Extreme lower probabilities*

Erik Quaeghebeur ^{*,1} Gert de Cooman

*SYSTeMS Research Group, EESA Department, Ghent University,
Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde, Belgium*

Abstract

We consider lower probabilities on finite possibility spaces as models for the uncertainty about the state. These generalizations of classical probabilities can have some interesting properties; for example: k-monotonicity, avoiding sure loss, coherence, permutation invariance. The sets formed by all the lower probabilities satisfying zero or more of these properties are convex. We show how the extreme points and rays of these sets – the extreme lower probabilities – can be calculated and we give an illustration of our results.

Key words: lower probabilities, extreme points, imprecise probabilities, non-additive measures, combinatorial problems
1991 MSC: 14N10, 52B11, 60A99

1 Introduction

We use and work on theories of imprecise probabilities. This means that we use concepts such as lower (and upper) probabilities to represent uncertainty. In this, lower probabilities serve the same purpose as classical probabilities, which they encompass. They can be used to express a larger variety of states of uncertainty, but generally at the cost of an increase in mathematical and computational complexity.

Usually, many different states of uncertainty – or, looking at it from another viewpoint, states of knowledge – are compatible with a given set of rationality criteria or any other property we want the state of uncertainty to satisfy. For many interesting such properties, the compatible states of uncertainty, when represented as lower probabilities, form convex sets. The extreme points of these convex sets are what we call *extreme lower probabilities*.

We first got interested in extreme lower probabilities when we read a paper by Maaß [11]. Apart from giving the theoretical basis and context for extreme lower probabilities, he also identifies them in an interesting basic case; he does this by hand and in an ad-hoc way. Both the theoretical basis as well as the example piqued our interest: we reasoned that if we could let a computer calculate them in a systematic way for less academic cases, then we could use them to our advantage in some other research domain of ours.

So, as a result of the research that originated with that idea, we show in this paper, for some important rationality criteria and a few other interesting properties, how extreme lower probabilities can be calculated in a systematic way. We do this for the case of a finite possibility space. In Section 2, we give a description of our systematic approach. In Section 3 we look at the rationality criteria and properties we have investigated (k-monotonicity, avoiding sure loss, coherence, and permutation invariance) and derive the results necessary for applying our systematic approach to these cases.

This paper also contains more practical results: in Section 4 we present and comment on the extreme lower probabilities for a number of especially interesting cases. Dealing with possibility spaces of cardinality 4 or less, these results are admittedly still somewhat academic. This is mainly due to the computational complexity of the problem for the most interesting properties (2-monotonicity and coherence).

To get started, let us next clear up some terminology and introduce some basic concepts, notation, and assumptions.

1.1 *Basic concepts from imprecise probability theory and convex analysis*

A *lower probability* $\underline{P}$ is a concept used in almost all the theories that make up the field of imprecise probabilities; it generalizes the classical concept of a probability P. Important examples of such theories are the ones described by Dempster [7] and Shafer [15], Fine [8], Walley [18], and Weichselberger [19].

Although the definition and interpretation of a lower probability may differ between the theories, the idea is similar. Like a probability, it is a real-valued set

function defined on the set $\wp(\Omega)$ of all *events* (subsets A, B, C) of a *possibility space* Ω of *states*.[2] We are uncertain about the state. As mentioned before, we restrict ourselves to finite possibility spaces Ω of cardinality $|\Omega| = n \in \mathbb{N}$.[3]

A probability P is

  (i) positive: for any event A, $P(A) \geq 0$,
  (ii) normed: $P(\Omega) = 1$, and
  (iii) additive: for disjoint events B and C, $P(B \cup C) = P(B) + P(C)$.

These are the properties a classical probability has to satisfy: They are a consequence of the typical rationality criteria when interpreted as a degree of belief or of its interpretation as a frequency otherwise. Similarly, a lower probability $\underline{P}$ must satisfy some properties, but they are weaker than those for probabilities. A lower probability is usually required to be at least

  (i) normed, and
  (ii) super-additive: for disjoint events B and C, $\underline{P}(B \cup C) \geq \underline{P}(B) + \underline{P}(C)$.

In this paper, we do not assume a priori that a lower probability satisfies any property (not even positivity).

