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

Lev V. Utkin[a]

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

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. However, the classical probabilistic interpretation used for computing belief and plausibility measures may be unreasonable in many real applications when the number of observations or measurements is rather small. 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. Combination rules taking into account reliability of sources of data are studied. Various numerical examples illustrate the proposed extension.

Keywords: Approximate reasoning; Belief measures; Dirichlet distribution; Multinomial model; Possibility theory; Dempster-Shafer theory; Imprecise probability theory

1. Introduction

Evidence theory or Dempster-Shafer theory [9,21] provides 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 [13,20], 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 they require 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 using Walley's imprecise Dirichlet model [26]. Since the possibility distribution function [10] can be regarded as a special case of belief and plausibility functions, 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. The imprecise Dirichlet model is considered in Section 3. A relationship between a set of multinomial models and statistical data in the form of subsets of a universal set is established in Section 4. The extended belief and plausibility functions are proposed in the same section. The extended belief and plausibility functions are explained in terms of a multivalued sampling process in Section 5. Some important properties of the obtained extensions are investigated in Section 6. The extended possibility distribution is described in Section 7. Two combination rules are studied in Section 8. Lower and upper probability distributions and expectations produced by the extended 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 an element $u \in U$, 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, called *basic probability assignment*, can be defined such that [9,16,21]:

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

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

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 [9], 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, then A_i is called a *focal element*.

According to [21], the *belief* $Bel(A)$ and *plausibility* $Pl(A)$ measures of an event $A \subseteq \Omega$ can be defined as

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

As pointed out in [13], 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 [9] 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., $Bel(A) \leq \Pr(A) \leq Pl(A)$.

Example 1 *A system consists of four components (*$U = \{1, 2, 3, 4\}$*). After a system failure,* 100 *sensors (*$N = 100$*) detect its cause.* 60 *sensors (*$c_1 = 60$*) indicate that the first component failed (*$A_1 = \{1\}$*),* 20 *sensors (*$c_2 = 20$*) indicate that the first or the second component failed (*$A_2 = \{1, 2\}$*),* 20 *sensors (*$c_3 = 20$*) indicate that the third component failed (*$A_3 = \{3\}$*). Then*

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

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

$$Bel(A) = m(A_1) = 0.6,\ Pl(A) = m(A_1) + m(A_2) = 0.8.$$

Now suppose that only 5 *sensors (*$N = 5$*) supply the same information approximately in the same proportion. Three sensors (*$c_1 = 3$*) indicate that the first component failed (*$A_1 = \{1\}$*), one sensor (*$c_2 = 1$*) indicates that the first or the second component failed (*$A_2 = \{1, 2\}$*), one sensor (*$c_3 = 1$*) indicates that the third component failed (*$A_3 = \{3\}$*). Then*

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

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 the same belief and plausibility functions in both cases. However, it is obvious that the number of sensors in the second case 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 [22]. If N is small, inferences become too precise and incautious. In order to overcome this difficulty, I shall apply the imprecise Dirichlet model [26], extending belief and plausibility functions such that a lack of sufficient statistical data can be taken into account.

3. Multinomial sampling and the imprecise Dirichlet model

Let $U = \{u_1, ..., u_K\}$ be a set of possible outcomes u_j. Assume the *standard multinomial model*: N observations are independently chosen from U with an identical probability distribution $\Pr\{u_j\} = \theta_j$ for $j = 1, ..., K$, where each $\theta_j \geq 0$ and $\sum_{j=1}^{K} \theta_j = 1$. Denote $\theta = (\theta_1, ..., \theta_K)$. Let n_j denote the number of observations of u_j in the N trials, so that $n_j \geq 0$ and $\sum_{j=1}^{K} n_j = N$. Under the above assumptions the random variables $n_1, ..., n_K$ have a multinomial distribution and the observed multinomial likelihood function generated by the data $\mathbf{n} = (n_1, ..., n_K)$ is

$$L(\mathbf{n}|\theta) \propto \prod_{j=1}^{K} \theta_j^{n_j}.$$

The *Dirichlet* $(s,\mathbf{t})$ *prior distribution* for θ, where $\mathbf{t} = (t_1,...,t_K)$, has probability density function [8,28]

$$p(\theta) = \Gamma(s)\left(\prod_{j=1}^{K}\Gamma(st_j)\right)^{-1}\cdot\prod_{j=1}^{K}\theta_j^{st_j-1}.$$

Here the parameter $t_i \in (0,1)$ is the mean of θ_i under the Dirichlet prior; the hyperparameter $s > 0$ determines the influence of the prior distribution on posterior probabilities; the vector $\mathbf{t}$ belongs to the interior of the K-dimensional unit simplex denoted by $S(1,K)$; $\Gamma(\cdot)$ is the Gamma-function which satisfies $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(1) = 1$.

When multiplied by the multinomial likelihood function $L(\mathbf{n}|\theta)$, the Dirichlet $(s,\mathbf{t})$ prior density generates a posterior density function

$$p(\theta|\mathbf{n}) \propto p(\theta)L(\mathbf{n}|\theta) = \prod_{j=1}^{K}\theta_j^{n_j+st_j-1},$$

which is seen to be the probability density function of a Dirichlet $(N+s,\mathbf{t}^*)$ distribution, where $t_j^* = (n_j + st_j)/(N+s)$.

