Constructing Consonant Belief Functions from Sample Data using Confidence Sets of Pignistic Probabilities [1]

Astride Aregui*,† and Thierry Denœux*

* HEUDIASYC, UTC, CNRS

Centre de Recherche de Royallieu

BP 20529, F-60205 Compiègne, France

† CIRSEE, Suez Environnement

38 rue du Pdt Wilson, F-78230 Le Pecq, France

June 4, 2008

[1]This paper is an extended version of [1].

Abstract

A new method is proposed for building a predictive belief function from statistical data in the Transferable Belief Model framework. The starting point of this method is the assumption that, if the probability distribution $\mathbb{P}_X$ of a random variable X is known, then the belief function quantifying our belief regarding a future realization of X should have its pignistic probability distribution equal to $\mathbb{P}_X$. When $\mathbb{P}_X$ is unknown but a random sample of X is available, it is possible to build a set $\mathcal{P}$ of probability distributions containing $\mathbb{P}_X$ with some confidence level. Following the Least Commitment Principle, we then look for a belief function less committed than all belief functions with pignistic probability distribution in $\mathcal{P}$. Our method selects the most committed consonant belief function verifying this property. This general principle is applied to arbitrary discrete distributions as well as exponential and normal distributions. The efficiency of this approach is demonstrated using a simulated multisensor classification problem.

Keywords: Dempster-Shafer theory, Evidence theory, Transferable Belief Model, possibility distribution, confidence region, statistical data.

1 Introduction

The Transferable Belief Model (TBM) [43, 40] is a two-level mental model in which the beliefs held by an agent are represented at the *credal level* by belief functions [34], whereas decision making is based on probability distributions and takes place at the *pignistic level* [42]. The TBM is gaining increasing interest as a formal framework for information fusion [29, 3, 32], pattern recognition [11, 12, 15, 24, 44] and imprecise data analysis [14, 27, 30]. However, it is not always clear how to quantify various uncertainties using belief functions as required in this framework, especially when statistical data are involved. A first approach to this problem was presented in [13] and [2] in the cases of discrete and continuous distributions, respectively. A different solution, more in line with the two-level structure of the TBM, is presented here.

More precisely, the problem considered in this paper can be described as follows. Let X be a random variable (r.v.) on a domain $\mathcal{X}$ with unknown probability distribution $\mathbb{P}_X$. We would like to quantify the beliefs held by an agent about a future realization of X from past independent observations $X_1, \ldots, X_n$ drawn from the same distribution. In [13], it was argued that a belief function $bel(\cdot; X_1, \ldots, X_n)$ solution to this problem should verify two properties:

1. $bel(\cdot; X_1, \ldots, X_n)$ should be *less committed* than $\mathbb{P}_X$ with a given probability, i.e., the event

$$bel(\cdot; X_1, \ldots, X_n) \leq \mathbb{P}_X$$

should be realized for a given proportion of realizations of the random sample, and

2. $bel(\cdot; X_1, \ldots, X_n)$ should converge towards $\mathbb{P}_X$ in probability as the sample size tends to infinity.

Several methods for constructing such belief functions (referred to as *predictive belief functions*) were proposed in [13] in the special case where X is discrete, based on multinomial confidence intervals. This approach was recently extended to the continuous case using confidence bands on the unknown cumulative probability distribution instead of multinomial confidence intervals [2], and a similar approach in the context of Possibility Theory was presented in [28].

In the above approach, the two requirements are derived from Hacking's frequency principle [26, 38], which equates the degree of belief of an event to its probability (long run frequency), when the latter is known. In other words, when $\mathbb{P}_X$ is known, we should have, according to this principle, $bel = \mathbb{P}_X$. When only partial information is available, it is then reasonable to demand that $bel(\cdot; X_1, \ldots, X_n)$ be *weaker* (less informative) than $\mathbb{P}_X$, hence the first requirement. Additionally, as an infinite sample size is equivalent to complete knowledge of the distribution, the predictive belief function should, in the long run, become closer to $\mathbb{P}_X$, hence the second requirement.

The relevance of Hacking's principle, however, can be questioned. For instance, consider the result X of a coin-tossing experiment, with $X \in \{H, T\}$, where H and T stand for "Heads" and "Tails", respectively. If the coin is known to be perfectly balanced, then $\mathbb{P}_X(\{H\}) = \mathbb{P}_X(\{T\}) = 0.5$. If asked about our opinion regarding the result of the next toss, it is natural to assign the same degree of belief to both events H and T for symmetry reasons. In the Bayesian approach, this degree of belief has to be equal to 0.5 because of the additivity property for probability measures. However, this property is not imposed to degrees of belief in the TBM, so that there does not seem to be any compelling reason for setting $bel(\{H\}) = bel(\{H\}) = 0.5$. Yet, if we are forced to bet on the result of this random experiment, it seems reasonable to assign equal odds to the two elementary events. In the TBM, degrees of chance are not equated with degrees of belief: as emphasized above, decision making is assumed to be handled at the *pignistic level*, which is distinguished from the *credal level* at which beliefs are entertained [43, 42]. The pignistic transformation converts each belief function *bel* into a *pignistic* probability distribution $BetP$ that is used for decision making. As a consequence, we may replace Hacking's principle by the weaker requirement that *the pignistic probability of an event be equal to its long run frequency, when the latter is known.* Coming back to the coin example, this requirement leads to the constraint $BetP(\{H\}) = BetP(\{T\}) = 0.5$, which defines a set of admissible belief functions. Among this set, the Least Commitment Principle [37] dictates to choose the least committed one (i.e., the least informative), which is here the vacuous belief function.

In the above example, the probability distribution of X was assumed to be known. In the more realistic situation considered here, we only have partial information about this distribution, in the form of a random sample $X_1, \ldots, X_n$. In that case, it is possible to construct a set $\mathcal{P}$ of probability distributions defined, e.g., by a parametric confidence region. A natural extension of the above line of reasoning is then to require that *bel* be less committed than any belief function with pignistic probability distribution in $\mathcal{P}$. This leads to the definition of a set of admissible belief functions, among which the most committed one can be chosen. This is the principle of the approach presented in this paper.

The rest of this paper is organized as follows. The background on the TBM will first be recalled in Section 2. The proposed approach will be formalized in Section 3. It will then be applied to the case of a discrete r.v. in Section 4, and to continuous parametric models in Section 5. In particular, the exponential and normal distributions will be treated in Section 5.1 and 5.2, respectively. An application to classification with simulation results will then be presented in Section 6, and Section 7 will conclude the paper.

2 Background on the TBM

This section provides a short introduction to the main notions pertaining to the theory of belief functions that will be used throughout the paper, and in particular, its TBM interpretation. We first consider the case of belief functions defined on a finite domain [34], and then address the case of a continuous domain [41].

2.1 Belief Functions on a Finite Domain

Let $\mathcal{X} = \{\xi_1, \ldots, \xi_K\}$ be a finite set, and let X be a variable taking values in $\mathcal{X}$. Given some evidential corpus, the knowledge held by a given agent at a given time over the actual value of variable X can be modeled by a so-called *basic belief assignment* (bba) m defined as a mapping from $2^{\mathcal{X}}$ into $[0, 1]$ such that:

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

Each mass $m(A)$ is interpreted as the part of the agent's belief allocated exactly to the hypothesis that X takes some value in A [34, 43] or, stated differently, as the weight assigned to the assumption of *only* knowing that X takes some value in A. The subsets $A \subseteq \mathcal{X}$ such that $m(A) > 0$ are called the *focal sets* of m. When the focal sets are nested, m is said to be *consonant*. In particular, a *simple* bba has only two focal sets: $\mathcal{X}$ and a strict subset $A \subset \mathcal{X}$.

Equivalent representations of m include the belief, plausibility and commonality functions defined, respectively, as:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \tag{2}$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \tag{3}$$

and

$$q(A) = \sum_{B \supseteq A} m(B), \tag{4}$$

for all $A \subseteq \mathcal{X}$.

When m is consonant, then the plausibility function is a possibility measure: it verifies $pl(A \cup B) = \max(pl(A), pl(B))$ for all $A, B \subseteq \mathcal{X}$. The corresponding possibility distribution is defined by $\text{poss}(x) = pl(\{x\}) = q(\{x\})$ for all $x \in \mathcal{X}$, and the commonality function verifies $q(A \cup B) = \min(q(A), q(B))$ for all $A, B \subseteq \mathcal{X}$. Conversely, any possibility measure Π with possibility distribution $\text{poss}(x) = \Pi(\{x\})$ for all $x \in \mathcal{X}$ is a plausibility function corresponding to a consonant bba m defined as follows [17]. Let $\pi_k = \text{poss}(\xi_k)$, and let us assume that the elements of $\mathcal{X}$ have been arranged in such a way that $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_K$. Then, we have:

$$m(A) = \begin{cases} 1 - \pi_1 & \text{if } A = \emptyset, \\ \pi_k - \pi_{k+1} & \text{if } A = \{\xi_1, \ldots, \xi_k\} \text{ for some } k \in \{1, \ldots, K-1\}, \\ \pi_K & \text{if } A = \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Two bbas m_1 and m_2 induced by distinct items of evidence can be combined using the TBM conjunctive rule (also referred to as the unnormalized Dempster's rule of combination) [34, 40]. The resulting bba $m_{1 \circledcap 2} = m_1 \circledcap m_2$ is defined by

$$m_{1 \circledcap 2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega. \tag{6}$$

This rule is commutative and associative [34]. Let B be a subset of Ω and m_B the bba defined by $m_B(B) = 1$. Given a bba m, its conditioning by B using the unnormalized Dempster's rule of conditioning is defined by $m(\cdot|B) = m \circledcap m_B$. The corresponding plausibility measure is

$$pl(A|B) = pl(A \cap B), \quad \forall A \subseteq \Omega. \tag{7}$$

In the TBM, the *Least Commitment Principle* (LCP) plays a role similar to the principle of maximum entropy in Bayesian Probability Theory. As explained in [37], the LCP states that, given two belief functions compatible with a set of constraints, the most appropriate is the least informative. To make this principle operational, it is necessary to define ways of comparing belief functions according to their information content. Several such partial orderings, generalizing set inclusion, have been proposed [46, 19]. Among them, the q- and pl-ordering relations are defined as follows:

- m_1 is said to be q-more committed than m_2 (noted $m_1 \sqsubseteq_q m_2$) if $q_1(A) \leq q_2(A)$, for all $A \subseteq \mathcal{X}$;
- m_1 is said to be pl-more committed than m_2 (noted $m_1 \sqsubseteq_{pl} m_2$) if $pl_1(A) \leq pl_2(A)$, for all $A \subseteq \mathcal{X}$;

The interpretation of these and other ordering relations is discussed in [19] from a set-theoretical perspective, and in [21, 22] from the point of view of the TBM. In general, q- and pl-orderings are distinct notions, and none of them implies the other. However, these two orderings are equivalent in the special case of consonant belief functions: if m_1 and m_2 are consonant, then

$$m_1 \sqsubseteq_q m_2 \Leftrightarrow m_1 \sqsubseteq_{pl} m_2 \Leftrightarrow \mathrm{poss}_1 \leq \mathrm{poss}_2. \tag{8}$$

All the above notions are related to the credal level of the TBM, whereas the pignistic level [43] concerns decision making. To make decisions, any bba m such that $m(\emptyset) < 1$ is mapped into a pignistic probability function $Betp = Bet(m)$ defined by

$$Betp(x) = \sum_{A \subseteq \mathcal{X}, A \neq \emptyset} \frac{m(A)}{1 - m(\emptyset)} \frac{1_A(x)}{|A|}, \quad \forall x \in \mathcal{X}, \tag{9}$$

where 1_A denotes the indicator function of A. Note that this definition is mathematically identical to that of the Shapley value introduced in cooperative game theory [35].

