Interpreting Belief Functions as Dirichlet Distributions

Audun Jøsang[1] and Zied Elouedi[2]

[1] Information Security Institute
Queensland University of Technology
Brisbane, Australia.
a.josang@qut.edu.au
[2] Tunisian Institute of Management
Tunisia
zied.elouedi@gmx.fr

Abstract. Traditional Dempster Shafer belief theory does not provide a simple method for judging the effect of statistical and probabilistic data on belief functions and vice versa. This puts belief theory in isolation from probability theory and hinders fertile cross-disciplinary developments, both from a theoretic and an application point of view. It can be shown that a bijective mapping exists between Dirichlet distributions and Dempster-Shafer belief functions, and the purpose of this paper is to describe this correspondence. This has three main advantages; belief based reasoning can be applied to statistical data, statistical and probabilistic analysis can be applied to belief functions, and it provides a basis for interpreting and visualizing beliefs for the purpose of enhancing human cognition and the usability of belief based reasoning systems.

1 Introduction

Belief theory has its origin in a model for upper and lower probabilities proposed by Dempster in [1]. Shafer later proposed a model for expressing beliefs described in his book [8]. The main idea behind belief theory is to abandon the additivity principle of probability theory, i.e. that the sum of probabilities on all pairwise exclusive possibilities must add up to one. Instead, belief theory gives observers the ability to assign so-called belief mass to any subset of the frame of discernment, i.e. non-exclusive subsets as well as the whole frame itself. The advantage of this approach is that ignorance, i.e. the lack of information, can be explicitly expressed by assigning belief mass to subsets of the frame, or to the whole frame.

While classic Dempster-Shafer belief theory represents a powerful framework for representing partial ignorance in reasoning, there is no simple connection to probability theory and statistics allowing statistical data to be interpreted as belief functions and vice versa. A more direct relationship between belief theory and probability calculus would make it possible to compute and compare expected consequences and utilities[3] of the various courses of action.

In this paper we show how belief functions can be directly interpreted as Dirichlet probability density functions and vice versa. Models for representing uncertainty using the Dirichlet Distribution have been presented in the literature, with for example

[3] see e.g. M.R.B. Clarke's response to Philippe Smets' Chapter on belief functions [9].

mappings to upper and lower probabilities which in some sense can represent belief functions, but none of the previously described models provide a direct mapping to belief functions on the form of bbas. This is precisely the contribution if this paper. Our method for mapping belief functions to statistical data is simple and direct. This makes it possible to perform belief based reasoning directly on statistical data, and statistical reasoning directly on belief functions. This also provides a basis for various visualizations of belief functions that facilitate human understanding of beliefs and improve the usability of belief based reasoning systems.

The remainder of this paper is organized as follows: Sec.2 gives an overview of previous approaches. Sec.3 gives an overview of the belief function theory necessary for this presentation. Then, Sec.4 presents the Dirichlet multinomial model. In Sec.5, the mapping between Dirichlet distribution and belief distribution functions is detailed, and Sec.6 describes some applications of the mapping.

2 Previous Approaches

The Imprecise Dirichlet Model (IDM) for multinomial data is described by Walley [11] as a method for determining upper and lower probabilities. The model is based on varying the base rate over all possible outcomes. The probability expectation value of an outcome resulting from assigning the total base rate (i.e. equal to one) to that outcome produces the upper probability, and the probability expectation value of an outcome resulting from assigning a zero base rate to that outcome produces the lower probability. The upper and lower probabilities are 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, as will be shown in Sec.4.

Utkin (2005) [10] defines a method for deriving beliefs and plausibilities based on the IDM, where the lower probability is interpreted as the belief and the upper probability is interpreted as the plausibility. This method can produce unreasonable results in practical applications, and Utkin provides extensions to the Imprecise Dirichlet Model to overcome some of these problems. In our view the belif and plausibility functions can not be based on the base rate uncertainty of the Dirichlet distributions. The base rates are determined by the structure of the state space when it is known, and must be estimated on a subjective basis when not known [7]. In belief theory, the state space structure is used when e.g. computing the pignistic probability expectations, but it is independent of the bba.