A property of the Dirichlet distribution is that its marginal distributions are also Dirichlet distributions. In particular, its univariate marginals are Beta distributions ($K = 2$) and the Dirichlet distribution can be seen as a multivariate generalization of the Beta distribution.

Walley [26] pointed out several reasons for using a set of Dirichlet distributions to model prior ignorance about probabilities θ:

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 θ are Dirichlet distributions.

The *imprecise Dirichlet model* (IDM) is defined by Walley [26] as the set of all Dirichlet $(s,\mathbf{t})$ distributions such that $\mathbf{t} \in S(1,K)$. We will see below that the choice of this model allows us to model the fact that we do not know the a priori probabilities of events.

For the IDM, the *hyperparameter* s determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Walley [26] 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 K or a number of observations. The detailed discussion concerning the parameter s and the IDM can be found in [1–3,23,26]. The application of the model in reliability was studied in [6]. This model was also applied to the game theory for choosing a strategy in a two-player game by Quaeghebeur and de Cooman [18]. An approach to dealing with incomplete sets of multivariate categorical data by exploiting Walley's IDM was studied by Zaffalon [29]. Hutter [15] provides an approach for deriving closed form expressions for interval estimates in the IDM.

Let A be any non-trivial subset of a sample space $\{u_1, ..., u_K\}$, i.e., A is not empty and $A \neq U$, and let $n(A)$ denote the observed number of occurrences of A in the N trials, $n(A) = \sum_{u_j \in A} n_j$. Then, according to [26], the predictive probability $P(A|\mathbf{n}, \mathbf{t}, s)$ under the Dirichlet posterior distribution is

$$P(A|\mathbf{n}, \mathbf{t}, s) = (n(A) + st(A)) \,/\, (N + s)\,,$$

where $t(A) = \sum_{u_j \in A} t_j$.

It should be noted that $P(A|\mathbf{n}, \mathbf{t}, s) = 0$ if A is empty and $P(A|\mathbf{n}, \mathbf{t}, s) = 1$ if $A = U$.

By maximizing and minimizing $P(A|\mathbf{n}, \mathbf{t}, s)$ over $\mathbf{t} \in S(1, K)$, we obtain the posterior upper and lower probabilities of A:

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

Before making any observations, $n(A) = N = 0$, so that $\underline{P}(A|\mathbf{n}, s) = 0$ and $\overline{P}(A|\mathbf{n}, s) = 1$ for all non-trivial events A. This is the *vacuous* probability model. Therefore, by using the IDM, we do not need to choose one specific prior. In contrast, the objective Bayesian approach [5] aims at modeling prior ignorance about the chances θ by characterizing prior uncertainty by a single prior probability distribution.

4. Extended belief and plausibility functions

4.1. A set of multinomial models

Suppose that there is a set of N imprecise observations $A_i \subseteq U$ and c_i is the number of occurrences of the subset A_i, $i = 1, ..., n$. The aim of this section is to consider a set of multinomial models produced by the available information. We construct this set by means of the following modified urn model.

Suppose that there are K urns $u_1, ..., u_K$ containing balls with numbers $1, ..., K$, respectively. We randomly choose a subset A_i of l_i urns such that $A_i = \{u_j : j \in J_i\}$, where J_i is a set of indices. Then we take randomly c_i balls from the urns numbered by elements of J_i. The same procedure is repeated n times, i.e., for each of the focal elements. What can we say about possible numbers of balls chosen from each urn now? It is obvious that there exist different combinations of numbers of balls except in 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. The number M is determined as

$$M = \prod_{i=1}^{n} \binom{c_i + |A_i| - 1}{c_i}.$$

Here $|A_i|$ denotes the cardinality of A_i. Denote the k-th possible vector of balls by $\mathbf{n}^{(k)} = (n_1^{(k)}, ..., n_K^{(k)})$, $k = 1, ..., M$, and the vector of the numbers c_i of occurrences of A_i by $\mathbf{c} = (c_1, ..., c_n)$. If we assume that the subsets A_i are independently chosen from the set of all subsets of U and the probability of selecting a ball from the i-th urn is θ_i, then every combination of balls produces the standard multinomial model (see Section 3). M possible combinations of balls produce M standard multinomial models. Moreover, we can not prefer one model over another.

Example 2 *Let $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$. Suppose that there are three focal sets $A_1 = \{u_4, u_5, u_6\}$, $A_2 = \{u_3, u_4\}$, $A_3 = \{u_2, u_3, u_4, u_5\}$ and $c_1 = c_3 = 1$, $c_2 = 2$. One of the possible vectors of balls is $\mathbf{n}^{(k)} = (0, 0, 1, 2, 0, 1)$. This vector is obtained by choosing u_6 from the subset A_1 (we take randomly 1 ball from the urns corresponding to the subset A_1 because $c_1 = 2$), u_3 and u_4 from A_2 (we take randomly 2 balls from the urns corresponding to the subset A_2 because $c_2 = 2$), and u_4 from A_3. As a result, we have one ball with the number 3, two balls with the number 4, and one ball with the number 6. Let θ_i be the probability of selecting a ball from the i-th urn. Then the observed multinomial likelihood function generated by the data $\mathbf{n}^{(k)} = (0, 0, 1, 2, 0, 1)$ is*

$$L((0, 0, 1, 2, 0, 1) \mid (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)) \propto \theta_3 \theta_4^2 \theta_6.$$

In the same way, we can obtain all 36 vectors $\mathbf{n}^{(k)}$.

