinion, then the averaging rule should be applied. In case the testimonies can be considered as two different logical statements that are to be conjunctively combined as a function of their respective belief values, then Dempster's rule should be applied. In our view, an adequate model for the court case in Zadeh's example should reflect how a judge or a jury would weigh the testimonies against each other, which implies that the averaging rule is the best alternative.

## 7.2 Zadeh's Example Modified

Fusion of highly conflicting beliefs is problematic when applying Dempster's rule as in the original Zadeh's example. Many authors [20] explain this by saying that problem to certain degree can be remedied by discounting the testimonies before being fused, in order to reduce their degree of conflict. This approach is illustrated by the *Modified Zadeh's example* [21] below.

By introducing a small amount of uncertainty in the witnesses testimonies (see Table 2), the cumulative and averaging rules produce different but still very similar results. Dempster's Rule now produces almost the same results as the cumulative and averaging rules, but which are very different from those it produced in the original Zadeh's Example.

	$W_A$	$W_B$	CR	AR	DR
$m(\text{Peter}) =$	0.98	0.00	0.4925	0.4900	0.4900
$m(\text{Paul}) =$	0.01	0.01	0.0100	0.0100	0.0150
$m(\text{Mary}) =$	0.00	0.98	0.4925	0.4900	0.4900
$m(\Theta) =$	0.01	0.01	0.0050	0.0100	0.0050

Table 2

Modified Zadeh's example (averaging situation)

The introduction of a small amount of ignorance in the bbas was sufficient to make Dempster's rule produce intuitive results. Although it can be argued that witness testimonies should always be considered with a degree of uncertainty, it is problematic that Dempster's rule breaks down when the argument bbas are certain as in the case of Zadeh's original example. The discontinuity of the results demonstrates that Dempster's rule is inappropriate in this situation. The correct model for both the original and the modified Zadeh's example is the averaging rule because it does not produce any discontinuity between the two situations, and because testimonies of witnesses having observed the same murder should be considered as dependent statistical evidence.

## 7.3 Fusing Independent Sensor Evidence

This example will illustrate the fusion of cumulative evidence.

Assume that a GE (Genetical Engineering) process can produce fish eggs that are either Male (M) or Female (F), and that two sensors observe whether Male and Female eggs are produced. For the purpose of independence it is here assumed that the sensors observe the processes at different time periods, which means that none of the eggs observed by the first sensor are also observed by the second sensor. The

frame has two elements:  $\Theta = \{M, F\}$  and so does the corresponding powerset  $X$  which can be expressed as  $X = 2^\Theta \setminus \Theta = \{M, F\}$ .

Assume that the observations by the two sensors produce equal and non-conflicting beliefs as given in the Table 3 below over the two time periods:

	$S_A$	$S_B$	CR	AR	DR
$m(M) =$	0.99	0.99	0.994975	0.99	0.9999
$m(F) =$	0.00	0.00	0.00	0.00	0.00
$m(\Theta) =$	0.01	0.01	0.005025	0.01	0.0001

Table 3

Fusion of independent beliefs from two sensors (cumulative situation)

Applying the cumulative rule and Dempster's rule to these beliefs results in a reduction in uncertainty, and a convergence towards the largest belief mass of the sensor outputs. Applying the averaging rule preserves the values of the input observation bbas, because both are equal.

Intuitively, fusion should reduce the ignorance because more evidence is taken into account in this situation. The averaging rule can therefore be dismissed because it does not reduce the ignorance.

In case of the cumulative rule, the input beliefs are equivalent to each sensor having observed 198 Male eggs ( $r(M) = 198$ ), and no Female eggs ( $r(F) = 0$ ), with the uncertainty computed as  $m_{A \circ B}(\Theta) = 2/(r(M) + r(F) + 2) = 0.01$ . The output beliefs is equivalent to the observation of  $2r(M) = 396$  Male eggs, with the uncertainty computed as  $m(\Theta) = 2/(396 + 2) = 0.005025$ . The output uncertainty of the cumulative rule is thus halved, which is what one would expect in the case of cumulative fusion.

In case of Dempster's rule, the uncertainty is reduced to the product of the input uncertainties computed as:

$$[m_A \odot m_B](\Theta) = m_A(\Theta) m_B(\Theta) = 0.0001 \quad (26)$$

i.e. by a factor of  $1/100$  which represents a very fast convergence. In fact, when considering that the amount of evidence has only been doubled, a reduction in uncertainty by a factor of  $1/100$  is too fast when considering this as a cumulative situation. When fusing two equal sensor outputs, one should expect the uncertainty to be reduced by  $1/2$ , because the double amount of observations have been made.

