Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model

Lev V. Utkin
Department of Computer Science, St. Petersburg Forest Technical Academy, Institutski per.5, 194021, St. Petersburg, Russia
E-mail: lvu@utkin.usr.etu.spb.ru

Abstract.
A belief function can be viewed as a generalized probability function and the belief and plausibility measures can be regarded as lower and upper bounds for the probability of an event. From this point of view, the common approach for computing belief and plausibility measures may be unreasonable in many real applications because there are cases where the probability function that governs the random process is not exactly known. In order to overcome this difficulty, Walley's imprecise Dirichlet model is used to extend the belief, plausibility and possibility measures. An interesting relationship between belief measures and sets of multinomial models is established. A combination rule taking into account reliability of sources of data is proposed. Various numerical examples illustrate the proposed extension.

Keywords: belief measure, Dirichlet distribution, multinomial model, possibility distribution, combination rule, Dempster-Shafer theory, imprecise probability theory

1. Introduction

Evidence theory or Dempster-Shafer theory [6,14] provide us with an appropriate mathematical model of uncertainty when information is not complete or when the result of each observation is not point-valued but set-valued, so that it is not possible to assume the existence of a unique probability measure. From this point of view, Dempster-Shafer theory can be interpreted as a generalization of probability theory where probabilities are assigned to sets as opposed to mutually exclusive singletons. There are three main functions in Dempster-Shafer theory: the basic probability assignment function, the belief function, and the plausibility function. As indicated in [9,13], generally, the term "basic probability assignment" does not refer to probability in the classical sense, but it is useful to interpret the basic probability assignment as a classical probability, and the framework of Dempster-Shafer theory can support this interpretation. Under this circumstance, Dempster-Shafer theory allows one to calculate upper (plausibility) and lower (belief) bounds for a probability of occurrence of an outcome, and many real applications use this interpretation.

At the same time, real applications meet some difficulties by using the classical probabilistic interpretation because it requires a large number of observations of events. If this number is small, inferences become too precise and incautious. Therefore, an approach to extend the belief and plausibility functions in order to take into account a lack of sufficient statistical data is proposed in the paper. This approach is based on Walley's imprecise Dirichlet model [19]. Since the possibility distribution function [7] can be regarded as a special case of belief and plausibility functions, then the same extension is applied to the possibility measure.

The paper is organized as follows. Basic formal definitions of Dempster-Shafer theory are given in Section 2. A relationship between a set of multinomial models and statistical data in the form of intervals is established in Section 3. The imprecise Dirichlet model is considered in Section 4. The extensions of belief and plausibility functions are proposed in Section 5. Some important properties of the obtained extensions are investigated in Section 6. The extended possibility distribution is described in Section 7. A combination rule is proposed in Section 8. Lower and upper probability distributions and expectations produced by extensions of belief and plausibility functions are considered in Section 9.

2. Belief functions

Let U be a universal set under interest, usually referred to in evidence theory as the *frame of discernment*. Suppose N observations were made of a parameter $u \in U$, each of which resulted in an imprecise (non-specific) measurement given by a set A of values. Let c_i denote the number of occurrences of the set $A_i \subseteq U$, and $\mathcal{P}(U)$ the set of all subsets of U (power set of U). A frequency function m can be defined, called *basic probability assignment*, such that [6,10,14]:

$$m : \mathcal{P}(U) \to [0,1],$$

$$m(\varnothing) = 1, \quad \sum_{A \in \mathcal{P}(U)} m(A) = 1.$$

Note that the domain of basic probability assignment, $\mathcal{P}(U)$, is different from the domain of a probability density function, which is U. According to [6], this function can be obtained as follows:

$$m(A_i) = c_i/N. \tag{1}$$

If $m(A_i) > 0$, i.e. A_i has occurred at least once, A_i is called a *focal element*.

According to [14], the *belief* $Bel(A)$ and *plausibility* $Pl(A)$ measures of an event $A \subseteq \Omega$ can be defined as

$$Bel(A) = \sum_{A_i : A_i \subseteq A} m(A_i),$$

$$Pl(A) = \sum_{A_i : A_i \cap A \neq \varnothing} m(A_i). \tag{2}$$

As pointed out in [9], a belief function can formally be defined as a function satisfying axioms which can be viewed as a weakening of the Kolmogorov axioms that characterize probability functions. Therefore, it seems reasonable to understand a belief function as a generalized probability function [6] and the belief $Bel(A)$ and plausibility $Pl(A)$ measures can be regarded as lower and upper bounds for the probability of A, i.e., $Bel(A) \le \Pr(A) \le Pl(A)$.