Remark 1 *It should be noted that the set of possible vectors of balls $\mathbf{n}^{(k)}$ produced by imprecise observations is very closely related to a set of all possible completions of missing attributes in models of missing data proposed by de Cooman and Zaffalon [7,29]. Moreover, the set of possible completions of missing attributes coincide with the set of $\mathbf{n}^{(k)}$ when the observations are of a special type. Therefore, the models considered in [7,29] can be regarded as a special case of the models produced by imprecise observations. Dealing with missing data by means of interval probabilities was also considered by Ramoni and Sebastiani [19].*

Since we have a set of vectors $\mathbf{n}^{(k)}$, there exists a set of likelihood functions $L(\mathbf{n}^{(k)}|\theta)$, $k = 1, ..., M$. This implies that, the probability of an arbitrary event $A \subseteq U$ depends on $\mathbf{n}^{(k)}$, that is, we can find $P(A|\mathbf{n}^{(k)})$. Even if we know precisely the conditional probabilities $P(A|\mathbf{n}^{(k)})$ for every event A in U and every possible vector $\mathbf{n}^{(k)}$, still we can only compute lower and upper probabilities for events A:

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

In particular, if all sets A_i consist of single elements, then $M = 1$ and

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

Since the vectors $\mathbf{n}^{(k)}$ depend on $\mathbf{c}$, the resulting lower and upper probabilities (after minimizing and maximizing $P(A|\mathbf{n}^{(k)})$) depend on $\mathbf{c}$ and they are denoted $\underline{P}(A|\mathbf{c})$ and $\overline{P}(A|\mathbf{c})$.

How can we obtain $P(A|\mathbf{n}^{(k)})$? In the case of multinomial samples, the Dirichlet distribution is the traditional choice. We choose the Dirichlet prior as a distribution on the probabilities $P(u_j) = \theta_j$, $j = 1, ..., K$, and this prior is updated using Bayes rule, by multiplying it with the multinomial likelihood function $L(\mathbf{n}^{(k)}|\theta)$.

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 finite set U^* of non-intersecting intervals B_k such that $B_1 \cup ... \cup B_K = U$, $K = 2^{n+1}$. Moreover, every interval A_i can be represented as the union of a finite number of intervals B_k. Therefore, by associating every interval B_k with an urn, we can construct the set of multinomial models considered above. It should be noted that different divisions of U can be considered. However, the proposed division produces the finite set U^* with the minimal value K.

4.2. Bounds for belief and plausibility functions

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

$$\underline{P}(A|\mathbf{c}, s) = \min_{k=1,...,M} \inf_{\mathbf{t}\in S(1,K)} \frac{n^{(k)}(A) + st(A)}{N+s},$$

$$\overline{P}(A|\mathbf{c}, s) = \max_{k=1,...,M} \sup_{\mathbf{t}\in S(1,K)} \frac{n^{(k)}(A) + st(A)}{N+s},$$

where $t(A) = \sum_{j\in J} t_j$, $n^{(k)}(A) = \sum_{j\in J} n_j^{(k)}$.

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

$$\underline{P}(A|\mathbf{c}, s) = \frac{\min_{k=1,...,M} n^{(k)}(A) + s \cdot \inf_{\mathbf{t}\in S(1,K)} t(A)}{N+s}, \tag{3}$$

$$\overline{P}(A|\mathbf{c}, s) = \frac{\max_{k=1,...,M} n^{(k)}(A) + s \cdot \sup_{\mathbf{t}\in S(1,K)} t(A)}{N+s}. \tag{4}$$

Note that $\inf_{\mathbf{t}\in S(1,K)} t(A)$ is achieved at $t(A) = 0$ and $\sup_{\mathbf{t}\in S(1,K)} t(A)$ is achieved at $t(A) = 1$ except when $A = U$ or $A = \varnothing$. If $A = U$, then the minimum and maximum are achieved at $t(A) = 1$. If $A = \varnothing$, then the minimum and maximum are achieved at $t(A) = 0$.

For achieving the minimum of $n^{(k)}(A)$ denoted $\Phi_1(A)$, we consider the only sets A_i for which the balls can not be placed all outside of A. These sets are subsets of A, i.e., $A_i \subseteq A$. Therefore, we can write

$$\Phi_1(A) = \min_{k=1,...,M} n^{(k)}(A) = \sum_{i:A_i\subseteq A} c_i.$$

For achieving the maximum of $n^{(k)}(A)$ denoted $\Phi_2(A)$, we consider the only sets A_i for which the balls can be placed inside of A. All these sets have at least one common urn with A, i.e., $A_i \cap A \neq \varnothing$. Therefore, we can write

$$\Phi_2(A) = \max_{k=1,\dots,M} n^{(k)}(A) = N - \sum_{i:A_i \cap A = \varnothing} c_i = \sum_{i:A_i \cap A \neq \varnothing} c_i.$$

It can be seen from (1) that

$$\frac{\Phi_1(A)}{N} = \sum_{i:A_i \subseteq A} \frac{c_i}{N} = \sum_{A_i:A_i \subseteq A} m(A_i),$$

$$\frac{\Phi_2(A)}{N} = \sum_{i:A_i \cap A \neq \varnothing} \frac{c_i}{N} = \sum_{A_i:A_i \cap A \neq \varnothing} m(A_i).$$