The difference in convergence i.e. the rate of uncertainty reduction, is noteworthy, and clearly illustrates that Dempster's rule in fact is not applicable to this situation, even in the case of non-conflicting beliefs. This is because the example describes a cumulative situation, and that it would be meaningless to model it with a conjunctive fusion rule.

## 7.4 Conjunctive Fusion of Beliefs

In this example we consider the case of a loaded dice, where  $\Theta = \{1, 2, 3, 4, 5, 6\}$  is the set of possible outcomes. An informant has special knowledge about the loaded dice, and an observer is trying to predict the outcome of throwing the dice based on hints from the informant.

First the informant provides hint  $A$  which says that the dice will always produce an even number. The observer translates this into the belief mass  $m_A(\{2, 4, 6\}) = 1$ . Then the informant provides hint  $B$  which says that the dice will always produce a prime number. The observer translates this into the belief mass  $m_B(\{2, 3, 5\}) = 1$ .

Table 4 shows the results of applying the cumulative, averaging and Dempster's rule to these bbas.

	$m_A$	$m_B$	CR	AR	DR
$m(\{2, 4, 6\}) =$	1.00	0.00	0.50	0.50	0.00
$m(\{2, 3, 5\}) =$	0.00	1.00	0.50	0.50	0.00
$m(\{2\}) =$	0.00	0.00	0.00	0.00	1.00
$m(\Theta) =$	0.00	0.00	0.00	0.00	0.00

Table 4

Fusion of hints about loaded dice (conjunctive situation)

The two hints are sufficient to determine that  $\{2\}$  is the only possible outcome. This result is obtained by conjunctive combination of the input evidence, as dictated by Dempster's rule. The conjunctive approach of DR is appropriate in this situation because the input evidence applies to different and orthogonal focal sets, which would make it problematic to apply either the cumulative or averaging rules.

The correct answer could be obtained by applying a normalised version of the cumulative rule [22], but this approach should be considered as ad-hoc. The situation is clearly conjunctive in nature, which means that Dempster's rule or the conjunctive rule is appropriate.

## 8 Discussion

The cumulative and averaging rules of belief fusion make it possible to use the theory of belief functions for modelling situations where evidence is combined in a cumulative or averaging fashion. Such situations could previously not be correctly modelled within the framework of belief theory. It is worth noticing that the cumulative, averaging rules and Dempster's rule apply to different types of belief fusion,

and that, strictly speaking, is meaningless to compare their performance in the same examples. The notion of cumulative and averaging belief fusion as opposed to conjunctive belief fusion has therefore been introduced in order to make this distinction explicit.

There is however considerable confusion regarding the applicability of Dempster's rule, which e.g. is illustrated by applying Dempster's rule to the court case situation in Zadeh's example. Often the problem is to identify which model best fits a particular situation. The court case of Zadeh's example intuitively requires an averaging approach, the fusion of independent evidence from sensors that measure the same phenomenon intuitively requires a cumulative approach, and logical conjunction of evidence intuitively requires a conjunctive approach as through Dempster's rule.

To be more specific about the applicability of Dempster's rule, the two bbas to be fused are required to be orthogonal. Being orthogonal means that the bbas are obtained by deliberately considering different subsets of the frame. This is for example the case in the application of Dempster's rule in the framework of Kohlas' theory of hints [23]. In the example of the loaded dice from Sec.7.4, the informant deliberately provides only a part of the truth when giving the hints, and each hint focuses on a specific subset.

The following scenario will illustrate using the cumulative and the averaging fusion rules, as well as Dempster's rule. Assume again that GE process can produce Male (M) or Female (F) eggs, and that in addition, each egg can have genetical mutation 1 or 2 independently of its gender. This constitutes the quaternary frame  $\Theta = \{M1, M2, F1, F2\}$ . Sensors IA and IB simultaneously observe whether each egg is M or F, and Sensor II observes whether the egg has mutation 1 or 2.

Assume that Sensors IA and IB have derived two separate bbas regarding the gender of a specific egg, and that Sensor II has produced a bba regarding its mutation. Because Sensors IA and IB have observed the same aspect simultaneously, the bbas should be fused with the averaging rule. Sensor II has observed a different and orthogonal aspect, so the output of the AR fusion and the bba of Sensor II should be combined with Dempster's Rule. This is illustrated in Fig.5.