Conversely, let us assume that we only know the pignistic probability function p_0 of an agent and we would like to recover its corresponding belief function. This problem is underdetermined, as the set $\mathcal{M}(p_0) = Bet^{-1}(p_0)$ of bbas m such that $Bet(m) = p_0$ is usually infinite. However, we may invoke the LCP and search for the *q-least committed* (q-LC) belief function associated to p_0 (see Figure 1). As shown in [22], the solution to this problem always exists and is unique. It is a consonant belief function, called the q-LC *isopignistic* belief function, and defined by the following possibility distribution:

$$\mathrm{poss}(x) = \sum_{x' \in \mathcal{X}} \min(p_0(x), p_0(x')). \tag{10}$$

Note that the above formula was first introduced in [18] as a probability-possibility transformation. If m is the bba associated to poss, we note $m = Bet_{LC}^{-1}(p_0)$.

Example 1 Let us consider a frame $\mathcal{X} = \{\xi_1, \xi_2, \xi_3\}$ and the probability distribution p_0 such that $p_0(\xi_1) = 0.7$, $p_0(\xi_2) = 0.2$ and $p_0(\xi_3) = 0.1$. We have

$$\begin{aligned}
\mathrm{poss}(\xi_1) &= \min(0.7, 0.7) + \min(0.7, 0.2) + \min(0.7, 0.1) = 0.7 + 0.2 + 0.1 = 1 \\
\mathrm{poss}(\xi_2) &= \min(0.2, 0.7) + \min(0.2, 0.2) + \min(0.2, 0.1) = 0.2 + 0.2 + 0.1 = 0.5 \\
\mathrm{poss}(\xi_3) &= \min(0.1, 0.7) + \min(0.1, 0.2) + \min(0.1, 0.1) = 0.1 + 0.1 + 0.1 = 0.3.
\end{aligned}$$

Using (5), we obtain the corresponding bba $m = Bet_{LC}^{-1}(p_0)$ as

$$m(\{\xi_1\}) = 0.5, \quad m(\{\xi_1, \xi_2\}) = 0.2, \quad m(\mathcal{X}) = 0.3.$$

2.2 Continuous Belief Functions on $\mathbb{R}$

Belief functions on $\mathbb{R}$ may be defined by replacing the concept of bba by that of basic belief density (bbd) [10, 36, 41]. A normal bbd m is a function taking values from the set of closed real intervals into $[0, +\infty)$, such that

$$\iint_{x \leq y} m([x, y])\, dx\, dy = 1. \tag{11}$$

The belief, plausibility and commonality functions can be defined in the same way as in the finite case, replacing finite sums by integrals. In particular,

$$bel([x, y]) = \iint_{[u,v] \subseteq [x,y]} m([u, v]) du dv, \tag{12}$$

$$pl([x, y]) = \iint_{[u,v] \cap [x,y] \neq \emptyset} m([u, v]) du dv, \tag{13}$$

$$q([x, y]) = \iint_{[u,v] \supseteq [x,y]} m([u, v]) du dv, \tag{14}$$

for all $x \leq y$.

A pignistic probability distribution $Betf = Bet(m)$ can be defined as in the discrete case. It is a continuous distribution with the following probability density [41, 7]:

$$Betf(x) = \lim_{\epsilon \to 0} \int_{-\infty}^{x} \int_{x+\epsilon}^{+\infty} \frac{m([u,v])}{v-u} dvdu. \tag{15}$$

The expression of the q-LC isopignistic bbd $m = Bet_{LC}^{-1}(f_0)$ associated with a unimodal (or "bell-shaped") probability density f_0 with mode ν was also derived in [41]. The focal sets of m are the level sets of the density function f_0. Consequently, m is consonant and the associated plausibility function is a possibility measure. The corresponding possibility distribution poss is given by:

$$\text{poss}(x) = pl(\{x\}) = \int_{-\infty}^{+\infty} \min\left(f_0(t), f_0(x)\right) dt, \tag{16}$$

which is a clearly the continuous counterpart of (10). Note that this expression was first derived in [20] as a continuous probability-possibility transformation. This transformation is illustrated in Figure 2. If f_0 is symmetrical, then poss has the following expression:

$$\text{poss}(x) = \begin{cases} 2(x-\nu)f_0(x) + 2\int_x^{+\infty} f_0(t)dt & \text{if } x \geq \nu \\ 2(\nu - x)f_0(x) + 2\int_{-\infty}^{x} f_0(t)dt & \text{otherwise.} \end{cases} \tag{17}$$

Example 2 Let f_0 be the density function of the exponential distribution $\mathcal{E}(\mu)$ with mean $\mu > 0$:

$$f_0(x;\mu) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

This is a unimodal density with mode $\nu = 0$. The corresponding q-LC distribution may be computed from (16). It is equal to

$$\text{poss}(x;\mu) = x f_0(x;\mu) + \int_x^{+\infty} f_0(t;\mu)dt \tag{19}$$

$$= e^{-x/\mu}\left(1 + \frac{x}{\mu}\right) \tag{20}$$

for $x \geq 0$ and $\text{poss}(x;\mu) = 0$ for $x < 0$. This function is plotted in Figure 3 for different values of μ.

Example 3 Let f_0 be the density function of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2:

$$f_0(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

This is a symmetrical unimodal density with mode μ. The corresponding q-LC distribution may be computed from (17). It is equal to

$$\operatorname{poss}(x;\mu,\sigma) = \begin{cases} \frac{2(x-\mu)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + 2\left(1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right) & \text{if } x \geq \mu \\ \frac{2(\mu-x)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + 2\Phi\left(\frac{x-\mu}{\sigma}\right) & \text{otherwise,} \end{cases} \tag{21}$$

where Φ is the standard normal cumulative distribution function.

This function is plotted in Figure 4 for $\mu = 0$ and three different values of σ.

3 Consonant Belief Function Induced by a Set of Pignistic Probabilities

Let us now assume that the pignistic probability distribution p_0 of an agent is only known to belong to a set $\mathcal{P}$ of probability distributions and, as before, we seek to approximate the agent's bba m_0. The problem is again underdetermined, as we can only say that m_0 belongs to the set $\mathcal{M}(\mathcal{P}) = Bet^{-1}(\mathcal{P})$ defined by

$$\begin{aligned} \mathcal{M}(\mathcal{P}) &= \{m \mid Bet(m) \in \mathcal{P}\} \\ &= \bigcup_{p\in\mathcal{P}} \mathcal{M}(p), \end{aligned}$$

where $\mathcal{M}(p) = Bet^{-1}(p)$ denotes the set of bbas whose pignistic probability distribution is equal to p.

According to the LCP, m_0 should be approximated by a bba m^* less committed than m_0, with respect to some ordering $\sqsubseteq$. In general, the set $\mathcal{M}(\mathcal{P})$ does not contain a LC element. However, we may define the *admissible* set $\mathcal{M}^*(\mathcal{P})$ as the set of bbas *dominating* (i.e., less committed than) all bbas in $\mathcal{M}(\mathcal{P})$:

$$\mathcal{M}^*(\mathcal{P}) = \{m' \mid m \sqsubseteq m', \forall m \in \mathcal{M}(\mathcal{P})\}.$$

It is then natural to choose m^* as the *most committed* element in $\mathcal{M}^*(\mathcal{P})$, if this element exists. The solution of this problem is not obvious in the general case. However, a simple solution can be found if we restrict the search to the subset $\mathcal{C}^*(\mathcal{P}) \subset \mathcal{M}^*(\mathcal{P})$ of *consonant* bbas less committed than all bbas in $\mathcal{M}(\mathcal{P})$ (see Figure 5), and we consider the q-ordering. (Note that, in Figure 5, $\mathcal{M}(p)$ and $\mathcal{M}^*(\mathcal{P})$ are represented as disjoint sets, but this may not be the case).

For all $p \in \mathcal{P}$, let $m_p = Bet_{LC}^{-1}(p)$ be the q-LC isopignistic bba induced by p. It is consonant. Let poss_p denote the corresponding possibility distribution. The following proposition holds.

Proposition 1 The bba m^* induced by possibility distribution

$$\text{poss}^*(x) = \sup_{p \in \mathcal{P}} \text{poss}_p(x), \quad \forall x \in \mathcal{X}, \tag{22}$$

is the the q-most committed element in $\mathcal{C}^*(\mathcal{P})$.

Proof. Obviously, m^* is consonant. As $\text{poss}^* \geq \text{poss}_p$ for all $p \in \mathcal{P}$, m^* is q-less committed than each m_p, which is itself the q-least committed bba in the set $\mathcal{M}(p)$ of bbas whose pignistic probability distribution is p. Consequently, m^* is q-less committed than all bbas in $\mathcal{M}(p)$ and it thus belongs to $\mathcal{C}^*(\mathcal{P})$. Now, this is also true for all bbas m associated from a possibility distribution poss such that $\text{poss} \geq \text{poss}_p$ for all $p \in \mathcal{P}$, and $\text{poss}^*(x)$ is obviously the most specific, i.e., the q-least committed element in that set. □

Possibility distribution poss^* will be referred to as the *q-most committed dominating* (q-MCD) possibility distribution associated to $\mathcal{P}$. The corresponding bba will be denoted as m^*. In the discrete case, it can be computed from poss^* using (5).

Note that the idea leading to (22) is also present in [5], where the authors consider the most committed possibility transform of each probability in a set of probabilities with density having fixed support and mode: a different problem, and a similar solution.

Example 4 Let us consider a frame $\mathcal{X} = \{\xi_1, \xi_2, \xi_3\}$ with three elements, and a set $\mathcal{P} = \{p, p', p''\}$ of three probability distributions shown in Table 1. The corresponding q-LC possibility distributions $\text{poss}, \text{poss}', \text{poss}''$ computed from (10) are displayed in Table 1. Note that there is no q-LC element among these three bbas, as poss'' is not comparable according to the $\sqsubseteq_q$ ordering with poss and poss'. Possibility distribution poss^* computed using (22) is shown in the last column of Table 1. The corresponding bba is

$$m^*(\{\xi_1\}) = 0.35, \quad m^*(\{\xi_1, \xi_2\}) = 0.05, \quad m^*(\mathcal{X}) = 0.6.$$

This bba is q-less committed than all bbas, whose pignistic distribution is in $\mathcal{P} = \{p, p', p''\}$, and it is the q-most committed among all consonant bbas verifying this property.

