Conditioning in Dempster-Shafer theory: prediction vs. revision

Didier Dubois and Thierry Denœux

Abstract We recall the existence of two methods for conditioning belief functions due to Dempster: one, known as Dempster conditioning, that applies Bayesian conditioning to the plausibility function and one that performs a sensitivity analysis on a conditional probability. We recall that while the first one is dedicated to revising a belief function, the other one is tailored to a prediction problem when the belief function is a statistical model. We question the use of Dempster conditioning for prediction tasks in Smets generalized Bayes theorem approach to the modeling of statistical evidence and propose a modified version of it, that is more informative than the other conditioning rule.

1 Introduction

Probabilistic conditioning is used both for *prediction* from observations and *revision* of uncertain information. When dealing with prediction, we have at our disposal a model of the world under the form of a probability distribution P issued for instance from a representative set of statistical data (e.g., medical knowledge). Moreover we get some new observations C on the current state of the world (e.g., test results for a patient). Then, one tries to predict some property A of the current world with its associated degree of belief (e.g. predict the disease of the patient). The conditional probability $P(A \mid C)$ (the frequency of observation of A in context C) is used for estimating the degree of belief that the current world satisfies A.

The revision scenario is different: given a probability distribution P (which is often a subjective probability), one learns that an event C occurred, which makes its subjective probability equal to 1 (and not to $P(C) < 1$ as it was supposed before).

Didier Dubois: IRIT, CNRS and Université de Toulouse, France, dubois@irit.fr. ·
Thierry Denœux: HEUDIASYC, CNRS and Université de Technologie de Compiègne, France, thierry.denoeux@hds.utc.fr.

The problem is to determine the new subjective probability measure P', such that $P'(C) = 1$, which is the closest to P in some sense, in order to comply with a minimal change principle. Then, it can be shown that if we use an appropriate relative information measure, it follows that $P'(A) = P(A \mid C), \forall A$ again [3]. However, in the prediction problem, generic knowledge remains unaffected by singular evidence, which is handled apart : predictions can be computed beforehand on the basis of the statistical probability, for each possible set of observations. Note that we do not consider the question of learning a statistical model.

For belief functions, not so much has been done from a statistical point of view, because the main references (Shafer's book [8], and most papers by Smets) present the theory of evidence as an approach to the merging of unreliable testimonies, and consider the mass function at work as a form of subjective probability. Shafer's book contains a single chapter on statistical evidence, then severely criticized, idualincluding by Shafer himself [9] (but more recently rehabilitated [2]). So, the mainstream literature on belief functions is a theory of handling singular uncertain evidence, and not so much an extension of Bayesian statistical prediction. Interestingly, Dempster's pioneering works on upper and lower probabilities are motivated by statistical reasoning and the rehabilitation of ideas of Fisher. In Dempster's setting, a probability space is given that corresponds to the usual setting of random observations for statisticians, along with a random set. Dempster [1] proposes a merging rule and two conditioning rules. However, only one of them was retained in the theory of belief functions. It was originally motivated by the presence of several independent sample spaces, not by a prediction problem. While Shafer is interested by merging independent unreliable isolated testimonies, Dempster considers the problem of merging independent bodies of statistical information. What is crucial is to notice that the merged items are of the same nature, whether singular or generic. As what is known as "Dempster conditioning" is a special case of Dempster rule of combination, this conditioning, widely used in evidence theory, can be viewed as a revision process, understood as a prioritized merging of a sure piece of information with an uncertain one. However the scenario of prediction, which involves a model reflecting a population and a piece of evidence pertaining to a single situation, features of which are to be predicted, totally escapes this revision scheme. In this paper, we distinguish between revision and prediction conditionings. The latter is applied to predicting the class θ of an instance x based on a statistical model $P(\cdot|\theta)$.

- We claim that in general, prediction cannot be achieved using Dempster conditioning. The most cautious prediction method is based on sensitivity analysis on conditional probabilities.
- Using Dempster conditioning in prediction relies on a bold application of the maximum likelihood principle [6]. The two first points were previously discussed by Dubois Prade and Smets [4].
- As a consequence, the approach to statistical prediction proposed by Smets as the Generalized Bayes Theorem (GBT) is questionable, and the prediction conditioning is inefficient. We propose a trade-off approach assuming some information on the training populations.