This implies that $\Phi_1(A)/N$ and $\Phi_2(A)/N$ (see (2)) are nothing else but the belief $Bel(A)$ and plausibility $Pl(A)$ measures of the set $A \subseteq U$. Then

$$\underline{P}(A|\mathbf{c}, s) = \frac{N \cdot Bel(A)}{N + s}, \quad \overline{P}(A|\mathbf{c}, s) = \frac{N \cdot Pl(A) + s}{N + s}. \tag{5}$$

It is obvious from the above expressions that $\underline{P}(A|\mathbf{c}, s) \leq Bel(A)$ and $\overline{P}(A|\mathbf{c}, s) \geq Pl(A)$, i.e., $\underline{P}(A|\mathbf{c}, s)$ and $\overline{P}(A|\mathbf{c}, s)$ 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:

$$\underline{P}(A|\mathbf{c}, s) = \varkappa Bel(A), \quad \overline{P}(A|\mathbf{c}, s) = 1 - \varkappa(1 - Pl(A)),$$

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

From this representation, lower $\underline{P}(A|\mathbf{c}, s)$ and upper $\overline{P}(A|\mathbf{c}, s)$ probabilities can be regarded as some correction of belief and plausibility functions. The same extensions in an implicit form have been obtained by Zaffalon [7,29] for models of missing data.

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 information supplied by 5 sensors:*

$$\underline{P}(A|\mathbf{c}, 1) = \frac{5 \cdot 0.6}{5 + 1} = 0.5, \quad \overline{P}(A|\mathbf{c}, 1) = \frac{5 \cdot 0.8 + 1}{5 + 1} = 0.833.$$

If $s = 2$, then $\underline{P}(A|\mathbf{c}, 2) = 0.428$, $\overline{P}(A|\mathbf{c}, 2) = 0.857$. By comparing these results with the bounds $[0.6, 0.8]$ 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 information supplied by 100 sensors:

$$\underline{P}(A|\mathbf{c}, 1) = \frac{100 \cdot 0.6}{100 + 1} = 0.594, \quad \overline{P}(A|\mathbf{c}, 1) = \frac{100 \cdot 0.8 + 1}{100 + 1} = 0.802.$$

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

5. Multivalued mapping

Let us explain the obtained extended belief and plausibility functions in terms of a multinomial multivalued sampling process[1]. Consider a probability measure $P(\omega)$ defined on a universal set Ω (which can be thought of as the set of our observations) related to U (the set of the values of our measurements) through a multivalued mapping $G:\Omega\to\mathcal{P}(U)$. Then the basic probability assignment is [9]:

$$m(A_i)=P(\omega_i)=c_i/N,\ \omega_i\in\Omega.$$

This multivalued mapping expresses the imprecision of the measurement experienced during each observation, i.e., our inability to attach a single number to each observation. So, for each set $A_i\in\mathcal{P}(U)$, the value $m(A_i)$ expresses the probability of $\omega_i=G^{-1}(A_i)$ ($\omega_i\in\Omega$). A *random set* is the pair $(\mathcal{F},m)$, where $\mathcal{F}$ is the family of all N focal elements.

Let A be a subset of U. If we define X_* as the subset of Ω whose elements must lead to A: $X_*=\{\omega\in\Omega:G(\omega)\subseteq A\}$, then the lower probability of A, according to Dempster's principle of inductive reasoning, is defined by $\underline{P}(A|\mathbf{c})=P(X_*)$. If we define X^* as the subset of Ω whose elements may lead to A: $X^*=\{\omega\in\Omega:G(\omega)\cap A\neq\varnothing\}$, then the upper probability of A is given by $\overline{P}(A|\mathbf{c})=P(X^*)$.

Suppose that the set Ω consists of K points $\omega_1,...,\omega_K$. Every observed subset A_j, corresponds to one point ω_j, $i=1,...,K$. Suppose that $P(\omega_j)=\theta_j$. By having N observations of $\omega_1,...,\omega_K$ independently chosen from Ω with probabilities $P\{\omega_j\}=\theta_j$, $j=1,...,K$, we deal with the multinomial model. By assuming that the probabilities $\theta=(\theta_1,...,\theta_K)$ have the Dirichlet $(s,\mathbf{t})$ distribution, we obtain the lower probability of A as follows:

$$\underline{P}(A|\mathbf{c},s)=\frac{n(X_*)+st(X_*)}{N+s}.$$

where $t(X_*)=\sum_{\omega_j\in X_*}t_j$, $n(X_*)=\sum_{\omega_j\in X_*}c_j$.

By using the IDM, we get for $A\subset U$

$$\begin{aligned}\underline{P}(A|\mathbf{c},s)&=\min_{\mathbf{t}\in S(1,K)}\frac{n(X_*)+st(X_*)}{N+s}=(N+s)^{-1}\sum_{\omega_j\in X_*}c_j\\&=N\cdot Bel(A)/(N+s)=\varkappa\cdot Bel(A).\end{aligned}$$

If $A=U$, then $\underline{P}(A|\mathbf{c},s)=1$.

The upper probability of A can be obtained in the same way:

$$\begin{aligned}\overline{P}(A|\mathbf{c},s)&=\max_{\mathbf{t}\in S(1,K)}\frac{n(X^*)+st(X^*)}{N+s}=\left(\sum_{\omega_j\in X^*}c_j+s\right)/(N+s)\\&=\left(N\cdot Pl(A)+s\right)/\left(N+s\right)=1-\varkappa(1-Pl(A)).\end{aligned}$$

[1] The idea to explain the results in terms of a multinomial multivalued sampling process belongs to one of the reviewers of the paper.