Remark 1 By definition, the q-MCD bba m^* is the q-most committed element among all *consonant* bbas that are q-less committed than all bbas in $\mathcal{M}(\mathcal{P})$. The restriction to consonant bbas is justified by the existence and unicity of a solution in $\mathcal{C}^*(\mathcal{P})$, whereas the existence of a q-most committed element in $\mathcal{M}^*(\mathcal{P})$ is not guaranteed in general. Additionally, finding the solution in $\mathcal{C}^*(\mathcal{P})$ is computationally tractable in several cases of practical interest, as will be shown below, and the result usually has a very simple expression. It may happen, however, that a q-most committed element in $\mathcal{M}^*(\mathcal{P})$ exists, and that it is strictly more committed than m^*. This is the case, in particular, when function q_{max} defined by

$$q_{max}(A) = \max_{p \in \mathcal{P}} q_p(A), \quad \forall A \subseteq \mathcal{X}$$

is a commonality function, q_p being the commonality function associated to m_p, . In that case, the corresponding bba m_{max} is obviously the q-most committed element in $\mathcal{M}^*(\mathcal{P})$. This is the case in Example 4: as shown in Table 2, $q_{max} = \max(q, q', q'')$ is a commonality function, and the corresponding bba m_{max} is strictly q-more committed than m^*.

Remark 2 The approach presented here is different from that introduced in [13] and [2], in which we searched for the pl-most committed bba m°, in the set $\mathcal{M}^\circ(\mathcal{P})$ of bbas that are less committed than *all probability measures* in $\mathcal{P}$ (see Figure 6). In this alternative approach, the solution is obtained as the lower envelope P_* of $\mathcal{P}$, when it is a belief function. This is the case, in particular, when $\mathcal{P}$ is a p-box [2], or when it is constructed from a multinomial confidence region with $K \leq 3$ [13]. Different heuristics were introduced in [13] for constructing a belief function less committed than P_* when P_* is not a belief function. The difference between the two approaches arises from different interpretations of the probability measures in $\mathcal{P}$: in [13] and [2], they are viewed as Bayesian belief functions, whereas in the present work they are viewed as pignistic probabilities. When $\mathcal{P}$ is a confidence set for an unknown probability measure underlying a random experiment, the former approach is thus based on Hacking's frequency principle, whereas the latter is based on a weaker form of this principle that only assumes that, if chances where known to the agent, then it would bet according to these chances.

Remark 3 Given a set $\mathcal{P}$ of probability measures, our approach computes a consonant belief function, which is formally equivalent to a possibility measure. It should be emphasized, however, that our purpose here in *not* to approximate probability families by possibility measures, as done in [5] and [28], for instance. In [5], the authors address the problem of finding the most precise possibility measure that dominates a set of probability measures. The resulting possibility measure thus has the semantics of a family of probabilities, which departs fundamentally with the semantics of belief functions in the TBM that is adopted in this paper. Consequently, the two approaches cannot be rigorously compared as they pursue different goals within two distinct theories of uncertainty. The approach described in [28] is also grounded in Possibility theory as it is based on the probability-possibility transformation introduced in [20], which is consistent with the principle of information preservation. This transformation is distinct from the inverse pignistic transformation (10) and usually yields more specific possibility measures. Technically, this approach and the one introduced in the present paper are based on similar mechanisms with two different probability-possibility transformations. However, this formal similarity should not hide the fundamental differences in the interpretations of possibility measures in the two methods. In [21], the authors advocate the use of the probability-possibility transformation introduced in [20] for transforming "objective" probabilities into possibilities, whereas the inverse pignistic transformation would be more appropriate in the case of "subjective" probabilities. While this distinction makes sense from a possibility-theoretic perspective, it does not seem to be relevant from the point of view of the TBM. In particular, the problem of *approximating* a probability measure by a possibility measure does not arise in the TBM, as a probability measure is already a belief function. Consequently, the role of the probability-possibility transformation introduced in [20] is not clear in the TBM, whereas the inverse pignistic transformation used in this paper does have a well-defined meaning in this framework.

4 Application to a Sample of a Discrete Random Variable

In this section, we consider the application of the above methodology to the construction of a predictive belief function based on an independent and identically distributed (iid) sample $X_1, \ldots, X_n$ from a discrete variable X defined on a finite domain $\mathcal{X}$. We first show that a set $\mathcal{P}$ of possible probability distributions of X can be constructed using multinomial simultaneous confidence intervals. An algorithm for finding the q-MCD possibility distribution poss* induced by $\mathcal{P}$ is then presented.

4.1 Construction of $\mathcal{P}$

Let X be a discrete r.v. on a finite domain $\mathcal{X} = \{\xi_1, \ldots, \xi_K\}$, with unknown probability distribution $\mathbb{P}_X$. Given an iid random sample $X_1, \ldots, X_n$ from $\mathbb{P}_X$, let $n_k = \sum_{i=1}^{n} 1_{\xi_k}(X_i)$ denote the number of observations in category ξ_k. The random vector $\mathbf{n} = (n_1, \ldots, n_K)$ has a multinomial distribution with parameters n and $\mathbf{p} = (p_1, \ldots, p_K)$, with $p_k = \mathbb{P}_X(\{\xi_k\})$.

Let $S(\mathbf{n})$ be a random subset of the parameter space $\Theta = \{\mathbf{p} = (p_1, \ldots, p_K) \in [0,1]^K \mid \sum_{k=1}^{K} p_k = 1\}$. $S(\mathbf{n})$ is said to be a confidence region for $\mathbf{p}$ at confidence level $1 - \alpha$, if

$$\mathbb{P}(S(\mathbf{n}) \ni \mathbf{p}) \geq 1 - \alpha,$$

i.e., the random region $S(\mathbf{n})$ contains the constant parameter vector $\mathbf{p}$ with probability (long-run frequency) $1 - \alpha$. It is an asymptotic confidence region if the above inequality only holds in the limit, as $n \to \infty$.

Of particular interest are simultaneous confidence intervals, i.e., regions defined as a Cartesian product of intervals:

$$S(\mathbf{n}) = [p_1^-, p_1^+] \times \ldots \times [p_K^-, p_K^+],$$

which have easy interpretation. Such asymptotic confidence regions were proposed by Quesenberry and Hurst [31], and Goodman [25]. Goodman's intervals are defined as:

$$p_k^- = \frac{a + 2n_k - \sqrt{\Delta_k}}{2(n+a)} \tag{23}$$

$$p_k^+ = \frac{a + 2n_k + \sqrt{\Delta_k}}{2(n+a)}, \tag{24}$$

where a is the quantile of order $1 - \alpha/K$ of the chi-square distribution with one degree of freedom (for $K > 2$), and

$$\Delta_k = a\left(a + \frac{4n_k(n - n_k)}{n}\right).$$

When $K = 2$, a should be defined as quantile of order $1 - \alpha$ of the chi-square distribution with one degree of freedom. Note that p_k^- and p_k^+ both converge in probability towards p_k as $n \to +\infty$, for $k = 1, \ldots, K$.

As remarked in [13, 28], $S(\mathbf{n})$ can be seen as defining a family $\mathcal{P}$ of probability measures. Such a family, obtained by bounding the probability of each singleton, is called a *set of probability intervals* in [8]. Each vector $\mathbf{p}$ of probabilities corresponds to a possible probability measure p for X.

Example 5 The data analyzed in [31] and [25] describe the frequency of ten modes of failure as recorded in a study of 870 machines that failed. These data are shown in Table 3, together with the corresponding Goodman confidence intervals at confidence level $1-\alpha = 0.90$.

4.2 Determination of the q-MCD Possibility Distribution

Following the approach outlined in the previous section, assume that $\mathcal{P}$ is interpreted as a set of pignistic probabilities. For each $Betp$ in $\mathcal{P}$, the q-LC isopignistic belief function is defined by (10). Consequently, the q-MCD possibility distribution poss* defined by (22) can be obtained by solving the following maximization problems for $k = 1, \ldots, K$:

$$\mathrm{poss}^*_k = \max_{\mathbf{p}} \sum_{\ell=1}^{K} \min(p_k, p_\ell) \tag{25}$$

under the constraints

$$p_\ell^- \leq p_\ell \leq p_\ell^+, \quad \ell = 1, \ldots, K \tag{26}$$

$$\sum_{\ell=1}^{K} p_\ell = 1. \tag{27}$$

This problem is similar to the one addressed in [28] for a different probability-possibility transformation. It may be remarked that the constraints defined by equations (26) and (27) are always feasible when the probability bounds are computed using (23) and (24), as Goodman's confidence region is never empty. The domain $\mathcal{R}$ of $\mathbb{R}^K$ defined by these constraints is closed and bounded, and the function to be maximized is continuous: consequently, it has a global maximum in $\mathcal{R}$, which explains why the supremum in (22) has been replaced by a maximum in (25). We may also notice that the solution has a simple upper bound $\widetilde{\mathrm{poss}}^*_k$ given by

$$\widetilde{\mathrm{poss}}^*_k = \min\left(1, \sum_{\ell=1}^{K} \min(p_k^+, p_\ell^+)\right), \tag{28}$$

which can be used as an approximation. It also has a lower bound given by $\sum_{\ell=1}^K \min(p_k^-, p_\ell^-)$, which is, however, less useful as an approximation.

The exact solution to optimization problem (25)-(27) may be found by reasoning as follows.

We first observe that (25) can be written as

$$\mathrm{poss}_k^* = \max_{\mathbf{p}} \sum_{\ell \in \overline{S_k(\mathbf{p})}} p_\ell + |S_k(\mathbf{p})| p_k, \tag{29}$$

where $S_k(\mathbf{p}) = \{\ell \in \{1, \ldots, K\} \mid p_\ell \geq p_k\}$ is the set of indices of probabilities p_ℓ at least equal to p_k, $\overline{S_k(\mathbf{p})}$ is the complement of $S_k(\mathbf{p})$, and $|S_k(\mathbf{p})|$ is its cardinality. For any $G \subseteq \{1, \ldots, K\}$, let $\mathcal{P}_{k,G} = \{\mathbf{p} \in \mathcal{P} | S_k(\mathbf{p}) = G\}$. If $\mathcal{P}_{k,G}$ is nonempty, then the maximum of $\sum_{\ell=1}^K \min(p_k, p_\ell)$ over $\mathcal{P}_{k,G}$ may be found by maximizing

$$\sum_{\ell \in \overline{G}} p_\ell + |G| p_k,$$

under the constraints (26)-(27) and

$$p_\ell \geq p_k, \quad \forall \ell \in G. \tag{30}$$

$$p_\ell < p_k, \quad \forall \ell \in \overline{G}, \tag{31}$$

which is a linear optimization problem that may be solved using a standard linear programming algorithm. Let

$$\mathcal{G}_k = \{G \subseteq \{1, \ldots, K\} | \mathcal{P}_{k,G} \neq \emptyset\}.$$

We may then write

$$\mathcal{P} = \bigcup_{G \in \mathcal{G}_k} \mathcal{P}_{k,G}$$

and

$$\max_{\mathbf{p} \in \mathcal{P}} \sum_{\ell=1}^K \min(p_k, p_\ell) = \max_{G \in \mathcal{G}_k} \max_{\mathbf{p} \in \mathcal{P}_{k,G}} \sum_{\ell=1}^K \min(p_k, p_\ell).$$

The optimization problem defined by (25)-(27) may thus be decomposed into a series of linear programming problems.

To enumerate the elements of $\mathcal{G}_k$, we may observe that any $G \in \mathcal{G}_k$ necessarily contains k and indices ℓ such that $p_\ell^- \geq p_k^+$, and cannot contain indices ℓ such that $p_\ell^+ < p_k^-$. All other indices may be included in G or not. Formally, let

$$S_k^* = \{k\} \cup \{\ell \in \{1, \ldots, K\} \mid p_\ell^- \geq p_k^+\}, \tag{32}$$

$$I_k^* = \{\ell \in \{1, \ldots, K\} \mid p_\ell^+ < p_k^-\}, \tag{33}$$

and

$$P_k^* = \{1, \ldots, K\} \setminus (S_k^* \cup I_k^*). \tag{34}$$

Then, all possible sets G are of the form $G = S_k^* \cup A$ for $A \subseteq P_k^*$.

