Decision Making with Imprecise Second-Order Probabilities

L.V. UTKIN
St.Petersburg State Forest Technical Academy, Russia

TH. AUGUSTIN
University of Munich, Germany

Abstract

In this paper we consider decision making under hierarchical imprecise uncertainty models and derive general algorithms to determine optimal actions. Numerical examples illustrate the proposed methods.

Keywords

decision making, generalized expected utility, imprecise probabilities, second-order uncertainty, natural extension, linear programming

1 Introduction

Consider the basic model of decision theory: One has to choose an *action* from a non-empty, finite set $\mathbb{A} = \{a_1, ..., a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state* of nature $\vartheta \in \Theta = \{\vartheta_1, ..., \vartheta_m\}$. The corresponding outcome is evaluated by the *utility function*

$$u: \quad \begin{array}{l} (\mathbb{A} \times \Theta) \to \mathbb{R} \\ (a, \vartheta) \longmapsto u(a, \vartheta) \end{array}$$

and by the associated random variable $\mathbf{u}(\mathbf{a})$ on $(\Theta, \mathcal{P}o(\Theta))$ taking the values $u(a, \vartheta)$. Often it makes sense to study randomized actions, which can be understood as a probability measure $\lambda = (\lambda_1, ..., \lambda_n)$ on $(\mathbb{A}, \mathcal{P}o(\mathbb{A}))$. Then $u(\cdot)$ and $\mathbf{u}(\cdot)$ are extended to randomized actions by defining $u(\lambda, \vartheta) := \sum_{s=1}^{n} u(a_s, \vartheta)\lambda_s$.

This model contains the essentials of every (formalized) decision situation under uncertainty and is applied in a huge variety of disciplines. If the states of nature are produced by a perfect random mechanism (e.g. an ideal lottery), and the corresponding probability measure $\pi(\cdot)$ on $((\Theta, \mathcal{P}o(\Theta)))$ is completely known, the Bernoulli principle is nearly unanimously favored. One chooses that action λ^* which maximizes the expected utility $\mathbb{E}_\pi \mathbf{u}(\lambda) := \sum_{j=1}^m (u(\lambda, \vartheta_j) \cdot \pi(\vartheta_j))$ among all λ.

In most practical applications, however, the true state of nature can not be understood as arising from an ideal random mechanism. And even if so, the corresponding probability distribution will be not known exactly. An efficient approach for solving this problem in the framework of imprecise probability theory (Kuznetsov [13], Walley [18], Weichselberger [20]) has been proposed by Augustin in [1, 2].

A related, quite commonly used, way to deal with complex uncertainty is to apply *second-order uncertainty models* (*hierarchical uncertainty models*). These models describe the uncertainty of a random quantity by means of two levels. Many papers are devoted to the theoretical [4, 5, 11, 14, 19] and practical [7, 9, 12] aspects of second-order uncertainty models. A comprehensive review of hierarchical models is given in [6] where it is argued that the most common hierarchical model is the Bayesian one [3, 10, 21]. At the same time, the Bayesian hierarchical model is unrealistic in applications where there is available only partial information about the system behavior.

Most proposed second-order uncertainty models assume that there is a precise second-order probability distribution (or possibility distribution). Unfortunately, such information is often absent and making additional assumptions may lead to wrong results. A new hierarchical uncertainty model for combining different types of evidence was proposed by Utkin [15, 16], where the second-order probabilities can be regarded as confidence weights and the first-order uncertainty is modelled by lower and upper previsions of different gambles. We will call these hierarchical models second-order probabilities of type 1.

It is worth noticing that there are cases when the type of the probability distribution of the states of nature is known, for example, from their physical nature, but parameters or a part of the parameters of the distribution are defined by experts. In reality, there is some degree of our belief to each expert's judgement whose value is determined by experience and competence of the expert. Therefore, it is necessary to take into account the available information about experts to obtain more credible decisions. This model can be also considered in the framework of hierarchical models and will be called second-order probabilities of type 2.

Decision making for both models of type 1 and type 2 are studied in the paper. In particular, we give general and efficient algorithms for calculating optimal actions and illustrate them in detailed examples.

One should note explicitly that throughout the paper we assume the utility and the description of the uncertainty on the state of nature are given. Alternatively, there are quite sophisticated approaches directly extending the Neumann-Morgenstern point of view. They *construct* separated utility and imprecise probability from axioms on behaviour and preferences (see, e.g., the work of [8] and the references therein).

