Josang, Audun (2007) Probabilistic Logic under Uncertainty. In Gudmundsson, J. and Jay, B., Eds. *Proceedings Thirteenth Computing : The Australasian Theory Symposium (CATS2007)* CRPIT, 65, pages pp. 101-110, Ballarat, Australia.

Probabilistic Logic under Uncertainty

Audun Jøsang

School of Software Engineering and Data Communications
QUT, Brisbane, Australia
Email: a.josang@qut.edu.au

Abstract

Probabilistic logic combines the capability of binary logic to express the structure of argument models with the capacity of probabilities to express degrees of truth of those arguments. The limitation of traditional probabilistic logic is that it is unable to express uncertainty about the probability values themselves. This paper provides a brief overview subjective logic which is a probabilistic logic that explicitly takes uncertainty about probability values into account. More specifically, we describe equivalent representations of uncertain probabilities, and their interpretations. Subjective logic is directly compatible with binary logic, probability calculus and classical probabilistic logic. The advantage of using subjective logic is that real world situations can be more realistically modelled, and that conclusions more correctly reflect the ignorance and uncertainties about the input arguments.

Keywords: Subjective logic, probabilistic logic, uncertainty, belief theory, opinion, Dirichlet, Beta,

1 Introduction

In standard logic, propositions are considered to be either true or false. However, a fundamental aspect of the human condition is that nobody can ever determine with absolute certainty whether a proposition about the world is true or false. In addition, whenever the truth of a proposition is assessed, it is always done by an individual, and it can never be considered to represent a general and objective belief. This indicates that important aspects are missing in the way standard logic captures our perception of reality, and that it is more designed for an idealised world than for the subjective world in which we are all living.

Probabilistic logic was first defined by Nilsson [12] with the aim of combining the capability of deductive logic to exploit the structure and relationship of arguments and events, with the capacity of probability theory to express degrees of truth about those arguments and events. This results in more realistic models of real world situations than is possible with binary logic.

The additivity principle of classical probability requires that the probability of mutually disjoint elements in a state space add up to 1. This requirement makes it impossible to express ignorance about the likelihoods of possible states or outcomes. If somebody wants to express ignorance as *"I don't know"* this would be impossible with a simple scalar probability value. A probability 0.5 would for example mean that the event will take place 50% of the time, which in fact is quite informative, and very different from ignorance. Alternatively, a uniform probability density function over all possible states would more closely express the situation of ignorance about the outcome of an event, and it will be shown below that this can be interpreted as equivalent to the type of belief functions which we will study here.

Belief theory represents an extension of classical probability by allowing explicit expression of ignorance. Belief theory has its origin in a model for upper and lower probabilities proposed by Dempster in 1960. Shafer later proposed a model for expressing beliefs [15]. 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 disjoint states must add up to one. Instead belief theory gives observers the ability to assign so-called belief mass to any subset of the state space, i.e. to non-exclusive possibilities including the whole state space itself. The main advantage of this approach is that ignorance, i.e. the lack of information, can be explicitly expressed e.g. by assigning belief mass to the whole state space.

Uncertainty about probability values can be interpreted as ignorance, or second order uncertainty about the first order probabilities. In this paper, the term "uncertainty" will be used in the sense of *"uncertainty about the probability values"*. A probabilistic logic based on belief theory therefore represents a generalisation of traditional probabilistic logic.

Classical belief representation is quite general, and allows complex belief structures to be expressed on arbitrary large state spaces. Shafer's book [15] describes many aspects of belief theory, but the two main elements are 1) a flexible way of expressing beliefs, and 2) a method for combining beliefs, commonly known as Dempster's Rule. We will not be concerned with Dempster's rule here.

In order to have a simpler representation of beliefs, which also can be mapped to probability density functions, special types of belief functions called *"opinions"* will be used. We will show that the definition of logic operators on opinions is very simple, thereby resulting in a rich set of operators that defines subjective logic. Through the equivalence between opinions and probability density functions, subjective logic also provides a calculus for probability density functions.

In previous presentations of subjective logic, operators are defined with binomial opinions as input and with a binomial opinion as output. A binomial opinion is a belief function defined over a binary state space. In this paper we describe the general principles of subjective logic, and we show how operators can be generalised to take multinomial opinions as input and thereby produce a multinomial opinion as output. We also describe three different but equivalent representations of opinions, and discuss their interpretation. This allows uncertain probabilities to be seen from different angles, and allows an analyst to define models according to the formalisms that they are most familiar with, and that most naturally represents a specific real world situation. Subjective logic contains the same set of basic operators known from binary logic and classical probability calculus, but also contains some non-traditional operators which are specific to subjective logic.

The advantage of subjective logic over traditional probabilistic logic is that real world situations can be modeled and analysed more realistically. The analyst's partial ignorance and lack of information can be taken into account during the analysis, and explicitly expressed in the conclusion. Applications can for example be decision making, where the decision makers will be better informed about uncertainties underlying a given model.

2 Representing Beliefs

2.1 Classical Belief Representation

Belief representation in classic belief theory[15] is based on an exhaustive set of mutually exclusive atomic states which is called the *frame of discernment* denoted by Θ. In our presentation below we will use the term *state space* in the sense of frame of discernment. The power set 2^{Θ} is the set of all sub-sets of Θ. A bba (basic belief assignment[1]) is a belief mass distribution function m_Θ mapping 2^{Θ} to $[0,1]$ such that

$$\sum_{x \subseteq \Theta} m_\Theta(x) = 1\ , \quad \text{where} \quad m_\Theta(\emptyset) = 0\ . \qquad (1)$$

The bba distributes a total belief mass of 1 amongst the subsets of Θ such that the belief mass for each subset is positive or zero. Each subset $x \subseteq \Theta$ such that $m_\Theta(x) > 0$ is called a focal element of m_Θ. In the case of total ignorance, $m_\Theta(\Theta) = 1$, and m_Θ is called a *vacuous* bba. In case all focal elements are atoms (i.e. one-element subsets of Θ) then we speak about *Bayesian* bba. A *dogmatic* bba is when $m_\Theta(\Theta) = 0$ [16]. Let us note that, trivially, every Bayesian bba is dogmatic.

