Constructing Predictive Belief Functions from Continuous Sample Data Using Confidence Bands

Astride Aregui
Suez Environnement and
UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France

Thierry Denœux
UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France

Abstract

We consider the problem of quantifying our belief in future values of a random variable X with unknown distribution P_X, based on the observation of a random sample from the same distribution. The adopted uncertainty representation framework is the Transferable Belief Model, a subjectivist interpretation of belief function theory. In a previous paper, the concept of predictive belief function at a given confidence level was introduced, and it was shown how to build such a function when X is discrete. This work is extended here to the case where X is a continuous random variable, based on step or continuous confidence bands.

Keywords. Dempster-Shafer Theory, Evidence Theory, Transferable Belief Model, p-box, distribution band.

1 Introduction

In the past few years, belief function theory has been developed as a tool for data fusion, but also for the management of uncertainty and various aspects of data mining or decision making. Different interpretations of this theory have been proposed [19]. In this paper, we shall adopt the Transferable Belief Model (TBM) interpretation [21], in which a belief function is considered as representing weighted opinions of an agent regarding some question of interest. This model provides a flexible framework even when the available information (data or expert knowledge) is poor. However, it is not always clear how to construct belief functions for a given problem.

In this paper, we consider the special case where the variable X of interest is defined from the result of a random experiment. It is thus a random variable, with unknown probability distribution P_X. The available information is assumed to consist in past observations collected from n independent repetitions of the same experiment, forming an independent random sample from P_X. Based on this information, we would like to express our beliefs regarding future values to be generated from P_X.

As the probability distribution of X is unknown, the available information is incomplete and the precision of the obtained belief function should depend on the number of observations. In [5], a formalization of this problem was suggested, using the concept of *predictive belief function* (PBF). A PBF was defined as a belief function less committed than P_X with some user-defined probability, and converging in probability towards P_X as the size of the sample tends to infinity. Practical methods for building belief functions were presented for the case where the domain $\mathcal{X}$ of X is discrete, based on multinomial confidence regions.

In this article, the above approach is extended to the case where X is a *continuous random variable.* The extension is based on confidence bands, which play a role similar to that of multinomial confidence regions in the discrete case. When a confidence band is defined by step upper and lower bounding functions, it is known to be equivalent to a belief function on the real line with a finite number of focal intervals. We first show that this belief function is a predictive belief function as defined in [5]. We then consider the generalization to continuous confidence band. In that case, the corresponding belief function is continuous, and we derive the expression of its basic belief density.

The paper is organized as follows. In Section 2, the reader is first reminded with the principles of belief functions theory and of the definition of predictive belief functions as introduced in [5]. The construction of a discrete predictive belief function from a step confidence band is then exposed in Section 3, and the construction of a continuous predictive belief function with a basic belief density from a continous confidence band is described in Section 4. Section 5 concludes the paper.

2 Background on Belief Functions

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 [16], and then address the case of a continuous domain [20]. The concept of predictive belief function as introduced in [5] is then recalled.

2.1 Belief Functions on a Finite Frame

2.1.1 Definition of a Basic Belief Assignment

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:

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

Each mass $m(A)$ is interpreted as the part of the agent's belief allocated to the hypothesis that X takes some value in A [16, 21]. The mass $m(\mathcal{X})$ is often regarded as representing a degree of ignorance.

2.1.2 Belief Updating

A fundamental mechanism for belief updating in the TBM is the unnormalized *Dempster's rule of conditioning*, which is defined as follows [21]. Assume that the agent's beliefs about X are represented by a bba m, and the agent learns that the true value of X lies in $B \subseteq \mathcal{X}$. Then, m is transformed into the conditional bba $m[B]$ defined as:

$$m[B](A) = \sum_{C: C \cap B = A} m(C). \quad (2)$$

Upon learning that the truth lies in B, each mass of belief given to C is thus *transferred* to $C \cap B$, hence the term "*Transferable* Belief Model". Equivalent representations of a bba m include the belief, plausibility and commonality functions [16] defined as follows.

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad (3)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad (4)$$

and

$$q(A) = \sum_{B \supseteq A} m(B), \quad (5)$$

for all $A \subseteq \mathcal{X}$. In the TBM, $bel(A)$ represents the agent's total degree of belief in A. The plausibility $pl(A) = bel[A](A)$ may be interpreted as the maximal degree of belief that could be given to A after acquiring new information. Similarly, we observe that $q(A) = m[A](A)$. The commonality of A is thus the mass of belief that remains attached to A (i.e., the degree of ignorance) after conditioning by A.

2.1.3 Decision Making

The TBM is a two-level model in which belief representation and updating take place at a first level termed *credal level*, whereas decision making takes place at a second level called *pignistic level* [21]. To make decisions, any bba m such that $m(\emptyset) < 1$ is mapped into a pignistic probability function *Betp* defined by

$$Betp(x) = \sum_{A \subseteq \mathcal{X}, A \neq \emptyset} \frac{m(A)}{1 - m(\emptyset)} \frac{1_A(x)}{|A|}, \quad \forall x \in \mathcal{X}, \quad (6)$$

where 1_A denotes the indicator function of A. A decision can then be made, based on *Betp* and on a loss function, just as is done in Bayesian Probability Theory.

2.2 Belief Functions on Real Numbers

Let us now assume that variable X takes values in $\mathcal{X} = \mathbb{R}$. The above formalism can then be extended in at least two different ways.

2.2.1 Discrete Bba on $\mathbb{R}$

In the simplest approach, a bba is defined as above, with the constraint that the set $\mathcal{F}(m) = \{A_1, \ldots, A_n\}$ of focal elements is finite. This will be referred to as a *discrete* bba. Typically, focal elements are chosen among intervals or, more generally, Borel sets [23, 6, 24, 13]. Denoting $m_i = m(A_i)$, with $\sum_{i=1}^n m_i = 1$, and assuming $A_i \neq \emptyset$ for all i, Equations (3)-(5) become:

$$bel(A) = \sum_{A_i \subseteq A} m_i, \quad (7)$$

$$pl(A) = \sum_{A_i \cap A \neq \emptyset} m_i, \quad (8)$$

and

$$q(A) = \sum_{A_i \supseteq A} m_i, \quad (9)$$

for all $A \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel sigma-algebra on $\mathbb{R}$.

