Novelty detection in the belief functions framework

Astride Aregui
Suez Environnement and
Université de Technologie de Compiègne
UMR CNRS 6599 Heudiasyc, BP 20529
F-60205 Compiègne Cedex France
astride.aregui@hds.utc.fr

Thierry Denoeux
Université de Technologie de Compiègne
UMR CNRS 6599 Heudiasyc, BP 20529
60205 Compiègne Cedex France
thierry.denoeux@hds.utc.fr

Abstract

The problem of testing whether an observation may be deemed to correspond to a given model is a difficult issue. Variants of the problem have been widely studied in Statistics and Pattern Recognition. We build a solution in the belief function framework and demonstrate its advantages over other approaches in situations where the available information is particularly scarce.

Keywords: Belief functions, novelty detection, outliers detection, one-class classification, hypothesis test.

1 Introduction

In industrial process plants, recent regulation has led to the record of numerous variables at many time points with the purpose of providing information for the *a posteriori* analysis of possible failures. Such data constitute a very rich source of information that may also be exploited on-line to detect or prevent faults in the system. Unfortunately, it is often the case that the normal state of the system is well represented by the available data while the other states are not. As a consequence, it is seldom possible to study all possible states of the system. However, it is most of the time sufficient to refer to a particular state (a model representative of "normality" for instance) and to check continuously whether the data are departing from it.

This problem consists in assessing to what extent an observation may be deemed to correspond to a given model. The issue is cross-disciplinary: variants of it have been studied under different names, such as novelty detection, one-class classification, outliers detection [5], pure significance tests [1], [12], single hypothesis tests [4] ...

Let T be a statistic varying over $\mathcal{T}$, representative of the state of a system at a given time. Let $\Omega = \{\omega_0, \omega_1\}$ be the set of possible states of the system. The problem under consideration is the assessment of the hypothesis that a system is in class ω_0 when the only available information about the system concerns the distribution of statistic T conditioned on ω_0.

In this paper, our objective is to sketch the outlines of a solution to this problem, in the belief function framework.

The reader will first be reminded of the theory of belief functions, both in the discrete and continuous cases. In a second part, the Generalized Bayes Theorem will be introduced, and the associated solution will be described. In Part III, we will show how to alter the reasoning process in order to take additional information into account. Part IV provides examples both for the discrete and continuous cases and Part V concludes the paper.

2 The Transferable Belief Model

The approach to belief functions used throughout this paper is Smets's Transferable Belief Model (TBM) [11, 8].

2.1 Discrete Case

2.1.1 The Foundations

Given some evidential corpus EC, the knowledge held by a given agent at a given time over the actual value of variable T can be modeled by a so called *belief structure*. The *basic belief assignment* (bba) $m^{\mathcal{T}}[EC]$ is a function that associates, to each subset S of $\mathcal{T}$, the part of the agent's belief allocated to the hypothesis that T takes some value in S [3, 11, 8]. Bba $m^{\mathcal{T}}[EC]$, denoted m where there is no ambiguity, is a mapping from $2^{\mathcal{T}}$ to $[0,1]$ and the following property necessarily holds:

$$\sum_{S\subseteq\mathcal{T}} m^{\mathcal{T}}[EC](S) = 1. \quad (1)$$

Equivalent representations of m include:

- the *belief* function, representing the amount of belief in S that is entirely justified by the evidential corpus EC:

$$bel(S) = \sum_{\emptyset\neq A\subseteq S} m(A), \quad \forall S\subseteq\mathcal{T} \quad (2)$$

- and the *plausibility* function, corresponding to the amount of belief that is not in contradiction with S given EC:

$$pl = \sum_{A\cap S\neq\emptyset} m(A), \quad \forall S\subseteq\mathcal{T} \quad (3)$$

The absence of knowledge is easily represented in the TBM framework by the so called *vacuous belief function* defined by : $m(\mathcal{T}) = 1$, and equivalently $pl(S) = 1, \forall S\subseteq\mathcal{T},\ S\neq\emptyset$.

2.1.2 Combination of Information

Two distinct [10] pieces of evidence m_1 and m_2 given by two different sources may be combined according to the conjunctive combination rule :

$$\begin{aligned} m_{12}(S) &= \left(m_1 \textcircled{\scriptsize$\cap$} m_2\right)(S) \\ &= \sum_{A\cap B=S} m_1(A)m_2(B), \forall S\subseteq\mathcal{T}. \end{aligned} \quad (4)$$

Note that a necessary condition for using this rule is the distinctness of the two pieces of evidence m_1 and m_2 [7, 10].

