Second-order uncertainty calculations by using the imprecise Dirichlet model

Lev V. Utkin
Department of Computer Science
St.Petersburg State Forest Technical Academy
Institutski per. 5, 194021, St.Petersburg, Russia
e-mail: lev.utkin@mail.ru

Abstract

Natural extension is a powerful tool for combining the expert judgments in the framework of imprecise probability theory. However, it assumes that every judgment is "true" and this fact leads to some difficulties in many applications. Therefore, a second-order uncertainty model is considered in the paper where probabilities on the second-order level are taken by using the imprecise Dirichlet model. The approach proposed in the paper is illustrated by application and auxiliary examples.

Keywords: Dirichlet distribution, expert judgments, interval-valued expectations, decision making, reliability.

1 Introduction

Judgments elicited from human experts may be a very important part of information in various applications. Several methods for elicitation, assessment and pooling of this type of information have been proposed by many authors [3, 8]. In order to get useful information from the experts, a proper uncertainty modeling of pieces of data supplied by experts has to be used. Judgments elicited from experts are usually imprecise due to the limited precision of human assessments. When several experts supply judgments or assessments, their responses are pooled so as to derive a single measure. One of the powerful and mathematically strong tools for combining the expert judgments is the imprecise probability theory (also called theory of lower previsions [23, 25], theory of interval statistical models [11], theory of interval probabilities [27]), whose general framework is provided by upper and lower previsions (expectations). In order to combine the expert judgments in the framework of imprecise probabilities, a general procedure called natural extension is used. The natural extension can be viewed as an optimization problem. It produces lower and upper bounds for a new aggregated measure.

At the same time, judgments elicited from experts are usually unreliable. Judgments of reliable experts should be more important than those of unreliable ones. However, the natural extension in its standard form [23] assumes that all judgments are of the same quality. Moreover, it assumes that all judgments are "true". This is a shortcoming of the natural extension in some cases. Models of aggregating expert judgments taking into account the quality of experts can be considered in a framework of *hierarchical uncertainty models* (second-order models), which are rather common in uncertainty theory. Different application examples and a comprehensive review of hierarchical models can be found in [4, 5]. These models assume that there exist second-order probabilities, which can be regarded as the measures of the judgment quality. Hierarchical uncertainty models for aggregating expert judgment based on the imprecise probabilities have been considered by Troffaes and de Cooman [15] and by Utkin [16, 17, 18, 21]. These models assume the lack of information about probability distributions on the first and second levels of the hierarchy. However, their usage is restricted by difficulties in finding the second-order probabilities under condition that the provided judgments are different in kind, i.e., they represent different probabilistic measures, for instance, probabilities, expectations, etc. How to estimate the quality of experts in this case? One of the approaches for computing the second-order probabilities has been proposed by Utkin [19], where it is shown that weights of experts as measures of their quality widely used in the literature [3, 8, 28] can not be measures of the quality of provided judgments if these judgments are imprecise, for instance, interval-valued. This approach uses the so-called imprecise Dirichlet model [24], which can roughly be viewed as a set of Dirichlet distributions. However, the aggregation of expert judgments different in kind is considered in the work [19] very briefly and mainly as a problem for further research. Therefore, the present paper discusses in detail this approach and illustrates it by various application and auxiliary examples.

Roughly, the main idea of the approach is that every judgment in the form of the interval-valued expectation (prevision) can be represented as a part of the unit simplex whose every point is some probability distribution. If we assume that some unknown "true" distribution is random itself and every judgment is an outcome of an experiment, then probabilities that judgments cover this "true" distribution (second-order probabilities of judgments) can be described by the Dirichlet distribution. To take into account the limited number of judgments, we use Walley's imprecise Dirichlet model to make cautious inference. It is shown in the paper that this approach can be regarded as a parametric extension of the belief and plausibility functions from Dempster-Shafer theory [7, 13] and can be interpreted in terms of robust models (ε-contaminated models) [10]. The approach proposed in the paper can be regarded as a combination of the models proposed in [17] and [19].

2 Some preliminary definitions

2.1 Imprecise probabilities

Suppose that there is a continuous random variable X defined on the sample space Ω and information about this variable is represented as a set of N interval-valued expectations of functions $h_1(X),...,h_N(X)$. Denote these lower and upper expectations $\underline{\mathbb{E}}h_i$ and $\overline{\mathbb{E}}h_i$, $i=1,...,N$. In terms of the theory of imprecise probabilities the corresponding functions $h_i(x)$ and interval-valued expectations $\underline{\mathbb{E}}h_i$ and $\overline{\mathbb{E}}h_i$, $i=1,...,N$, are called *gambles* and *lower and upper previsions*, respectively. The lower and upper previsions $\underline{\mathbb{E}}h_i$ and $\overline{\mathbb{E}}h_i$ can be viewed as bounds for an unknown precise prevision $\mathbb{E}h_i$ which will be called a *linear prevision*. Various types of information can be modelled by means of lower and upper previsions. For example, if h_i is the indicator function of an event A, $I_A(X)$, then previsions $\underline{\mathbb{E}}h_i$ and $\overline{\mathbb{E}}h_i$ can be regarded as lower and upper probabilities of the event A. If $h_i(X)=X$, then $\underline{\mathbb{E}}X$ and $\overline{\mathbb{E}}X$ are bounds for the mean value of the corresponding random variable.

For computing new previsions $\underline{\mathbb{E}}g$ and $\overline{\mathbb{E}}g$ of a gamble $g(X)$ from the available information, the *natural extension* can be used in the following primal form:

$$\underline{\mathbb{E}}g=\min_{\mathcal{A}}\sum_{x\in\Omega}g(x)p(x),\ \overline{\mathbb{E}}g=\max_{\mathcal{A}}\sum_{x\in\Omega}g(x)p(x), \quad (1)$$

subject to

$$\sum_{x\in\Omega} p(x) = 1,\ p(x) \geq 0,$$

$$\underline{\mathbb{E}}h_i \leq \sum_{x\in\Omega} h_i(x)p(x) \leq \overline{\mathbb{E}}h_i,\ i = 1, ..., N. \tag{2}$$

Here the minimum and maximum are taken over the set $\mathcal{A}$ of all possible probability mass functions $\{p(x)\}$ satisfying conditions (2). Solutions for problems (1)-(2) exist if all constraints (2) form a non-empty subset $\mathcal{A}_0 \subseteq \mathcal{A}$. If subset $\mathcal{A}_0$ is empty, this means that the set of evidence is conflicting.

It should be noted that problems (1)-(2) are linear and the dual optimization problems can be written as follows [11]:

$$\overline{\mathbb{E}}g = \min_{c_0,c_i,d_i} \left(c_0 + \sum_{i=1}^{N} (c_i \overline{a}_i - d_i \underline{a}_i) \right),\ \underline{\mathbb{E}}g = -\overline{\mathbb{E}}(-g), \tag{3}$$

subject to $c_i,\ d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1, ..., N$, and $\forall x \in \Omega$,

$$c_0 + \sum_{i=1}^{N} (c_i - d_i)\, h_i(x) \geq g(x). \tag{4}$$

If the considered random variables are continuous, then problems (3)-(4) are not changed. In problems (1)-(2), sums and probability mass functions are replaced by integrals and densities, respectively.

Natural extension is a general mathematical procedure for calculating new previsions from initial judgements. It produces a coherent overall model from a certain collection of imprecise probability judgements and may be seen as the basic constructive step in interval-valued statistical reasoning. At the same time, it realizes too strong inference from initial data because it assumes that all judgments involved are "true", i.e., they have unitary probabilities. As a result, we might have too precise results or just lack of solutions when judgments are inconsistent. This is illustrated by the following examples.

