A method for processing the unreliable expert judgments about parameters of probability distributions

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

A method for combining two types of judgments about an object analyzed, which are elicited from experts, is considered in the paper. It is assumed that the probability distribution of a random variable is known, but its parameters may be determined by experts. The method is based on the use of the imprecise probability theory and allows us to take into account the quality of expert judgments, heterogeneity and imprecision of information supplied by experts. An approach for computing "cautious" expert beliefs under condition that the experts are unknown is studied. Numerical examples illustrate the proposed method.

Keywords: uncertainty modelling, imprecise probabilities, upper and lower probabilities, linear programming, expert judgments

1. Introduction

A set of judgments elicited from human experts in various application areas is a very important part of information when limited experimental observations are possible. Several methods for elicitation, assessment and pooling of this type of information have been proposed in [2]. In order to get useful information from the experts, a proper uncertainty modeling of pieces of data supplied by experts has to be used. As pointed out in [3], the uncertainty models play a central role in the use of expert judgments, because no human being would claim that he is absolutely sure about his judgments or advice. Therefore, it is necessary to incorporate into any model the individual expert's uncertainty about his advice, the decision maker's uncertainty about the quality of the expert(s), and how these two kinds of uncertainty interact and impact on the credibility of the final results.

Judgements elicited from experts are usually imprecise and unreliable 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. It is clear, however, that the opinion of reliable experts should be more important than those of unreliable ones. Various methods of the pooling of assessments, taking into account the quality of experts, are available in the literature [2,12,24]. These methods use the concept of precise probabilities for modelling the uncertainty. A framework of the possibility theory has been also applied to combining judgments [3] because this theory allows us to consider the lack of precision of the expert knowledge.

However, there are cases when a probability distribution of a random variable is known, for example, from physical nature of objects considered, but its parameters or a part of parameters are defined by experts. In this case, experts may be asked for providing judgments about parameters of this distribution. Moreover, experts may provide also direct information about the random variable. So, there are heterogeneous judgments which have to be combined by means of a unified framework. In order to cope with uncertainty and heterogeneity of available information, it is proposed to apply the *imprecise probability theory* (also called the theory of lower previsions [20], the theory of interval statistical models [11], the theory of interval probabilities [22]), whose general framework is provided by upper and lower previsions. The theory has been applied to a number of application areas, for instance, some examples of the successful application of imprecise probabilities to reliability analysis can be found in [9,18], where it is assumed that probability distributions of the component lifetimes are unknown and there exists only some partial information about the component reliability behavior in the form of interval probabilities, interval moments, etc.

It should be noted that uncertainty of parameters can be considered in a framework of *hierarchical uncertainty models* which are rather common in uncertainty theory. The various application examples of the models can be found in [5,15]. A comprehensive review of hierarchical models is given in [4], where it is argued that the most common hierarchical model is the Bayesian one [1,6–8]. A Bayesian hierarchical approach to aggregating the expert judgments has been considered in [10]. At the same time, the Bayesian hierarchical model is unrealistic in problems where there is available only partial information about the system behavior. The hierarchical models in cases when probability distributions are unknown at all levels of hierarchy are considered in [14–16]. However, the obtained resulting measures in this case are too imprecise because the proposed models do not take into account the fact that the type of probability distributions at the first level of hierarchical models is often known. Therefore, the models studied in this paper can be regarded as an extension of the Bayesian hierarchical model to the case of imprecise parameters of probability distributions.

As an exhaustive review of expert's elicitation procedures is given in [2], these questions are remained outside the scope of this paper. It is supposed here that assessments of parameters of probability distributions in the form of their intervals are available and each assessment is characterized by some probability (belief) or some interval-valued probability that a true value of the assessed parameter is in the given interval. This implies that the term "expert information" is used in a more general sense in the paper. For example, confidence intervals of parameters with corresponding confidence probabilities also can be considered as the "expert information". Moreover, it is supposed that judgments about intervals of the random variables are available. Therefore, the main aim of this paper is to consider a method for combining these different judgments. An approach for computing "cautious" expert judgment beliefs under condition that the experts are unknown is also studied.

2. Preliminary definitions

Suppose there is a continuous random variable $X : \Omega \to \mathbb{R}$ defined on the sample space Ω and information about this variable is represented as a set of m interval-valued expectations of real-valued functions $f_1(X), ..., f_m(X)$. These functions are Riemann integrable. Denote these lower and upper expectations $\underline{\mathbb{E}}f_i$ and $\overline{\mathbb{E}}f_i$, $i = 1, ..., m$. In terms of the theory of imprecise probabilities the corresponding functions $f_i(X)$ and interval-valued expectations $\underline{\mathbb{E}}f_i$ and $\overline{\mathbb{E}}f_i$, $i = 1, ..., m$, are called *gambles* and *lower and upper previsions*, respectively [20]. Various types of information can be modeled by means of lower and upper previsions. For example, if f_i is the indicator function of an event A, then previsions $\underline{\mathbb{E}}f_i$ and $\overline{\mathbb{E}}f_i$ can be regarded as lower and upper probabilities of the event A. If $f_i(X) = X$, then $\underline{\mathbb{E}}X$ and $\overline{\mathbb{E}}X$ are bounds for the mean value of the corresponding random variable. The lower and upper previsions $\underline{\mathbb{E}}f_i$ and $\overline{\mathbb{E}}f_i$ can be regarded as bounds for an unknown precise prevision $\mathbb{E}f_i$ which will be called a *linear prevision*.

