Self Conditional Probabilities and Probabilistic Interpretations of Belief Functions.

Roger Cooke
Department of Mathematics
Delft University of Technology
Delft, The Netherlands
r.m.cooke@its.tudelft.nl

Philippe Smets*
IRIDIA
Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6, 1050 Bruxelles, Belgium
psmets@ulb.ac.be
http://iridia.ulb.ac.be/~psmets

March 15, 2001

Abstract: We present an interpretation of belief functions within a pure probabilistic framework, namely as normalized self conditional expected probabilities, and study their mathematical properties. Interpretations of belief functions appeal to partial knowledge. The self-conditional interpretation does this within the traditional probabilistic framework by considering surplus belief in an event emerging from a future observation, conditional on the event occurring. Dempster's original interpretation, in contrast, involves partial knowledge of a belief state. The modal interpretation, currently gaining popularity, models the probability of a proposition being believed (or proved, or known). The versatility of the belief function formalism is demonstrated by the fact that it accommodates very different intuitions.

Keywords: belief functions, self conditional expected probabilities, Dempster's model, probability of modal propositions.

1 Introduction

The interpretation of Dempster-Shafer belief functions continues to be an object of research and debate. (see the special issues of the Intern. J. Approx. Reasoning, vol. 4, 1990 and vol. 6, 1992, (Smets, 1994) and (Gabbay & Smets, 1998))

*This work has been partially funded by the CEC-ESPRIT IV Working Group FUSION (project 20001) for both authors, by by the Communauté Française de Belgique, ARC 92/97-160 (BELON) for the second author.

[1]. Usually, the label Dempster-Shafer is given to a theory where belief functions 'weight' sets and where belief functions are combined by the so-called Dempster rule of combination and conditionalized with the Dempster rule of conditionalization. The Dempster Shafer theory does not offer an interpretation of belief functions. Belief functions can be interpreted as quantified beliefs, as done in the transferable belief model (Smets & Kennes, 1994; Smets, 1998) or in the hint model (Kohlas & Monney, 1995). They can also be interpreted as random sets, probability of provability or convex families of probability measures (Smets, 1994). In some of these interpretations, Dempster's rules of combination and conditionalization do not apply.

This article proposes an interpretation of belief functions in terms of 'self-conditional expected probabilities'. It focuses on the interpretation of the basic belief assignment, without considering Dempster rule of combination. Dempster's original idea was that belief functions reflect constraints on belief-states (i.e. probabilities) induced by partial knowledge of a special kind. This is reviewed in the second section. The third section introduces self-conditional expected probabilities and shows how these capture some of the intuitions surrounding partial knowledge. Necessary and sufficient conditions that a basic belief mass can be interpreted in terms of self-conditional expected probabilities emerge. The fourth section briefly reviews the interpretation of belief functions as probability defined on modal propositions. Section five explores the relation with the belief functions of section three. A final section draws conclusions.

2 Upper and lower probabilities induced by a multiple valued mapping

Dempster (1967a, 1967b, 1968b, 1968a, 1969, 1972) studied belief functions while trying to solve the problem of fiducial inference. Dempster's approach assumes two finite spaces X and Y, a probability measure P^X on X, and a mapping $M : X \to 2^Y$ from X to the power set of Y. P^X induces a 'random set over Y' whose mass function is $P^{2^Y}(A) = P^X(\{x \in X | M(x) = A\})$ for $A \in 2^Y$. Note that P^{2^Y} is not a mass function on 2^Y induced by a mass function P^Y on Y in the normal way: $P(A) = \sum_{y \in A} P_y$. It is rather a mass function on the powerset 2^Y of Y induced by M which need not be constrained in any way be the subset algebra on 2^Y. P^{2^Y} is a *basic belief assignment* to the subsets of Y. This basic belief assignment is used to extract information regarding mass functions $P^Y(y)$

[1]As with any new theory, terminology regarding belief functions is rather volatile. We suggest the following definitions: *Belief Function Theory* is the theory of Choquet capacities monotone of order infinite. *Dempster Shafer Theory* is Belief Function Theory supplemented with a combination rule called Dempster's rule of combination. The *transferable Belief Model* is a special case of Dempster Shafer Theory in which the belief function represents quantified beliefs. It largely extends the model initially presented by Shafer. The *Hint Model* is a developed form of Dempster Shafer theory that fits with Probabilistic Argumentation Systems, where arguments are weighted by probabilities and support sets of hypothesis. It largely extends the model initially presented by Dempster. There are other theories that enter in the Belief Function Theory framework, like the *Random Sets Theory*, the *Upper and Lower Probability* theories, the probability theory extended to modal propositions like in the theory of *Probability of Knowing* and the *Probability of Proving* etc... In these theories, Dempster's rule is not always appropriate or justified, hence they would not fall under the Dempster Shafer theory as used here.

on the elements y of Y. We write P^Y for the probability *measure* on Y ... not to be confused with the mass function P^{2^Y} on 2^Y. All that can be stated about the probability $P^Y(A)$ that the actual (but unknown) value y_0 of Y is in $A \subseteq Y$ is that

$$P^Y(y_0 \in A) \in [bel^Y(A), pl^Y(A)]$$

where

$$\begin{aligned} bel^Y(A) &= P_*(A) = P^X(M_*(A)) = \sum_{M(x) \subseteq A} P^X(x), \\ pl^Y(A) &= P^*(A) = P^X(M^*(A)) = \sum_{M(x) \cap A \neq \emptyset} P^X(x); \end{aligned}$$

and where

$$\begin{aligned} M^*(A) &= \{x : x \in X, M(x) \cap A \neq \emptyset\} \\ M_*(A) &= \{x : x \in X, M(x) \subseteq A, M(x) \neq \emptyset\} \end{aligned}$$

By construction, these functions P_* and P^* so defined on Y are indeed belief and plausibility functions, respectively. This mathematical analogy explains why Shafer's theory of evidence (Shafer, 1976) was often understood as a special form of upper and lower probability theory.