Example 1 *Suppose two experts provide judgments* $p \leq 0.1$ *and* $p \geq 0.1$ *about probabilities of rainy weather tomorrow. By using (1)-(2), we get the combined probability* $p = 0.1$. *It is difficult to believe that two quite different and rather imprecise judgments produce absolutely precise probability.*

Example 2 *Let us consider a special case of three judgments*

$$p_3 \geq 0.7,\ p_2 \geq 0.5,\ 1 \leq \mathbb{E}X \leq 2$$

about a discrete random variable X *taking three values with unknown probabilities* $p = (p_1, p_2, p_3)$. *Here* $N = 3$, $\Omega = \{x_1, x_2, x_3\}$. *The set of constraints corresponding to these judgments is*

$$\begin{array}{c} 0 \leq p_1 \leq 1 \\ 0.5 \leq p_2 \leq 1 \\ 0.7 \leq p_3 \leq 1 \\ 1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2 \\ p_1 + p_2 + p_3 = 1 \end{array}.$$

It is impossible to infer new judgments from the available ones because the above inequalities are inconsistent.

2.2 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 identical probability distributions $\Pr\{u_j\} = \theta_j$ 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 [6]

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

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 the 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 $t_j^* = (n_j + st_j)/(N + s)$.

Walley [24] 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 [24] 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 [24] 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, 14, 24]. Most authors propose to take $s = 1$ or $s = 2$. At the same time, we have to point out that the proper choice of the hyperparameter s is an open question and further work is needed to find some strong rules for determining s.

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 [24], 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 [2] aims at modeling prior ignorance about the chances θ by characterizing prior uncertainty by a single prior probability distribution.

3 Second-order probabilities of judgments

We have pointed out in the previous sections some difficulties that can be met by using the natural extension. Moreover, we have conjectured that the main reason of the difficulties is unitary probabilities of judgments. Therefore, we consider further how to assign probabilities to judgments of the form (2).

3.1 Sets of multinomial models produced by judgments

Suppose that we have N judgments of the form

$$\underline{\mathbb{E}}h_k \leq \mathbb{E}h_k \leq \overline{\mathbb{E}}h_k,\ k = 1, ..., N. \tag{5}$$

Here h_k is a function of the random variable X, $\underline{\mathbb{E}}h_k$ and $\overline{\mathbb{E}}h_k$ are lower and upper expectations of the function h_k. For example, the first expert gives an interval of a probability of an event (h_1 is the indicator function of the event) and the second expert provides an interval of the expectation of a random variable (h_2 is the linear function). When judgments are of the same type, the second-order model is significantly simplified and reduced to the model studied by Utkin [19].

Our aim in this section is to determine the lower and upper probabilities of these judgments.

Suppose for simplicity that the random variable X is discrete and takes L values. Then the set of all probability distributions of X can be represented as the L-dimensional unit simplex $S(1, L)$ having L vertices where every point is a probability distribution $p = (p_1, p_2, ..., p_L)$. At the same time, every judgment (5) can be written as follows:

$$\underline{\mathbb{E}}h_k \leq \sum_{i=1}^{L} h_k(x_i)p_i \leq \overline{\mathbb{E}}h_k,$$

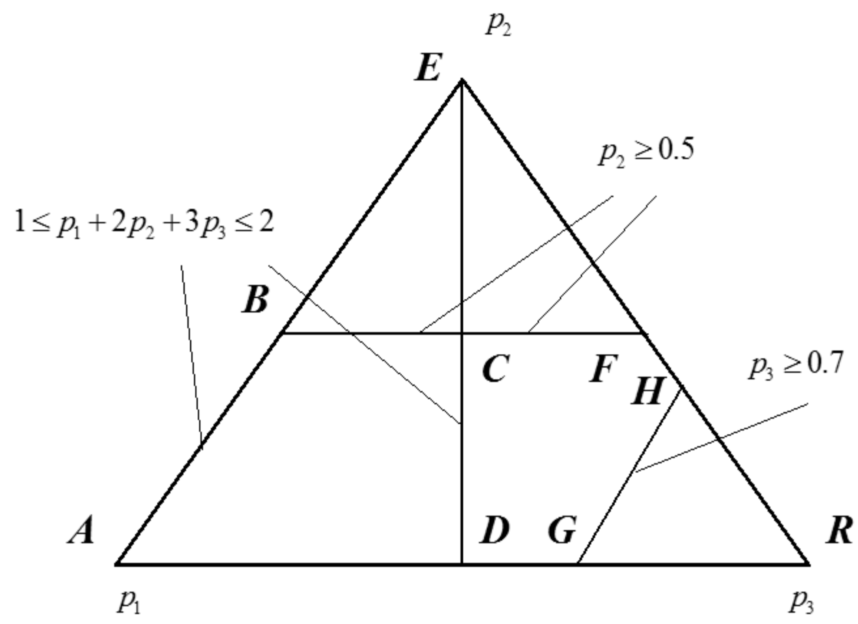

Figure 1: Illustration to Example 3

and, therefore, it produces some subset of the simplex in the form of a polyhedron which we denote by $\mathcal{A}_k$.

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 subset $\mathcal{B}_k$ $(k = 1, ..., 2^N)$ as follows:

$$\mathcal{B}_k = \left(\bigcap_{j:i_j=1} \mathcal{A}_j \right) \bigcap \left(\bigcap_{j:i_j=0} \mathcal{A}_j^c \right), \; i_j \in \mathbf{i}.$$

Here $\mathcal{A}_{n+1} = S(1, L)$ and $\mathcal{A}_j^c$ is the complement of $\mathcal{A}_j$ such that $\mathcal{A}_j^c \cup \mathcal{A}_j = S(1, L)$. As a result, we divide the simplex $S(1, L)$ into a set of non-intersecting subsets (polyhedrons) $\mathcal{B}_k$ such that $\mathcal{B}_1 \cup ... \cup \mathcal{B}_M = S(1, L)$, $M = 2^n$. Moreover, every subset $\mathcal{A}_i$ can be represented as the union of a finite number of subsets $\mathcal{B}_k$ such that $\mathcal{A}_i = \bigcup_{k \in J_i} \mathcal{B}_k$, where J_i is a set of indices.

Example 3 *Let us consider three judgments*

$$p_3 \geq 0.7, \; p_2 \geq 0.5, \; 1 \leq \mathbb{E}X \leq 2$$

about a discrete random variable X taking three values ($L = 3$) with unknown probabilities $p = (p_1, p_2, p_3)$. These judgments induce three subsets $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ of probability distributions which are depicted in Fig.1 in the form of polyhedrons GHR, BEF, AED, respectively. Then we can define subsets $\mathcal{B}_k$ as follows: $\mathcal{B}_1$ is the set of all points in the polyhedron BEC, i.e. $\mathcal{B}_1 = \mathcal{A}_1^c \cap \mathcal{A}_2 \cap \mathcal{A}_3$; $\mathcal{B}_2$ is the set of all points in the polyhedron $ABCD$, i.e. $\mathcal{B}_2 = \mathcal{A}_1^c \cap \mathcal{A}_2^c \cap \mathcal{A}_3$; $\mathcal{B}_3$ is the set of all points in the polyhedron CEF, i.e. $\mathcal{B}_3 = \mathcal{A}_1^c \cap \mathcal{A}_2 \cap \mathcal{A}_3^c$; $\mathcal{B}_4$ is the set of all points in the polyhedron GHR, i.e. $\mathcal{B}_4 = \mathcal{A}_1 \cap \mathcal{A}_2^c \cap \mathcal{A}_3^c$; $\mathcal{B}_5$ is the set of all points in the polyhedron $DCFHG$, i.e. $\mathcal{B}_5 = \mathcal{A}_1^c \cap \mathcal{A}_2^c \cap \mathcal{A}_3^c$. It should be noted that there hold $\mathcal{B}_1 \cup ... \cup \mathcal{B}_5 = S(1, 3)$, $\mathcal{A}_1 = \mathcal{B}_4$, $\mathcal{A}_2 = \mathcal{B}_1 \cup \mathcal{B}_3$, $\mathcal{A}_3 = \mathcal{B}_1 \cup \mathcal{B}_2$.