An indirect quantitative method for determining belief functions from statistical data is described in Shafer's book [8] (p.237). Shafer's method requires the specification of an auxiliary frame of discernment of possible probability values for the elements of the primary frame of discernment. The method can then be used to determine belief functions on the auxiliary frame of discernment based on statistical data. The awkwardness of this method makes it difficult to use in practical applications.

3 Belief Theory

In this section several concepts of the Dempster-Shafer theory of evidence [8] are recalled in order to introduce notations used throughout the article. Let $\Theta = \{\theta_i;\ i = 1, \cdots k\}$ denote a finite set of exhaustive and exclusive possible values for a state variable of interest. The frame of discernment 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[4]), denoted by m, is defined as a belief distribution function from the power set 2^{Θ} to $[0, 1]$ satisfying:

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

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

A bba m can be equivalently represented by a non additive measure: a belief function Bel: $2^{\Theta} \to [0, 1]$, defined as

$$\mathrm{Bel}(x) \triangleq \sum_{\emptyset \neq y \subseteq x} m(y) \qquad \forall\, x \subseteq \Theta\ . \tag{2}$$

The quantity $\mathrm{Bel}(x)$ can be interpreted as a measure of one's total belief committed to the hypothesis that x is true. Note that functions m and Bel are in one-to-one correspondence [8] 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 Θ. This bba expresses the total ignorance. A *Bayesian* bba is when all the focal elements are singletons, i.e. one-element subsets of Θ. If all the focal elements are nestable (i.e. linearly ordered by inclusion) then we speak about *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.

We will use X to denote the power set of Θ, defined by $X = 2^{\Theta} \backslash \Theta$, which can also be expressed as $X = \{x_i; x_i \subset \Theta\}$. Thus all proper subsets of Θ are elements of X. By considering X as a state space in itself, a general bba on Θ becomes a particular bba on X called a *Dirichlet bba*. We define $m(X) = m(\Theta)$. A belief mass on a proper subsets of Θ then becomes a belief mass on an element of X. Dirichlet bba's on X are characterised by having mutually disjoint focal elements, except the whole state space X itself. This is defined as follows.

Definition 1 (Dirichlet bba). *Let X be a state space. 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.*

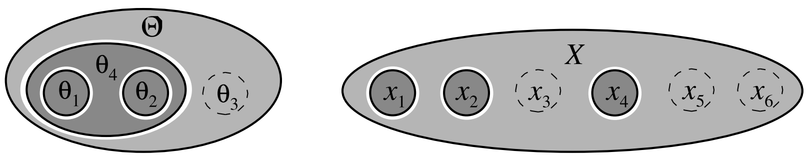
Fig.1.b below illustrates a Dirichlet bba on X, where the shaded circles around singletons and the shaded ellipse around X represent belief masses on those subsets. The focal elements in this example are X, x_1, x_2 and x_4.

[4] Called *basic probability assignment* in [8], and *Belief Mass Assignment* (BMA) in [3, 4].

The number of elements in X is $|X| = 2^{|\Theta|} - 2$ when excluding $\emptyset$. For example, Fig.1.b illustrates X as having cardinality 6, meaning that it is the power set of a ternary frame of discernment. The subsets of Θ 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 Θ and elements of X.

$$\begin{array}{lll} x_1 = \theta_1 & x_4 = \theta_1 \cup \theta_2 & X = \Theta \\ x_2 = \theta_2 & x_5 = \theta_1 \cup \theta_3 & \\ x_3 = \theta_3 & x_6 = \theta_2 \cup \theta_3 & \end{array} \tag{3}$$

Under this correspondence between the belief masses on X and on Θ, the focal elements of Θ are θ_1, θ_2, θ_4 and Θ as shown in Fig.1.a



(a) Focal elements of Θ (b) Focal elements of X

Fig. 1. Correspondence between belief masses on state space Θ and its power set X

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 bba's of this type are equivalent to Dirichlet probability density functions under a specific mapping. A bijective mapping between Dirichlet bba's and Dirichlet probability density functions is defined in [5], and is also described in Section 5 below.

4 The Dirichlet Multinomial Model

The cumulative rule of combination, to be described in detail in the following sections, is firmly rooted in the classical Bayesian inference theory, and is equivalent to the combination of multinomial observations. For self-containment, we briefly outline the Dirichlet multinomial model below, and refer to [2] for more details.