A probability P *dominates* a lower probability $\underline{P}$ if $\underline{P}(A) \leq P(A)$ for all events A. A probability P is called *degenerate* when it assumes the value 1 on a singleton.

The set of all lower probabilities defined on some possibility space can be seen as a *closed* and *convex* subset of a finite dimensional space. This is also the case for the set of all lower probabilities additionally satisfying some interesting properties (as discussed further on). Any such closed convex set (that contains no lines) is fully determined by its *extreme points* and *extreme rays* (directions), and vice-versa [14, Thm 18.5]: all elements of the set can be written as a sum of

  (i) a positive linear combination of extreme rays, and
  (ii) a convex combination of extreme points.

The extreme points and extreme rays of a convex set of lower probabilities are its *extreme lower probabilities*.

---

[2] An *upper probability* $\overline{P}$ can be defined using its so-called conjugate lower probability $\underline{P}$: for any event A, $\overline{P}(A) = 1 - \underline{P}(\Omega \setminus A)$. Because $\wp(\Omega)$ is closed under complementation, we can – and do so here – work with lower probabilities only.
[3] Notation for number sets: $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ respectively denote the nonnegative integers, the rationals, and the reals. To denote common subsets of these, we use predicates as subscripts; e.g., $\mathbb{R}_{>0} = \{r \in \mathbb{R} | r > 0\}$ denotes the strictly positive reals.

If you can see that a triangle can be described by its three vertices, you have understood the main idea behind extreme lower probabilities. Of course, we will be talking about things that are a bit more complicated than triangles.

We will look at sets of lower probabilities satisfying some interesting properties; for example 2-monotonicity, which is characterized by the following finite number of constraints:

(i) for any two events B and C such that $B \subseteq C$, $\underline{P}(B) \leq \underline{P}(C)$, and

(ii) for any events B and C, $\underline{P}(B \cup C) + \underline{P}(B \cap C) \geq \underline{P}(B) + \underline{P}(C)$.

It turns out that for finite possibility spaces, not surprisingly, the number of extreme points and extreme rays of the set of all lower probabilities that satisfy one or more of these properties is finite.

In the next section, we will outline our approach to calculating extreme lower probabilities: calculating a set of constraints, and using these to compute extreme points. Then, we look at how we can obtain *manageable* sets of constraints for the properties that are of interest to us.

2 On constraints and vertex enumeration

A (lower) probability can be required to satisfy a variety of interesting properties; most can be expressed using sets of linear constraints. (We do not consider requirements related to things like independence and conditioning, which typically lead to nonlinear constraints.) In the Introduction we have already seen three simple examples: additivity, super-additivity, and 2-monotonicity. In general, the constraints for these properties are linear inequalities; equalities are expressed using two inequalities.

Sometimes, not all constraints are necessary: a constraint can be implied by another constraint, or a set of other constraints. Such a so-called *redundant* constraint can be removed from the set of constraints. If a constraint makes another one redundant, we call the former more *stringent* than the latter.

It is useful to look at this problem in a geometrical framework. Lower probabilities can be viewed as points in a 2^n-dimensional vector space – one dimension per subset in $\wp(\Omega)$. A linear inequality then generally corresponds to a half-space delimited by a hyperplane. With some property, there corresponds a set of half-spaces, and the intersection of these is the convex set of all lower probabilities satisfying that property.

The geometrical approach is illustrated in Figure 1 below using a toy example.

We consider $n = 1$, so the dimension of the vector space is 2 (this is actually the only one we can draw directly). The constraints are given using a set of hyperplanes $\{h_i | i = 1, \ldots, 6\}$, the 'hairs' indicate the half-spaces the constraints correspond to. Constraints h_3 and h_6 are redundant; the former because of h_1 and h_2, the latter because h_5 is more stringent. The set of points satisfying the constraints is filled.
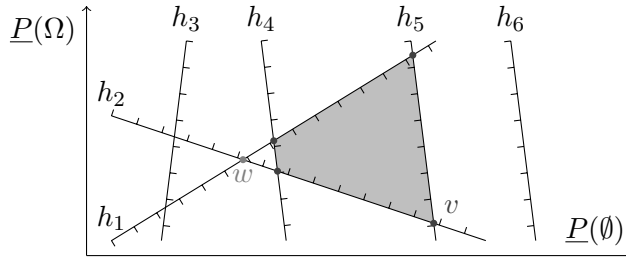

Fig. 1. Illustration of constraints and vertices.