2 Second-Order Probabilities of Type 1

Suppose that there is a set of weighted expert judgements related to some measures of the states of nature $\mathbb{E}f_i(\vartheta_j)$, $i = 1,...,r$, i.e., there are values $\underline{b}_i$, $\overline{b}_i$ of lower and upper previsions. Suppose that the credibility of each of r experts is characterized by a subjective probability γ_i or interval of probabilities $[\underline{\gamma}_i, \overline{\gamma}_i]$, $i = 1,...,r$. It should be noted that the second-order probabilities $\underline{\gamma}_i$ and $\overline{\gamma}_i$ form an imprecise probability, described by a set $\mathcal{N}$ of distributions on the set $\mathcal{M}$ of all distributions π on $(\Theta, \mathcal{P}o(\Theta))$. We assume that the second-order imprecise probability is avoiding sure loss, i.e., $\mathcal{N}$ is not empty. Denote for any gamble f the lower (upper) second-order expectations by ${}^L\mathbb{E}_{\mathcal{N}}f$ (${}^U\mathbb{E}_{\mathcal{N}}f$), respectively. Generally, the judgements can be written as follows:

$$\Pr\left\{\underline{b}_i \leq \mathbb{E}_\pi f_i \leq \overline{b}_i\right\} \in [\underline{\gamma}_i, \overline{\gamma}_i],\ i = 1,...,r, \tag{1}$$

or

$${}^L\mathbb{E}_{\mathcal{N}} I_{B_i}(\mathbb{E}_\pi f_i) = \underline{\gamma}_i,\ \ {}^U\mathbb{E}_{\mathcal{N}} I_{B_i}(\mathbb{E}_\pi f_i) = \overline{\gamma}_i,\ i = 1,...,r.$$

Here the set $\{\underline{b}_i, \overline{b}_i\}$ contains the first-order previsions, $B_i = [\underline{b}_i, \overline{b}_i]$, the set $\{\underline{\gamma}_i, \overline{\gamma}_i\}$ contains the second-order probabilities and $\mathbb{E}_\pi f_i = \sum_{j=1}^m f_i(\vartheta_j)\pi(\vartheta_j)$.

The problem here is that the resulting set of distributions may be rather complex because the functions f_i are different, especially, if the value of m is large.

2.1 Decision Making

Since there exists the set $\mathcal{N}$ of distributions on the set $\mathcal{M}$ of all distributions π, the expected utility $\mathbb{E}_\pi \mathbf{u}(\lambda)$ can be considered as a random variable described by distributions from $\mathcal{N}$, and there exist lower ${}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda))$ and upper ${}^U\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda))$ expectations of this random variable, which depend on the action λ. These expectations can be roughly called also by lower and upper "average" expected utilities. With this respect, we can assert that every action is evaluated by its minimal "average" expected utility. By representing the interval $[{}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda)), {}^U\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda))]$ by the lower interval limit alone, we can write the criterion of decision making.

Throughout the paper we evaluate interval-valued expectations by their lower interval-limits only — more complex interval orderings are a topic of further research, see also Section 4. Therefore, an action λ^* is optimal iff for all λ

$${}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda^*)) \geq {}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda)). \tag{2}$$

Then the optimal action λ^* can be obtained by maximizing ${}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda))$ subject to $\sum_{s=1}^n \lambda_s = 1$, $\lambda_s \geq 0$, $s = 1,...,n$. In other words, the following optimization problem has to be solved:

$${}^L\mathbb{E}_{\mathcal{N}}(\mathbb{E}_\pi \mathbf{u}(\lambda)) \to \max_{\lambda_s}$$

under the constraints

$$\sum_{s=1}^{n} \lambda_s = 1,\ \lambda_s \geq 0,\ s = 1,...,n.$$

Due to arguments similar to those used in [17], this problem can be rewritten as

$$^{L}\mathbb{E}_{\mathcal{N}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^{*})\right) = \max_{c\in\mathbb{R},c_k\in\mathbb{R}_+,d_k\in\mathbb{R}_+,\lambda_s\in\mathbb{R}_+} \left\{ c + \sum_{k=1}^{r}\left(c_k\underline{\gamma}_k - d_k\overline{\gamma}_k\right)\right\} \quad (3)$$

subject to

$$c + \sum_{k=1}^{r}(c_k - d_k)\, I_{B_k}\left(\mathbb{E}_{\pi}f_k\right) \leq \mathbb{E}_{\pi}\mathbf{u}(\lambda), \quad (4)$$

$$\sum_{s=1}^{n} \lambda_s = 1. \quad (5)$$

By substituting the expressions for $\mathbb{E}_{\pi}f_i$ and $\mathbb{E}_{\pi}\mathbf{u}(\lambda^{*})$ into the constraints, we get

$$c + \sum_{k=1}^{r}(c_k - d_k)\, I_{B_k}\left(\sum_{j=1}^{m} f_k(\vartheta_j)\pi(\vartheta_j)\right) \leq \sum_{j=1}^{m}\left(u(\lambda,\vartheta_j)\cdot\pi(\vartheta_j)\right),\ \forall\pi\in\mathcal{M}. \quad (6)$$

It is worth noticing that the maximal number of different expressions for the left sides of the constraints (6) is 2^r because they involve indicator functions. Let us write a vector $\mathbf{i} = (i_1,...,i_r)$, $i_j \in \{0,1\}$, whose values correspond to those situations. In accordance with possible values of the binary vector $\mathbf{i}$, the set $\mathcal{M}$ can be divided into 2^r subsets $\mathcal{M}_1,...,\mathcal{M}_{2^r}$ such that the i-th subset is formed by the set of constraints

$$\mathbb{E}_{\pi}f_k \in \begin{cases} B_k, & i_k = 1 \\ B_k^c, & i_k = 0 \end{cases},\ k = 1,...,r. \quad (7)$$

Here $B_k^c = [\inf\mathbb{E}_{\pi}f_k, \sup\mathbb{E}_{\pi}f_k]\backslash B_k$ is the (relative) complement of the interval B_k.