For computing new previsions $\underline{\mathbb{E}}g$ and $\overline{\mathbb{E}}g$ of a gamble $g(X)$ from the available information, *natural extension* can be used. Natural extension is a general mathematical procedure for calculating new previsions from initial judgments. It produces a coherent overall model from a certain collection of imprecise probability judgments and may be seen as the basic constructive step in interval-valued statistical reasoning. It is written as the following optimization problems:

$$\underline{\mathbb{E}}g\left(\overline{\mathbb{E}}g\right) = \min_{\rho}\left(\max_{\rho}\right)\int_{\mathbb{R}} g(x)\rho(x)\mathrm{d}x, \quad (1)$$

subject to

$$\rho(x) \geq 0,\ \int_{\mathbb{R}} \rho(x)\mathrm{d}x = 1, \quad (2)$$

$$\underline{\mathbb{E}}f_i \leq \int_{\mathbb{R}} f_i(x)\rho(x)\mathrm{d}x \leq \overline{\mathbb{E}}f_i,\ i \leq m. \quad (3)$$

Here the minimum and maximum are taken over a set of all possible probability density functions $\{\rho(x)\}$ satisfying conditions (3).

Optimization problems (1)-(3) can be explained as follows. The linear prevision $\mathbb{E}g$ is computed as

$$\mathbb{E}g = \int_{\mathbb{R}} g(x)\rho(x)\mathrm{d}x.$$

However, we do not know the density ρ because our initial information is restricted only by the lower and upper previsions $\underline{\mathbb{E}}f_i$ and $\overline{\mathbb{E}}f_i$, $i = 1, ..., m$, and there is no information about distributions of X. At the same time, the available lower and upper previsions produce a set of possible densities that are consistent with these previsions. This means that we can find the largest and smallest possible values of $\mathbb{E}g$ for all densities from the set $\{\rho(x)\}$. It can be carried out by solving optimization problems (1)-(3).

It should be noted that problems (1)-(3) are linear and the dual optimization problems can be written as follows [18,11]:

$$\overline{\mathbb{E}}g = \min_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^{m} \left(c_i \overline{\mathbb{E}} f_i - d_i \underline{\mathbb{E}} f_i \right) \right), \tag{4}$$

$$\underline{\mathbb{E}}g = -\overline{\mathbb{E}}(-g), \tag{5}$$

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

$$c_0 + \sum_{i=1}^{m} (c_i - d_i) f_i(x) \geq g(x). \tag{6}$$

Here c_0, c_i, d_i are the optimization variables such that c_0 corresponds to the constraint $\int_{\mathbb{R}} \rho(x)\mathrm{d}x = 1$, c_i corresponds to the constraint $\int_{\mathbb{R}} f_i(x)\rho(x)\mathrm{d}x \leq \overline{\mathbb{E}}f_i$, and d_i corresponds to the constraint $\underline{\mathbb{E}}f_i \leq \int_{\mathbb{R}} f_i(x)\rho(x)\mathrm{d}x$. It turns out that the dual optimization problems are more simple in comparison with problems (1)-(3) in many applications because this representation allows to avoid the situation when a number of the optimization variables is infinite. Of course, the dual optimization problems generally have an infinite number of constraints each of them is defined by a value of x. However, as it will be shown below, the number of constraints can often be reduced to a finite number.

In order to indicate that expectations are found in accordance with the density ρ, we will note them below by $\mathbb{E}_\rho$.

3. The problem statement

Suppose that a continuous random variable $X : \Omega \to \mathbb{R}$ is governed by a probability density function $\pi(x|\Theta)$, where $\Theta = (\theta_1, ..., \theta_k)$ is a vector of k parameters.

It is assumed that information about the random variable is represented in the following form:

$$\underline{p}_i \leq \mathbb{E}_{\pi(x|\Theta)} f_i(X) \leq \overline{p}_i,\ i = 1, ..., m, \tag{7}$$

or

$$\underline{p}_i \leq \int_{\mathbb{R}} f_i(x)\pi(x|\Theta)\mathrm{d}x \leq \overline{p}_i,\ i = 1, ..., m. \tag{8}$$

This means that we know lower and upper expectations of some real-valued functions of the random variable, such that m judgments are available about its statistical characteristics. In particular, if $\underline{p}_j$ and $\overline{p}_j$ are lower and upper probabilities that the random variable takes values in $[a, b]$, then $f_i(X)$ is the indicator function $I_{[a,b]}(X)$ such that $I_{[a,b]}(X) = 1$ if $X \in [a, b]$ and $I_{[a,b]}(X) = 0$ if $X \notin [a, b]$. By formalizing the information about random variable in the form of (7), we assume that there is a set of possible densities satisfying (7). At the same time, it is known that the random variable is governed by a density of a certain type, for example, exponential one, but with unknown parameters. This additional information significantly restricts the set of densities.

Suppose that there is some additional information about parameters Θ of the probability distribution of X in the following form:

$$\underline{\alpha}_j \leq \mathbb{E}_{\rho(\Theta)} h_j(\Theta) \leq \overline{\alpha}_j,\ j = 1, ..., n, \tag{9}$$

or

$$\underline{\alpha}_j \leq \int_{\Lambda} h_j(\Theta)\rho(\Theta)\mathrm{d}\Theta \leq \overline{\alpha}_j,\ j = 1, ..., n. \tag{10}$$

Here ρ is an unknown joint density function of the vector Θ satisfying constraints (9) and Λ is a set of values of the vector Θ, in particular, $\Lambda = \mathbb{R}^k$. This means that we know lower and upper expectations of some functions of parameters Θ, such that n judgments are available about parameters of the distribution. By formalizing the information about parameters in the form of (9), we assume that parameters are random variables having the density ρ. If parameters are statistically independent as random variables, then the joint density is represented as a product of marginal ones and this condition can be considered as some additional information about parameters.

How to find in this case the expectation of some new function $g(X)$, i.e., $\mathbb{E}_\pi g$?

