Errors Bounds for Finite Approximations of Coherent Lower Previsions

Damjan Škulj DAMJAN.SKULJ@FDV.UNI-LJ.SI
University of Ljubljana, Faculty of Social Sciences
Ljubljana (Slovenia)

Abstract

Coherent lower previsions are general probabilistic models allowing incompletely specified probability distributions. However, for complete description of a coherent lower prevision – even on finite underlying sample spaces – an infinite number of assessments is needed in general. Therefore, they are often only described approximately by some less general models, such as coherent lower probabilities or in terms of some other finite set of constraints. The magnitude of error induced by the approximations has often been neglected in the literature, despite the fact that it can be significant with substantial impact on consequent decisions. An apparent reason is that no widely used general method for estimating the error seems to be available at the moment. The goal of this paper is to provide such a method. The proposed method allows calculating an upper bound for the error of a finite approximation of coherent lower prevision on a finite underlying sample space. An estimate of the maximal error is especially useful in the cases where calculating assessments is computationally demanding. Our method is based on convex analysis applied to credal sets, which in the case of finite sample spaces correspond to convex polyhedra.

Keywords: lower prevision; partially specified lower prevision; credal set; convex polyhedron; quadratic programming.

1. Introduction

One of the most popular and also most general models of imprecise probabilities are *coherent lower previsions* (see, e.g., Miranda, 2008; Troffaes and De Cooman, 2014). A coherent lower prevision $\underline{P}$ is an imprecise probability model based on judgements about the lower or upper expectations on a set of bounded maps $\mathcal{K}$ from a sample space $\mathcal{X}$ to real numbers, also called *gambles*. The set of all gambles on a given underlying sample space will be denoted by $\mathcal{L}$. In this paper, all sample spaces are finite, therefore, we do not address any measurability or countable additivity conditions. The *judgement* or *assessment* $\underline{P}(f) = a$ states that every precise probability distribution P compatible with $\underline{P}$ must satisfy $E_P(f) \geq a$, that is $\underline{P}(f)$ means that the expectation of f is at least a. *Coherence* in this context means that the judgements on the set of gambles allow, for every gamble f, the existence of at least one precise probability distribution P compatible with $\underline{P}$ for which $E_P(f) = \underline{P}(f)$. The expectation functionals with respect to precise (finitely additive) probability distributions are often called *linear previsions*.

A coherent lower prevision $\underline{P}$ specified on a set of gambles $\mathcal{K}$ can have multiple possible extensions to a larger set, say $\mathcal{H} \supset \mathcal{K}$. In other words, there can be multiple coherent lower previsions that coincide on a set of gambles. In particular, a coherent lower prevision may be approximated by a more specific model, such as *coherent lower probability* (see, e.g., Antonucci and Cuzzolin, 2010), in which case its restriction to *indicator gambles* is only known, i.e. an *indicator gamble* 1_A

is a map $\mathcal{X} \to \mathbb{R}$ such that $1_A(x)$ equals 1 if $x \in A$ and 0 otherwise. We will write 1_x instead of $1_{\{x\}}$ for elements $x \in \mathcal{X}$.

In this paper we investigate the following problem. Let $\underline{P}$ be a coherent lower prevision on the set $\mathcal{L}$ of all gambles on a finite sample space $\mathcal{X}$. Its full description would in general require detailed information on the set of compatible precise models, called *credal set*, which often is unavailable. Suppose that instead we know the values of $\underline{P}$ on a set of gambles $\mathcal{K}$. The restriction $\underline{P}_{\mathcal{K}}$ approximates $\underline{P}$ and the natural question arises, how accurate is this approximation. Given the restriction, $\underline{P}$ is an extension of $\underline{P}_{\mathcal{K}}$, which in general is not unique. Therefore, we would like to know by how much can another extension deviate from $\underline{P}$. That is, we want to find the maximal distance between two arbitrary extensions of a coherent lower prevision on a finite set $\mathcal{K}$ to the set of all gambles.

In our analysis we first show that the maximal possible distance is always reached when one of the extensions is the *natural extension*. Consequently, much of the analysis is done on the credal set of the natural extension with the special emphasis on its extreme points. Our main result gives an upper bound for the maximal distance in terms of distances between the extreme points.

The paper is structured as follows. In Section 2 we review basic concepts of imprecise probabilities with the emphasis on coherent lower previsions. In Section 3 we analyze basic properties of credal sets as convex polyhedra and apply some general concepts of convex analysis to the case of credal sets. Our main results are stated in Section 4.

2. Notation and basic results

In this section we introduce the notation and review the concepts used in the paper. When possible we will stick with the standard terminology used in the theory of imprecise probabilities, which will sometimes be supplemented by the standard terminology of convex analysis, linear algebra and optimization.

GAMBLES.

Throughout this paper let $\mathcal{X}$ represent a finite set, a *sample space*, and $\mathcal{L}$ the set of all real-valued maps on $\mathcal{X}$, also called *gambles*. Equivalently, $\mathcal{L}$ may be viewed as the set of vectors in $\mathbb{R}^{|\mathcal{X}|}$. The set of gambles will be endowed by the standard inner product $f \cdot g = \sum_{x\in\mathcal{X}} f(x)g(x)$, which generates the l^2 norm: $\|f\| = \sqrt{f \cdot f} = \sqrt{\sum_{x\in\mathcal{X}} f(x)^2}$, and the Euclidean distance between vectors: $d(f, g) = \|f - g\|$, which will be used by default throughout the paper.