Introduce the set $K_j \subseteq \{1,...,r\}$ corresponding to the set $\mathcal{M}_j$ such that for any $\pi\in\mathcal{M}_j$ and $k\in K_j$ there holds $I_{B_k}(\mathbb{E}_{\pi}f_k) = 1$, and for $l \notin K_j$ there holds $I_{B_l}(\mathbb{E}_{\pi}f_l) = 0$.

Let $\pi = (\pi(\vartheta_1),...,\pi(\vartheta_m))$ be a probability distribution belonging to $\mathcal{M}_j$. It should be noted that some elements from the set $\{\mathcal{M}_j, j = 1,...,2^r\}$ may be empty, i.e., there are no such distributions π that satisfy all constraints (7). This means that the corresponding vector of indices $\mathbf{i}$ provides inconsistent judgements (7) and corresponding constraints (4) must be removed from the list of 2^r constraints. Therefore, as the first step, it is necessary to determine the consistency of judgements. The consistency of the set of constraints, corresponding to a realization of the vector $\mathbf{i}$, can be determined by solving a linear programming problem with an arbitrary objective function and constraints (7). If any solution exists, then the feasible region is non-empty and there exists at least one probability distribution π satisfying all constraints (7), i.e., $\mathcal{M}_j \neq \emptyset$. Otherwise, $\mathcal{M}_j = \emptyset$ and the corresponding constraint (4) must be removed.

Let $L \subseteq \{1,...,2^r\}$ be a set of indices for all consistent constraints or all non-empty sets. Suppose that $\pi_1 \in \mathcal{M}_j$ and $\pi_2 \in \mathcal{M}_j$ are two distributions from $\mathcal{M}_j$, $j \in L$, such that $\mathbb{E}_{\pi_1}\mathbf{u}(\lambda) \geq \mathbb{E}_{\pi_2}\mathbf{u}(\lambda)$. Since $\pi_1 \in \mathcal{M}_j$ and $\pi_2 \in \mathcal{M}_j$, then the constraint

$$c + \sum_{k \in K_j} (c_k - d_k) \leq \mathbb{E}_{\pi_1}\mathbf{u}(\lambda),$$

follows from the constraint

$$c + \sum_{k \in K_j} (c_k - d_k) \leq \mathbb{E}_{\pi_2}\mathbf{u}(\lambda),$$

because the left sides of constraints are the same. This implies that from all constraints, corresponding to the set $\mathcal{M}_j$, we have to keep only one constraint

$$c + \sum_{k \in K_j} (c_k - d_k) \leq \min_{\pi \in \mathcal{M}_j} \mathbb{E}_{\pi}\mathbf{u}(\lambda).$$

So, problem (3)-(5) becomes

$$^{L}\mathbb{E}_{\mathcal{N}}(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*)) = \max_{c \in \mathbb{R}, c_k \in \mathbb{R}_+, d_k \in \mathbb{R}_+, \lambda_s \in \mathbb{R}_+} \left\{ c + \sum_{k=1}^{r} \left(c_k \underline{\gamma}_k - d_k \overline{\gamma}_k \right) \right\} \tag{8}$$

subject to

$$c + \sum_{k \in K_j} (c_k - d_k) \leq \min_{\pi \in \mathcal{M}_j} \mathbb{E}_{\pi}\mathbf{u}(\lambda), \ \forall j \in L, \tag{9}$$

$$\sum_{s=1}^{n} \lambda_s = 1. \tag{10}$$

Write $G_j = \min_{\pi \in \mathcal{M}_j} \mathbb{E}_{\pi}\mathbf{u}(\lambda)$, $j \in L$. Then there holds

$$^{L}\mathbb{E}_{\mathcal{N}}(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*)) = \max_{c \in \mathbb{R}, c_k \in \mathbb{R}_+, d_k \in \mathbb{R}_+, \lambda_s \in \mathbb{R}_+, G_j} \left\{ c + \sum_{k=1}^{r} \left(c_k \underline{\gamma}_k - d_k \overline{\gamma}_k \right) \right\} \tag{11}$$

subject to

$$c + \sum_{k \in K_j} (c_k - d_k) \leq G_j, \tag{12}$$

$$\mathbb{E}_{\pi}\mathbf{u}(\lambda) \geq G_j, \ \pi \in \mathcal{M}_j, \ \forall j \in L, \ \sum_{s=1}^{n} \lambda_s = 1. \tag{13}$$

One can see that the variables G_k are linear for all $k \in L$. This implies that the optimization problem (11)-(13) is linear, but, in the way it is written, it contains infinitely many constraints. In order to overcome this difficulty, note, however, that the set of distributions $\mathcal{M}_j$ for every j can be viewed as a simplex in a finite dimensional space. According to some general results from linear programming theory, an optimal solution to the above problem is achieved at extreme points of the simplex, and the number of its extreme points is finite. This implies, similar to the solution in the first-order decision problem [1, 2], that the infinite set of constraints (13) is reduced to some finite number, and standard routines for linear programming can be used to determine optimal actions. If one wants to concentrate on unrandomized actions (pure actions), where $\lambda_s \in \{0,1\}$, then Boolean optimization can be used.

