Powerful algorithms for decision making under partial prior information and general ambiguity attitudes

L.V. Utkin
Department of Computer Science
St.Petersburg State Forest Technical Academy
Russia
lvu@utkin.usr.etu.spb.ru

Th. Augustin
Department of Statistics
University of Munich
Germany
Thomas@stat.uni-muenchen.de

Abstract

This paper discusses decision making in the practically important situation where only partial prior information on the stochastic behavior of the states of nature expressed by imprecise probabilities (interval probability) is available. For this situation, in literature several optimality criteria have been suggested and investigated theoretically. Practical computation of optimal solutions, however, is far from being straightforward. The paper develops powerful algorithms for determining optimal actions under arbitrary ambiguity attitudes and the criterion of E-admissibility. The algorithms are based on linear programming and can be implemented by standard software.

Keywords: ambiguity attitudes, Choquet expected utility, decision making, E-admissibility, imprecise probabilities, interval probability, interval statistical models, linear programming, MaxEMin criterion, maximality, maxmin expected utility model, minimality, partial prior information, structure dominance

1 Introduction

Decision theory provides an elegant formal framework for determining optimal actions under uncertainty on the states of nature. Practical application of the popular Bernoulli criterion of maximal expected utility, however, is limited by the fact that quite often the uncertainty is too complex to be adequately described by a classical, i.e. precise, probability distribution. For instance, Ellsberg's experiments (cf. [7]) make it clear without any ifs and buts that ambiguity, i.e. the extent of deviation from ideal stochasticity, plays a constitutive role in decision making that must not be neglected: simply applying criteria successful in classical decision theory to situations of complex uncertainty, where only partial information about the states of nature is available, would lead to deceptive conclusions.

To take ambiguity into account properly, a substantial generalization of the concept of probability has been developed, known as the theory of credal sets [15], imprecise probabilities [23], interval probability [25] or interval statistical models [14], also comprising belief functions (e.g. [19], [27]) and convex (or two-monotone) capacities (e.g. [5]). Based on this comprehensive framework to model partial information alternative optimality criteria for decision making have been developed to generalize the classical notion of expected utility (cf., in particular, the recent survey by [20]).

The main issue is that now, in general, the expected utility is interval-valued. Two types of criteria may then be suggested: The first one, like the criterion of maximality (as proposed by [23]) or the concept of E-admissibility (advocated by [15], [16]), renounces the completeness of the ordering and generalizes the concept of admissibility by distinguishing a set of actions as being not inferior. This set consists of all actions that are optimal with respect to at least one classical, i.e. real-valued, probability distribution on the states of nature consistent with the available partial information.

In contrast, the second branch of criteria aims at a complete ordering of the actions, and so the interval-valued expected utility eventually has to be transformed to the real line. Which represen-

tation of intervals is used depends on the ambiguity attitude. The most conservative choice is strict ambiguity aversion, where one concentrates on the lower interval limit only, leading to a popular criterion known under different names (see Section 3).

Just as the minimax criterion used in classical decision theory, this criterion, however, turns out to be overpessimistic, and more sophisticated representations of interval-valued expected utility are highly desirable that take additionally the decision maker's attitude towards ambiguity into consideration (see [25, Chapter 2.6] for a detailed discussion of this topic). It is often overseen that already Ellsberg [7, p. 664] himself was aware of this problem and suggested to consider a Hurwicz-like compromise between lower and upper expected utility (see also, e.g., [11], [17], [26], [1]).

In more complex situations practical application of all these criteria is limited by the fact that the computation of optimal actions is far from being straightforward. [1] gives a general algorithm to calculate optimal solutions under strict ambiguity aversion by linear programming. In large problems, however, this procedure may become intractable since the constraints are defined by the vertices of the underlying polyhedron of classical probabilities, and so this number may become very large. Moreover, the vertices often have to be calculated in advance by separate optimization problems.

This paper presents substantial further developments on computational issues in generalized decision making: Section 2 briefly states the decision problem more formally and introduces the notation used throughout the paper. The pessimistic decision making based on strict ambiguity aversion is considered in Section 3, where, in particular, a general algorithm based on partial dualization of the original problem is given. It still allows to calculate optimal decisions by a single linear programming problem, but now manages to achieve this without relying on the vertices. Then in Section 4 both algorithms, the one based on the vertices as well the dual one, are extended to arbitrary ambiguity attitudes expressed by a caution parameter – a step that seems to be easy at a first glance but proves to demand some care. Both ways to proceed are compared, generally and in an example. Section 5 turns to the first type of criteria, derives a general algorithm for calculating all E-admissible actions and briefly glances at the criterion of maximality.

2 Problem statement

Consider the basic model of decision theory: one has to choose an optimal *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*,$u : (\mathbb{A} \times \Theta) \to \mathbb{R}$, which can be represented in the following table:

	ϑ_1	...	ϑ_j	...	ϑ_m
a_1	$u(a_1, \vartheta_1)$		...		$u(a_1, \vartheta_m)$
$\vdots$		$\ddots$			
a_s	$\vdots$		$u(a_i, \vartheta_j)$		$\vdots$
$\vdots$				$\ddots$	
a_n	$u(a_n, \vartheta_1)$		...		$u(a_n, \vartheta_m)$

and by the associated random variable $\mathbf{u}(a)$ on $(\Theta, \mathcal{P}o(\Theta))$ taking the values $u(a, \vartheta)$.[1] Often it makes sense to study *randomized actions*,[2] which can be understood as a classical probability measure $\lambda = (\lambda_1, ..., \lambda_n)$ on $(\mathbb{A}, \mathcal{P}o(\mathbb{A}))$, where λ_i is interpreted as the probability with which action a_i is taken. 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 if the corresponding probability mass function $\pi(\cdot)$ on Θ is completely known, the Bernoulli principle is nearly unanimously favored. One chooses the action λ^* maximizing the expected utility $\mathbb{E}_\pi \mathbf{u}(\lambda) := \sum_{j=1}^{m} (u(\lambda, \vartheta_j) \cdot \pi(\vartheta_j))$ among all λ.