Example 1 *A system consists of four components ($U = \{1,2,3,4\}$). After a system failure,* 100 *experts ($N = 100$) try to define its cause.* 54 *experts ($c_1 = 54$) suppose that the first component is failed ($A_1 = \{1\}$),* 16 *experts ($c_2 = 16$) suppose that the first or the second component is failed ($A_2 = \{1,2\}$),* 30 *experts ($c_3 = 30$) suppose that the third component is failed ($A_3 = \{3\}$). Then there hold*

$$m(A_1) = 0.54, \; m(A_2) = 0.16, \; m(A_3) = 0.3.$$

Let us find lower and upper bounds for the probability that the first component is failed ($A = \{1\}$). By using (2), we obtain

$$Bel(A) = m(A_1) = 0.54,$$

$$Pl(A) = m(A_1) + m(A_2) = 0.7.$$

Now suppose that only 5 *experts ($N = 5$) supply the same judgments approximately in the same proportion. Three experts ($c_1 = 3$) suppose that the first component is failed ($A_1 = \{1\}$), one expert ($c_2 = 1$) supposes that the first or the second component is failed ($A_2 = \{1,2\}$), one expert ($c_3 = 1$) supposes that the third component is failed ($A_3 = \{3\}$). Then there hold*

$$m(A_1) = 0.6, \; m(A_2) = 0.2, \; m(A_3) = 0.2.$$

Now we can write

$$Bel(A) = m(A_1) = 0.6,$$

$$Pl(A) = m(A_1) + m(A_2) = 0.8.$$

As a result, we have quite different bounds. It is obvious that the number of experts is rather small to consider the obtained belief and plausibility functions as correct bounds for the probability of A.

Thus, definition (1) can be used when N is rather large. However, this condition may be violated in many real applications. There are cases where the probability function that governs the random process is not exactly known. This difficulty has been investigated by Smets [15]. If N

is small, inferences become too precise and incautious. In order to overcome this difficulty, I try to apply the imprecise Dirichlet model [19] and to extend belief and plausibility function to take into account a lack of sufficient statistical data.

3. A set of multinomial models

Suppose $U = \{u_1, ..., u_L\}$ and the set $A_i \subseteq U$ contains elements from U with indices J_i, i.e., $A_i = \{u_j : j \in J_i\}$. At that, the number of elements in J_i is l_i. Let n be the number of focal elements. Then $N = \sum_{i=1}^{n} c_i$. Let us calculate possible numbers of occurrences of every element of U. Associate the set A_i with an oblong box of size l_i with one open side and the set U with L small empty boxes of size 1. The i-th oblong box contains c_i balls which can move inside the box and we do not know location of balls in the i-th box because its open side is behind. Then we cover small boxes by the i-th oblong box and c_i balls enter in l_i small boxes with numbers from a set J_i. We do not know exact location of balls, but we know that they are in boxes with numbers from J_i. The same procedure is repeated n times. What can we say about possible numbers of balls in the small boxes now? It is obvious that there exist different combinations of numbers of balls except the case when $l_i = 1$ for $i = 1, ..., n$, i.e., all sets A_i consist of one element. Suppose that the number of the possible combinations is M. Denote the k-th possible vector of balls by $\mathbf{n}^{(k)} = (n_1^{(k)}, ..., n_L^{(k)})$, $k = 1, ..., M$. If to assume that the sets A_i occurred independently and a ball in the i-th small box has some unknown probability π_i, then every combination of balls in small boxes produces the *standard multinomial model*. M possible combinations of balls produce M equivalent standard multinomial models. The models are equivalent in the sense that we can not choose one of them as a more preferable case.

For every model, the probability of an arbitrary event $A \subseteq U$ depends on $\mathbf{n}^{(k)}$, that is, we can find $P(A|\mathbf{n}^{(k)})$. So far as all the models are equivalent, even by precise probabilities of all categories only lower and upper probabilities of A can be computed

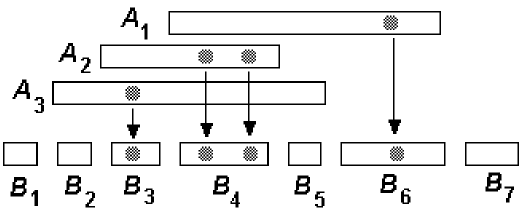

Figure 1. Illustration of balls and boxes

$$\underline{P}(A) = \min_{k=1,...,M} P(A|\mathbf{n}^{(k)}),$$

$$\overline{P}(A) = \max_{k=1,...,M} P(A|\mathbf{n}^{(k)}).$$

In particular, if all sets A_i consist of single elements, that is, all oblong boxes are of size 1, then $M = 1$ and

$$\underline{P}(A) = P(A|\mathbf{n}^{(k)}), \quad \overline{P}(A) = P(A|\mathbf{n}^{(k)}).$$

The following problem is to define $\mathbf{n}^{(k)}$ and $P(A|\mathbf{n}^{(k)})$. In the case of multinomial samples, the Dirichlet distribution is the traditional choice.