The proposed algorithm may be summarized as follows:

1. Initialize $\text{poss}_k^* = 0$.
2. Compute S_k^*, I_k^* and P_k^* using (32)-(34).
3. For all $A \subseteq P_k^*$:
   (a) Let $G = S_k^* \cup A$.
   (b) If constraints (26)-(27) and (30)-(31) are feasible, then
      i. Compute $\text{poss}_k^*(G) = \max_{\mathbf{p}} \sum_{\ell \in \overline{G}} p_\ell + |G| p_k$ under constraints (26)-(27) and (30)-(31) using a linear programming procedure.
      ii. $\text{poss}_k^* = \max(\text{poss}_k^*, \text{poss}_k^*(G))$.
   (c) End if.
4. End For.

Regarding the complexity of this algorithm, it may be noticed that the number of the linear programming problems to be solved is, in the worst case, exponential with respect to the size of the set P_k^*. For large K, it may thus be necessary to resort to the approximation given by (28).

Example 6 Let us come back to the data of Example 5 reported in Table 3. Detailed calculations for $k = 7$ are presented in Appendix A. The values of poss_k^* for $k = 1, \ldots, 10$ are shown in Table 4, together with the approximations $\widetilde{\text{poss}}_k^*$ computed using (28). The q-LC possibility distribution $\widehat{\text{poss}}$ computed from the sample frequencies n_k/n is also shown in Table 4. This possibility distribution is more committed than poss^* as it does not take into account sampling uncertainty.

Remark 4 Example 6 above may be used to illustrate an important point regarding the way a belief function constructed from sample data should be updated to account for new evidence pertaining to a given situation under study. Assume that, after inspecting a given machine, we arrive at the conclusion that its failure mode is in $A = \{\xi_7, \xi_8, \xi_9, \xi_{10}\}$. How should we update our beliefs based on this evidence? We could either condition our q-MCD possibility distribution by A using Dempster's rule of conditioning (7), or we could build a new q-MCD possibility distribution using a reduced sample composed only of observations of the last four failure modes. In the first case we would get a consonant belief function with associated possibility distribution:

$$\mathrm{poss}^{*\prime}{}_7 = 0.804, \quad \mathrm{poss}^{*\prime}{}_8 = 0.867, \quad \mathrm{poss}^{*\prime}{}_9 = 0.935, \quad \mathrm{poss}^{*\prime}{}_{10} = 1,$$

and $\mathrm{poss}^{*\prime}{}_k = 0$ for $k \in \{1, \ldots, 6\}$. In the second case, applying the above algorithm to the $92 + 118 + 173 + 297$ observations corresponding to failure modes ξ_7 to ξ_{10} yields the following possibility distribution:

$$\mathrm{poss}^{*\prime\prime}{}_7 = 0.670, \quad \mathrm{poss}^{*\prime\prime}{}_8 = 0.793, \quad \mathrm{poss}^{*\prime\prime}{}_9 = 0.899, \quad \mathrm{poss}^{*\prime\prime}{}_{10} = 1,$$

and $\mathrm{poss}^{*\prime\prime}{}_k = 0$ for $k \in \{1, \ldots, 6\}$. The two methods thus produce different results. Intuitively, the second approach seems preferable: if we are sure that the failure mode is in A, then statistical data related to other modes are irrelevant. The process of discarding irrelevant statistical data seems to be related to a general mechanism for updating generic knowledge based on specific evidence pertaining to a particular case, referred to as "focussing" in [16]. From the point of view of the TBM, the fact that Dempster's rule of conditioning is not suitable in this case means that the statistical sample and factual evidence about the particular case under study are not distinct items of evidence: acquiring one piece of evidence (partial information about the true failure mode) changes the way a belief function is selected based on the other piece of evidence (the statistical observations). This is exactly the definition of non distinctness given by Smets in [39, pages 279-280].

Remark 5 Other methods for generating a predictive belief functions in the case of discrete sample data have been proposed by Walley [45], Denœux [13] and Masson and Denœux [28]. Although each of these methods basically address the same problem, they do it in different frameworks or using different hypotheses. The methods introduced in [13] and [28] have already been discussed from a conceptual point of view in Remarks 2 and 3, respectively. Walley's imprecise Dirichlet model [45, 6] extends Bayesian inference by considering a set of Dirichlet priors on the parameters of the multinomial model. It turns out that the resulting family of predictive distributions of X is characterized by a lower probability measure which is formally a belief function. However, as argued in [13], the asumption that one's prior knowledge on the probability distribution of X is represented by a family of Dirichlet distribution has no justification in the TBM. This model makes sense in the imprecise probability framework, but it seems to be totally unrelated to the TBM. Each of these methods is thus consistent within a given uncertainty representation framework. A comparison of the efficiency of these frameworks for handling various problems involving uncertainty goes beyond the scope of this paper.

5 Application to Continuous Parametric Models

The general approach introduced in Section 3 can also be applied to the construction of a predictive belief function based on a sample from a continuous r.v. X with unimodal probability density function $f(x;\theta)$ depending on a parameter θ. For each value of θ, the q-LC possibility distribution $\mathrm{poss}(x;\theta)$ may be computed using (16) or (17). Given a confidence region $\mathcal{R}$ for θ, one may then compute the q-MCD possibility distribution poss^* as

$$\mathrm{poss}^*(x) = \sup_{\theta\in\mathcal{R}} \mathrm{poss}(x;\theta),$$

for all $x \in \mathbb{R}$.

This approach is illustrated below in the cases of exponential and normal distributions.

5.1 Exponential Distribution

Let us assume that X has an exponential distribution $\mathcal{E}(\mu)$ with density function $f(x;\mu)$ defined by (18). As shown in Example 2, the corresponding q-LC possibility distribution is defined for fixed μ by (20).

Here, we assume that μ is unknown but we have observed an iid sample $X_1, \ldots, X_n$ from $\mathcal{E}(\mu)$. It is well known from standard textbooks (see, e.g. [23]) that the sample average $\overline{X}$ is an unbiased estimator for μ, and its variance is μ^2/n. From the Central Limit Theorem, the statistics

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\mu}$$

converges in distribution to a r.v. that is normally distributed with mean 0 and variance 1. For large n and $\alpha \in (0,1)$, we thus have

$$\mathbb{P}\left(-u_{1-\alpha/2} \leq \frac{\sqrt{n}(\overline{X}-\mu)}{\mu} \leq u_{1-\alpha/2}\right) \approx 1-\alpha,$$

where $u_{1-\alpha/2}$ is the upper $\alpha/2$ percentile of a standard normal distribution. Equivalently,

$$\mathbb{P}\left(\frac{\overline{X}}{1+u_{1-\alpha/2}/\sqrt{n}} \leq \mu \leq \frac{\overline{X}}{1+u_{1+\alpha/2}/\sqrt{n}}\right) \approx 1-\alpha.$$

The interval

$$\mathcal{R}(X_1,\ldots,X_n) = \left\{\mu : \frac{\overline{X}}{1+u_{1-\alpha/2}/\sqrt{n}} \leq \mu \leq \frac{\overline{X}}{1-u_{1-\alpha/2}/\sqrt{n}}\right\}$$

is thus an approximate confidence interval for μ at level $1-\alpha$.

To compute the supremum of $\text{poss}(x;\mu)$ for $\mu \in \mathcal{R}(X_1,\ldots,X_n)$, we observe that

$$\frac{\partial \text{poss}(x;\mu)}{\partial \mu} = \frac{x^2}{\mu^3} e^{-x/\mu} > 0.$$

Consequently,

$$\text{poss}^*(x) = \text{poss}\left(x; \widehat{\mu}^+\right)$$

with

$$\widehat{\mu}^+ = \frac{\overline{X}}{1-u_{1-\alpha/2}/\sqrt{n}}.$$

Figure 7 shows the possibility distribution $\text{poss}^*(x)$ for $\overline{x} = 1$ and various values of n. The case $n = \infty$ corresponds to the situation where parameter μ is known: in that case, poss^* is simply the q-LC isopignistic possibility distribution induced by the exponential pignistic distribution with $\mu = 1$.

Example 7 Suppose that the life time X of light bulbs manufactured by a certain company follows an exponential distribution $\mathcal{E}(\mu)$ with unknown μ. For $n = 20$ bulbs, the average observed life time was $\overline{x} = 30.5$ hours. What are the belief and plausibility that the life time of a new bulb will exceed 50 hours ?

For $\alpha = 0.05$, we have $u_{1-\alpha/2} = 1.96$ and $\widehat{\mu}^+ = 30.5/(1 - 1.96/\sqrt{20}) = 54.3$. The q-MCD possibility distribution is thus

$$\text{poss}^*(x) = e^{-x/54.3}\left(1 + \frac{x}{54.3}\right).$$

We have

$$pl([50, +\infty)) = \sup_{x \geq 50} \text{poss}^*(x) = \text{poss}^*(50) = 0.76$$

and

$$bel([50, +\infty)) = 1 - pl([0, 50)) = 1 - \sup_{0 \leq x < 50} \text{poss}^*(x) = 1 - \text{poss}^*(0) = 0.$$

5.2 Normal Distribution

Let us now assume that X has a normal distribution with mean μ and variance σ^2. If these two parameters are known, then the possibility distribution $\text{poss}(\cdot; \mu, \sigma)$ is given by (21).

When μ and σ^2 are unknown but an iid sample $X_1, \ldots, X_n$ is available, then it is possible to define a joint confidence region for μ and σ^2 [4]. In particular, the Mood exact confidence region at level $1 - \alpha = (1 - \alpha_1)(1 - \alpha_2)$ is defined by

$$\mathcal{R}(X_1, \ldots, X_n) = \Bigg\{ (\mu; \sigma^2) : \overline{X} - u_{1-\alpha_1/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{1-\alpha_1/2}\frac{\sigma}{\sqrt{n}}, \\ \frac{nS^2}{\chi^2_{n-1;1-\alpha_2/2}} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{n-1;\alpha_2/2}} \Bigg\}, \quad (35)$$

where $\overline{X}$ is the sample mean, $S^2 = (1/n)\sum_{i=1}^n (X_i - \overline{X})^2$ is the sample variance, $u_{1-\alpha_1/2}$ is the upper $\alpha_1/2$ percentile of a standard normal distribution, and $\chi^2_{n-1;\alpha_2/2}$ and $\chi^2_{n-1;1-\alpha_2/2}$ are the lower and upper $\alpha_2/2$ percentiles of a χ^2_{n-1} distribution. The shape of that region is illustrated in Figure 8. Values of α_1 and α_2 yielding a region of smallest possible size for a fixed confidence level are given in [4].

Let $\mathcal{P}$ denote the set of Gaussian distributions with parameters contained in confidence region $\mathcal{R}$. Applying the principle outlined in Section 3, we may obtain the q-MCD possibility distribution poss^* for any x by maximizing $\text{poss}(x; \mu, \sigma)$ given by (21) with respect to μ and σ, under the constraint $(\mu, \sigma^2) \in \mathcal{R}$. The result is given by the following proposition.