As argued in the introduction, in many applications however it is not possible to describe the prior knowledge on the stochastic behavior of the states of nature by a classical probability, and a more general description of uncertainty is needed, as provided by imprecise probabilities and related approaches.

[1] For a simple example see Section 4.3.

[2] Next to simplifying calculations, under some criteria the optimal randomized action may be really superior to the optimal unrandomized ones, cf. e.g. [1].

From the technical point of view, the usual concepts of imprecise probability lead to sets of classical probabilities. In the sequel, we additionally assume that these sets are (convex) polyhedra, i.e. that the available information about the states of nature can be represented by a set of r lower and upper expectations (previsions), $\underline{b}_i$ and $\overline{b}_i$, of functions f_i, $i = 1, ..., r$, defined on Θ. So we have

$$\underline{b}_i \leq \mathbb{E}_\pi f_i = \sum_{j=1}^{m} f_i(\vartheta_j)\pi(\vartheta_j) \leq \overline{b}_i,\ i = 1, ..., r, \quad (1)$$

such that also the trivial constraints $\pi(\vartheta_j) \geq 0$ for all j and $\sum_{j=1}^{n} \pi(\vartheta_j) = 1$ are included. Interval probabilities fit into this frame by considering indicator functions, and also bounds on conditional classical probabilities can be brought into this form. Moreover, comparative probabilities, i.e. assignments of the form $\pi(\vartheta_j) \leq \pi(\vartheta_l)$ are included, too. This set of inequalities restricts all mass functions π on Θ to a set $\mathcal{M}$ such that every π from $\mathcal{M}$ satisfies all the inequalities. Moreover, every π from $\mathcal{M}$ produces a classical expected utility $\mathbb{E}_\pi \mathbf{u}(\lambda)$. Since the set $\mathcal{M}$ is convex, assuming $\mathcal{M} \neq \varnothing$, all possible expected utilities $\mathbb{E}_\pi \mathbf{u}(\lambda)$ range within the interval

$$\left[\, \underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda)\,,\, \overline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) \,\right], \quad (2)$$

$$\text{where} \qquad \underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) = \min_{\pi \in \mathcal{M}} \sum_{j=1}^{m} u_j(\lambda) \cdot \pi_j\,, \quad (3)$$

$$\text{and} \qquad \overline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) = \max_{\pi \in \mathcal{M}} \sum_{j=1}^{m} u_j(\lambda) \cdot \pi_j\,. \quad (4)$$

This interval-valued quantity is called *generalized expected utility*[3]. The difference between upper and lower limit reflects the ambiguity.

In contrast to the classical case of real-valued expected utility, where the usual ordering on the real line can be used directly to judge which action is optimal, now different criteria make sense, depending on the attitude towards ambiguity. (See the discussions in the introduction and at the beginnings of the Section 3, 4 and 5.) Before we discuss their efficient handling, let us introduce for simplicity some abbreviations:

$$\begin{array}{llll} \pi & :=(\pi_1, \ldots, \pi_m), & \pi_j & :=\pi(\vartheta_j), \\ u_{ij} & :=u(a_i, \vartheta_j), & \underline{u}_j(\lambda) & :=\underline{u}(\lambda, \vartheta_j), \\ \underline{\mathbf{B}} & :=(\underline{b}_1, ..., \underline{b}_r), & \overline{\mathbf{B}} & :=(\overline{b}_1, ..., \overline{b}_r), \\ \mathbf{1} & :=(1, ..., 1)^{\mathrm{T}}, & \mathbf{F}_j & :=(f_1(\vartheta_j), ..., f_r(\vartheta_j)). \end{array}$$

The obvious constraints $\lambda_i \geq 0$ will be omitted in most places.

3 Two algorithms for pessimistic decision making

Under strict ambiguity aversion every action is evaluated by its minimal expected utility, and so, the interval-valued expectations in (2) are represented by the lower interval limits alone.
Then an action λ^* is optimal iff for all λ

$$\underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda). \quad (5)$$

This criterion has been proposed under different names. It corresponds to the Gamma-Minimax criterion (as considered, e.g., in [4, Section 4.7.6], [22]), to the Maxmin expected utility model [10], to the MaxEMin criterion investigated by [13] (cf. also [12] and the references therein) and the notion of maximinity in [23]. In the case of two-monotone capacities it is equivalent to maximizing Choquet expected utility (as studied, e.g., in [5]).

The optimal randomized action λ^* can be obtained by maximizing $\underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda)$ subject to $\lambda \cdot \mathbf{1} = 1$. In other words, the following optimization problem has to be solved:

$$\underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) \to \max_{\lambda}$$

under the restrictions $\lambda \cdot \mathbf{1} = 1$. By substituting the expression for $\mathbb{E}_\pi \mathbf{u}(\lambda)$ into the objective function, we get

$$\min_{\pi \in \mathcal{M}} \sum_{j=1}^{m} (u_j(\lambda) \cdot \pi_j) \to \max_{\lambda} \quad (6)$$

subject to $\lambda \cdot \mathbf{1} = 1$.[4] Two approaches for solving the above problem will be proposed.

[3] For rigorous axiomatic justification of generalized expected utility in the sense of (2) and different criteria derived from it see among others [8], [9], [18], as well as the references therein, who manage to extend Neumann-Morgenstern and Anscombe-Aumann theory to the situation of complex uncertainty with partial prior information.