Equation (6) can be replaced by

$$Betp(x) = \sum_{i=1}^{n} m_i \frac{1_{A_i}(x)}{|A_i|}, \quad \forall x \in \mathbb{R}, \tag{10}$$

where $|A_i|$ now denotes the Lebesgue measure of A_i and we assume that $0 < |A_i| < \infty$ for all i. Equation (10) defines a probability density function [13]. In particular, if the A_is are bounded intervals, $Betp$ is a finite mixture of continuous uniform distributions.

2.2.2 Basic Belief Density

A more complex generalization of the finite case is obtained by replacing the concept of bba by that of basic belief density (bbd) [4, 17, 20]. 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 can be defined in the same way as in the finite case, replacing finite sums by integrals. The following definitions hold:

$$bel(A) = \iint_{[x,y] \subseteq A} m([x, y])\, dx\, dy, \tag{12}$$

$$pl(A) = \iint_{[x,y] \cap A \neq \emptyset} m([x, y])\, dx\, dy, \tag{13}$$

$$q(A) = \iint_{[x,y] \supseteq A} m([x, y])\, dx\, dy, \tag{14}$$

for all $A \in \mathcal{B}(\mathbb{R})$. In particular, when $A = [x, y]$,

$$bel([x, y]) = \int_x^y \int_u^y m([u, v]) dv du, \tag{15}$$

$$pl([x, y]) = \int_{-\infty}^{y} \int_{\max(x,u)}^{+\infty} m([u, v]) dv du, \tag{16}$$

$$q([x, y]) = \int_{-\infty}^{x} \int_{y}^{+\infty} m([u, v]) dv du, \tag{17}$$

for all $x \leq y$. The domains of these integrals may be represented as in Figure 1, where each point in the triangle corresponds to an interval with upper and lower bounds indicated on the horizontal and vertical axes, respectively.

Conversely, m may be recovered from bel or q as:

$$m([x, y]) = -\frac{\partial^2 bel([x, y])}{\partial x \partial y} = -\frac{\partial^2 q([x, y])}{\partial x \partial y}, \tag{18}$$

provided these derivatives exist.

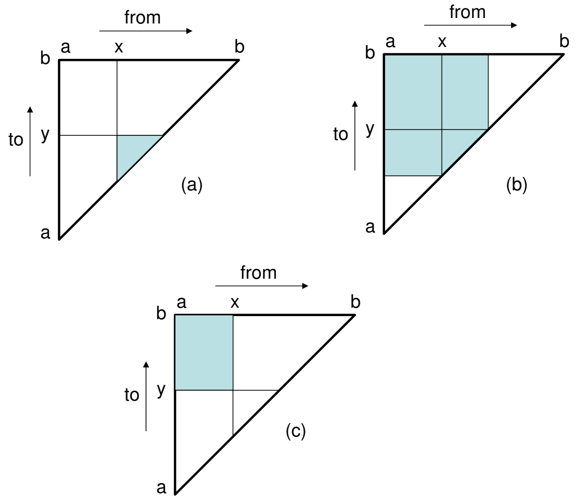

Figure 1: The belief, plausibility and commonality functions are defined as integrals of the bbd with support $[a, b]$ on the shaded area of triangles (a), (b) and (c), respectively.

The pignistic probability density becomes [20]:

$$Betp(x) = \lim_{\epsilon \to 0} \int_{-\infty}^{x} \int_{x+\epsilon}^{+\infty} \frac{m([u, v])}{v - u} dv du. \tag{19}$$

2.3 Predictive Belief Functions

In this section, we summarize the concept of predictive belief function introduced in [5]. Assume X is a random variable with unknown probability distribution P_X, and we have observed a realization $\mathbf{x} = (x_1, \ldots, x_n)$ of an independent and identically distributed (iid) random sample $\mathbf{X} = (X_1, \ldots, X_n)$ with parent distribution P_X. Based on this information, we would like to quantify our beliefs about the next value of X. As a toy example, consider the case where X denotes the color of a ball taken from an urn containing balls of different colors. Having observed the colors of n balls randomly taken from the urn with replacement, we would like to quantify our belief regarding the color of the next ball.

Let $bel(\cdot; \mathbf{X})$ denote a belief function on $\mathcal{X}$ constructed using $\mathbf{X}$. This is a function taking values from a sigma algebra $\mathcal{A}$ into $[0, 1]$. Typically, $\mathcal{A} = 2^{\mathcal{X}}$ if $\mathcal{X}$ is finite, and $\mathcal{A} = \mathcal{B}(\mathbb{R})$ if $\mathcal{X} = \mathbb{R}$ (only these two cases will be considered in this paper). In [5], we postulated that such a belief function should satisfy the following two requirements:

$$\forall A \in \mathcal{A}, \quad bel(A; \mathbf{X}) \xrightarrow{P} P_X(A), \text{ as } n \to \infty, \tag{20}$$

where $\xrightarrow{P}$ denotes convergence in probability, and

$$P\left\{bel(A; \mathbf{X}) \leq P_X(A), \forall A \in \mathcal{A}\right\} \geq 1 - \alpha, \tag{21}$$

where $\alpha \in (0,1)$.

Requirement (20) means that $bel(\cdot;\mathbf{X})$ should become closer to P_X as the sample size tends to infinity.

For finite n, $bel(\cdot;\mathbf{X})$ should be less informative than P_X, hence the condition $bel(\cdot;\mathbf{X}) \leq P_X$. However, this condition cannot be satisfied for all realizations of the random sample[1], hence requirement (21), which states that it should be satisfied asymptotically for at least a fraction $1-\alpha$ of the samples.

