THE TRANSFERABLE BELIEF MODEL AND RANDOM SETS.

Philippe SMETS[1]
IRIDIA, Université Libre de Bruxelles
50 av. F. Roosevelt, CP194/6, 1050 Brussels, Belgium.

1. Introduction.

The Transferable Belief Model (TBM) is a model for quantified belief based on the use of belief functions. It corresponds essentially to Shafer's initial proposal as described in his book (1976), except for the following adaptations and explicitations.

1) Quantified beliefs are **point-valued**, not interval-valued
2) Any **connection with randomization or necessary additivity** as encountered within probability theory has been explicitly eliminated.
3) A difference has been established between **open and closed world assumptions** (Smets 1988). The **normalization** after conditioning and combination is not performed in the open world context.
4) A **two-level model** for the beliefs has been proposed (Smets 1989a). It consists of a **credal level** where beliefs are entertained and a **pignistic level** where beliefs are used to make decisions. At the credal level, beliefs are quantified by belief functions. At the pignistic level, beliefs are quantified by probability functions. When a decision must be made, the beliefs at the credal level are transformed into beliefs at the pignistic level, i.e. there exists a transformation from the belief functions to the probability functions. It is called the **pignistic transformation** (Smets 1989b). It corresponds to the Generalized Insufficient Reason Principle.
5) The justification of the TBM is based on the idea that the impact of an evidence consists in allocating parts m(A) of an initial unitary amount of belief among the propositions A of a given algebra. m(A) is that part of our belief that supports A and that, due to lack of information, does not support any strict subproposition of A. The m are called the basic belief masses (bbm).
6) The definition of bel (and pl) are derived from the bbm, and the inequalities among the belief functions are deduced.

Other interpretations for the use of belief functions to quantify belief have been proposed. Among them, it has been suggested that belief functions based models are analogous to models based on random sets (Nguyen 1978). If correct it would permits to use the

[1] The following text presents research results of the Belgian National incentive-program for fundamental research in artificial intelligence initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsability is assumed by its author. These researches have been partially supported by the project DRUMS which are funded by grants from the Commission of the European Communities under the ESPRIT Basic Research Project 3085.

probability theory apparatus in order to justify the use of belief functions. We study in this paper if such relation exists and under what conditions. Our conclusions will be that the two models are different and should not be confused because they behave differently when conditioning is involved.

We summarizes the mathematic of beliefs functions and random sets. We then insist on the existence of two components, the static and the dynamic, in any model for quantified beliefs. We finally show that the static components of the two models are identical but not their dynamic parts.

2. The Transferable Belief Model.

Let Ω be a non empty finite set called the frame of discernment equipped with the Boolean algebra 2^{Ω} of its subsets. Every element of 2^{Ω} is called a proposition. $\emptyset$ represents the contradiction.

On Ω, we define a "valuation" i.e. a map from Ω to {true, false} such that under the closed-world assumption one and only one element of Ω is true, this element is called the 'truth'. Under the open-world assumption at most one element of Ω is true. A proposition is true iff one of its elements is true. The elements of Ω can be seen as possible answers to a certain question, one and only one answer - the truth - being correct.

A **credal state** on a frame of discernment Ω is a normative description of our subjective, personal judgment that propositions $A \in 2^{\Omega}$ are true, i.e. contain the correct answer, for all A in 2^{Ω}. It results from known, possibly inconclusive pieces of evidence that induce partial beliefs on the propositions of 2^{Ω}.

The **transferable belief model** postulates that the impact of an evidence consists in allocating parts of an initial unitary amount of belief among the propositions of Ω. For $A \in 2^{\Omega}$, m(A) is the part of our belief that supports A i.e. that the 'truth' is in A, and that, due to lack of information, does not support any strict subproposition of A. The m's are called the **basic belief masses** (bbm).