Remark 1 *It is worth noticing that the Dirichlet distribution should be regarded as one of the possible multinomial models that can be applied here.*

Remark 2 *If U is some interval of real numbers, then we can always transform this universal set to a set with finite numbers of elements. Let $\{\mathbf{i}\} = \{(i_1, ..., i_n, i_{n+1})\}$ be a set of all binary vectors consisting of $n+1$ components such that $i_j \in \{0, 1\}$. For every vector $\mathbf{i}$, we define the interval B_k ($k = 1, ..., 2^{n-1}$) as follows:*

$$B_k = \left(\bigcap_{j:i_j=1} A_j \right) \bigcap \left(\bigcap_{j:i_j=0} A_j^c \right), \ i_j \in \mathbf{i}.$$

Here $A_{n+1} = U$. As a result, we divide the set U into a set of non-intersecting intervals B_k such *that $B_1 \cup ... \cup B_L = U$, $L = 2^{n-1}$. Moreover, every interval A_i can be represented as the union of a finite number of intervals B_k.*

Example 2 *Suppose that there are three focal sets A_1, A_2, A_3 and $c_1 = c_3 = 1$, $c_2 = 2$. The corresponding oblong and small boxes are shown in Fig.1. One of the possible vectors of balls is $\mathbf{n}^{(k)} = (0,0,1,2,0,1)$. Here we unite intervals B_1 and B_7 because they never contain balls.*

4. Imprecise Dirichlet model

The *Dirichlet (s,α) prior distribution* for π, where $\alpha = (\alpha_1, ..., \alpha_L)$, has probability density function [5,22]

$$p(\pi) = C(s,\alpha) \cdot \prod_{j=1}^{L} \pi_j^{s\alpha_j - 1},$$

where $s > 0$, $0 < \alpha_j < 1$ for $j = 1, ..., L$, $\alpha \in S(1,L)$, and the proportionality constant C is determined by the fact that the integral of $p(\pi)$ over the simplex of possible values of π is 1 and

$$C(s,\alpha) = \Gamma(s) \left(\prod_{j=1}^{L} \Gamma(s\alpha_j) \right)^{-1}.$$

Here α_i is the mean of π_i under the Dirichlet prior and s determines the influence of the prior distribution on posterior probabilities. $\Gamma(\cdot)$ is the Gamma-function which satisfies $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(1) = 1$. $S(1,L)$ denotes the interior of the unit simplex.

Walley [19] pointed out several reasons for using a set of Dirichlet distributions to model prior ignorance about probabilities π:

1. Dirichlet prior distributions are mathematically tractable because they generate Dirichlet posterior distributions;

2. sets of Dirichlet distributions are very rich, because they produce the same inferences as their convex hulls and any prior distribution can be approximated by a finite mixture of Dirichlet distributions;

3. the most common Bayesian models for prior ignorance about probabilities π are Dirichlet distributions.

   The *imprecise Dirichlet model* is defined by Walley [19] as the set of all Dirichlet (s,α) distributions such that $\alpha \in S(1,L)$.

For the imprecise Dirichlet model, the *hyperparameter s* determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Walley [19] defined s as a number of observations needed to reduce the imprecision (difference between upper and lower probabilities) to half its initial value. Smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. At the same time, the value of s must not depend on L or a number of observations. The detailed discussion concerning the parameter s and the imprecise Dirichlet model can be found in [1,2,4,16,19]. The application of the model in reliability was studied in [4]. This model was also applied to the game theory for choosing a strategy in a two-player game by Quaeghebeur and de Cooman [12]. An approach to dealing with incomplete sets of multivariate categorical data by exploiting Walley's imprecise Dirichlet model was studied by Zaffalon [23].

Let A be any non-trivial subset of a sample space $\{\omega_1, ..., \omega_L\}$, and let $n(A)$ denote the observed number of occurrences of A in the N trials, $n(A) = \sum_{\omega_j \in A} n_j$. Then, according to [19], the predictive probability $P(A,s)$ under the Dirichlet posterior distribution is

$$P(A,s) = \frac{n(A) + s\alpha(A)}{N + s},$$

where $\alpha(A) = \sum_{\omega_j \in A} \alpha_j$.

By maximizing and minimizing α_j under restriction $\alpha \in S(1,L)$, we obtain the posterior upper and lower probabilities of A:

$$\underline{P}(A,s) = \frac{n(A)}{N + s}, \ \overline{P}(A,s) = \frac{n(A) + s}{N + s}.$$

By returning to the multinomial models considered in the example with boxes and balls and assuming that probabilities of balls are governed by the Dirichlet distribution, we can write the lower $\underline{P}(A,s)$ and upper $\overline{P}(A,s)$ probabilities of an event A, whose elements have indices from a set J, as follows:

$$\underline{P}(A,s) = \min_{k=1,\dots,M} \inf_{\alpha\in S(1,L)} \frac{n^{(k)}(A) + s\alpha(A)}{N+s},$$

$$\overline{P}(A,s) = \max_{k=1,\dots,M} \sup_{\alpha\in S(1,L)} \frac{n^{(k)}(A) + s\alpha(A)}{N+s},$$

where

$$\alpha(A) = \sum_{j\in J} \alpha_j, \quad n^{(k)}(A) = \sum_{j\in J} n_j^{(k)}.$$