2.1.3 Marginalization and Vacuous Extension

From now on, we will work on the product space $\mathcal{T}\times\Omega$. T is a random variable varying over $\mathcal{T}$, representative of the state of a system at a given time. $\Omega = \{\omega_0,\ldots,\omega_K\}$ is a finite set describing all possible states of the system. The ω_i are termed *classes*, and they are mutually exclusive.

Let $m^{\mathcal{T}\times\Omega}$ denote a bba defined on the Cartesian product $\mathcal{T}\times\Omega$ of the two variables T and ω. The *marginal bba* $m^{\mathcal{T}\times\Omega\downarrow\mathcal{T}}$ on T is defined, for all $S\subseteq\mathcal{T}$ as:

$$m^{\mathcal{T}\times\Omega\downarrow\mathcal{T}}(S) = \sum_{\{A\subseteq(\mathcal{T}\times\Omega)|\mathrm{Proj}(A\downarrow\mathcal{T})=S\}} m^{\mathcal{T}\times\Omega}(A), \quad (5)$$

where $\mathrm{Proj}(A\downarrow\mathcal{T})$ denotes the projection of A onto $\mathcal{T}$:

$$\mathrm{Proj}(A\downarrow\mathcal{T}) = \{t\in\mathcal{T}|\exists\omega\in\Omega,\ (t,\omega)\in A)\}. \quad (6)$$

The inverse operation is the *vacuous extension* [6]. Let m^{Ω} be a bba on Ω. Its vacuous extension on $\mathcal{T}\times\Omega$ is defined as:

$$m^{\Omega\uparrow\mathcal{T}\times\Omega}(A) = \begin{cases} m^{\Omega}(B) \\ if\ A = B\times\mathcal{T} \text{for some } B\subseteq\Omega \\ 0 \text{ otherwise.} \end{cases} \quad (7)$$

2.1.4 Conditioning

When a given hypothesis $h\subseteq\Omega$ is ascertained, the beliefs are altered to reflect the new state of knowledge. The conditioning operation consists in combining masses conjunctively with a categorical[1] bba supporting hypothesis h. Hence, the mass of belief allocated to $S\subseteq\mathcal{T}$ knowing that hypothesis $h\subseteq\Omega$ holds, i.e. $m_h^{\Omega}(h) = 1$, is:

$$m^{\mathcal{T}}[h] = \left(m^{\mathcal{T}\times\Omega} \textcircled{\scriptsize$\cap$} m_h^{\Omega\uparrow\mathcal{T}\times\Omega}\right)^{\downarrow\mathcal{T}} \quad (8)$$

Now let $m^{\mathcal{T}}[h]$ be the bba on $\mathcal{T}$ conditioned with respect to $h\subseteq\Omega$. Assume we now learn h finally does not necessarily hold and all previous states of knowledge have been lost. Masses associated with any non-empty set S of $\mathcal{T}$ are then transferred onto $(S\times h)\cup(\mathcal{T}\times(\Omega\setminus h))$. The *balooning*

[1] A *categorical bba* m is a bba that gives total support to a single hypothesis h i.e. $m(h) = 1$.

extension process [7], opposite of the conditioning operation, thus yields to:

$$m^{\mathcal{T}}[h]^{\Uparrow(\mathcal{T}\times\Omega)}(A) = \begin{cases} m^{\mathcal{T}}[h](S) \\ \quad if\ A = (S \times h) \cup (\mathcal{T} \times (\Omega \setminus h)), \\ 0\ otherwise. \end{cases} \quad (9)$$

2.2 Continuous Case

The extension of the above described tools to the continuous case is fairly straightforward [9].

Let us consider a non empty interval in $\mathbb{R}$ denoted $[a,b]$, $a < b$, and let $I_{[a,b]}$ be the set of closed intervals in [a,b]. A convenient representation of $I_{[a,b]}$ is the triangle shown in Figure 1a. In effect, each point of coordinates $(\underline{t}, \bar{t})$ in this triangle corresponds to a unique subinterval $t = [\underline{t}; \bar{t}] \in I_{[a,b]}$ and vice versa.

Any function $m^{I_{[a,b]}} : I_{[a,b]} \to [0, \infty[$ verifying:

$$\int_a^b \int_x^b m^{I_{[a,b]}}(x,y)dydx \le 1, \quad (10)$$

is a *basic belief density (bbd)* on $I_{[a,b]}$. The term bbd is used instead of bba (or mass) in order to account for the fact that m is now continuous. As a convention, the one's complement of integral (10) is allocated to the empty set.