4.1 The Dirichlet Distribution

We are interested in knowing the probability distribution over the disjoint elements of a state space. In case of a binary state space, it is determined by the Beta distribution. In the general multinomial 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 simple additivity requirement

$$\sum_{i=1}^{k} p(x_i) = 1 \ . \tag{4}$$

Because of this additivity requirement, the Dirichlet distribution has only $k-1$ degrees of freedom. This means that knowing $k-1$ probability variables and their density uniquely determines the last probability variable and its density.

The Dirichlet distribution captures a sequence of observations of the k possible outcomes with k positive real parameters $\alpha(x_i),\ i = 1 \dots k$, 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 $\vec{\alpha} = \{\alpha_i \mid 1 \leq i \leq k\}$ to denote the k-component random evidence variable $[\alpha(x_i)]_{i=1}^{k}$.

The $\vec{\alpha}$ vector represents the *a priori* as well as the observation evidence. The weight of the *a priori* evidence can be expressed as a constant C, and this weight is distributed over all the possible outcomes as a function of the base rate.

The elements in a state space of cardinality k can have a base rate different from the default value $a = 1/k$. It is thereby possible to define a base rate as a vector $\vec{a}$ with arbitrary distribution over the k mutually disjoint elements x_i with $i = 1 \dots k$, as long as the simple additivity requirement expressed as $\sum_{i=1}^{k} a(x_i) = 1$ is satisfied. The total evidence $\alpha(x_i)$ for each element x_i can then be expressed as:

$$\alpha(x_i) = r(x_i) + C\,a(x_i)\ , \quad \text{where the constant } C \text{ is the } \textit{a priori} \text{ weight.} \tag{5}$$

The selection of the *a priori* weight C will be discussed below. The Dirichlet distribution over a set of k possible states x_i can thus be represented as a function of the base rate vector $\vec{a}$, the *a priori* weight C and the observation evidence $\vec{r}$.

Definition 2. Dirichlet Distribution

Let Θ be a state space consisting of k mutually disjoint elements. Let $\vec{r}$ represent the evidence vector over the elements of Θ and let $\vec{a}$ represent the base rate vector over the same elements. Then the multinomial Dirichlet density function over Θ can be expressed as:

$$f(\vec{p} \mid \vec{r}, \vec{a}) = \frac{\mathit{\Gamma}\left(\sum\limits_{i=1}^{k}(r(x_i) + Ca(x_i))\right)}{\prod\limits_{i=1}^{k}\mathit{\Gamma}\left(r(x_i) + Ca(x_i)\right)} \prod_{i=1}^{k} p(x_i)^{(r(x_i)+Ca(x_i)-1)} \tag{6}$$

where $r(x_1), \dots r(x_k) \geq 0, \quad a(x_1), \dots a(x_k) \in [0,1], \quad \sum_{i=1}^{k} a(x_i) = 1, \quad .$

The notation of Eq.(6) is useful, because it allows the determination of the probability distribution over state spaces where each element can have an arbitrary base rate.

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

$$\mathrm{E}(p(x_i) \mid \vec{r}, \vec{a}) = \frac{r(x_i) + Ca(x_i)}{C + \sum_{i=1}^{k} r(x_i)} \;. \tag{7}$$

It is normally required that the *a priori* distribution in case of a binary state space $X = \{x, \overline{x}\}$ is uniform, which means that $\alpha(x) = \alpha(\overline{x}) = 1$. Because of the additivity of $\vec{a}$, then necessarily the *a priori* weight $C = 2$. Should one assume an *a priori* uniform distribution over state spaces other than binary, the constant, and also the common value would be different. The *a priori* weight C will always be equal to the cardinality of the state space over which a uniform distribution is assumed.

Selecting $C > 2$ will result in new observations having relatively less influence over the Dirichlet distribution. This could be meaningful e.g. as a representation of specific *a priori* information provided by a domain expert. It can be noted that it would be unnatural to require a uniform distribution over arbitrary large state spaces because it would make the sensitivity to new evidence arbitrarily small.