LINEAR PREVISIONS.

A *linear prevision* P is an expectation functional with respect to some probability mass vector p on $\mathcal{X}$. It maps a gamble f into a real number $P(f)$. Usually, we will write $P(f) = \sum_{x\in\mathcal{X}} p(x)f(x) =: P \cdot f$. The set of linear previsions is therefore a subset of the dual space of $\mathcal{L}$. The inner product notation is introduced because we will often use linear functionals of the form $f \mapsto p \cdot f$ where the vector p will not necessarily be a probability mass vector. We will then use the inner product notation to avoid misinterpretations. Without danger of confusion we will therefore interpret a linear prevision P as a vector with the same length as gambles in $\mathcal{L}$.

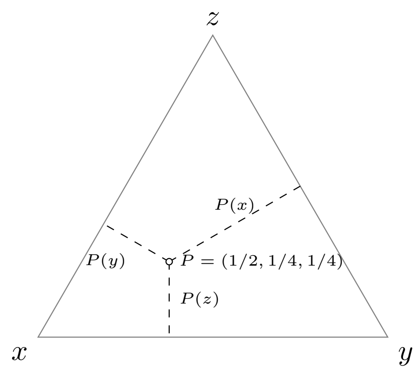

Figure 1: Probability simplex: the distance from a side denotes the probability of the element at the opposite vertex.

PROBABILITY SIMPLEX.

If the sample space $\mathcal{X}$ contains exactly three elements, say $\mathcal{X} = \{x, y, z\}$, the probability mass vectors can be represented as points of the form $(p(x), p(y), p(z))$ in $\mathbb{R}^3$. However, since the restriction $p(x)+p(y)+p(z) = 1$ applies, they in fact form a two dimensional space, which can be depicted as an equilateral triangle with vertices x, y and z. Given any point in this triangle, the sum of distances to its sides is constantly equal to its altitude, which equals $\frac{\sqrt{3}}{2}a$, where a is the common length of the sides. Taking $a = \frac{2}{\sqrt{3}}$ makes the altitude equal to 1. The distance of a point from each side now denotes the probability of the point in the opposite vertex. (See Figure 1.) Probability simplex diagrams are very useful to illustrate concepts of imprecise probabilities; however, one needs to be cautious not to be mislead by specifics of low dimensional probability spaces.

COHERENT LOWER PREVISIONS.

A *coherent lower prevision* on an arbitrary set of gambles $\mathcal{K}$ is a mapping $\underline{P}\colon \mathcal{K} \to \mathbb{R}$ that allows the representation

$$\underline{P}(f) = \min_{\mathcal{M}} P(f) \tag{1}$$

for every $f \in \mathcal{K}$, where $\mathcal{M}$ is a closed and convex set of linear previsions. Note that unless $\mathcal{K}$ is the set of all gambles, there may be multiple sets $\mathcal{M}$ that fit into equation (1) (this is also one of the motivations for this paper); however, there is a unique maximal such set. We will denote the maximal such set with $\mathcal{M}(\underline{P})$ and call it the *credal set* of $\underline{P}$.

THE NATURAL EXTENSION.

Given a coherent lower prevision $\underline{P}$ on $\mathcal{K}$, it is possible to extend it to the set of all gambles $\mathcal{L}$ in possibly several different ways, and again, there is unique minimal extension, called the *natural extension*:

$$\underline{E}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f). \tag{2}$$

Note that replacing $\mathcal{M}(\underline{P})$ with another set $\mathcal{M}$ of linear previsions satisfying the equation (1) would result in some other extension of $\underline{P}$.

A mapping $\underline{P}\colon \mathcal{K} \to \mathbb{R}$, where $\mathcal{K}$ is a linear (vector) space, is a coherent lower prevision if and only if it satisfies the following axioms (Miranda, 2008) for all $f, g \in \mathcal{K}$ and $\lambda \geq 0$:

(P1) $\underline{P}(f) \geq \inf_{x\in\mathcal{X}} f(x)$ [accepting sure gains];

(P2) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [non-negative homogeneity];

(P3) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ [superlinearity].

An easy consequence of the definitions is :

(P4) $\underline{P}(f + \lambda 1_{\mathcal{X}}) = \underline{P}(f) + \lambda$ for any $\lambda \in \mathbb{R}$ and $f \in \mathcal{L}$ [constant additivity].

3. Credal set as a convex polyhedron

A credal set is a closed and convex set of linear previsions. Since every linear prevision can be uniquely represented as a probability mass vector, a credal set can be represented as a convex set of probability mass vectors. The set $\mathcal{M}$ is therefore the maximal set of $|\mathcal{X}|$-dimensional vectors p satisfying:

$$p \cdot f \geq \underline{P}(f) \qquad \text{for every } f \in \mathcal{K}, \tag{3}$$

$$p \cdot 1_x \geq 0 \qquad \text{for every } x \in \mathcal{X} \text{ and} \tag{4}$$

$$p \cdot 1_{\mathcal{X}} = 1. \tag{5}$$

In the sequel we will assume that the set $\mathcal{K}$ is finite. When needed, we will index its elements as f_i for $i \in \{1, \dots, n\}$.