This can be looked at in a different way. We are interested in a probability measure P^{XY} on the product space $X \times Y$. The marginal measure P^X is known and P^{XY} is constrained to satisfy $P^{XY}(x, y) = 0$ if $y \notin M(x)$. The conditional measures $P_x^Y(y) = P^{XY}(x, y)/P^X(x)$ which can arise in this way, for fixed x, are denoted

$$\widetilde{M(x)} = \{P_x^Y | P_x^Y(M(x)) = 1\}.$$

It is easy to see that

$$\sum_{x \in X} min_{P_x^Y \in \widetilde{M(x)}} P_x^Y(A) P^X(x) = \sum_{x : M(x) \subseteq A} P^X(x) = P_*(A) = bel^Y(A).$$

Similarly

$$\sum_{x \in X} max_{P_x^Y \in \widetilde{M(x)}} P_x^Y(A) P^X(x) = \sum_{x : M(x) \cap A \neq \emptyset} P^X(x) = P^*(A) = pl^Y(A).$$

This suggests an analogy with upper and lower probabilities. However, it must be noted that the minimum resp. maximum in the above equations can be realized in many ways.

We note that Dempster's approach requires a very particular partial knowledge of the belief state P^{XY}. From a strict' subjectivist' viewpoint, degree of belief in every event can be quantified, and it is difficult to imagine circumstances in which one would know P^X and the function $M(x)$ but would not know anything more. This is not to say that such situations cannot be constructed (see e.g. the hint models of (Kohlas & Monney, 1995)). The function $M(x)$ is not given a concrete interpretation. In the following section we propose an interpretation of belief functions which is purely probabilisitc, that is, which does not invoke partial knowledge of a belief state.

2.1 Conditioning

There are two obvious ways of conditionalizing in the context of Dempster's belief functions; conditioning on $\Xi \subseteq X$ or conditioning on $\Psi \subseteq Y$.

If one learns that the actual value of X belongs to $\Xi \subseteq X$, one computes the conditional probability function $P^X(.|\Xi)$ on X given Ξ, and then recomputes *bel* and *pl* from this conditional probability function.

If one learns that the actual value of Y belongs to $\Psi \subseteq Y$, then the M relation is changed into M_Ψ with $M_\Psi(x) = M(x) \cap \Psi$. In that case, for $A \subseteq Y$:

$$bel(A|\Psi) = bel(A \cup \overline{\Psi}) - bel(\overline{\Psi}) \qquad pl(A|\Psi) = pl(A \cap \Psi)).$$

on which case $m(\emptyset)$ can be positive. If one further conditions P^X on the x's such that $M_\Psi(x) \neq \emptyset$, one gets

$$bel(A|\Psi) = \frac{bel(A \cup \overline{\Psi}) - bel(\overline{\Psi})}{1 - bel(\overline{\Psi})} \qquad pl(A|\Psi) = \frac{pl(A \cap \Psi)}{pl(\Psi)}$$

This second conditioning rule is the classical Dempster's rule of conditioning, in which case $m(\emptyset) = 0$. Dempster's rule of combination has been proposed to combine two belief functions induced by two distinct pieces of evidence. Using the concept of specialization, other rules for combination have been described in order to cope with arbitrary pieces of evidence, Dempster's rule of combination being just a special case of these (Smets, 1998). Furthermore, all these combination rules correspond to conditioning on an uncertain event (Smets, 1993a).

It is evident that these notions of conditioning refer essentially to partial knowledge of the belief state P^{XY}. In the interpretation proposed below, there is no appeal to partial knowledge of a belief state, and the interpretation is not based on a mapping $M : X \mapsto 2^Y$. Hence these notions of conditioning, and their generalization as combination rules, do not apply. Rather, combination and conditionalization would proceed within the normal probabilistic framework, but these issues are not pursued further here.

3 The self conditional interpretation

In this section we propose a probabilistic reconstruction of intuitions underlying the notion of a basic belief assignment. S denotes the set of possible worlds or possible outcomes, and is assumed finite and $A \subseteq S$. The 'basic belief mass' $m(A)$ associated with A is glossed as (Smets & Kennes, 1994) ($\subset$ denotes strict subset).

$m(A)$ is the mass that supports A, and does not support any $A^* \subset A$.

$m(A)$ is a mass that could freely be given to any subsets of A if we were given new information.

The *basic belief assignment* $m : 2^S \rightarrow [0,1]$ satisfies

$$\sum_{A \subseteq S} m(A) = 1.$$

The *degree of belief* $bel(A)$ is defined as

$$bel(A) = \sum_{X \subseteq A} m(X). \tag{1}$$

$P(A)$	0.5
$P(x_1)$	0.1
$P(x_2)$	0.5
$P(x_3)$	0.4
$P(A\|X = x_1)$	0.1
$P(A\|X = x_2)$	0.3
$P(A\|X = x_3)$	0.85

Table 1: Probabilities and conditional probabilities for figure 1.