For example, requiring a uniform *a priori* distribution over a state space of cardinality 100, would force the constant C to be $C = 100$. 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 counterintuitive. In contrast, when a uniform distribution is assumed in the binary case, and the same evidence is analysed, the derived probability expectation of the event of interest would be close to 1, as intuition would dictate. A good discussion of the choice of C can be found in [11].

It is here timely to revisit the Imprecise Dirichlet Model (IDM) described by Walley [11]. According to this model, the upper and lower probability values for an outcome x_i are defined as:

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

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

It can easily be shown that these values can not be literally interpreted as upper and lower bounds for for the relative frequency. 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. In other words, if $\underline{P}(\text{black}) > p(\text{black})$ then $\underline{P}(\text{black})$ can impossibly be the lower bound. This case shows that the upper and lower probabilities defined by the IDM should be interpreted as an expectation value range, because that would make it consistent with the fact that actual relative frequencies can be outside the range.

4.2 Visualizing Dirichlet Distributions

Visualising Dirichlet distributions is challenging because it is a density function over $k-1$ dimensions, where k is the state space cardinality. For this reason, Dirichlet distributions over ternary state spaces are the largest that can be practically visualised.

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

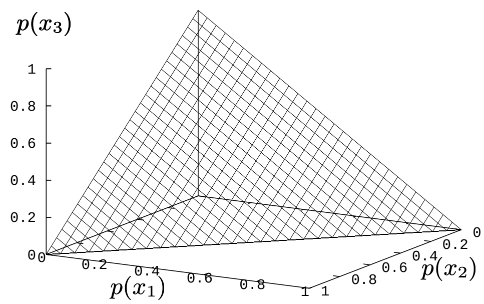


Fig. 2. Triangular plane

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.

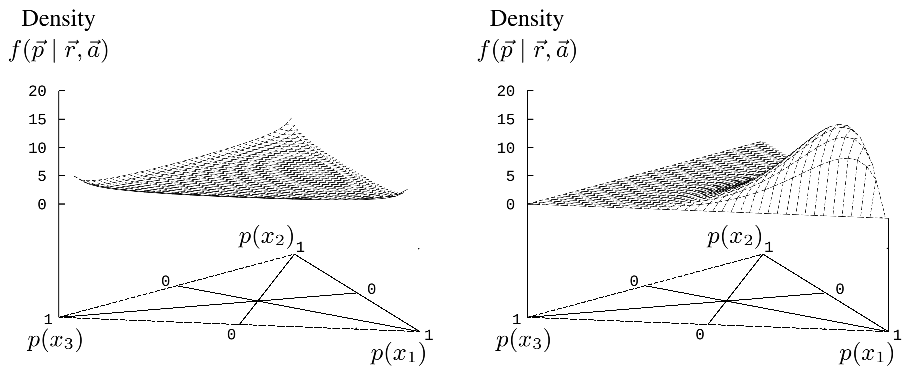
Let us consider the example of an urn containing balls of the three different types: the balls can be marked with x_1, x_2 or x_3 (i.e. $k=3$). Let us first assume that no other information than the cardinality is available, meaning that the default base rate is $a=1/3$, and that $r(x_1)=r(x_2)=r(x_3)=0$. Then Eq.(7) dictates that the expected *a priori* probability of picking a ball of any specific colour is the default base rate probability, which is $\frac{1}{3}$. The *a priori* Dirichlet density function is illustrated on the left side of Fig.3.

Let us now assume that an observer has picked (with return) 6 balls of type x_1, 1 ball of type x_2 and 1 ball of type x_3, i.e. $r(x_1)=6,\ r(x_2)=1,\ r(x_3)=1$, then the *a posteriori* expected probability of picking a ball of type x_1 can be computed as $\mathrm{E}(p(x_1))=\frac{2}{3}$. The *a posteriori* Dirichlet density function is illustrated on the right side in Fig.3.

4.3 Coarsening Example: From Ternary to Binary

We reuse the example of Sec.4.2 with the urn containing red, black and yellow balls, but this time we create a binary partition of $x_1=\{\text{red}\}$ and $x_2=\{\text{black, yellow}\}$. The base rate of picking a red ball is set to the relative atomicity of red balls, expressed as $a(x_1)=\frac{1}{3}$.