2 Conditioning for belief functions and imprecise probabilities

In standard probability theory, even if their solution is given by the same conditioning rule, the problem of revising a probability function is different from the one of predicting on the basis of generic statistical information. In the following we examine these two problems in the setting of belief functions.

Prediction conditioning In the case where the generic knowledge of the agent is represented by imprecise probabilities, Bayesian prediction is generalized by performing a sensitivity analysis on the conditional probability. It represents all predictions that could have been made, had the probabilistic model been precise. Let $\mathscr{P}$ be a family of probability measures on S. For each proposition A, a lower bound $P_*(A)$ and an upper bound $P^*(A)$ of the probability degree of A are known. In presence of singular observations summarized under the form of a context C, the belief of an agent in a proposition A is represented by the interval $[P_*(A \mid C), P^*(A \mid C)]$ (already proposed in [1]) where:

$$P_*(A \mid C) = \inf\{P(A \mid C) : P(C) > 0, P \in \mathscr{P}\}; \tag{1}$$

$P^*(A \mid C)$ is obtained by replacing inf by sup. It may happen that the interval $[P_*(A \mid C), P^*(A \mid C)]$ is larger than $[P_*(A), P^*(A)]$, which corresponds to a loss of information in specific contexts. This property reflects the idea that the more singular information is available about a situation, the less informative is the application of generic information to it (since the number of statistical data that fit this situation may become very small). We see that this form of conditioning does not correspond at all to the idea of enriching a statistical model, it is only a matter of querying it.

Belief and plausibility functions in the sense of Shafer [8] are mathematically speaking important particular cases of lower and upper probabilities, although these functions were independently introduced in Shafer's book without any reference to the idea of imprecise probability. Information is supposed to be represented by the assignment of non-negative weights $m(E)$ to subsets E of S. In a generic knowledge representation perspective, $m(E)$ is, for instance, the proportion of imprecise results of the form $x \in E$, in a statistical experiment on a random variable x (in Dempster work, it stems from a known random variable relating observations and parameters). It is clear that in that kind of situation, there exists a real probabilistic model underlying the belief function representation. It contrasts with belief functions on unique events, where there is no underlying subjective probability.

In the frequentist framework, prediction in context C requires evaluating the proportion of the population lying in C, taken as the new frame, from the information on the mass function m. Three cases should be considered for a focal set E:

- $E \subseteq C$: In this case, the frequency $m(E)$ remains assigned to E.
- $E \cap C = \emptyset$: In this case, E no longer matters and $m(E)$ is eliminated.
- $E \cap C \neq \emptyset$ and $E \cap \overline{C} \neq \emptyset$: In this case, there is a proportion $\alpha_E \cdot m(E)$ of the population that satisfies $E \cap C$ and the rest, i.e., $(1-\alpha_E) \cdot m(E)$, satisfies $E \cap \overline{C}$. But these proportions may be unknown.

It is clear that $\alpha_E = 1$ and $\alpha_E = 0$ in the first and second above cases respectively. The third case corresponds to incomplete observations E that neither confirm nor disconfirm C. Suppose that all values $\alpha = \{\alpha_E : E \subseteq S\}$ were known. Then, we can build a mass function obtained as

$$m_\alpha^C(B) = \sum_{E:B=C\cap E} \alpha_E m(E). \tag{2}$$

Note that a renormalisation of this mass function is necessary, in general, as soon as $Pl_\alpha^C(C) < 1$, letting $m_\alpha(\cdot \mid C) = \frac{m_\alpha^C(\cdot)}{Pl_\alpha^C(C)}$. If one denotes by $Bel_\alpha(A \mid C)$ and $Pl_\alpha(A \mid C)$ the corresponding conditional belief and plausibility functions, based on the allocation vector α, we can define conservative belief and plausibility degrees conditional to C by considering all possible weight vectors α. One still obtains belief and plausibility functions (Jaffray [7]). Moreover the following result holds:

$$Bel(A \mid C) = \inf_\alpha Bel_\alpha(A \mid C) = P_*(A \mid C) = \frac{Bel(A\cap C)}{Bel(A\cap C)+Pl(\overline{A}\cap C)} \tag{3}$$

$$Pl(A \mid C) = \sup_\alpha Pl_\alpha(A \mid C) = P^*(A \mid C) = \frac{Pl(A\cap C)}{Pl(A\cap C)+Bel(\overline{A}\cap C)} \tag{4}$$

It is easy to see that $Pl(A \mid C) = 1 - Bel(\overline{A} \mid C)$, and that the closed form formulas generalize probabilistic conditioning. Note that if $Bel(C) = 0$ and $Pl(C) = 1$ (complete ignorance regarding C) then all the focal sets of m overlap C without being contained in C. In this case, $Bel(A \mid C) = 0$ and $Pl(A \mid C) = 1, \forall A \neq \emptyset, A \subset C$: one loses all information in context C.