Remark 1 *It is necessary to point out that the number M of subsets $\mathcal{B}_k$ does not depend on L and depends only on the number N of available judgments. If the number N of judgments is finite, then M is also finite.*

Imagine that every judgment $\mathcal{A}_i$ is associated with a large box whose form is the corresponding polyhedron. This box contains one ball which can move inside the box and we do not know its location. Every subset $\mathcal{B}_k$ is associated with an urn. If we cover subsets $\mathcal{B}_k$ with indices $k \in J_i$ by the box $\mathcal{A}_i$, then the ball enters randomly in one of the urns $\mathcal{B}_k$. After covering the urns by all (N) boxes, N balls are in some urns. What can we say about possible numbers of balls in the urns? It is obvious that there exist different combinations of numbers of balls except the case when all sets J_i consist of one element. Suppose that the number of the possible combinations is Q. Denote the k-th possible vector of balls in urns by $\mathbf{n}^{(k)} = (n_1^{(k)}, ..., n_M^{(k)})$, $k = 1, ..., Q$, $\sum_{i=1}^{M} n_i^{(k)} = N$. If we assume that judgments are independent and a ball in the i-th urn has some unknown probability π_i, then every combination of balls in urns produces the *standard multinomial model.* Q possible combinations of balls produce Q equivalent standard multinomial models. The models are equivalent in the sense that we can not choose one of them as a more preferable case. So, the set of judgments (5) can be represented as a set of standard multinomial models.

Example 4 *In Example 3, the following vectors* $\mathbf{n}^{(k)} = (n_1^{(k)}, n_2^{(k)}, n_3^{(k)}, n_4^{(k)})$ *of balls correspond to three available judgments:*

$$(1, 0, 1, 0, 1),\ (1, 1, 0, 0, 1),\ (0, 1, 1, 0, 1),\ (0, 2, 0, 0, 1).$$

We have 4 *multinomial models here.*

For every model, the probability of an arbitrary event $\mathcal{A} \subseteq S(1, L)$ depends on $\mathbf{n}^{(k)}$, that is, we can find this probability: $P(\mathcal{A}|\mathbf{n}^{(k)})$. So far as all the models are equivalent, even by precise probabilities of all categories only lower and upper probabilities of $\mathcal{A}$ can be computed:

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

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

In particular, if all sets J_i consist of single elements, then $Q = 1$ and

$$\underline{P}(\mathcal{A}) = P(\mathcal{A}|\mathbf{n}^{(1)}),\ \ \overline{P}(\mathcal{A}) = P(\mathcal{A}|\mathbf{n}^{(1)}).$$

The next problem is to define $\mathbf{n}^{(k)}$ and $P(\mathcal{A}|\mathbf{n}^{(k)})$. In the case of multinomial samples, the Dirichlet distribution is the traditional choice. However, the number of judgments may be rather small. Therefore, we use the imprecise Dirichlet model.

3.2 Computing the probabilities $\underline{P}(\mathcal{A}, s)$ and $\overline{P}(\mathcal{A}, s)$

By returning to the multinomial models considered in the example with boxes and balls and using notations introduced for the imprecise Dirichlet model, we can assume that the sample space considered now is the finite set of subsets $\mathcal{B}_1, ...\mathcal{B}_M$, events are the subsets $\mathcal{A}_1, ..., \mathcal{A}_N$. Then the lower and upper probabilities of an event (subset of the simplex $S(1, L)$) $\mathcal{A}$, whose representation as the union of elements $\mathcal{B}_1, ...\mathcal{B}_M$ has indices from a set J, can be written as follows:

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

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

where

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

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

$$\underline{P}(\mathcal{A}, s) = \frac{\min\limits_{k=1,\dots,Q} n^{(k)}(\mathcal{A}) + s \cdot \inf\limits_{\alpha \in S(1,M)} \alpha(\mathcal{A})}{N + s},$$

$$\overline{P}(\mathcal{A}, s) = \frac{\max\limits_{k=1,\dots,Q} n^{(k)}(\mathcal{A}) + s \cdot \sup\limits_{\alpha \in S(1,M)} \alpha(\mathcal{A})}{N + s}.$$

Proposition 1 *Denote*

$$L_1(\mathcal{A}) = \min_{k=1,\dots,Q} n^{(k)}(\mathcal{A}) = \sum_{i:\mathcal{A}_i \subseteq \mathcal{A}} 1,$$

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

Then there hold

$$\underline{P}(\mathcal{A}, s) = \frac{L_1(\mathcal{A})}{N + s}, \quad \overline{P}(\mathcal{A}, s) = \frac{L_2(\mathcal{A}) + s}{N + s}.$$

Proof. Note that $\inf_{\alpha \in S(1,M)} \alpha(\mathcal{A})$ is achieved at $\alpha(\mathcal{A}) = 0$ and $\sup_{\alpha \in S(1,M)} \alpha(\mathcal{A})$ is achieved at $\alpha(\mathcal{A}) = 1$ except a case when $\mathcal{A} = S(1, L)$. If $\mathcal{A} = S(1, L)$, then $\alpha(\mathcal{A}) = 1$ for the minimum and maximum. In order to find the minimum and maximum of $n^{(k)}(\mathcal{A})$ we consider three types of subsets $\mathcal{A}_i^{(1)}$, $\mathcal{A}_j^{(2)}$, $\mathcal{A}_k^{(3)}$ produced by judgments such that $\mathcal{A}_i^{(1)} \subseteq \mathcal{A}$, $\mathcal{A}_j^{(2)} \cap \mathcal{A} = \varnothing$, $\mathcal{A}_k^{(3)} \cap \mathcal{A} \neq \varnothing$ and $\mathcal{A}_k^{(3)} \subsetneq \mathcal{A}$. It is obvious that balls corresponding to sets $\mathcal{A}_i^{(1)}$ belong to the set $\mathcal{A}$ and $n^{(k)}(\mathcal{A})$ can not be less than $\sum_{i:\mathcal{A}_i^{(1)} \subseteq \mathcal{A}} 1$. On the other hand, balls corresponding to sets $\mathcal{A}_j^{(2)}$ do not belong to $\mathcal{A}$. This implies that $n^{(k)}(\mathcal{A})$ can not be greater than $N - \sum_{i:\mathcal{A}_j^{(2)} \cap \mathcal{A} = \varnothing} 1$. Balls corresponding to $\mathcal{A}_k^{(3)}$ may belong to $\mathcal{A}$, but it is not necessary. Therefore, $\min n^{(k)}(\mathcal{A}) = \sum_{i:\mathcal{A}_i^{(1)} \subseteq \mathcal{A}} 1$ and $\max n^{(k)}(\mathcal{A}) = N - \sum_{i:\mathcal{A}_j^{(2)} \cap \mathcal{A} = \varnothing} 1$, as was to be proved. ■

If the set $\mathcal{A}$ coincides with the set $\mathcal{A}_k$, then $\underline{P}(\mathcal{A}_k, s)$ and $\overline{P}(\mathcal{A}_k, s)$ can be regarded as the lower and upper second-order probabilities of the k-th judgment [16, 17].