Figure 1 also shows the vertices – v, for example – of the convex set defined by the set of half-spaces (constraints). These are the geometrical equivalent of what we call extreme lower probabilities.

In general, the *vertex enumeration problem* – finding the vertices corresponding to a given set of half-spaces – is hard: no polynomial time algorithm (in dimension, number of constraints, and number of vertices) is known [9]. To get a feeling for the complexity, realize that not only all intersections of nonredundant constraints have to be found, but we must also decide which of these to discard (such as w in Figure 1).

To obtain the set of constraints for different cardinalities and for the various properties we have looked at, we have written our own program, `constraints` [13]. The properties in question will be described in the next section, where we will also give the theorems that form the basis for this program.

We have used publicly available programs (`redund`, `lrs`, and `cdd`) to remove redundant constraints and do vertex enumeration. These are maintained by Avis [1] and Fukuda [10].

3   From properties to constraints

Although for some properties – such as avoiding sure loss (see below) and additivity plus positivity plus normalization – it is possible to obtain the corresponding extreme points directly, we were not able to do this for the most interesting ones (k-monotonicity and coherence). For these, we use the

vertex enumeration approach described in the previous section: generate the constraints corresponding to the property of interest and then use vertex enumeration to obtain the corresponding extreme lower probabilities. With this approach it is also easy to combine properties; one just has to combine the corresponding sets of constraints. The (big) downside is that it cannot be used in practice for 'large' possibility spaces – large here means $n \geq 5$. The reasons for this are, first of all, as mentioned in the previous section, that vertex enumeration is hard, and secondly, that the dimension of the space in which we work rises exponentially with the cardinality of the possibility space, i.e., as 2^n.

We will now look at how we can obtain the constraints for some interesting properties. The typical definitions for these properties either give a lot of redundant constraints or even an infinite number of constraints. We reformulate them to reduce the number of redundant constraints and, if necessary, to ensure that they generate only a finite number of constraints.

To start out with a clean slate, we assume nothing about lower probabilities at this point beyond the fact that they are real-valued set functions defined on $\wp(\Omega)$; we do not even assume that they are positive, normed, or super-additive.

3.1 k-Monotonicity

In the theory of Dempster [7] and Shafer [15] lower probabilities are completely monotone. This is an extreme case of a mathematically interesting type of property: *k-monotonicity*, where $k \in \mathbb{N}_{>0}$.

A formal definition (adapted from De Cooman et al. [5,6] and based on Choquet [3]): [4]

Definition 1 *A lower probability $\underline{P}$ is k-monotone if and only if for all $\ell = 1, \ldots, k$, any event A and any ℓ-tuple of events $(B_i | i \in \mathbb{N}_{<\ell})$, it holds that*

$$\sum_{I \subseteq \mathbb{N}_{<\ell}} (-1)^{|I|} \underline{P}(A \cap \bigcap_{i \in I} B_i) \geq 0.$$

In this definition, we used the convention $\bigcap_{i \in \emptyset} B_i = \Omega$. You can see that a k-monotone lower probability is also ℓ-monotone, for $\ell = 1, \ldots, k-1$. We have seen the case $k = 2$ in the introduction (in a different, but equivalent form).

---

[4] The relationships with other definitions in the literature are implicitly clarified by the steps we take in the proof of Theorem 2.

Monotonicity is the same as 1-monotonicity: $\underline{P}(B) \le \underline{P}(A)$ for any event A and all $B \subseteq A$. *Completely monotone* means k-monotone for all $k \in \mathbb{N}_{>0}$.

The above definition gives rise to a lot of redundant constraints. Many of them are equivalent or are trivially satisfied. Removing them allows us to formulate the following definition, leading to a more efficient program.

Theorem 2 (Constraints for k-monotonicity) *A lower probability $\underline{P}$ is called k-monotone if and only if it is monotone, and for all nonempty events A and all sets of events $\mathcal{A} \subseteq \wp(A) \setminus \{A, \emptyset\}$ that satisfy*

(i) $0 < |\mathcal{A}| \le k$,
(ii) $\bigcup_{B \in \mathcal{A}} B = A$, *and such that*
(iii) *no event $C \in \mathcal{A}$ exists for which $C = \bigcap_{B \in \mathcal{A}} B$,*

it holds that $\underline{P}(A) + \sum_{\mathcal{B} \subseteq \mathcal{A}, \mathcal{B} \ne \emptyset} (-1)^{|\mathcal{B}|} \underline{P}(\bigcap_{B \in \mathcal{B}} B) \ge 0$.