2.2 Numerical Example

Suppose that 2 experts evaluate 3 states $\{1,2,3\}$ of nature as follows: the probability that either the first state or the second one is true is less than 0.4; the mean value of states is between 1 and 2. The belief to the first expert is 0.5. This means that he (she) provides 50% of true judgements. The belief to the second expert is between 0.3 and 1. This means that he (she) provides more than 30% of true judgements. Values of the utility function $u(a_s,\vartheta_j)$ are given in Table 1.

Table 1: Values of the utility function $u(a_s,\vartheta_j)$

	ϑ_1	ϑ_2	ϑ_3
a_1	6	3	1
a_2	2	7	4

Table 2: Consistency of constraints

i	set	consistent
$(1,1)$	$\mathbb{E}_\pi I_{\{1,2\}}(\vartheta) \in [0,0.4]$, $\mathbb{E}_\pi \vartheta \in [1,2]$	no
$(1,0)$	$\mathbb{E}_\pi I_{\{1,2\}}(\vartheta) \in [0,0.4]$, $\mathbb{E}_\pi \vartheta \in [2,3]$	yes
$(0,1)$	$\mathbb{E}_\pi I_{\{1,2\}}(\vartheta) \in [0.4,1]$, $\mathbb{E}_\pi \vartheta \in [1,2]$	yes
$(0,0)$	$\mathbb{E}_\pi I_{\{1,2\}}(\vartheta) \in [0.4,1]$, $\mathbb{E}_\pi \vartheta \in [2,3]$	yes

The above judgements can be written in the formal form as follows:

$$\Pr\left\{0 \le \mathbb{E}_\pi I_{\{1,2\}}(\vartheta) \le 0.4\right\} = 0.5, \quad \Pr\{1 \le \mathbb{E}_\pi \vartheta \le 2\} \in [0.3,1].$$

Let us find the set $L \subseteq \{1,2,3,4\}$. It can be seen from Table 2 that $L = \{2,3,4\}$. Let us find the optimal strategies λ_1^*, λ_2^*. For doing so, it is necessary to find extreme points for subsets $\mathcal{M}_2$, $\mathcal{M}_3$, $\mathcal{M}_4$.

Subset 2:

$$\{\pi_1 = 0, \pi_2 = 0, \pi_3 = 1\}$$
$$\{\pi_1 = 0, \pi_2 = 0.4, \pi_3 = 0.6\}$$
$$\{\pi_1 = 0.4, \pi_2 = 0, \pi_3 = 0.6\}$$

Subset 3:

$$\{\pi_1 = 1, \pi_2 = 0, \pi_3 = 0\}$$
$$\{\pi_1 = 0, \pi_2 = 1, \pi_3 = 0\}$$
$$\{\pi_1 = 0.5, \pi_2 = 0, \pi_3 = 0.5\}$$

Subset 4:

$$\{\pi_1 = 0, \pi_2 = 1, \pi_3 = 0\}$$
$$\{\pi_1 = 0.5, \pi_2 = 0, \pi_3 = 0.5\}$$
$$\{\pi_1 = 0, \pi_2 = 0.4, \pi_3 = 0.6\}$$
$$\{\pi_1 = 0.4, \pi_2 = 0, \pi_3 = 0.6\}$$

So, the following optimization problem has to be considered:

$$^{L}\mathbb{E}_{\mathcal{N}}(\mathbb{E}_{\pi}\mathbf{u}(\lambda^{*})) = \max_{c,c_k,d_k,\lambda_s,G_j} \{c + 0.5c_1 - 0.5d_1 + 0.3c_2 - 1d_2\}$$

subject to $c_i \geq 0, d_i \geq 0, \lambda_i \geq 0$, $i = 1,2$,

$$\begin{aligned} c + 1\cdot(c_1 - d_1) + 0\cdot(c_2 - d_2) &\leq G_2, \\ c + 0\cdot(c_1 - d_1) + 1\cdot(c_2 - d_2) &\leq G_3, \\ c + 0\cdot(c_1 - d_1) + 0\cdot(c_2 - d_2) &\leq G_4, \end{aligned}$$