Remark 2 *It can be seen from the above expressions for $\underline{P}(\mathcal{A}, s)$ and $\overline{P}(\mathcal{A}, s)$ that we do not need to find subsets $\mathcal{B}_k$. Only conditions $\mathcal{A}_i \subseteq \mathcal{A}$ and $\mathcal{A}_i \cap \mathcal{A} = \varnothing$ are important.*

Remark 3 *It is noteworthy that the obtained probabilities can be regarded as some extension of the* belief $Bel(\mathcal{A})$ *and* plausibility $Pl(\mathcal{A})$ functions *of $\mathcal{A}$ in the framework of* random set theory *[7, 13]. According to [20], there hold*

$$\underline{P}(\mathcal{A}, s) = \frac{N \cdot Bel(\mathcal{A})}{N + s}, \quad \overline{P}(\mathcal{A}, s) = \frac{N \cdot Pl(\mathcal{A}) + s}{N + s}.$$

At that, if $s = 0$, then $\underline{P}(\mathcal{A}, 0) = Bel(A)$, $\overline{P}(\mathcal{A}, 0) = Pl(A)$. However, by using the probabilistic interpretation of the belief and plausibility functions [9], we need to have a lot of judgments in *order to compute the so-called* basic probability assignments *for events and this is just unrealistic in many applications. At the same time, by increasing the hyperparameter s, we can make the results to be more cautious and the possible large imprecision of results reflects insufficiency of available information. A detailed discussion of the extended belief and plausibility functions can be found in [20].*

Remark 4 *The obtained probabilities also can be considered in the framework of an ε-contaminated model [10]. As pointed out by Seidenfeld and Wasserman [24] in the discussion part of the Walley's paper [24], the imprecise Dirichlet model 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 is the set $\{(1-\varepsilon)P+\varepsilon Q : Q \text{ is arbitrary}\}$). Here $P = n(A)/N$ and $\varepsilon = s/(N+s)$. This implies that the lower and upper probabilities of judgments can be regarded as extended (contaminated) probabilities. On the one hand, this representation of the probabilities has a clear explanation and allows us to control the bounds by changing the value of ε. On the other hand, it has a shortcoming. As new judgments are obtained, the value of ε has to be changed and it is difficult to define rules for doing that. As pointed out by Walley [24], s must not depend on the number of judgments. This implies that the hyperparameter s remains without changes and the lower and upper probabilities depend only on N.*

Example 5 *Let us find the probabilities of judgments given in Example 3. If can be seen from Fig.1 that $\mathcal{A}_1 \subseteq \mathcal{A}_1$, $\mathcal{A}_2 \subseteq \mathcal{A}_2$, $\mathcal{A}_3 \subseteq \mathcal{A}_3$, $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$, $\mathcal{A}_1 \cap \mathcal{A}_3 = \varnothing$. This implies that $L_1(\mathcal{A}_1) = 1$, $L_1(\mathcal{A}_2) = 1$, $L_1(\mathcal{A}_3) = 1$, $L_2(\mathcal{A}_1) = 3-2 = 1$, $L_2(\mathcal{A}_2) = 3-1 = 2$, $L_2(\mathcal{A}_3) = 3-1 = 2$, and*

$$\underline{P}(\mathcal{A}_1, s) = \frac{1}{3+s},\ \overline{P}(\mathcal{A}_1, s) = \frac{1+s}{3+s},$$

$$\underline{P}(\mathcal{A}_2, s) = \frac{1}{3+s},\ \overline{P}(\mathcal{A}_2, s) = \frac{2+s}{3+s},$$

$$\underline{P}(\mathcal{A}_3, s) = \frac{1}{3+s},\ \overline{P}(\mathcal{A}_3, s) = \frac{2+s}{3+s}.$$

For instance, if we take $s = 1$, then

$$\underline{P}(\mathcal{A}_1, s) = \frac{1}{4},\ \overline{P}(\mathcal{A}_1, s) = \frac{1}{2},$$

$$\underline{P}(\mathcal{A}_2, s) = \underline{P}(\mathcal{A}_3, s) = \frac{1}{4},\ \overline{P}(\mathcal{A}_2, s) = \overline{P}(\mathcal{A}_3, s) = \frac{3}{4}.$$

3.3 Conditions $\mathcal{A}_i \subseteq \mathcal{A}$ and $\mathcal{A}_i \cap \mathcal{A} = \varnothing$

In order to compute the lower and upper second-order probabilities of judgments, we have to find algorithms for testing conditions $\mathcal{A}_i \subseteq \mathcal{A}$ and $\mathcal{A}_i \cap \mathcal{A} = \varnothing$. The first condition means that the set $\mathcal{A}_i$ of probability distributions produced by the constraint $\underline{\mathbb{E}}h_i \leq \mathbb{E}h_i \leq \overline{\mathbb{E}}h_i$ is a part of a set $\mathcal{A}$ of probability distributions produced by the constraint $\underline{\mathbb{E}}h \leq \mathbb{E}h \leq \overline{\mathbb{E}}h$. The second condition means that sets $\mathcal{A}_i$ and $\mathcal{A}$ produced by the i-th judgment and $\underline{\mathbb{E}}h \leq \mathbb{E}h \leq \overline{\mathbb{E}}h$ do not have common distributions.

Proposition 2 *Let $C = [\underline{c}, \overline{c}]$ be an interval such that $\underline{c} = \inf_{\mathcal{A}_i} \mathbb{E}h$, $\overline{c} = \sup_{\mathcal{A}_i} \mathbb{E}h$. Then $\mathcal{A}_i \subseteq \mathcal{A}$ if $C \subseteq [\underline{\mathbb{E}}h, \overline{\mathbb{E}}h]$.*

Proof. It is obvious that if $\underline{\mathbb{E}}h$ is achieved at a distribution from $\mathcal{A}_i$, then $\underline{c} = \underline{\mathbb{E}}h$. If $\underline{\mathbb{E}}h$ is achieved at a distribution from $\mathcal{A}\backslash\mathcal{A}_i$, then $\underline{c} \geq \underline{\mathbb{E}}h$. The same can be said for the upper bounds. ■

Proposition 2 implies that for satisfying the first condition $\mathcal{A}_i \subseteq \mathcal{A}$, we have to extend $\underline{\mathbb{E}}h_i$ and $\overline{\mathbb{E}}h_i$ on the bounds $\underline{c}$ and $\overline{c}$ of the linear expectation $\mathbb{E}h$ by means of the natural extension [11, 23, 27]

$$\underline{c} = \inf_{p\in S(1,L)} \mathbb{E}h, \ \overline{c} = \sup_{p\in S(1,L)} \mathbb{E}h,$$

subject to $\underline{\mathbb{E}}h_i \leq \mathbb{E}h_i \leq \overline{\mathbb{E}}h_i$.

These linear optimization problems can be written in the dual form:

$$\underline{c} = \sup \left(b_0 + b_1\underline{\mathbb{E}}h_i - d_1\overline{\mathbb{E}}h_i\right)$$

subject to $b_0 \in \mathbb{R}$, $b_1, d_1 \in \mathbb{R}_+$, and $\forall x_j$, $j = 1, ..., L$,

$$b_0 + b_1 h_i(x_j) - d_1 h_i(x_j) \leq h(x_j),$$

and

$$\overline{c} = \inf \left(b_0 + b_1\overline{\mathbb{E}}h_i - d_1\underline{\mathbb{E}}h_i\right)$$

subject to $b_0 \in \mathbb{R}$, $b_1, d_1 \in \mathbb{R}_+$, and $\forall x_j$, $j = 1, ..., L$,

$$b_0 + b_1 h_i(x_j) - d_1 h_i(x_j) \geq h(x_j),$$

Proposition 3 *If the linear programming problem with an arbitrary objective function and constraints*

$$\underline{\mathbb{E}}h_i \leq \mathbb{E}h_i \leq \overline{\mathbb{E}}h_i, \ \underline{\mathbb{E}}h \leq \mathbb{E}h \leq \overline{\mathbb{E}}h$$

does not have any solution, then there holds $\mathcal{A}_i \cap \mathcal{A} = \varnothing$.