We propose to interpret the 'new information available to a subset of A ' as an observation in the normal probabilistic sense. If A were true, then $m(A)$ is the additional belief in A which we expect from the performing the observation. The key notion is termed a *self-conditional expected probability.* It is defined in section 3.1 below and its mathematical properties are explored in section 3.3. In section 3.4 we relate the self-conditional probabilities to the basic belief masses $m(A)$. While it is not the case that every basic belief assignment can be represented in this way, it does seem to capture properties of the underlying intuition.

3.1 Self conditional expected probabilities

The basic notion in the interpretation proposed below is the probability of an event A given some observation X, conditional on A. To explain this notion, let X be an observation taking one of 3 values x_1, x_2, x_3. The set of possible worlds is pictured in figure 1. The event A is the shaded area, and the sets $X = x_i$ are pictured as columns.

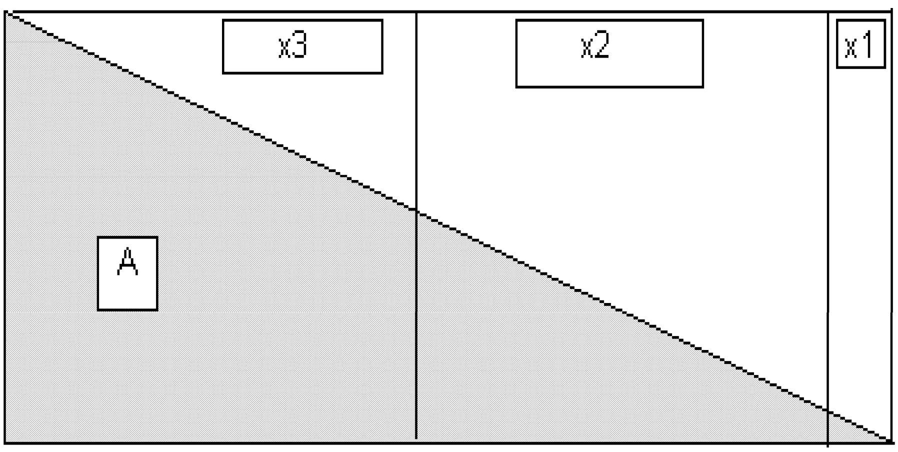

Figure 1: Example of event A with observation taking values x_1, x_2, x_3.

Suppose the values of the probabilities and conditional probabilities are as given in table 1.

Clearly, the posterior probability of A given X is a random variable whose value depends on the value of X. Of course, the expectation of this variable is just the probability of A:

$$E(P(A|X)) = \sum_{i=1}^{3} P(A|X = x_i)P(X = x_i) = P(A).$$

However, if we consider this expectation conditional on A occurring, then we compute :

$$\begin{aligned} E(P(A|X)|A) &= \sum_{i=1}^{3} P(A|x_i)P(x_i|A) = \sum_{i=1}^{3} P(A|x_i)^2 P(x_i)/P(A) \quad (2) \\ &= 0.67. \quad (3) \end{aligned}$$

By definition $E(P(A|X)|A) = 0$ if $P(A) = 0$. Supposing that A holds, we expect our probability of A to rise upon performing an observation X.

Assume that one given observation, say X, is being contemplated, but not yet performed. Supposing A to hold, $E(P(A|X)|A)$ then reflects a surplus in belief, beyond the current degree of belief $P(A)$, which we expect to become available to subsets of A, 'if we were given new information', i.e. the information obtained by performing observation X. The quantity $E(P(A|X)|A)$, after standardization, will be proposed as a probabilistic interpretation of $m(A)$. $E(P(A|X)|A)$ is not a belief state after observing X, but it does play a role in deliberating about possible observations, prior to deciding which observation to perform. This is illustrated in the following example.

3.2 Example

Suppose A is the event that a tumor is malignant. The complement $\overline{A}$ is the event that the tumor is benign. The a priori probability of A based on population data is 0.1. A diagnostic test X with values $0, 1$ may be performed. From clinical trials on patients whose tumors are known to be malignant resp. benign, it is determined that

$$P(X = 1|A) = 0.9;\ P(X = 0|A) = 0.1$$
$$P(X = 1|\overline{A}) = 0.2;\ P(X = 0|\overline{A}) = 0.8.$$

With Bayes' theorem we compute $P(A|X = 1) = 0.333$, $P(A|X = 0) = 0.013$ $E(P(A|X)|A)$ measures the belief in A which we expect to have after performing test X, given A:

$$E(P(A|X)|A) = P(A|X = 1)P(X = 1|A) + P(A|X = 0)P(X = 0|A) = 0.301$$

In reasoning about the value of test X for diagnosing A, it is natural to consider the difference $E(P(A|X)|A) - P(A) = 0.201$. If we had to choose between X and another equally expensive test Y with $E(P(A|Y)|A) = 0.5$, we should certainly prefer Y. If the tumor is malignant, then our expected increase of belief in A is greater for test Y than for test X. Self-conditional expected probabilities afford a natural way to compare different observations.

3.3 Mathematical properties

We explore some properties of $E(P(A|X)|A)$. Throughout, all random variables are finite; that is, they take finitely many possible values. $\overline{A}$ denotes the complement of $A : \overline{A} = S - A$. The proof of proposition (3.1) is similar to a proof in Savage (1954).