Let us again assume that an observer has picked (with return) 6 red balls, and 2 "black or yellow" balls, i.e. $r(x_1)=6,\ r(x_2)=2$.

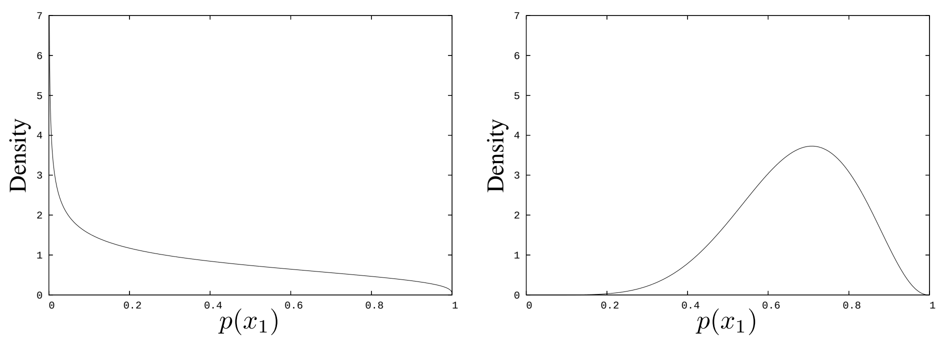


(a) *A priori* Dirichlet distribution

(b) Dirichlet distrib. after 6 red (x_1), 1 black (x_2) and 1 yellow (x_3) balls

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

Since the state space 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.4.



(a) *A priori* Beta distribution, $a(x_1) = \frac{1}{3}$

(b) After 6 red and 2 yellow or black balls

Fig. 4. Visualising prior and posterior Beta distributions

The *a posteriori* expected probability of picking a red ball can be computed with Eq.(7) as $\mathrm{E}(p(x_1)) = \frac{2}{3}$, which is the same as before the coarsening, as described in Sec.4.2. This shows that the coarsening does not influence the probability expectation value of specific events.

5 Mapping Between Dirichlet Distribution and Belief Distribution Functions

In this section we will define a bijective mapping between Dirichlet probability distributions described in Sec.4, and Dirichlet bba's described in Sec.3.

Let $X = \{x_i;\ i = 1, \cdots k\}$ be a state space where each singleton represents a possible outcome of a state variable. It is assumed that X is the power set of a frame of discernment Θ. Let m be a general bba on Θ and therefore a Dirichlet bba on X, and let $f(\vec{p} \mid \vec{r}, \vec{a})$ be a Dirichlet probability distribution function over X.

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

$$\wp(x_i) = \mathrm{E}(p(x_i) \mid \vec{r}, \vec{a}) \tag{10}$$

$$\Updownarrow$$

$$m(x_i) + a(x_i)m(X) = \frac{r(x_i) + a(x_i)C}{C + \sum_{t=1}^{k} r(x_t)} \tag{11}$$

We also require that $m(x_i)$ be an increasing function of $r(x_i)$, and that $m(X)$ 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 the belief mass on that outcome. Furthermore, the less evidence available in general, the more vacuous the bba (i.e. the greater $m(X)$). These intuitive requirements together with Eq.(11) imply the bijective mapping defined by Eq.(12).

For $\quad m(X) \neq 0$:

$$\begin{cases} m(x_i) = \frac{r(x_i)}{C + \sum_{t=1}^{k} r(x_t)} \\ \\ m(X) = \frac{C}{C + \sum_{t=1}^{k} r(x_t)} \end{cases} \quad \Leftrightarrow \quad \begin{cases} r(x_i) = \frac{C m(x_i)}{m(X)} \\ \\ 1 = m(X) + \sum_{i=1}^{k} m(x_i) \end{cases} \tag{12}$$

Next, we consider the case of zero uncertainty. In case $m(X) \longrightarrow 0$, then necessarily $\sum_{i=1}^{k} m(x_i) \longrightarrow 1$, and $\sum_{i=1}^{k} r(x_i) \longrightarrow \infty$, 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.(13).

For $m(X) = 0$:

$$\begin{cases} m(x_i) = \eta(x_i) \\ \\ m(X) = 0 \end{cases} \Leftrightarrow \begin{cases} r(x_i) = \eta(x_i) \sum\limits_{t=1}^{k} r(x_t) = \eta(x_i)\infty \\ \\ 1 = \sum\limits_{t=1}^{k} m(x_t) \end{cases} \tag{13}$$

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 $j = 1 \ldots l$, then all the evidence parameters are all equally infinite.