Remark 3 *The obtained probabilities can also be considered in the framework of an ε-contaminated model [14]. As pointed out by Seidenfeld and Wasserman in the discussion part of Walley's paper [26], the IDM has the same lower and upper probabilities as the ε-contaminated model (a class of probabilities which for fixed $\varepsilon \in (0,1)$ and $P(\omega_i)$ is the set $\{(1-\varepsilon)P(\omega_i) + \varepsilon Q(\omega_i)\}$). Here $P(\omega_i) = c_i/N$, Q is an arbitrary distribution and $\varepsilon = s/(N+s) = 1 - \varkappa$. Then*

$$\begin{aligned}\underline{P}(A|\mathbf{c},\varepsilon) &= \min_Q \sum_{\omega_i \in X_*} \{(1-\varepsilon)P(\omega_i) + \varepsilon Q(\omega_i)\} \\ &= \sum_{\omega_j \in X_*} (1-\varepsilon)P(\omega_i) = \sum_{\omega_j \in X_*} (1-\varepsilon)c_i/N = (1-\varepsilon)Bel(A),\end{aligned}$$

$$\begin{aligned}\overline{P}(A|\mathbf{c},s) &= \max_Q \sum_{\omega_i \in X^*} \{(1-\varepsilon)P(\omega_i) + \varepsilon Q(\omega_i)\} \\ &= \varepsilon + \sum_{\omega_i \in X^*} (1-\varepsilon)P(\omega_i) = \varepsilon + \sum_{\omega_i \in X^*} (1-\varepsilon)c_i/N \\ &= (1-\varepsilon)Pl(A) + \varepsilon.\end{aligned}$$

6. Properties of lower and upper probabilities

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

1. For any $s \geq 0$, we have $\underline{P}(A|\mathbf{c},s) \leq Bel(A)$, $\overline{P}(A|\mathbf{c},s) \geq Pl(A)$.

2. $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ are belief and plausibility functions with the basic probability assignment $m^*(A_i) = c_i/(N+s)$ for every A_i and the additional basic probability assignment $m^*(U) = s/(N+s)$, i.e., $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ can be obtained as standard belief and plausibility functions under condition that there are s additional observations $A_{n+1} = U$. If we denote $m(A_i) = c_i/N$, then $m^*(A_i) = m(A_i) \cdot N/(N+s) = \varkappa \cdot m(A_i)$, and

$$\underline{P}(A|\mathbf{c},s) \sum_{A_i : A_i \subseteq A} m^*(A_i),\ \overline{P}(A|\mathbf{c},s) = m^*(U) + \sum_{A_i : A_i \cap A \neq \varnothing} m^*(A_i).$$

   The above also follows from an interpretation of the hyperparameter s as the number of *hidden* observations [26]. At the same time, the value $1 - \varkappa$ can be regarded as discount rate [21] characterizing the reliability of a source of data. This implies that the application of Walley's IDM leads to a discounting scheme with discounting rates strongly defined by the number of observations N and by the hyperparameter s.

3. It follows from Property 2 that

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

$$\underline{P}(A|\mathbf{c},s) + \underline{P}(B|\mathbf{c},s) \leq \underline{P}(A \cup B|\mathbf{c},s),$$

   where A^c is the complement of the set A and $A, B \subseteq U$

4. 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 $s \geq 0$, we have

$$\underline{P}(A|\mathbf{c}, s) = Bel(A) = 0, \quad \overline{P}(A|\mathbf{c}, s) = Pl(A) = 1.$$

5. 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: $\underline{P}(A|\mathbf{c}, s) = Bel(A)$, $Pl(A) = \overline{P}(A|\mathbf{c}, s)$.

6. If $s_1 \leq s_2$, then $[\underline{P}(A|\mathbf{c}, s_1), \overline{P}(A|\mathbf{c}, s_1)] \subseteq [\underline{P}(A|\mathbf{c}, s_2), \overline{P}(A|\mathbf{c}, s_2)]$.

Let us consider two special cases when belief and plausibility functions may give improper conclusions. Suppose that we have a multivalued estimate $A \subset U$ ($\mathbf{c} = (1)$). According to the rule, given by Eq. (1), the corresponding belief and plausibility functions are $Bel(A) = Pl(A) = 1$. This implies that we completely believe this one estimate: this conclusion may be too risky. This can be avoided by using the IDM ($s > 0$). Indeed, if we take $s = 1$, then

$$\underline{P}(A|\mathbf{c}, s) = 1/(1 + s) = 0.5, \ \overline{P}(A|\mathbf{c}, s) = 1.$$

It is worth noting that if we have r identical estimates ($\mathbf{c} = (r)$), 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 $\underline{P}(A|\mathbf{c}, s)$ and $\overline{P}(A|\mathbf{c}, s)$ are

$$\underline{P}(A|\mathbf{c}, s) = r/(r + s), \ \overline{P}(A|\mathbf{c}, s) = 1.$$