A belief function $bel(\cdot;\mathbf{X})$ satisfying requirements (20) and (21) is called a *predictive belief function at confidence level* $1-\alpha$. Methods for constructing such belief functions in the case where random variable X is discrete were described in [5], based on multinomial confidence regions.

The construction of predictive belief functions in the continuous case ($\mathcal{X} = \mathbb{R}$) is the main topic of this paper. It will be addressed in the following two sections.

3 Discrete Predictive Belief Functions on $\mathbb{R}$

In this section, the construction of a discrete predictive belief function on $\mathbb{R}$ from a step confidence band is addressed. Basic definitions related to confidence bands are first recalled in Section 3.1, and the construction of Kolmogorov confidence bands is exposed in Section 3.2. In Section 3.3, we show that the discrete belief function with interval focal sets equivalent to a Kolmogorov confidence band is a predictive belief function. The random set interpretation of a p-box is finally recalled in Section 3.4, as a way to introduce the continuous generalization presented in the next section.

3.1 Confidence Bands: Definitions

Let us assume that we have a random variable X with cumulative distribution function (cdf) F_X. In some cases, F_X is not precisely known, but we can specify a lower bounding function $\underline{F} : \mathbb{R} \to [0,1]$ and an upper bounding function $\overline{F} : \mathbb{R} \to [0,1]$ such that $\underline{F}(x) \leq F_X(x) \leq \overline{F}(x)$ for all $x \in \mathbb{R}$. The convex set of probabilities compatible with these constraints

$$\Gamma_X(\underline{F},\overline{F}) = \{P | \forall x \in \mathbb{R}, \underline{F}(x) \leq P((-\infty,x]) \leq \overline{F}(x)\}$$

is called a *distribution band* [11].

In the special case where $\underline{F}$ and $\overline{F}$ are step functions, then $\Gamma_X(\underline{F},\overline{F})$ is called a *probability box*[2], or p-box for short [6]. A continuous distribution bound can always be enclosed in a p-box. The smallest discrete approximation is always obtained by choosing the lower and upper bounding step functions to be right and left-continous, respectively [6]. From now on, only p-boxes possessing this property will be considered.

Suppose now that the available information about F_X takes the form of an iid random sample $\mathbf{X} = (X_1,\ldots,X_n)$ with parent distribution F_X. Let $\underline{F}(\cdot;\mathbf{X})$ and $\overline{F}(\cdot;\mathbf{X})$ be two functions computed from $\mathbf{X}$ and such that $\underline{F}(\cdot;\mathbf{X}) \leq \overline{F}(\cdot;\mathbf{X})$. The distribution band $\Gamma_X(\underline{F}(\cdot;\mathbf{X}),\overline{F}(\cdot;\mathbf{X}))$ is called a *confidence band at level* $\alpha \in (0,1)$ [12, page 334] iff

$$P\left\{\underline{F}(x;\mathbf{X}) \leq F_X(x) \leq \overline{F}(x;\mathbf{X}),\ \forall x \in \mathbb{R}\right\} = 1-\alpha,$$

or, equivalently:

$$P\left\{P_X \in \Gamma_X\left(\underline{F}(\cdot;\mathbf{X}),\overline{F}(\cdot;\mathbf{X})\right)\right\} = 1-\alpha.$$

Note that, in the above equalities, F_X and P_X are fixed unknown functions, whereas $\underline{F}(\cdot;\mathbf{X})$ and $\overline{F}(\cdot;\mathbf{X})$ depend on random sample $\mathbf{X}$.

3.2 Kolmogorov Confidence Bands

Let us assume that X is a continuous random variable. The simplest way to obtain a confidence band for F_X is to use Kolmogorov's statistic

$$D_n = \sup_x |S_n(x;\mathbf{X}) - F_X(x)|,$$

where $S_n(\cdot;\mathbf{X})$ is the sample distribution function defined by

$$S_n(x;\mathbf{X}) = \begin{cases} 0, & x < X_{(1)} \\ k/n, & X_{(k)} \leq x < X_{(k+1)} \\ 1, & X_{(n)} \leq x, \end{cases} \tag{22}$$

for all $x \in \mathbb{R}$, where $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the observations sorted in increasing order.

The distribution of D_n does not depend on F_X. It was computed for fixed n by Kolmogorov [10], who also computed the asymptotic distribution of D_n. Let $d_{n,\alpha}$ denote the critical value of D_n defined as $P(D_n > d_{n,\alpha}) = \alpha$. Thus,

$$P\{\ S_n(x;\mathbf{X}) - d_{n,\alpha} \leq F_X(x) \leq S_n(x;\mathbf{X}) + d_{n,\alpha}, \forall x \in \mathbb{R}\ \} = 1-\alpha, \tag{23}$$

which implies that $S_n \pm d_{n,\alpha}$ defines a confidence bound at level $1-\alpha$ [9, page 481]. This band may

[1] Indeed, such a requirement would lead to the vacuous belief function.

[2] Ferson *et al.* [6] actually used the term "p-box" as a synonym to "distribution band". However, following Kriegler and Held [11], we prefer to reserve the term "p-box" for the important case where the bounding functions are step functions.

be narrowed by using the inequalities $0 \leq F_X(x) \leq 1$ for all x. Hence, we have:

$$\underline{F}(x;\mathbf{X}) = \max(0, S_n(x;\mathbf{X}) - d_{n,\alpha}), \quad (24)$$
$$\overline{F}(x;\mathbf{X}) = \min(1, S_n(x;\mathbf{X}) + d_{n,\alpha}). \quad (25)$$

If the support of X is bounded and known to be included in $[b, B]$, then the above bounds can be further narrowed.