6 Applications of the bba-Dirichlet Correspondence

Having established the mapping between Dirichlet distributions and belief mass distributions in the form of bbas, it is possible to investigate how tools from traditional probability theory can be used in belief theory and vice versa.

Bayesian updating is for example performed by simple addition of observation variables. Let $\vec{r}_A$ and $\vec{r}_B$ be two sets of observations of the same set of outcomes. The Dirichlet distribution of the combined observation is obtained by simple vector addition of $\vec{r}_A$ and $\vec{r}_B$. Mapping this vector sum to the bba space results in an operator called the cumulative fusion rule for belief [6] which represents a generalization of the consensus operator used in subjective logic [4].

Similarly, the average of statistical observations can be computed by taking the average of two sets of observations represented as vectors. Mapping the average vector to the bba space results in an operator called the averaging fusion rule for beliefs [6].

Any operator from belief theory can be translated and be applied to Dirichlet distributions, such as e.g. Dempster's orthogonal rule. Interestingly this rule becomes very different from traditional Bayesian updating. For a binary state space $X = \{x, \overline{x}\}$, Dempster's orthogonal rule can be expressed as

$$m_A \odot m_B : \begin{cases} m(x) = \frac{m_A(x)m_B(x)+m_A(x)m_B(X)+m_A(X)m_B(x)}{1-m_A(x)m_B(\overline{x})-m_A(\overline{x})m_B(x)} \\ \\ m(\overline{x}) = \frac{m_A(\overline{x})m_B(\overline{x})+m_A(\overline{x})m_B(X)+m_A(X)m_B(\overline{x})}{1-m_A(x)m_B(\overline{x})-m_A(\overline{x})m_B(x)} \\ \\ m(X) = \frac{m_A(X)m_B(X)}{1-m_A(x)m_B(\overline{x})-m_A(\overline{x})m_B(x)} \end{cases} \tag{14}$$

The Beta distribution represents the binary case of the Dirichlet distribution. Let r represent the observation parameter of x and let s represent the observation parameter of $\overline{x}$. Bayesian updating dictates that $(r_A, s_A) + (r_B, s_B) = (r_A + r_B,\ s_A + s_B)$ whereas Dempster's orthogonal rule is expressed as:

$$(r_A, s_A) \odot (r_B, s_B) : \begin{cases} r = \frac{r_A r_B + 2(r_A + r_B)}{(r+s+2)^2 - r_A s_B - s_A r_B} \\ \\ s = \frac{s_A s_B + 2(s_A + s_B)}{(r+s+2)^2 - r_A s_B - s_A r_B} \end{cases} \tag{15}$$

Combining statistical observation evidence according to the binomial Dempster's orthogonal rule of Eq.(15) is certainly new in the field of Bayesian probability theory. Generalisation to a multinomial expression is straightforward.

Belief functions on binary state spaces can be expressed as opinions in subjective logic, and visualised with the opinion triangle of subjective logic. Opinions correspond to Beta distributions which are also convenient to visualise. A simple online demonstration shows the correspondence between opinions and Beta density functions.This is shown in Fig.5, which is a screen capture of an online demonstration[5].