PROOF. We are going to start from the definition of k-monotonicity and show that the characterization given in the theorem is equivalent.

A lower probability is $\underline{P}$ is k-monotone if and only if (Definition 1)

$$\forall A \subseteq \Omega, \quad \forall \ell = 1, \dots, k, \quad \forall (B_i | i \in \mathbb{N}_{<\ell}) \subseteq \Omega^\ell,$$
$$\sum_{I \subseteq \mathbb{N}_{<\ell}} (-1)^{|I|} \underline{P}(A \cap \bigcap_{i \in I} B_i) \ge 0,$$

where the convention $\bigcap_{i \in \emptyset} B_i = \Omega$ is used. We call A the *generating event*.

First we observe that only intersections with A play any role in this definition. So we only need to consider ℓ-tuples $(B_i | i \in \mathbb{N}_{<\ell})$ of events in $\wp(A)$. When, for $\ell > 1$, there is a duplicate in the ℓ-tuple, the corresponding constraint reduces to a constraint for the $(\ell - 1)$-tuple formed by removing one of the duplicates. To see this, assume without loss of generality that $B_{\ell-1} = B_{\ell-2}$ and rewrite the left-hand side of the constraint as follows ($\Diamond_I$ stands for the expression $A \cap \bigcap_{i \in I} B_i$):

$$\sum_{I \subseteq \mathbb{N}_{<\ell}} (-1)^{|I|} \underline{P}(\Diamond_I) = \sum_{I \subseteq \mathbb{N}_{<\ell-2}} (-1)^{|I|} \Big(\underline{P}(\Diamond_I) - \underline{P}(B_{\ell-2} \cap \Diamond_I) $$
$$+ \underline{P}(B_{\ell-2} \cap B_{\ell-1} \cap \Diamond_I)$$
$$- \underline{P}(B_{\ell-1} \cap \Diamond_I) \Big)$$
$$= \sum_{I \subseteq \mathbb{N}_{<\ell-2}} (-1)^{|I|} \Big(\underline{P}(\Diamond_I) - \underline{P}(B_{\ell-2} \cap \Diamond_I) \Big)$$
$$= \sum_{I \subseteq \mathbb{N}_{<\ell-1}} (-1)^{|I|} \underline{P}(\Diamond_I).$$

So, rather than working with ℓ-tuples of events for all $\ell = 1, \dots, k$, we can also work more efficiently with *sets* of events $\mathcal{A} \subseteq \wp(A)$ where $0 < |\mathcal{A}| \le k$. These

observations lead us to the reformulated condition [5]

$$\forall A \subseteq \Omega, \quad \forall \mathcal{A} \subseteq \wp(A) \text{ s.t. } 0 < |\mathcal{A}| \leq k, \quad \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{|\mathcal{B}|} \underline{P}(\bigcap_{B \in \mathcal{B}} B) \geq 0,$$

where now the convention $\bigcap_{B \in \emptyset} B = A$ is used.

Whenever $A \in \mathcal{A}$, the resulting constraint

$$\sum_{\mathcal{B} \subseteq \mathcal{A} \setminus \{A\}} \left((-1)^{|\mathcal{B}|} \underline{P}(\bigcap_{B \in \mathcal{B}} B) + (-1)^{|\mathcal{B} \cup \{A\}|} \underline{P}(A \cap \bigcap_{B \in \mathcal{B}} B) \right) = 0 \geq 0$$

is trivially satisfied. A special instance of this obtains when $A = \emptyset$ and therefore $\mathcal{A} = \{A\}$. So we can write the condition as

$$\forall A \subseteq \Omega \text{ s.t. } A \neq \emptyset, \quad \forall \mathcal{A} \subseteq \wp(A) \setminus \{A\} \text{ s.t. } 0 < |\mathcal{A}| \leq k,$$
$$\underline{P}(A) + \sum_{\mathcal{B} \subseteq \mathcal{A}, \mathcal{B} \neq \emptyset} (-1)^{|\mathcal{B}|} \underline{P}(\bigcap_{B \in \mathcal{B}} B) \geq 0,$$

where we have written the case $\mathcal{B} = \emptyset$ as a separate term.