The corresponding belief function *bel* is the sum of all masses assigned to intervals included in $[\underline{t}, \bar{t}]$ (Figure 1a).

$$bel^{I_{[a,b]}}([\underline{t}; \bar{t}]) = \int_{\underline{t}}^{\bar{t}} \int_x^{\bar{t}} m^{I_{[a,b]}}(x,y)dydx \quad (11)$$

Similarly, the associated plausibility function *pl* is the sum of all masses allocated to intervals whose intersection with $[\underline{t}; \bar{t}]$ is non-empty (Figure 1b). Hence,

$$pl^{I_{[a,b]}}([\underline{t}; \bar{t}]) = \int_a^{\bar{t}} \int_{\underline{t}\vee x}^b m^{I_{[a,b]}}(x,y)dydx \quad (12)$$

It can be shown that the restriction to closed intervals may be relaxed, the whole real line $\mathbb{R}$ being used instead.

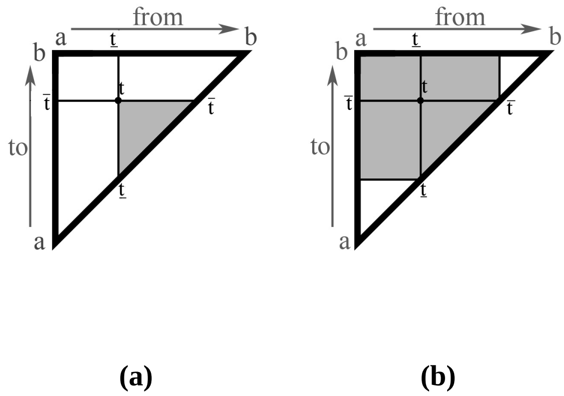

Figure 1: The belief and plausibility functions are defined as integrals of the bbd defined on the shaded area of triangle (a) and (b) respectively.

3 The Generalized Bayes Theorem

The Generalized Bayes' Theorem (GBT) was introduced by Smets [7]. It generalizes Bayes's theorem in that, whenever the belief functions are Bayesian, and we also have a Bayesian a priori on the classes, the two theorems are exactly equivalent. However, the power of the GBT lies in the fact that it does not require any prior knowledge on Ω (for instance, no prior class probabilities).

3.1 Definition of the GBT

Let us suppose we know all the conditional bbas $m^{\mathcal{T}}[\omega_k]$, $k = 0, \ldots, n$, we have no prior knowledge on Ω, and we observe $t_* \subseteq \mathcal{T}$. From that, we would like to derive our belief in the fact that the system is in a particular state ω_i, knowing the value of statistic T. In other words, we seek $m^{\Omega}[t_*]$. The GBT allows us to find the answer in three steps.

We shall first calculate the ballooning extension of each of the functions $m^{\mathcal{T}}[\omega_k]$, that is to say, *"de-condition"* them in order to obtain a belief on $\mathcal{T} \times \Omega$. The obtained bbas $m^{\mathcal{T}}[\omega_k]^{\Uparrow\mathcal{T}\times\Omega}$ are distinct, as the original $m^{\mathcal{T}}[\omega_k]$ were distinct. Hence, the $m^{\mathcal{T}}[\omega_k]^{\Uparrow\mathcal{T}\times\Omega}$ can be combined by applying the conjunctive combination rule: this will be the second step. We now have a global and unconditioned belief function on $\mathcal{T} \times \Omega$. Conditioning with respect to t_* returns the belief function we need, namely $m^{\Omega}[t_*]$. The GBT may then be defined as follows:

$$m^{\Omega}[t_*] = \left(\circledcirc_{k=0}^{K} m^{\mathcal{T}}[\omega_k]^{\Uparrow \mathcal{T}\times\Omega}\right)[t_*]. \qquad (13)$$

The equivalent formulation in terms of plausibility functions is sometimes easier to manipulate:

$$pl^{\Omega}[t_*](A) = 1 - \prod_{\omega_k \in A}(1 - pl^{\mathcal{T}}[\omega_k](t_*)), \quad \forall A \subseteq \Omega \qquad (14)$$

3.2 The problem (P_0) at hand

Now let us consider the case (P_0) where $\Omega = \{\omega_0, \omega_1\}$ and we only know the behaviour of T when ω_0 holds. In the sequel, $pl_0^{\mathcal{T}}$ will stand for $pl[\omega_0]^{\mathcal{T}}$ and $pl_1^{\mathcal{T}}$ for $pl[\omega_1]^{\mathcal{T}}$; $m_0^{\mathcal{T}}$ and $m_1^{\mathcal{T}}$ will (respectively) be the associated bbas.