Note that $S_n(\cdot|\mathbf{X})$ as defined by (22) and, consequently, both $\underline{F}(\cdot;\mathbf{X})$ and $\overline{F}(\cdot;\mathbf{X})$ are right-continuous step functions. However, $\overline{F}(\cdot;\mathbf{X})$ can be replaced by the left-continuous function $\overline{F}'(\cdot;\mathbf{X})$ taking the same values everywhere except at sample points, defined as $\overline{F}'(x;\mathbf{X}) = \lim_{h\to x^-} \overline{F}(h;\mathbf{X})$. The pair $(\underline{F}, \overline{F}')$ still defines a confidence band at level $1-\alpha$, i.e.,

$$P\left\{P_X \in \Gamma_X(\underline{F}, \overline{F}')\right\} = 1 - \alpha. \quad (26)$$

Example 1. *The data reported in [14] consists in the operational lives (in hours) of 20 bearings. These are 2398, 2812, 3113, 3212, 3523, 5236, 6215, 6278, 7725, 8604, 9003, 9350, 9460, 11584, 11825, 12628, 12888, 13431, 14266, 17809. Here, the variable of interest, denoted X (the lifetime of a bearing), has a lower bound $b = 0$ and no upper bound ($B = \infty$). Figure 2 shows the sample cdf of this data, together with the lower and upper bounding functions defining the Kolmogorov confidence band at level $1 - \alpha = 0.95$.*

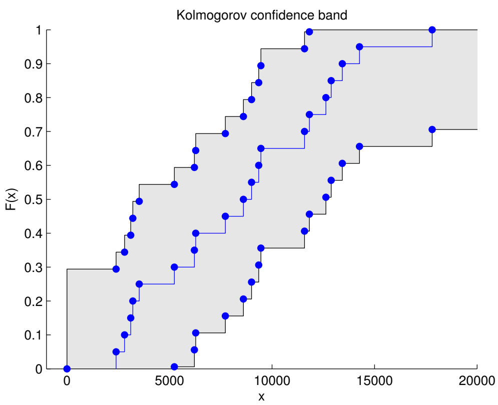

Figure 2: Sample cdf S_n and Kolmogorov confidence band at level $1 - \alpha = 0.95$ for the bearings data.

3.3 Predictive Belief Function Induced by a Kolmogorov Confidence Band

The above method for constructing a confidence band yields a pair of lower and upper step functions, i.e., a p-box. The relationship between p-boxes and belief functions has been studied by several authors [23, 6, 22]. Recently, the exact correspondance between p-boxes with bounded support and discrete belief functions was proved by Kriegler and Held [11], who also proposed an algorithm for the rigorous construction of a discrete mass function m on $\mathbb{R}$ equivalent to a p-box.

The principle of this construction is illustrated in Figure 3. The lower and upper bounding functions are assumed to be right and left continuous, respectively. Each rectangle A_i in this figure corresponds to a focal interval $[a_i, b_i)$, with mass $m([a_i, b_i)) = d_i - c_i$.

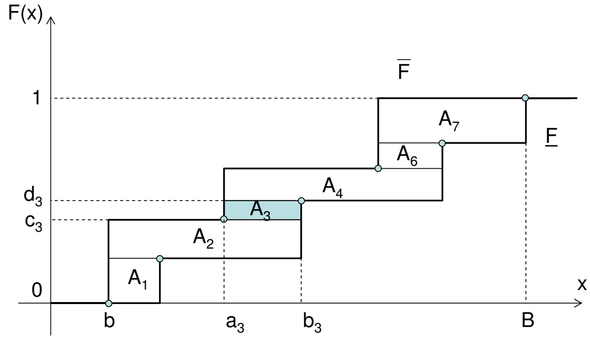

Figure 3: Principle of the construction of a basic belief assignment m from a p-box $(\underline{F}, \overline{F})$. Each rectangle A_i in the area between the lower and upper bounding functions corresponds a focal interval $[a_i, b_i)$ of m, with mass $d_i - c_i$.

Let $\Gamma_X(bel)$ denote the set of probability measures compatible with bel, the belief function induced by m, i.e.,

$$\Gamma_X(bel) = \{P | bel(A) \leq P(A), \forall A \in \mathcal{B}(\mathbb{R})\}.$$

Kriegler and Held [11] proved that $(\underline{F}, \overline{F})$ and bel are two equivalent representations of a unique family of probabilities, i.e.,

$$\Gamma_X(bel) = \Gamma_X(\underline{F}, \overline{F}). \quad (27)$$

If bel and pl denote the corresponding belief and plausibility functions, and if $\underline{P}$ and $\overline{P}$ denote the lower and upper envelopes of $\Gamma_X(\underline{F}, \overline{F})$, we have $bel = \underline{P}$ and $pl = \overline{P}$. In particular, $bel((-\infty, x]) = \underline{F}(x)$ and $pl((-\infty, x]) = \overline{F}(x)$ for all $x \in \mathbb{R}$.

Note that, although Kriegler and Held only considered the case of p-boxes with bounded support, their algorithm and result may be applied directly to the case of p-boxes with unbounded support.

Let us now consider the case where $\underline{F}$ and $\overline{F}$ are the lower and upper bounding functions of Kolmogorov confidence band at level $1 - \alpha$, as defined by (24)-(25). Let $bel(\cdot;\mathbf{X})$ denote the belief function on $\mathbb{R}$ constructed from p-box $(\underline{F}, \overline{F}')$ using Kriegler and Held's algorithm. The following proposition holds.