Proposition 2 The q-MCD possibility distribution poss* associated with the Mood confidence confidence region $\mathcal{R}$ at level $(1-\alpha_1)(1-\alpha_2)$ is

$$\text{poss}^*(x) = \begin{cases} \text{poss}(x; \widehat{\mu}^-, \widehat{\sigma}^+) & \text{if } x < \widehat{\mu}^- \\ 1 & \text{if } \widehat{\mu}^- \leq x \leq \widehat{\mu}^+ \\ \text{poss}(x; \widehat{\mu}^+, \widehat{\sigma}^+) & \text{if } x > \widehat{\mu}^+, \end{cases} \tag{36}$$

with

$$\widehat{\sigma}^+ = \left(\frac{nS^2}{\chi^2_{n-1;\alpha_2/2}} \right)^{1/2},$$

$$\widehat{\mu}^- = \overline{X} - u_{1-\alpha_1/2} \frac{\widehat{\sigma}^+}{\sqrt{n}}, \quad \widehat{\mu}^+ = \overline{X} + u_{1-\alpha_1/2} \frac{\widehat{\sigma}^+}{\sqrt{n}}.$$

Proof. We have by definition

$$\text{poss}^*(x) = \sup_{(\mu,\sigma^2)\in\mathcal{R}} \text{poss}(x; \mu, \sigma).$$

If $x \in [\widehat{\mu}^-, \widehat{\mu}^+]$, then we can get $\text{poss}(x;\mu,\sigma) = 1$ by setting $\mu = x$ and $\sigma = \widehat{\sigma}^+$. If $x < \widehat{\mu}^-$, then the value 1 cannot be reached. However, we obtain using standard calculus for $x < \mu$:

$$\frac{\partial \text{poss}(x;\mu,\sigma)}{\partial \mu} = -\frac{(x-\mu)^2}{\sigma^3\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) < 0$$

and

$$\frac{\partial \text{poss}(x;\mu,\sigma)}{\partial \sigma} = \frac{(\mu-x)^3}{\sigma^4\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) > 0.$$

Consequently, $\text{poss}(x;\mu,\sigma)$ is maximized by jointly minimizing μ and maximizing σ, and the maximum is reached for $(\mu,\sigma) = (\widehat{\mu}^-, \widehat{\sigma}^+)$. Similarly, we get for $x > \widehat{\mu}^+$:

$$\frac{\partial \text{poss}(x;\mu,\sigma)}{\partial \mu} = \frac{(x-\mu)^2}{\sigma^3\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) > 0$$

and

$$\frac{\partial \text{poss}(x;\mu,\sigma)}{\partial \sigma} = \frac{(x-\mu)^3}{\sigma^4\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) > 0.$$

Consequently, the maximum of $\text{poss}(x;\mu,\sigma)$ for $x > \widehat{\mu}^+$ is reached for $(\mu,\sigma) = (\widehat{\mu}^+, \widehat{\sigma}^+)$. □

Figure 9 shows the possibility distribution $\text{poss}^*(x)$ for $\overline{x} = 0$, $s^2 = 1$, $\alpha = 0.1$, $\alpha_1 = \alpha_2$ and various values of n. The case $n = \infty$ corresponds to the situation where parameters μ and σ^2 are known: in that case, poss* is simply the q-LC isopignistic possibility distribution induced by the normal pignistic distribution with $\mu = 0$ and $\sigma^2 = 1$.

6 Application to Classification

To demonstrate the usefulness of the proposed approach for constructing a belief function from sample data, let us consider the following multi-sensor classification problem.

6.1 Problem Statement and Solution in the TBM

Let Σ denote a system that can be in two states (classes) ω_1 and ω_2 corresponding, e.g., to the normal state and a faulty state. Let $\Omega = \{\omega_1, \omega_2\}$. The system is equipped with two sensors S_x and S_y that deliver measurements X and Y, considered to be r.v.'s with distribution depending on the system state. Both r.v.'s are assumed to be normally distributed and independent conditionally on the system state.

Let us further assume that sensor S_x has been available for a long time, so that we have gathered a learning set $\mathcal{L}_x$ of $n_x = 1000$ observations of X from each class. In contrast, sensor S_y is recent and we have only a much smaller learning set $\mathcal{L}_y$ of $n_y \ll n_x$ observations of Y from each class.

Based on this information, we would like to construct a decision rule for predicting the system state from measurements x_0 and y_0 delivered by the two sensors.

In the TBM, the solution of this problem goes through the following steps [37, 9, 15]:

1. Compute the plausibilities $pl(x_0|\omega_k)$ and $pl(y_0|\omega_k)$ of observing x_0 and y_0, respectively, when the system is in state ω_k $(k = 1, 2)$ using the learning data;

2. As X and Y are conditionally independent, let $pl(x_0, y_0|\omega_k) = pl(x_0|\omega_k)pl(y_0|\omega_k)$.

3. Using the General Bayesian Theorem (GBT) [37], compute the conditional bba $m^\Omega(\cdot|x_0, y_0)$ on Ω given $X = x_0$ and $Y = y_0$ using the following formula:

$$m^\Omega(\cdot|x_0, y_0) = \{\omega_1\}^{pl(x_0,y_0|\omega_2)} \circledcirc \{\omega_2\}^{pl(x_0,y_0|\omega_1)},$$

   where the notation $\{\omega_k\}^w$ stands for the simple bba m such that $m(\{\omega_k\}) = 1 - w$

and $m(\Omega) = w$. We thus have

$$m^\Omega(\emptyset|x_0, y_0) = (1 - pl(x_0, y_0|\omega_1))\,(1 - pl(x_0, y_0|\omega_2)) \quad (37)$$

$$m^\Omega(\{\omega_1\}|x_0, y_0) = pl(x_0, y_0|\omega_1)\,(1 - pl(x_0, y_0|\omega_2)) \quad (38)$$

$$m^\Omega(\{\omega_2\}|x_0, y_0) = (1 - pl(x_0, y_0|\omega_1))\,pl(x_0, y_0|\omega_2) \quad (39)$$

$$m^\Omega(\Omega|x_0, y_0) = pl(x_0, y_0|\omega_1)pl(x_0, y_0|\omega_2). \quad (40)$$

4. Compute the pignistic probability $BetP^\Omega(\cdot|x_0, y_0)$ induced by $m^\Omega(\cdot|x_0, y_0)$:

$$\begin{aligned} BetP^\Omega(\omega_1|x_0, y_0) &= \frac{m^\Omega(\{\omega_1\}|x_0, y_0) + m^\Omega(\Omega|x_0, y_0)/2}{1 - m^\Omega(\emptyset|x_0, y_0)}, \\ BetP^\Omega(\omega_2|x_0, y_0) &= 1 - BetP^\Omega(\omega_1|x_0, y_0). \end{aligned}$$

5. Select the system state with the highest pignistic probability.

The approach exposed in this paper concerns step 1. The plausibilities $pl(x_0|\omega_k)$ and $pl(y_0|\omega_k)$ may either be computed from (21) by substituting the mean and standard deviation by their sample estimates (this method will be referred to as LC), or from (36) using Mood confidence regions (MCD method). In the latter case, the plausibility values will be higher, reflecting the additional sampling uncertainty.

6.2 Illustrative Example

Figures 10 and 11 show typical learning sets $\mathcal{L}_x$ and $\mathcal{L}_y$ with, respectively, $n_x = 1000$ and $n_y = 50$ observations for each class, as well as the corresponding possibility distributions computed using each of the two methods. For the MCD method, the confidence level of the Mood regions were fixed at $1-\alpha = 0.8$. The values $x_0 = 1.5$ and $y_0 = -1$ are indicated as vertical lines in the upper parts of Figures 10 and 11.

Let us first do the computations for the LC method. We have $pl(x_0|\omega_1) = 0.517$, $pl(x_0|\omega_2) = 0.966$, $pl(y_0|\omega_1) = 0.828$, $pl(y_0|\omega_2) = 0.537$. Hence (step 2),

$$\begin{aligned} pl(x_0, y_0|\omega_1) &= 0.517 \times 0.828 = 0.428 \\ pl(x_0, y_0|\omega_2) &= 0.966 \times 0.537 = 0.519. \end{aligned}$$

Using (37)-(40) we get (step 3):

$$\begin{aligned}
m^{\Omega}(\emptyset|x_0, y_0) &= (1 - 0.428)\,(1 - 0.519) = 0.275 \\
m^{\Omega}(\{\omega_1\}|x_0, y_0) &= 0.428 \times (1 - 0.519) = 0.206 \\
m^{\Omega}(\{\omega_2\}|x_0, y_0) &= (1 - 0.428) \times 0.519 = 0.297 \\
m^{\Omega}(\Omega|x_0, y_0) &= 0.428 \times 0.519 = 0.222.
\end{aligned}$$

The corresponding pignistic probability function is:

$$\begin{aligned}
BetP^{\Omega}(\omega_1|x_0, y_0) &= 0.437, \\
BetP^{\Omega}(\omega_2|x_0, y_0) &= 0.563.
\end{aligned}$$

Using the MCD method, we have $pl(x_0|\omega_1) = 0.579$, $pl(x_0|\omega_2) = 0.978$, $pl(y_0|\omega_1) = 0.960$, $pl(y_0|\omega_2) = 0.807$. We thus get

$$\begin{aligned}
pl(x_0, y_0|\omega_1) &= 0.579 \times 0.960 = 0.556 \\
pl(x_0, y_0|\omega_2) &= 0.978 \times 0.807 = 0.789,
\end{aligned}$$

and

$$\begin{aligned}
m^{\Omega}(\emptyset|x_0, y_0) &= (1 - 0.556)\,(1 - 0.789) = 0.094 \\
m^{\Omega}(\{\omega_1\}|x_0, y_0) &= 0.556 \times (1 - 0.789) = 0.117 \\
m^{\Omega}(\{\omega_2\}|x_0, y_0) &= (1 - 0.556) \times 0.789 = 0.350 \\
m^{\Omega}(\Omega|x_0, y_0) &= 0.556 \times 0.789 = 0.439.
\end{aligned}$$

Finally, the corresponding pignistic probability functions is

$$\begin{aligned}
BetP^{\Omega}(\omega_1|x_0, y_0) &= 0.372, \\
BetP^{\Omega}(\omega_2|x_0, y_0) &= 0.628.
\end{aligned}$$

We observe that the observation x_0 tends to point to class ω_2 (as $pl(x_0|\omega_2) > pl(x_0|\omega_1)$), whereas y_0 points to class ω_1 (as $pl(y_0|\omega_1) > pl(y_0|\omega_2)$). Using the LC method, the two observations counterbalance each other, and the resulting pignistic probabilities are close to 0.5. However, using the MCD method, the plausibilities $pl(y_0|\omega_1)$, $pl(y_0|\omega_2)$ are significantly closer to unity, reflecting weak knowledge of the distribution of Y in both classes, due to the small number of training examples in $\mathcal{L}_y$. As a consequence, the impact of observation y_0 is less important, resulting in a higher pignistic probability assigned to class ω_2.

In this simple example, the final decision does not change. However, it is clear that the two methods for computing the plausibilities of observations in each class may lead to different decisions. As the MCD method takes into account the different sizes of $\mathcal{L}_x$ and $\mathcal{L}_y$ and, as a consequence, gives less importance to sensor S_y in the decision, it may be expected to result in better performance. This will be verified in the following section.