[4] We formulate all algorithms in terms of randomized actions. If consideration is confined to unrandomized actions, the derived optimization problems still can be used by passing over to Boolean optimization. In this case, however, depending on the situation, it may be more efficient to calculate the lower and upper expectation (by linear programming, or by Choquet-integration in the case of two-monotone lower probabilities) for every action a_i and then to compare

3.1 An approach based on extreme points

The first approach has been suggested by [1, 3]. After having introduced a new variable $G = \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi \mathbf{u}(\lambda)$, the problem (6) can be rewritten as

$$\max_{\lambda, G} G \tag{7}$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$ and

$$G \leq \sum_{j=1}^{m} \left(\sum_{s=1}^{n} u_{sj} \lambda_s \right) \pi_j, \quad \forall \pi \in \mathcal{M}. \tag{8}$$

The optimization problem (7)-(8) is linear, but, in the way it is written, it contains one constraint for every $\pi \in \mathcal{M}$, i.e. infinitely many constraints in general. In order to overcome this difficulty, note, however, that the set of distributions $\mathcal{M}$ can be viewed as a simplex in a finite dimensional space. According to some general results from linear programming theory, for every fixed λ, $\sum_{j=1}^{m} u_j(\lambda) \cdot \pi_j$ attains its minimum at an extreme point of the convex polyhedron $\mathcal{M}$. Since the set $\mathcal{E}(\mathcal{M})$ of extreme points is finite, this implies that the infinite set of constraints (8) is reduced to some finite set, and standard routines for linear programming can be used to determine optimal actions. Finally, we have the following linear programming problem for computing the optimal randomized action:

$$\max_{\lambda, G} G \tag{9}$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$, and

$$G \leq \sum_{j=1}^{m} \left(\sum_{s=1}^{n} u_{sj} \lambda_s \right) \pi_j, \quad \forall \pi \in \mathcal{E}(\mathcal{M}). \tag{10}$$

It is important to recall that this approach for solving the decision problem requires the extreme points of $\mathcal{M}$. In the case where prior information is given by two-monotone probabilities or belief functions closed expressions are available (as used in the corollaries in [1],[3]). In general, however, this task is computationally expensive and may lead to the need to solve a number of linear optimization problems in advance. Moreover, the number of extreme points may become as large as $m!$ [24], and so the number of constraints may become very large. The following approach allows us to avoid this difficulty; solving one linear programming problem only will prove to be sufficient.

3.2 An approach based on partial dualization

This solution is based on the idea of partial dualization of the problem in (6) with variables π and λ. Take for the moment λ as fixed and replace the remaining optimization problem with variables π by the dual one (cf., e.g. ,[6]). Writing $\mathbf{C} = (c_1, ..., c_r)^{\mathrm{T}}$, $\mathbf{D} = (d_1, ..., d_r)^{\mathrm{T}}$, the dual problem is of the form

$$\max_{c, \mathbf{C}, \mathbf{D}} \left\{ c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D} \right\}$$

subject to $c \in \mathbb{R}$, $\mathbf{C}, \mathbf{D} \in \mathbb{R}^r_+$, and

$$c + \mathbf{F}_j \left(\mathbf{C} - \mathbf{D} \right) \leq u_j(\lambda), \ j = 1, ..., m.$$

Here $c, \mathbf{C}, \mathbf{D}$ are optimization variables such that the variable c corresponds to the constraint $\sum_{j=1}^{m} \pi_j = 1$ in the primal form, c_i corresponds to the constraints $\underline{b}_i \leq E_\pi f_i$ and d_i corresponds to the constraints $E_\pi f_i \leq \overline{b}_i$. As pointed out in [21], the dual problem has the same form as the natural extension [14, 23] in the framework of imprecise probabilities.

Using the fact that at the optimum the objective functions of the primal and the dual problem attain the same value, i.e., that for fixed λ,

$$\min_{\pi \in \mathcal{M}} \sum_{j=1}^{m} \left(u_j(\lambda) \cdot \pi_j \right) = \max_{c, \mathbf{C}, \mathbf{D}} \left\{ c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D} \right\},$$

the maximization over λ can be combined with the dual optimization problem as follows:

$$\max_{c, \mathbf{C}, \mathbf{D}, \lambda} \left\{ c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D} \right\} \tag{11}$$

subject to $c \in \mathbb{R}$, $\mathbf{C}, \mathbf{D} \in \mathbb{R}^r_+$, $\lambda \cdot \mathbf{1} = 1$ and

$$c + \mathbf{F}_j \left(\mathbf{C} - \mathbf{D} \right) \leq u_j(\lambda), \ j = 1, ..., m, \tag{12}$$

We obtain a linear optimization problem with variables c, $\mathbf{C}$, $\mathbf{D}$, λ. This implies that the optimal

---

the resulting values in order to find an optimal action.
In addition we would like to recall that – next to the issue of easy computation – the representation as a linear programming problem also provides plenty of theoretical insight (see for instance the proof Lemma 2 here or the argumentation in [1]).

randomized action is found by solving the single linear programming problem (11)-(12).[5]

4 Decision making with the caution parameter

Already in the introduction it has been emphasized that relying on strict ambiguity aversion may be overpessimistic, and the upper interval limit of the generalized expected utility should also be taken into account. With respect to this more sophisticated criterion, an action λ^* is optimal iff for all λ

$$\begin{aligned}&\eta\underline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda^*) + (1-\eta)\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda^*)\\ &\geq \eta\underline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda) + (1-\eta)\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)\,,\end{aligned} \tag{13}$$

with $\underline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)$ and $\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)$ from (3) and (4). The caution parameter η reflects the degree of ambiguity aversion (e.g., [25, Chapter 2.6]); the more ambiguity averse the decision maker is, the higher is the influence of the lower interval limit of the generalized expected utility. $\eta = 1$ corresponds to strict ambiguity aversion (pessimistic decision), i.e. the case considered above, $\eta = 0$ expresses maximal ambiguity seeking attitudes (optimistic decision). The value $1-\eta$ can be interpreted as the probability that nature will turn out as favorably as possibly towards us as decision makers [17].