Let m: $2^{\Omega} \rightarrow [0,1]$ with

$$\sum_{A \in 2^{\Omega}} m(A) = 1 \qquad \text{and} \qquad m(\emptyset) = 0.$$

The difference with classical probability models is that masses can be given to any proposition of 2^{Ω} whereas within probability models masses are given only to the elements of Ω.

If further evidence becomes available and implies that the truth is in a subset B of A, then the mass m(A) initially allocated to A is transferred to B. Hence the name of the transferable belief model. This transfer of belief corresponds to the so-called **Dempster's rule of conditioning**. Let m be a bbm on Ω and suppose the conditioning evidence that the truth is in $B \in 2^{\Omega}$, the bbm m is transformed into m_B: $2^{\Omega} \rightarrow [0,1]$ with:

$$m_B(A) = c \sum_{X \subseteq \overline{B}} m(A \cup X) \quad \text{for } A \subseteq B$$

$$m_B(A) = 0 \quad \text{for } A \not\subseteq B$$

with $c = 1$ (open-world assumption)

$$c = \frac{1}{1 - \sum_{X \subseteq \overline{B}} m(X)}$$ (closed-world assumption)

Given Ω, the **degree of belief** of $A \in 2^{\Omega}$, bel(A), quantifies the total amount of belief supporting A without supporting $\overline{A}$. It is obtained by summing all the basic belief masses given to proposition $X \in 2^{\Omega}$ with $X \subseteq A$.

$$bel(A) = \sum_{X \subseteq A} m(X) \qquad bel(\emptyset) = 0$$

The Dempster's rule of conditioning expressed with bel is:

$$bel(A|B) = bel(A \cup \overline{B}) - bel(\overline{B})$$ (open-world assumption)

$$bel(A|B) = \frac{bel(A \cup \overline{B}) - bel(\overline{B})}{1 - bel(\overline{B})}$$ (closed-world assumption)

Suppose two distinct pieces of evidence E_1 and E_2 and the corresponding belief functions bel_1 and bel_2 induced on Ω. The two belief functions are then combined by Dempster's rule of combination into bel_{12} with:

$m_{12}(A)$ = k \i\su(X\S\UP1(⊆)Ω, Y\S\UP1(⊆)Ω, X∩Y=Ø,, $m_1(A \cup X)$ $m_2(A \cup Y)$)

with k = 1 (open-world assumption`)

k = 1 / (1 - \i\su(X\S\UP1(⊆)Ω, Y\S\UP1(⊆)Ω, X∩Y=Ø,, $m_1(X)$ $m_2(Y)$)) (closed-world assumption)

To study a model for quantified belief, one must consider not only its **static component** (how beliefs are assigned to propositions?) but also its **dynamic component** (how beliefs are updated?). In the TBM, the static component corresponds to the basic belief masses assignment and the dynamic component to the transfer of those basic belief masses among the propositions.

It is important to note that the TBM includes the **two components**. Many authors working on Dempster-Shafer's model, when they do not confound it with some pure upper and lower probabilities model, consider only the static component (the bbm assignment). When necessary they introduce Dempster's rule of combination and Dempster's rule of conditioning as a special case of the last. We think that this approach can be criticized, as far as the concept of updating on true facts (conditioning) is more fundamental then the concept of combining the belief functions induced by two distinct pieces of evidence. A comparison with probability theory is worth considering.

Suppose bel_1 and bel_2 (with m_1 and m_2 the corresponding bbm) are two probability functions (they correspond to so-called Bayesian belief functions). Dempster's rule of combination applied to these two probability functions reduces itself under closed-world assumption (the only one considered in classical probability theory) to

$$m_{12}(x) = \frac{m_1(x)\ m_2(x)}{\sum_{y\in\Omega} m_1(y)\ m_2(y)} \quad \text{for } x\in\Omega \qquad \text{and} \qquad m_{12}(A) = 0 \text{ for } |A|>1.$$

If 1) the two belief functions bel_1 and bel_2 are based on the same equi a priori probability on Ω and 2) the two pieces of evidence that induced bel_1 and bel_2 are conditionaly independent for each $x\in\Omega$, then $m_{12}(x)$ derived by the application of Dempster's rule of combination is the solution one will obtain by the application of probability theory.