6.3 Numerical Experiment

To study the impact of the MCD method for computing the class-conditional plausibilities in the above scheme, a numerical experiment was carried out as follows. The following conditional distributions of X and Y were assumed: $f(x|\omega_1) \sim \mathcal{N}(0,1)$, $f(x|\omega_2) \sim \mathcal{N}(2,1)$, $f(y|\omega_1) \sim \mathcal{N}(0,1)$, $f(y|\omega_2) \sim \mathcal{N}(0.5,1)$.

A test set of 1000 examples for each class was randomly generated. The size of $\mathcal{L}_x$ was fixed to $n_x = 1000$, while the size n_y of $\mathcal{L}_y$ was varied in $\{10, 50, 100\}$. For each value of n_y, the following procedure was repeated 50 times:

- Generate randomly a learning set $\mathcal{L}_x$ of size $n_x = 1000$;
- Generate randomly a learning set $\mathcal{L}_y$ of size n_y;
- Classify each test example using the approach described in Section 6.1 and each of the following options
  - use only $pl(x_0|\omega_k)$ $(k = 1,2)$ computed using the MCD method (the decision is the same if the LC method is used instead);
  - use $pl(x_0|\omega_k)$ and $pl(y_0|\omega_k)$ $(k = 1,2)$ computed using the LC method;
  - use $pl(x_0|\omega_k)$ and $pl(y_0|\omega_k)$ $(k = 1,2)$ computed using the MCD method;
- Compute the error rates err_x, err_{LC} and err_{MCD} using the three methods.

Additionally, we also computed for comparison the classification results of the standard Bayesian approach with estimated class parameters (as in the LC and MCD approaches) and equal prior probabilities.

The results are shown in Figure 12. In this figure, each box plot represents a distribution over 50 trials. Each box has lines at the lower quartile, median, and upper quartile values. The whiskers extending from each end of the box show the extent of the rest of the data (except outliers represented separately). Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level.

We can see that the MCD method significantly outperforms the LC method, especially for small values of n_y. For $n_y = 50$ and $n_y = 100$, both methods take advantage of information from sensor S_y, as they reach significantly lower error rates than that those obtained using sensor S_x alone. For $n_y = 10$, the LC method exhibits very poor performances and a very high variance. In contrast, the MCD method has uniformly good performances for all values of n_y, and a much lower variance for small sample size. The MCD and LC methods both outperform the Bayesian method for all values of n_y.

7 Conclusion

A new method for generating a belief function from statistical data in the TBM framework has been presented. The starting point of this method is the assumption that, if the probability distribution $\mathbb{P}_X$ of a random variable is known, then the belief function quantifying our belief regarding a future realization of X should be such that its pignistic probability distribution equals $\mathbb{P}_X$. In the realistic situation where $\mathbb{P}_X$ is unknown but a random sample of X is available, it is possible to build a set $\mathcal{P}$ of probability distributions containing $\mathbb{P}_X$ with some confidence level. Following the LCP, it is then reasonable to impose that the sought belief function be q-less committed than all belief functions whose pignistic probability distribution is in $\mathcal{P}$. Our method selects the q-most committed consonant belief function verifying this property, referred to as the q-MCD possibility distribution induced by $\mathcal{P}$. This general principle has been illustrated in three special cases of general interest involving discrete, exponential and normal distributions, respectively.

In conjunction with the General Bayesian Theorem [37], the q-LC isopignistic transformation has proved useful to tackle classification problems using the TBM [33]. In this approach, the parameters of the pignistic distributions are usually assumed to be given by experts or estimated using large samples. Using the tools presented in this paper, it is possible to apply this methodology to a wider range of problems where only small datasets are available. A first numerical experiment demonstrating the advantages of this approach has been presented here. Future work in this direction should allow a more widespread application of the TBM to classification and statistical learning problems.

Acknowledgment

The authors thank the three anonymous referees for their careful reading of the manuscript and their constructive comments, as well as Marco Zaffalon for handling the review of this paper as area editor.

A Detailed Calculations for Example 6

Let us consider the calculation of poss_7^* in Example 6 presented in Section 4.2. We have $S_7^* = \{7, 9, 10\}$, $I_7^* = \{1, 2, 3, 4\}$ and $P_7^* = \{5, 6, 8\}$. Using the algorithm described in Section 4.2, we have to solve a distinct linear optimization problem for each of the $2^3 = 8$ subsets A of P_7^*. Let us consider these eight cases:

- For $A = \emptyset$ we have $G = \{7, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{7, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 5, 6, 8\}$$

  are consistent. The maximum of $\sum_{\ell=1}^{6} p_\ell + 3p_7 + p_8$ under these constraints is 0.804; it is achieved for

$$\mathbf{p} = (0.013, 0.021, 0.030, 0.043, 0.076, 0.086, 0.136, 0.128, 0.166, 0.301).$$

- For $A = \{5\}$ we have $G = \{5, 7, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{5, 7, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 6, 8\}$$

  are not consistent, so the optimization problem is not feasible.

- For $A = \{6\}$ we have $G = \{6, 7, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{6, 7, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 5, 8\}$$

are not consistent, so the optimization problem is not feasible.

- For $A = \{8\}$ we have $G = \{7, 8, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{7, 8, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 5, 6\}$$

are consistent. The maximum of $\sum_{\ell=1}^{6} p_\ell + 4p_7$ under these constraints is 0.804; it is achieved for

$$\mathbf{p} = (0.013, 0.020, 0.029, 0.042, 0.074, 0.083, 0.136, 0.136, 0.166, 0.301).$$

- For $A = \{5, 6\}$ we have $G = \{5, 6, 7, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{5, 6, 7, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 8\}$$

are consistent. The maximum of $\sum_{\ell=1}^{4} p_\ell + 5p_7 + p_8$ under these constraints is 0.659; it is achieved for

$$\mathbf{p} = (0.010, 0.017, 0.026, 0.038, 0.092, 0.098, 0.092, 0.109, 0.192, 0.327).$$

- For $A = \{5, 8\}$ we have $G = \{5, 7, 8, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{5, 7, 8, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 6\}$$

are consistent. The maximum of $\sum_{\ell=1}^{4} p_\ell + p_6 + 5p_7$ under these constraints is 0.688; it is achieved for

$$\mathbf{p} = (0.017, 0.027, 0.039, 0.054, 0.092, 0.092, 0.092, 0.112, 0.170, 0.306).$$

- For $A = \{6, 8\}$ we have $G = \{6, 7, 8, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{6, 7, 8, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4, 5\}$$

are consistent. The maximum of $\sum_{\ell=1}^{5} p_\ell + 5p_7$ under these constraints is 0.735; it is achieved for

$$\mathbf{p} = (0.014, 0.024, 0.037, 0.054, 0.089, 0.104, 0.104, 0.109, 0.166, 0.301).$$

- For $A = \{5, 6, 8\}$ we have $G = \{5, 6, 7, 8, 9, 10\}$. The constraints (26), (27) and

$$p_\ell \geq p_k, \quad \forall \ell \in \{5, 6, 7, 8, 9, 10\},$$

$$p_\ell < p_k, \quad \forall \ell \in \{1, 2, 3, 4\}$$

are consistent. The maximum of $\sum_{\ell=1}^{4} p_\ell + 6p_7$ under these constraints is 0.688; it is achieved for

$$\mathbf{p} = (0.017, 0.027, 0.039, 0.054, 0.092, 0.093, 0.092, 0.111, 0.170, 0.305).$$

The highest value obtained in these eight linear optimization problems is 0.804. We thus have $\text{poss}_7^* = 0.804$.

References

[1] A. Aregui and T. Denœux. Consonant belief function induced by a confidence set of pignistic probabilities. In K. Mellouli, editor, *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, pages 344–355, Hammamet, Tunisia, October/November 2007. Springer-Verlag.

[2] A. Aregui and T. Denœux. Constructing predictive belief functions from continuous sample data using confidence bands. In G. De Cooman, J. Vejnarová, and M. Zaffalon, editors, *Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '07)*, pages 11–20, Prague, Czech Republic, July 2007.

[3] A. Aregui and T. Denœux. Fusion of one-class classifiers in the belief function framework. In *Proceedings of the 10th Int. Conf. on Information Fusion*, Quebec, Canada, July 2007.

[4] B. C. Arnold and R. M. Shavelle. Joint confidence sets for the mean and variance of a normal distribution. *The American Statistician*, 52(2):133–140, 1998.

[5] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.

[6] J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2–3):123–150, 2005.

[7] F. Caron, B. Ristic, E. Duflos, and P. Vanheeghe. Least committed basic belief density induced by a multivariate Gaussian: Formulation with applications. *International Journal of Approximate Reasoning (in press)*, 2007. doi:10.1016/j.ijar.2006.10.003.

[8] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(2):167–196, 1994.

[9] F. Delmotte and P. Smets. Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model. *IEEE Transactions on Systems, Man and Cybernetics A*, 34(4):457–471, 2004.

[10] A. P. Dempster. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39(3):957–966, 1968.

[11] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

[12] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics A*, 30(2):131–150, 2000.

[13] T. Denœux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42(3):228–252, 2006.

[14] T. Denœux and M.-H. Masson. EVCLUS: Evidential clustering of proximity data. *IEEE Trans. on Systems, Man and Cybernetics B*, 34(1):95–109, 2004.

[15] T. Denœux and P. Smets. Classification using belief functions: the relationship between the case-based and model-based approaches. *IEEE Transactions on Systems, Man and Cybernetics B*, 36(6):1395–1406, 2006.

[16] D. Dubois, S. Moral, and H. Prade. *Belief change*, volume 3 of *Handbook of defeasible reasoning and uncertainty management systems*, chapter Belief change rules in ordinal and numerical uncertainty theories. Kluwer Academic Publishers, Boston, 1998.

[17] D. Dubois and H. Prade. On several representations of an uncertainty body of evidence. In M. M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, New-York, 1982.

[18] D. Dubois and H. Prade. Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. *Fuzzy sets and systems*, 10(1):15–20, 1983.

[19] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[20] D. Dubois, H. Prade, and S. A. Sandri. On possibility-probability transformations. In R. Lowen and M. Lowen, editors, *Fuzzy Logic: State of the Art*, pages 103–112. Kluwer Academic Publisher, 1993.

[21] D. Dubois, H. Prade, and P. Smets. New semantics for quantitative possibility theory. In S. Benferhat and P. Besnard, editors, *Proc. of the 6th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 2001)*, pages 410–421, Toulouse, France, 2001. Springer-Verlag.

[22] D. Dubois, H. Prade, and P. Smets. A definition of subjective possibility. *International Journal of Approximate Reasoning (In press)*, 2007. doi:10.1016/j.ijar.2007.01.005.

[23] E. J. Dudewicz and S. N. Mishra. *Modern Mathematical Statistics*. Wiley, 1988.

[24] Z. Elouedi, K. Mellouli, and P. Smets. Belief decision trees: Theoretical foundations. *International Journal of Approximate Reasoning*, 28:91–124, 2001.

[25] L. A. Goodman. On simultaneous confidence intervals for multinomial proportions. *Technometrics*, 7(2):247–254, 1965.

[26] I. Hacking. *Logic of Statistical Inference*. Cambridge University Press, Cambridge, 1965.

[27] M.-H. Masson and T. Denœux. Clustering interval-valued data using belief functions. *Pattern Recognition Letters*, 25(2):163–171, 2004.

[28] M.-H. Masson and T. Denœux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, 2006.