According to the above, it would be suitable to extend the domain of $\underline{P}$ with the gambles of the form 1_x for every $x \in \mathcal{X}$. Doing so, though, may result in a non-coherent lower prevision, because other constraints my already imply that $\underline{P}(1_x) \geq 0$, where the inequality may even be strict. Therefore we adopt the following convention:

Convention 1 *The domain $\mathcal{K}$ of all lower previsions used will contain all gambles of the form 1_x together with the value $\underline{P}(1_x) = 0$, unless $\underline{P}(1_x) \geq 0$ is already implied by other values of $\underline{P}$ on $\mathcal{K}$.*

Assuming the above convention, the credal set of coherent lower prevision $\underline{P}$ is the set of vectors p satisfying constraints (3) and (5).

In the case where $\mathcal{K}$ is finite, the corresponding credal set is a *convex polyhedron*. Strictly speaking, it is an $\mathcal{H}$-polyhedron, which means that it is bounded and an intersection of a finite number of half spaces. According to Theorem 14.3 in Gruber (2007) every $\mathcal{H}$-polyhedron in an $\mathbb{R}^m$ is also a $\mathcal{V}$-polyhedron, which means that it is a convex combination of a finite number of extreme points.

Example 1 *Let $\underline{P}$ be a lower prevision on $\mathcal{K} = \{f_1, \dots, f_5\}$ where*

$$f_1 = (0, 1, 0.5) \qquad f_2 = (0, 0.5, 1) \qquad f_3 = (0.15, 0, 1)$$
$$f_4 = (1, 0, 0.6) \qquad f_5 = (0.2, 1, 0)$$

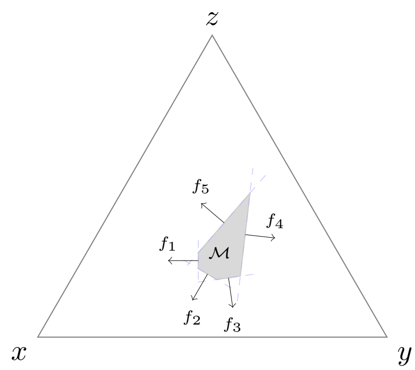

Figure 2: Credal set from Example 1 as an intersection of half planes: their support lines are dashed, gambles $f_i \in \mathcal{K}^+$ are depicted as normal vectors to faces.

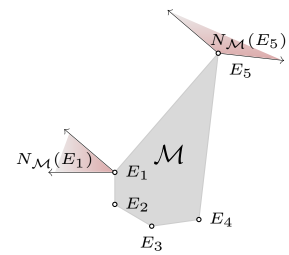

Figure 3: Normal cones at extreme points are the non-negative hulls of the normal vectors of adjacent faces.

and

$$\underline{P}(f_1) = 0.46 \qquad \underline{P}(f_2) = 0.4 \qquad \underline{P}(f_3) = 0.25$$
$$\underline{P}(f_4) = 0.44 \qquad \underline{P}(f_5) = 0.4$$

The credal set corresponding to $\underline{P}$ is depicted in Figure 2 as an intersection of half-planes.

FACES AND EXTREME POINTS OF A FINITELY GENERATED CREDAL SET.

The *faces* of a credal set $\mathcal{M}$ are the sets of the form $\mathcal{M}_f = \{P \in \mathcal{M} \colon P(f) = \underline{E}(f)\}$, where f is an arbitrary gamble. The smallest faces are exactly the extreme points and the faces of codimension 1 are called *facets*[1]. The set of all extreme points of $\mathcal{M}$ will be denoted by $\mathcal{E}(\mathcal{M})$ or simply $\mathcal{E}$. The set of extreme points of a face $\mathcal{M}_f$ will be denoted by $\mathcal{E}_f$, and $\mathcal{E}_f \subseteq \mathcal{E}$ holds.

Example 2 *The extreme points of the credal set from Example 1 are*

$$E_1 = (0.4, 0.32, 0.28) \qquad E_2 = (0.43, 0.35, 0.23) \qquad E_3 = (0.39, 0.42, 0.19)$$
$$E_4 = (0.32, 0.48, 0.20) \qquad E_5 = (0.15, 0.37, 0.48)$$

(See Figure 3.)

Let $f \in \mathcal{K}$ be a gamble and $\underline{P}(f)$ its lower prevision. Then the lower prevision of the gamble $f - \underline{P}(f)1_{\mathcal{X}}$ equals 0. Moreover, setting $\underline{P}(f - \underline{P}(f)1_{\mathcal{X}}) = 0$ is equivalent to setting the lower prevision of f to $\underline{P}(f)$, by constant additivity. Following this idea, we extend a credal set $\mathcal{M}$ to the set of vectors

$$\hat{\mathcal{M}} = \{p \colon p \cdot (f - \underline{P}(f)1_{\mathcal{X}}) \geq 0, \text{ for every } f \in \mathcal{K}\}, \tag{6}$$

1. The codimension 1 is meant relative to the dimension of $\mathcal{M}$. That is $\dim \mathcal{M}_f = \dim \mathcal{M} - 1$. Note also that a credal set is at most of dimension $|\mathcal{X}| - 1$ because of the constraint $P(1_{\mathcal{X}}) = 1$.

which is a convex cone, with the *basis* $\mathcal{M}$. This means that every $p \in \hat{\mathcal{M}}$ is of the form $p = \lambda P$ for some $\lambda \geq 0$ and $P \in \mathcal{M}$. This is easily seen by noticing that every $p \in \hat{\mathcal{M}}$ has non-negative components, which is guaranteed by Convention 1. Dividing $\mathbf{0} \neq p \in \hat{\mathcal{M}}$ by the sum of its components then results in a vector P whose components are non-negative, sum to one and clearly satisfy the same linear constraints as p, except (5).