By using the above criterion, the following optimization problem has to be solved:

$$\eta\underline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda) + (1-\eta)\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda) \to \max_{\lambda} \tag{14}$$

subject to the constraints $\lambda \cdot \mathbf{1} = 1$.
The generalization of the algorithms developed in the previous section needs some attention. Applying the method from Section 3.1 now leads to a bilinear optimization problem (cf. [1, Section 5.2]). The idea to bound the objective function by a penalty function [1, p. 19], however, fails, and it seems not possible to reduce the calculation of the optimum to a single linear optimization problem. However, still solutions based on linear programming can be obtained; again there are two ways, depending whether the primal program or its dual is used:

[5] Some remarks on the comparison of the two algorithms will be made at the end of Section 4 in the extended setting considered there.

4.1 A solution based on the extreme points

For the generalization of the procedure described in Section 3.1 based on the variable $G = \min_{\pi\in\mathcal{M}} \mathbb{E}_\pi \mathbf{u}(\lambda)$, relation (14) can be rewritten as

$$\eta \cdot G + (1-\eta)\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda) \to \max_{\lambda, G}$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$, and

$$G \leq \sum_{j=1}^{m}\left(\sum_{s=1}^{n} u_{sj}\lambda_s\right)\pi_j, \quad \forall \pi \in \mathcal{E}(\mathcal{M}). \tag{15}$$

Note, however, that now the following asymmetry occurs: Although $\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)$ is attained at an element of $\mathcal{E}(\mathcal{M})$, it is not possible to introduce a slack variable for describing the value of $\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)$, too. The objective function would be unbounded. Instead, one has to consider linear programming problems where in the objective function $\overline{\mathbb{E}}_{\mathcal{M}}\mathbf{u}(\lambda)$ is replaced by $\mathbb{E}_{\pi^{(z)}}\underline{\mathbf{u}}(\lambda)$ for some extreme point $\pi^{(z)} \in \mathcal{E}(\mathcal{M})$, i.e.,

$$\eta \cdot G + (1-\eta)\sum_{j=1}^{m}\left(\sum_{s=1}^{n} u_{sj}\lambda_s\right)\pi_j^{(z)} \to \max_{\lambda}.$$

Taking the maximum over all these objective functions arising from all elements of $\mathcal{E}(\mathcal{M})$ must give the optimal value of (14), and the corresponding solution $\lambda(z)$ gives the optimal action under the criterion (13).

In sum, in order to find an optimal randomized action, we have to find all extreme points of the set $\mathcal{M}$ and to solve the linear optimization problem for every extreme point.

4.2 Dual form for solving the problem

An alternative approach for solving the problem is to replace optimization problems (3)-(4) by dual ones. Denote additionally

$$\mathbf{V} = (v_1, ..., v_r)^{\mathrm{T}},\ \mathbf{W} = (w_1, ..., w_r)^{\mathrm{T}}.$$

By using notations introduced in the previous sections and the approach described in Section 3.2, we can rewrite (14) as follows:

$$\begin{aligned}&\eta \cdot \max_{c,\mathbf{C},\mathbf{D}} \left\{c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D}\right\} +\\ &(1-\eta) \cdot \min_{v,\mathbf{V},\mathbf{W}} \left\{v + \overline{\mathbf{B}}\mathbf{V} - \underline{\mathbf{B}}\mathbf{W}\right\} \to \max_{\lambda}\end{aligned}$$

subject to $c, v \in \mathbb{R}$, $\mathbf{C}, \mathbf{D}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^r_+$, and

$$c + \mathbf{F}_j(\mathbf{C} - \mathbf{D}) \le u_j(\lambda),\ j = 1, ..., m,$$
$$v + \mathbf{F}_j(\mathbf{V} - \mathbf{W}) \ge u_j(\lambda),\ j = 1, ..., m,$$
$$\lambda \cdot \mathbf{1} = 1.$$

By introducing a new variable H such that

$$H = \min\left\{v + \overline{\mathbf{B}}\mathbf{V} - \underline{\mathbf{B}}\mathbf{W}\right\},$$

we get the following equivalent problem:

$$\eta\left\{\max_{c,\mathbf{C},\mathbf{D}}\left(c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D}\right)\right\} + (1-\eta)H \to \max_{\lambda}$$

subject to $c, v \in \mathbb{R}$, $\mathbf{C}, \mathbf{D}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^r_+$, and

$$c + \mathbf{F}_j(\mathbf{C} - \mathbf{D}) \le u_j(\lambda),\ j = 1, ..., m, \quad (16)$$
$$v + \mathbf{F}_j(\mathbf{V} - \mathbf{W}) \ge u_j(\lambda),\ j = 1, ..., m, \quad (17)$$
$$v + \overline{\mathbf{B}}\mathbf{V} - \underline{\mathbf{B}}\mathbf{W} \ge H,\ \ \lambda \cdot \mathbf{1} = 1. \quad (18)$$

This problem is reduced to

$$\eta\left\{c + \underline{\mathbf{B}}\mathbf{C} - \overline{\mathbf{B}}\mathbf{D}\right\} + (1-\eta)H \to \max_{c,v,\mathbf{C},\mathbf{D},\mathbf{V},\mathbf{W},\lambda,H} \quad (19)$$

subject to $c, v \in \mathbb{R}$, $\mathbf{C}, \mathbf{D}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^r_+$, and (16)-(18).