Example : Ellsberg urn Consider a bag of balls containing $1/3$ red balls, the rest being black or white. So $S = \{w, b, r\}$ and the corresponding frequentist mass function is $m(r) = 1/3$, $m(wb) = 2/3$ (we omit the brackets to denote sets). The prediction problem consists in guessing the colour of a ball x picked at random in the urn. Before knowing anything about x, $Bel(r) = Pl(r) = 1/3; Bel(w) = 0, Pl(w) = 2/3$. Suppose one hears that x is *not black* ($C = \bar{b}$). Applying the prediction conditioning yields $Bel(r|\overline{b}) = \frac{Bel(r)}{Bel(r)+Pl(w)} = 1/3$, $Pl(r|\overline{b}) = \frac{Pl(r)}{Pl(r)+Bel(w)} = 1$, $Bel(w|\overline{b}) = \frac{Bel(w)}{Bel(w)+Pl(r)} = 0$, $Pl(w|\overline{b}) = \frac{Pl(w)}{Pl(w)+Bel(r)} = 2/3$. So the piece of information "the ball is not black" does not alter our beliefs about x being white or not. One may indeed argue it should not, as, hearing x is not black, nothing forbids the urn to contain no white ball, nor no black ball. But the plausibility of the ball being red strongly increases. This is a loss of information on the probability of the ball being red or not.

Revision conditioning The other conditioning, called 'Dempster conditioning' systematically assumes $\alpha_E = 1$ as soon as $E \cap C \neq \emptyset$ in the above mass transfer process. It transfers the full mass of each focal set E to $E \cap C \neq \emptyset$ (followed by a renormalisation). This means that we interpret the new information C as modifying the initial belief function in such a way that $Pl(\overline{C}) = 0$: situations where C is false are considered as impossible. If one denotes by $Pl(A \mid\mid C)$ the plausibility function after revision, we have:

$$Pl(A \mid\mid C) = \frac{Pl(A \cap C)}{Pl(C)} \tag{5}$$

The conditional belief is then obtained by duality as $Bel(A \mid\mid C) = 1 - Pl(\overline{A} \mid\mid C)$. Note that with this conditioning, the size of focal sets diminishes, thus information becomes more precise, and the intervals $[Bel, Pl]$ may become tighter than those obtained by prediction conditioning. Dempster conditioning thus corresponds to a process where information is enriched, which contrasts with prediction conditioning. If $Bel(C) = 0$ and $Pl(C) = 1$ (complete ignorance about C), Dempster conditioning on C will often significantly increases the precision of resulting beliefs.

In the more general framework of imprecise probabilities, the revision by a piece of information C consists in adding the extra constraint $P(C) = Pl(C)$ to the family $\mathscr{P} = \{P \geq Bel\}$. It has been shown by Gilboa and Schmeidler [6] that:

$$P_*(A \mid\mid C) = \inf\{P(A \mid C), P(C) = Pl(C), P \geq Bel\}; \tag{6}$$

$$P^*(A \mid\mid C) = \sup\{P(A \mid C), P(C) = Pl(C), P \geq Bel\}. \tag{7}$$

They indicate that Dempster conditioning comes down to applying the maximal likelihood principle in the imprecise probability setting.

In the view of Shafer and Smets, this type of conditioning is little related with the previous prediction problem, since, in their setting, the mass function m does not represent generic knowledge, but rather uncertain information collected about a particular situation (non fully reliable testimonies, more or less safe clues). It is a form of reasoning under uncertainty where generic knowledge is not taken into account, but where all the pieces of information are singular (as in the Peter Paul and Mary example [4]).