Given a credal set $\mathcal{M}$, the *cone of (almost) desirable gambles* contains exactly those gambles in $\mathcal{L}$ whose lower prevision is non-negative:

$$\mathcal{D} = \{f \in \mathcal{L} \colon P(f) \geq 0 \text{ for every } P \in \mathcal{M}\} = \{f \in \mathcal{L} \colon \underline{P}(f) \geq 0\}. \tag{7}$$

The gambles f with $\underline{P}(f) = 0$ are sometimes called *marginally desirable*.

3.1 Normal cones of credal sets

THE NORMAL CONE.

Let

$$\mathcal{C} = \{x \in \mathbb{R}^n \colon Ax \leq b\}, \tag{8}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$ a vector, be a convex polyhedron and x a point on its boundary. According to Gruber (2007), the *normal cone* at x is the set

$$N_{\mathcal{C}}(x) = \{u \colon u \cdot y \leq u \cdot x \text{ for all } y \in \mathcal{C}\} = \{u \colon u \cdot (y - x) \leq 0 \text{ for all } y \in \mathcal{C}\}. \tag{9}$$

In our case, let $\mathcal{M}$ be a credal set defined by constraints of the form (3) and (5) and E its boundary point. The normal cone of $\mathcal{M}$ at E is the set

$$N_{\mathcal{M}}(E) = \{f \colon E(f) \leq P(f) \text{ for every } P \in \mathcal{M}\}. \tag{10}$$

The normal cone is thus the set of gambles f that satisfy $E(f) = \underline{P}(f)$.

Proposition 2 (Gruber (2007) Proposition 14.1.) *Let $\mathcal{C}$ be a convex polyhedron defined as in (8) and x its boundary point. Let $a_i \cdot x = b_i$ hold for exactly $i \in I \subseteq \{1, \ldots, m\}$, where a_i denotes i-th row of the matrix A. Then $N_{\mathcal{C}}(x) = \operatorname{pos}\{a_i \colon i \in I\}$, where* pos *denotes the* non-negative hull.

Corollary 3 *Let $\mathcal{M}$ be a credal set defined by constraints (3) and (5). Then the set of (almost) desirable gambles $\mathcal{D}$ corresponding to $\mathcal{M}$ is the normal cone of $\hat{\mathcal{M}}$ at $\mathbf{0}$ and we have that $\mathcal{D} = \operatorname{pos}\{f - \underline{P}(f)1_{\mathcal{X}} \colon f \in \mathcal{K}\}$.*

Proof The set $\hat{\mathcal{M}}$ is a convex cone whose support hyperplanes are exactly the sets of the form $H_f = \{p \colon p \cdot (f - \underline{P}(f)1_{\mathcal{X}}) = 0\}$ for $f \in \mathcal{K}$, and the origin is exactly the intersection of all support hyperplanes: $\mathbf{0} \cdot (f - \underline{P}(f)1_{\mathcal{X}}) = 0$ for every $f \in \mathcal{K}$. We can therefore apply Proposition 2. ■

Remark 4 *In Augustin et al. (2014) Chapter 1, the set constructed as $\mathcal{D}$ in Corollary 3 is called the natural extension of the assessment $\mathcal{K}$. The fact that the set of desirable gambles is the non-negative hull of marginally desirable assessments in $\mathcal{K}$ with included strictly positive gambles can also be found in Chapter 2 of the mentioned book. In our case, strictly positive gambles are included because of Convention 1.*

Corollary 5 *Let $\mathcal{M}$ be a credal set defined by constraints of the form (3) and (5), $E \in \mathcal{M}$ a linear prevision and h a gamble such that $E(h) = \underline{P}(h)$. For every gamble $f \in \mathcal{K}$ let $\tilde{f} = f - \underline{P}(f)$, and thus $\underline{P}(\tilde{f}) = 0$ for all $f \in \mathcal{K}$.*

Suppose that $E(\tilde{f}_i) = 0$ for exactly $i \in I \subseteq \{1, \dots, n\}$. Then there exist $\alpha_i \geq 0$ for every $i \in I$ and $\beta \in \mathbb{R}$ so that

$$h = \sum_{i \in I} \alpha_i \tilde{f}_i + \beta 1_{\mathcal{X}}. \tag{11}$$

Proof Let $h \in \mathcal{L}$ be a gamble such that $E(h) = \underline{P}(h)$. Set $g = h - \underline{P}(h)$. Then, for every $p \in \hat{\mathcal{M}}$ (see (6)), $p = \alpha P$ for some $P \in \mathcal{M}$ and $\alpha \geq 0$. Therefore $p \cdot g = \alpha P \cdot g \geq 0 = E \cdot g$, whence $g \in N_{\hat{\mathcal{M}}}(E)$. By Proposition 2, $g = \sum_{i \in I} \alpha_i \tilde{f}_i$ for some non-negative constants α_i. Hence $h = \sum_{i \in I} \alpha_i \tilde{f}_i + \underline{P}(h) 1_{\mathcal{X}}$, which proves the proposition. ■

Note that Equation (11) still holds if $\tilde{f}_i$ are replaced by f_i.