Contruction of $m_0^{\mathcal{T}}$: This belief function on T under hypothesis ω_0 may have been elicited from an expert, or it may have been built from a set of data collected when we took for certain that ω_0 was true. This belief function may or may not be a probability function.

If the information at hand is a belief function, no further processing is required before applying the GBT. If the available information takes the form of a data set, it is possible to build *the least committed belief function* (LCBF) fulfilling the constraints provided by the available information [2].

Solution: We will derive the solution of problem (P_0) without focusing on how the $m_k^{\mathcal{T}}$ were obtained.

As we know nothing on the behaviour of T when ω_1 holds, the information we have at our disposal in this respect can be modeled with the vacuous belief function:

$$pl_1^{\mathcal{T}}(t_*) = pl^{\Omega}[t_*](\{\omega_1\}) = 1 \qquad (15)$$

On the other hand, we do know $m_0^{\mathcal{T}}$. The plausibility-related form of the GBT [11] thus yields to:

$$\begin{aligned} pl^{\Omega}[t_*](\{\omega_0\}) &= m^{\Omega}[t_*](\{\omega_0\}) + m^{\Omega}[t_*](\Omega) = pl_0^{\mathcal{T}}(t_*) \\ pl^{\Omega}[t_*](\{\omega_1\}) &= m^{\Omega}[t_*](\{\omega_1\}) + m^{\Omega}[t_*](\Omega) = 1 \\ m^{\Omega}[t_*](\{\omega_0\}) &+ m^{\Omega}[t_*](\{\omega_1\}) + m^{\Omega}[t_*](\Omega) = 1. \end{aligned} \qquad (16)$$

Hence, from (15) and (16),

$$m^{\Omega}[t_*](\{\omega_0\}) = 0 \qquad (17a)$$
$$m^{\Omega}[t_*](\{\omega_1\}) = 1 - pl_0^{\mathcal{T}}(t_*) \qquad (17b)$$
$$m^{\Omega}[t_*](\Omega) = pl_0^{\mathcal{T}}(t_*). \qquad (17c)$$

Interpretation: If the value of T is completely plausible assuming ω_0 to be true ($pl_0^{\mathcal{T}}(t_*) = 1$), it is not possible to say whether the system is in state ω_0 or in any other state that yields similar values of T. Thus, no value of T ever supports ω_0 only, leading to (17a). Moreover, the nearer the values of T to that obtained under ω_0, the more plausible is Ω, hence (17c). Finally, the more the value of T differs from that obtained when ω_0 holds, the greater the belief we have in ω_1: from that we get (17b).

3.3 Continuous Case

The same reasoning holds for the continuous case, and equations (17) remain valid. A drawback of this method is that, in the specific case where $pl_0^{\mathcal{T}}$ is continuous and Bayesian, and t_* is a singleton, then $pl(t_*)$ equals zero. Equation (17b) thus becomes:

$$m^{\Omega}[t_*](\{\omega_1\}) = 1 - pl_0^{\mathcal{T}}(t_*) = 1, \qquad (18)$$

and the conclusion is that we always assign full belief to ω_1, without taking the value of t_* into account. There is a paradox there, but we argue that the problem is not in formula (17). In effect, when the belief about T is represented by a probability density function, it does not really make sense to assume that $pl(t) = 0$ for all $t \subseteq \mathcal{T}$. As an alternative, it seems more reasonable to use the plausibility function whose pignistic transform equals p [9].

4 Introduction of a Priori Information

We considered up to now the case were the only available information is related to ω_0. Nevertheless, some sort of a priori information is quite often available about ω_1, though it may be very weak. As any piece of knowledge can be turned into a belief function, no matter how incomplete or scarce it might be, there is no reason not to use it when it is available. Let $\mathcal{T}$ be an ordered set:

$\mathcal{T} = \{t_{(i),\, i\in\{1,\dots,n\}} : t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}\}$. Let us suppose we know how large variable T will be under hypothesis ω_1 in comparison with how large it is under hypothesis ω_0; e.g. we think that T tends to be larger when ω_1 holds than when ω_0 is true. This statement may be turned into the following constraint:

$$pl_1(T \leq t) \leq pl_0(T \leq t), \quad \forall i \in 1, ...n. \tag{19}$$

Note that (19) generalizes the stochastic inequality in that, when both pl_0 and pl_1 are probabilities, it turns into stochastic inequality. Property (19) will thus be termed *cognitive inequality*.