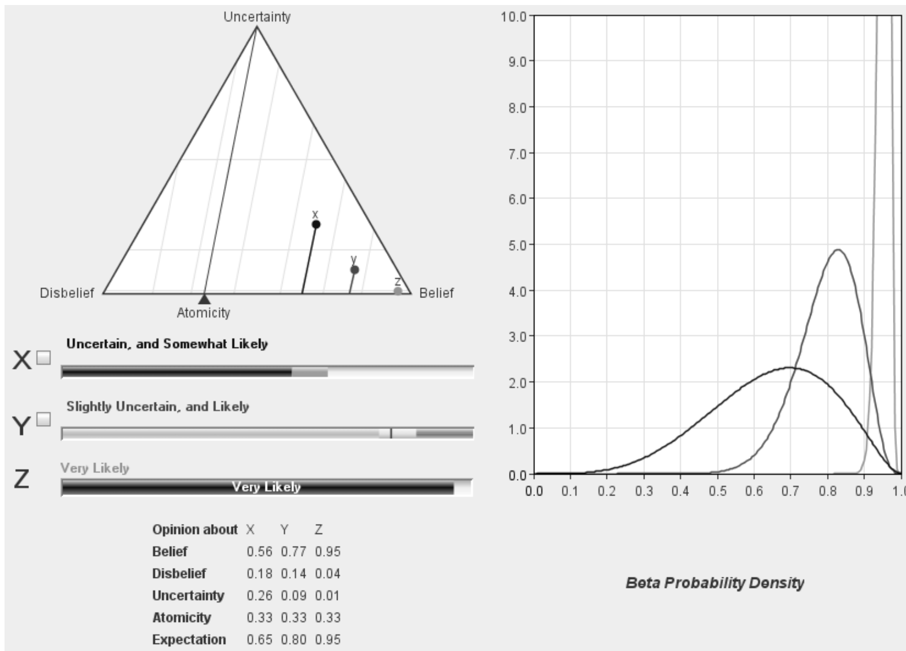


Fig. 5. Example visualisations of binomial opinions

The example visualises opinions about three different statements x, y and z. Each belief is visualised in different ways, i.e. in the form of 1) points in an opinion triangle, 2) beta density functions, 3) coloured/shaded bars, and 4) fuzzy verbal categories.

The interpretation of the opinion triangle and the beta PDF need no further explanation, as they have been described in the previous sections. Suffice to mention that the leftmost PDF refers to x, the middle PDF refers to y and the rightmost PDF refers to z.

The horizontal shaded bars are actually coloured in the online demonstration, which makes them easier to interpret. The first horizontal bar, representing the belief in x, consists of a dark shaded area representing b_x, and a light shaded area representing

[5] http://www.fit.qut.edu.au/ josang/sl/demo/BV.html

$a_x u_x$ (i.e. the amount of uncertainty that contributes to $\mathrm{E}(x)$, so that the total length of the dark and light shaded areas together represent $\mathrm{E}(x)$.

The second horizontal bar, representing the belief in y, consists of a green (leftmost) area representing b_y, an amber (middle) area representing u_y, and a red (rightmost) area representing d_y, as well as a black vertical line within the amber area indicating $\mathrm{E}(y)$. The second horizontal bar thus uses the traffic light metaphor, where green indicates "go", red indicates "stop" and amber indicates "caution".

The third horizontal bar, representing the belief in z, simply has a single dark shaded area representing $\mathrm{E}(z)$.

7 Conclusion

The mapping between beliefs and probability distribution functions puts belief theory and statistical theory 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. A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, 38:325–339,, 1967.
2. A. Gelman et al. *Bayesian Data Analysis, 2nd ed.* Chapman and Hall/CRC, Florida, USA, 2004.
3. A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.
4. A. Jøsang. The Consensus Operator for Combining Beliefs. *Artificial Intelligence Journal*, 142(1–2):157–170, October 2002.
5. A. Jøsang. Probabilistic Logic Under Uncertainty. In *In the Proceedings of Computing: The Australian Theory Symposium (CATS2007), CRPIT Volume 65*, Ballarat, Australia, January 2007.
6. A. Jøsang and S. Marsh. The Cumulative Rule of Belief Fusion (to appear). In *In the Proceedings of the International Conference on Information Fusion*, Quebec, Canada. July 2007.
7. Simon Pope and Audun Jøsang. Analsysis of Competing Hypotheses using Subjective Logic. In *Proceedings of the 10th International Command and Control Research and Technology Symposium (ICCRTS)*. United States Department of Defense Command and Control Research Program (DoDCCRP), 2005.
8. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
9. Ph. Smets. Belief Functions. In Ph. Smets et al., editors, *Non-Standard Logics for Automated Reasoning*, pages 253–286. Academic Press, 1988.
10. L.V. Utkin. Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model. *Fuzzy Sets and Systems*, 154:413–431, 2005.
11. P. Walley. Inferences from Multinomial Data: Learning about a Bag of Marbles. *Journal of the Royal Statistical Society*, 58(1):3–57, 1996.