5. Bounds for belief and plausibility functions

Now we have to find $n^{(k)}(A)$ and $\alpha(A)$. The lower and upper probabilities $\underline{P}(A,s)$ and $\overline{P}(A,s)$ can be rewritten as follows:

$$\underline{P}(A,s) = \frac{\min_{k=1,\dots,M} n^{(k)}(A) + s\cdot \inf_{\alpha\in S(1,L)} \alpha(A)}{N+s}, \tag{3}$$

$$\overline{P}(A,s) = \frac{\max_{k=1,\dots,M} n^{(k)}(A) + s\cdot \sup_{\alpha\in S(1,L)} \alpha(A)}{N+s}. \tag{4}$$

Note that $\inf_{\alpha\in S(1,L)} \alpha(A)$ is achieved at $\alpha(A) = 0$ and $\sup_{\alpha\in S(1,L)} \alpha(A)$ is achieved at $\alpha(A) = 1$ except a case when $A = U$. If $A = U$, then $\alpha(A) = 1$ for the minimum and maximum.

In order to find the minimum and maximum of $n^{(k)}(A)$ we consider three sets A_1, A_2, A_3 such that $A_1 \subseteq A$, $A_2 \cap A = \varnothing$, $A_3 \cap A \neq \varnothing$ and $A_3 \subsetneq A$. Numbers of their occurrences are c_1, c_2, c_3, respectively. It is obvious that all balls (c_1) corresponding to the set A_1 belong to the set A and $n^{(k)}(A)$ can not be less than c_1. On the other hand, all balls (c_2) corresponding to the set A_2 do not belong to A. This implies that $n^{(k)}(A)$ can not be greater than $N - c_2$. A part of balls corresponding to A_3 may belong to A, but it is not necessary. Therefore, $\min n^{(k)}(A) = c_1$ and $\max n^{(k)}(A) = N - c_2$. By extending this reasoning on an arbitrary set of A_i, we get the minimal, $L_1(A)$, and maximal, $L_2(A)$, values of $n^{(k)}(A)$:

$$L_1(A) = \min_{k=1,\dots,M} n^{(k)}(A) = \sum_{i:A_i\subseteq A} c_i,$$

$$L_2(A) = \max_{k=1,\dots,M} n^{(k)}(A) = N - \sum_{i:A_i\cap A=\varnothing} c_i = \sum_{i:A_i\cap A\neq\varnothing} c_i.$$

It can be seen from (1) that

$$\frac{L_1(A)}{N} = \sum_{i:A_i\subseteq A} \frac{c_i}{N} = \sum_{A_i:A_i\subseteq A} m(A_i),$$

$$\frac{L_2(A)}{N} = \sum_{i:A_i\cap A\neq\varnothing} \frac{c_i}{N} = \sum_{A_i:A_i\cap A\neq\varnothing} m(A_i).$$

This implies that $L_1(A)/N$ and $L_2(A)/N$ (see (2)) are none other than the belief $Bel(A)$ and plausibility $Pl(A)$ measures of the set $A \subseteq U$. Then there hold

$$\underline{P}(A,s) = \frac{N\cdot Bel(A)}{N+s},$$

$$\overline{P}(A,s) = \frac{N\cdot Pl(A) + s}{N+s}. \tag{5}$$

It is obvious from the above expressions that $\underline{P}(A,s) \leq Bel(A)$ and $\overline{P}(A,s) \geq Pl(A)$, i.e., $\underline{P}(A,s)$ and $\overline{P}(A,s)$ are extensions of the belief and plausibility measures of A taking into account the fact that N is restricted.

Let us rewrite equalities (5) as follows:

$$\underline{P}(A,s) = \varepsilon Bel(A),$$

$$\overline{P}(A,s) = 1 - \varepsilon(1 - Pl(A)),$$

where $\varepsilon = (1 + s/N)^{-1}$ and $\varepsilon \in [0,1]$.

From this representation, lower $\underline{P}(A,s)$ and upper $\overline{P}(A,s)$ probabilities can be regarded as some correction of belief and plausibility functions.

Example 3 *Let us return to Example 1. By assuming $s = 1$, we obtain new bounds for the probability of the event $A = \{1\}$ on the basis of judgments supplied by 5 experts:*

$$\underline{P}(A, 1) = \frac{5 \cdot 0.6}{5 + 1} = 0.5,$$

$$\overline{P}(A, 1) = \frac{5 \cdot 0.8 + 1}{5 + 1} = 0.833.$$

If $s = 2$, then $\underline{P}(A, 2) = 0.428$, $\overline{P}(A, 2) = 0.857$. By comparing these results with the bounds $[0.54, 0.7]$ obtained in Example 1, one can conclude that the extended bounds, for instance, $[0.5, 0.833]$ by $s = 1$, allow us to take into account the fact that N is rather small.

Now let us compute bounds for the probability of the event $A = \{1\}$ by $s = 1$ on the basis of judgments supplied by 100 experts:

$$\underline{P}(A, 1) = \frac{100 \cdot 0.54}{100 + 1} = 0.535,$$

$$\overline{P}(A, 1) = \frac{100 \cdot 0.7 + 1}{100 + 1} = 0.703.$$

It can be seen from the results that $\underline{P}(A, 1)$ and $\overline{P}(A, 1)$ are closely equal to $Bel(A)$ and $Pl(A)$.