Example: Ellsberg again Hearing that there is no black ball in the urn is a piece of generic information of the same nature as the prior knowledge about the urn. There are then two independent sample spaces as assumed by Dempster [1] in his pioneering paper (one with the possibility of back balls, one without it). We then revise the content of the urn, which in turn impacts a change of belief about the colour of the picked ball x. We then legitimately conclude that since the urn does not contain black balls, the probability of x being white is $2/3$. It may look questionable to apply this conditioning rule to the problem of predicting the colour of a ball x drawn from the urn, based on the fact that it is not black. Using the maximum likelihood interpretation of Dempster conditioning, doing so comes down to assuming that since the ball is known not to be black, we restrict to the probabilistic models P making this event maximally likely (the ones such that $P(\bar{b}) = Pl(\bar{b})$). Then, $P(\bar{b})$ is viewed as the likelihood of the probabilistic model $P \geq Bel$ if the ball is not black.

There is one case when it is easy to show that the two forms of conditionings coincide:

Proposition 1. *If the conditioning event C is such that for any focal set E of m it either contains it or is disjoint from it, then* $\forall A \subseteq S, Pl(A \mid\mid C) = Pl(A \mid C)$.

3 Application to Smets Generalized Bayes Theorem

Despite the warning of Shafer regarding the fact that his theory of evidence deals with unique uncertain events, it has been applied to statistical prediction problems, and the appropriateness of Dempster conditioning for such a task is most of the time not even questioned. A typical example is the Generalized Bayes Theorem of Smets [10], but it is true as well for various other similar approaches to the estimation of parameters based on a belief function model (see [2] for more bibliography).

In the simplest setting, the parametric inference problem is stated as follows : Given a **finite** parameter space Θ and a set of parametric belief functions $Bel_X(\cdot|\theta)$, $\theta \in \Theta$, and some observation $x \in X$, compute $Bel_\Theta(\cdot|x)$.

The most usual situation is when a finite number of probabilistic likelihood functions $\{P(\cdot|\theta), \theta \in \Theta\}$, are available, each one coming from a different population representing a class θ. The GBT procedure specializes as follows:

1. **Conditional embedding** : Each likelihood function $P(\cdot|\theta)$ is modelled by a belief function Bel^θ on $X \times \Theta$ (ballooning): the associated mass function is defined by $m^\theta(\bar{\theta} \cup x) = P(x|\theta), x \in X$; Bel^θ on $X \times \Theta$ has a vacuous marginal on Θ and yields $P(\cdot|\theta)$ back when conditioned on θ.
2. **Conjunctive merging** of the belief functions $Bel^\theta, \theta \in \Theta$ on $X \times \Theta$. Consider a function $\phi : \Theta \to X$; we must assign mass $\prod_{\theta \in \Theta} P(\phi(\theta)|\theta)$ to $\bigcap_{\theta \in \Theta} \bar{\theta} \cup \phi(\theta) = \bigcup_{\theta \in \Theta} \{\theta\} \times \{\phi(\theta)\}$. So, each function ϕ is a focal element and the resulting mass function on $\Theta \times X$ is of the form $m(\phi) = \prod_{\theta \in \Theta} P(\phi(\theta)|\theta)$[1].
3. **Conditioning** of the result on the observation x using Dempster conditioning :

$$Pl_\Theta(\theta||x) = \frac{Pl(\{\theta\} \times \{x\})}{Pl(\Theta \times \{x\})} = \frac{\sum_{\phi:\phi(\theta)=x} m(\phi)}{\sum_{\phi:\phi^{-1}(x)\neq\emptyset} m(\phi)} = \frac{\sum_{\phi:\phi(\theta)=x} \prod_{\tau \in \Theta} P(\phi(\tau)|\tau)}{\sum_{\phi:\phi^{-1}(x)\neq\emptyset} \prod_{\tau \in \Theta} P(\phi(\tau)|\tau)}.$$

The result is a general belief function not fully representable by the $Pl_\Theta(\theta||x)$'s, since the mass function $m(T||x)$ is of the form $\frac{\sum_{\phi:\phi^{-1}(x)=T} m(\phi)}{Pl(\Theta \times \{x\})}$.