The Dempster-Shafer theory [15] defines a belief function $b(x)$. The probability transformation [1][2] projects a bba onto a probability expectation value denoted by $p(x)$. These functions are defined as:

$$b(x) = \sum_{\emptyset \neq y \subseteq x} m_\Theta(y) \qquad \forall\, x \subseteq \Theta\ , \qquad (2)$$

$$p(x) = \sum_{y \subseteq \Theta} m_\Theta(y) \frac{|x \cap y|}{|y|} \qquad \forall\, x \subseteq \Theta\ . \qquad (3)$$

In subjective logic, opinions express specific types of beliefs, and represent the input and output parameters of the subjective logic operators. Opinions expressed over binary state spaces are called binomial. Opinions defined over state spaces larger than binary are called multinomial. Multinomial opinions will be defined first.

2.2 Belief Notation of Opinions

As a specialisation of the general bba, we define the Dirichlet bba, and its cluster variant, as follows.

Definition 1 (Dirichlet bba) *A bba where the possible focal elements are Θ and/or singletons of Θ, is called a Dirichlet belief mass distribution function.*

Definition 2 (Cluster Dirichlet bba) *A bba where the possible focal elements are Θ and/or mutually disjoint subsets of Θ (singletons or clusters of singletons), is called a cluster Dirichlet belief mass distribution function.*

It can be noted that Bayesian bbas are a special case of Dirichlet bbas.

The name "Dirichlet" bba is used because bbas of this type can be interpreted as equivalent to Dirichlet probability density functions under a specific mapping described below. The same mapping in the case of binary state spaces is described in [3].

The probability transformation of Eq.(3) assumes that each element in the state space gets an equal share of belief masses that are assigned to (partly) overlapping elements. In case of Dirichlet bbas, the belief mass on the whole state space is the only belief mass to be distributed in this way. Let $a(x)$ represent the relative share that each element x receives. The function a will be called the *base rate function*, as defined below.

Definition 3 (Base Rate Function) *Let $\Theta = \{x_i | i = 1, \ldots k\}$ be a state space and let a be a function from Θ to $[0,1]$ representing* a priori *probability expectation before any evidence has been received, satisfying:*

$$a(\emptyset) = 0 \quad \textit{and} \quad \sum_{x \in \Theta} a(x) = 1\ . \qquad (4)$$

Then a is called a base rate function.

The combination of a Dirichlet bba and a base rate function can be contained in a composite function called an *opinion*. In order to have a simple and intuitive notation, the Dirichlet bba is split into a belief vector $\vec{b}$ and an uncertainty parameter u. This is defined as follows.

Definition 4 (Belief Vector and Uncertainty Parameter) *Let m_Θ be a Dirichlet bba. The belief vector $\vec{b}_\Theta$ and the uncertainty parameter u_Θ are defined as follows:*

$$b_\Theta(x_i) = m_\Theta(x_i)\ , \quad \textit{where} \quad x_i \neq \Theta \qquad (5)$$

$$u_\Theta = m_\Theta(\Theta) \qquad (6)$$

It can be noted that $u_\Theta + \sum_{x=1}^{k} b_\Theta(x_i) = 1$ because of Eq.(1). The belief vector $\vec{b}_\theta$ and the uncertainty parameter u_θ are used in the definition of multinomial opinions below.

Definition 5 (Belief Notation of Opinions) *Let $\Theta = \{x_i | i = 1 \ldots k\}$ be a frame of discernment. Assume m_Θ to be a Dirichlet bba on Θ with belief vector $\vec{b}_\theta$ and uncertainty parameter u_Θ, and assume $\vec{a}_\Theta$ to be a base rate vector on Θ. The composite function $\omega_\Theta = (\vec{b}_\Theta, u_\Theta, \vec{a}_\Theta)$ is then an opinion on Θ represented in belief notation.*

[1] Called *basic probability assignment* in [15].

[2] Also known as the pignistic transformation [17, 18]

We use the convention that the subscript on the multinomial opinion symbol indicates the state space on which the opinion applies, and that a superscript indicates the subject owner of the opinion. Subscripts can be omitted when it is clear to which state space an opinion applies, and superscripts can be omitted when it is irrelevant who the owner is.

Assuming that the state space Θ has cardinality k, the belief vector $\vec{b}_\Theta$ and the base rate vector $\vec{a}_\Theta$ will have k parameters each. The uncertainty parameter u_Θ is a simple scalar. A multinomial opinion in belief notation over a state space of cardinality k will thus contain $2k+1$ parameters. However, given the constraints of Eq.(1) and Eq.(4), the multinomial opinion will only have $2k-1$ degrees of freedom. A binomial opinion will for example be 3-dimensional.

The introduction of the base rate function allows the probabilistic transformation to be independent from the internal structure of the state space. The probability transformation of multinomial opinions can be expressed as a function of the belief and the base rate vectors.

Definition 6 (Probability Expectation Function) *Let $\Theta = \{x_i | i = 1, \ldots k\}$ be a state space and let ω_Θ be an opinion on Θ with belief vector $\vec{b}$ and uncertainty u. Let $\vec{a}$ be a base rate vector on Θ. The function p_Θ from Θ to $[0,1]$ representing the* a posteriori *probability expectation expressed as:*

$$p_\Theta(x_i) = b_\Theta(x_i) + a_\Theta(x_i)u_\Theta \ . \tag{7}$$

is then called the probability expectation function.

It can be shown that p satisfies the additivity principle:

$$p_\Theta(\emptyset) = 0 \quad \text{and} \quad \sum_{x\in\Theta} p_\Theta(x) = 1 \ . \tag{8}$$

It is interesting to notice that the base rate function of Def.3 expresses *a priori* probability, where as the probability expectation function of Eq.(7) expresses *a posteriori* probability.

Given a state space of cardinality k, the default base rate function for each element in the state space is $1/k$, but it is possible to define arbitrary base rates for all mutually exclusive elements of the state space, as long as the additivity constraint is satisfied.