To complete the proof, we must add two more considerations.

Consider monotonicity separately. Then for $\mathcal{A}$ such that $\bigcup_{B \in \mathcal{A}} B \neq A$, the corresponding constraint is implied by $\underline{P}(A) \geq \underline{P}(\bigcup_{B \in \mathcal{A}} B)$ (monotonicity) and the constraint for $\bigcup_{B \in \mathcal{A}} B$ as the generating event.

Consider the case where $C = \bigcap_{B \in \mathcal{A}} B \in \mathcal{A}$. Then, because of the binomial theorem,

$$\sum_{\mathcal{B} \subseteq \mathcal{A}, C \in \mathcal{B}} (-1)^{|\mathcal{B}|} \underline{P}(\bigcap_{B \in \mathcal{B}} B) = \underline{P}(C) \sum_{\mathcal{B} \subseteq \mathcal{A}, C \in \mathcal{B}} (-1)^{|\mathcal{B}|} = 0,$$

so this leads to the same constraint as when we use $\mathcal{A} \setminus \{C\}$ rather than $\mathcal{A}$. The situation $\emptyset \in \mathcal{A}$ is a special case of this. $\square$

It may also be interesting to remark that k-monotonicity for any $k \geq n$ is equivalent to complete monotonicity [2, Cor. 1]. Note that the conditions $\underline{P}(\emptyset) = 0$ and $\underline{P}(\Omega) = 1$ are not included in our definition for k-monotonicity; they are commonly added, however.

------

[5] Note the close correspondence of this form with the definition given by Chateauneuf and Jaffray [2].

3.2   Avoiding sure loss

Avoiding sure loss is a property that is of central importance in Walley's [18, §2.4.1] behavioral theory.[6]

Definition 3 *A lower probability $\underline{P}$ avoids sure loss if and only if for all $\mathcal{B} \subseteq \wp(\Omega)$, and all $\lambda \in (\mathbb{R}_{\geq 0})^{\mathcal{B}}$, it holds that*

$$\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) \leq \sup \sum_{B \in \mathcal{B}} \lambda_B I_B.$$

In this definition, I_B is the indicator function of B, which is 1 for $\omega \in B$ and 0 elsewhere.

If we only require a lower probability to avoid sure loss, and nothing more, the extreme lower probabilities can be determined almost directly: Because a lower probability avoids sure loss if and only if it is dominated by some probability [18, §3.3.3], the set of extreme lower probabilities consists of the degenerate probabilities as extreme points and all negative main directions as extreme rays.

If we want to use vertex enumeration to obtain the extreme lower probabilities, the problem arises that Definition 3 gives an infinite number of constraints (because λ can take on an infinite number of values). This is of course unmanageable for any computer. Dealing with this infinite amount of constraints becomes inevitable when, in addition to having to avoid sure loss, the lower probabilities also have to be positive, something which is usually required. Moreover, determining the extreme points and rays of the set of lower probabilities that avoid sure loss using the vertex enumeration approach is a good preparation for the next subsection about coherence.

Luckily – by removing redundant constraints –, we can reduce the set of constraints for avoiding sure loss to a finite set. It was shown by Walley [18, §A] that in Definition 3, to get the most stringent constraints, we only need to consider

(i) $\mathcal{B}$ such that $\{I_B | B \in \mathcal{B}\}$ is a linearly independent set and $\bigcup_{B \in \mathcal{B}} B = \Omega$, and

(ii) λ such that the function on right-hand side is the constant function 1.

Walley [18, §A3] assumes positivity ($\underline{P}$ is in the first orthant), but mentions this is not a substantive assumption. When we want to do the vertex enumeration

---

[6] In Walley's theory [18], *lower previsions* – expectation operators – play a central role. Here, we restrict ourselves to the less general lower probabilities.

approach without this assumption (as we do here), we need to add extra constraints for the cases where some (or all) of the components of $\underline{P}$ are negative. This can be done by taking every original constraint, and creating a new one for each of the 2^{2^n} possible orthants $\underline{P}$ can be located in. We do this by setting the λ_B for which $\underline{P}(B) < 0$ to 0. (More details can be found in the proof of Theorem 4, which is based on Walley's [18, §A] work, some of which he left implicit.)