However, a problem similar to above occurs: by maximizing the objective function, the variable H will unrestrictedly increase with variables v, v_k, and w_k. How to restrict these variables? Let us consider in detail the problem

$$H^* = \min\left\{v + \overline{\mathbf{B}}\mathbf{V} - \underline{\mathbf{B}}\mathbf{W}\right\} \quad (20)$$

subject to $v \in \mathbb{R}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^r_+$, and

$$v + \mathbf{F}_j(\mathbf{V} - \mathbf{W}) \ge u_j(\lambda),\ j = 1, ..., m. \quad (21)$$

This linear programming problem has $2r+1$ variables and $2r+m$ constraints, where $2r$ constraints are of the form: $v_k \ge 0$ and $w_k \ge 0$, $i = 1, ..., r$. This implies that $2r+1$ equalities among $2r+m$ inequalities have to take place. These equalities restrict the values of v_k, w_k, and H. So, it is necessary to solve $\binom{2r+m}{2r+1}$ linear optimization problems having $2r+1$ equalities in constraints and to choose the maximal solution from all possible ones. However, not all solutions to problem (19) satisfy (20)-(21), that is, the optimal value of H in (19) may be different from H^* in (20)-(21) by the same optimal λ. Therefore, after solving problem (19), optimal values of λ are substituted into (21) and problem (20)-(21) has to be solved. If $H \ne H^*$, then the obtained optimal solution λ must be removed from the list of possible ones. If $H = H^*$, then the obtained optimal solution λ is a candidate for the final optimal solution to (19). This implies that $\binom{2r+m}{2r+1}$ linear optimization problems (20)-(21) have to be additionally solved.

4.3 Numerical toy example

Let $\Theta = \{1, 2, 3\}$ and suppose that an expert provides the following two judgements: the mean value of states is less than 2 ($\pi_1 + 2\pi_2 + 3\pi_3 \le 2$); the probability of the third state is less than 0.3 ($\pi_3 \le 0.3$). In order to use the proposed algorithms, this information is formally represented as $1 \le \mathbb{E}_\pi \vartheta \le 2$ and $0 \le \mathbb{E}_\pi I_{\{3\}}\vartheta \le 0.3$ (cp. equation (1).). Here $I_{\{3\}}\vartheta$ is the indicator function taking the value 1 if $\vartheta = \vartheta_3$ and 0 if $\vartheta \ne \vartheta_3$. So, we have $r = 2$, $m = 3$, $\underline{b}_1 = 1$, $\overline{b}_1 = 2$, $\underline{b}_2 = 0$, $\overline{b}_2 = 0.3$. The utility function u_{sj} is given in Table 1.

Table 1: Values of utilities $u(a_s, \vartheta_j)$

	ϑ_1	ϑ_2	ϑ_3
a_1	6	3	1
a_2	2	7	4

Suppose $\eta = 0.6$. Let us firstly find the optimal action by using the approach relying on the extreme points. The set of classical probabilities induced by the given judgements has the following four extreme points $(\pi_1^{(l)}, \pi_2^{(l)}, \pi_3^{(l)})$:

$$(1, 0, 0),\ (0, 1, 0),\ (0.7, 0, 0.3),\ (0.3, 0.4, 0.3).$$

Then the following optimization problem can be written:

$$H = \eta G + (1-\eta)((6\pi_1^{(l)} + 3\pi_2^{(l)} + 1\pi_3^{(l)})\lambda_1$$
$$+(2\pi_1^{(l)} + 7\pi_2^{(l)} + 4\pi_3^{(l)})\lambda_2) \to \max_{\lambda_1,\lambda_2,G}$$

subject to $\lambda_1 + \lambda_2 = 1$ and

$$6\lambda_1 + 2\lambda_2 \ge G, \qquad 4.5\lambda_1 + 2.6\lambda_2 \ge G,$$
$$3\lambda_1 + 7\lambda_2 \ge G. \qquad 3.3\lambda_1 + 4.6\lambda_2 \ge G.$$

By solving 4 optimization problems for every $z = 1, ..., 4$, we choose the optimal randomized action

(λ_1, λ_2) which provides the maximal objective function

$$z = 1,\ G = \frac{31}{9}, \lambda_1 = \frac{8}{9}, \lambda_2 = \frac{1}{9},\ H = 4.29,$$
$$z = 2,\ G = \frac{22}{7}, \lambda_1 = \frac{2}{7}, \lambda_2 = \frac{5}{7},\ H = 4.23,$$
$$z = 3, 4,\ G = \frac{303}{80}, \lambda_1 = \frac{5}{8}, \lambda_2 = \frac{3}{8},\ H = 3.79.$$

It can be seen from the numerical results, that the optimal action is $\lambda_1 = \frac{8}{9}$, $\lambda_2 = \frac{1}{9}$.