Proposition 3.1 *Let $A \subset S$ and let X be a random variable, then*

$$E(P(A|X)|A) \geq P(A).$$

Equality holds if and only if A is independent of X.

Proof. To prove the inequality it suffices to show

$$E(\frac{P(A|X)}{P(A)}|A) \geq 1.$$

This is equivalent to

$$\log(E(\frac{P(A|X)}{P(A)}|A)) \geq 0;$$

which by Jensen's inequality is entailed by

$$\begin{aligned} E(\log(\frac{P(A|X)}{P(A)})|A) &\geq 0; \\ E(-\log(\frac{P(A)}{P(A|X)})|A) &\geq 0; \end{aligned}$$

Again, by Jensen's inequality ($-\log$ is convex), this is entailed by

$$-\log(E(\frac{P(A)}{P(A|X)}|A)) \geq 0.$$

Now,

$$\begin{aligned} E(\frac{P(A)}{P(A|X)}|A) &= P(A)\sum_{x_i}\frac{P(x_i|A)}{P(A|x_i)} \\ &= P(A)\sum_{x_i}\frac{P(x_i|A)P(x_i)}{P(x_i|A)P(A)} = 1. \end{aligned}$$

Hence,

$$-\log(E(\frac{P(A)}{P(A|X)}|A)) = -\log(1) = 0.$$

Independence of A and X is clearly sufficient for equality. To show that this is also necessary, note that equality holds in the first application of Jensen's inequality only if

$$\log(E(\frac{P(A|X)}{P(A)})|A)) = (E\log(\frac{P(A|X)}{P(A)})|A)) = 0.$$

Since the logarithmic function is nowhere linear, this holds only if $P(A|x) = P(A)$ for all x. □

Corollary 3.1 *Let X and Y be random variables, then :*

$$1 \geq \sum_x P(Y = y|X = x)P(X = x|Y = y) \geq P(Y = y)$$

with the upper bound holding if $X = Y$ and the lower bound holding if X and Y are independent.

Information from observations X and Y are combined by simply considering the joint observation (X, Y). The next proposition gives an obvious but important property of combined information.

Proposition 3.2 *Let X and Y be random variables, then*

$$E(P(A|X, Y)|A) \geq E(P(A|X)|A).$$

Proof. With Bayes' theorem it suffices to show:

$$\sum_{x,y} \frac{P(A, x, y)^2}{P(x, y)} \geq \sum_x \frac{P(A, x)^2}{P(x)}$$

We look at this term-wise in x, it suffices to show:

$$\begin{aligned} \sum_y \frac{P(A, x, y)^2}{P(x, y)} &\geq \frac{P(A, x)^2}{P(x)} \\ &= \frac{[\sum_y P(A, x, y)]^2}{\sum_y P(x, y)}. \end{aligned}$$

For $y \in \{y_1, ...y_n\}$ write:

$$\begin{aligned} d_i &= \sqrt{(P(x, y_i))} : \\ c_i &= \frac{P(A, x, y_i)}{d_i}; \end{aligned}$$

then it suffices to show

$$\begin{aligned} \sum c_i^2 &\geq \frac{[\sum c_i d_i]^2}{\sum d_i^2}; \text{or} \\ \sum d_i^2 \sum c_i^2 &\geq [\sum c_i d_i]^2 \end{aligned}$$

which is the Cauchy Schwarz inequality. □

Self-conditional expected probabilities are sub-additive, as shown in the following proposition.

Proposition 3.3 *If $A \cap B = \emptyset$, and X a random variable, then:*

$$E(P(A \cup B|X)|A \cup B) \leq E(P(A|X)|A) + E(P(B|X)|B);$$

with equality holding if and only if (i) A is independent of X and B is independent of X, or (ii) $P(X = x|A) = P(X = x|B)$, for all x.

Proof. Term-wise in X, it suffices to show that

$$\frac{P(A \cup B|x)^2}{P(A \cup B)} \leq \frac{P(A|x)^2}{P(A)} + \frac{P(B|x)^2}{P(B)}$$

Write $d_A = \sqrt{p(A)}, d_B = \sqrt{p(B)}$, $c_A(x) = p(A|X = x)/d_A, c_B(x) = p(B|X = x)/d_B$, then the above becomes

$$(c_A(x)d_A + c_B(x)d_B)^2 \leq (cA(x)^2 + c_B(x)^2)(d_A^2 + d_B^2)$$

which is just the Cauchy Schwarz inequality. Equality holds if and only if $c_A(x) = \beta(x)d_A$ and $c_B(x) = \beta(x)d_B$ for some non-zero function β. Case (i) for equality is trivial. For case (ii), substitution and Bayes' theorem give

$$\beta(x) = P(A|X = x)/P(A) = P(X = x|A)/p(X = x) = P(X = x|B)/P(X = x.)$$

This gives the second case in which equality holds. □

Corollary 3.2 *Let A and B be disjunct subsets of* $\{X = x\}$*, then equality holds in Proposition 3.3.*

Proof. $P(x|A) = P(x|B) = 1$; if $x' \neq x, P(x'|A) = P(x'|B) = 0$. □

Considering the observation X to be fixed, it is convenient to introduce the notation

$$\mu(A) = E(P(A|X)|A). \quad (4)$$

Proposition 3.4 *If* $P(\overline{A}) > 0$*, then*

$$1 - \mu(\overline{A}) = \frac{P(A)(1 - \mu(A))}{P(\overline{A})}$$

Proof. This follows immediately from

$$P(A) = E(P(A|X)) = E(P(A|X)|A)P(A) + E(P(A|X)|\overline{A})P(\overline{A}).$$

□

We now consider that the field F of events whose 'basic belief' is to be measured is possibly smaller than the set of subsets of S. F is sometimes called the 'frame of discernment'. If X is an observation, then F_X denotes the smallest field on which X is measurable. F_X is generated by the events $X = x_i$, where i indexes the (finitely many) possible values of X. In Figure 1, F_X is the field generated by the columns.