Two different multinomial opinions on the same state space will normally share the same base rate functions. However, it is obvious that two different observers can assign different base rate functions to the same state space, and this could naturally reflect two different analyses of the same situation by two different persons.

2.3 Binomial Opinions

A special notation will be used to denote a binomial opinion which consists of an ordered tuple containing the three specific belief masses *belief, disbelief, uncertainty* as well as the base rate of x_i.

Definition 7 (Binomial Opinion) *Let $\Theta = \{x_i | i = 1 \ldots k\}$ be a state space. Assume m_Θ to be a Dirichlet bba on Θ, and a_Θ to be a base rate function on Θ. The ordered quadruple ω_{x_i} defined as:*

$$\omega_{x_i} = (b_{x_i},\ d_{x_i},\ u_{x_i},\ a_{x_i}) \ ,$$

$$\textit{where} \begin{cases} \textit{Belief:} & b_{x_i} = m_\Theta(x_i) \\ \textit{Disbelief:} & d_{x_i} = m_\Theta(\overline{x}_i) \\ \textit{Uncertainty:} & u_{x_i} = m_\Theta(\Theta) \\ \textit{Base rate:} & a_{x_i} = a_\Theta(x_i)) \end{cases} \tag{9}$$

is then called a binomial opinion on the binary state space $X_i = \{x_i, \overline{x_i}\}$.

Binomial opinions represent a special case of multinomial opinions. Binomial opinions can be defined for binary state spaces or for state spaces larger that binary when coarsened to binary.

Binomial opinions are used in traditional subjective logic operators defined in [3, 4, 8, 9, 6, 13].

The probability expectation value of a binomial opinion can be derived from Eq.(7), is:

$$p(\omega_{x_i}) = b_{x_i} + a_{x_i}u_{x_i} \tag{10}$$

Binomial opinions can be mapped to a point in an equal-sided triangle. The relative distances from the left side edge to the point represent belief, from the right side edge to the point represent disbelief, and from the base line to the point represents uncertainty. For an arbitrary opinion $\omega_x = (b_x, d_x, u_x, a_x)$, the three parameters b_x, d_x and u_x thus determine the position of the opinion point in the triangle. The base line is the *probability axis*, and the base rate value can be indicated as a point on the probability axis.

Fig.1 illustrates an example opinion about x with the value $\omega_x = (0.7,\ 0.1,\ 0.2,\ 0.5)$ indicated by a black dot in the triangle.

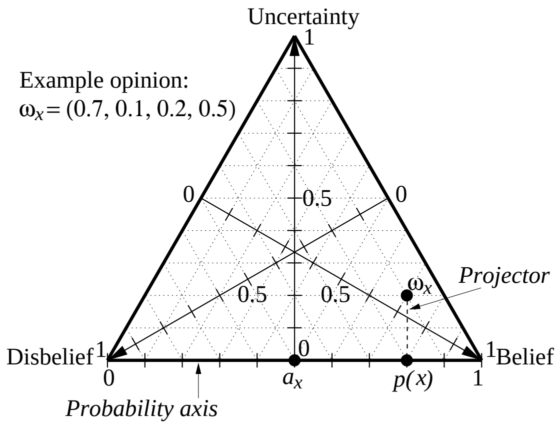

Figure 1: Opinion triangle with example opinion

The *projector* going through the opinion point, parallel to the line that joins the uncertainty corner and the base rate point, determines the probability expectation value $p(x) = b_x + a_x u_x$.

Although a binomial opinion has 4 parameters, it only has 3 degrees of freedom because the three components b_x, d_x and u_x are dependent through Eq.(1). As such they represent the traditional *Bel*(x) (Belief) and *Pl*(x) (Plausibility) pair of Shaferian belief theory through the correspondence *Bel*$(x) = b_x$ and *Pl*$(x) = b_x + u_x$.

The reason why a redundant parameter is kept in the opinion representation is that it allows for more compact expressions of opinion operators than otherwise would have been possible.

Various visualisations of opinions are possible to facilitate human interpretation. See for example http://sky.fit.qut.edu.au/∼josang/sl/demo/BV.html

General binomial opinions are equivalent to Beta distributions. Given that binomial opinions represent a special case of multinomial opinions, this fact is derived from the fact that Beta distributions represent a special case of Dirichlet distributions, and that multinomial opinions are equivalent to Dirichlet distributions.

3 Probabilistic Notation of Opinions

A disadvantage of the belief notation described in Sec.2.2 and Sec.2.3, is that it does not directly reflect the probability expectation values of the various elements in the state space. The classical probabilistic representation has the advantage that it is used in all areas of science and that people are familiar with it. The probability expectation can easily be derived with Eq.(7), but this still represents a mental barrier to a direct intuitive interpretation of opinions. An intuitive representation of multinomial opinions could therefore be to represent the probability expectation value directly, together with the degree of uncertainty and the base rate function. This will be called the *probabilistic notation* of opinions:

Definition 8 (Probabilistic Notation of Opinions) *Let Θ be a state space and let ω_Θ be an opinion on Θ in belief notation. Let $p(x)$ be a multinomial probability expectation function on Θ defined according to Def.6, let $a(x)$ be a multinomial base rate function on Θ defined according to Def.3, and let u be the uncertainty function on Θ as defined in Eq.(9). The probabilistic notation of opinions can then be expressed as the ordered tuple $\omega = (\vec{p}, u, \vec{a})$.*

In case $u = 0$, then $\vec{p}$ is a frequentist probability distribution. In case $u = 1$, then $\vec{p} = \vec{a}$, and no evidence has been received, so the posterior probability is equal to the prior probability.

The equivalence between the belief notation and the probabilistic notation of opinions is defined below.