Proposition 1. *$bel(\cdot; \mathbf{X})$ is a predictive belief function at level $1-\alpha$.*

Sketch of proof. First, requirement (21) is obviously satisfied as a direct consequence of (26) and (27): since $\Gamma_X(bel(\cdot;\mathbf{X})) = \Gamma_X(\underline{F}, \overline{F}')$, we have

$$P\left\{bel(A;\mathbf{X}) \leq P_X(A), \forall A \in \mathcal{A}\right\} = P\left\{P_X \in \Gamma_X(bel(\cdot;\mathbf{X}))\right\} = 1-\alpha.$$

Moreover, given that $\underline{F}(x) \xrightarrow{P} F_X(x)$ and $\overline{F}'(x) \xrightarrow{P} F_X(x)$ for all $x \in \mathbb{R}$, it can easily be shown that $bel(A;\mathbf{X}) \xrightarrow{P} P_X(A)$ for all interval A. Lastly, for any $B = \bigcup_{i\in I} A_i$ where $(A_i)_{i\in I}$ with $I \in \mathbb{N}$ is a countable family of intervals, we have

$$bel(B;\mathbf{X}) = \sum_{i\in I} bel(A_i;\mathbf{X}) \xrightarrow{P} \sum_{i\in I} P_X(A_i) = P_X(B),$$

which proves that requirement (20) is satisfied, and completes the proof. □

Example 2. *To illustrate the construction of a predictive belief function from a Kolmogorov confidence band, let us consider again the data of Example 1. Based on this data, we would like to express our beliefs regarding the lifetime X of a new bearing taken randomly from the same population. For commodity of representation, let us adopt the reasonable assumption that X has an upper bound, which will arbitrarily be set to 30000, so that the support of X is assumed to be $[0, 30000]$.*

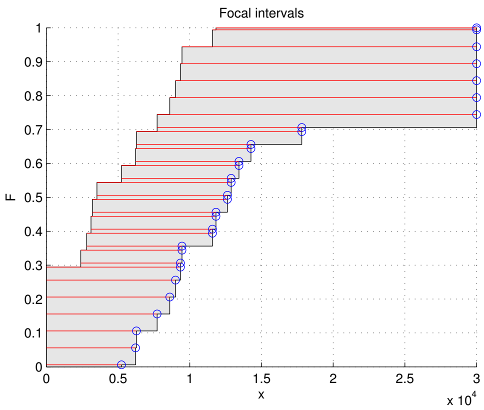

Figure 4: Focals intervals of the PBF constructed from the Kolmogorov confidence band at level $1-\alpha = 0.95$ (bearings data). The height of each segment representing a focal interval is equal to the cumulated mass.

The focal intervals of the corresponding PBF $bel(\cdot;\mathbf{X})$ are displayed in Figure 4. Figures 5 and 6 are examples of graphical displays that reveal different aspects of the information contained in the belief function $bel(\cdot;\mathbf{X})$. Figure 5 shows the plausibility profile function $x \rightarrow pl(\{x\};\mathbf{X})$ and the pignistic probability density function Betp computed from (6), which are two left-continuous real-valued step functions with simple interpretation. Figure 6 shows grey level representations of $bel([x,y];\mathbf{X})$, $pl([x,y];\mathbf{X})$ and $q([x,y];\mathbf{X})$ as two-dimensional functions of (x,y).

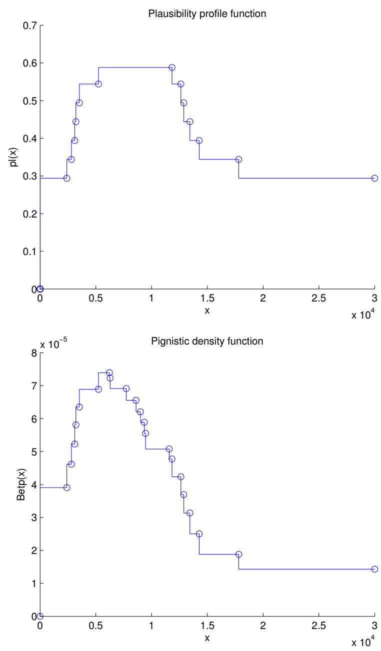

Figure 5: Plausibility profile function (up) and pignistic probability density function (down) of the discrete PBF constructed from the Kolmogorov confidence band (Bearings data).

3.4 Random Set Interpretation

The bba m associated to a p-box $(\underline{F}, \overline{F})$ may also be shown to correspond formally to a random set [1]. Let $\underline{F}^{-1}$ and $\overline{F}^{-1}$ be the pseudo-inverses of $\underline{F}$ and $\overline{F}$, defined, respectively, as:

$$\underline{F}^{-1}(\alpha) = \inf\{x \in \mathbb{R}, \underline{F}(x) \geq \alpha\},$$

$$\overline{F}^{-1}(\alpha) = \inf\{x \in \mathbb{R}, \overline{F}(x) \geq \alpha\},$$

for all $\alpha \in [0,1]$. Let us consider the mapping ρ from $[0,1]$ to the set of real intervals, such that $\rho(\alpha) = (\underline{F}^{-1}(\alpha), \overline{F}^{-1}(\alpha)]$, and let us consider the uniform probability distribution P_U on $[0,1]$. Then ρ is a random set, and it is formally equivalent to m. Let $\mathcal{F} = \{(\underline{F}^{-1}(\alpha), \overline{F}^{-1}(\alpha)], \alpha \in [0,1]\}$. For all $A \in \mathcal{F}$, we have

$$m(A) = P_U(\rho^{-1}(A)).$$

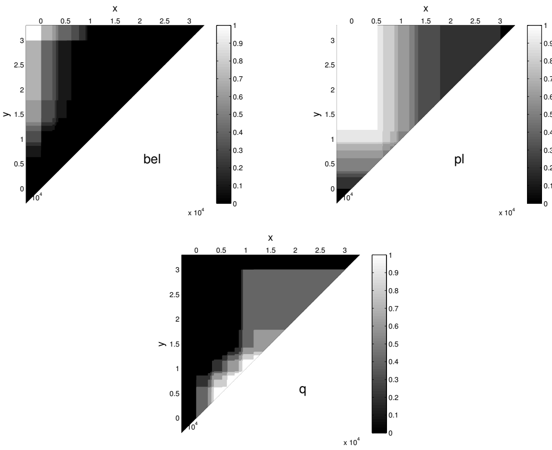

Figure 6: Contour plots of functions $bel[\mathbf{X}]([x,y])$, $pl[\mathbf{X}]([x,y])$ and $q[\mathbf{X}]([x,y])$ constructed from Kolmogorov's confidence band (Bearings data).