Proposition 3.5 *If F is a field with n atoms, then for any probability P on F:* $\sum_{A \in F} P(A) = 2^{n-1}$.

Proof. The set of probabilities for which the proposition holds is clearly convex. If P is concentrated on an atom, then the proposition holds, since exactly half of the events in F contain any given atom. Any probability can be expressed as a convex sum of measures concentrated on atoms. □

3.4 Basic belief assignments

Definition 3.1 *For a given frame of discernment F with n atoms and a given observation X, the* basic belief mass generated by X *is defined for all $A \in F$ as:*

$$
\begin{aligned}
m_X(A) &= \frac{\mu(A) - P(A)}{2^{n-1}}; A \neq \emptyset; &(5)\\
m_X(\emptyset) &= 1 - \sum_{A \in F, A \neq \emptyset} m_X(A). &(6)
\end{aligned}
$$

From a remark to proposition (3.6) it follows that $m_X(\emptyset) > 0$. We say that a basic belief assignment m is derived from self-conditional probabilities if there is a random variable X such that $m = m_X$, i.e. if there is a random variable which generates m.

Note that

$$m_X(A) = \frac{1}{2^{n-1}} \sum_{x \in X} (P(A|x) - P(A))P(x|A)$$

so $m_X(A)$ may be regarded as the expected increased in 'information on A' given A. Note also that $m_X(S) = 0$. There is an interesting analogy with the Kullback-Leibler relative information $I(P(\bullet|x), P(\bullet))$ with respect to a partition $\{A_1, ..A_n\}$ of S [2]

$$I(P(\bullet|x), P(\bullet)) = \sum_{i=1}^{n} P(A_i|x)(lnP(A_i|x) - lnP(A_i)). \quad (7)$$

Note that the above is defined with respect to a given partition, and the summation goes over the elements of this partition. In comparison, the summation in (5) goes over the values of X. $m_X(A)$ is not restricted to a partition of S, but is defined for all subsets A.

Proposition 3.6 *If F_X is independent of F then $m_X(\emptyset) = 1$; if $F_X = F$, then $m_X(\emptyset) = 1 - 2^{-n+1}$.*

Proof. The first statement follows from the corollary to Proposition 3.1. If $F_X = F$, then by the corollary to Proposition 3.1, $\mu(A) = 1$ and

$$
\begin{aligned}
\sum_{A \in F; A \neq \emptyset} m_X(A) &= \sum_{A \in F; A \neq \emptyset} \frac{1 - P(A)}{2^{n-1}} = \\
\sum_{A \in F; A \neq \emptyset} \frac{P(\overline{A})}{2^{n-1}} &= \frac{2^{n-1} - 1}{2^{n-1}} = 1 - 2^{-n+1}.
\end{aligned}
$$

□

Remark If equality in the first equation is replaced by $\leq$, then the statement holds generally, and shows that $m_X(\emptyset) > 0$

[2]We are grateful: to an anonymous referee for drawing our attention to this fact

The first statement of proposition 3.6 says that if F is independent of F_X, then there is no belief available to subsets of $A \in F$ after observing X, since X can tell us nothing new about A. Hence, all mass is assigned to the empty set. The second statement says that if $F = F_X$, then the empty set gets minimal mass. Significantly, this statement does not depend on P.

Proposition 3.7

$$\begin{aligned} m_X(A) &\geq 0, && (8)\\ m_X(\overline{A}) &= \frac{P(A)m_X(A)}{P(\overline{A})};\ \textit{if } P(\overline{A}) > 0 && (9)\\ P(A) &= \frac{m_X(\overline{A})}{m_X(\overline{A}) + m_X(A)};\ \textit{if } P(\overline{A}) > 0. && (10)\end{aligned}$$

Proof. The first statement is immediate. The third is an equivalent form of the second. For the second statement, Proposition 3.4 gives:

$$\begin{aligned} \mu(\overline{A}) - P(\overline{A}) &= 1 - \frac{P(A)}{P(\overline{A})}(1 - \mu(A)) - P(\overline{A}) = \\ \frac{P(A)}{P(\overline{A})}(\mu(A) - 1) + P(A) &= \frac{P(A)}{P(\overline{A})}(\mu(A) - 1 + P(\overline{A})) \\ &= \frac{P(A)}{P(\overline{A})}(\mu(A) - P(A)). \end{aligned}$$

□

Proposition 3.8 *Let P and $\tilde{P}$ be probability measures defined on F. Let $\mu(A)$ and $\tilde{\mu}(A)$ be defined as above, and suppose that the sets $A \in F$ such that $\mu(A) = 1$ generate a strict subfield F' of F. Suppose further that for all $A \in F$, $\mu(A) = \tilde{\mu}(A)$; then $\tilde{P} = P$.*