Proof. The proof is obvious since the optimization problem does not have any solution if the sets of probabilities produced by the above constraints are not intersecting. ■

Example 6 *Let us test the condition* $\mathcal{A}_2 \subseteq \mathcal{A}_3$ *in Example 3. Here* $[\underline{\mathbb{E}}h_3, \overline{\mathbb{E}}h_3] = [1, 2]$, $[\underline{\mathbb{E}}h_2, \overline{\mathbb{E}}h_2] = [0.5, 1]$ *and*

$$\underline{c} = \inf_{p\in S(1,3)} (1p_1 + 2p_2 + 3p_3), \ \overline{c} = \sup_{p\in S(1,3)} (1p_1 + 2p_2 + 3p_3),$$

subject to

$$\begin{array}{c} 0 \leq p_1 \leq 1 \\ 0 \leq p_2 \leq 1 \\ 0.5 \leq p_3 \leq 1 \\ p_1 + p_2 + p_3 = 1 \end{array}.$$

Hence there hold $\underline{c} = 2$ *and* $\overline{c} = 3$. *Therefore,* $[2, 3] \subsetneq [1, 2]$ *and* $\mathcal{A}_2 \subsetneq \mathcal{A}_3$. *This also can be seen from Fig.1.*

What do the obtained second-order probabilities mean? Several interpretations of the probabilities exist [4, 5, 12, 15, 25, 26]. In our case the second-order probabilities can measure our prior beliefs that the given interval-valued judgments cover some unknown "true" value of the corresponding statistical characteristics. For instance, the "true" value of the expectation of the random variable considered in Example 3 is in the interval $[1, 2]$ with the probability larger 0.25. The second-order probabilities can geometrically be depicted as some "mountains" over subsets of the simplex produced by expert judgments. These mountains are none other than a part of the set of probability distributions of the first-order "true" statistical measure defined on the sample space $S(1,L)$. Moreover, it should be noted that many authors assume that the second-order probabilities are *subjective* ones given by some subject called the *modeller* [15] who is over *experts* providing the first-order information. The model studied here is quite different. It considers *objective* second-order probabilities obtained as a result of statistical inference on the basis of the expert first-order judgments.

4 Computing "average" bounds for expectations

We have now the set of judgments in the form of first-order expectations (5) and the corresponding lower and upper second-order probabilities of the judgments. This information can formally be written as

$$\Pr\left\{\underline{\mathbb{E}}h_k \leq \mathbb{E}h_k \leq \overline{\mathbb{E}}h_k\right\} \in [\underline{P}(\mathcal{A}_k,s),\overline{P}(\mathcal{A}_k,s)],\ k=1,...,N.$$

For computing the lower and upper bounds for the expectation $\mathbb{E}g$, an algorithm proposed by Utkin [17] will be used. The main idea of this algorithm is that upper and lower second-order probabilities are represented in turn as the lower and upper expectations (previsions) of the indicator function $\mathbf{1}_{[\underline{\mathbb{E}}h_k,\overline{\mathbb{E}}h_k]}(\mathbb{E}h_k)$ which takes the value 1 if $\mathbb{E}h_k \in [\underline{\mathbb{E}}h_k,\overline{\mathbb{E}}h_k]$ and 0 otherwise. At that, every linear expectation $\mathbb{E}h_k$ is regarded as a random variable z_k taking values from the set $[\inf h_k, \sup h_k]$ such that N variables have a unknown joint density $\Psi(z_1,...,z_m)$. Then for computing the lower and upper expected values of a new linear expectation $\mathbb{E}g$ (for reducing the second-order model to the first-order one), the following linear optimization problems (natural extension) have to be solved:

$$\underline{\mathbb{E}}g\ (\overline{\mathbb{E}}g) = \min_{\Psi}\left(\max_{\Psi}\right)\int_{\Theta^N} g(z_1,..,z_N)\Psi(z_1,..,z_N)\mathrm{d}z_1\cdots\mathrm{d}z_N,$$

subject to

$$\underline{P}(\mathcal{A}_k,s) \leq \int_{\Theta^N}\mathbf{1}_{[\underline{\mathbb{E}}h_k,\overline{\mathbb{E}}h_k]}(\mathbb{E}h_k)\Psi(z_1,..,z_N)\mathrm{d}z_1\cdots\mathrm{d}z_N \leq \overline{P}(\mathcal{A}_k,s),\ k\leq N.$$

Here $\Psi(z_1,...,z_m)$ is the optimization variable, Θ^N is the set of all values of the vector $(z_1,...,z_N)$.

What does the new linear expectation $\mathbb{E}g$ mean? Our aim is to find some statistical characteristic of the random variable we have information in the form of judgments (5) about. This characteristic can be regarded as a new interval-valued judgment with some bounds. By taking certain bounds for the characteristic, we can find the probability of this new judgment. By changing the bounds, we get infinitely many probabilities that produce some set of probability distributions. The lower and upper expectations of these distributions are nothing else but "average" values of bounds for the statistical characteristic, which in turn are $\underline{\mathbb{E}}g$ and $\overline{\mathbb{E}}g$. For example, if we are interesting to find the value of the probability distribution at point t, then $\mathbb{E}g = \mathbb{E}\mathbf{1}_{(-\infty,t]}(z)$. Of course, we could find the probabilities $\underline{P}(\mathcal{A},s)$ and $\overline{P}(\mathcal{A},s)$ for different intervals $\underline{\mathbb{E}}g$ and $\overline{\mathbb{E}}g$ to construct the second-order distributions of $\mathbb{E}g$ by means of the approach proposed in the previous section and to compute lower and upper expectations for these distributions. However, in this case we had to solve infinitely many optimization problems. Another way is to consider dual optimization problems for computing $\underline{\mathbb{E}}g$ and $\overline{\mathbb{E}}g$. It turns out that this way is much simpler in spite of the apparent complexity of the dual optimization problems. The detailed description and justification of the algorithm can be found in [17]; only its scheme is given here.

Step 1. Let J be a subset of $\{1,2,...,N\}$ and $Y_J = (y_1,...,y_N)$ be a binary vector ($y_i \in \{0,1\}$) such that $y_i = 1$ if $i \in J$ and $y_i = 0$ if $i \notin J$. Note that the subset J can be empty.

Step 2. Let

$$\mathcal{R}_J = \left(\bigcap_{i\in J}\mathcal{A}_i\right)\bigcap\left(\bigcap_{i\notin J}\mathcal{A}_i^c\right).$$

Step 3. $\underline{\mathbb{E}}g$ is computed by solving the following linear optimization problem

$$\underline{\mathbb{E}}g = \max_{c_0,c_i,d_i}\left\{c_0 + \sum_{i=1}^{N}\left(c_i\cdot\underline{P}(\mathcal{A}_i,s) - d_i\cdot\overline{P}(\mathcal{A}_i,s)\right)\right\}, \tag{6}$$

subject to c_i, $d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1,...,N$, $\forall\mathcal{R}_J \neq \varnothing$

$$c_0 + \sum_{i\in J}(c_i - d_i) \leq \min_{\mathcal{R}_J}\mathbb{E}g. \tag{7}$$

Here $\min_{\mathcal{R}_J}\mathbb{E}g$ is defined as follows:

$$\min_{\mathcal{R}_J}\mathbb{E}g = \min_p\sum_{i=1}^{L}g(x_i)p_i \tag{8}$$

subject to $p_1 + ... + p_L = 1$ and

$$\sum_{i=1}^{L}h_k(x_i)p_i \in [\underline{\mathbb{E}}h_k, \overline{\mathbb{E}}h_k],\ \ k\in J, \tag{9}$$

$$\sum_{i=1}^{L}h_k(x_i)p_i \in [\underline{\mathbb{E}}h_k, \overline{\mathbb{E}}h_k]^c,\ \ k\notin J. \tag{10}$$

Step 4. $\overline{\mathbb{E}}g$ is computed by solving the following linear optimization problem

$$\overline{\mathbb{E}}g = \min_{c_0,c_i,d_i}\left\{c_0 + \sum_{i=1}^{N}\left(c_i\cdot\overline{P}(\mathcal{A}_i,s) - d_i\cdot\underline{P}(\mathcal{A}_i,s)\right)\right\}, \tag{11}$$

subject to c_i, $d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1,...,N$, $\forall\mathcal{R}_J \neq \varnothing$

$$c_0 + \sum_{i\in J}(c_i - d_i) \geq \max_{\mathcal{R}_J}\mathbb{E}g. \tag{12}$$

Here $\max_{\mathcal{R}_J}\mathbb{E}g$ is defined by solving optimization problem (8)-(10) if to replace the operation min by max.

The emptiness of the set $\mathcal{R}_J$ can be defined by solving problem (8)-(10). If the problem does not have any solution, then $\mathcal{R}_J = \varnothing$ and the corresponding constraint is removed from the list of constraints (7) and (12).