Now the dual approach for solving the problem will be illustrated: The optimization problem (19) is written as

$$\eta \{c + c_1 - 2d_1 + 0c_2 - 0.3d_2\} + (1-\eta)H \to \max_{c,v,\mathbf{C},\mathbf{D},\mathbf{V},\mathbf{W},\lambda,H}$$

subject to $c, v, H \in \mathbb{R}, \mathbf{C}, \mathbf{D}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^2_+$, and

$$\begin{aligned}
c + 1(c_1 - d_1) + 0(c_2 - d_2) &\le 6\lambda_1 + 2\lambda_2,\\
c + 2(c_1 - d_1) + 0(c_2 - d_2) &\le 3\lambda_1 + 7\lambda_2,\\
c + 3(c_1 - d_1) + 1(c_2 - d_2) &\le 1\lambda_1 + 4\lambda_2,\\
v + 1(v_1 - w_1) + 0(v_2 - w_2) &\ge 6\lambda_1 + 2\lambda_2,\\
v + 2(v_1 - w_1) + 0(v_2 - w_2) &\ge 3\lambda_1 + 7\lambda_2,\\
v + 3(v_1 - w_1) + 1(v_2 - w_2) &\ge 1\lambda_1 + 4\lambda_2,\\
v + 2v_1 - 1w_1 + 0.3v_2 - 0w_2 &\ge H,\\
\lambda_1 + \lambda_2 &= 1.
\end{aligned}$$

Since we have 5 variables v, v_1, w_1, v_2, w_2 (for the subproblem (20)-(21)), the number of optimization problems to be solved is 21. By replacing symbols "$\ge$" in inequalities containing at least one of the variables v, v_1, w_1, v_2, w_2 by symbols "$=$", we can write 21 linear optimization problems. The maximum of the objective function is achieved if $v_1 = v_2 = 0$ and

$$\begin{aligned}
v + 1(v_1 - w_1) + 0(v_2 - w_2) &= 6\lambda_1 + 2\lambda_2,\\
v + 2(v_1 - w_1) + 0(v_2 - w_2) &= 3\lambda_1 + 7\lambda_2,\\
v + 3(v_1 - w_1) + 1(v_2 - w_2) &= 1\lambda_1 + 4\lambda_2.
\end{aligned}$$

At that $\lambda_1 = \frac{8}{9}$, $\lambda_2 = \frac{1}{9}$, $H = \frac{50}{9}$, and the value of the objective function is 4.289. It is worth noticing that the same solution is obtained for some different combinations of equalities. In order to prove that the obtained solution is optimal it is necessary to solve optimization problem (20)-(21) with $\lambda = (\frac{8}{9}, \frac{1}{9})$ and to compare the values of the objective functions (H and H^*). In the considered case, problem (20)-(21) becomes

$$H^* = v + 2v_1 - 1w_1 + 0.3v_2 - 0w_2 \to \min_{v,v_k,w_k}$$

subject to $v \in \mathbb{R}, v_k, w_k \in \mathbb{R}_+$, and

$$\begin{aligned}
v + 1(v_1 - w_1) + 0(v_2 - w_2) &\ge 6 \cdot \frac{8}{9} + 2 \cdot \frac{1}{9},\\
v + 2(v_1 - w_1) + 0(v_2 - w_2) &\ge 3 \cdot \frac{8}{9} + 7 \cdot \frac{1}{9},\\
v + 3(v_1 - w_1) + 1(v_2 - w_2) &\ge 1 \cdot \frac{8}{9} + 4 \cdot \frac{1}{9}.
\end{aligned}$$

Hence $H^* = \frac{50}{9}$ and $H = H^*$. Therefore, $\lambda = (\frac{8}{9}, \frac{1}{9})$ is the optimal action.

4.4 On the comparison of both approaches

The superiority of one of the approaches for computing the optimal randomized action by using the caution parameter depends on the specific task considered. It is obvious that the complexity of searching the optimal action is mainly defined by the number of optimization problems to be solved and by the number of constraints to the optimization problems. If the number of extreme points is rather large, then the first approach may be preferable. For example, if the available information about states of nature $\{\vartheta_1, ..., \vartheta_m\}$ is restricted by a few points, say r points: $\vartheta_{l(1)}, ..., \vartheta_{l(r)}$, $r \ll m$, of an unknown cumulative probability distribution of states, then the number of extreme points can be shown to be up to $l(1) \times (l(2) - l(1)) \times ... \times (l(r) - l(r-1))$ and may be quite large. At the same time, by using the second approach, it is necessary to solve $\binom{2r+m}{2r+1}$ linear optimization problems (19) having $2m + 2r + 2r$ constraints and $2r + 1$ variables and $\binom{2r+m}{2r+1}$ linear optimization problems (20)-(21) having $m + 2r$ constraints and $2r + 1$ variables. This implies that the second approach is preferable.

Suppose now that $r \approx m$, i.e., we do not know only a few points of the probability distribution. The number of extreme points may be expected to be very small in this case, but the number of constraints in the second approach is rather large. Then the first approach should be used under this initial information.

5 E-admissibility and maximality

A fundamental alternative to the criteria considered up to now is to understand indecision as a natural consequence arising from the presence of ambiguity. Instead of ordering all the actions according to some real valued representation of the interval-valued expected utility, the aim is now to classify every action according to the dichotomy: 'acceptable' or 'not acceptable'.

We study the criterion of E-admissibility (cf. [15], [16]) in some detail and also briefly look at Walley's maximality criterion [23, Section 3.9]. An action a^* is said to be *E-admissible* with respect to a set of prior probabilities $\mathcal{M}$ iff there exists a classical prior $\pi(\cdot)$ such that a^* is Bayes with respect to $\pi(\cdot)$, i.e. iff for all other actions a under consideration

$$\sum_{j=1}^{m} u(a^*, \vartheta_j)\pi(\vartheta_j) \geq \sum_{j=1}^{m} u(a, \vartheta_j)\pi(\vartheta_j)\,. \qquad (22)$$

5.1 A lemma from classical Bayesian decision theory

Due to this direct connection to classical Bayesian decision theory based on precise probabilities it is beneficial to consider for a moment the classical basic decision problem $(\mathbb{A}, \Theta, u(\cdot))$ again with $|\Theta| < \infty$, but with a single classical prior $\pi(\cdot)$. In this context, the next lemma shows: If finding only one of the optimal actions is enough, then consideration may be confined to the set of unrandomized actions. Only if all optimal actions are needed, randomized actions have to be taken into account; they are obtained from convex combinations of optimal pure actions.