In order to give the reader the essence of the subject analyzed and make all the formulas more readable, throughout the paper the obvious constraints for densities ρ (or π) to the optimization problems such that $\rho(x) \geq 0$, $\int_{\mathbb{R}} \rho(x)\mathrm{d}x = 1$ will not be written. Moreover, integrals of the form $\int_{\mathbb{R}} g(x)\pi(x|\Theta)\mathrm{d}x$ and $\int_{\Lambda} g(\Theta)\rho(\Theta)\mathrm{d}\Theta$ will be denoted in terms of expectations $\mathbb{E}_{\pi(x|\Theta)}g$ and $\mathbb{E}_{\rho(\Theta)}g$, respectively.

4. Solution of the problem by $n = 0$

Let us consider a special case when $n = 0$, i.e., the available information is represented only in the form of (7). One of the possible ways for solving the problem of computing the expectation $\mathbb{E}_{\pi}g$ is the following. The i-th judgment in the form of (7) produces a set Ψ_i of possible values of parameters. This implies that all judgments produce the set $\Psi = \bigcap_{i=1}^{n} \Psi_i$ (the set satisfying all judgments). By using the set Ψ, it is possible to find lower and upper bounds for the expectation $\mathbb{E}_{\pi}g$ of the function g. The problem can formally be represented in the form of the following optimization problems:

$$\underline{\mathbb{E}}_{\pi}g \left(\overline{\mathbb{E}}_{\pi}g\right) = \min_{\Theta\in\Lambda} \left(\max_{\Theta\in\Lambda}\right) \mathbb{E}_{\pi(x|\Theta)}g(x), \quad (11)$$

subject to

$$\underline{p}_i \leq \mathbb{E}_{\pi(x|\Theta)}f_i(x) \leq \overline{p}_i, \ i = 1, ..., m. \quad (12)$$

Each constraint in (12) produces the set Ψ_i of parameters. The maximum and minimum are taken over the set Ψ of all parameters Θ simultaneously belonging to all sets Ψ_i, $i = 1, ..., m$.

Let us illustrate this approach by means of the following example.

Example 1 *Suppose that the following information about a random variable X is available:*

$$0.1 \leq \Pr\{X \in [0, 10]\} \leq 0.2, \quad (13)$$

$$0.3 \leq \Pr\{X \in [0, 20]\} \leq 0.5. \quad (14)$$

Let us find the probability $\Pr\{X \geq 30\}$ *under condition that X is governed by the exponential distribution with the parameter λ. First, we have to find possible sets of λ produced by two judgments:*

$$0.1 \leq 1 - \exp(-10 \cdot \lambda) \leq 0.2,$$
$$0.3 \leq 1 - \exp(-20 \cdot \lambda) \leq 0.5.$$

Hence

$$\Psi_1 = [0.0105, \ 0.0223],$$
$$\Psi_2 = [0.0178, \ 0.0346].$$

Then

$$\Psi = \Psi_1 \cap \Psi_2 = [0.0178, \ 0.0223].$$

Since the function

$$\mathbb{E}_{\pi}I_{[30,\infty)}(x) = \Pr\{X \geq 30\} = \exp(-30 \cdot \lambda)$$

is monotone (decreasing) with the parameter λ, then we obtain

$$\underline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x) = \exp(-30 \cdot 0.0223) = 0.512,$$
$$\overline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x) = \exp(-30 \cdot 0.0178) = 0.586.$$

It is interesting to note that if the probability distribution would be unknown, then

$$\underline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x) = 0, \ \overline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x) = 0.7.$$

The proposed example shows that lower and upper expectations of any function g can be found without difficulties if there is only one parameter of the distribution and if there exists an inverse function allowing us to express the parameter through bounds $\underline{p}_i$ and $\overline{p}_i$. However, it is difficult to expect that such the simple procedure can be used in the case of two and more parameters. Therefore, it is necessary to develop another approach.

5. Solution of the problem by $m = 0$

Now we consider a special case when $m = 0$, i.e., the available information is represented only in the form of (9).

Let us assume that parameters of a probability distribution are random variables governed by a joint density $\rho(\Theta)$. By knowing the density function $\pi(x|\Theta)$, we can find the conditional probability $P(B|\Theta)$ of an event $B(x)$ under condition that the vector of parameters of π is Θ as follows:

$$P(B|\Theta) = \mathbb{E}_{\pi(x|\Theta)}I_B(x).$$

Then the probability of the event B is computed as (total probability theorem):

$$\begin{aligned} P(B) &= \mathbb{E}_{\rho(\Theta)}(P(B|\Theta)) \\ &= \mathbb{E}_{\rho(\Theta)}(\mathbb{E}_{\pi(x|\Theta)}I_B(x)). \end{aligned}$$

This corresponds to the Bayesian uncertainty model which can always be reduced to a first-order model, by "integrating out the higher-order parameters" [1,6–8].

The approach for determining the probability of the event B can be extended to a more general case when the expectation of a function g has to be determined. In this case, we can write

$$\mathbb{E}_\pi g = \mathbb{E}_{\rho(\Theta)}(\mathbb{E}_{\pi(x|\Theta)} g(x)).$$

Denote $G(g, \Theta) = \mathbb{E}_{\pi(x|\Theta)} g(x)$. So, there are judgments in the form of (10) and we have to find

$$\mathbb{E}_\pi g = \mathbb{E}_{\rho(\Theta)} G(g, \Theta).$$

Since we have no information about $\rho(\Theta)$, then, this problem can be solved by using the natural extension in the form of the following optimization problems:

$$\underline{\mathbb{E}}_\pi g \left(\overline{\mathbb{E}}_\pi g \right) = \min_\rho \left(\max_\rho \right) \mathbb{E}_{\rho(\Theta)} G(g, \Theta), \quad (15)$$

subject to (10).