4.1 Discrete Case

The idea is to calculate the least committed belief function $m_1^{\mathcal{T}}$ that satisfies requirement (19) and to combine it with $m_0^{\mathcal{T}}$ for the purpose of calculating $m^{\Omega}[t_*]$.

4.1.1 Construction of $m_1^{\mathcal{T}}$

The least committed bba satisfying (19) corresponds to a maximization of plausibilities subject to constraints (19), and can easily be built from $m_0^{\mathcal{T}}(t)$. For all $i \in \{1, \dots, n\}$, $pl_1^{\mathcal{T}}(t_{(1)}, \dots t_{(i)}) = pl_0^{\mathcal{T}}(t_{(1)}, \dots, t_{(i)})$ implies:

$$m_1^{\mathcal{T}}(t_{(i)}, \dots, t_{(n)}) = pl_0^{\mathcal{T}}(t_{(1)}, \dots, t_{(i)}) - pl_0^{\mathcal{T}}(t_{(1)}, \dots, t_{(i-1)}). \tag{20}$$

The idea is to try and get an equality for relation (19), and to deduce bba m_1 from this. Deriving equation (19) for each i successively leads to the above result.

Note that m_1 is a consonant bba. Additionnally, if pl_0 is Bayesian, pl_1 is its cumulated distribution function.

It is possible to calculate $m_1^{\mathcal{T}}$ directly from $m_0^{\mathcal{T}}$:

$$\begin{aligned}
m_1^{\mathcal{T}}(\mathcal{T}) &= pl_0^{\mathcal{T}}(t_{(1)}) = \sum_{t_{(1)}\in S} m_0^{\mathcal{T}}(S), \\
\forall i > 1, & \\
m_1^{\mathcal{T}}(t_{(i)}, \dots, t_{(n)}) &= pl_0^{\mathcal{T}}(t_{(1)}, \dots, t_{(i)}) - pl_0^{\mathcal{T}}(t_{(1)}, \dots, t_{(i-1)}) \\
&= \sum_{\substack{\exists t_{(j)}\leq t_{(i)} \\ t_{(j)}\in S}} m_0^{\mathcal{T}}(S) - \sum_{\substack{\exists t_{(j)}\leq t_{(i-1)} \\ t_{(j)}\in S}} m_0^{\mathcal{T}}(S) \\
&= \sum_{\substack{t_{(i)}\in S \\ \forall j<i,\ t_{(j)}\notin S}} m_0^{\mathcal{T}}(S) \\
&= \sum_{S\subseteq\{t_{(i+1)},\dots,t_{(n)}\}} m_0^{\mathcal{T}}(\{t_{(i)}\} \cup S) \\
\text{and} & \\
m_1^{\mathcal{T}}(t_{(n)}) &= m_0^{\mathcal{T}}(t_n).
\end{aligned} \tag{21}$$

Thus, $m_1^{\mathcal{T}}$ may be obtained from $m_0^{\mathcal{T}}$ by transferring each mass $m_0^{\mathcal{T}}(S)$ onto $\{\min(S), \dots, t_{(n)}\}$.

4.1.2 Combination with $m_0^{\mathcal{T}}$

If we follow the reasoning of Section 3, we should now apply the GBT to $m_0^{\mathcal{T}}$ and $m_1^{\mathcal{T}}$ in order to obtain $m^{\Omega}[t]$. However, remember that a necessary condition for the application of the GBT is the independence of $m_0^{\mathcal{T}}$ and $m_1^{\mathcal{T}}$. It happens that, as we built $m_1^{\mathcal{T}}$ from $m_0^{\mathcal{T}}$, they are not independent. Consequently, the conjunctive combination rule cannot be applied here. We need to build $m^{\mathcal{T}\times\Omega}$ such that:

$$m^{\mathcal{T}\times\Omega}[\{\omega_0\} \times \mathcal{T}]^{\downarrow\mathcal{T}} = m_0^{\mathcal{T}} \tag{22}$$

$$\text{and } m^{\mathcal{T}\times\Omega}[\{\omega_1\} \times \mathcal{T}]^{\downarrow\mathcal{T}} = m_1^{\mathcal{T}} \tag{23}$$