The above and some other, minor, changes result in the following theorem.

Theorem 4 (Constraints for avoiding sure loss) *A lower probability $\underline{P}$ avoids sure loss if and only if $\underline{P}(\emptyset) \leq 0$, $\underline{P}(\Omega) \leq 1$ and for all*

(i) $\mathcal{B} \subseteq \wp(\Omega) \setminus \{\emptyset, \Omega\}$ *such that $\bigcup_{B \in \mathcal{B}} B = \Omega$, $\{I_\Omega\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly dependent set, and $\{I_B | B \in \mathcal{B}\}$ is a linearly independent set,*
(ii) $\lambda \in (\mathbb{Q} \cap {]0,1]})^{\mathcal{B}}$ *such that $\sum_{B \in \mathcal{B}} \lambda_B I_B = 1$,[7] and*
(iii) *binary masks $\beta \in \{0,1\}^{\mathcal{B}}$,*

it holds that $\sum_{B \in \mathcal{B}} \beta_B \lambda_B \underline{P}(B) \leq 1$.

PROOF. A lower probability $\underline{P}$ avoids sure loss if and only if (Definition 3)

$$\forall \mathcal{B} \subseteq \wp(\Omega), \quad \forall \lambda \in (\mathbb{R}_{\geq 0})^{\mathcal{B}}, \quad \sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) \leq \sup \sum_{B \in \mathcal{B}} \lambda_B I_B. \qquad (\star)$$

First of all, we can restrict λ to $(\mathbb{R}_{>0})^{\mathcal{B}}$; for if $\lambda_B = 0$ for some $B \in \mathcal{B}$, the condition reduces to one with a set $\mathcal{B}$ of lower cardinality.

Secondly, applying the definition with $\mathcal{B} = \{\emptyset\}$ and $\mathcal{B} = \{\Omega\}$ gives us $\underline{P}(\emptyset) \leq 0$ and $\underline{P}(\Omega) \leq 1$. By looking at how they affect the right-hand and left-hand side of the constraint in $(\star)$, these constraints can be seen to make any constraint for a $\mathcal{B}$ including $\emptyset$ or Ω redundant relative to the same constraint for which $\emptyset$ and Ω have been dropped from $\mathcal{B}$.

Each of the resulting constraints can now be replaced by a set of constraints written in some standard form.

First, we focus on the constraint's right-hand side only. By renormalizing λ, the right-hand side supremum can always be made to equal 1, causing λ to have components in ${]0,1]}$. Additionally, by increasing λ_B for some $B \in \mathcal{B}$, adding some well-chosen $C \in \wp(\Omega) \setminus \mathcal{B}$ to $\mathcal{B}$ (e.g., singletons not already in $\mathcal{B}$), and choosing λ_C appropriately, we can obtain $\sum_{B \in \mathcal{B}} \lambda_B I_B = I_\Omega = 1$. Moreover,

---

[7] We use pointwise extension of common binary operators, e.g., $I_\Omega = 1$ is shorthand for $\forall \omega \in \Omega, I_\Omega(\omega) = 1$ and $\beta \underline{P} = (\beta_B \underline{P}(B) | B \in \mathcal{B})$.

this procedure causes $\bigcup_{B \in \mathcal{B}} B = \Omega$ and $\{I_\Omega\} \cup \{I_B | B \in \mathcal{B}\}$ to be a linearly dependent set.

What is the effect of this procedure on the constraint as a whole? We need to ensure it results in a constraint that is either equivalent or more stringent. For $\underline{P}$ that are nonnegative everywhere, the left-hand side increases by increasing some λ_B, adding some C, and choosing λ_C, whereas the right-hand side remains constant (i.e., equal to 1), so the constraint becomes more stringent. However, when $\underline{P}$ has negative components, the left-hand side might decrease, leading to a possibly less stringent constraint. To counter this, we use binary masks $\beta \in \{0,1\}^\mathcal{B}$ to create a specific constraint for all possible orthants that $\underline{P}$ can be located in. So we get a new collection of constraints; to wit, for every $\beta \in \{0,1\}^\mathcal{B}$,

$$\sum_{B \in \mathcal{B}} \beta_B \lambda_B \underline{P}(B) \leq \sup \sum_{B \in \mathcal{B}} \lambda_B I_B = 1.$$