$$\begin{aligned} (\lambda_1 + 4\lambda_2)\cdot 1 &\geq G_2, \\ (3\lambda_1 + 7\lambda_2)\cdot 0.4 + (\lambda_1 + 4\lambda_2)\cdot 0.6 &\geq G_2, \\ (6\lambda_1 + 2\lambda_2)\cdot 0.4 + (\lambda_1 + 4\lambda_2)\cdot 0.6 &\geq G_2, \\ (6\lambda_1 + 2\lambda_2)\cdot 1 &\geq G_3, \\ (3\lambda_1 + 7\lambda_2)\cdot 1 &\geq G_3, \\ (6\lambda_1 + 2\lambda_2)\cdot 0.5 + (\lambda_1 + 4\lambda_2)\cdot 0.5 &\geq G_3, \\ (3\lambda_1 + 7\lambda_2)\cdot 1 &\geq G_4, \\ (6\lambda_1 + 2\lambda_2)\cdot 0.5 + (\lambda_1 + 4\lambda_2)\cdot 0.5 &\geq G_4, \\ (3\lambda_1 + 7\lambda_2)\cdot 0.4 + (\lambda_1 + 4\lambda_2)\cdot 0.6 &\geq G_4, \\ (6\lambda_1 + 2\lambda_2)\cdot 0.4 + (\lambda_1 + 4\lambda_2)\cdot 0.6 &\geq G_4, \\ \lambda_1 + \lambda_2 &= 1. \end{aligned}$$

Solution of the problem: $c = 3.143$, $G_2 = G_3 = G_4 = 3.143$, $c_1 = c_2 = d_1 = d_2 = 0$, $\lambda_1 = 0.2857$, $\lambda_2 = 0.7143$.

3 Second-Order Probabilities of Type 2

Suppose that the states of nature are described by a discrete probability distribution of a certain type, for example, binomial, hypergeometric or Poisson distributions. The certain type of the distribution is often known from some physical properties of the considered object. However, the parameters of the corresponding distribution may be uncertain. Denote by $\alpha = (\alpha_1, ..., \alpha_h)$ a vector of parameters for some discrete distribution $\pi(\vartheta, \alpha)$. Consider a case of continuous real parameters, i.e., $\alpha_i \in \mathbb{R}$. If we suppose that the experts provide some evidence about parameters, then the vector α can be considered, just as in classical Bayesian statistics, as a random variable. This is due to the following reasons: First, experts may provide some information about statistical characteristics of parameters, for example, about intervals of mean values or about some probability that the i-th parameter is in an interval. Second, even if experts provide only information about intervals of possible values of parameters, we can not totally believe in the experts because they may be unreliable. This implies that every expert is characterized by a probability or by an interval-valued probability of producing correct judgements. Generally, if we suppose that the vector of parameters is governed by some unknown joint density ρ, then the expert judgements can be formally written as follows:

$$\underline{\gamma}_{ij} \leq \mathbb{E}_\rho f_{ij}(\alpha_i) \leq \overline{\gamma}_{ij},\ i = 1, ..., h,\ j = 1, ..., r_i. \tag{14}$$

Here r_i is a number of judgements related to i-th parameter; f_{ij} is a function corresponding to information about the i-th parameter provided by the j-th expert. For example, if an expert offers information about the probability that the i-th parameter is in an interval B, then $f_{ij}(\alpha_i)$ is the indicator function of the event B, i.e., $f_{ij}(\alpha_i) = I_B(\alpha_i)$. If the expert provides the mean value of the i-th parameter, then there holds $f_{ij}(\alpha_i) = \alpha_i$. The values $\underline{\gamma}_{ij}$ and $\overline{\gamma}_{ij}$ are the bounds for the provided characteristic $\mathbb{E}_\rho f_{ij}(\alpha_i)$ of the i-th parameter[1].

3.1 Decision Making

We assume that there are some bounds for all parameters $[\underline{\alpha}_i, \overline{\alpha}_i]$, $i = 1, ..., h$. This means that the i-th parameter belongs to the interval $[\underline{\alpha}_i, \overline{\alpha}_i]$ with probability 1. Inside this interval, the parameter is distributed according to an unknown probability density ρ_i.

So, we have some infinite set of discrete probability distributions $\pi(\vartheta_j, \alpha)$ defined by different parameters. Then the expected utility corresponding to one

[1]For simplicity, it is assumed that either experts with weights provide intervals for unknown parameters or experts without weights provide some statistical characteristics of random parameters. Of course, we could consider more complex cases when experts with weights provide statistical characteristics of random parameters, but the study of these, so-to-say third-order level, cases may hide the main results behind complex notation.

realization of the vector α is

$$\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha)=\sum_{j=1}^{m}\left(u(\lambda,\vartheta_j)\cdot\pi(\vartheta_j,\alpha)\right).$$

By averaging the expected utilities $\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha)$ over all possible vectors α, we get

$$\mathbb{E}_{\rho}\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha)=\int_{\Omega^h}\left(\sum_{j=1}^{m}\left(u(\lambda,\vartheta_j)\cdot\pi(\vartheta_j,\alpha)\right)\right)\rho(\alpha)\mathrm{d}\alpha.$$

Here Ω^h is a sample space and $\Omega^h=[\underline{\alpha}_1,\overline{\alpha}_1]\times\ldots\times[\underline{\alpha}_h,\overline{\alpha}_h]$.