Note that the uniform probability distribution on $[0,1]$ and the mapping ρ are only considered here as mathematical constructs. In the TBM, only belief functions have an interpretation, and an underlying multi-valued mapping is not assumed. However, the random set point of view will guide us in the following section to propose a generalization of the above results in the case of continuous distribution bands.

4 Continuous Predictive Belief Functions on $\mathbb{R}$

Kolmogorov's confidence bands have the advantage of being exact and non parametric. However, they have a constant vertical width, which makes them unnecessarily broad in the tails. As a result, the equivalent belief functions may be excessively imprecise. Narrower confidence bands can be computed using parametric methods, but they are defined by continuous bounding functions. The usual approach to continuous distribution bands is to approximate them using a p-box [6]. Here, we show that this approximation can be avoided, and a continuous predictive belief function on $\mathbb{R}$ can be constructed from a continuous confidence band, thus providing an extension to the results presented in the previous section. Parametric confidence bands are first briefly reviewed in the following section.

4.1 Parametric Confidence Bands

Methods for the construction of continuous confidence bands as described above were proposed by several authors, including Kanofsky and Srinivasan [8] and Cheng and Iles [3]. In the sequel, Cheng and Iles' method, which will be used later to demonstrate the main findings of this paper, will briefly be recalled.

Let us assume that X is a continuous random variable with cdf $F_X(x,\theta)$, where θ is vector of r unknown parameters. Cheng and Iles' approach consists in determining lower and upper bounds of the cdf when θ varies in a confidence region R. This confidence region is built from the statistics

$$Q(\theta) = (\widehat{\theta} - \theta)^T I(\theta)(\widehat{\theta} - \theta),$$

where $\widehat{\theta}$ is the maximum likelihood estimate of θ, and $I(\theta)$ is the Fisher information matrix. It is known that $Q(\theta)$ is asymptotically a chi-squared variable with r degrees of freedom. In [3], Cheng and Iles apply their method in the case of a general location-scale parametric model of the form:

$$F_X(x) = G\left(\frac{x-\mu}{\sigma}\right),$$

where G is a fixed distribution function, and μ and σ are the unknown location and scale parameters. In that case the Fisher information matrix is of the form

$$I(\mu,\sigma) = \frac{n}{\sigma^2}\begin{pmatrix} k_0 & -k_1 \\ -k_1 & k_2 \end{pmatrix},$$

where k_0, k_1 and k_2 are constants independent of μ and σ. The bounds of the confidance band then have the following expressions:

$$\overline{F}(x) = G(\xi + h), \tag{28}$$

$$\underline{F}(x) = G(\xi - h), \tag{29}$$

where $\xi = (x - \widehat{\mu})/\widehat{\sigma}$, $\widehat{\mu}$ and $\widehat{\sigma}$ are the maximum likelihood estimates of μ and σ, and

$$h = \sqrt{\frac{\gamma}{n\,k_0}\left(1 + \frac{(k_0\xi + k_1)^2}{k_0k_2 - k_1^2}\right)}. \tag{30}$$

Coefficient γ is the value for which $P(Q(\mu,\sigma) \leq \gamma) = 1-\alpha$. It can be approximated by the chi-squared quantile $\chi^2_2(\alpha)$. Cheng and Iles [3] demonstrate the application of these formula for the cases of the normal, lognormal, extreme-value (log-Weibull) and Weibull distributions. In the case of the normal distribution, $k_0 = 1$, $k_1 = 0$, and $k_2 = 2$.

4.2 PBF Induced by a Continuous Confidence Band

Let $(\underline{F}, \overline{F})$ be a continuous distribution band for some continuous random variable X, and assume that the lower and upper bounding functions $\underline{F}$ and $\overline{F}$ are strictly increasing. Consider the mapping ρ from $[0,1]$ to the set of real intervals, such that $\rho(\alpha) =$

$[\underline{F}^{-1}(\alpha), \overline{F}^{-1}(\alpha)]$, where $\underline{F}^{-1}$ and $\overline{F}^{-1}$ are the inverses of $\underline{F}$ and $\overline{F}$, respectively. If the $[0,1]$ interval is endowed with a uniform probability distribution, then mapping ρ defines a random set, which corresponds to a continuous belief function bel on $\mathbb{R}$ as described in Section 2.2.2.

This belief function is such that $bel([x,y]) = \underline{P}([x,y])$ for all $x \le y$, $\underline{P}$ being the lower envelope of the distribution band. In particular, we have $bel((-\infty, x]) = \underline{F}(x)$ and $pl((-\infty, x]) = \overline{F}(x)$, for all $x \in \mathbb{R}$. As we are working within the TBM, this random set is for us a purely mathematical construct, and we would like to express bel directly through its bbd $m([x,y])$, $x \le y$. This can be achieved using (18). The following proposition holds.

Proposition 2. *The bbd associated to a continuous distribution band* $(\underline{F}, \overline{F})$ *is defined by*

$$m([x,y]) = -\frac{\partial^2 bel([x,y])}{\partial x \partial y},$$

with:

$$\begin{aligned} \frac{\partial^2 bel([x,y])}{\partial x \partial y} &= -\overline{f}(x)\underline{f}(y)\delta(\underline{F}(y) - \overline{F}(x)), & (31)\\ &= -\overline{f}(x)\delta(y - \underline{F}^{-1} \circ \overline{F}(x)), & (32)\\ &= -\underline{f}(y)\delta(x - \overline{F}^{-1} \circ \underline{F}(y)), & (33) \end{aligned}$$

where $\underline{f}$ and $\overline{f}$ are the first derivatives of $\underline{F}$ and $\overline{F}$, respectively, and δ is the Dirac delta function.