Let us explain how these binary masks work: Whenever $\underline{P}(B) < 0$ for some $B \in \mathcal{B}$, there is a constraint for which $\beta_B = 0$, which will be more stringent than the original constraint. We need to be cautious, however, because only when $\beta = 1$ (suitable for $\underline{P}$ in the first orthant, with only nonnegative components) we have a constraint that is of the form given by Equation $(\star)$. The others might be *too* stringent, meaning that they are not implied by the criterion of avoiding sure loss.

Luckily, however, these constraints turn out not to be too stringent: We can show that the constraints for which β is not identically 1 are also implied by the original definition of avoiding sure loss. This follows by replacing the λ_B in the original definition by $\beta_B \lambda_B$ and realizing that $\sup \sum_{B \in \mathcal{B}} \beta_B \lambda_B I_B \leq \sup \sum_{B \in \mathcal{B}} \lambda_B I_B = 1$.

So we end up with a set of constraints for which the summation condition $\sum_{B \in \mathcal{B}} \lambda_B I_B = 1$ holds. We can now use Lemma 5 with $\pi = \beta \underline{P}$ and $r = 1$ to find that we only need to consider those $\mathcal{B} \in \wp(\Omega)$ such that $\{I_B - \underline{P}(B) | B \in \mathcal{B}\}$ is linearly independent; because all the components of λ are strictly positive, the summation condition stays satisfied (possibly after a renormalization of λ to ensure $r = 1$). Lemma 6 then shows that linear independence of $\{I_B - \underline{P}(B) | B \in \mathcal{B}\}$ is equivalent to linear independence of $\{I_B | B \in \mathcal{B}\}$. This means $\sum_{B \in \mathcal{B}} \lambda_B I_B = 1$ is a regular system of linear equations in λ having a unique rational solution, which finishes the proof. $\square$

Lemma 5 (Inspired by Walley [18, §A1]) *Given some $\pi \in \mathbb{R}^{\wp(\Omega)}$, the set of constraints defined by*

$$\Big\{ \sup \sum_{B \in \mathcal{B}} \lambda_B (I_B - \pi_B) \geq 0 \Big| \mathcal{B} \subseteq \wp(\Omega),$$
$$\lambda \in (\mathbb{R}_{\neq 0})^\mathcal{B} \text{ such that } \sum_{B \in \mathcal{B}} \lambda_B I_B = r \Big\},$$

where r is some real number, is equivalent to the set where only those $\mathcal{B} \subseteq \wp(\Omega)$ are considered for which the set $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly independent, if the value of r is allowed to change. Additionally, conditions on the signs of the components of λ can be maintained.

PROOF. It suffices to prove that any constraint determined

   (i) by some $\mathcal{B} \in \wp(\Omega)$ for which $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly dependent, and
   (ii) by some $\lambda \in (\mathbb{R}_{\neq 0})^{\mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \lambda_B I_B = r$, $r \in \mathbb{R}$,

is equivalent to a constraint determined by some $\mathcal{B}' \subset \mathcal{B}$ for which the set $\{I_B - \pi_B | B \in \mathcal{B}'\}$ is linearly independent, and by some $\lambda' \in (\mathbb{R}_{\neq 0})^{\mathcal{B}'}$ such that $\sum_{B \in \mathcal{B}'} \lambda'_B I_B = r'$, $r' \in \mathbb{R}$, and for which it holds for all $B \in \mathcal{B}'$ that λ_B and λ'_B have the same sign.

Linear dependence of the set $\{I_B - \pi_B | B \in \mathcal{B}\}$ means that there is some $\mu \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \mu_B (I_B - \pi_B) = 0$. So the constraint can be rewritten as $\sup \sum_{B \in \mathcal{B}} \lambda_B (1 + c\frac{\mu_B}{\lambda_B})(I_B - \pi_B) \geq 0$, where $c \in \mathbb{R}$ can be freely chosen. Choosing $c = -\frac{\lambda_C}{\mu_C}$, where C is an element of $\mathcal{B}$ such that $\mu_C \neq 0$, reduces the constraint to one not explicitly dependent on C, which can therefore be excluded from $\mathcal{B}$.