Step 2 comes down to applying the disjunctive combination rule to the conditional probabilities $P(\cdot|\theta)$: $Bel(A \times T) = \prod_{\theta \in T} P(A|\theta), \forall A \subseteq X$. For $T \subseteq \Theta, Pl_X(x|T)$ is a function of elementary likelihoods $P(x|\theta), \theta \in T$. The merging rule in step 2 assumes that the likelihood functions $P(\cdot|\theta)$, $\theta \in \Theta$ have been inferred from distinct sets of empirical data obtained from independent sources. Hence, each value θ corresponds to a specific class of objects, and is not a continuous parameter.

However, x is a single observation while m is a statistical model on $X \times \Theta$. Let us then compute the prediction conditioning $Pl_\Theta(\theta|x)$, the plausibility that θ is the class of x. Since a focal element is in the form of a mapping $\phi : \Theta \to X$, and observation x is modelled by $\{x\} \times \Theta$ three situations can be met for a focal ϕ:

1. $\nexists \theta \in \Theta, \phi(\theta) = x$, which means $\phi^{-1}(x) = \emptyset$; ϕ can be eliminated.

[1] Unions are intersections over different spaces consider cylindrical extensions of elements. In the more general case of conditional belief functions $Bel_X(\cdot|\theta)$, focal elements on $\Theta \times X$ are multimappings $\Gamma : \Theta \to 2^X$.

2. $\phi^{-1}(x) \neq \emptyset$ and $\neq \Theta$, then the line $\{x\} \times \Theta$ only overlaps the graph of ϕ.
3. $\phi(\theta) = x, \forall \theta \in \Theta$ then $\phi = \{x\} \times \Theta$ supports x but it gives no clue on θ.

When observing x and considering a focal ϕ consistent with it (case 2), it is not clear what part of $m(\phi)$ should be allocated to $\phi^{-1}(x)$ and what part should be allocated to its complement in Θ. Applying the prediction conditioning would then yield an empty prediction since $\forall T \subset \Theta, Bel(T \times \{x\}) = Bel(\overline{T} \times \{x\}) = 0$, because $\phi \subseteq T \times \{x\}$ never holds ($\forall \theta \in \Theta, \phi(\theta) \neq \emptyset$). However, it is possible to propose an alternative prediction rule that is less bold than Dempster conditioning and more useful than the plain prediction conditioning.

The prior probabilities $P(\theta)$'s are unknown (for Edwards [5], they are even meaningless). It is then tempting to replace $P(\theta)$ by the number of observations $n(\theta)$ available for class θ. This number does not necessarily correspond to a prior probability (as the actual probability $P(\theta)$, if any, is different from the number of cases actually at hand). Yet, $n(\theta)$ if available reflects the reliability of the information regarding the likelihood function $P(\cdot|\theta)$.

Since $m(\phi)$ accounts for all $m_X(\phi(\theta)|\theta)$, one may consider, when observing x, sharing $m(\phi)$ between all θ such that $\phi(\theta) = x$, and those such that $\phi(\theta) \neq x$, according to $n(\theta)$. Then consider the portion $\alpha_\phi(x) = \frac{\sum_{\theta \in \phi^{-1}(x)} n(\theta)}{\sum_{\theta \in \Theta} n(\theta)}$ of the available training data that pertains to observing x. It suggests the following modified conditional belief function for prediction:

$$\forall T \subseteq \Theta, m_\alpha(T|x) = \frac{\sum_{\phi:\phi^{-1}(x)=T} m(\phi)\alpha_\phi(x)}{\sum_{\phi:\phi^{-1}(x)\neq\emptyset} m(\phi)\alpha_\phi(x)}. \quad (8)$$

Example: The simplest example of the problem (actually studied by Shafer [9]) uses a space $\mathscr{S} = \{x, \bar{x}\} \times \{\theta, \bar{\theta}\}$. The available knowledge consists in the two likelihood values $a = P(x|\theta) > b = P(x|\bar{\theta})$. So there is a majority of x's in class θ and a majority of $\bar{x}$'s in class $\bar{\theta}$. Suppose that the likelihood functions are based on independent populations and that the number of available samples of class θ is much greater than those for class $\bar{\theta}$ (say 1000 times more). There are only 4 ϕ functions with their mass assignments on $\Theta \times X$ shown in the table below.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	
θ	x	x	$\bar{x}$	$\bar{x}$	$n(\theta) = 3000$
$\overline{\theta}$	x	$\bar{x}$	x	$\bar{x}$	$n(\overline{\theta}) = 3$
$P(\cdot|\theta)$	$a = 2/3$	$a = 2/3$	$1-a = 1/3$	$1-a = 1/3$	
$P(\cdot|\overline{\theta})$	$b = 1/3$	$1-b = 2/3$	$b = 1/3$	$1-b = 2/3$	
$m(\phi)$	$ab = 2/9$	$a(1-b) = 4/9$	$(1-a)b = 1/9$	$(1-a)(1-b) = 2/9$	
$\alpha_\phi(x)$	1	1000/1001	1/1001	0	
$\alpha_\phi(\bar{x})$	0	1/1001	1000/1001	1	