Suppose that there are 2 conflicting judgments A_1 and A_2 ($\mathbf{c} = (1, 1)$) such that $A_1 \cap A_2 = \varnothing$. Then, by Eq. (1), we have $Bel(A_i) = Pl(A_i) = 0.5$, $i = 1, 2$. For instance, let $U = [0, 1]$, $A_1 = [0, 0.999]$, $A_2 = [0.9991, 1]$. If A_1 and A_2 are interval-valued measurements or observations, then the above conclusion does not raise doubts. However, if the intervals are supplied by two experts, then we can observe that the first expert is overcautious and the second expert is overconfident [11]. In this case, the conclusion $Bel(A_i) = Pl(A_i) = 0.5$ may be too risky. At the same time, we can write

$$\underline{P}(A_i|\mathbf{c}, s) = 1/(2 + s), \ \overline{P}(A_i|\mathbf{c}, s) = (1 + s)/(2 + s).$$

If we take $s = 1$, then we get a bit more cautious bounds $\underline{P}(A_i|\mathbf{c}, 1) = 1/3$ and $\overline{P}(A_i|\mathbf{c}, 1) = 2/3$.

The above special cases show advantages of the 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 [10]. In this interpretation, a fuzzy set is a model of ambiguity and not of vagueness [24].

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

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

It is obvious that $\overline{\pi}_F(u|s) \geq \pi_F(u)$ for all u and for any $s \geq 0$. Note that the resulting extension of a possibility measure is again a possibility measure, and hence, it is indeed uniquely characterized by its extended possibility distribution.

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.1 (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.1 (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 rules

Combination rules are some types of aggregation methods for data obtained from multiple sources (bodies of evidence). A large literature has developed over the last years on the aggregation of evidence and a detailed overview of combination rules (Dempster's rule, the discount method, Yager's modified rule, Inagaki's unified rule, Zhang's center combination rule, Dubois and Prade's disjunctive consensus rule, "mixing" combination rule, etc.) is given in [20,12]. Below we describe two modified combination rules by using the IDM.

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

8.1. Linear combination of evidence

One of the properties of the Dirichlet distribution is that the Dirichlet prior density with parameters $\mathbf{t} = (t_1, ..., t_K)$ generates a posterior density function with parameters

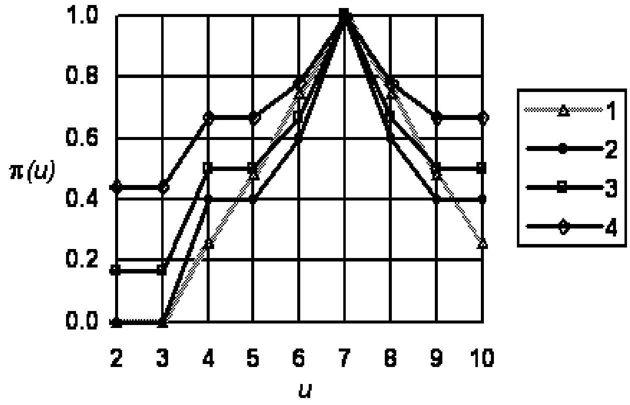

Figure 1. Possibility distribution functions by various s and N

$\mathbf{t}^* = (t_1^*, ..., t_K^*)$, where $t_i^* = (n_i + st_i)/(N + s)$. This implies that every source of data can be regarded as additional observations. If we have two sources, then the aggregated lower probability $\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s)$ of an event A is

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

Since

$$\sum_{i:A_i^{(1)} \subseteq A} c_i^{(1)} = N_1 Bel_1(A), \quad \sum_{i:A_i^{(2)} \subseteq A} c_i^{(2)} = N_2 Bel_2(A),$$

it follows that

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = \frac{N_1 Bel_1(A) + N_2 Bel_2(A)}{N_1 + N_2 + s}. \tag{6}$$

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

$$\underline{P}_i(A|\mathbf{c}^{(i)}, s_i) = N_i \cdot Bel_i(A) / \left(N_i + s_i\right), \ i = 1, 2.$$

Hence

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

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

$$\begin{aligned} \underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = & \frac{N_1 + s_1}{N_1 + N_2 + s} \cdot \underline{P}_1(A|\mathbf{c}^{(1)}, s_1) \\ & + \frac{N_2 + s_2}{N_1 + N_2 + s} \cdot \underline{P}_2(A|\mathbf{c}^{(2)}, s_2) \end{aligned} \tag{7}$$

and if $A = U$, then $\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = 1$.

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

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = \frac{N_1 + s_1}{N_1 + N_2 + s_1 + s_2}\underline{P}_1(A|\mathbf{c}^{(1)}, s_1) + \frac{N_2 + s_2}{N_1 + N_2 + s_1 + s_2}\underline{P}_2(A|\mathbf{c}^{(2)}, s_2).$$

Denote $\gamma = (N_1 + s_1)/(N_1 + N_2 + s_1 + s_2)$. Then

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = \gamma\underline{P}_1(A|\mathbf{c}^{(1)}, s_1) + (1-\gamma)\underline{P}_2(A|\mathbf{c}^{(2)}, s_2).$$

It is worth noting that

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = \underline{P}_1(A|\mathbf{c}^{(1)}, s_1) = \underline{P}_2(A|\mathbf{c}^{(2)}, s_2)$$

if $\underline{P}_1(A|\mathbf{c}^{(1)}, s_1) = \underline{P}_2(A|\mathbf{c}^{(2)}, s_2)$. Consequently, the above combination rule is idempotent.

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

$$\overline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s) = \gamma\overline{P}_1(A|\mathbf{c}^{(1)}, s_1) + (1-\gamma)\overline{P}_2(A|\mathbf{c}^{(2)}, s_2).$$

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.