Now we define an optimal action. An action λ^* is optimal iff

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*,\alpha)\right)\geq {}^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha)\right). \tag{15}$$

Here $\mathcal{P}$ is a set of all possible density functions $\rho(\alpha)$ satisfying the constraints

$$\underline{\gamma}_{ij}\leq\mathbb{E}_{\rho}f_{ij}(\alpha_i)\leq\overline{\gamma}_{ij},\ i=1,\ldots,h,\ j=1,\ldots,r_i,$$

or

$$\underline{\gamma}_{ij}\leq\int_{\underline{\alpha}_i}^{\overline{\alpha}_i}f_{ij}(\alpha_i)\rho_i(\alpha_i)\mathrm{d}\alpha_i\leq\overline{\gamma}_{ij},\ i=1,\ldots,h,\ j=1,\ldots,r_i.$$

Then the optimal action λ^* can be obtained by maximizing $^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha)\right)$ subject to $\sum_{s=1}^{n}\lambda_s=1$, $\lambda_s\geq 0$, $s=1,\ldots,n$. In other words, the following optimization problem has to be solved:

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*,\alpha)\right)\to\max_{\lambda_s} \tag{16}$$

under the constraints

$$\sum_{s=1}^{n}\lambda_s=1,\ \lambda_s\geq 0,\ s=1,\ldots,n. \tag{17}$$

If we assume that there is no information about independence of parameters, i.e., the joint density $\rho(\alpha)$ can not be represented as a product of marginal ones, then problem (16)-(17) can be rewritten as

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*,\alpha)\right)=\max_{c\in\mathbb{R},c_{kj}\in\mathbb{R}_+,d_{kj}\in\mathbb{R}_+,\lambda_s}\left\{c+\sum_{k=1}^{h}\sum_{j=1}^{r_k}\left(c_{kj}\underline{\gamma}_{kj}-d_{kj}\overline{\gamma}_{kj}\right)\right\} \tag{18}$$

subject to

$$c+\sum_{k=1}^{h}\sum_{j=1}^{r_k}\left(c_{kj}-d_{kj}\right)f_{kj}(\alpha_i)\leq\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha),\ \forall\alpha\in\Omega^h, \tag{19}$$

$$\sum_{s=1}^{n}\lambda_s=1,\ \lambda_s\geq 0. \tag{20}$$

This is a linear programming problem having an infinite number of constraints. However, for many special cases problem (18)-(20) can be simplified. Let us consider the most important and realistic case when experts provide h intervals $B_1,...,B_r$ for unknown parameters and each expert is characterized by some probability γ_{kj} or interval-valued probability $[\underline{\gamma}_{ij},\overline{\gamma}_{ij}]$. Moreover, in order to give the reader the essence of the subject analyzed and make all the formulas more readable, we will also assume that $h=1$ and $\alpha=(\alpha)$, i.e., there is only one parameter of the distribution $\pi(\vartheta_j,\alpha)$. We also denote r_1 by r. In other words, constraints (14) are represented as

$$\underline{\gamma}_j \leq \int_{\underline{\alpha}}^{\overline{\alpha}} I_{B_j}(\alpha)\rho(\alpha)\mathrm{d}\alpha \leq \overline{\gamma}_j,\ j=1,...,r. \tag{21}$$

Then problem (18)-(20) can be rewritten as

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^{*},\alpha)\right) = \max_{c\in\mathbb{R},c_k\in\mathbb{R}_+,d_k\in\mathbb{R}_+,\lambda_s}\left\{c+\sum_{k=1}^{r}\left(c_k\underline{\gamma}_k - d_k\overline{\gamma}_k\right)\right\} \tag{22}$$

subject to

$$c+\sum_{k=1}^{r}(c_k-d_k)\,I_{B_k}(\alpha) \leq \mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha),\ \forall\alpha\in[\underline{\alpha},\overline{\alpha}], \tag{23}$$

$$\sum_{s=1}^{n}\lambda_s = 1,\ \lambda_s \geq 0. \tag{24}$$

Denote $\mathbf{i}=(i_1,...,i_r),\ i_j\in\{0,1\}$. In accordance with possible values of the binary vector $\mathbf{i}$, the interval $B=[\underline{\alpha},\overline{\alpha}]$ of all values α can be divided into 2^r subintervals $B^{(1)},...,B^{(2^r)}$ such that the i-th subinterval is formed by

$$B^{(i)} = \bigcap_{k=1}^{r}\left\{\begin{array}{ll} B_k, & i_k=1 \\ B_k^c, & i_k=0 \end{array}\right. . \tag{25}$$

Let $L\subseteq\{1,...,2^r\}$ be a set of indices for all non-empty subintervals $B^{(j)}\neq\emptyset$. Then from all constraints corresponding to the subinterval $B^{(j)}$, we have to keep only one constraint

$$c+\sum_{k=1}^{r}(c_k-d_k)\,i_k \leq \min_{\alpha\in B^{(j)}}\mathbb{E}_{\pi}\mathbf{u}(\lambda,\alpha).$$