Proof. : From proposition 3.4 we have for $A \notin F', 0 < P(A) < 1$:

$$\begin{aligned} \frac{1 - \mu(\overline{A})}{1 - \mu(A)} &= \frac{P(A)}{P(\overline{A})} \\ P(A) &= \frac{1 - \mu(A)}{1 - \mu(\overline{A}) + 1 - \mu(A)}. \end{aligned}$$

Since $\tilde{P}$ satisfies the same relation, $\tilde{P}(A) = P(A)$ for $A \notin F'$. Since $F' \neq F$, there must be an atom α of F which is not an element of F'. For any distinct atom C of $F', \alpha \notin C$, we have $C \cup \alpha \notin F'$ since otherwise $C' \cap (C \cup \alpha) = \alpha \in F'$. Since

$$P(\alpha \cup C) - P(\alpha) = P(C).$$

It follows that $P(C) = \tilde{P}(C)$ and hence $P = \tilde{P}$. □

It is instructive to compare m_X with the basic belief assignment in Dempster's basic belief assignment P^{2^Y}. We observe that P^{2^Y} may be an arbitrary mass function on 2^Y, whereas m_X is constrained as indicated in the above propositions. Since P^{2^Y} is arbitrary, we can set $S = Y$ and find some construction such that $P^{2^Y}(A) = m_X(A)$ for all A. The point of the self-conditional interpretation of the basic belief assignment, however, is that we do not *need* to do this. We have only one space S equipped with a probability mass function, and the basic belief mass emerges from the self conditional expected probabilities.

The function *bel* may be computed according to (1); however it does not yield a probability bound. To see this, it suffices to consider $A \subseteq X = x$. It is not difficult to verify that

$$\sum_{B \subseteq A} \frac{\mu(B) - P(B)}{2^{n-1}} = \frac{2^{\#A} P(A) P\{X \neq x\}}{2^n P\{X = x\}}.$$

This may of course be greater than $P(A)$. It does not appear possible to recover m_X from this probability and some mapping from S to 2^S. The next section addresses the question when a basic belief assignment can be derived from self conditional expected probabilities.

3.5 Conditions for $m = m_X$

Although self-conditional probabilities yield basic belief assignments which seem to capture something of the intuitive meaning of 'basic belief assignment', it is *not* true that every basic belief assignment can be represented as m_X for some random variable X. Indeed, it suffices to note that $m_X(S) = 0$ whereas this need not hold for an arbitrary basic belief assignment. Further, if A is independent of X, then $m_X(A) = m_X(\overline{A}) = 0$ and m_X can yield no information about $P(A)$. It is therefore appropriate to restrict attention to basic belief assignments m for which $m(A) = 0$ entails $A = S$.

We consider a finite set S with n elements and a basic belief assignment m satisfying $m(A) = 0 \Leftrightarrow A = S$. We say that m is derived from a self-conditional probability if we can define a probability mass function P on S and random variable X such that $m = m_X$. F_X denotes the field generated by sets of the form $\{X = x\}$.

If $A \in F_X$ then $\mu(A) = 1$; hence for $A \neq S, m_X(A)$ must satisfy

$$2^{n-1} m_X(A) = 1 - P(A) = P(\overline{A})$$

A similar equation must hold for $m_X(\overline{A})$. Adding these two equations yields the first condition below. The second condition is just equation (10).

$$\{A \subseteq S | m_X(A) + m_X(\overline{A}) = 2^{1-n}\} \cup \{S, \emptyset\} \text{ is a field.} \qquad (11)$$

$$P(A) = \frac{m(\overline{A})}{m(\overline{A}) + m(A)} \text{ for } A \neq \emptyset; P(\emptyset) = 0; \text{ is a probability.} \qquad (12)$$

Suppose that m satisfies (11). Let F_X denote the corresponding field, and define a random variable X assigning distinct constants to the elements of distinct atoms of F_X. The above conditions are clearly necessary for $m = m_X$. To obtain sufficient conditions, the P given by (12) may be substituted into (2) and the result into (4), and this finally into (5).

The definition of m_X refers to one observation X. An evident generalization would consider mixtures of measures m_X, by introducing lotteries over possible observations. This will not be pursued further here.

4 Probability functions extended to modal propositions.

4.1 The probability of believing

Conventionally, probabilities are allocated to sets. This is easily adapted in order that probability be defined on propositions. Indeed, in the finite case, propositions and subsets are in one to one correspondence. So using one or the other approach is a matter of convenience. Nevertheless, once one considers propositions, one can as well start considering what would become probability theory if the domain was no more classical propositions, but modal propositions.

Ruspini (1986, 1987) has proposed to consider $bel^{\Omega}_{Y,t}(A)$ as the probability that the agent Y *knows* at time t that A holds. Pearl (1988) proposed to understand it as the probability that A is *provable*. Both approaches fit essentially with the same ideas. Smets (1991, 1993b) takes Pearl's approach, but discusses at length the nature of the conditioning process. More recently Tsiporkova *et al.* (1999b, 1999, 1999a) also study the extension of probability to modal propositions.

Let $\Box p$ denote the modal proposition where the $\Box$ operator, called the box operator, denotes that proposition p is necessary, known, proved, etc... depending on which modal theory one is interested in. Our discussion will use the modality 'believing' as it is simple. We prefer 'believing' as we do not want to force the T axiom : $\Box p \to p$ that is usually introduced when knowing and proving are considered. Further one should not confuse this modality 'believing', which is categorical, with the weighted beliefs encountered in the subjective probability theory and in the transferable belief model. Here, $\Box p$ means the agent believes p.