4. The distance between coherent lower previsions

4.1 The definition of the distance

Let $\underline{P}$ and $\underline{P}'$ be two coherent lower previsions on the set of all gambles $\mathcal{L}$ on a finite set $\mathcal{X}$. We define the distance[2] between $\underline{P}$ and $\underline{P}'$ as

$$d(\underline{P}, \underline{P}') = \max_{f \in \mathcal{L}} \frac{|\underline{P}(f) - \underline{P}'(f)|}{\|f\|}, \tag{12}$$

where the norm $\|f\| = \sqrt{f \cdot f}$ is the Euclidean norm in $\mathbb{R}^{|\mathcal{X}|}$. Clearly, the following alternative definition is equivalent: $d(\underline{P}, \underline{P}') = \max_{\substack{f \in \mathcal{L} \\ \|f\|=1}} |\underline{P}(f) - \underline{P}'(f)|$.

It is readily verified that the above distance function induces a metric in the set of all lower previsions on $\mathcal{L}$. In this section we will analyze the maximal possible distance between two coherent lower previsions that coincide on a finite set of gambles.

Suppose that $\underline{P}$ is a lower prevision on $\mathcal{L}$, and the only information about it are the values on a finite set of gambles $\mathcal{K} \subset \mathcal{L}$. That is, $\underline{P}(f)$ are given for every $f \in \mathcal{K}$. We denote the restriction of $\underline{P}$ to $\mathcal{K}$ by $\underline{P}_{\mathcal{K}}$. We also adopt Convention 1. The natural extension $\underline{E}$ is the minimal (or the least committal) extension of $\underline{P}_{\mathcal{K}}$. This implies that $\underline{P}(f) \geq \underline{E}(f)$ for every $f \in \mathcal{L}$. Therefore, given another extension $\underline{P}'$ of $\underline{P}_{\mathcal{K}}$, we have that

$$|\underline{P}(f) - \underline{P}'(f)| \leq \max\{\underline{P}(f) - \underline{E}(f), \underline{P}'(f) - \underline{E}(f)\}, \tag{13}$$

which implies that $d(\underline{P}, \underline{P}') \leq \max\{d(\underline{P}, \underline{E}), d(\underline{P}', \underline{E})\}$. As we are interested in the maximal possible distance between coherent lower previsions coinciding on $\mathcal{K}$, it will therefore be enough to focus to the case where one of them is the natural extension of $\underline{P}_{\mathcal{K}}$.

4.2 Maximal distance to the natural extension

Let $\underline{E}$ and $\underline{P}$ be respectively the natural extension of $\underline{P}_{\mathcal{K}}$ and another extension, and $\mathcal{M}$ and $\mathcal{C}$ respectively their credal sets. As described in previous sections, both are convex sets and the natural extension is a convex polyhedron with extreme points $\mathcal{E}(\mathcal{M})$.

2. For another distance function between coherent lower previsions, see, e.g., Škulj and Hable (2013).

Assuming the above notations, we start with the following proposition.

Proposition 6 *Take some $f \in \mathcal{K}$ and let $\mathcal{M}_f$ be the corresponding face of $\mathcal{M}$. Then $\mathcal{C} \cap \mathcal{M}_f \neq \emptyset$.*

Proof Clearly, $\mathcal{M}_f$ contains exactly all linear previsions P in $\mathcal{M}$ such that $P(f) = \underline{P}(f)$. If no $P \in \mathcal{C}$ belongs to $\mathcal{M}_f$, this then implies that $P(f) > \underline{P}(f)$ for every $P \in \mathcal{C}$, and since $\mathcal{C}$ is compact, this would imply that $\min_{P\in\mathcal{C}} P(f) > \underline{P}(f)$, which contradicts the assumptions. ■

Corollary 7 *Let $h \in \mathcal{L}$ be an arbitrary gamble. Then:*

(i) $\underline{P}(h) \leq \max_{P\in\mathcal{M}_f} P(h)$ *for every* $f \in \mathcal{K}$;

(ii) $\underline{P}(h) \leq \min_{f\in\mathcal{K}} \max_{P\in\mathcal{M}_f} P(h)$*; the inequality is tight in the sense that for every* $h \in \mathcal{L}$ *an extension of* $\underline{P}_{\mathcal{K}}$ *exists that gives equality in the equation.*

(iii) $\underline{P}(h) \leq \min_{f\in\mathcal{K}} \max_{E\in\mathcal{E}_f} E(h)$ *where* $\mathcal{E}_f$ *is the set of extreme points of the face* $\mathcal{M}_f$*; and the inequality is again tight.*

Proof (i) is an immediate consequence of Proposition 6.

The inequality in (ii) is a direct consequence of (i). It remains to prove that there is an extension of $\underline{P}_{\mathcal{K}}$ where the equality is reached.

Let $\mathcal{M}_f$ be a face of $\mathcal{M}$ and let $P_f \in \arg\max_{P\in\mathcal{M}_f} P(h)$. Let $\mathcal{M}'$ be the convex hull of $\{P_f : f \in \mathcal{K}\}$ and $\underline{P}'$ the corresponding coherent lower prevision, which coincides with $\underline{P}$ on $\mathcal{K}$ by construction, and thus must satisfy the inequality (ii). For every $P \in \mathcal{M}'$, on the other hand, we have that $P = \sum_{i\in\mathcal{K}} \alpha_f P_f$, for some collection of values $\alpha_f \geq 0$ for every $f \in \mathcal{K}$ and $\sum_{f\in\mathcal{K}} \alpha_f = 1$. Thus,

$$P(h) = \sum_{f\in\mathcal{K}} \alpha_f P_f(h) \geq \min_{f\in\mathcal{K}} P_f(h) = \min_{f\in\mathcal{K}} \max_{P\in\mathcal{M}_f} P(h) \tag{14}$$

Hence, $\underline{P}'(h) = \min_{P\in\mathcal{M}'} P(h) \geq \min_{f\in\mathcal{K}} \max_{P\in\mathcal{M}_f} P(h)$, which combined with the above reverse inequality gives the required equality.