Theorem 1 (Probabilistic Notation Equivalence) *Let $\omega_\Theta = (\vec{b}_\Theta, u_\Theta, \vec{a}_\Theta)$ be an opinion expressed in belief notation, and $\omega = (\vec{p}, u, \vec{a})$ be an opinion expressed in probabilistic notation, both over the same state space Θ. Then the following equivalence holds:*

$$\left\{ \begin{array}{ll} p(x_i) & = b_\Theta(x_i) + a_\Theta(x_i) u_\Theta \\ u & = u_\Theta \\ \vec{a} & = \vec{a}_\Theta \end{array} \right. \qquad (11)$$

$$\Updownarrow$$

$$\left\{ \begin{array}{ll} b_\Theta(x_i) & = p(x_i) - a_\Theta(x_i) u \\ u_\Theta & = u \\ \vec{a}_\Theta & = \vec{a} \end{array} \right. \qquad (12)$$

Binomial opinions in probabilistic notation will be written as $\omega = (p, u, a)$ where p represents the probability expectation value, u represents the uncertainty and a represents the base rate.

4 Beliefs from Observation of Evidence

The evidence notation of opinions is centered around the Dirichlet multinomial probability distribution. 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 $\sum_{i=1}^{k} p(x_i) = 1$.

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 observation variable $[\alpha(x_i)]_{i=1}^k$.

The Dirichlet probability density function is then given by Eq.(13) below.

$$f(\vec{p} \mid \vec{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^k \alpha(x_i)\right)}{\prod_{i=1}^k \Gamma(\alpha(x_i))} \prod_{i=1}^{k} p(x_i)^{\alpha(x_i)-1}, \qquad (13)$$

$$\text{where} \left\{ \begin{array}{l} p(x_1), \dots, p(x_k) \geq 0 \\ \sum_{i=1}^k p(x_i) = 1 \\ \alpha(x_1), \dots, \alpha(x_k) > 0. \end{array} \right.$$

The probability expectation value of any of the k random variables is defined as:

$$\mathrm{E}(p(x_i) \mid \vec{\alpha}) = \frac{\alpha(x_i)}{\sum_{i=1}^k \alpha(x_i)}. \qquad (14)$$

Because of the additivity requirement $\sum_{i=1}^k p(x_i) = 1$, 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 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 vector with arbitrary distribution over the k mutually disjoint elements x_j with $j = 1 \dots k$, as long as the simple additivity requirement is satisfied, expressed as:

$$\sum_{x_j \in \Theta} a(x_j) = 1. \qquad (15)$$

The total evidence $\alpha(x_j)$ for each element x_j can then be expressed as:

$$\alpha(x_j) = r(x_j) + a(x_j) \qquad (16)$$

In order to distinguish between the base rate, and the evidence, we introduce the augmented notation for Dirichlet distribution over a set of k singletons.

Definition 9 (Augmented Dirichlet Notation)
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 in augmented notation as:

$$f(\vec{p} \mid \vec{r}, \vec{a}) = \frac{\Gamma\left(\sum_{j=1}^{k} (r(x_j) + Ca(x_j))\right)}{\prod_{j=1}^{k} \Gamma\left(r(x_j) + Ca(x_j)\right)} \prod_{j=1}^{k} p(x_j)^{(r(x_j)+Ca(x_j)-1)} \tag{17}$$

$$\textit{where} \begin{cases} p(x_1), \ldots, p(x_k) \geq 0, \\ \sum_{j=1}^{k} p(x_j) = 1, \\ \alpha(x_1), \ldots, \alpha(x_k) > 0, \\ \sum_{j=1}^{k} a(x_j) = 1, \\ C = 2. \end{cases}$$

The augmented notation of Eq.17 is useful, because it allows the determination of the probability distribution over state spaces where each element can have an arbitrary base rate as long as the simple additivity principle is satisfied. Given the augmented Dirichlet distribution of Eq.(17), the probability expectation of any of the k random probability variables can now be written as:

$$\mathrm{E}(p(x_j) \mid \vec{r}, \vec{a}) = \frac{r(x_j) + Ca(x_j)}{C + \sum_{j=1}^{k} r(x_j)} \,. \tag{18}$$

4.2 Visualising 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.

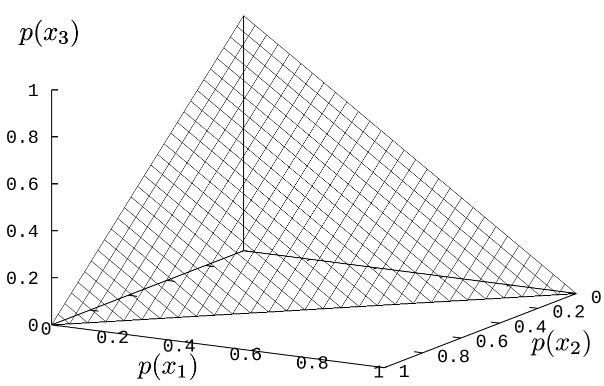

Figure 2: Triangular plane

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 of the three different colours: red, black and yellow (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 $r(\text{red}) = r(\text{black}) = r(\text{yellow}) = 0$. Then Eq.(18) 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 in Fig.3.

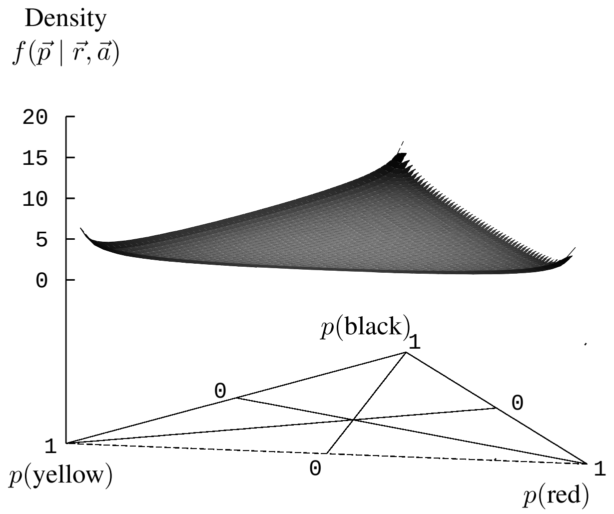

Figure 3: Prior Dirichlet distribution in case of urn with balls of 3 different colours

Let us now assume that an observer has picked (with return) 6 red, 1 black and 1 yellow ball, i.e. $r(\text{red}) = 6$, $r(\text{black}) = 1$, $r(\text{yellow}) = 1$, then the *a posteriori* expected probability of picking a red ball can be computed as $\mathrm{E}(p(\text{red})) = \frac{2}{3}$. The *a posteriori* Dirichlet density function is illustrated in Fig.4.