Let F_1 to F_K be the focal elements of $m_0^{\mathcal{T}}$. To each F_k is associated F'_k, focal element of $m_1^{\mathcal{T}}$, such that: $F'_k = [min_{t_i\in F_k} t_i, \dots, t_{(n)}]$. Thus,

$$m^{\mathcal{T}\times\Omega}\left(F_k \times \{\omega_0\} \cup F'_k \times \{\omega_1\}\right) = m_0^{\mathcal{T}}(F_k). \tag{24}$$

4.1.3 Conditioning with respect to $t_* \subseteq \mathcal{T}$

Note that $|t_*|$ may be greater than 1 and that t_* is not necessarily an interval. The following relations hold:

$$\begin{aligned}
pl^{\Omega}[t_*](\omega_0) &= pl_0^{\mathcal{T}}(t_*) \\
pl^{\Omega}[t_*](\omega_1) &= pl_0^{\mathcal{T}}(t_{(1)}, \dots, \max(t_*)) \\
pl^{\Omega}[t_*](\emptyset) &= 1 - pl_0^{\mathcal{T}}(t_{(1)}, \dots, \max(t_*))
\end{aligned} \tag{25}$$

Hence,

$$\begin{array}{rcl} m^{\Omega}[t_*](\omega_0) & = & 0 \\ m^{\Omega}[t_*](\omega_1) & = & pl_0^{\mathcal{T}}(t_{(1)},\ldots,\max(t_*)) - pl_0^{\mathcal{T}}(t_*) \\ m^{\Omega}[t_*](\Omega) & = & pl_0^{\mathcal{T}}(t_*) \\ m^{\Omega}[t_*](\emptyset) & = & 1 - pl_0^{\mathcal{T}}(t_{(1)},\ldots,\max(t_*)) \end{array} \quad (26)$$

This end result may be easily interpreted:

- When the values of T are similar to those obtained under hypothesis ω_0, nothing can be said about them being from one class or the other, and the belief is thus spread onto Ω.
- When the values of T are smaller than those we get when ω_0 holds, there is an inconsistency with our original information according to which, when data are departing from ω_0, they should tend to be bigger than when ω_0 is true. The corresponding amount of belief is thus allocated to the empty set, reflecting this conflict.
- When T gets bigger than its usual values under ω_0, then our belief turns to ω_1, in agreement with the above piece of information.
- Finally, no value of T ever supports ω_0 only.

4.2 Continuous Case

The solution exposed in the previous paragraph easily extends to the continuous case: Ω remains discrete and still equals $\{\omega_0,\omega_1\}$, $\mathcal{T}=\mathbb{R}$, and $m_0^{\mathcal{T}}$ is a continuous bbd on $\mathbb{R}$.

Our information according to which the values of T tend to be bigger under hypothesis ω_1 than when ω_0 holds imposes that :

$$pl_1^{\mathcal{T}}((-\infty;t]) \le pl_0^{\mathcal{T}}((-\infty;t]),\ \forall t\in\mathbb{R}. \quad (27)$$

From this, we deduce the general expression of $m_1^{\mathcal{T}}$. The equality $pl_1^{\mathcal{T}}((-\infty;t]) = pl_0^{\mathcal{T}}((-\infty;t])$ is actually required for all t in $\mathbb{R}$, with analogy to (21), leading to:

$$m_1^{\mathcal{T}}([t;+\infty)) = \int_t^{+\infty} m_0^{\mathcal{T}}([t;v])dv, \quad (28)$$

Subsequently, masses $m_1^{\mathcal{T}}$ are all allocated to intervals of the form $[u;+\infty)$, with $u\in(-\infty;t]$. As a result, it may be shown that requirement (27) is met:

$$\begin{array}{rcl} pl_1^{\mathcal{T}}((-\infty;t]) & = & \int_{-\infty}^{t} m_1^{\mathcal{T}}([u;+\infty))du \\ & = & \int_{-\infty}^{t}\int_{t}^{\infty} m_0^{\mathcal{T}}([u;v])dudv \\ & = & pl_0^{\mathcal{T}}((-\infty;t]). \end{array} \quad (29)$$

Note that it can also easily be shown that $pl_1^{\mathcal{T}}([a;b]) = pl_0^{\mathcal{T}}((-\infty;b])$. In the probabilistic case, $pl([a;b])$ is the cumulated distribution function of $pl_0^{\mathcal{T}}((-\infty;b])$, as was already demonstrated in the discrete case.