The above rule is similar to the discount method [21] proposed for combining the belief functions.

It is worth noting that $\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s)$ and $\overline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), s)$ are again belief and plausibility functions with the basic probability assignments

$$m_{12}^*(A_i^{(1)}) = \gamma c_i^{(1)}/(N_1 + s_1),$$

$$m_{12}^*(A_j^{(2)}) = (1-\gamma)m_2^*(A_j^{(2)}) = (1-\gamma)c_j^{(2)}/(N_2 + s_2),$$

$$m_{12}^*(U) = \gamma s_1/(N_1 + s_1) + (1-\gamma)s_2/(N_2 + s_2).$$

If $A_k^{(1)} = A_k^{(2)} = A_k$, then

$$m_{12}^*(A_k) = \gamma c_k^{(1)}/(N_1 + s_1) + (1-\gamma)c_k^{(2)}/(N_2 + s_2).$$

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

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), 2) = \frac{100+1}{105+2}\cdot 0.594 + \frac{5+1}{105+2}\cdot 0.5 = 0.589,$$

$$\overline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), 2) = \frac{100+1}{105+2}\cdot 0.802 + \frac{5+1}{105+2}\cdot 0.833 = 0.804.$$

8.2. Modified Dempster's rule

Dempster's rule combines multiple belief functions through their basic probability assignments. Denote

$$m_1(A_i^{(1)}) = c_i^{(1)}/N_1,\ m_2(A_j^{(2)}) = c_j^{(2)}/N_2.$$

A combined probability assignment (m_{12}) is given by

$$m_{12}(A) = \frac{1}{1-K} \sum_{A_i^{(1)} \cap A_j^{(2)} = A} m_1(A_i^{(1)}) m_2(A_j^{(2)}),$$

where

$$K = \sum_{A_i^{(1)} \cap A_j^{(2)} = \varnothing} m_1(A_i^{(1)}) m_2(A_j^{(2)})$$

and $m_{12}(\varnothing) = 0$. For simplicity, we assume that $A_i^{(1)} \neq U$ and $A_j^{(2)} \neq U$ for all i and j.

K represents basic probability mass associated with conflict. Note that Dempster's rule can not be used in case of $K = 1$, i.e., conflicting evidence can not be combined. To handle conflict of information sources, the discounting scheme has been introduced in Dempster-Shafer theory. By using Property 2 given in Section 6, we can modify Dempster's rule if we replace the basic probability assignments $m_1(A_i^{(1)})$ and $m_2(A_j^{(2)})$ by the (discounted) probability assignments $m_1^*(A_i^{(1)}) = \varkappa_1 m_1(A_i^{(1)})$, $m_2^*(A_j^{(2)}) = \varkappa_2 m_2(A_j^{(2)})$ and additional assignments $m_1^*(U) = 1 - \varkappa_1$, $m_2^*(U) = 1 - \varkappa_2$, where

$$\varkappa_1 = N_1/(N_1 + s_1),\ \varkappa_2 = N_2/(N_2 + s_2).$$

Hence

$$\begin{aligned} 1 - K^* &= \sum_{A_i^{(1)} \cap A_j^{(2)} \neq \varnothing} m_1^*(A_i^{(1)}) m_2^*(A_j^{(2)}) \\ &= (1-K)\varkappa_1\varkappa_2 + \sum_i m_1^*(A_i^{(1)}) m_2^*(U) \\ &+ \sum_j m_2^*(A_j^{(2)}) m_1^*(U) - m_1^*(U) m_2^*(U) \\ &= 1 - \varkappa_1\varkappa_2 K. \end{aligned}$$

If $A \neq U$, then

$$\begin{aligned} m_{12}^*(A) &= \frac{\varkappa_1\varkappa_2}{1-K^*} \sum_{A_i^{(1)} \cap A_j^{(2)} = A} m_1(A_i^{(1)}) m_2(A_j^{(2)}) \\ &+ \frac{\varkappa_1(1-\varkappa_2)}{1-K^*} m_1(A) + \frac{\varkappa_2(1-\varkappa_1)}{1-K^*} m_2(A). \end{aligned}$$

If A does not belong to the sets of focal elements $\{A_i^{(1)}\}$, $\{A_j^{(2)}\}$, then $m_1(A) = 0$ and $m_2(A) = 0$. If $A = U$, then

$$m_{12}^*(U) = (1 - \varkappa_1)(1 - \varkappa_2)/(1 - K^*).$$

It can be seen from the above that $1 - K^* > 0$ if $s_i > 0$. This implies that conflicting evidence can always be combined by means of the modified rule and this rule offers a way to handle large disagreement between bodies of evidence.