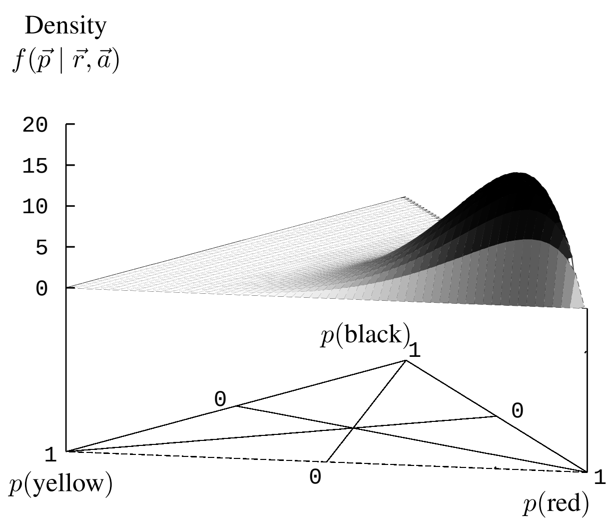

Figure 4: *A posteriori* Dirichlet distribution after picking 6 red, 1 black and 1 yellow ball

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. $r(x_1) = 6,\ r(x_2) = 2$.

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.4.3 and Fig.4.3 .

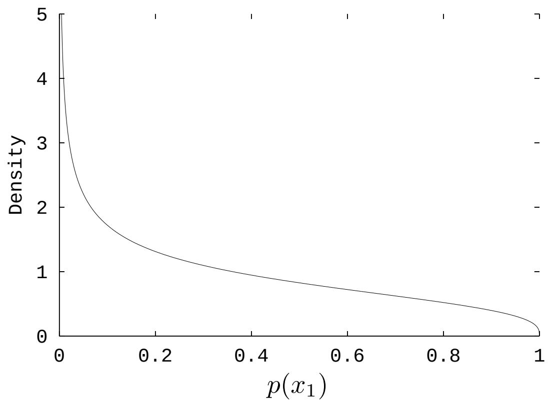

Figure 5: *A priori* Beta distribution

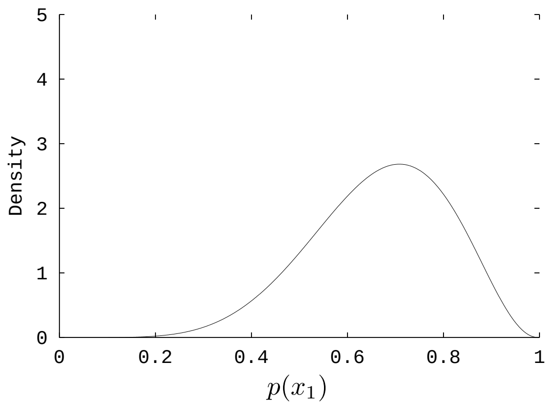

Figure 6: Updated Beta distribution after 6 red balls

The *a posteriori* expected probability of picking a red ball can be computed with Eq.(18) as $\mathrm{E}(p(x_1)) = \frac{2}{3}$, which is the same as before the coarsening, as illustrated in Sec.4.2. This shows that the coarsening does not influence the probability expectation value of specific events.

4.4 Evidence Notation of Opinions

Dirichlet distributions translate observation evidence directly into probability density functions. The representation of evidence, together with the base rate, can be used to denote opinions.

Definition 10 (Evidence Notation of Opinions) *Let Θ be a state space with a Dirichlet distribution $f(\vec{p} \mid \vec{r}, \vec{a})$. The evidence notation of opinions can then be expressed as the ordered tuple $\omega = (\vec{r}, \vec{a})$.*

Let $\omega = (\vec{r}, \vec{a})$ be an opinion in evidence notation over a state space with cardinality k. Then all the k parameters of $\vec{r}$ are independent, whereas only $k-1$ parameters of $\vec{a}$ are independent because of Eq.(4). As expected, the evidence notation therefore has $k-1$ dimensions, as do the belief and probabilistic notations.

It is possible to define a bijective mapping between opinions expressed in evidence notation based on the probability distributions described in Sec.4.1, and opinions expressed in belief notation as described in Sec.2.2.

Let $\Theta = \{x_i \mid i = 1, \cdots, k\}$ be a state space. Let $\omega_\Theta = (\vec{b}_\Theta, u_\Theta, \vec{a}_\Theta)$ be an opinion on Θ in belief notation, and let $\omega = (\vec{r}, \vec{a})$ be an opinion on Θ in evidence notation.

For the bijective mapping between ω_Θ and ω, we require equality between the pignistic probability values $p_\Theta(x_j)$ derived from ω_Θ, and the probability expectation values $\mathrm{E}(p(x_j))$ derived from $\omega = (\vec{r}, \vec{a})$. This constraint is expressed as:

For all $x_j \in \Theta_X$:

$$p_\Theta(x_j) = \mathrm{E}(\ p(x_j) \mid \vec{r}, \vec{a}) \qquad (19)$$

$$\Updownarrow$$

$$b_\Theta(x_j) + a_\Theta(x_j)u_\Theta = \frac{r(x_j)}{C + \sum_{j=1}^{k} r(x_j)} + \frac{Ca(x_j)}{C + \sum_{j=1}^{k} r(x_j)} \qquad (20)$$

We also require that $b_\Theta(x_j)$ be an increasing function of $r(x_j)$, and that u_Θ be a decreasing function of $\sum_{j=1}^{k} r(x_j)$. 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, the less certain the opinion (i.e. the greater u_Θ).

In case $u_\Theta \longrightarrow 0$, then $\sum_{j=1}^{k} b_\Theta(x_j) \longrightarrow 1$, and $\sum_{j=1}^{k} r(x_j) \longrightarrow \infty$, meaning that at least some, but not necessarily all, of the evidence parameters $r(x_j)$ are infinite. We define $\eta(x_j)$ as the the relative degree of infinity between the corresponding infinite evidence parameters $r(x_j)$ such that $\sum_{j=1}^{k} \eta(x_j) = 1$. When infinite evidence parameters exist, any finite evidence parameter $r(x_j)$ can be assumed to be zero in any practical situation because it will have $\eta(x_j) = 0$, i.e. it will carry zero weight relative to the infinite evidence parameters.