From these equations, one could define the conditioning process. Suppose $m_2(x) = 1/\ |X|$ for $x\in X\subseteq\Omega$ and 0 otherwise. Then

$$m_{12}(x) = \frac{m_1(x)}{\sum_{y\in X} m1(y)} \quad \text{for } x\in X \qquad \text{and } m_{12}(x) = 0 \text{ otherwise.}$$

bel_{12} is a probability function and for any $A\subseteq\Omega$, one has

$$bel_{12}(A) = bel_1(A|X) = \frac{bel_1(A\cap X)}{bel_1(X)}$$

thus the conditioning rule encountered in probability theory.

Is this an adequate definition of the concept of a conditional probability function. The conditional probability $P(x|X)$ is no more defined as $P(x)/P(X)$ for $x\in X$. It is defined as the result of the combination (with the equi prior and the conditional independence being postulated) of the probability function P with a probability function P' such that $P'(x) = 1/\ |X|$ for $x\in X$ and 0 otherwise. Could one accept such a new definition of the conditional probability? It is mathematically correct but it looks like a surrealistic definition. To define conditioning with belief functions in general as a special form of combination is equivalently odd and thus inadequate.

3. Random set model.

Some authors (Nguyen 1978) propose to interpret the basic belief masses as probability densities on 2^Ω. We present the theory of random sets through some examples.

Suppose an urn U with n numbered balls. Let $\Omega = \{x_i: i=1, 2\ldots n\}$ be the set of balls in U. I will take an handful of balls out of U. Suppose m(A) is the probability that I will extract the set $A\subseteq\Omega$, with $\sum_{A\subseteq\Omega} m(A) = 1$. Beware that $m(\emptyset)$ might be positive (just as under the open-world assumption), as I might leave all the balls in the urn.

Let S be the set randomly extracted from the urn. E.g. one has:

$$\Pr(S\in \{\{x_1\},\{x_1,x_2\}\}) = \Pr(S\in \{\{x_1\}\}) + \Pr(S\in \{\{x_1,x_2\}) = m(\{x_1\}) + m(\{x_1,x_2\})$$

where Pr is the probability that the set S belongs to a set of subsets of Ω. $\Pr : 2^{2^\Omega}\rightarrow[0, 1]$, whereas $m : 2^\Omega\rightarrow[0, 1]$. m is a probability density on 2^Ω and Pr is the corresponding probability distribution on 2^{2^Ω}.

For our comparison with belief functions, we are interested in two types of probabilities: those concerning $2^A = \{B: B\subseteq A\}$, the power set of $A\subseteq\Omega$, and those concerning $\kappa(A) = 2^A - \emptyset = \{B: B\subseteq A, B\neq\emptyset\}$, the non empty sets of the power set of $A\subseteq\Omega$.

$\Pr(S\in 2^A)$ is the probability that the randomly selected set is a subset of A.
$\Pr(S\in \kappa(A))$ is the probability that the randomly selected set is a non-empty subset of A.
One has: $\Pr(S\in 2^A) = \Pr(S\in \kappa(A)) + m(\emptyset)$

Let bel(A), $A\subseteq\Omega$, be defined such that bel(A) = $\Pr(S\in \kappa(A))$ = \i\su(B⊆A,B≠Ø,, m(B)).
bel is an unnormalized belief function as $bel(\Omega) = 1 - m(\emptyset)$ and not 1.
The plausibility function is given by $pl(A) = bel(\Omega) - bel(\overline{A})$.

Those who assimilate belief functions with probabilities of random sets use this relation to justify themselves (and postulate $m(\emptyset) = 0$, i.e. the handful is never empty, an innocuous assumption at this level of the presentation). The static component of the two theories are similar.