Optimization problems (15) may be difficult to be solved. Therefore, it is useful to write the corresponding dual optimization problems

$$\underline{\mathbb{E}}_\pi g = \max_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^{n} (c_i \underline{\alpha}_i - d_i \overline{\alpha}_i) \right),$$

subject to c_i, $d_i \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, $i = 1, ..., n$, and $\forall \Theta \in \Lambda$,

$$c_0 + \sum_{i=1}^{n} (c_i - d_i) \, h_i(\Theta) \leq G(g, \Theta),$$

and

$$\overline{\mathbb{E}}_\pi g = \min_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^{n} (c_i \overline{\alpha}_i - d_i \underline{\alpha}_i) \right),$$

subject to c_i, $d_i \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, $i = 1, ..., n$, and $\forall \Theta \in \Lambda$,

$$c_0 + \sum_{i=1}^{n} (c_i - d_i) \, h_i(\Theta) \geq G(g, \Theta).$$

Example 2 *Suppose that the following information about the parameter λ of the exponential distribution is available:*

$$0.2 \leq \Pr\{\lambda \in [0.02,\ 0.03]\} \leq 0.5,$$
$$0.7 \leq \Pr\{\lambda \in [0.04,\ 0.05]\} \leq 0.8.$$

At first sight, experts provide conflicting information because $[0.02,\ 0.03] \cap [0.04,\ 0.05] = \varnothing$. However, we will see that the beliefs of expert judgments make this information to be consistent. Let us find the probability $\Pr\{X \geq 30\}$ under condition that X is governed by the exponential distribution. Then

$$\underline{\mathbb{E}}_\pi g \left(\overline{\mathbb{E}}_\pi g \right) = \min_\rho \left(\max_\rho \right) \mathbb{E}_{\rho(\lambda)} G(I_{[30,\infty)}(x), \lambda),$$

subject to

$$0.2 \leq \mathbb{E}_{\rho(\lambda)} I_{[0.02,\ 0.03]}(\lambda) \leq 0.5,$$
$$0.7 \leq \mathbb{E}_{\rho(\lambda)} I_{[0.04,\ 0.05]}(\lambda) \leq 0.8.$$

Here $\lambda \in \Lambda = \mathbb{R}_+$ and

$$G(I_{[30,\infty)}(x), \lambda) = \exp(-30 \cdot \lambda).$$

The dual optimization problem for computing $\underline{\mathbb{E}}_\pi g$ is

$$\underline{\mathbb{E}}_\pi g = \max_{c_0, c_i, d_i} (c_0 + 0.2c_1 - 0.5d_1$$
$$+ 0.7c_2 - 0.8d_2),$$

subject to c_i, $d_i \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, $i = 1, 2$, and $\forall \lambda \geq 0$,

$$\exp(-30 \cdot \lambda) \geq c_0 + (c_1 - d_1) \, I_{[0.02,\ 0.03]}(\lambda)$$
$$+ (c_2 - d_2) \, I_{[0.04,\ 0.05]}(\lambda).$$

Let us divide the set of values of λ into intervals $[0, 0.02]$, $[0.02, 0.03]$, $[0.03, 0.04]$, $[0.04, 0.05]$, $[0.05, \infty)$. The indicator functions are not changed if λ lies inside these intervals. This implies that an infinite number of constraints is reduced to 5 constraints with minimal right sides, i.e. with minimal values of $\exp(-30 \cdot \lambda)$. The function $\exp(-30 \cdot \lambda)$ decreases as λ increases. Consequently, for each interval of λ, we have to take the maximal value of λ. So, we can write all constraints as follows:

$$c_0 \leq \exp(-30 \cdot 0.02),$$
$$c_0 + (c_1 - d_1) \leq \exp(-30 \cdot 0.03),$$
$$c_0 \leq \exp(-30 \cdot 0.04),$$
$$c_0 + (c_2 - d_2) \leq \exp(-30 \cdot 0.05),$$
$$c_0 \leq 0.$$

The optimization problem for computing $\overline{\mathbb{E}}_{\pi}g$ can similarly be written. By solving the problems, we get

$$\underline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x)=0.242,\ \overline{\mathbb{E}}_{\pi}I_{[30,\infty)}(x)=0.421.$$

6. General case ($m>0$ and $n>0$)

Now we suppose that both types of judgments (7) and (9) are available. In this case, optimization problems for computing $\underline{\mathbb{E}}_{\pi}g$ and $\overline{\mathbb{E}}_{\pi}g$ are of the form:

$$\underline{\mathbb{E}}_{\pi}g\left(\overline{\mathbb{E}}_{\pi}g\right)=\min_{\rho}\left(\max_{\rho}\right)\mathbb{E}_{\rho(\Theta)}G(g,\Theta), \quad (16)$$

subject to (7) and (9).

Proposition 1 *Problems (16) are equivalent to optimization problems with the same objective functions, with n constraints (10), and m constraints*

$$\underline{p}_i\leq\mathbb{E}_{\rho(\Theta)}G(f_i,\Theta)\leq\overline{p}_i,\ i=1,...,m. \quad (17)$$

Proof. Consider one of the problems (16), for example, the problem for computing $\overline{\mathbb{E}}_{\pi}g$. By introducing a new variable C, we can rewrite (11)-(12) as follows:

$$\overline{\mathbb{E}}_{\pi}g=\max C,$$

subject to (9) and

$$C\geq\mathbb{E}_{\pi(x|\Theta)}g(x)=G(g,\Theta),$$
$$\underline{p}_i\leq\mathbb{E}_{\pi(x|\Theta)}f_i(x)=G(f_i,\Theta)\leq\overline{p}_i,\ i\leq m.$$

By assuming that the integrals in constraints are nonnegative and taking into account that $\mathbb{E}_{\rho(\Theta)}1=1$, we can write

$$\mathbb{E}_{\rho(\Theta)}C\geq\mathbb{E}_{\rho(\Theta)}G(g,\Theta),$$

$$\mathbb{E}_{\rho(\Theta)}\underline{p}_i\leq\mathbb{E}_{\rho(\Theta)}G(f_i,\Theta)\leq\mathbb{E}_{\rho(\Theta)}\overline{p}_i,\ i\leq m.$$

Hence

$$C\geq\mathbb{E}_{\rho(\Theta)}G(g,\Theta),$$
$$\underline{p}_i\leq\mathbb{E}_{\rho(\Theta)}G(f_i,\Theta)\leq\overline{p}_i,\ i\leq m,$$

as was to be proved.