These intuitive requirements together with Eq.(21) imply the following bijective mapping:

Table 1: Example values with the three equivalent notations of binomial opinion, and their interpretations.

Belief notation $\omega = (b, d, u, a)$	Probabilistic notation $\omega = (p, u, a)$	Density notation $\omega = (r, s, a)$	Interpretation
$(1, 0, 0, a)$	$(1, 0, a)$	$(\infty, 0, a)$	Binary logic TRUE, and probability $p = 1$
$(0, 1, 0, a)$	$(0, 0, a)$	$(0, \infty, a)$	Binary logic FALSE, and probability $p = 0$
$(0, 0, 1, a)$	$(a, 1, a)$	$(0, 0, a)$	Vacuous opinion, Beta distribution with prior a
$(0, 0, 1, \frac{1}{2})$	$(\frac{1}{2}, 1, \frac{1}{2})$	$(0, 0, \frac{1}{2})$	Vacuous opinion, uniform Beta distribution over binary state space
$(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(1, 1, \frac{1}{2})$	Symmetric Beta distribution after 1 positive and 1 negative observation, binary state space

Theorem 2 (Evidence Notation Equivalence) *Let $\omega_\Theta = (\vec{b}_\Theta, u_\Theta, \vec{a}_\Theta)$ be an opinion expressed in belief notation, and $\omega = (\vec{r}, \vec{a})$ be an opinion expressed in evidence notation, both over the same state space Θ. Then the following equivalence holds:*

For $u_\Theta \neq 0$:

$$\begin{cases} b_\Theta(x_j) &= \frac{r(x_j)}{C + \sum_{i=1}^{k} r(x_i)} \\ u_\Theta &= \frac{C}{C + \sum_{i=1}^{k} r(x_i)} \end{cases} \quad (21)$$

$$\Updownarrow$$

$$\begin{cases} r(x_j) = \frac{C b_\Theta(x_j)}{u_\Theta} \\ 1 = u_\Theta + \sum_{i=1}^{k} b_\Theta(x_i) \end{cases} \quad (22)$$

For $u_\Theta = 0$:

$$\begin{cases} b_\Theta(x_j) &= \eta(x_j) \\ u_\Theta &= 0 \end{cases} \quad (23)$$

$$\Updownarrow$$

$$\begin{cases} r(x_j) &= \eta(x_j) \sum_{i=1}^{k} r(x_i) \\ &= \eta(x_j)\infty \\ 1 = \sum_{j=1}^{k} m(x_j) \end{cases} \quad (24)$$

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

Binomial opinions in evidence notation will be written as $\omega = (r, s, a)$, where r and s represent the amount of positive and negative evidence respectively, and a represents the base rate.

5 Basic Properties of Subjective Logic

Subjective logic consists of a set of operators where opinions are input and output parameters. Table 2 provides a brief overview of the main operators. Additional operators exist for modeling special situations such as fusion of evidence from multiple observers. Most of the operators correspond to well-known operators from binary logic and probability calculus, whereas others are specific to subjective logic. An online demonstration of subjective logic can be accessed at: http://sky.fit.qut.edu.au/∼josang/sl/.

Subjective logic is a generalisation of binary logic and probability calculus. This means that when a corresponding operator exists in binary logic, and the input parameters are equivalent to binary logic TRUE or FALSE, then the result opinion is equivalent to the result that the corresponding binary logic expression would have produced.

For example consider the case of binary logic AND, which corresponds to multiplication of opinions which is described in [8]. Assume the pair of binomial opinions in probabilistic notation

$$\begin{cases} \omega_x = (1, 0, a_x) \\ \omega_y = (0, 0, a_y) \end{cases} \text{ then } w_{x \wedge y} = (0, 0, a_x a_y) \quad (25)$$

which corresponds to TRUE $\wedge$ FALSE = FALSE.

Similarly, when a corresponding operator exists in probability calculus, then the probability expectation value of result opinion is equal to the result that the corresponding probability calculus expression would have produced with input parameters equal to the probability expectation values of the input opinions.

For example, assume the following pair of argument opinions in probabilistic notation:

$$\begin{cases} \omega_x = (p_x, 0, a_x) \\ \omega_y = (p_y, 0, a_y) \end{cases} \quad (26)$$
$$\text{then } \omega_{x \wedge y} = (p_x p_y, 0, a_x a_y)$$

which corresponds to $p(x \wedge y) = p(x)p(y)$.

Table 1 provides the equivalent interpretation in binary logic and probability for a small set of binomial opinions represented in belief notation, probabilistic notation and in evidence notation.

It can be seen that certain values correspond to binary logic and probability values, whereas other values correspond to probability density distributions. This richness

Table 2: Correspondence between probability, set and logic operators.