This result from fusing the two orthogonal beliefs with Dempster's Rule can now be considered as a single observation. By combining beliefs from multiple observations it is possible to express the most likely status of future eggs as a predictive belief. We are now dealing with two different situations which must be considered separately. The first situation relates to the state of a given egg that the sensors have already observed. The second situation relates to the possible state of eggs that will be produced in the future. A bba in the first situation is based on the sensors as illustrated inside Observation 1 in Fig.5. The second situation relates to combining multiple observations, as illustrated by fusing the beliefs from Observation 1 and Observation 2 in Fig.5.

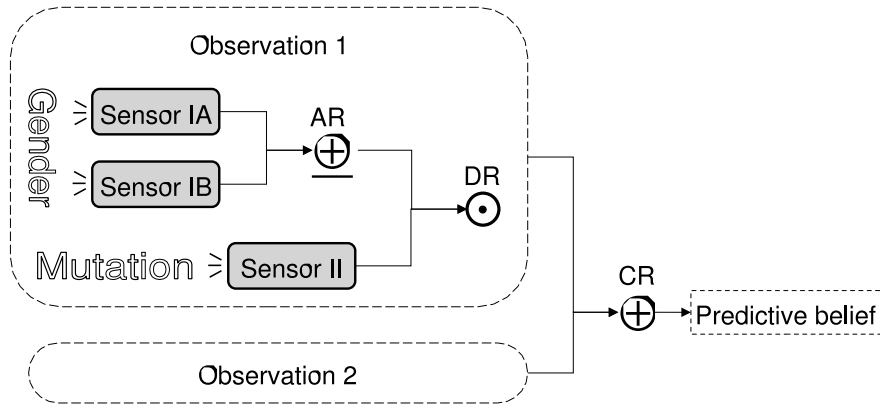


Figure 5. Applying different types of belief fusion according to the situation

In order to fuse observations, the bba for each observation must be normalised to carry the weight corresponding to a single Bayesian observation. This is done by multiplying all belief masses on proper subsets of  $\Theta$  with the factor  $\lambda$  expressed as:

$$\lambda = \frac{m(\Theta) + C(1 - m(\Theta))}{2C} \quad (27)$$

The belief mass on  $\Theta$  is then increased to compensate for the decreased belief masses on the proper subsets of  $\Theta$ . This produces the normalised bba:

$$m' : \begin{cases} m(x'_i) = \lambda m(x_i) \\ m'(\Theta) = m(\Theta) + (1 - \lambda)(1 - m(\Theta)) \end{cases} \quad (28)$$

Table 5 provides a numerical example that relates directly to the situations of Fig.5. Table entries with zero value are omitted. For simplicity it is assumed that Observation 2 produces the same beliefs as observation 1.

From Table 5 it can be seen that most likely egg in Observation 1 is Male with mutation 1, which is supported with belief mass  $b(M1) = 0.7$ .

To proceed from the results of DR to the values of Obs.1, the computation of the normalisation factor  $\lambda$  is required. By using Eq.(27) and setting  $C = 2$  we get  $\lambda = 0.498125$ . By using Eq.(28) the values in the column of Obs.1 can be computed, which represents the predictive bba resulting from one observation. For simplicity the bba of Obs.2 is set identical to that of Obs.1. It is then possible to fuse the two observations with the cumulative rule, as expressed by the rightmost column, to get a more accurate predictive bba of future observations.

It can be observed that the application of Dempster's rule in the examples above did not require any normalisation. In that case Dempster's rule is equivalent to the conjunctive rule of Def.2 could have been used.

	$S_{IA}$	$S_{IB}$	AR	$S_{II}$	DR	Obs.1	Obs.2	CR
$m(\{M1,M2\})$	0.90	0.80	0.875		0.0875	0.0436	id.	0.0583
$m(\{F1,F2\})$	0.05	0.05	0.050		0.0050	0.0025	id.	0.0033
$m(\{M1,F1\})$				0.80	0.0600	0.0299	id.	0.0400
$m(\{M2,F2\})$				0.10	0.0075	0.0037	id.	0.0050
$m(\{M1\})$					0.7000	0.3487	id.	0.4667
$m(\{M2\})$					0.0875	0.0436	id.	0.0583
$m(\{F1\})$					0.0400	0.0199	id.	0.0267
$m(\{F2\})$					0.0050	0.0025	id.	0.0033
$m(\Theta)$	0.05	0.15	0.075	0.10	0.0075	0.5056	id.	0.3384

Table 5

Application of different types of belief fusion according to the situations in Fig.5

## 9 Conclusion

Different situations require different types of belief fusion. We have described the cumulative and the averaging belief fusion rules which can be used for belief fusion in situations where Dempster's rule is inadequate. The two new rules represent generalisations of the corresponding fusion operators for opinions used in subjective logic, and have been derived from classical Bayesian analysis through a bijective mapping between Dirichlet distributions and belief functions.