Proof. We have

$$\begin{aligned} bel([x,y]) &= \underline{P}([x,y]) & (34)\\ &= \max(0, \underline{F}(y) - \overline{F}(x)) & (35)\\ &= (\underline{F}(y) - \overline{F}(x))H(\underline{F}(y) - \overline{F}(x)), & (36) \end{aligned}$$

where H is the Heaviside function. By differentiating with respect to x and y, we get:

$$\begin{aligned} \frac{\partial^2 bel([x,y])}{\partial x \partial y} = -\overline{f}(x)\big(&\delta(\underline{F}(y) - \overline{F}(x))\underline{f}(y)\\ &+ \underline{f}(y)\delta(\underline{F}(y) - \overline{F}(x))\\ &+(\underline{F}(y) - \overline{F}(x))\delta'(\underline{F}(y) - \overline{F}(x))\underline{f}(y)\big). \quad (37) \end{aligned}$$

Now, from the property of the delta function: $x\delta'(x) = -\delta(x)$, we have:

$$(\underline{F}(y) - \overline{F}(x))\delta'(\underline{F}(y) - \overline{F}(x)) = -\delta(\underline{F}(y) - \overline{F}(x)).$$

Hence, (37) is equivalent to (31).

In order to prove that (32) and (33) can be deduced from (31), the following property of the delta function can be used. For all function g,

$$\delta(g(x)) = \sum_i \frac{\delta(x - x_i)}{|g'(x_i)|},$$

where the x_i are the roots of g. For fixed x, $\underline{F}(y) - \overline{F}(x)$ is a function of y with a unique root $\underline{F}^{-1} \circ \overline{F}(x)$. Hence,

$$\overline{f}(x)\underline{f}(y)\delta(\underline{F}(y) - \overline{F}(x)) = \overline{f}(x)\underline{f}(y)\frac{\delta(y - \underline{F}^{-1} \circ \overline{F}(x))}{\underline{f}(\underline{F}^{-1} \circ \overline{F}(x))} \quad (38)$$

The left-hand side of (38) is equal to 0 if $y \neq \underline{F}^{-1} \circ \overline{F}(x)$, and $\overline{f}(x)\delta(y - \underline{F}^{-1} \circ \overline{F}(x))$ otherwise. Consequently,

$$\overline{f}(x)\underline{f}(y)\delta(\underline{F}(y) - \overline{F}(x)) = \overline{f}(x)\delta(y - \underline{F}^{-1} \circ \overline{F}(x)).$$

Equation (33) can be deduced from (31) in a similar way, by fixing y and treating $\underline{F}(y) - \overline{F}(x)$ as a function of x. □

It can be checked that (35) may be recovered from $m([x,y])$ using (15). Similarly, the expressions of $pl([x,y])$ and $q([x,y])$ can be obtained from $m([x,y])$ using (16) and (17). The following proposition holds.

Proposition 3. *Let m be the bbd associated to a continuous distribution band $(\underline{F}, \overline{F})$. The plausibility and the commonality of any real interval $[x,y]$ are given by:*

$$pl([x,y]) = \overline{F}(y) - \underline{F}(x), \quad (39)$$

$$q([x,y]) = \max(0, \overline{F}(x) - \underline{F}(y)). \quad (40)$$

Proof. Let us prove (40). We have

$$\begin{aligned} q([x,y]) &= \int_{-\infty}^{x} \int_{y}^{+\infty} m([u,v]) dv du\\ &= \int_{-\infty}^{x} \overline{f}(u) I(u) du, \end{aligned}$$

with

$$I(u) = \int_{y}^{+\infty} \delta\left(v - \underline{F}^{-1} \circ \overline{F}(u)\right) dv.$$

Now, $I(u) = 1$ if $\underline{F}^{-1} \circ \overline{F}(u) \ge y$, i.e., if $u \ge \overline{F}^{-1} \circ \underline{F}(y)$, and 0 otherwise. Hence $q([x,y]) = 0$ if $\overline{F}^{-1} \circ \underline{F}(y) \ge x$, i.e., if $\underline{F}(y) \ge \overline{F}(x)$; otherwise,

$$q([x,y]) = \int_{\overline{F}^{-1} \circ \underline{F}(y)}^{x} \overline{f}(u) du = \overline{F}(x) - \underline{F}(y).$$

The proof of (39) is similar. □

Finally, the expression of the pignistic probability density associated to bbd m is given by the following proposition.

Proposition 4. *Let m be the bbd associated to a continuous distribution band $(\underline{F}, \overline{F})$. The associated pignistic probability density $Betp$ is given by*

$$Betp(x) = \int_{\overline{F}^{-1} \circ \underline{F}(x)}^{x} \frac{\overline{f}(u)}{\underline{F}^{-1} \circ \overline{F}(u) - u} du.$$

Proof. From (19), we get

$$Betp(x) = \lim_{\epsilon \to 0} \int_{-\infty}^{x} J(u) du,$$

with

$$\begin{aligned} J(u) &= \overline{f}(u) \int_{x+\epsilon}^{+\infty} \frac{\delta\left(v - \underline{F}^{-1} \circ \overline{F}(u)\right)}{v-u} dv \\ &= \begin{cases} \frac{\overline{f}(u)}{\underline{F}^{-1} \circ \overline{F}(u) - u} & \text{if } \underline{F}^{-1} \circ \overline{F}(u) \geq x + \epsilon \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The condition $\underline{F}^{-1} \circ \overline{F}(u) \geq x + \epsilon$ can be expressed as $u \geq \overline{F}^{-1} \circ \underline{F}(x+\epsilon)$, hence

$$\begin{aligned} Betp(x) &= \lim_{\epsilon \to 0} \int_{\overline{F}^{-1} \circ \underline{F}(x+\epsilon)}^{x} \frac{\overline{f}(u)}{\underline{F}^{-1} \circ \overline{F}(u) - u} du \\ &= \int_{\overline{F}^{-1} \circ \underline{F}(x)}^{x} \frac{\overline{f}(u)}{\underline{F}^{-1} \circ \overline{F}(u) - u} du. \end{aligned}$$

□

The above results are valid for any continuous distribution band $(\underline{F}, \overline{F})$. When $(\underline{F}, \overline{F})$ is a confidence band at level $1-\alpha$, then it is easy to see, using the same line of reasoning as in Section 3.3, that the corresponing belief function is a predictive belief function at level $1-\alpha$.