Opinion operator	Symbol	Set operator	Logic operator	Symbol	Notation
Addition[11]	$+$	Union	XOR [3]	$\cup$	$\omega_{x\cup y} = \omega_x + \omega_y$
Subtraction[11]	-	Difference	n.a.	$\setminus$	$\omega_{x\setminus y} = \omega_x - \omega_y$
Multiplication[8]	$\cdot$	Conjunction	AND	$\wedge$	$\omega_{x\wedge y} = \omega_x \cdot \omega_y$
Division[8]	/	Unconjunction	UN-AND	$\overline{\wedge}$	$\omega_{x\overline{\wedge} y} = \omega_x / \omega_y$
Comultiplication[8]	$\sqcup$	Disjunction	OR	$\vee$	$\omega_{x\vee y} = \omega_x \sqcup \omega_y$
Codivision[8]	$\overline{\sqcup}$	Undisjunction	UN-OR	$\overline{\vee}$	$\omega_{x\overline{\vee} y} = \omega_x \overline{\sqcup} \omega_y$
Complement[3]	$\neg$	Negation	NOT	$\overline{x}$	$\omega_{\overline{x}} = \neg\omega_x$
Deduction[9]	$\circledcirc$	Conditional Inference	MP	$\|$	$\omega_{Y\|X} = \omega_X \circledcirc \omega_{Y\|X}$
Abduction[13]	$\overline{\circledcirc}$	Reverse conditional Inference	MT	$\overline{\|}$	$\omega_{Y\overline{\|}X} = \omega_Y \overline{\circledcirc} \omega_{X\|Y}$
Discounting[10]	$\otimes$	Transitivity	n.a.	:	$\omega_x^{A:B} = \omega_B^A \otimes \omega_x^B$
Cumulation[5, 10]	$\oplus$	Cumulative Fusion	n.a.	$\diamond$	$\omega_X^{A\diamond B} = \omega_X^A \oplus \omega_X^B$
Average[10]	$\underline{\oplus}$	Averaging Fusion	n.a.	$\underline{\diamond}$	$\omega_x^{A\underline{\diamond} B} = \omega_x^A \underline{\oplus} \omega_x^B$

of expression represents the advantage of subjective logic over other probabilistic logic logic frameworks.

It is interesting to note that subjective logic represents a calculus for Dirichlet distributions because opinions are equivalent to Dirichlet distributions. Analytical manipulations of Dirichlet distributions is complex but can be done for simple operators, such as multiplication in which case it is called a joint distribution.

The multiplicative product of two opinions is not equal to the joint distribution of Dirichlet distributions in general, but it can be shown that the approximation is very good. This means that the simplicity of some subjective logic operators comes at the cost of reducing those operators to approximations of the analytically correct operators.

The analytical result of joint Dirichlet distributions will in general involve the Gauss hypergeometric function, see e.g. [14] for the details. However, this analytical method will quickly become unmanageable when applied to the more complex operators of Table 2 such as conditional deduction and abduction. Subjective logic therefore has the advantage of providing advanced operators for Dirichlet distributions for which no practical analytical solutions exist.

6 Fusion of Multinomial Opinions

In many situations there will be multiple sources of evidence, and fusion can be used to combine evidence from different sources.

In order to provide an interpretation of fusion in subjective logic it is useful to consider a process that is observed by two sensors. A distinction can be made between two cases.

1. The two sensors observe the process during disjoint time periods. In this case the observations are independent, and it is natural to simply add the observations from the two sensors, and the resulting fusion is called *cumulative fusion*.
2. The two sensors observe the process during the same time period. In this case the observations are dependent, and it is natural to take the average of the observations by the two sensors, and the resulting fusion is called *averaging fusion*.

Fusion of binomial opinions have been described in [3, 4]. The two types of fusion for multinomial opinions are described in the following subsections. When observations are partially dependent, a hybrid fusion operator can be defined [10].

6.1 Cumulative Fusion

The cumulative fusion rule is equivalent to *a posteriori* updating of Dirichlet distributions. Its derivation is based on the bijective mapping between the belief and evidence notations described in Sec.4.4.

Assume a state space Θ containing k elements. Assume two observers A and B who observe the outcomes of the process over two separate time periods.

Let the two observers' respective observations be expressed as $\vec{r}^A$ and $\vec{r}^B$. The evidence opinions resulting from these separate bodies of evidence can be expressed as $(\vec{r}^A, \vec{a})$ and $(\vec{r}^B, \vec{a})$

The cumulative fusion of these two bodies of evidence simply consists of vector addition of $\vec{r}^A$ and $\vec{r}^B$, expressed as:

$$(\vec{r}^A, \vec{a}) \oplus (\vec{r}^B, \vec{a}) = ((\vec{r}^A + \vec{r}^B), \vec{a}) \ . \tag{27}$$

The symbol "$\diamond$" denotes the 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 3 (Cumulative Fusion Rule)
Let ω^A and ω^B be opinions respectively held by agents A and B over the same state space $\Theta = \{x_j \mid j = 1, \cdots, l\}$. Let $\omega^{A\diamond B}$ be the opinion such that:

Case I: *For* $u^A \neq 0 \ \vee \ u^B \neq 0$:

$$\begin{cases} b^{A\diamond B}(x_j) & = \frac{b^A(x_j)u^B + b^B(x_j)u^A}{u^A + u^B - u^A u^B} \\ u^{A\diamond B} & = \frac{u^A u^B}{u^A + u^B - u^A u^B} \end{cases} \tag{28}$$

Case II: *For* $u^A = 0 \ \wedge \ u^B = 0$:

$$\begin{cases} b^{A\diamond B}(x_j) & = \gamma^A b^A(x_j) + \gamma^B b^B(x_j) \\ u^{A\diamond B} & = 0 \end{cases} \tag{29}$$

$$\text{where} \quad \gamma^A = \lim_{\substack{u^A \to 0 \\ u^B \to 0}} \frac{u^B}{u^A + u^B}$$

$$\text{and} \quad \gamma^B = \lim_{\substack{u^A \to 0 \\ u^B \to 0}} \frac{u^A}{u^A + u^B}$$

Then $\omega^{A\diamond B}$ is called the cumulatively fused bba of ω^A and ω^B, representing the combination of independent opinions of A and B. By using the symbol '$\oplus$' to designate this belief operator, we define $\omega^{A\diamond B} \equiv \omega^A \oplus \omega^B$.

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

Proof 1 *Let ω^A and ω^B be opinions held by observer A and B respectively. The mapping from opinions in belief notation to opinions in evidence notation is done according Eq.(2) and Eq.(24), expressed as:*

$$\begin{aligned} \omega^A_\Theta &\longmapsto (\vec{r}^A, \vec{a}) \\ \omega^B_\Theta &\longmapsto (\vec{r}^B, \vec{a}) \end{aligned} \tag{30}$$

These opinions in evidence notation can now be fused according to Eq.(27), expressed as:

$$(\vec{r}^A, \vec{a}) \oplus (\vec{r}^B, \vec{a}) = ((\vec{r}^A + \vec{r}^B), \vec{a}) \tag{31}$$

Finally, the result of Eq.(31) is mapped back to an opinion in belief notation again using Eq.(21) and Eq.(23). This can be written as:

$$((\vec{r}^A + \vec{r}^B), \vec{a}) \longmapsto \omega^{A\diamond B}_\Theta \tag{32}$$

By inserting the full expressions for the parameters in Eqs.(30), (31) and (32), the expressions of Eqs.(28) and (29) in Theorem 3 emerge.