Let $\Box$ satisfies the classical KD system (Chellas, 1980):

$$D : \Box p \to \neg\Box\neg p$$

$$K : \Box(p \to q) \to (\Box p \to \Box q)$$

Axiom D states that if you believe p, you do not believe its negation. Axiom K states: if you believe an implication and its antecedent, then you believe its consequent.

Under KD, one deduces:

$$AND : \Box(p \wedge q) = \Box p \wedge \Box q$$

that states that believing the conjunction of two propositions is equivalent to believing each of them individually.

The behavior of $\Box$ when the modal operators are nested are not considered here as regards the probability over sets of worlds, as they are not used.

Formally, let L be a finite propositional language, supplemented by the tautology and the contradiction. Let SL be the sentences of L formed by repeated applications of the connectives $\vee, \wedge, \neg$.

Let Ω be the set of worlds that correspond to the interpretations of L. For any $p \in SL$, let $[p] \subseteq \Omega$ be the set of worlds identified by p, i.e., those worlds where p is true:

$$[p] = \{\omega : \omega \in \Omega, \omega \models p\}.$$

Suppose the language L is enriched with the modal operator $\Box$ with the meaning of 'believing' (other meanings like 'proving' or 'knowing' could as well be considered). So each world is not only characterized by what is true in it, but also by what is believed in it. Formally, it means that there is an 'accessibility relation' R on $\Omega \times \Omega$ so that:

$$\omega \models \Box p \text{ iff } \forall \omega' \text{ such that } (\omega, \omega') \in R, \omega' \models p.$$

where p is a non-modal proposition in L (nested modalities are not considered).

We define the set $[!\Box p] \subseteq \Omega$ as the set of worlds where all the agent believes is p (and thus nothing more 'specific'). So:

$$[!\Box p] = \{\omega : \omega \in \Omega, \omega \models \Box p \text{ and } \forall q \neq p, q = p \wedge q, \omega \not\models \Box q\}.$$

By construction, if $p, q \in SL$, then either $[!\Box p] = [!\Box q]$ (when $p \equiv q$) or $[!\Box p] \cap [!\Box q] = \emptyset$. The sets $[!\Box p]$ for $p \in SL$ constitute a partition of Ω. We also have:

$$[\Box p] = \bigcup_{q \in SL, q = p \wedge q} [!\Box q].$$

Let P be a probability measure on Ω. Then define $bel : 2^{\Omega} \to [0,1]$ such that:

$$bel([p]) = P(\{\omega : \omega \in \Omega, \omega \models \Box p\}) = P([\Box p]).$$

This function is indeed a belief function on Ω.

To see this, it is sufficient to find the basic belief assignment that corresponds to bel. Let the basic belief assignment $m : 2^{\Omega} \to [0,1]$ be defined by:

$$m([p]) = P(\{\omega : \omega \in \Omega, \omega \models !\Box p\}) = P([!\Box p]).$$

The $m([p])$ terms are non negative. As the sets $[!\Box p]$ are the elements of a partition of Ω, their sum is 1. Therefore m is indeed a basic belief assignment on Ω.

Then

$$bel[p]) = P([\Box p]) = \sum_{[q] \subseteq \Omega, [q] \subseteq [p]} P([!\Box q]) = \sum_{[q] \subseteq \Omega, [q] \subseteq [p]} m([q])$$

what is just the relation linking bel to its basic belief assignment m. So bel is indeed a belief function.

We note that this interpretation links the belief functions with a much broader domain of potential application than the interpretation in the previous section; namely, the domain of modal discourse. However, for this interpretation to work, we need more than simply a probability measure P on Ω, we need the relation R providing the semantic interpretation of the modal operator.

5 Comparison of Modal and Self-conditional interpretations

The intuitions underlying the self-conditional and modal interpretations are quite different. This may be illustrated by considering a simple, though generic, realization of the modal interpretation. Consider a formal propositional calculus generated by n logically independent propositional constants $p_1, ...p_n$. We consider the free algebra generated by these; this is an algebra with 2^n atoms, each atom being a conjunction $\bigwedge_{i=1...n} r_i$, where $r_i = p_i$ or $r_i = \neg p_i$, the negation of p_i. A model of this language is just an ultrafilter in this algebra; in this case that is the set of propositions entailed by a single atom. For model ω, denote this atomic proposition as a_ω.

Suppose that the 'accessibility relation' R is symmetric. This is obtained by adding schema B : $\neg p \rightarrow \Box\neg\Box p$, that states: 'if p is false, You believe that You do not believe p'. For $\omega \in \Omega$, call the set of models accessible from ω, the *orbit* of $\omega, O(\omega)$. Then $\omega' \in O(\omega) \Rightarrow O(\omega') = O(\omega)$. The orbits partition Ω. Define

$$q_\omega = \bigvee_{\omega' \in O(\omega)} a_{\omega'}$$

Each orbit is the set of models satisfying the disjunction of the atoms of models in the orbit, and q_ω is the strongest proposition with that property.

Let $A(R)$ be the subalgebra (or equivalently subfield) generated by the orbits. It is easy to verify the following facts. The modal basic belief assignment assigns zero belief to any proposition which is not an atom of $A(R)$. To atoms of $A(R)$, the basic belief assignment is just the probability antecedently defined over Ω .