Lemma 1 *Let $\mathbb{A}^*_\pi$ be the set of all pure Bayes actions with respect to $\pi(\cdot)$, and Λ^*_π the set of all randomized Bayes actions with respect to $\pi(\cdot)$. Then*

i) $\mathbb{A}^*_\pi \neq \emptyset$

ii) $\Lambda^*_\pi = \mathrm{conv}(\mathbb{A}^*_\pi)$.[6]

Proof. The task of finding a Bayes action can be written as a linear programming problem

$$\sum_{j=1}^{m} \left(\sum_{i=1}^{n} u(a_i, \vartheta_j)\, \lambda(a_i) \right) \pi(\vartheta_j) \longrightarrow \max_{\lambda}$$

subject to $\sum_{i=1}^{n} \lambda(a_i) = 1$, and $\lambda(a_i) \geq 0$, for all i. Noting that the pure actions correspond to the vertices of the polyhedron defined by the constraints, both parts of the lemma follow immediately from general results on linear programming: The fact that one optimal solution must be attained at a vertex gives the non-emptiness of $\mathbb{A}^*_\pi$, while the second statement is deduced from the convexity of the set of optimal solutions. ■

5.2 An efficient algorithm for calculating E-admissible actions

Based on these considerations we can develop a general algorithm to calculate all E-admissible actions. To do so, we turn the problem around and fix the actions for a moment. More precisely, for every action a_i, $i = 1, \ldots n$, we look at the set Π_i of all priors $\pi(\cdot) \in \mathcal{M}$ under which a_i is Bayes action with respect to $\pi(\cdot)$. According to Lemma 1, for every $\pi(\cdot)$, the maximum of expected utility with respect to every $\pi(\cdot)$ is also attained at a pure action, and so it suffices to compare $\sum u(a_i, \vartheta_j)\pi(\vartheta_j)$ with $\sum u(a_l, \vartheta_j)\pi(\vartheta_j)$ for all pure actions a_l:

$$\Pi_i = \Big\{ \pi(\cdot) \in \mathcal{M} \;\Big|\; \sum_{j=1}^{m} u(a_i, \vartheta_j)\, \pi(\vartheta_j) \geq \sum_{j=1}^{m} u(a_l, \vartheta_j)\, \pi(\vartheta_j), \quad \forall\, l = 1, \ldots, n \Big\}$$

After having introduced an artificial auxiliary variable z the following linear programming problem is considered:[7]

$$\begin{aligned} z \quad &\longrightarrow \quad \max_{(\pi^T, z)^T} \\ \sum_{j=1}^{m} u(a_i, \vartheta_j)\, \pi(\vartheta_j) \quad &\geq \quad \sum_{j=1}^{m} u(a_l, \vartheta_j)\, \pi(\vartheta_j), \quad (23) \\ \forall l = 1, ..., n& \end{aligned}$$

$$\sum_{j=1}^{m} \pi(\vartheta_j) = z\,,\; z \leq 1\,, \quad \pi(\vartheta_j) \geq 0\,,\; j = 1, \ldots, m\,,$$

$$\underline{b}_l \leq \sum f_l(\vartheta_j)\pi(\vartheta_j) \leq \overline{b}_l\,,\; l = 1, \ldots, r\,.$$

Iff the optimal value of z equals 1, then Π is not empty and a_i is E-admissible.

[6] Here every pure action $a_i \in \mathbb{A}$ is identified with the randomized action $\lambda(a) = 1$ if $a = a_i$ and $\lambda(a) = 0$ else, and with the corresponding $(n \times 1)$ vector.

[7] Alternatively, if the set $\mathcal{E}(\mathcal{M})$ of vertices $\mathcal{M}$ is available, then $\Pi_i = \mathrm{conv}(\pi_{\mathcal{E}} \in \mathcal{E}(\mathcal{M}) | \sum_{j=1}^{m} u(a_i, \vartheta_j)\, \pi_{\mathcal{E}}(\vartheta_j) \geq \sum_{j=1}^{m} u(a_l, \vartheta_j)\, \pi(\vartheta_j), \forall l = 1, ..., n)$.

Therefore, for determining the set $\mathbb{IA}^*$ of all E-admissible pure actions $n = |\mathbb{IA}|$ simple linear optimization problems have to be solved. In the light of Lemma 1, the set $\mathbb{IA}^*$ is so-to-say essentially complete even in the set of all randomized actions: $\mathbb{IA}^*$ contains, for every $\pi(\cdot) \in \mathcal{M}$, an action that is optimal with respect to $\pi(\cdot)$; by randomization further E-admissible actions may be produced, but they can not have a higher expected utility with respect to $\pi(\cdot)$.

When the set $\mathbb{IA}^{**}$ of *all* the E-admissible actions in the set of randomized actions is of interest, Part ii) of Lemma 1 has to be finally applied to those $\pi(\cdot)$ where several pure actions are optimal. For achieving this the algorithm above can be used in an extended form. Instead of considering single actions a_i and the corresponding sets Π_i one has to check for arbitrary subsets $I \subseteq \{1, \ldots, m\}$ whether there is a prior π under which all a_i, $i \in I$ are simultaneously optimal, i.e. one replaces (23) by

$$\sum_{j=1}^{m} u(a_i, \vartheta_j)\, \pi(\vartheta_j) \quad \geq \quad \sum_{j=1}^{m} u(a_l, \vartheta_j)\, \pi(\vartheta_j)$$
$$\forall i \in I,\, l = 1, \ldots, n.$$

If the corresponding set $\Pi_I \neq \varnothing$ all the elements of $\mathrm{conv}(a_i | i \in I)$ are E-admissible actions. Of course, if $\Pi_I = \varnothing$ for some I then all index sets $J \supset I$ need not be considered anymore.