The construction of the final solution, via the form of combination described in §5.1 and conditioning with respect to a subset t_* yields the same result as in the discrete case, with similar interpretation:

$$\begin{array}{rcl} m^{\Omega}[t_*](\omega_0) & = & 0 \\ m^{\Omega}[t_*](\omega_1) & = & pl_0^{\mathcal{T}}(-\infty,\sup(t_*)) - pl_0^{\mathcal{T}}(t_*) \\ m^{\Omega}[t_*](\Omega) & = & pl_0^{\mathcal{T}}(t_*) \\ m^{\Omega}[t_*](\emptyset) & = & 1 - pl_0^{\mathcal{T}}(-\infty,\sup(t_*)) \end{array} \quad (30)$$

where $pl_0^{\mathcal{T}}(-\infty,\ldots,\sup(t_*)) = pl_1(t_*)$.

Note that, if pl_0 is Bayesian we may end up always deciding in favour of ω_1, but the remark of §3.3 still holds. If our information with respect to ω_0 is a probability, then we should use the belief function whose pignistic transform is this probability, and not the belief function whose bba is this probability.

5 Example

Let us consider a plant for which the set of possible states is $\Omega=\{\omega_0,\omega_1\}$. Hypothesis ω_0 represents the fact that the plant is working in secure conditions (also termed safe or nominal mode) and ω_1 represents the fact that it is not, all types of failure taken together. An observed variable T is assumed to carry information regarding the state of the plant. We know the distribution of T when ω_0 is true, but we know nothing about the distribution of T under ω_1. We observe a value of T equal to t_* at instant i. Now, knowing t_*, what is our belief that the plant is in a safe mode ? We will derive the solutions of this problem, with and without a priori knowledge on ω_1, for an example where $m_0^{\mathcal{T}}$ is continuous.

5.1 Construction of $m_0^{\mathcal{T}}$

Let us suppose we dispose of a great number number of data, allowing us to estimate the mean and variance of the distribution with sufficient precision, or we have elicited these first two moments from an expert's opinion. On a normality test, we decide to use a Gaussian model for the distribution. The result of the estimation permits to conclude that the expert's bet on T under hypothesis ω_0 would follow a normal distribution of mean $\mu = 4$ and variance $\sigma^2 = 4$. As mentioned in §3.3 and 4.2, the associated belief function is the belief function whose pignistic transform equals $\mathcal{N}(\mu; \sigma^2)$. The details of the construction of this belief function are given in [9].

5.2 Solution without additional information

Without any prior knowledge on ω_1, the belief function on Ω obtained using the GBT (17) is shown, as a function of t, in Figure 2.

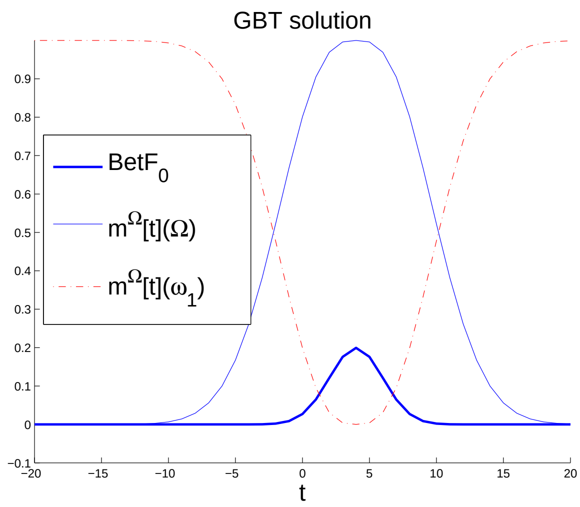

Figure 2: GBT solution in the continuous case, with $BetF_0 \sim \mathcal{N}(4; 4)$

The interpretation is similar to that given in §3.2:

If t is likely to be a nominal value of T given p_0, i.e. if $pl_0^{\mathcal{T}}(t)$ is fairly high, we can say nothing about the plant being safe or not. In effect, the plant might be working in safe mode, but it might also very well be working in a mode leading to values of T similar to those obtained in nominal mode. Thus, the nearer the observed value of T to the likeliest nominal values, the more belief we assign to Ω. This is why the curves representing $pl[t](\Omega)$ (continuous line) and $BetF_0$ (bold continuous line) have the same shape and mode.