6. Properties of lower and upper probabilities

Let us consider some special cases and properties of the proposed extended belief and plausibility measures.

1. For any $s \geq 0$ and $A \subseteq U$, there hold $\underline{P}(A, s) \leq Bel(A)$ and $\overline{P}(A, s) \geq Pl(A)$.
2. In the case of total ignorance about elements of U, i.e., before making any observations, it can be stated $c_i = N = 0$. Then for any $s \geq 0$, there hold

   $$\underline{P}(A, s) = Bel(A) = 0,$$

   $$\overline{P}(A, s) = Pl(A) = 1.$$

3. In the case $N \to \infty$, i.e., basic probability assignment coincides with probability in its classical sense, it can be stated for any s: $\underline{P}(A, s) = Bel(A)$, $Pl(A) = \overline{P}(A, s)$. Moreover, if $N \to \infty$ and all the sets A_i are singletons and $A_1 \cup ... \cup A_L = U$, then

   $$\underline{P}(A, s) = Bel(A) = Pl(A) = \overline{P}(A, s).$$

4. If $s = 0$, then $\underline{P}(A, 0) = Bel(A)$, $\overline{P}(A, 0) = Pl(A)$.
5. Suppose that $s_1 \leq s_2$, then there holds

   $$[\underline{P}(A, s_1), \overline{P}(A, s_1)] \subseteq [\underline{P}(A, s_2), \overline{P}(A, s_2)].$$

6. If A^c is the complement of the set A, then there hold

   $$\overline{P}(A, s) = 1 - \underline{P}(A^c, s),$$

   $$\underline{P}(A, s) = 1 - \overline{P}(A^c, s).$$

   The first equality follows from

   $$\underline{P}(A^c, s) = \varepsilon \cdot Bel(A^c) = \varepsilon \cdot (1 - Pl(A)) = 1 - \overline{P}(A, s).$$

   The second equality is similarly proved.

7. (*Avoiding sure loss* [18,20]) For any $s \geq 0$ and $A \subseteq U$, there holds $\underline{P}(A, s) \leq \overline{P}(A, s)$.
8. (*Coherence* [18,20]) For any $s \geq 0$ and $A, B \subseteq U$, there holds

   $$\underline{P}(A, s) + \underline{P}(B, s) \leq \underline{P}(A \cup B, s).$$

   Indeed, due to coherence of belief functions we can write

   $$\underline{P}(A, s) + \underline{P}(B, s) = \varepsilon Bel(A) + \varepsilon Bel(B) \leq \varepsilon Bel(A \cup B) = \underline{P}(A \cup B, s).$$

Let us consider two special cases when belief and plausibility functions may give improper conclusions. Suppose that there is one estimate A. According to Dempster-Shafer theory, the corresponding belief and plausibility functions are $Bel(A) = Pl(A) = 1$. This implies that we believe completely to one estimate. The conclusion is very unlikely. This contradiction can be avoided by using the imprecise Dirichlet model ($s > 0$). Indeed, if we take $s = 1$, then there hold

$$\underline{P}(A, s) = \frac{1}{1+s} = 0.5, \ \overline{P}(A, s) = 1.$$

It is worth noticing that if we have r identical estimates, then the belief and plausibility functions are the same $Bel(A) = Pl(A) = 1$. This implies that the belief and plausibility functions do not depend on the value r while $\underline{P}(A, s)$ and $\overline{P}(A, s)$ are

$$\underline{P}(A, s) = \frac{r}{r+s}, \ \overline{P}(A, s) = 1.$$

Suppose that there are 2 conflicting judgments A_1 and A_2 such that $A_1 \cap A_2 = \varnothing$. Then we have $Bel(A_i) = Pl(A_i) = 0.5$, $i = 1, 2$. This also is an incorrect result in some cases. Let $U = [0, 1]$, $A_1 = [0, 0.99]$, $A_2 = [0.991, 1]$. It is difficult to expect that both intervals have the same precise probabilities 0.5. At the same time, we can write

$$\underline{P}(A_i, s) = \frac{1}{2+s}, \ \overline{P}(A_i, s) = \frac{1+s}{2+s}.$$

If to take $s = 1$, then $\underline{P}(A_i, s) = 1/3$ and $\overline{P}(A_i, s) = 2/3$.

The above special cases show an advantage of extended belief and plausibility measures.

7. Extended possibility distribution

When the set of focal elements can be ordered in such a way that $A_i \subseteq A_{i+1}$, $i = 1, ..., n-1$, then this set is said to be *consonant*. The focal elements can thus be seen as the α-cuts of the fuzzy set F whose membership function, or possibility distribution is:

$$\pi_F(u) = Pl(\{u\}) = \sum_{A_i : u \in A_i} m(A_i) = \sum_{A_i : u \in A_i} \frac{c_i}{N}.$$

Then plausibility coincides with possibility and belief with necessity in possibility theory [7]. In this interpretation, a fuzzy set is a model of ambiguity and not of vagueness [17].