To compare the dynamic components, we show the forms of conditioning rules that can be defined with random sets. We restrict ourselves to the events of the form 2^A and $\kappa(A)$.

Case 1a.

$$\Pr(S\in 2^A|S\in 2^B) = \frac{\Pr(S\in 2^{A\cap B})}{\Pr(S\in 2^B)} = \frac{bel(A\cap B) + m(\emptyset)}{bel(B) + m(\emptyset)}$$

Case 1b.

$$\Pr(S\in\kappa(A)|S\in\kappa(B)) = \frac{\Pr(S\in\kappa(A\cap B))}{\Pr(S\in\kappa(B))} = \frac{bel(A\cap B)}{bel(B)}$$

Case 1b is the geometric rule of conditioning (Suppes and Zanottii, 1977) also called the strong conditioning in Planchet (1989).

Case 2a.

$$\Pr(S\in 2^A|S\notin 2^{\overline{B}}) = \frac{\Pr(S\in 2^A\cap S\notin 2^{\overline{B}})}{1 - \Pr(S\in 2^{\overline{B}})}$$

$$= \frac{\Pr(S\in 2^A) - \Pr(S\in 2^{A\cap\overline{B}})}{1 - \Pr(S\in 2^{\overline{B}})} = \frac{bel(A) - bel(A\cap\overline{B})}{pl(B)}$$

Case 2b.

$$\Pr(S\in\kappa(A)|S\notin\kappa(\overline{B})) = \frac{\Pr(S\in\kappa(A)\cap S\notin\kappa(\overline{B}))}{1 - \Pr(S\in\kappa(\overline{B}))} = \frac{bel(A) - bel(A\cap\overline{B})}{1 - bel(\overline{B})}$$

When $m(\emptyset) = 0$, cases 2a and 2b are identical and correspond to the weak rule of conditioning in Planchet (1989). When only one ball can be extracted from the urn, rule 1b and 2b become identical: to be in B or not in $\overline{B}$ are equivalent events.

To get Dempster's rule of conditioning, one must adapt somehow the process. Suppose the random set S is selected from a list of all the subsets of Ω with m(A) being the probability that the selected random set is $A\subseteq\Omega$ (as before). Let $B\subseteq\Omega$. Suppose all balls in $\overline{B}$ are eliminated from the urn and only those in B remain in it. Given the selected random set S, we extract from the urn all those balls still in the urn (those in B) that are member of S. Let Z be the set of balls really extracted from the urn. Then $Z = S\cap B$. Let $\Pr_B$ be the probabilities on 2^{2^Ω} derived in such context.

Case 3a.

$$\Pr_B(Z\in 2^A) = \Pr(S\in 2^{A\cup\overline{B}}) = bel(A\cup\overline{B}) + m(\emptyset)$$

Case 3b.

$$\Pr_B(Z\in\kappa(A)) = \Pr((S\in 2^{A\cup\overline{B}})\cap(S\cap B\neq\emptyset))$$

$$= \Pr(S\in\kappa(A\cup\overline{B})) - \Pr(S\in\kappa(\overline{B})) = bel(A\cup\overline{B}) - bel(\overline{B})$$

what is the unnormalized Dempster's rule of conditioning (open-world assumption)

Case 3c.

$$\Pr_B(Z\in\kappa(A)|S\cap B\neq\emptyset) = \frac{\Pr((S\in 2^{A\cup\overline{B}})\cap(S\cap B\neq\emptyset))}{P(S\cap B\neq\emptyset)} = \frac{bel(A\cup\overline{B}) - bel(\overline{B})}{bel(\Omega) - bel(\overline{B})}$$

what is the normalized Dempster's rule of conditioning (closed-world assumption)

So the conditionings encountered in the TBM correspond to the probability that the subset of the randomly selected set S that belongs to B is a subset of A, (given it is not empty if one works under the closed-world assumption).