The fact that extremal values are reached in extreme points easily implies (iii). ■

Now we can express the maximal possible distance between two arbitrary extensions of $\underline{P}_{\mathcal{K}}$ in terms of its natural extension alone.

Corollary 8 *Let $\underline{E}$ be the natural extension and $\underline{P}$ and $\underline{P}'$ two other extensions of $\underline{P}_{\mathcal{K}}$, and $h \in \mathcal{L}$ a gamble. Then $|\underline{P}(h) - \underline{P}'(h)| \leq \min_{f\in\mathcal{K}} \max_{P\in\mathcal{E}_f} P(h) - \underline{E}(h)$ and*

$$d(\underline{P}, \underline{P}') \leq \max_{\|h\|=1} \min_{f\in\mathcal{K}} \max_{P\in\mathcal{E}_f} P(h) - \underline{E}(h). \tag{15}$$

Proof The first inequality is a direct consequence of Corollary 7(iii) and Eq. (13). The second inequality is an immediate consequence of the first one, definition of the distance between two coherent lower previsions and the fact that $\underline{E}(h)$ is less than $P(h)$ for every feasible P. ■

Equation (15) gives the maximal possible distance between two unknown extensions of $\underline{P}_{\mathcal{K}}$ entirely in terms of its natural extension. However, as an optimization problem it is not solvable in any apparently applicable way. We will therefore apply it to derive a practically computable upper bounds.

By the definition of $\underline{E}$ we have:

$$d(\underline{P}, \underline{P}') \leq \max_{\|h\|=1} \max_{E \in \mathcal{E}} \min_{f \in \mathcal{K}} \max_{P \in \mathcal{E}_f} P(h) - E(h) \tag{16}$$

$$= \max_{E \in \mathcal{E}} \max_{\|h\|=1} \min_{f \in \mathcal{K}} \max_{P \in \mathcal{E}_f} P(h) - E(h) \tag{17}$$

by interchanging $\max_{\|h\|=1}$ and $\min_{f \in \mathcal{K}}$:

$$\leq \max_{E \in \mathcal{E}} \min_{f \in \mathcal{K}} \max_{P \in \mathcal{E}_f} \max_{\|h\|=1} P(h) - E(h) \tag{18}$$

$$= \max_{E \in \mathcal{E}} \min_{f \in \mathcal{K}} \max_{P \in \mathcal{E}_f} d(P, E), \tag{19}$$

where $d(P, E)$ is the Euclidean distance between extreme points P and E.

Now denote

$$\bar{d}(E, f) = \max_{P \in \mathcal{E}_f} d(P, E), \tag{20}$$

which is the maximal Euclidean distance between an extreme point E and a face $\mathcal{M}_f$. Thus we obtain the following formula:

$$d(\underline{P}, \underline{P}') \leq \max_{E \in \mathcal{E}} \min_{f \in \mathcal{K}} \bar{d}(E, f). \tag{21}$$

Since E and P in the above expressions are (extreme) points in $\mathbb{R}^{|\mathcal{X}|}$, their Euclidean distances can be found easily by calculating the Euclidean norms $\|P - E\|$. Particularly, calculating $\bar{d}(E, f)$ requires calculating the Euclidean distances between E and all extreme points of the face $\mathcal{M}_f$. Finally, the RHS expression in (21) is calculated by finding $\bar{d}(E, f)$ for all pairs of extreme points and gambles in $\mathcal{K}$.

4.3 Improved bounds

Equation (21) gives an upper bound for the difference between coherent lower previsions coinciding on a set of gambles, however, the estimate is systematically too conservative. This is caused by the fact that extreme points E can only maximize expression (16) for some h if $E(h) = \underline{E}(h)$. This means that the domain for h in (18) should be restricted to those gambles h that reach the lowest value $\underline{E}(h)$ in E. In other words, h should belong to the normal cone $N_{\mathcal{M}}(E)$.

Therefore, instead of taking the Euclidean distance between E and P in (20), we should take the following distance:

$$d_E(E, P) = \max_{h \in N_{\mathcal{M}}(E)} \frac{|P(h) - E(h)|}{\|h\|}, \tag{22}$$

which we call the *normed distance* between E and P.

The geometrical intuition behind replacing Euclidean distance with the above distance function is the following. Given a gamble h, the difference $P(h) - E(h)$ can be viewed as the inner product $(P - E) \cdot h$, which depends on the angle between $(P - E)$ and h. As the normal cone contains elements that are orthogonal to $P - E$ for adjacent extreme points P, we may expect that the other elements are nearly orthogonal too, especially in the case of narrow normal cones. In Figure 3 such situation can be observed in the case of the normal cone of E_1, in contrast to the case of E_5, where the normal cone is wide. Therefore, we would, for instance, expect that the normed distances between E_1 and its adjacent extreme points would be significantly smaller than the Euclidean distance, in contrast the case of E_5. Analytically we demonstrate this in Example 3.

In the sequel we represent the calculation of the normed distance in the form of a quadratic programming problem.