Let us prove now that constraints (17) can be replaced by constraints (7). According to [19], the optimal density $\rho^{*}(\Theta)$ is a weighted sum of Dirac functions $\delta(\Theta-\Theta_k)$ which have unit area concentrated in the immediate vicinity of points Θ_k, i.e.,

$$\rho^{*}(\Theta)=\sum_{k=1}^{m+n+1}c_k\delta(\Theta-\Theta_k),$$

where $\sum_{k=1}^{m+n+1}c_k=1$. Then (17) can be rewritten as follows:

$$\underline{p}_i\leq\sum_{k=1}^{m+n+1}c_kG(f_i,\Theta_k)\leq\overline{p}_i$$

or

$$\underline{p}_i\leq\int_{\mathbb{R}}f_i(x)\left(\sum_{k=1}^{m+n+1}c_k\pi(x|\Theta_k)\right)\mathrm{d}x\leq\overline{p}_i.$$

Hence $\underline{p}_i\leq\mathbb{E}_{\pi(x|\Theta)}f_i(x)\leq\overline{p}_i$, where $\pi(x|\Theta)=\sum_{k=1}^{m+n+1}c_k\pi(x|\Theta_k)$. ■

Proposition 1 allows us to combine both types of judgments for computing bounds for some measure. We now illustrate the usefulness of this proposition in the following example.

Example 3 *Suppose that information about a random variable X governed by the exponential distribution with the parameter λ is the same as given in Example 1. However, the following judgment about the parameter λ is added:*

$$0.7\leq\Pr\{\lambda\in[0.04,\ 0.05]\}\leq0.8.$$

Let us find the probability $\Pr\{X\geq30\}$. The optimization problems for computing lower and upper probabilities of failure after time 30 are of the form:

$$\underline{\mathbb{E}}_{\pi}g\left(\overline{\mathbb{E}}_{\pi}g\right)=\min_{\rho}\left(\max_{\rho}\right)\mathbb{E}_{\rho(\lambda)}G(I_{[30,\infty)}(x),\lambda),$$

subject to

$$0.1\leq\mathbb{E}_{\rho(\lambda)}G(I_{[0,10]}(x),\lambda)\leq0.2,$$
$$0.3\leq\mathbb{E}_{\rho(\lambda)}G(I_{[0,20]}(x),\lambda)\leq0.5,$$
$$0.7\leq\mathbb{E}_{\rho(\lambda)}I_{[0.04,\ 0.05]}(\lambda)\leq0.8.$$

Here

$$G(I_{[30,\infty)}(x),\lambda)=\exp(-30\cdot\lambda),$$
$$G(I_{[0,10]}(x),\lambda)=1-\exp(-10\cdot\lambda),$$
$$G(I_{[0,20]}(x),\lambda)=1-\exp(-20\cdot\lambda).$$

By solving the problems, we obtain

$$\underline{\mathbb{E}}_\pi I_{[30,\infty)}(x) = 0.574,\ \overline{\mathbb{E}}_\pi I_{[30,\infty)}(x) = 0.581.$$

By comparing the obtained results with those of Example 1, one can conclude that the additional information reduces the imprecision of the final results.

Example 4 *Let us consider a more complex example. Suppose that information about the random variable X is the same as given in Example 1. However, it is known that the distribution of X is normal with parameters μ (expectation) and σ (standard deviation). Then*

$$G(I_{[30,\infty)}(x),\mu,\sigma) = 1 - \Phi\left((30-\mu)/\sigma\right),$$
$$G(I_{[0,10]}(x),\mu,\sigma) = \Phi\left((10-\mu)/\sigma\right),$$
$$G(I_{[0,20]}(x),\mu,\sigma) = \Phi\left((20-\mu)/\sigma\right).$$

Here Φ is the standard normal distribution function. By assuming that parameters μ and σ are statistically independent, we can write the following optimization problems for computing $\underline{\mathbb{E}}_\pi I_{[30,\infty)}(x)$ and $\overline{\mathbb{E}}_\pi I_{[30,\infty)}(x)$:

$$\underline{\mathbb{E}}_\pi g\left(\overline{\mathbb{E}}_\pi g\right) = \min_{\rho_1,\rho_2}\left(\max_{\rho_1,\rho_2}\right) \times \mathbb{E}_{\rho_1(\mu)\cdot\rho_2(\sigma)} G(I_{[30,\infty)}(x),\mu,\sigma),$$

subject to

$$0.1 \leq \mathbb{E}_{\rho_1(\mu)\cdot\rho_2(\sigma)} G(I_{[0,10]}(x),\mu,\sigma) \leq 0.2,$$
$$0.3 \leq \mathbb{E}_{\rho_1(\mu)\cdot\rho_2(\sigma)} G(I_{[0,20]}(x),\mu,\sigma) \leq 0.5.$$

Solutions to problems are $\underline{\mathbb{E}}_\pi g = 0.1$, $\overline{\mathbb{E}}_\pi g = 0.63$.

7. Cautious expert judgment beliefs

The main difficulty of using the approach presented in the previous sections is how to assign the bounds $\underline{\alpha}_j$ and $\overline{\alpha}_j$ (beliefs of expert judgments) when experts provide intervals A_j of parameters, i.e., $h_j(\Theta) = I_{A_j}(\Theta)$, $A_j \subset \Lambda$. One of the ways to overcome this difficulty is to find subsets of parameters produced by judgments (7). In this case we get a number of subsets of possible parameters and can deal with their intersection (see Section 4). However, the expert judgments might be contradictory and their intersection in this case produces the empty set. Therefore, we propose another way.