[29] D. Mercier, B. Quost, and T. Denœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion (In press)*, 2006. doi:10.1016/j.inffus.2006.08.001.

[30] S. Petit-Renaud and T. Denœux. Nonparametric regression analysis of uncertain and imprecise data using belief functions. *International Journal of Approximate Reasoning*, 35(1):1–28, 2004.

[31] C. P. Quesenberry and D. C. Hurst. Large sample simultaneous confidence intervals for multinomial proportions. *Technometrics*, 6(2):191–195, 1964.

[32] B. Quost, T. Denœux, and M.-H. Masson. Pairwise classifier combination using belief functions. *Pattern Recognition Letters*, 28(5):644–653, 2007.

[33] B. Ristic and P. Smets. Target classification approach based on the belief function theory. *IEEE Transactions on Aerospace and Electronic Systems*, 41(2):574 – 583, 2005.

[34] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[35] L. S. Shapley. A value for n-person games. In H. Kuhn and A. W. Tucker, editors, *Contributions to the theory of games*, volume 2, pages 307–317. Princeton University Press, Princeton, NJ, US, 1953.

[36] P. Smets. *Un modèle mathématico-statistique simulant le processus du diagnostic médical*. PhD thesis, Université Libre de Bruxelles, Brussels, Belgium, 1978. (in French).

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

[38] P. Smets. Belief induced by the partial knowledge of the probabilities. In D. H. *et al.*, editor, *Uncertainty in AI'94*, pages 523–530. Morgan Kaufmann, San Mateo, 1994.

[39] 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 Academic Publishers, Dordrecht, 1998.

[40] P. Smets. Practical uses of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 15 (UAI99)*, pages 612–621, Stockholm, Sweden, 1999.

[41] P. Smets. Belief functions on real numbers. *International Journal of Approximate Reasoning*, 40(3):181–223, 2005.

[42] P. Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38:133–147, 2005.

[43] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[44] S. Trabelsi, Z. Elouedi, and K. Mellouli. Pruning belief decision tree methods in averaging and conjunctive approaches. *International Journal of Approximate Reasoning*, 46(3):568–595, 2007.

[45] P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1):3–57, 1996.

[46] R. R. Yager. The entailment principle for Dempster-Shafer granules. *Int. J. of Intelligent Systems*, 1:247–262, 1986.

Figures

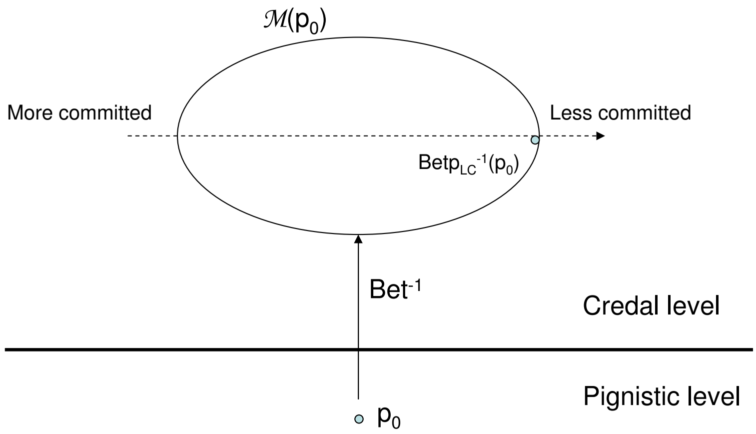

Figure 1: Definition of the q-LC isopignistic bba $m = Bet_{LC}^{-1}(p_0)$ associated to a pignistic probability function p_0.

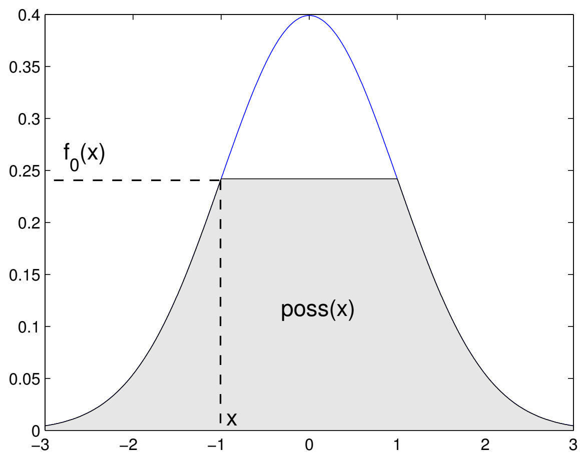

Figure 2: Calculation of the q-LC possibility distribution poss induced by a probability density function f_0.

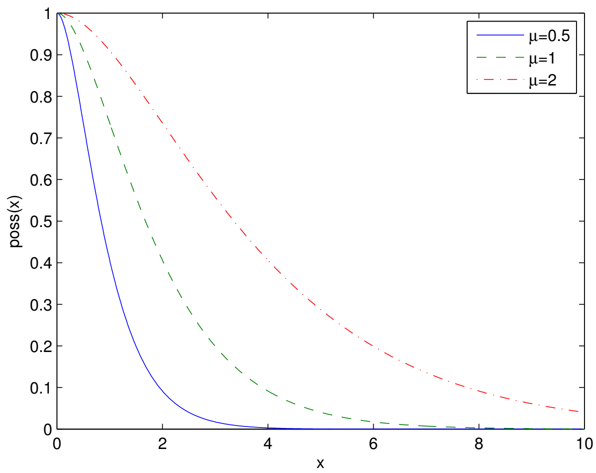

Figure 3: q-LC possibility distribution induced by an exponential probability density $\mathcal{E}(\mu)$ for three different values of μ.

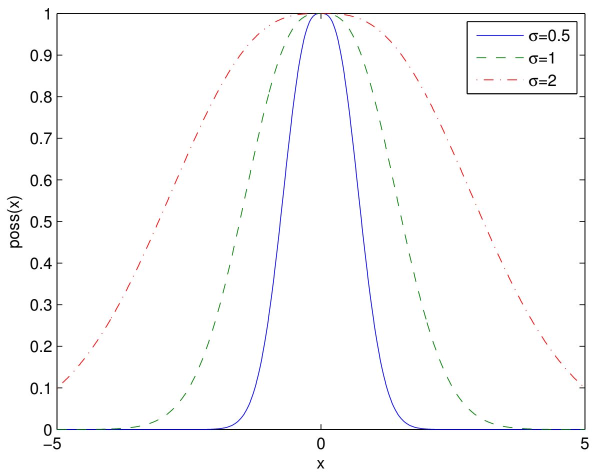

Figure 4: q-LC possibility distribution induced by a normal probability density $\mathcal{N}(\mu, \sigma^2)$ for $\mu = 0$ three different values of σ.

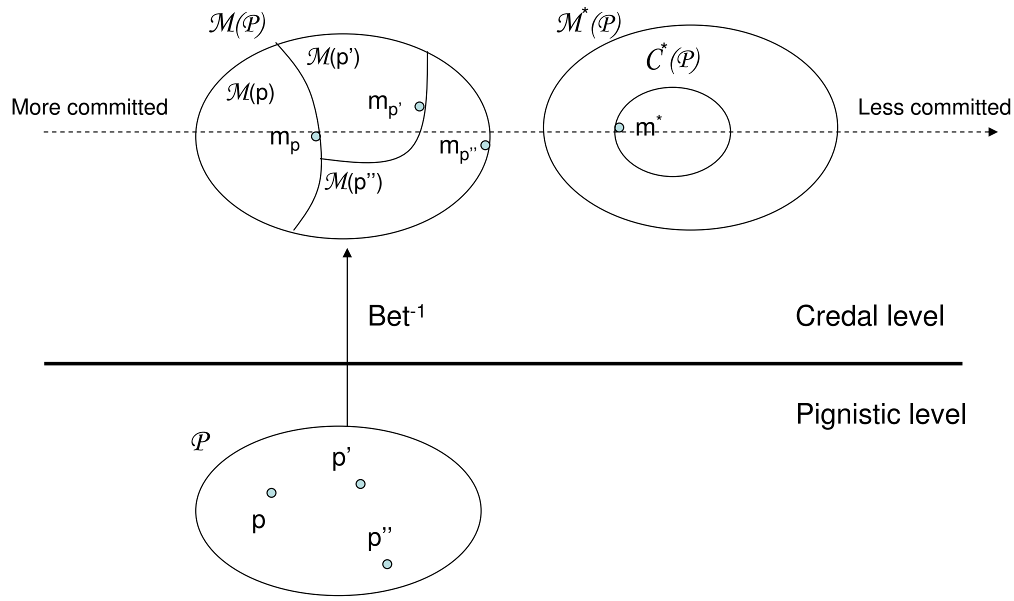

Figure 5: Definition of the q-most committed dominating (q-MCD) bba m^* associated to a set $\mathcal{P}$ of probability distribution. The set $\mathcal{M}(\mathcal{P})$ contains all bbas with pignistic probability functions in $\mathcal{P}$. The set $\mathcal{M}^*(\mathcal{P})$ contains all bbas dominating (i.e., less committed than) all bbas in $\mathcal{M}(\mathcal{P})$. The q-MCD bba m^* is the q-most committed consonant bba in $\mathcal{M}^*(\mathcal{P})$.

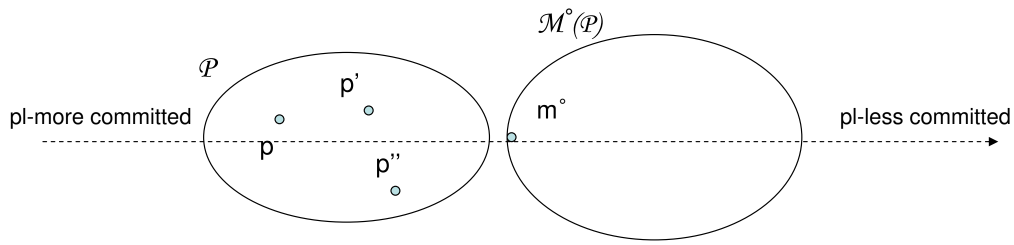

Figure 6: Illustration of the approach introduced in [13]: m° is the pl-most committed bba in the set $\mathcal{M}^\circ(\mathcal{P})$ of bbas that are less committed than all probability measures in $\mathcal{P}$. This approach does not distinguish between the pignistic and credal levels (compare with Figure 5).

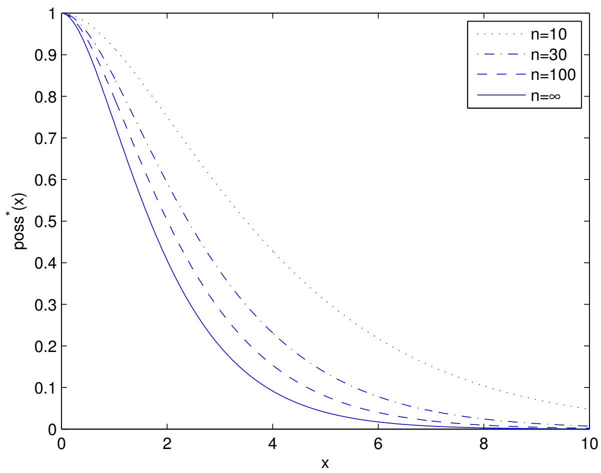

Figure 7: Plot of $\text{poss}^*(x)$ for the exponential distribution with $\overline{x} = 1$, $\alpha = 0.1$, and $n = 10$, 30, 100 and ∞.