MINIMUM NORM ELEMENTS OF THE NORMAL CONE.

Consider an element h of the form (11). Given a pair of expectation functionals E and P, the distance $P(h) - E(h)$ does not depend on β. In order to maximize the normed distance (22), we must consider the representative with the minimum norm, as the norm appears in the denominator of the expression. The characterization of the minimal norm element of the form (11) follows.

Proposition 9 *Let h be a gamble. Then $\|h+\beta 1_{\mathcal{X}}\| \geq \|h\|$ for every $\beta \in \mathbb{R}$ if and only if $h \cdot 1_{\mathcal{X}} = 0$.*

Proof We have that $\|h + \beta 1_{\mathcal{X}}\|^2 = \|h\| + \beta^2 + 2\beta h \cdot 1_{\mathcal{X}}$, which has minimum in $\beta = -h \cdot 1_{\mathcal{X}}$. Hence the minimizing β equals 0 exactly if $h \cdot 1_{\mathcal{X}}$ does. ■

Corollary 10 *Let E, h and I be as in Corollary 5 and let f_i' be the unique vectors such that $f_i - f_i' = c1_{\mathcal{X}}$ and $f_i' \cdot 1_{\mathcal{X}} = 0$ for every $i \in I$. Then, as follows from Corollary 5, there exist some $\alpha_i' \geq 0$ for every $i \in I$ and $\beta' \in \mathbb{R}$ so that*

$$h = \sum_{i \in I} \alpha_i' f_i' + \beta' 1_{\mathcal{X}}. \tag{23}$$

Moreover,

$$\left\| \sum_{i \in I} \alpha_i' f_i' \right\| \leq \left\| \sum_{i \in I} \alpha_i' f_i' + \beta 1_{\mathcal{X}} \right\| \text{ for every } \beta \in \mathbb{R}. \tag{24}$$

Proof Since $f_i' \cdot 1_{\mathcal{X}} = 0$, we have that $\left(\sum_{i \in I} \alpha_i' f_i'\right) \cdot 1_{\mathcal{X}} = 0$, whence by Proposition 9 it follows that this is the minimal-norm gamble of the form (23). ■

Let I and f_i', for $i \in I$, be as in Corollary 10 and let $\underline{\alpha}\colon I \to [0, \infty)$ be a map and $\beta \in \mathbb{R}$ a constant (we will write α_i instead of $\alpha(i)$). Then we define $h(\underline{\alpha}, \beta) = \sum_{i \in I} \alpha_i f_i' + \beta 1_{\mathcal{X}}$. Clearly, $h(\underline{\alpha}, \beta) \in N_{\mathcal{M}}(E)$ and every element of $N_{\mathcal{M}}(E)$ is of the form $h(\underline{\alpha}, \beta)$, by Corollary 5.

Corollary 11 *The following equality holds:*

$$\max_{(\underline{\alpha}, \beta)} \frac{|E(h(\underline{\alpha}, \beta)) - P(h(\underline{\alpha}, \beta))|}{\|h(\underline{\alpha}, \beta)\|} = \max_{\underline{\alpha}} \frac{|E(h(\underline{\alpha}, 0)) - P(h(\underline{\alpha}, 0))|}{\|h(\underline{\alpha}, 0)\|} \tag{25}$$

Proof Since $|E(h + \beta 1_{\mathcal{X}}) - P(h + \beta 1_{\mathcal{X}})| = |E(h) - P(h)|$, the maximum of the expression is achieved at h with the minimum norm, which is the one with $\beta = 0$. ■

THE CALCULATION OF THE NORMED DISTANCE BETWEEN EXPECTATION FUNCTIONALS.

Take two linear expectation functionals P and $E \in \mathcal{M}$ and let I and f'_i for $i \in I$ be as in Corollary 10. Our goal is to find the normed distance (22). The absolute value in the numerator of (22) can be omitted because $E(h) = \min_{P\in\mathcal{M}} P(h)$ for every $h \in N_{\mathcal{M}}(E)$. By Corollary 11, every $h \in N_{\mathcal{M}}(E)$ that can minimize the above expression is of the form $h(\underline{\alpha}, 0)$. Since E and P are themselves vectors too, we can denote $D = P - E$, and write $P(h) - E(h) = (P - E) \cdot h = D \cdot h$.

Now we can decompose every f'_i for $i \in I$ as $f'_i = \lambda_i D + u_i$, so that $D \cdot u_i = 0$. Given that $h = \sum_{i\in I} \alpha_i f'_i$, we obtain $h = (\underline{\alpha} \cdot \underline{\lambda})D + \underline{\alpha} \cdot U$, where U is the matrix whose rows are u_i, $\underline{\lambda}$ is the column vector with components λ_i and the vectors f'_i are also written as row vectors. We also assume $\underline{\alpha}$ to be a column vector.

Further we have that $\|h\|^2 = h \cdot h = \|D\|^2 \underline{\alpha}\,\underline{\lambda}\,\underline{\lambda}^t \underline{\alpha}^t + \underline{\alpha} U U^t \underline{\alpha}^t$. Now denote $\Pi = \|D\|^2 \underline{\lambda}\,\underline{\lambda}^t + UU^t$ and write $\|h\|^2 = \underline{\alpha}\Pi\underline{\alpha}^t$. Clearly, Π is a symmetric and positive semi-definite matrix.