Let us extend the possibility distribution $\pi_F(u)$ by taking into account the restriction of N. By using (5), we can write

$$\overline{\pi}_F(u, s) = \overline{P}(\{u\}, s) = \frac{N \cdot Pl(\{u\}) + s}{N+s} = \frac{N \cdot \pi_F(u) + s}{N+s}.$$

It is obvious that $\overline{\pi}_F(u, s) \geq \pi_F(u)$ for all u and for any $s \geq 0$.

Example 4 *Suppose that a system consists of* 10 *components (*$U = \{u_1, ..., u_{10}\}$*) and a diagnostic system detects failures of the system components. The diagnostic system can indicate only some subsets of failed components due to imperfect fault diagnostics. A statistical report about the system components contains statistical data after* 100 *failures:* 26 *failures (*$c_4 = 26$*) of components from the set* $A_4 = \{u_4, ..., u_{10}\}$*,* 22 *failures (*$c_3 = 22$*) of components from the set* $A_3 = \{u_5, ..., u_9\}$*,* 27 *failures (*$c_2 = 27$*) of components from the set* $A_2 = \{u_6, ..., u_8\}$*,* 25 *failures (*$c_1 = 25$*) of components from the set* $A_1 = \{u_7\}$*. The possibility distribution of failed components* $\pi_F^*(u)$ *is shown in Fig.2 (curve 1).*

Now suppose that the statistical report contains statistical data after 5 *failures:* 1 *failure (*$c_4 = 2$*) of components from* A_4*,* 1 *failure (*$c_2 = 1$*) of components from* A_2*,* 2 *failures (*$c_1 = 2$*) of components from the set* A_1*. The possibility distribution* $\pi_F(u)$ *obtained through the plausibility function (*$s = 0$*) is shown in Fig.2 (curve 2). It can be seen from the picture that this distribution differs from the first distribution. Possibility distributions 3 and 4 are corrections of* $\pi_F(u)$ *by* $s = 1$ *and* $s = 4$*, respectively. Note that* $\overline{\pi}_F(u, 1) \ngeq \pi_F^*(u)$ *and* $\overline{\pi}_F(u, 4) \geq \pi_F^*(u)$ *for all* $u \in U$*. This implies that* $s = 1$ *is insufficient to make a cautious decision about failed components.*

8. Combination rule

Combination rules are some types of aggregation methods for data obtained from multiple sources (bodies of evidence). These sources provide different assessments for the same frame of discernment and Dempster-Shafer theory is based on the assumption that the sources are independent [13,8]. A large literature has developed over

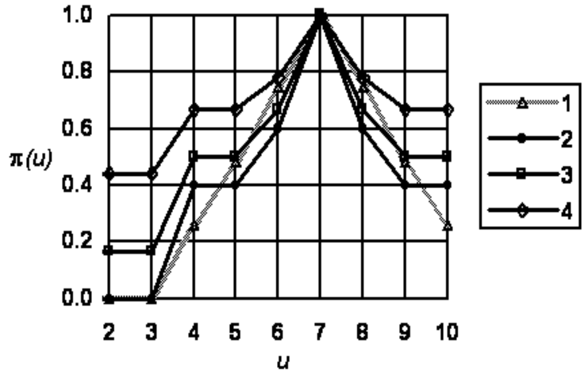

Figure 2. Possibility distribution functions by various s and N

the last years on the aggregation of evidence. In particular, Dempster's rule is extensively used in the theory to combine and update belief functions.

Suppose that there are two sources of data. The first source supplies N_1 observations of a parameter u and $c_i^{(1)}$ is the number of occurrences of the set $A_i^{(1)}$, $i = 1, ..., n_1$. The second source supplies N_2 observations of the parameter u and $c_i^{(2)}$ is the number of occurrences of the set $A_i^{(2)}$, $i = 1, ..., n_2$.

The main property of the Dirichlet distribution is that the Dirichlet prior density with parameters $\alpha = (\alpha_1, ..., \alpha_L)$ generates a posterior density function with parameters $\alpha^* = (\alpha_1^*, ..., \alpha_L^*)$, where $\alpha_i^* = (n_i + s\alpha_i)/(N + s)$. This implies that every source of data can be regarded as additional observations. If we have two sources, then the aggregated lower probability $\underline{P}(A, s)$ of an event A is

$$\underline{P}(A, s) = \frac{\sum_{i:A_i^{(1)} \subseteq A} c_i^{(1)} + \sum_{i:A_i^{(2)} \subseteq A} c_i^{(2)}}{N_1 + N_2 + s}.$$

Since

$$\sum_{i:A_i^{(1)} \subseteq A} c_i^{(1)} = N_1 Bel_1(A),$$

$$\sum_{i:A_i^{(2)} \subseteq A} c_i^{(2)} = N_2 Bel_2(A),$$

then

$$\underline{P}(A, s) = \frac{N_1 Bel_1(A) + N_2 Bel_2(A)}{N_1 + N_2 + s}. \quad (6)$$