The quality of experts is often modelled by means of *weights* assigned to every expert in accordance to some rules. However, if the experts are unknown and they supply interval-valued judgments, the weights as measures of their quality can not measure the quality of provided opinions. Therefore, we consider an approach for computing the bounds $\underline{\alpha}_j$ and $\overline{\alpha}_j$ taking into account judgments of the form (7).

The i-th judgment (7) produces a subset of possible values of parameters $\Psi_i \subseteq \Lambda$ (see Section 4). Therefore, after obtaining m judgments, we have m subsets Ψ_i, $i = 1,...,m$. Moreover, we have n judgments about parameters in the form of intervals $A_j \subseteq \Lambda$, $j = 1,...,n$. Let $\{\mathbf{i}\} = \{(i_1,...,i_{n+m+1})\}$ be a set of all binary vectors consisting of $n+m+1$ components such that $i_j \in \{0,1\}$. For every vector $\mathbf{i}$, we define the subsets B_k ($k = 1,...,2^{n+m+1}$) as follows:

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

Here $C_{n+m+1} = \Lambda$, $C_j = \Psi_j$ if $j \leq m$ and $C_{j+m} = A_j$ if $j = 1,...,n$. As a result, we divide the set Λ into a set of non-intersecting subsets B_k such that $B_1 \cup ... \cup B_L = \Lambda$, $L \leq 2^{n+m+1}$. Moreover, every subset A_j or Ψ_i can be represented as the union of a finite number of subsets B_k.

Let us imagine that every subset A_j or Ψ_i is associated with a large box. This box contains one ball which can move inside the box and we do not know its location. Every subset B_k is associated with a urn. If we cover subsets B_k with indices $k \in J_i$ by the box A_i or Ψ_i, then the ball enters randomly in one of the urns B_k. Here J_i is a set of indices k such that $B_k \subseteq A_i$ or Ψ_i. After covering the urns by all $(n+m)$ boxes, $n+m$ balls are in some urns. What can we say about possible numbers of balls in every urn? It is obvious that there exist different combinations of numbers of balls except a 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 $\mathbf{n}^{(k)} = (n_1^{(k)},...,n_L^{(k)})$, $k = 1,...,Q$, $\sum_{i=1}^{L} n_i^{(k)} = n+m$. If to 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. For every model, the probability $P(A|\mathbf{n}^{(k)})$ of an arbitrary event $A \subseteq \Lambda$ depends on $\mathbf{n}^{(k)}$. So far as all the models are equivalent, even by precise probabilities of all categories only lower and upper probabilities of A can be computed

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

In case of multinomial samples, the Dirichlet distribution is the traditional choice. The *Dirichlet* (s,t) *prior distribution* for π, where $t = (t_1,...,t_L)$, has probability density function [23]

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

where $s > 0$, $t \in S(1,L)$, C is the proportionality constant, $S(1,L)$ denotes the interior of the unit simplex.

Since the number of judgments might be rather small, then the *imprecise Dirichlet model* should be used. This model is defined by Walley [21] as the set of all Dirichlet (s,t) distributions such that $t \in S(1,L)$. For the imprecise Dirichlet model, the *hyperparameter* s determines how quickly upper and lower probabilities of events converge as statistical data accumulate.

Let A be any non-trivial subset of a sample space $\{\omega_1,...,\omega_L\}$, and let $n(A)$ denote the observed number of occurrences of A in the $m+n$ trials, $n(A) = \sum_{\omega_j\in A} n_j$. Then, according to [21], the predictive probability $P(A,s)$ under the Dirichlet posterior distribution is

$$P(A,s) = \frac{n(A)+st(A)}{m+n+s},\ t(A) = \sum_{\omega_j\in A} t_j.$$

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

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

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 write that the sample space considered now is the finite set of subsets $B_1,...,B_L$, events are the subsets $C_1,...,C_{m+n}$. Then the lower and upper probabilities of A, whose representation as the union of elements $B_1,...,B_L$ has indices from a set J, can be written as follows:

$$\underline{P}(A,s) = \min_{k=1,...,Q}\inf_{t\in S(1,L)}\frac{n^{(k)}(A)+st(A)}{m+n+s},$$

$$\overline{P}(A,s) = \max_{k=1,...,Q}\sup_{t\in S(1,L)}\frac{n^{(k)}(A)+st(A)}{m+n+s},$$

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

$$\max n^{(k)}(A) = m+n-\sum_{i:C_j^{(2)}\cap A=\varnothing} 1 = \sum_{i:C_j^{(2)}\cap A\neq\varnothing} 1.$$

Hence

$$\underline{P}(A,s) = \frac{\sum_{i:C_i\subseteq A} 1}{m+n+s},$$
$$\overline{P}(A,s) = \frac{s+\sum_{i:C_i\cap A\neq\varnothing} 1}{m+n+s}.$$

These probabilities can be regarded as a parametric extension [17] of belief and plausibility functions described in Dempster-Shafer theory [13].

Since A is arbitrary, then

$$\underline{\alpha}_j = \underline{P}(A_j,s),\ \overline{\alpha}_j = \overline{P}(A_j,s).$$

Now we have judgments (10) and can use the approach given in Section 6 for computing the lower and upper bounds for the expectation $\mathbb{E}_\pi g$.