These three cases show nicely the origin of the three terms that appear in Dempster's rule of conditioning. The subtraction of bel($\overline{B}$) is due to the fact we restrict ourselves to $\kappa(A)$ instead of 2^A. The divisor reflects the conditioning on the fact that the selected random set S and the set B of balls left in the urn have a non-empty intersection (i.e. some balls will be in the handful after the extraction from the urn).

4. The sublimable ball paradigm.

To show how the case 3c can be observed, we describe in detail an example of random sets that leads to the conditioning described under case 3c.

Let us consider an urn full of cold water, with four balls, three of them (x, y, z) made of stone, the fourth (w) made of a product that is solid in cold water but sublimates (disappears by becoming gaseous) instantaneously at room temperature leaving no trace. I am going to extract from the urn a non empty set of balls with m(A) being the probability that the set $A \neq \emptyset$, $A \subseteq \Omega = \{x,y,z,w\}$ is selected. You want to bet that the selected set is a subset of some given set $X \subseteq \Omega$. Given its nature, if ball w is selected, you will not know it as w will disappear before you can observe it. So you will observe the set {x} whenever the randomly selected set is {x} or {x,w}. Suppose furthermore that I tell you that there is still at least one ball in my hand after sublimation, so the extracted non empty set is not {w}. The conditional probability that the observed set is a subset of {x,y} is

$$\frac{P(\{\{x\},\{y\},\{x,y\},\{x,w\},\{y,w\},\{x,y,w\}\})}{1 - P(\{\{w\}\})}$$

$$= \frac{P(\kappa(\{x,y,w\})) - P(\kappa(\{w\}))}{1 - P(\kappa(\{w\}))}$$

5. Conclusions.

Is there any connection between paradigms like the sublimable ball one and what the TBM tends to model? It is obviously an open question. I cannot make up my mind about such a relation. But maybe somebody will someday. The aim of the TBM is to quantify our degree of belief that some propositions are true, being accepted that one and only one of the elementary propositions of Ω is true (or at least one being true under the open-world assumption) .

There is a morphism between the TBM and the theory of random sets when the conditioning is defined as in case 3c.[1] But this doesn't mean that the TBM is a particular

[1] In fact even Dempster's rule of combination can be derived in both theory: take two handfuls of balls, and consider the set of balls in both handfuls.

form of of the random set theory. We see a serious difference between the two theories that can be illustrated through the following linguistic argument.

The following statements x1 (x = A, B, C) deal with the TBM (i.e. the fact that the true elementary proposition belongs to a subset of Ω), statements x2 deal with classical probability theory (i.e. an element of Ω is selected randomly), statements x3 deal with probabilities of random sets (i.e. a subset of Ω is selected randomly).

A1: bel(A) is the belief that the truth is in A
A2: Pr(A) is the probability that the randomly selected element is in A.
A3: $\Pr(S\in \kappa(A))$ is the probability that the randomly selected subset S is a non empty subset of A.

Suppose A and B are two subsets of Ω and $A\cap B = \emptyset$
B1: $\mathrm{bel}(A\cup B)$ is the belief that the truth is in A or B

$$\mathrm{bel}(A\cup B) \geq \mathrm{bel}(A) + \mathrm{bel}(B)$$

B2: $\Pr(A\cup B)$ is the probability that the randomly selected element is in A or B

$$\Pr(A\cup B) = \Pr(A) + \Pr(B)$$

B3: $\Pr(S\in \kappa(A\cup B))$ is the probability that the randomly selected subset S is a non empty subset of $A\cup B$.

$$\Pr(S\in \kappa(A\cup B)) \geq \Pr(S\in \kappa(A)) + \Pr(S\in \kappa(B))$$

In fact the equivalence between relations B1 and B3 can be generalized for all the inequalities encountered with belief functions. It is the origin of the assimilation between belief functions models and random sets models. The static components of the two models are identical.