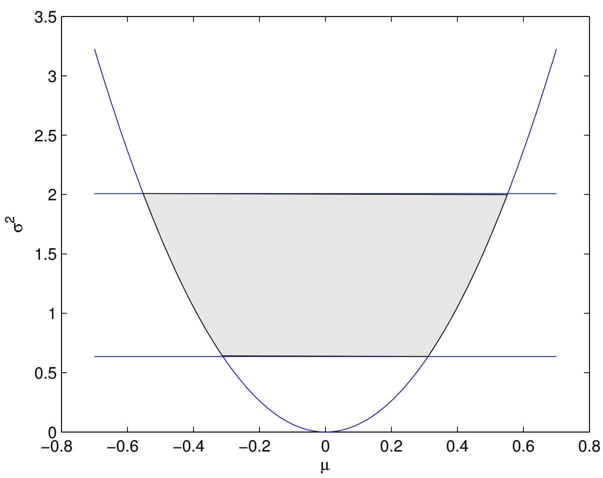

Figure 8: Shape of Mood's exact region: the Mood Exact Region for $\alpha = 0.1$, $\alpha_1 = \alpha_2$ and $n = 25$. Without loss of generality, $\overline{x} = 0$ and $s^2 = 1$.

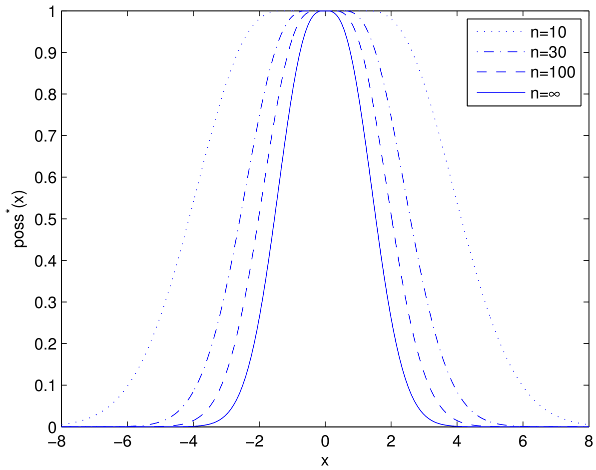

Figure 9: Plot of $\text{poss}^*(x)$ for the normal distribution with $\overline{x} = 0$, $s^2 = 1$, $\alpha = 0.1$, $\alpha_1 = \alpha_2$, and $n = 10$, 30, 100 and ∞.

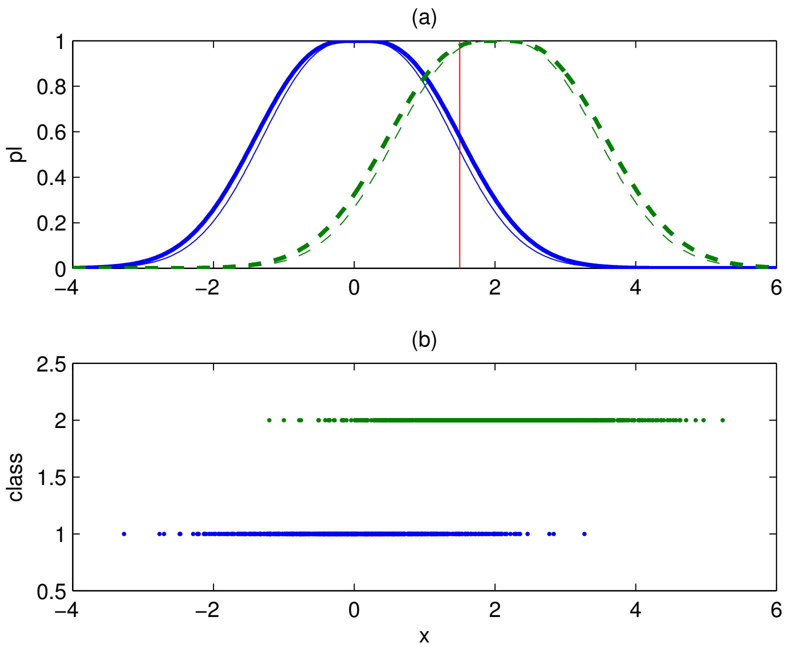

Figure 10: (a): Plot of $pl(x|\omega_1)$ (solid lines) and $pl(x|\omega_2)$ (broken lines) computed using the LC and MCD methods (thin and thick lines, respectively), as functions of x. (b) Dot plots of training data from each class in learning set $\mathcal{L}_x$.

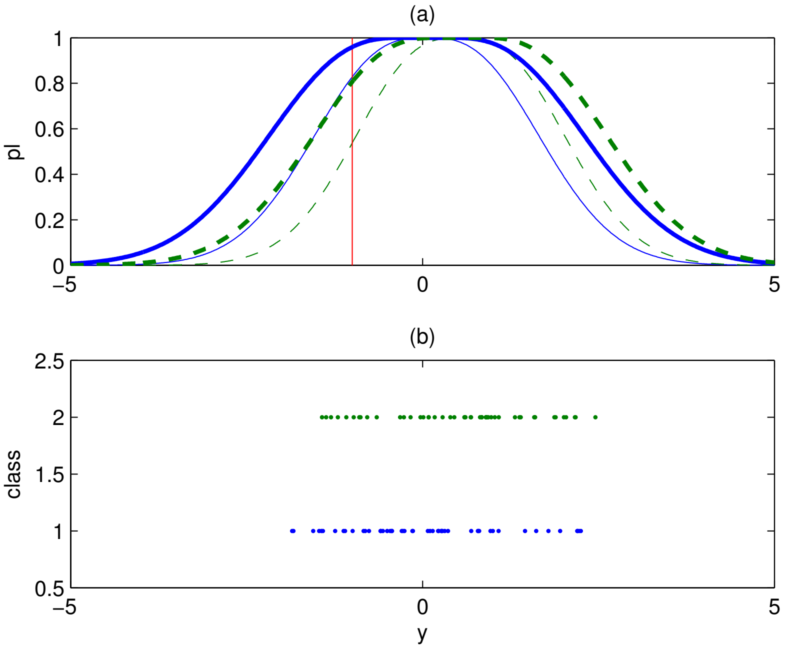

Figure 11: (a): Plot of $pl(y|\omega_1)$ (solid lines) and $pl(y|\omega_2)$ (broken lines) computed using the LC and MCD methods (thin and thick lines, respectively), as functions of y. (b) Dot plots of training data from each class in learning set $\mathcal{L}_y$.

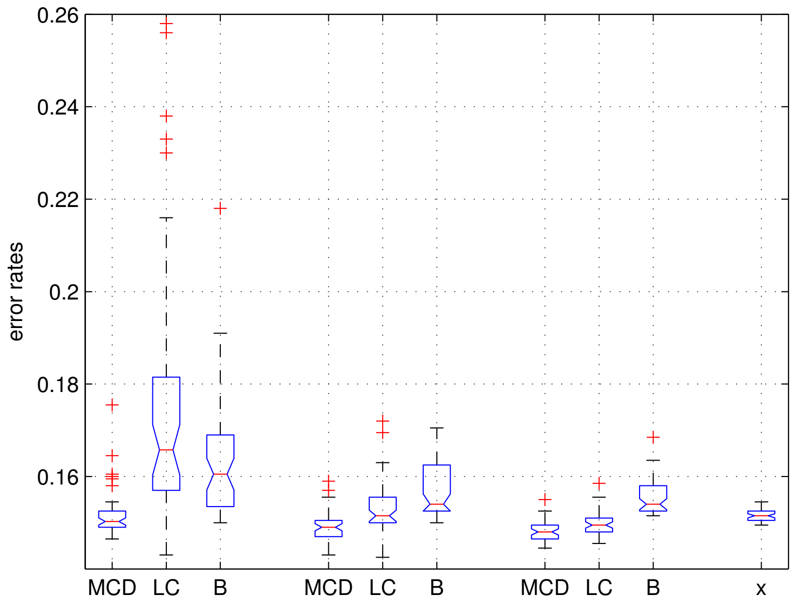

Figure 12: Box plots of error rates for the LC, MCD and Bayes (B) methods for different sizes of training set $\mathcal{L}_y$ ($n_y = 10, 50, 100$ from left to right), and results with sensor S_x alone (rightmost box).

Tables

Table 1: Pignistic probabilities and corresponding q-LC isopignistic possibility distributions of Example 4.

x	$p(x)$	$p'(x)$	$p''(x)$	$\mathrm{poss}(x)$	$\mathrm{poss}'(x)$	$\mathrm{poss}''(x)$	$\mathrm{poss}^*(x)$
ξ_1	0.7	0.6	0.65	1	1	1	1
ξ_2	0.2	0.25	0.1	0.5	0.65	0.3	0.65
ξ_3	0.1	0.15	0.25	0.3	0.45	0.6	0.6

Table 2: Calculation of q_{max} for the data of Example 4. In that case, q_{max} is a commonality function, and the corresponding bba m_{max} is strictly q-more committed than m^*, as we have $q_{max}(A) < q^*(A)$ for $A = \{\xi_2, \xi_3\}$ and for $A = \mathcal{X}$.

A	$\{\xi_1\}$	$\{\xi_2\}$	$\{\xi_1, \xi_2\}$	$\{\xi_3\}$	$\{\xi_1, \xi_3\}$	$\{\xi_2, \xi_3\}$	$\mathcal{X}$
$q(A)$	1	0.5	0.5	0.3	0.3	0.3	0.3
$q'(A)$	1	0.65	0.65	0.45	0.45	0.45	0.45
$q''(A)$	1	0.3	0.3	0.6	0.6	0.3	0.3
$q^*(A)$	1	0.65	0.65	0.6	0.6	0.6	0.6
$q_{max}(A)$	1	0.65	0.65	0.6	0.6	0.45	0.45
$m_{max}(A)$	0.2	0	0.2	0	0.15	0	0.45

Table 3: Goodman simultaneous confidence intervals for the data of Example 5, at confidence level $1 - \alpha = 0.90$.

Mode ξ_k	1	2	3	4	5	6	7	8	9	10
n_k	5	11	19	30	58	67	92	118	173	297
n_k/n	0.0057	0.013	0.022	0.035	0.067	0.077	0.106	0.136	0.199	0.341
p_k^-	0.002	0.006	0.012	0.022	0.048	0.057	0.082	0.109	0.166	0.301
p_k^+	0.017	0.027	0.039	0.054	0.092	0.104	0.136	0.168	0.236	0.384

Table 4: Possibility distributions computed for the failure mode data of Example 6: q-LC possibility distribution computed from the sample frequencies ($\widehat{\text{poss}}$), q-MCD possibility distribution computed from the multinomial confidence intervals shown in Table 3 (poss^*), and approximation computed using (28) ($\widetilde{\text{poss}}$).

Mode ξ_k	1	2	3	4	5	6	7	8	9	10
n_k	5	11	19	30	58	67	92	118	173	297
n_k/n	0.0057	0.013	0.022	0.035	0.067	0.077	0.106	0.136	0.199	0.341
$\widehat{\text{poss}}_k$	0.058	0.120	0.193	0.282	0.475	0.526	0.641	0.731	0.858	1
poss_k^*	0.171	0.258	0.353	0.462	0.688	0.735	0.804	0.867	0.935	1
$\widetilde{\text{poss}}_k^*$	0.171	0.258	0.353	0.462	0.688	0.747	0.875	0.973	1	1