On the other hand, lower probabilities of the event A separately obtained from the first and second sources are

$$\underline{P}_i(A, s_i) = \frac{N_i Bel_i(A)}{N_i + s_i}, \ i = 1, 2.$$

Hence

$$Bel_i(A) = \frac{N_i + s_i}{N_i} \underline{P}_i(A, s_i), \ i = 1, 2.$$

It is assumed here that the hyperparameters s_1 and s_2 of Dirichlet distributions, corresponding to different sources, may vary. By substituting the above equalities into (6), we get

$$\underline{P}(A, s) = \frac{N_1 + s_1}{N_1 + N_2 + s} \cdot \underline{P}_1(A, s_1) + \frac{N_2 + s_2}{N_1 + N_2 + s} \cdot \underline{P}_2(A, s_2). \quad (7)$$

Note that the hyperparameter s can be interpreted as the number of *hidden* observations. This implies that, by having s_1 hidden observations from the first source and s_2 hidden observations from the second source, we can expect $s = s_1 + s_2$ hidden observations in their combination. Then there holds

$$\underline{P}(A, s) = \frac{N_1 + s_1}{N_1 + N_2 + s_1 + s_2} \underline{P}_1(A, s_1) + \frac{N_2 + s_2}{N_1 + N_2 + s_1 + s_2} \underline{P}_2(A, s_2).$$

Denote

$$\gamma = \frac{N_1 + s_1}{N_1 + N_2 + s_1 + s_2}.$$

Then

$$\underline{P}(A, s) = \gamma \underline{P}_1(A, s_1) + (1 - \gamma) \underline{P}_2(A, s_2).$$

It is worth noticing that $\underline{P}(A, s) = \underline{P}_1(A, s_1) = \underline{P}_2(A, s_2)$ if $\underline{P}_1(A, s_1) = \underline{P}_2(A, s_2)$. Consequently, the above combination rule is idempotent.

The similar combination rule can be written for the upper probability:

$$\overline{P}(A,s) = \gamma \overline{P}_1(A,s_1) + (1-\gamma)\overline{P}_2(A,s_2).$$

The parameter γ can be viewed as a relative degree of reliability of sources. Furthermore, γ increases as N_1 increases. Indeed, large numbers of observations (judgments) produce more reliable estimates.

Example 5 *Let us return to Examples 1 and 3. Suppose that the first* 100 *experts are regarded as the first source of data and the second* 5 *experts are the second source. If to assume that all hyperparameters are identical* $(s_1 = s_2 = 1)$*, then it follows from (7):*

$$\begin{aligned}\underline{P}(A,2) &= \frac{100+1}{105+2}\cdot 0.535 + \frac{5+1}{105+2}\cdot 0.703 \\ &= 0.544,\end{aligned}$$

$$\begin{aligned}\overline{P}(A,2) &= \frac{100+1}{105+2}\cdot 0.703 + \frac{5+1}{105+2}\cdot 0.833 \\ &= 0.710\,.\end{aligned}$$

9. Bounds for distribution functions and expectations

Suppose that U is the real line restricted by some values $\inf U$ and $\sup U$. Then we can define lower and upper cumulative probability distribution functions of a random variable X, about which we have data in the form of intervals A_i, $i = 1, ..., n$. By using (5) and taking into account the fact that $\alpha(U) = 1$, we get

$$\begin{aligned}\underline{F}(t,s) &= \underline{P}(\{u \le t\}, s) \\ &= \begin{cases} C \sum\limits_{i:\sup A_i \le t} c_i, & t < \sup U \\ 1, & t = \sup U \end{cases},\end{aligned}$$

$$\begin{aligned}\overline{F}(t,s) &= \overline{P}(\{u \le t\}, s) \\ &= \begin{cases} C\left(s + \sum\limits_{i:\inf A_i \le t} c_i\right), & t > \inf U \\ 0, & t = \inf U \end{cases},\end{aligned}$$

where $C = (N+s)^{-1}$.

These expressions coincide with similar ones given in [3] by $s = 0$. The distribution functions are an envelope of all the possible cumulative distribution functions compatible with the data and they allow us to calculate the lower and upper expectations of X. If X is a continuous random variable, then the following hold

$$\begin{aligned}\underline{\mathbb{E}}_s X &= \int_U u \frac{\mathrm{d}\overline{F}(t,s)}{\mathrm{d}t}\mathrm{d}u \\ &= (N+s)^{-1}\left(s\cdot \inf U + \sum_{i=1}^{n} c_i \cdot \inf A_i\right),\end{aligned}$$

$$\begin{aligned}\overline{\mathbb{E}}_s X &= \int_U u \frac{\mathrm{d}\underline{F}(t,s)}{\mathrm{d}t}\mathrm{d}u \\ &= (N+s)^{-1}\left(s\cdot \sup U + \sum_{i=1}^{n} c_i \cdot \sup A_i\right).\end{aligned}$$