Moreover, we have that $P(h) - E(h) = D \cdot (\underline{\alpha} \cdot \underline{\lambda}) D = (\underline{\alpha} \cdot \underline{\lambda})\|D\|^2$. Our goal is the maximization of expression (22). Thus we need to maximize

$$\varphi(\underline{\alpha}) = \frac{(\underline{\alpha} \cdot \underline{\lambda})\|D\|^2}{\sqrt{\underline{\alpha}\Pi\underline{\alpha}^t}} \tag{26}$$

over the set of all I-vectors $\underline{\alpha}$ with non-negative components. Clearly, for every non negative constant k we have that $\varphi(k\underline{\alpha}) = \varphi(\underline{\alpha})$. Moreover, only those $\underline{\alpha}$ for which the numerator in $\varphi(\underline{\alpha})$ is positive are of interest, and then multiplying $\underline{\alpha}$ by a suitable positive constant can ensure that the numerator is 1. Maximizing $\varphi(\underline{\alpha})$ is then equivalent to minimizing the nominator, which yields the following quadratic programming problem:

Minimize:

$$\underline{\alpha}\Pi\underline{\alpha}^t \tag{27}$$

subject to

$$(\underline{\alpha} \cdot \underline{\lambda})\|D\|^2 = 1 \tag{28}$$

$$\underline{\alpha} \geq 0 \tag{29}$$

Example 3 *Consider the lower prevision $\underline{P}$ from Example 1. We will calculate the distance $d_{E_1}(E_1, E_5)$, where $E_1 = (0.4, 0.32, 0.28)$ and $E_5 = (0.15, 0.37, 0.48)$. First we have:*

$$D = E_5 - E_1 = (-0.2462, 0.0492, 0.1969),$$

and its norm, which is the Euclidean distance between the two extreme points is $\|D\| = 0.3191$. The positive basis of $N_{\mathcal{M}}(E_1)$ consists of the transformed gambles

$$f'_1 = f_1 - f_1 \cdot 1_{\mathcal{X}}/3 = (-0.5, 0.5, 0)$$
$$f'_5 = f_5 - f_5 \cdot 1_{\mathcal{X}}/3 = (-0.2, 0.6, -0.4).$$

(see Corollary 10).

We have $f_1' = 1.451D + (-0.1429, 0.4286, -0.2857)$*, and since* f_5' *is orthogonal to* D*, it follows that* $u_5 = f_5'$ *and* $\lambda_2 = 0$*. Thus* $\underline{\lambda} = \begin{bmatrix} 1.451 \\ 0 \end{bmatrix}$ *and* $U = \begin{bmatrix} -0.14 & 0.43 & -0.29 \\ -0.20 & 0.60 & -0.40 \end{bmatrix}$ *which gives* $\Pi = \|D\|^2 \underline{\lambda}\,\underline{\lambda}^t + UU^t = \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.56 \end{bmatrix}$*. Taking* $\underline{\alpha} = (\alpha_1, \alpha_2)^t$*, we obtain the objective function to be minimized:* $\underline{\alpha}\Pi\underline{\alpha}^t = 0.5\alpha_1^2 + 0.8\alpha_1\alpha_2 + 0.56\alpha_2^2$ *subject to* $\|D\|^2 \underline{\alpha} \cdot \underline{\lambda} = \|D\|^2 \lambda_1 \alpha_1 = 1$ *whence* $\alpha_1 = 6.7708$*. Substituting* α_1 *in the objective function we obtain* $\underline{\alpha}\Pi\underline{\alpha}^t = 22.9219 + 5.41664\alpha_2 + 0.56\alpha_2^2$*, which has to be minimized subject to* $\alpha_2 \geq 0$*. The minimum is obtained for* $\alpha_2 = 0$*, with the minimal value of objective function* $\underline{\alpha}\Pi\underline{\alpha}^t$ *equal to* 22.9219*. Now* $d_{E_1}(E_1, E_5) = \varphi(\underline{\alpha}) = 1/\sqrt{22.9219} = 0.2089$*. Note that this is significantly less than the Euclidean distance between the points, which is equal to* $\|D\| = 0.3191$.

5. Conclusions and further work

This paper provides as its main contribution a practically computable upper bound for the difference between any two extensions of a coherent lower prevision given on an arbitrary finite set of gambles. The problem is relevant for many applications of the theory of imprecise probabilities, where complete description of lower previsions or their credal sets is often infeasible.

A drawback of the proposed method is that it requires finding all extreme points of credal sets in question. The number of the extreme points in general grows exponentially with the number of constraints, which makes the method computationally demanding.

In future, faster and perhaps less accurate methods could be developed to quickly asses maximal possible error of finite approximation of coherent lower previsions could be developed based on the results proposed in this paper. The method might also be simplified for special cases of approximations, such as coherent lower probabilities.

References

A. Antonucci and F. Cuzzolin. Credal sets approximation by lower probabilities: application to credal networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 716–725. Springer, 2010.

T. Augustin, F. P. Coolen, G. de Cooman, and M. C. Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

P. Gruber. *Convex and Discrete Geometry*. Springer-Verlag Berlin Heidelberg, 2007. doi:10.1007/978-3-540-71133-9.

E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628 – 658, 2008. ISSN 0888-613X. doi:10.1016/j.ijar.2007.12.001. In Memory of Philippe Smets (1938-2005).

M. C. Troffaes and G. De Cooman. *Lower previsions*. John Wiley & Sons, 2014.

D. Škulj and R. Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, dec 2013. ISSN 0026-1335. doi:10.1007/s00184-011-0378-0.