Example 7 *Let us correct bounds for the probability p_3 on the basis of judgments given in Example 3 by taking into account that by $s = 1$, we have the second-order probabilities (see previous examples) $\underline{P}(\mathcal{A}_1,1) = 1/4$, $\overline{P}(\mathcal{A}_1,1) = 1/2$, $\underline{P}(\mathcal{A}_2,1) = 1/4$, $\overline{P}(\mathcal{A}_2,1) = 3/4$, $\underline{P}(\mathcal{A}_3,1) = 1/4$, $\overline{P}(\mathcal{A}_3,1) = 3/4$.*

Let $J = \{1,2,3\}$. Then

$$\min_{\mathcal{R}_J}\left(\max_{\mathcal{R}_J}\right)\mathbb{E}g = \min_p\left(\max_p\right)p_3$$

Table 1: Constraints to optimization problems by different sets J

J	Constraints	$\mathcal{R}_J$	$\min_{\mathcal{R}_J} \mathbb{E}g$	$\max_{\mathcal{R}_J} \mathbb{E}g$
$\{1,2\}$	$0.5 \leq p_2 \leq 1$ $0.7 \leq p_3 \leq 1$ $2 \leq 1p_1 + 2p_2 + 3p_3 \leq 3$	$\varnothing$		
$\{2,3\}$	$0.5 \leq p_2 \leq 1$ $0 \leq p_3 \leq 0.7$ $1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2$		0	0.25
$\{1,3\}$	$0 \leq p_2 \leq 0.5$ $0.7 \leq p_3 \leq 1$ $1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2$	$\varnothing$		
$\{1\}$	$0 \leq p_2 \leq 0.5$ $0.7 \leq p_3 \leq 1$ $2 \leq 1p_1 + 2p_2 + 3p_3 \leq 3$		0.7	1
$\{2\}$	$0.5 \leq p_2 \leq 1$ $0 \leq p_3 \leq 0.7$ $2 \leq 1p_1 + 2p_2 + 3p_3 \leq 3$		0	0.5
$\{3\}$	$0 \leq p_2 \leq 0.5$ $0 \leq p_3 \leq 0.7$ $1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2$		0	0.5
$\{\varnothing\}$	$0 \leq p_2 \leq 0.5$ $0 \leq p_3 \leq 0.7$ $2 \leq 1p_1 + 2p_2 + 3p_3 \leq 3$		0.25	0.7

subject to $p_1 + p_2 + p_3 = 1$ *and*

$$\begin{array}{c} 0.5 \leq p_2 \leq 1 \\ 0.7 \leq p_3 \leq 1 \\ 1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2 \end{array}.$$

The problems do not have any solution and $\mathcal{R}_{\{1,2,3\}} = \varnothing$.

Optimization problems by different sets J *have identical objective functions* p_3. *Their constraints and solutions are given in Table 1.*

Substituting the above results into (6)-(7), we can write

$$\underline{\mathbb{E}}g = \max_{c_0, c_i, d_i} \left\{ c_0 + 0.25c_1 - 0.5d_1 + 0.25c_2 - 0.75d_2 + 0.25c_3 - 0.75d_3 \right\},$$

subject to c_i, $d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1, 2, 3$,

$$\begin{aligned} c_0 + (c_2 - d_2) + (c_3 - d_3) &\leq 0 \\ c_0 + (c_1 - d_1) &\leq 0.7 \\ c_0 + (c_2 - d_2) &\leq 0 \\ c_0 + (c_3 - d_3) &\leq 0 \\ c_0 &\leq 0.25 \end{aligned}$$

Hence $\underline{\mathbb{E}}g = 0.175$. *The upper bound for* p_3 *can be found from the problem*

$$\overline{\mathbb{E}}g = \min_{c_0, c_i, d_i} \left\{ c_0 + 0.5c_1 - 0.25d_1 + 0.75c_2 - 0.25d_2 + 0.75c_3 - 0.25d_3 \right\},$$

Table 2: Values of the utility function u_{rj}

		States of nature			
		growth	medium growth	no change	low
Actions		x_1	x_2	x_3	x_4
bonds	a_1	6	9	9	8
stocks	a_2	12	7	3	−2
deposit	a_3	7	7	7	7

subject to $c_i,\ d_i \in \mathbb{R}^+,\ c_0 \in \mathbb{R},\ i = 1,2,3,$

$$\begin{aligned} c_0 + (c_2 - d_2) + (c_3 - d_3) &\geq 0.25 \\ c_0 + (c_1 - d_1) &\geq 1 \\ c_0 + (c_2 - d_2) &\geq 0.5 \\ c_0 + (c_3 - d_3) &\geq 0.5 \\ c_0 &\geq 0.7 \end{aligned}$$

Hence $\overline{\mathbb{E}}g = 0.75$.

Thus, the corrected interval-valued probability p_3 *is* $[0.175, 0.75]$. *Why is the corrected probability quite different from the initial one (*$[0.7, 1]$*)? The fact is that the first judgment concerning the probability* p_3 *contradicts with the second and third judgments. This fact is determined by the small interval-valued second-order probability (*$[0.25, 0.5]$*) of the first judgment. By using this second-order probability, we can say that the "complementary" judgment* $p_3 \in [0, 0.7]$ *has the larger interval-valued probability* $[0.5, 0.75]$. *It should be noted that the judgment* $p_3 \in [0, 0.7]$ *does not contradict with the second and third judgments. This implies that the largest part of the second-order probability distribution and the corresponding expectation are concentrated in the area from* 0 *to* 0.7. *That is why the corrected probability is quite different from the initial one.*

It should be noted that such the conclusion is based on the analysis of three initial judgments. Now suppose that we have 15 *judgments identical to the first one, i.e., there are* 17 *judgments (*15 *judgments produce the same subset* $\mathcal{A}_1$*). It follows from Proposition 1 that* $\underline{P}(\mathcal{A}_1, 1) = 15/18$, $\overline{P}(\mathcal{A}_1, 1) = 16/18$, $\underline{P}(\mathcal{A}_2, 1) = 1/18$, $\overline{P}(\mathcal{A}_2, 1) = 3/18$, $\underline{P}(\mathcal{A}_3, 1) = 1/18$, $\overline{P}(\mathcal{A}_3, 1) = 3/18$. *Hence* $\underline{\mathbb{E}}g = 0.583$, $\overline{\mathbb{E}}g = 0.944$. *One can see from these results that the new judgments supporting the first judgments lead to the larger interval-valued probability* p_3.

5 Application examples

5.1 Decision making

Consider the following investment decision-making example. The *states of nature* $\{x_1, ..., x_4\}$ are the states of economy during one year. The problem is to decide what action to take among three possible *courses of action* $(a_1, ..., a_3)$ with the given rates of return $u_{ij} = u(a_i, x_j)$ (*utility function*) as shown in Table 2.

The decision making assumes that an action k is optimal iff for all $r \in \{1, ..., 4\}$, $\mathbb{E}\mathbf{u}_k \geq \mathbb{E}\mathbf{u}_r$, where

$$\mathbb{E}\mathbf{u}_r = \sum_{i=1}^{4} u_{ri} p_i$$

Suppose that 6 experts supply the following judgments concerning the states of economy: two experts believe that the "medium growth" of economy is more probable than the "no change"; three experts suppose that the "low" is more probable than the "growth"; one expert supposes that the "no change" is more probable than the "medium growth". Here $N = 6$, $L = 4$ and the probability distribution of states of nature $(p_1, ..., p_4)$. The available judgments can formally be written as $p_2 > p_3$, $p_4 > p_1$, $p_3 > p_2$ or $p_2 - p_3 > 0$, $p_4 - p_1 > 0$, $p_3 - p_2 > 0$. It is obvious that the above judgments are contradictory and it is impossible to find a distribution function satisfying all the conditions. Therefore, we use the second-order probabilities. Define three subsets $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ of distributions $(p_1, ..., p_4)$ produced by every judgment, respectively. In order to find $L_1(\mathcal{A}_i)$ and $L_2(\mathcal{A}_i)$, we have to determine relations $\mathcal{A}_i \subseteq \mathcal{A}_j$ and $\mathcal{A}_i \cap \mathcal{A}_j = \varnothing$. It is obvious that $\mathcal{A}_1 \cap \mathcal{A}_3 = \varnothing$ because the first and the third types of judgments are contradictory. Therefore, $\mathcal{A}_1 \subsetneqq \mathcal{A}_3$ and $\mathcal{A}_3 \subsetneqq \mathcal{A}_1$. Let us test the condition $\mathcal{A}_1 \subseteq \mathcal{A}_2$. For doing it, we solve the linear optimization problems