Similarly, whatever the value of T may be, it will never strengthen the hypothesis that the plant can only be in safe mode as there may be many modes leading to similar values of T. Hence, function $pl[t](\omega_0)$ is always null (it is not represented in the figure).

On the other hand, if the value of T is nowhere near the values corresponding to the safe mode (it is either greater or smaller), we will tend to think that some failure is occurring. In other words, the more t differs from the nominal values, the greater belief we have in the fact that the plant is malfunctioning (dash-dotted line).

5.3 Solution with a priori knowledge on ω_1

Let us now introduce some more information. The a priori according to which values of T get larger under ω_1 leads to the curves of Figure 3, obtained via equation (30).

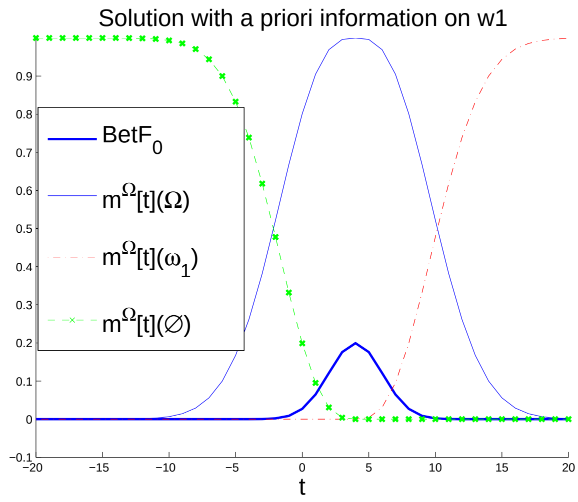

Figure 3: Solution with introduction of a priori information on m_1 in the continuous case, with $BetF_0 \sim \mathcal{N}(4; 4)$.

The first two aforementioned points still hold: no belief is ever allocated to ω_0 alone and if the value of t is likely to be a safe mode value, the most important part of our belief is again, assigned to Ω (continuous line).

The main difference with the case where we do not dispose of any additional information lies in the fact that we now assume that T takes greater values under failure than in safe mode. As a consequence, our belief in failure increases as t becomes significantly larger than safe mode values of T (dash-dotted line), but it does not increase when it becomes smaller.

Conversely, whenever t is smaller than safe values of T tend to be, we face an inconsistency with the information according to which, when the plant is malfunctioning, values of T are larger than they are liable to be in nominal mode. Thus, some belief is assigned to the empty set (dashed crossed line), representing conflict, or contradiction.

Suppose we need to stop the plant whenever our belief in a possible state of failure reaches a certain level. From the above short example, it can be seen that, with as little information as m_0, it is already possible to weight the different costs of the decisions to stop the plant or not for a given value of T. It is also possible to measure the influence of an additional, qualitative piece of information (for instance, the fact that t gets bigger under failure) provided by an expert.

6 Conclusion

We built a solution to the problem of testing an hypothesis versus its exact opposite under the belief function framework. Our solution takes advantage of the facilities offered by this theory to work with partial knowledge without addition of any assumption. It thus allows us to make a decision when very little information is available. However, as mentioned in sections 3.3 and 4.2, this solution sometimes leads to a surprising decision when probabilities are involved. The resolution of this problem is presently under study. Another perspective of this work is the description of the whole process that leads from raw data to the decision stage through the TBM framework.

References

[1] D. Cox and D. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974.

[2] T. Denoeux. Construction of predictive belief functions using a frequentist approach. In *IPMU 2006 (submitted)*.

[3] B. Dubuisson et al. *Approches non probabilistes*, chapter 3, pages 159–162. Hermes, Paris, 2001.

[4] K. Fukunaga. *Statistical Pattern Recognition*. Academic Press, Inc., London, 1990.

[5] B. Ripley. *Pattern Recognition and neural networks*. Cambridge University Press, Cambridge, 1996.

[6] G. Shafer. *A theory of statistical evidence*, volume 2, pages 365–436. 1976.

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

[8] P. Smets. The transferable belief model for quantified belief representation. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer, Doordrecht, The Netherlands, 1998.

[9] P. Smets. Belief functions on real numbers. *Int. J. Approx. Reasoning*, 40:181–223, 2005.

[10] P. Smets and R. Kennes. The concept of distinct evidence. In *Proceedings of the 4th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 789–794, Palma de Mayorca, 1992.

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

[12] D. Sprott. *Statistical Inference in Science*. Springer, New York, 2000.