So, problem (22)-(24) becomes

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^{*},\alpha)\right) = \max_{c\in\mathbb{R},c_k\in\mathbb{R}_+,d_k\in\mathbb{R}_+,\lambda_s}\left\{c+\sum_{k=1}^{r}\left(c_k\underline{\gamma}_k - d_k\overline{\gamma}_k\right)\right\} \tag{26}$$

subject to

$$c + \sum_{k=1}^{r} (c_k - d_k)\, i_k \leq \min_{\alpha \in B^{(j)}} \mathbb{E}_\pi \mathbf{u}(\lambda, \alpha),\ \forall \mathbf{i}, \tag{27}$$

$$\sum_{s=1}^{n} \lambda_s = 1,\ \lambda_s \geq 0. \tag{28}$$

Let us introduce the variable $G_j = \min_{\alpha \in B^{(j)}} \mathbb{E}_\pi \mathbf{u}(\lambda, \alpha)$. Then problem (26)-(28) can be rewritten as

$$^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_\pi \mathbf{u}(\lambda^*, \alpha)\right) = \max_{c \in \mathbb{R}, c_k \in \mathbb{R}_+, d_k \in \mathbb{R}_+, \lambda_s, G_j} \left\{ c + \sum_{k=1}^{r} \left(c_k \underline{\gamma}_k - d_k \overline{\gamma}_k \right) \right\} \tag{29}$$

subject to

$$c + \sum_{k=1}^{r} (c_k - d_k)\, i_k \leq G_j,\ \forall \mathbf{i}, \tag{30}$$

$$\mathbb{E}_\pi \mathbf{u}(\lambda, \alpha) \geq G_j,\ \forall \alpha \in B^{(j)},\ \forall \mathbf{i}, \tag{31}$$

$$\sum_{s=1}^{n} \lambda_s = 1,\ \lambda_s \geq 0. \tag{32}$$

In this case, we obtain the linear programming problem with infinite number of constraints. However, if it is known that the function $\mathbb{E}_\pi \mathbf{u}(\lambda, \alpha)$ is monotone with α, then it is sufficient to consider only boundary points of intervals $B^{(j)}$. Constraints (31) can be written as

$$\sum_{j=1}^{m} \left(\sum_{s=1}^{n} (u(a_s, \vartheta_j)\lambda_s) \cdot \pi(\vartheta_j) \right) \leq G_j,$$

or

$$\sum_{s=1}^{n} \left(\sum_{j=1}^{m} (u(a_s, \vartheta_j)\pi(\vartheta_j)) \right) \lambda_s \leq G_j.$$

Hence it is obvious that the constraints are linear with λ_s.

3.2 Numerical Example

Suppose that 3 states $\{1, 2, 3\}$ of nature are governed by the binomial distribution

$$\pi(\vartheta_j, \alpha) = \binom{3-1}{j-1} \alpha^{j-1} (1-\alpha)^{3-j-1},\ j = 1, 2, 3.$$

Two experts provide their judgements about the parameter $\alpha \in [0, 1]$ as follows:

1. the parameter α is in interval $[0.8,1]$;
2. the parameter α is in interval $[0.7,1]$.

The belief in the correctness of the first expert is 0.5. The belief in the second expert is between 0.3 and 1 (see Section 2.2). The above judgements can be written in the formal form as follows:

$$\int_0^1 I_{[0.8,1]}(\alpha)\rho(\alpha)\mathrm{d}\alpha = 0.5,\ \int_0^1 I_{[0.7,1]}(\alpha)\rho(\alpha)\mathrm{d}\alpha \in [0.3,1].$$

Let us find the set $L \subseteq \{1,2,3,4\}$.

i	intervals	non-empty
$(1,1)$	$[0.8,1]\cap[0.7,1]$	yes
$(1,0)$	$[0.8,1]\cap[0,0.7]$	no
$(0,1)$	$[0,0.8]\cap[0.7,1]$	yes
$(0,0)$	$[0,0.8]\cap[0,0.7]$	yes

Table 3: Intersections of intervals

It can be seen from Table 3 that $L = \{1,3,4\}$.
Let us find λ_1, λ_2. In this case, there holds

$${}^{L}\mathbb{E}_{\mathcal{P}}\left(\mathbb{E}_{\pi}\mathbf{u}(\lambda^*,\alpha)\right) = \max_{c\in\mathbb{R},c_k\in\mathbb{R}_+,d_k\in\mathbb{R}_+,\lambda_s,G_j}\{c+0.5c_1-0.5d_1+0.3c_2-1d_2\}$$

subject to

$$\begin{aligned} c+1\cdot(c_1-d_1)+1\cdot(c_2-d_2) &\leq G_1,\\ c+0\cdot(c_1-d_1)+1\cdot(c_2-d_2) &\leq G_3,\\ c+0\cdot(c_1-d_1)+0\cdot(c_2-d_2) &\leq G_4,\\ \left(\alpha^2-6\alpha+6\right)\lambda_1+\left(10\alpha-8\alpha^2+2\right)\lambda_2 &\geq G_1,\ \alpha\in[0.8,1],\\ \left(\alpha^2-6\alpha+6\right)\lambda_1+\left(10\alpha-8\alpha^2+2\right)\lambda_2 &\geq G_3,\ \alpha\in[0.7,0.8],\\ \left(\alpha^2-6\alpha+6\right)\lambda_1+\left(10\alpha-8\alpha^2+2\right)\lambda_2 &\geq G_3,\ \alpha\in[0,0.7],\\ \lambda_1+\lambda_2 = 1,\ \lambda_1 &\geq 0, \lambda_2 \geq 0. \end{aligned}$$

By solving this problem approximately (for a finite number of values of α), we get $c = 3.636$, $G_1 = 2.773$, $G_3 = G_4 = 3.636$, $c_1 = c_2 = d_2 = 0$, $d_1 = 0.864$, $\lambda_1 = 0.409$, $\lambda_2 = 0.591$.