$$\min(\max)\ \ (p_4 - p_1)$$

subject to $p_i \geq 0$, $i = 1, ..., 4$,

$$p_2 > p_3,$$
$$p_1 + p_2 + p_3 + p_4 = 1.$$

Hence $p_4 - p_1 \in (-1, 1)$. However, $p_2 - p_3 \in (0, 1)$ and $(-1, 1) \subsetneqq (0, 1)$. This implies that $\mathcal{A}_1 \subsetneqq \mathcal{A}_2$. Similarly, we get $\mathcal{A}_2 \subsetneqq \mathcal{A}_1$, $\mathcal{A}_2 \cap \mathcal{A}_1 \neq \varnothing$, $\mathcal{A}_3 \subsetneqq \mathcal{A}_2$, $\mathcal{A}_2 \subsetneqq \mathcal{A}_3$, $\mathcal{A}_2 \cap \mathcal{A}_3 \neq \varnothing$. Now we can compute

$$L_1(\mathcal{A}_1) = 2,\ L_2(\mathcal{A}_1) = 6 - 1 = 5,$$
$$L_1(\mathcal{A}_2) = 3,\ L_2(\mathcal{A}_2) = 6 - 0 = 6,$$
$$L_1(\mathcal{A}_3) = 1,\ L_2(\mathcal{A}_3) = 6 - 2 = 4.$$

If to take $s = 1$, then there hold

$$\underline{P}(\mathcal{A}_1, 1) = 2/7,\ \overline{P}(\mathcal{A}_1, 1) = 6/7,$$
$$\underline{P}(\mathcal{A}_2, 1) = 3/7,\ \overline{P}(\mathcal{A}_2, 1) = 1,$$
$$\underline{P}(\mathcal{A}_3, 1) = 1/7,\ \overline{P}(\mathcal{A}_3, 1) = 5/7.$$

The statistical characteristic g whose average value has to be found for decision making is the expected utility $\mathbb{E}\mathbf{u}_r$. This implies that $g = \mathbb{E}\mathbf{u}_r$. Now the criterion of decision making can be rewritten as follows. An action k is optimal iff for all $r \in \{1, ..., 4\}$, $\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_k \geq \underline{\mathbb{E}}\mathbb{E}\mathbf{u}_r$, where

$$\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_r = \min_{\mathcal{P}} \mathbb{E}\mathbf{u}_r,$$

where $\mathcal{P}$ is the set of second-order probability distributions.

A detailed analysis of the hierarchical decision making problem is given by Utkin and Augustin [22].

Let $J = \{1, 2, 3\}$. Then

$$\min_{\mathcal{R}_J} \left(\max_{\mathcal{R}_J} \right) \mathbb{E}\mathbf{u}_r = \min_{p} \left(\max_{p} \right) \sum_{i=1}^{4} u_{ri} p_i$$

subject to $p_1 + p_2 + p_3 = 1$ and

$$\begin{array}{c} p_2 - p_3 > 0 \\ p_4 - p_1 > 0 \\ p_3 - p_2 > 0 \\ 1 \leq 1p_1 + 2p_2 + 3p_3 \leq 2 \end{array}.$$

Table 3: Calculation results

J	$\underline{\mathbb{E}}\mathbf{u}_1$	$\underline{\mathbb{E}}\mathbf{u}_2$	$\underline{\mathbb{E}}\mathbf{u}_3$
$\{2,3\}$	7	-2	7
$\{1,2\}$	7	-2	7
$\{1\}$	6	5	7
$\{2\}$	7	-2	7
$\{3\}$	6	3	7
$\{\varnothing\}$	6	5	7

The problems do not have any solution and $\mathcal{R}_{\{1,2,3\}} = \varnothing$. Similarly, we can consider all possible sets J and find lower and upper bounds for $\mathbb{E}\mathbf{u}_r$ by taking into account that the complement of the set $\mathcal{A}_i$ produced by the inequality $p_j - p_k > 0$ is formed by the inequality $p_j - p_k \leq 0$. The calculation results corresponding to non-empty sets $\mathcal{R}_{\{\cdot\}}$ are shown in Table 3.

Substituting the above results into (6)-(7), we can write the optimization problem for computing the lower "average" expected utility of the first action

$$\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_1 = \max_{c_0,c_i,d_i} \left\{ c_0 + \frac{2}{7}c_1 - \frac{6}{7}d_1 + \frac{3}{7}c_2 - d_2 + \frac{1}{7}c_3 - \frac{5}{7}d_3 \right\},$$

subject to c_i, $d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1,2,3$,

$$\begin{aligned} c_0 + (c_2 - d_2) + (c_3 - d_3) &\leq 7 \\ c_0 + (c_1 - d_1) + (c_2 - d_2) &\leq 7 \\ c_0 + (c_1 - d_1) &\leq 6 \\ c_0 + (c_2 - d_2) &\leq 7 \\ c_0 + (c_3 - d_3) &\leq 6 \\ c_0 &\leq 6 \end{aligned}$$

Hence $\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_1 = 6.428$. Similarly, we get $\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_2 = -2$, $\underline{\mathbb{E}}\mathbb{E}\mathbf{u}_3 = 7$. The optimal action is the third one.

5.2 Reliability of a system

Suppose that 3 experts evaluate a system as follows:

1. the probability that the number of failures in a predefined period of time is between 0 and 10 is less than 0.4;
2. the probability that the number of failures in a predefined period of time is between 0 and 8 is less than 0.2;
3. the mean number of failures is between 0 and 2.

Let us find the "average" interval of the probability that the number of failures between 10 and 12 in the same time period. This example has been considered by Utkin [17] under condition of known probabilities of expert judgments. Here we use the proposed approach and find the interval of the probability under condition that the experts are unknown.

By denoting the random number of failures X, the above judgements can be written formally as follows:

$$0 \le \mathbb{E}I_{\{0,...,10\}}(X) \le 0.4,$$
$$0 \le \mathbb{E}I_{\{0,...,8\}}(X) \le 0.2,$$
$$0 \le \mathbb{E}X \le 2,$$
$$g = \mathbb{E}I_{\{10,11,12\}}(X).$$

Constraints corresponding to the above judgments are written as

$$0 \le \sum_{i=0}^{\infty} I_{\{0,...,10\}}(i)p_i \le 0.4,$$
$$0 \le \sum_{i=0}^{\infty} I_{\{0,...,8\}}(i)p_i \le 0.2,$$
$$0 \le \sum_{i=0}^{\infty} ip_i \le 2,$$
$$1 = \sum_{i=0}^{\infty} p_i \tag{13}$$

Let $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ be subsets produced by the judgments. By solving a set of linear programming problems, we get $\mathcal{A}_1 \cap \mathcal{A}_2 \neq \varnothing$, $\mathcal{A}_1 \cap \mathcal{A}_3 = \varnothing$, $\mathcal{A}_2 \cap \mathcal{A}_3 = \varnothing$, $\mathcal{A}_i \subsetneq \mathcal{A}_j$, for all i and j such that $i \neq j$. This implies that

$$L_1(\mathcal{A}_1) = 1, \ L_2(\mathcal{A}_1) = 3 - 1 = 2,$$
$$L_1(\mathcal{A}_2) = 1, \ L_2(\mathcal{A}_2) = 3 - 1 = 2,$$
$$L_1(\mathcal{A}_3) = 1, \ L_2(\mathcal{A}_3) = 3 - 2 = 1.$$

If to take $s = 1$, then there hold

$$\underline{P}(\mathcal{A}_1, 1) = 1/4, \ \overline{P}(\mathcal{A}_1, 1) = 3/4,$$
$$\underline{P}(\mathcal{A}_2, 1) = 1/4, \ \overline{P}(\mathcal{A}_2, 1) = 3/4,$$
$$\underline{P}(\mathcal{A}_3, 1) = 1/4, \ \overline{P}(\mathcal{A}_3, 1) = 2/4.$$

It should be noted that all linear programming problems are approximately solved with the restricted number (100) of variables.