The difference with the self-conditional basic belief assignment m_X is striking. According to proposition 1, $m_X(B) = 0$ only if B is independent of X, i.e. if B is independent of all sets of the form $\{X = x\}$. In particular, $m_X(S) = 0$; that is, new information can never yield more belief in S. On the other hand, in the modal interpretation with symmetric accessibility, $m(B) = 0$ for any B which is not an atom of $A(R)$. If the relation R is empty, such that no world is accessible from any other, then the orbits are just the singletons $\{\omega\}$, and only the atomic propositions have non-zero basic belief mass. If the trivial proposition has positive basic belief mass, then Ω must be an orbit of R, and since the orbits partition Ω, this is the only orbit. In other words, every world is accessible from every other. In this case the basic belief mass of the trivial proposition is one.

6 Conclusions

Belief functions model partial knowledge. We have argued that the original interpretation of belief functions suggested by Dempster is restricted in so far as it models only one very particular state of partial knowledge of a belief state. The self-conditional expected probability interpretation does not invoke partial knowledge of a belief state. Instead, it interprets the basic belief assignment as *new* belief becoming available from new information via an observation. The modal interpretation models the probability that a proposition is believed.

The intuitions underlying these interpretations all find expression within the Dempster Shafer formalism, but in many respects they are antithetical. This is illustrated by the basic belief mass assigned to the trivial proposition. In Dempster's interpretation this mass can take any value between zero and one. In the self-conditional interpretation this mass is zero. In the modal interpretation with symmetric accessibility, the trivial proposition has basic belief mass zero or one. In the same vein, any event can have basic belief mass zero in Dempster's approach, in the self conditional approach zero basic belief mass is equivalent with independence with respect to the observational field F_X. In the symmetric modal approach, every proposition which is not an atom of the orbit algebra has zero basic belief mass.

It emerges that the belief function formalism allows expression of very different intuitions having very different properties. Probability theory also has different interpretations, e.g. the frequency interpretation, the subjective interpretation, the logical interpretation and the propensity interpretation. The formal properties of probability are not identical in these interpretations. For example, subjective probabilities are sometimes said to be finitely and not countably additive, propensities are said to be Renyi spaces rather than normalized probabilities. Nonetheless, these formal differences are not great and have no impact on applications. With regard to belief functions, in contrast, the basic belief assignments in different interpretations have very different formal properties. In general, it is impossible that a 'Dempster basic belief mass' could be a self-conditional basic belief mass, and neither could be a modal basic belief mass. Appreciating this fact may help clarify the debate surrounding belief functions.

References

Chellas, B. F. (1980). *Modal logic.* Cambridge Univ. Press, G.B.

Dempster, A. P. (1967a). Upper and lower probabilities induced by a multiple valued mapping. *Ann. Math. Statistics*, *38*, 325–339.

Dempster, A. P. (1967b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika*, *54*, 515–528.

Dempster, A. P. (1968a). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statistics*, *39*, 957–966.

Dempster, A. P. (1968b). A generalization of Bayesian inference. *J. Roy. Statist. Soc. B.*, *30*, 205–247.

Dempster, A. P. (1969). Upper and lower probability inferences for families of hypothesis with monotone density ratios. *Ann. Math. Statistics*, *40*, 953–969.

Dempster, A. P. (1972). A class of random convex polytopes. *Annals of Mathematical Statistics*, *43*, 250–272.

Gabbay, D. M., & Smets, P. (Eds.). (1998). *Handbook of defeasible reasoning and uncertainty management systems, vol. 1.* Kluwer, Doordrecht, The Netherlands.

Kohlas, J., & Monney, P. A. (1995). *A mathematical theory of hints: An approach to Dempster-Shafer theory of evidence.* Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Pub. San Mateo, Ca, USA.

Ruspini, E. H. (1986). *The logical foundations of evidential reasoning* (Tech. Rep.). Technical note 408, SRI International, Menlo Park, Ca, US.

Ruspini, E. H. (1987). Epistemic logics, probability and the calculus of evidence. In IJCAI-87 (Ed.), *Intern. joint conf. on artificial intelligence* (pp. 924–931). Morgan Kaufman, San Mateo, Ca.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton Univ. Press. Princeton, NJ.

Smets, P. (1991). Probability of provability and belief functions. *Logique et Analyse*, *133*, 177–195.

Smets, P. (1993a). Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, *9*, 1–35.

Smets, P. (1993b). Probability of deductibility and belief functions. In M. Clarke, R. Kruse, & S. Moral (Eds.), *Symbolic and quantitative approaches to uncertainty* (pp. 332–341). Springer-Verlag, Berlin.

Smets, P. (1994). What is Dempster-Shafer's model? In R. R. Yager, J. Kacprzyk, & M. Fedrizzi (Eds.), *Advances in theDempster-Shafer theory of evidence* (pp. 5–34). Wiley, New York.

Smets, P. (1998). The transferable belief model for quantified belief representation. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems, vol. 1* (pp. 267–301). Doordrecht, The Netherlands: Kluwer.

Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, *66*, 191–234.

Tsiporkova, E., Boeva, V., & De Baets, B. (1999). Dempster-Shafer's theory framed in modal logic. *Int. J. Approx. Reasoning*, *to appear.*

Tsiporkova, E., De Baets, B., & Boeva, V. (1999a). Dempster's rule of conditioning translated into modal logic. *Fyzzy Sets and Systems*, *102*, 371–383.

Tsiporkova, E., De Baets, B., & Boeva, V. (1999b). Evidence theory in multivalued models of modal logic. *J. Appl. Non-Classical Logics*, *to appear.*