Example 6 *A system consists of four components ($U = \{1,2,3,4\}$) (see Example 1). After a system failure, 200 sensors (the first source: $N_1 = 200$, $c_1^{(1)} = 200$, $m_1(A_1^{(1)}) = 1$) indicate that the first component failed ($A_1^{(1)} = \{1\}$), 3 sensors (the second source: $N_2 = 3$, $c_1^{(2)} = 3$, $m_2(A_1^{(2)}) = 1$) indicate that the second component failed ($A_1^{(2)} = \{2\}$). If we assume that $s_1 = s_2 = 0$, then $K = 1$ (evidence are conflicting). The combined beliefs can not be calculated despite the fact that the first source provides the large number (200) of observations in comparison with the second one. If we assume that $s_1 = s_2 = 1$, then $\varkappa_1 = 200/201$, $\varkappa_2 = 3/4$,*

$$1 - K^* = 1 - \varkappa_1\varkappa_2 = 0.2537,$$

$$m_{12}^*(A_1^{(1)}) = \varkappa_1(1 - \varkappa_2)m_1(A_1^{(1)})/0.2537 = 0.98,$$
$$m_{12}^*(A_1^{(2)}) = \varkappa_2(1 - \varkappa_1)m_2(A_1^{(2)})/0.2537 = 0.015,$$
$$m_{12}^*(U) = (1 - \varkappa_1)(1 - \varkappa_2)/0.2537 = 0.005.$$

If $A = A_1^{(1)} = \{1\}$, then

$$\underline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), 1) = m_{12}^*(A_1^{(1)}) = 0.98,$$

$$\overline{P}(A|(\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), 1) = m_{12}^*(A_1^{(1)}) + m_{12}^*(U) = 0.985.$$

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 $t(U) = 1$, we get

$$\underline{F}(x|\mathbf{c}, s) = \underline{P}(\{u \leq x\}|\mathbf{c}, s) = \begin{cases} C\sum_{i:\sup A_i \leq x} c_i, & x < \sup U \\ 1, & x = \sup U \end{cases},$$

$$\overline{F}(x|\mathbf{c}, s) = \overline{P}(\{u \leq x\}|\mathbf{c}, s) = \begin{cases} C\left(s + \sum_{i:\inf A_i \leq x} c_i\right), & x > \inf U \\ 0, & x = \inf U \end{cases},$$

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

These expressions coincide with similar ones given in [4] if we set $s = 0$. 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.

The lower and upper probability distributions can be considered in a framework of the so-called p-boxes [12]. A probability box (p-box) is a class of distribution functions delimited by some upper and lower bounds which collectively represent the epistemic uncertainty about the distribution function of a random variable. As indicated in [12], p-boxes are a somewhat coarser way to describe uncertainty than are Dempster-Shafer structures on the real line. Every Dempster-Shafer structure specifies a unique p-box and every p-box specifies an equivalence class of Dempster-Shafer structures.

If X is a continuous random variable, then

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

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

Similar expressions can be obtained for the case of the discrete random variable X.

Note that the case $s = 0$ can be found in [9]:

$$\underline{\mathbb{E}}_0 X = \sum_{i=1}^{n} m(A_i) \cdot \inf A_i, \quad \overline{\mathbb{E}}_0 X = \sum_{i=1}^{n} m(A_i) \cdot \sup A_i.$$

Example 7 *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

The extension of belief and plausibility measures when there is little information about elements of the frame of discernment has been proposed in the paper. However, this extension can not be called a new one. One can see that it is equivalent to the well-known discounting scheme [21] introduced to handle conflict of information sources. The extension provides the same lower and upper probabilities as the ε-contaminated model [14]. Nevertheless, it is often difficult to determine the values of ε in the ε-contaminated model or discount rates in the discounting scheme. In contrast, the extended belief and plausibility functions are strongly defined by the number of observations or judgments N. They can be simply updated after observing new events or obtaining new expert judgments because Dirichlet prior distributions generate Dirichlet posterior distributions. Moreover, they may be more realistic in many applications and possible large imprecision of results reflects insufficiency of available information. The extended belief and plausibility functions (possibility distributions) are again belief and plausibility functions (possibility distributions) and, therefore, they preserve all their properties. 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 [17,25,27]. It is worth noting that the Dirichlet distribution should be regarded as one of the possible multinomial models that can be applied here. Therefore, another direction for further work could be possible extensions of belief and plausibility measures with using multinomial probability distributions different from the Dirichlet distribution.

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 Reasoning in Uncertainty Quantification”. I would like to express my appreciation to the anonymous referees whose very valuable comments have improved the paper.

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. J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 2005. To appear.
4. 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.
5. J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley, Chichester, 1994.
6. 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.
7. G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1-2):75-125, 2004.
8. M. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.
9. A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annales of Mathematical Statistics*, 38:325–339, 1967.
10. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
11. D. Dubois and H. Prade. Possibility theory and data fusion in poorly informed environments. *Control Engineering Practice*, 2(5):811–823, 1994.
12. 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.

13. J.Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.
14. P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.
15. M. Hutter. Robust estimators under 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 274–289, Lugano, Switzerland, July 2003. Carleton Scientific.
16. G.J. Klir and M.J. Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory*. Physica-Verlag, Heidelberg, 1998.
17. V. P. Kuznetsov. *Interval Statistical Models*. Radio and Communication, Moscow, 1991. in Russian.
18. 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.
19. M. Ramoni and P. Sebastiani. Robust learning with missing data. *Machine Learning*, 45(2):147–170, 2001.
20. K. Sentz and S. Ferson. Combination of evidence in Dempster-Shafer theory. Report SAND2002-0835, Sandia National Laboratories, April 2002.
21. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
22. 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.
23. 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.
24. 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, 2000.
25. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
26. P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996. with discussion.
27. K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung*, volume I Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg, 2001.
28. S. Wilks. *Mathematical Statistics*. Wiley and Sons, New York, 1962.
29. M. Zaffalon. Exact credal treatment of missing data. *Journal of Statistical Planning and Inference*, 105(1):105–122, 2002.