However, we can assign the lower $\underline{\beta}_j = \underline{P}(\Psi_j, s)$ and upper $\overline{\beta}_j = \overline{P}(\Psi_j, s)$ probabilities to m judgments (7) and this is some additional information which should be taken into account. In this case, the approach proposed in Section 5 is more preferable. At that, judgments (7) are represented in the form of the following inequalities

$$\underline{\beta}_j \leq \mathbb{E}_{\rho(\Theta)} I_{\Psi_j}(\Theta) \leq \overline{\beta}_j, \ j = 1, ..., m.$$

Here $I_{\Psi_j}(\Theta)$ is the indicator function taking the value 1 if $\Theta \in \Psi_j$.

As a result, we have the following optimization problems for computing $\underline{\mathbb{E}}_\pi g$ and $\overline{\mathbb{E}}_\pi g$:

$$\underline{\mathbb{E}}_\pi g \left(\overline{\mathbb{E}}_\pi g\right) = \min_\rho \left(\max_\rho\right) \mathbb{E}_{\rho(\Theta)} G(g, \Theta),$$

subject to

$$\underline{\alpha}_j \leq \mathbb{E}_{\rho(\Theta)} h_j(\Theta) \leq \overline{\alpha}_j, \ j = 1, ..., n,$$

$$\underline{\beta}_j \leq \mathbb{E}_{\rho(\Theta)} I_{\Psi_j}(\Theta) \leq \overline{\beta}_j, \ j = 1, ..., m.$$

The corresponding dual optimization problems are

$$\underline{\mathbb{E}}_\pi g = \max_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^n (c_i \underline{\alpha}_i - d_i \overline{\alpha}_i) + \sum_{i=1}^m \left(v_i \underline{\beta}_i - w_i \overline{\beta}_i \right) \right),$$

subject to $c_i, d_i, v_j, w_j \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, $i = 1, ..., n$, $j = 1, ..., m$, and $\forall \Theta \in \Lambda$,

$$G(g, \Theta) \geq c_0 + \sum_{i=1}^n (c_i - d_i) h_i(\Theta) + \sum_{i=1}^m (v_i - w_i) I_{\Psi_i}(\Theta),$$

and

$$\overline{\mathbb{E}}_\pi g = \min_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^n (c_i \overline{\alpha}_i - d_i \underline{\alpha}_i) + \sum_{i=1}^m \left(v_i \overline{\beta}_i - w_i \underline{\beta}_i \right) \right),$$

subject to $c_i, d_i, v_j, w_j \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, $i = 1, ..., n$, $j = 1, ..., m$, and $\forall \Theta \in \Lambda$,

$$G(g, \Theta) \leq c_0 + \sum_{i=1}^n (c_i - d_i) h_i(\Theta) + \sum_{i=1}^m (v_i - w_i) I_{\Psi_i}(\Theta).$$

Example 5 *Let us return to Example 3 under condition that the judgment about the parameter λ is of the form $\lambda \in [0.04, \ 0.05]$ and the lower and upper probabilities of this judgment are unknown. As shown in Example 1 judgments (13) and (14) produce the sets $\Psi_1 = [0.0105, \ 0.0223]$ and $\Psi_2 = [0.0178, \ 0.0346]$ of λ. By using the notation introduced in this section, we write $A_1 = [0.04, \ 0.05]$. One can see that the above judgments are conflicting. Note that $\Psi_1 \cap \Psi_2 \neq \varnothing$, $\Psi_i \cap A_1 = \varnothing$, $i = 1, 2$, $\Psi_1 \subsetneq \Psi_2$, $\Psi_2 \subsetneq \Psi_1$, $A_1 \subsetneq \Psi_i$, $\Psi_i \subsetneq A_1$, $i = 1, 2$. This implies that*

$$\underline{P}(A_1, s) = \frac{1}{3+s}, \ \overline{P}(A_1, s) = \frac{s+1}{3+s},$$

$$\underline{P}(\Psi_i, s) = \frac{1}{3+s}, \ \overline{P}(\Psi_i, s) = \frac{s+2}{3+s}, \ i = 1, 2.$$

Let us take $s = 1$. Then

$$\underline{\mathbb{E}}_\pi g \left(\overline{\mathbb{E}}_\pi g\right) = \min_\rho \left(\max_\rho\right) \mathbb{E}_{\rho(\lambda)} \exp(-30 \cdot \lambda),$$

subject to

$$0.25 \leq \mathbb{E}_{\rho(\lambda)} I_{[0.0105, \ 0.0223]}(\lambda) \leq 0.75,$$
$$0.25 \leq \mathbb{E}_{\rho(\lambda)} I_{[0.0178, \ 0.0346]}(\lambda) \leq 0.75,$$
$$0.25 \leq \mathbb{E}_{\rho(\lambda)} I_{[0.04, \ 0.05]}(\lambda) \leq 0.5.$$

The dual problem for computing the lower bound is

$$\underline{\mathbb{E}}_\pi g = \max_{c_0, c_i, d_i} (c_0 + 0.25 c_1 - 0.5 d_1 + 0.25 v_1 - 0.75 w_1 + 0.25 v_2 - 0.75 w_2),$$

subject to $c_i, d_i, v_j, w_j \in \mathbb{R}_+$, $c_0 \in \mathbb{R}$, and $\forall \lambda \geq 0$,

$$\exp(-30 \cdot \lambda) \geq c_0 + (c_1 - d_1) I_{[0.04, \ 0.05]}(\lambda) + (v_1 - w_1) I_{[0.0105, \ 0.0223]}(\lambda) + (v_2 - w_2) I_{[0.0178, \ 0.0346]}(\lambda).$$

By solving the problems, we get

$$\underline{\mathbb{E}}_\pi I_{[30, \infty)}(x) = 0.184, \ \overline{\mathbb{E}}_\pi I_{[30, \infty)}(x) = 0.775.$$

The imprecision of the results depends on the number of available judgments and on the parameter s. Smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. So, by increasing the value of s, we calculate ”cautious“ expert judgment beliefs in case when the number of available judgments is small.