Example 3. *This method for computing a continuous predictive belief function was applied to the bearings data of examples 1 and 2 As in [3], we assumed these data have a lognormal distribution. Figure 7 shows the 95 % confidence band and the estimated cdf. The plausibility profile function* $x \to pl(\{x\}; \mathbf{X})$ *is shown in Figure 8, and contour plots of* $bel([x,y]; \mathbf{X})$, $pl([x,y]; \mathbf{X})$ *and* $q([x,y]; \mathbf{X})$ *are shown in Figure 9. These figures should be compared to Figures 2, 5 and 6, respectively.*

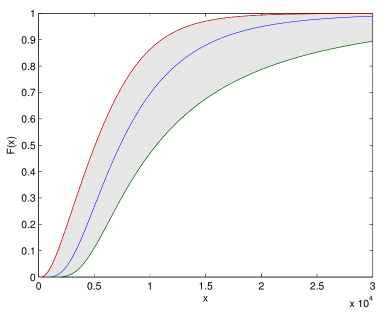

Figure 7: Continuous confidence band and cumulative density function estimated through Cheng and Iles' algorithm.

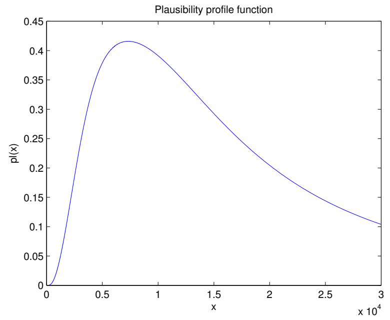

Figure 8: Plausibility profile function obtained from the continuous confidence band of Figure 7.

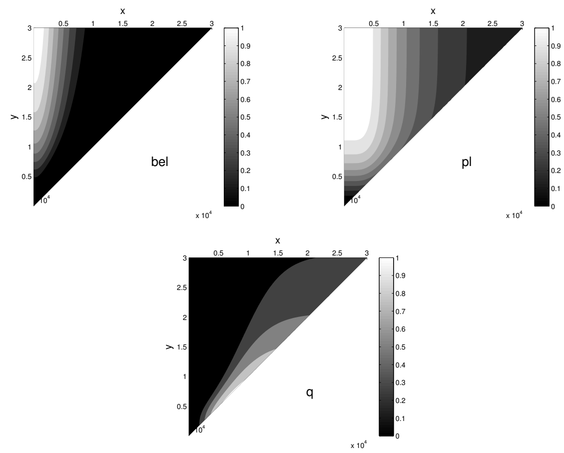

Figure 9: Contour plots of functions $bel([x,y]; \mathbf{X})$, $pl([x,y]; \mathbf{X})$ and $q([x,y]; \mathbf{X})$ constructed from Cheng and Iles' confidence band.

5 Conclusion

We have addressed the problem of constructing predictive belief functions as defined in [5], in the case where the random variable X is continuous. We have shown that such belief functions can be constructed from confidence bands, which play the same role as multinomial intervals in [5]. The methods yields a discrete BF with a finite number of interval focal sets when applied to a Kolomogov confidence band, and a basic belief density as studied in [20] when applied to a continuous parametric confidence band. These belief functions are interpreted as quantifying our belief in a future realization of X, based on a realization of a random sample from the same distribution.

An application of these results to novelty detection is described in [2]. Assume that we have defined a novelty measure T using, e.g., one-class support vector machines [15] or kernel principal component analysis [7]. Based on observations $T_1, \ldots, T_n$ of T for data recorded while the system under study was in the normal state ω_0, we may compute a predictive belief function on T, *given that the system is in the normal state*. Using the General Bayesian Theorem [18] with some assumptions, it is then possible to build a belief function on $\Omega = \{\omega_0, \overline{\omega}_0\}$ (where $\overline{\omega}_0$ denotes the hypothesis that the system is not in the normal state), given T. This belief function may be combined with other information or used for decision making.

References

[1] D. A. Alvarez. On the calculation of the bounds of probability of events using infinite random sets. *International Journal of Approximate Reasoning*, 43(3):241–267, 2006.

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

[3] R. C. H. Cheng and T. C. Iles. Confidence bands for cumulative distribution functions of continuous random variables. *Technometrics*, 25(1):77–86, 1983.

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

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

[6] S. Ferson, V. Kreinovitch, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015, Sandia National Laboratories, Albuquerque, NM, 2003.

[7] Heiko Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40:863–874, March 2007.

[8] P. Kanofsky and R. Srinivasan. An approach to the construction of parametric confidence bands on cumulative distribution functions. *Biometrica*, 59(3):623–631, 1972.

[9] M. Kendall and A. Stuart. *The advanced theory of statistics*, volume 2. Charles Griffin and Co Ltd, London, fourth edition, 1979.

[10] A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Istituto Italiano degli Attuari*, 4:83–91, 1933.

[11] E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning*, 39(2–3):185–209, 2005.

[12] E. L. Lehman. *Testing statistical hypotheses*. Springer-Verlag, New-York, 2nd edition, 1986.

[13] 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.

[14] R. E. Schafer and J. J. Angus. Estimation of Weibull quantiles with minimum error in the distribution function. *Technometrics*, 21:367–370, 1979.

[15] B. Schölkopf and A.J. Smola. *Learning with kernels*. MIT Press, 2002.

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

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

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

[19] Ph. Smets. What is Dempster-Shafer's model ? In R. R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer theory of evidence*, pages 5–34. Wiley, New-York, 1994.

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

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

[22] F. Tonon. Using random set theory to propagate epistemic uncertainty through a mechanical system. *Reliability Engineering & System Safety*, 85(1–3):169–181, 2004.

[23] R. R. Yager. Arithmetic and other operations on Dempster-Shafer structures. *Int. J. Man-Machines Studies*, 25:357–366, 1986.

[24] R. R. Yager. Cumulative distribution functions from Dempster-Shafer belief structures. *IEEE Trans. on Systems, Man and Cybernetics B*, 34(5):2080–2087, 2004.