But once conditioning is introduced the analogy disappears.
C1: $\mathrm{bel}(A|B)$ is the belief that the truth is in A given that the truth is in B.

$$\mathrm{bel}(A|B) = \mathrm{bel}(A\cup \overline{B}) - \mathrm{bel}(\overline{B})$$

C2: $\Pr(A|B)$ is the probability that the randomly selected element is in A given he randomly selected element is in B.

$$\Pr(A|B) = \Pr(A\cap B) / \Pr(B)$$

C3a: $\Pr(S\in \kappa(A)|S\in \kappa(B))$ is the probability that the randomly selected subset S is a non empty subset of A given that the randomly selected subset S is a non empty subset of B

$$\Pr(S\in \kappa(A)|S\in \kappa(B)) = \Pr(S\in \kappa(A\cap B)) / \Pr(S\in \kappa(B))$$

(see case 1b)
C3b. $\Pr_B(Z\in \kappa(A))$ is the probability that the intersection between B and the randomly selected set S is a non empty subset of A.

$$\Pr_B(Z\in \kappa(A)) = \Pr((S\in 2^{A\cup \overline{B}})\cap(S\cap B\neq\emptyset)) = \Pr(S\in \kappa(A\cup \overline{B})) - \Pr(S\in \kappa(\overline{B}))$$

(see case 3b)

Statements C1 and C3a are analogous but the conditioning rules are different. To get Dempster's rule of conditioning one must use C3b, but it is not a statement linguistically equivalent to C1.

The TBM speaks about a set of elementary propositions, one and only one being true under the closed-world assumption. So one element of Ω is somehow labelled by 'true', but we can only assess beliefs about which element is labelled, and the fact that the labelled element belongs to some given sets. The problem studied within the TBM is exactly the same as the one studied in subjective probability theory. This is easily seen by comparing A1 and A2, B1 and B2 and C1 and C2. But the mathematical rules are different, being more general with belief functions.

In random set, a subset (S) is labelled and we can assess our belief that the labelled subset will be a subset of some other given set of subsets on Ω.

How to pass from one model to the other is not clear for the author, if only such a relation exists. Further more how to explain the conditioning process is still not clear.

The mathematical analogy between the TBM and the theory of random sets (when using conditioning according to case 3c) might nevertheless be useful to create some yardsticks to assess degrees of belief. This mathematical analogy can be compared with the one between bayesian probabilities and objective probabilities as encountered when exchangeable bets are introduced. This might provide a method to assess the numerical values of degrees of belief.

Acknowledgement.

The author is indebted to H.T. Nguyen for his useful comments and suggestions.

Bibliography.

NGUYEN H.T. (1978) On random sets and belief functions. J.Math.Anal. and Applic. 65:531-542.

PLANCHET B. (1989) Credibility and conditioning. J. Theor. Probabil. 2:289-299.

SHAFER G. (1976) A mathematical theory of evidence. Princeton Univ. :Press. Princeton, NJ.

SMETS Ph. (1988) Belief functions. in SMETS Ph, MAMDANI A., DUBOIS D. and PRADE H. ed. Non standard logics for automated reasoning. Academic Press, London pg 253-286.

SMETS Ph. and KENNES R. (1989a). The transferable belief model: comparison with Bayesian models. Technical report, IRIDIA-89

SMETS Ph. (1989b) Construucting the pignistic probability function in a context of uncertainty. Proc. Fifth Workshop on Uncertainty in AI, Windsor, Canada, 319-326. also in: Uncertainty in Artificial Itelligence 5, Henrion M., Shachter R.D., Kanal L.N. and Lemmer J.F. ed., North Holland, Amsterdam 1990, 29-39.

SMETS Ph. (1990) Belief functions and their combinations.IEEE PAMI 12, 447-458.

SUPPES P. and ZANOTTI M. (1977) On using random relations to generate upper and lower probabilities. Synthesis 36:427-440.