4 Concluding Remarks

Two models of decision making based on different types of initial hierarchical information about states of nature have been studied in the paper. We have shown that both models can be brought into a form which allows us to give general algorithms to determine optimal solutions.

It should be noted that we have focused in this paper on the basic decision problem. However, the fundamental ideas of this paper should be also applicable to more complex decision problems, like multi-criteria decision making and data-based decision problems. Another topic of furhter research is to extend the results obtained here to other optimality criteria which are more sophisticated than the criteria from (2) and (15), which take into account only the lower interval limits.

Acknowledgement

The first author gratefully acknowledges financial support of the Alexander von Humboldt Foundation. We both thank Kurt Weichselberger, Franz Ferschl, Martin Gümbel, Bernhard Rüger, Heinrich Strecker and Anton Wallner for numerous comments and stimulating discussions during the Munich weekly seminar on interval probability.

References

[1] AUGUSTIN, T. On decision making under ambiguous prior and sampling information. In *Imprecise Probabilities and Their Applications. Proc. of the 2nd Int. Symposium ISIPTA'01* (Ithaca, USA, June 2001), G. de Cooman, T. Fine, and T. Seidenfeld, Eds., Shaker Publishing, pp. 9–16.

[2] AUGUSTIN, T. Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity. *Statistical Papers 43* (2002), 5–22.

[3] BERGER, J. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 1985.

[4] DE COOMAN, G. Possibilistic previsions. In *Proceedings of IPMU'98* (Paris, 1998), vol. 1, Editions EDK, pp. 2–9.

[5] DE COOMAN, G. Precision–imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. *Journal of Statistical Planning and Inference 105*, 1 (2002), 175–198.

[6] DE COOMAN, G., AND WALLEY, P. A possibilistic hierarchical model for behaviour under uncertainty. *Theory and Decision 52*, 4 (2002), 327–374.

[7] EKENBERG, L., AND THORBIÖRNSON, J. Second-order decision analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9* (2 2001), 13–38.

[8] GHIRARDATO, P., AND MARINACCI, M. Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research 26* (2001), 864–890.

[9] GILBERT, L., DE COOMAN, G., AND KERRE, E. Practical implementation of possibilistic probability mass functions. In *Proceedings of Fifth Workshop on Uncertainty Processing (WUPES 2000)* (Jindvrichouv Hradec, Czech Republic, June 2000), pp. 90–101.

[10] GOOD, I. Some history of the hierarchical Bayesian methodology. In *Bayesian Statistics*, J. Bernardo, M. DeGroot, D. Lindley, and A. Smith, Eds. Valencia University Press, Valencia, 1980, pp. 489–519.

[11] GOODMAN, I. R., AND NGUYEN, H. T. Probability updating using second order probabilities and conditional event algebra. *Information Sciences 121*, 3-4 (1999), 295–347.

[12] KOZINE, I., AND UTKIN, L. Processing unreliable judgements with an imprecise hierarchical model. *Risk Decision and Policy 7*, 3 (2002), 325–339.

[13] KUZNETSOV, V. P. *Interval Statistical Models*. Radio and Communication, Moscow, 1991. in Russian.

[14] NAU, R. F. Indeterminate probabilities on finite sets. *The Annals of Statistics 20* (1992), 1737–1767.

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

[16] UTKIN, L. Imprecise second-order hierarchical uncertainty model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11*, 3 (2003). To Appear.

[17] UTKIN, L., AND KOZINE, I. Different faces of the natural extension. In *Imprecise Probabilities and Their Applications. Proc. of the 2nd Int. Symposium ISIPTA'01* (Ithaca, USA, June 2001), G. de Cooman, T. Fine, and T. Seidenfeld, Eds., Shaker Publishing, pp. 316–323.

[18] WALLEY, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

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

[20] WEICHSELBERGER, K. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung*, vol. I Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg, 2001.

[21] ZELLNER, A. *An introduction to Bayesian Inference in Econometrics*. Wiley, New York, 1971.

Lev Utkin is with the Department of Computer Science, St.Petersburg Forest Technical Academy, Institutski per. 5, 194021 St.Petersburg, Russia.
Phone +7/812/247693
E-mail: lvu@utkin.usr.etu.spb.ru

Thomas Augustin is with the Department of Statistics, Ludwig-Maximilians-University of Munich, Ludwigstr. 33, D-80539 Munich, Germany.
Phone +49/89/2180-3520 Fax +49/89/2180-5044
E-mail: augustin@stat.uni-muenchen.de