The following results are obtained if x is observed using Dempster maximal likelihood conditioning:

$$Pl_\Theta(\theta||x) = \frac{a}{a+b-ab} = \frac{6}{7}; Pl_\Theta(\bar{\theta}||x) = \frac{b}{a+b-ab} = \frac{3}{7}. \quad (9)$$

Note that if $\bar{x}$ is observed, the figures are exchanged. However, this symmetry is surprising: since the data set of class $\overline{\theta}$ is very poor, the observation of $\bar{x}$ should suggest class $\overline{\theta}$ to a lesser extent than the one to which observing x should suggest class θ. The last two lines of the table above show the proportions of the overall available population of examples concerned when observing x and $\bar{x}$, given any focal element ϕ. It comes down to assuming that if we observe x, the portion of weight of ϕ_3 to be transferred to $\bar{\theta} \wedge x$ should be much less than the portion to be transferred to $\theta \wedge x$. Conversely if we observe $\bar{x}$, the portion of weight of ϕ_2 to be transferred to $\bar{\theta} \wedge \bar{x}$ should be much less than the portion to be transferred to $\theta \wedge \bar{x}$. So, if x is observed the modified conditional mass $m(\phi_3|\overline{x})$ is reduced to $1/9009$ and becomes negligible: $Pl_\alpha(\theta|x) \simeq 1; Pl_\alpha(\bar{\theta}|x) \simeq 1/3$. If $\overline{x}$ is observed, the modified conditional mass $m(\phi_2|\overline{x})$ is reduced to $4/9009$ so $Pl_\alpha(\theta|\overline{x}) \simeq 1; Pl_\alpha(\bar{\theta}|\overline{x}) \simeq 2/3$. It makes it clear that, as expected, we become more confident about class θ when observing x than about class $\overline{\theta}$ when observing $\bar{x}$. We even still believe in class θ in the latter case, due to the overwhelming number of θ examples. Note that having many more examples of class θ than of class $\overline{\theta}$ does not mean that class $\overline{\theta}$ is rare, but only that we could not have access to many examples of it. So we should not confuse the size of available samples with prior probabilities of classes. The above discussion also lays bare that the GBT approach presupposes not only independent populations of samples for each class, but also that such populations are (approximately) equal.

To conclude, we suggest how to improve the GBT so as to make a trade-off between prediction and revision conditioning. The case of prediction based on several pieces of observations can be addressed by merging the information coming from each observation. Other techniques for modelling statistical evidence [2] should be studied in the light of the above discussion as well.

References

1. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat., 38:325-339 (1967)
2. D. Dubois, T. Denoeux. Statistical inference with belief functions and possibility measures: a discussion of basic assumptions. SMPS 2010, Springer, Advances in Intelligent and Soft Computing 77, p. 217-225 (2010).
3. Z. Domotor. Probability kinematics - conditional and entropy principles. Synthese, 63:74-115, (1985).
4. D. Dubois, H. Prade, P. Smets: Representing partial ignorance. IEEE Trans. on Systems, Man and Cybernetics, 26(3), 1996, 361-377.
5. W. F. Edwards, Likelihood, Cambridge University Press, Cambridge, U.K. (1972)
6. I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. In Y. Moses, ed., Proc. of the 4th Conf. Theor. Aspects of Reasoning About Knowledge (TARK'92), 143-162, (1992).
7. J.Y. Jaffray. Bayesian updating and belief functions. IEEE Trans. on Systems, Man and Cybernetics, 22:1144-1152, 1992.
8. G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, (1976).
9. G Shafer. Belief Functions and Parametric Models, Journal of the Royal Statistical Society. Series B, 44: 322-352 (1982)
10. P. Smets. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approx. Reasoning 9(1): 1-35 (1993)