The same expressions can be obtained for the case of the discrete random variable X. It is worth noticing that the same expectations by $s = 0$ can be found in [6]:

$$\underline{\mathbb{E}}_s X = \sum_{i=1}^{n} m(A_i)\cdot \inf A_i,$$

$$\overline{\mathbb{E}}_s X = \sum_{i=1}^{n} m(A_i)\cdot \sup A_i.$$

Example 6 *Suppose that there are six expert judgments* $(N = 6)$ *about possible values of a random quantity* X*, say, shares of a firm in dollars next day, which can be changed in the interval* $U = [0, 10]$*. Three experts* $(c_1 = 3)$ *provide the interval* $A_1 = [4, 5]$*, two experts* $(c_2 = 2)$ *give* $A_2 = [2, 4]$*, and one expert* $(c_3 = 1)$ *gives* $A_3 = [1, 5]$*. Then bounds for the aggregated expected interval by* $s = 1$ *are* $\underline{\mathbb{E}}_1 X = 2.43$, $\overline{\mathbb{E}}_1 X = 5.43$*. Note that if* $s = 0$*, we get the following bounds:* $\underline{\mathbb{E}}_0 X = 2.83$, $\overline{\mathbb{E}}_0 X = 4.67$.

10. Conclusion

Extension of belief and plausibility measures when there is little information about elements of the frame of discernment has been proposed in the paper. Let me pointed out main virtues and shortcomings of the approach. First, the extended belief and plausibility functions can be simply updated after observing new events or obtaining new expert judgments because Dirichlet prior distributions generate Dirichlet posterior distributions. Second, the extended functions may be more realistic in many applications and possible large imprecision of results reflects insufficiency of available information. At the same time, the main open question here is how to determine the value of the hyperparameter s because different values of s lead to different results and to large imprecision. This is a direction for further work, which, in my opinion, could be solved in the framework of imprecise probability theory [11,18,21].

Acknowledgement

I am very grateful to George Klir (Binghamton State University, USA), Igor Kozine (Risoe National Laboratory, Denmark), Gert de Cooman, Matthias Troffaes (Ghent University, Belgium) for comments and stimulating discussions of some statements of this paper during the St.Petersburg Mini-Workshop "Imprecise Reasonong in Uncertainty Quantification".

REFERENCES

1. J.-M. Bernard. Non-parametric inference about an unknown mean using the imprecise Dirichlet model. In G. de Cooman, T.L. Fine, and T. Seidenfeld, editors, *Imprecise Probabilities and Their Applications. Proc. of the 2nd Int. Symposium ISIPTA'01*, pages 40–50, Ithaca, USA, June 2001. Shaker Publishing.
2. J.-M. Bernard. Implicative analysis for multivariate binary data using an imprecise Dirichlet model. *Journal of Statistical Planning and Inference*, 105(1):83–104, 2002.
3. A. Bernardini. What are the random and fuzzy sets and how to use them for uncertainty modelling in engineering systems? In I. Elishakoff, editor, *Whys and hows in uncertainty analysis: probability, fuzziness and anti-optimization*, pages 63–125. Springer, Vienna, 1999.
4. F.P.A. Coolen. An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, 56:61–68, 1997.
5. M. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.
6. A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annales of Mathematical Statistics*, 38:325–339, 1967.
7. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
8. S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Report SAND2002-4015, Sandia National Laboratories, January 2003.
9. J.Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.
10. G.J. Klir and M.J. Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory*. Physica-Verlag, Heidelberg, 1998.
11. V. P. Kuznetsov. *Interval Statistical Models*. Radio and Communication, Moscow, 1991. in Russian.
12. E. Quaeghebeur and G. de Cooman. Game-theoretic learning using the imprecise Dirichlet model. In J.-M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Proc. of the 3rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'03*, pages 450–464, Lugano, Switzerland, July 2003. Carleton Scientific.
13. K. Sentz and S. Ferson. Combination of evidence in Dempster-Shafer theory. Report SAND2002-0835, Sandia National Laboratories, April 2002.
14. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
15. P. Smets. Belief induced by the knowledge of the probabilities. Technical report 94-4.1, Institut de Recherches Interdisciplinaires

et de Developpements en Intelligence Artificielle, Brussels, 1994.
16. M. Smithson, T. Bartos, and K. Takemura. Human judgment under sample space ignorance. In G. de Cooman, F.G. Cozman, S. Moral, and P. Walley, editors, *ISIPTA '99 - Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, pages 324–332, Zwijnaarde, Belgium, June 1999.
17. F. Tonon, A. Bernardini, and A. Mammino. Determination of parameters range in rock engineering by means of random set theory. *Reliability Engineering and System Safety*, 70:241–261, 3 2000.
18. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
19. P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:357, 1996. with discussion.
20. P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.
21. K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung*, volume I Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg, 2001.
22. S. Wilks. *Mathematical Statistics*. Wiley and Sons, New York, 1962.
23. M. Zaffalon. Exact credal treatment of missing data. *Journal of Statistical Planning and Inference*, 105(1):105–122, 2002.