□

It can be verified that the cumulative rule is commutative, associative and non-idempotent. In Case II of Theorem 3, the associativity depends on the preservation of relative weights of intermediate results, which requires the additional weight variable γ. 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* constant 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 and ad hoc choices.

The cumulative rule represents a generalisation of the consensus operator [4, 3] which emerges directly from Theorem 3 by assuming a binary state space.

6.2 The Average Rule of Belief Fusion

The average rule is equivalent to averaging the evidence of Dirichlet distributions. Its derivation is based on the bijective mapping between the belief and evidence notations described in Sec.4.4.

Assume a state space Θ containing k elements. Assume two observers A and B who observe the outcomes of the process over the same time periods.

Let the two observers' respective observations be expressed as $\vec{r}^A$ and $\vec{r}^B$. The evidence opinions resulting from these separate bodies of evidence can be expressed as $(\vec{r}^A, \vec{a})$ and $(\vec{r}^B, \vec{a})$

The averaging fusion of these two bodies of evidence simply consists of averaging $\vec{r}^A$ and $\vec{r}^B$. Expressed in terms of Dirichlet distributions, this can be expressed as:

$$(\vec{r}^A, \vec{a}) \underline{\oplus} (\vec{r}^B, \vec{a}) = ((\frac{\vec{r}^A + \vec{r}^B}{2}), \vec{a}) \ . \tag{33}$$

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

Theorem 4 (Averaging Fusion Rule)
Let ω^A and ω^B be opinions respectively held by agents A and B over the same state space $\Theta = \{x_j \mid j = 1, \cdots, l\}$. Let $\omega^{A\underline{\diamond}B}$ be the opinion such that:

Case I: *For* $u^A \neq 0 \ \vee \ u^B \neq 0$:

$$\begin{cases} b^{A\underline{\diamond}B}(x_j) & = \frac{b^A(x_j)u^B + b^B(x_j)u^A}{u^A + u^B} \\ u^{A\underline{\diamond}B} & = \frac{2u^A u^B}{u^A + u^B} \end{cases} \tag{34}$$

Case II: *For* $u^A = 0 \ \wedge \ u^B = 0$:

$$\begin{cases} b^{A\diamond B}(x_j) & = \gamma^A b^A(x_j) + \gamma^B b^B(x_j) \\ u^{A\diamond B} & = 0 \end{cases} \tag{35}$$

$$\text{where} \quad \gamma^A = \lim_{\substack{u^A \to 0 \\ u^B \to 0}} \frac{u^B}{u^A + u^B}$$

$$\text{and} \quad \gamma^B = \lim_{\substack{u^A \to 0 \\ u^B \to 0}} \frac{u^A}{u^A + u^B}$$

Then $\omega^{A\underline{\diamond}B}$ is called the averaged opinion of ω^A and ω^B, representing the combination of the dependent opinions of A and B. By using the symbol '$\underline{\oplus}$' to designate this belief operator, we define $\omega^{A\underline{\diamond}B} \equiv \omega^A \underline{\oplus} \omega^B$.

The proof of Theorem 4 is very similar to that of Theorem 3 and is omitted here.

It can be verified that the averaging fusion rule is commutative, associative and non-idempotent.

The cumulative rule represents a generalisation of the consensus rule for dependent opinions defined in [7].

7 Conclusion

Subjective logic represents a generalisation of probability calculus and logic under uncertainty. Subjective logic will always be equivalent to traditional probability calculus when applied to traditional probabilities, and will be equivalent to binary logic when applied to TRUE and FALSE statements.

While subjective logic has traditionally been applied to binary state spaces, we have shown that it can easily be extended and be applicable to state spaces larger than binary. The input and output parameters of subjective logic are beliefs in the form of opinions. We have described three different equivalent notations of opinions which provides rich interpretations of opinions. This also allows the analyst to choose the opinion representation that best suits a particular situation.

References

[1] D. Dubois and H. Prade. On Several Representations of an Uncertain Body of Evidence. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, 1982.

[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. The Cumulative Rule for Combining Beliefs. (Submitted to) *Information Fusion Journal*, 00(0):0–0, 2006. Available from: http://www.fit.qut.edu.au/∼josang/.

[6] A. Jøsang, R. Hayward, and S. Pope. Trust Network Analysis with Subjective Logic. In *Proceedings of the* 29^{th} *Australasian Computer Science Conference (ACSC2006), CRPIT Volume 48*, Hobart, Australia, January 2006.

[7] A. Jøsang and S.J. Knapskog. A Metric for Trusted Systems (full paper). In *Proceedings of the 21st National Information Systems Security Conference*. NSA, October 1998.

[8] A. Jøsang and D. McAnally. Multiplication and Comultiplication of Beliefs. *International Journal of Approximate Reasoning*, 38(1):19–51, 2004.

[9] A. Jøsang, S. Pope, and M. Daniel. Conditional deduction under uncertainty. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, 2005.

[10] 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.

[11] D. McAnally and A. Jøsang. Addition and Subtraction of Beliefs. In *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004)*, Perugia, July 2004.

[12] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28(1):71–87, 1986.

[13] 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.

[14] A.P. Prudnikov, Yu.A. Brychkov, and O.I. Marichev. *Integrals and Series,* (translated from Russian), volume 1–3. Gordon and Breach Science Publishers, Amsterdam, New York, 1986.

[15] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[16] Ph. Smets. Belief Functions. In Ph. Smets et al., editors, *Non-Standard Logics for Automated Reasoning*, pages 253–286. Academic Press, 1988.

[17] Ph. Smets. Decision Making in the TBM: the Necessity of the Pignistic Transformation. *Int. J. Approximate Reasoning*, 38:133–147, 2005.

[18] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.