This simple mapping positions belief theory and statistical theory closely and firmly together. This is important in order to make belief theory more practical and easier to interpret, and to make belief theory more acceptable in the main stream statistics and probability communities.

## References

- [1] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [2] M. Daniel, "Associativity in combination of belief functions," in *Proceedings of 5th Workshop on Uncertainty Processing. - Praha, Edicni oddeleni VSE 2000*. Springer, 2000, pp. 41–54.
- [3] J. Dezert, "Foundations for a new theory of plausible and paradoxical reasoning," *Information and Security Journal*, 2002.
- [4] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, pp. 244–264, 1988.

- [5] A. Jøsang, “The Consensus Operator for Combining Beliefs,” *Artificial Intelligence Journal*, vol. 142, no. 1–2, pp. 157–170, October 2002.
- [6] E. Lefevre, O. Colot, and P. Vannoorenberghe, “Belief functions combination and conflict management,” *Information Fusion*, vol. 3, no. 2, pp. 149–162, June 2002.
- [7] C. K. Murphy, “Combining belief functions when evidence conflicts,” *Decision Support Systems*, vol. 29, pp. 1–9, 2000.
- [8] P. Smets, “The combination of evidence in the transferable belief model,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.
- [9] R. Yager, “On the Dempster-Shafer framework and new combination rules,” *Information Sciences*, vol. 41, pp. 93–137, 1987.
- [10] F. Smarandache, “An In-Depth Look at Information Fusion Rules & the Unification of Fusion Theories,” *Computing Research Repository (CoRR)*, *Cornell University arXiv*, vol. cs.OH/0410033, 2004.
- [11] A. Jøsang, S. Pope, J. Diaz, and B. Bouchon-Meunier, “Dempster’s Rule as Seen by Little Coloured Balls (submitted to),” *International Journal of Approximate Reasoning*, 2007, available from <http://www.fit.qut.edu.au/~josang>.
- [12] A. Jøsang, S. Pope, and S. Marsh, “Exploring Different Types of Trust Propagation,” in *Proceedings of the 4th International Conference on Trust Management (iTrust)*, Pisa, May 2006.
- [13] A. Jøsang, “Probabilistic Logic Under Uncertainty,” in *The Proceedings of Computing: The Australian Theory Symposium (CATS2007)*, *CRPIT Volume 65*, Ballarat, Australia, January 2007.
- [14] —, “Cumulative and Averaging Unfusion of Beliefs,” in *The Proceedings of the International Conference on Information Processing and Management of Uncertainty (IPMU2008)*, Malaga, June 2008.
- [15] A. Jøsang and Z. Elouedi, “Interpreting Belief Functions as Dirichlet Distributions,” in *The Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, Hammamet, Tunisia, November 2007.
- [16] A. Jøsang, “A Logic for Uncertain Probabilities,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, June 2001.
- [17] P. Walley, “Inferences from Multinomial Data: Learning about a Bag of Marbles,” *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 3–57, 1996.
- [18] A. Gelman *et al.*, *Bayesian Data Analysis, 2nd ed.* Florida, USA: Chapman and Hall/CRC, 2004.
- [19] L. Zadeh, “Review of Shafer’s A Mathematical Theory of Evidence,” *AI Magazine*, vol. 5, pp. 81–83, 1984.



- [20] R. Haenni, “Shedding New Light on Zadeh’s Criticism of Dempster’s Rule of Combination,” in *Proceedings of Information Fusion*, Philadelphia, July 2005.
- [21] A. Jøsang, M. Daniel, and P. Vannoorenberghe, “Strategies for Combining Conflicting Dogmatic Beliefs,” in *Proceedings of the 6th International Conference on Information Fusion*, X. Wang, Ed., 2003.
- [22] A. Jøsang and S. Pope, “Normalising the Consensus Operator for Belief Fusion,” in *Proceedings of the International Conference on Information Processing and Management of Uncertainty (IPMU2006)*, Paris, July 2006.
- [23] J. Kohlas and P. Monney, *A Mathematical Theory of Hints. An Approach to Dempster-Shafer Theory of Evidence*, ser. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1995, vol. 425.