One can see from the above example that computations are rather simple. However, this takes place if there is one unknown parameter and efficient computational methods should be developed in case of several unknown parameters.

8. Conclusion

The method for combining unreliable judgments has been studied in the paper. This method has the following advantages in comparison with the classical probabilistic and possibilistic approaches:

1. As pointed out in [3], experts better supply intervals rather than point-values because their knowledge is not only of limited reliability, but also imprecise. The proposed method supports this fact.

2. By assessing the quality of expert judgments, the decision maker does not need to take into account a very common case when experts are too cautious because the intervals they supply are too large to be informative. This case is automatically taken into account. Therefore, beliefs of expert judgments are assigned only on the base of their accuracy, i.e., by analyzing how values given by experts are consistent with the actual information about a variable.

3. The proposed method allows us to combine heterogeneous judgments, i.e., judgments about parameters of probability distributions and directly about random variables.

4. The proposed method also takes into account conflicting judgments of experts in a group. In this case, the optimization problems considered above do not have solutions.

5. The proposed method provides the best intervals of expectations of arbitrary functions under given information elicited from experts.

6. The proposed method is mathematically strong and does not use operators which are often applied intuitively like operators ”min” and ”max” in the possibilistic approach to combining the expert judgments.

The main deficiency of the method is its possible computational complexity.

It should be noted in conclusion that the obtained results can be applied to various application areas where information elicited from experts is imprecise. Moreover, the results are easily extended to a case of discrete random variables. In this case, the probability density functions and integrals are replaced by probability mass functions and sums, respectively.

Acknowledgement

The work was supported by the Alexander von Humboldt Foundation (Germany). I am very grateful to Prof. Dr. Kurt Weichselberger (Munich University, Institute of Statistics) for his help and support. I would like to express my appreciation to the anonymous referees whose comments have improved the paper.

REFERENCES

1. J. Berger, Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York, 1985.
2. R. Cook, Experts in Uncertainty. Opinion and Subjective Probability in Science, Oxford University Press, New York, 1991.
3. D. Dubois, H. Kalfsbeek, Elicitation, assessment and pooling of expert judgement using possibility theory, in: C. Manikopoulos (Ed.), Proc. of the 8th Inter. Congress of Cybernetics and Systems, New Jersey Institute of Technology Press, Newark, NJ, 1990, pp. 360–367.
4. G. de Cooman, Precision–imprecision equivalence in a broad class of imprecise hierarchical uncertainty models, Journal of Statistical Planning and Inference 105 (1) (2002) 175–198.
5. G. de Cooman, P. Walley, A possibilistic hierarchical model for behaviour under uncertainty, Theory and Decision 52 (4) (2002) 327–374.
6. M. Goldstein, The prevision of a prevision, J. Amer. Statist. Soc. 87 (1983) 817–819.

7. I. Good, Some history of the hierarchical Bayesian methodology, in: J. Bernardo, M. DeGroot, D. Lindley, A. Smith (Eds.), Bayesian Statistics, Valencia University Press, Valencia, 1980, pp. 489–519.
8. I. R. Goodman, H. T. Nguyen, Probability updating using second order probabilities and conditional event algebra, Information Sciences 121 (3-4) (1999) 295–347.
9. S. Gurov, L. Utkin, Reliability of Systems under Incomplete Information, Lubavich Publ., Saint Petersburg, 1999, in Russian.
10. L. Joseph, G. Parmigiani, V. Hasselblad, Combining expert judgment by hierarchical modeling: An application to physician staffing, Management Science 44 (1998) 149–161.
11. V. P. Kuznetsov, Interval Statistical Models, Radio and Communication, Moscow, 1991, in Russian.
12. P. Morris, Combining expert judgements, a Bayesian approach, Management Science 23 (1977) 679–692.
13. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
14. M. Troffaes, G. de Cooman, Uncertainty and conflict: A behavioural approach to the aggregation of expert opinions, in: J. Vejnarova (Ed.), Proc. of the 6th Workshop on Uncertainty Processing (WUPES ’03), VSE - Oeconomica Publishers, 2003, pp. 263–277.
15. L. Utkin, Imprecise second-order hierarchical uncertainty model, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11 (3) (2003) 301–317.
16. L. Utkin, A second-order uncertainty model for the calculation of the interval system reliability, Reliability Engineering and System Safety 79 (3) (2003) 341–351.
17. L. Utkin, Belief functions and the imprecise Dirichlet model, in: I. Batyrshin, J. Kacprzyk, L. Sheremetov (Eds.), Proc. of the Int. Conf. on Fuzzy Sets and Soft Computing in Economics and Finance, Vol. 1, Mexican Petroleum Institute, St.Petersburg, Russia, 2004, pp. 178–185.
18. L. Utkin, S. Gurov, New reliability models based on imprecise probabilities, in: C. Hsu (Ed.), Advanced Signal Processing Technology, World Scientific, 2001, Ch. 6, pp. 110–139.
19. L. Utkin, I. Kozine, Different faces of the natural extension, in: G. de Cooman, T.L. Fine and T. Seidenfeld (Eds.), Proc. of the 2nd Int. Symposium on Imprecise Probabilities and Their Applications (ISIPTA’01), Shaker Publishing, Ithaca, USA, 2001, pp. 316-323.
20. P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
21. P. Walley, Inferences from multinomial data: Learning about a bag of marbles, Journal of the Royal Statistical Society, Series B 58 (1996) 3–57, with discussion.
22. K. Weichselberger, Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, Vol. I Intervallwahrscheinlichkeit als umfassendes Konzept, Physika, Heidelberg, 2001.
23. S. Wilks, Mathematical Statistics, Wiley and Sons, New York, 1962.
24. R. Winkler, The consensus of subjective probability distributions, Management Science 15 (1968) 61–7.