Let $J = \{1, 2, 3\}$. Then

$$\min_{\mathcal{R}_J} \left(\max_{\mathcal{R}_J} \right) g = \min_{p} \left(\max_{p} \right) \sum_{i=0}^{\infty} I_{\{10,11,12\}}(i) \cdot p_i$$

subject to $p_i \ge 0$, $i = 0, ..., \infty$, and (13).

The problem does not have any solution. All cases of non-empty sets $\mathcal{R}_J$ are shown in Table 4.

Let us find the "average" interval of the prevision $\mathbb{E}I_{\{10,11,12\}}(X)$. By using the preliminary results, we get the following optimization problem for computing the upper probability that the number of failures between 10 and 12:

$$\min_{c_0,c_i,d_i} \{c_0 + \frac{3}{4}c_1 - \frac{1}{4}d_1 + \frac{3}{4}c_2 - \frac{1}{4}d_2 + \frac{2}{4}c_3 - \frac{1}{4}d_3\},$$

Table 4: Lower and upper bounds for $\mathbb{E}I_{\{10,11,12\}}(X)$

J	$\min_{\mathcal{R}_J} \mathbb{E}I_{\{10,11,12\}}(x)$	$\max_{\mathcal{R}_J} \mathbb{E}I_{\{10,11,12\}}(x)$
$\{1,2\}$	0	1
$\{3\}$	0	0.2
$\{2\}$	0	0.6
$\{1\}$	0	0.8
$\{\varnothing\}$	0	0.4

subject to c_i, $d_i \in \mathbb{R}^+$, $c_0 \in \mathbb{R}$, $i = 1, 2, 3$, and

$$\begin{aligned} c_0 + (c_3 - d_3) &\geq 0.2, \\ c_0 + (c_1 - d_1) + (c_2 - d_2) &\geq 1, \\ c_0 + (c_2 - d_2) &\geq 0.6, \\ c_0 + (c_1 - d_1) &\geq 0.8, \\ c_0 &\geq 0.4. \end{aligned}$$

The optimal solution is $d_1 = c_3 = d_2 = 0$, $c_0 = 0.4$, $c_1 = 0.4$, $c_2 = 0.2$, $d_3 = 0.2$. Hence $\overline{\mathbb{E}}\mathbb{E}I_{\{10,11,12\}}(X) = 0.8$. The lower bound for this probability is 0.

Suppose that additionally 10 experts confirm the third judgments. Then we have the following probabilities of judgments:

$$\begin{aligned} &\underline{P}(\mathcal{A}_1, 1) = 1/14, \ \overline{P}(\mathcal{A}_1, 1) = 3/14, \\ &\underline{P}(\mathcal{A}_2, 1) = 1/14, \ \overline{P}(\mathcal{A}_2, 1) = 3/14, \\ &\underline{P}(\mathcal{A}_3, 1) = 11/14, \ \overline{P}(\mathcal{A}_3, 1) = 12/14. \end{aligned}$$

Hence $\overline{\mathbb{E}}\mathbb{E}I_{\{10,11,12\}}(X) = 0.37$.

6 Conclusion

This paper can be regarded as an attempt to find a way for constructing the hierarchical uncertainty model and for combining the interval-valued expert judgments without additional, sometimes erroneous, prior assumptions concerning probabilities or weights of experts. An approach for computing the second-order probabilities proposed in the paper also has the following additional advantages:

1. The obtained probabilities can be simply updated after observing new events or obtaining new expert judgments because Dirichlet prior distributions generate Dirichlet posterior distributions.

2. The obtained hierarchical uncertainty model is more realistic in many applications in the case of a small number of judgments, incomplete, and imprecise data. The possible large imprecision of results reflects insufficiency of available information.

3. The model allows us to make cautious inference about probabilities of judgments due to the controllable hyperparameter s of the imprecise Dirichlet model. As a result, the obtained probabilities of judgments take into account the incompleteness of the available information and total ignorance.

4. The model allows us to deal with conflicting judgments without removing them.

References

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

[2] J.M. Bernardo and A.F.M. Smith. Bayesian Theory. Wiley, Chichester, 1994.

[3] R.M. Cook. Experts in Uncertainty. Opinion and Subjective Probability in Science. Oxford University Press, New York, 1991.

[4] G. de Cooman. Precision–imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. Journal of Statistical Planning and Inference 105 (2002), 175–198.

[5] G. de Cooman and P. Walley. A possibilistic hierarchical model for behaviour under uncertainty. Theory and Decision 52 (2002), 327–374.

[6] M. DeGroot. Optimal Statistical Decisions. McGraw-Hill, New York, 1970.

[7] A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. Annales of Mathematical Statistics 38 (1967), 325–339.

[8] C. Genest and J. Zidek. Combining probability distributions: A critique and an annotated bibliography. Statistical Science 1 (1986), 114–148.

[9] J.Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. Artificial Intelligence 54 (1992), 275–317.

[10] P.J. Huber. Robust Statistics. Wiley, New York, 1981.

[11] V. P. Kuznetsov. Interval Statistical Models. Radio and Communication, Moscow, 1991. In Russian.

[12] R.F. Nau. Uncertainty aversion with second-order probabilities and utilities. In G. de Cooman, T.L. Fine, and T. Seidenfeld, editors, Imprecise Probabilities and Their Applications. Proc. of the 1st Int. Symposium ISIPTA'01, pp. 273–283, Ithaca, USA, June 2001. Shaker Publishing.

[13] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, 1976.

[14] 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, pp. 324–332, Zwijnaarde, Belgium, June 1999.

[15] M.C.M. Troffaes and G. de Cooman. Uncertainty and conflict: A behavioural approach to the aggregation of expert opinions. In J. Vejnarova, editor, Proc. of the 6th Workshop on Uncertainty Processing (WUPES '03), pp. 263–277. VSE - Oeconomica Publishers, September 2003.

[16] L.V. Utkin. A hierarchical uncertainty model under essentially incomplete information. In P. Grzegorzewski, O. Hryniewicz, and M.A. Gil, editors, Soft Methods in Probability, Statistics and Data Analysis, pp. 156–163. Phisica-Verlag, Heidelberg, New York, 2002.

[17] L.V. Utkin. Imprecise second-order hierarchical uncertainty model. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11 (2003), 301–317.

[18] L.V. Utkin. A second-order uncertainty model for the calculation of the interval system reliability. Reliability Engineering and System Safety 79 (2003), 341–351.

[19] L.V. Utkin. Probabilities of judgments provided by unknown experts by using the imprecise Dirichlet model. Risk, Decision and Policy 9 (2004), 391 – 400.

[20] L.V. Utkin. Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model. Fuzzy Sets and Systems 154 (2005), 413–431.

[21] L.V. Utkin. Imprecise second-order uncertainty model for a system of independent random variables. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 13 (2005), 177–194.

[22] L.V. Utkin and T. Augustin. Decision making with imprecise second-order probabilities. In J.-M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, Proc. of the 3rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'03, pp. 545–559, Lugano, Switzerland, July 2003. Carleton Scientific.

[23] P. Walley. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London, 1991.

[24] P. Walley. Inferences from multinomial data: Learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58 (1996), 3–57.

[25] P. Walley. Measures of uncertainty in expert systems. Artificial Intelligence 83 (1996), 1–58.

[26] P. Walley. Statistical inferences based on a second-order possibility distribution. International Journal of General Systems 9 (1997), 337–383.

[27] K. Weichselberger. Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, volume I Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg, 2001.

[28] R.L. Winkler. The consensus of subjective probability distributions. Management Science 15 (1968), 61–75.