5.3 Maximality and structure dominance

The results just obtained are also of interest for other criteria. If Π contains an element $\pi(\cdot)$ with $\pi(\vartheta_j) > 0$, for all j, then the corresponding action a_i is admissible (in terms of classical decision theory) and therefore, as can easily be shown, maximal in the sense of Walley (cf. [23, Section 3.9]).

Furthermore, if $\Pi_i = \mathcal{M}$ for some i then a_i is uniformly optimal, a very strong criterion related to Weichselberger's concepts of structure dominance (cf. [25, Chapter 4.2]).

6 Concluding remarks

The paper presented powerful algorithms for handling sophisticated criteria for decision making under partial information. Hopefully, the availability of sound computational methods will stimulate application to larger problems, and may so contribute to gather experience on the performance of different criteria in substantive scientific fields.

Also for such applications it is of interest to mention explicitly that the results obtained here extend to the case where additionally data from a sample are available. One can either consider posteriori partial information or use decision functions. However, attention has to be paid to the fact that – in contrast to classical decision theory based on precise probabilities – both ways to proceed usually do not coincide under partial information (cf. [2]).

Acknowledgement We thank the referees for very helpful remarks.

References

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

[2] T. Augustin. On the suboptimality of robust Bayesian procedures from the decision theoretic point of view – a cautionary note on updating imprecise priors. In: J.M. Bernard, T. Seidenfeld, M. Zaffalon (eds.): ISIPTA 03: Proc. of the 3rd Int. Symposium on Imprecise Probabilities and their Applications, Lugano. Carleton Scientific, Waterloo, 31–45, 2003.

[3] T. Augustin. Optimal decisions under complex uncertainty - basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability. *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik*, 10–11:1–10, 2004.

[4] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis.* (2nd edition). Springer. New York, 1984.

[5] A. Chateauneuf, M. Cohen, I. Meilijson, New tools to better model behavior under risk and uncertainty: An overview. *Finance*, 18:25–46, 1997.

[6] G. Dantzig, *Linear Programming and Extensions.* Princeton University Press, Princeton, NJ, 1963.

[7] D. Ellsberg, Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.

[8] P. Ghirardato and M. Marinacci. Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research*, 26:864-890, 2001.

[9] P. H. Giang and P. P. Shenoy. Statistical decisions using likelihood information without prior probabilities. *Darwiche and N. Friedman, editors, Uncertainty in Artificial Intelligence, 2002*, pages 170-178, San Francisco, CA, 2002. Morgan Kaufmann.

[10] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

[11] J.Y. Jaffray, Linear utility theory and belief functions. *Operations Research Letters*, 8:107–122, 1989.

[12] E. Kofler, *Prognosen und Stabilität bei unvollständiger Information.* Campus, Frankfurt/Main, 1989.

[13] E. Kofler, G. Menges, *Entscheidungen bei unvollständiger Information.* Springer, Berlin (Lecture Notes in Economics and Mathematical Systems, 136), 1976.

[14] V.P. Kuznetsov. *Interval Statistical Models.* Radio and Communication, Moscow, 1991 (in Russian).

[15] I. Levi, On indeterminate probabilities, *Journal of Philosophy* 71:391–418, 1974.

[16] M. Schervish, T. Seidenfeld, J. Kadane, J., I. Levi, Extensions of expected utility theory and some limitations of pairwise comparisons. In: J.M. Bernard, T. Seidenfeld, M. Zaffalon (eds.): ISIPTA 03: Proc. of the 3rd Int. Symposium on Imprecise Probabilities and their Applications, Lugano. Carleton Scientific, Waterloo, pages 496–510, 2003.

[17] J. Schubert. On ρ in a decision-theoretic apparatus of Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 13:185–200, 1995.

[18] T. Seidenfeld, M.J. Schervish and J.B. Kadane, A representation of partially ordered preferences. *Annals of Statistics*, 23:2168-2217.

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

[20] M.C.M. Troffaes, Decisions making with imprecise probabilities: a short review. In: F.G. Cozman (ed.): The SIPTA Newsletter, Vol. 2 No. 1: Society for Imprecise Probability Theory and Applications, Manno, Switzerland. `http://www.sipta.org,` pages 4–7, December 2004.

[21] L.V. Utkin and I.O. Kozine. Different faces of the natural extension. In: G. de Cooman, T.L. Fine, and T. Seidenfeld (eds.): ISIPTA'01: Proc. of the 2nd Int. Symposium on Imprecise Probabilities and their Applications, Ithaca, USA. Shaker Publishing, pages 316–323, 2001.

[22] B. Vidakovic, Γ-minimax: A paradigm for conservative robust Bayesians. In: Insua, D.R., and Ruggeri, F. (eds.) *Robust Bayesian Analysis.* Springer, 2000.

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

[24] A. Wallner. Maximal number of vertices of polytopes defined by F-probabilit ies. Conditionally accepted for *ISIPTA05.*

[25] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, Volume I Intervallwahrscheinlichkeit als umfassendes Konzept.* Physika, Heidelberg, 2001.

[26] K. Weichselberger and T. Augustin. Analysing Ellsberg's Paradox by means of interval-probability. In R. Galata and H. Küchenhoff (eds.) *Econometrics in Theory and Practice. (Festschrift for Hans Schneeweiß)*, pages 291–304. Physika, Heidelberg, 1998.

[27] R. Y. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence.* Wiley, New York, 1994.