For all $B \in \mathcal{B}'$, the sign of λ_B and $\lambda'_B = \lambda_B(1 + c\frac{\mu_B}{\lambda_B})$ are the same when $0 \leq 1 + c\frac{\mu_B}{\lambda_B} = 1 - \frac{\lambda_C}{\mu_C}\frac{\mu_B}{\lambda_B}$. This can be guaranteed by not only choosing $C \in \mathcal{B}$ such that $\mu_C \neq 0$, but moreover such that $|\frac{\lambda_C}{\mu_C}| = \min_{B \in \mathcal{B}, \mu_B \neq 0} |\frac{\lambda_B}{\mu_B}|$.

To see that the new coefficients $\lambda' = \lambda + c\mu$ introduced above also satisfy the summation condition $\sum_{B \in \mathcal{B}'} \lambda'_B I_B = r'$ and discover the value of r', first note that $\lambda'_C = 0$ by construction, so

$$
\begin{aligned}
\sum_{B \in \mathcal{B}'} \lambda'_B I_B &= \sum_{B \in \mathcal{B}} \lambda_B(1 + c\frac{\mu_B}{\lambda_B})I_B \\
&= \sum_{B \in \mathcal{B}}(\lambda_B I_B + c\mu_B I_B) \\
&= r + c\sum_{B \in \mathcal{B}} \mu_B \pi_B = r',
\end{aligned}
$$

where, to obtain the third equality, we used $\sum_{B \in \mathcal{B}} \mu_B(I_B - \pi_B) = 0$.

Possibly after repeating this procedure, one eventually obtains a $\mathcal{B}' \subset \mathcal{B}$ for which $\{I_B - \pi_B | B \in \mathcal{B}'\}$ is linearly independent. $\quad\square$

Lemma 6 *Consider $\mathcal{B} \subseteq \wp(\Omega)$ and $\lambda \in \mathbb{R}^{\mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \lambda_B I_B = 1$. Then for any $\pi \in \mathbb{R}^{\mathcal{B}}$ it holds that $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly (in)dependent if and only if $\{I_B | B \in \mathcal{B}\}$ is linearly (in)dependent.*

PROOF. Fix π.

First we prove that linear dependence of $\{I_B - \pi_B | B \in \mathcal{B}\}$ implies linear dependence of $\{I_B | B \in \mathcal{B}\}$.

If $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly dependent, there is a $\mu \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \mu_B I_B = \sum_{B \in \mathcal{B}} \mu_B \pi_B$. Define $c(\mu) = \sum_{B \in \mathcal{B}} \mu_B \pi_B$. Under the assumption of the lemma, $\sum_{B \in \mathcal{B}} c(\mu) \lambda_B I_B = \sum_{B \in \mathcal{B}} \mu_B \pi_B$. So with $\mu' = \mu - c(\mu)\lambda$ we find that $\sum_{B \in \mathcal{B}} \mu'_B I_B = 0$, so $\{I_B | B \in \mathcal{B}\}$ is linearly dependent.

Now we prove that linear independence of $\{I_B - \pi_B | B \in \mathcal{B}\}$ implies linear independence of $\{I_B | B \in \mathcal{B}\}$.

If $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly independent, then for all $\mu \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$, it holds that $\sum_{B \in \mathcal{B}} \mu_B I_B \neq c(\mu) = \sum_{B \in \mathcal{B}} \mu_B \pi_B$. Again using the assumption of the lemma, we find that for all $\mu \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$, $\sum_{B \in \mathcal{B}} (\mu_B - c(\mu)\lambda_B) I_B \neq 0$. So if for all $\mu' \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$ there is a $\mu \in \mathbb{R}^{\mathcal{B}} \setminus \{0\}^{\mathcal{B}}$ such that $\mu' = \mu - c(\mu)\lambda$, then $\{I_B | B \in \mathcal{B}\}$ is linearly independent and the lemma is proven.

To finish the proof, we have to show, under the given conditions on μ and μ', that the linear system $\mu' = \mu - c(\mu)\lambda$ of equations in μ can always be solved for μ when $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly independent. Or, using the expression for $c(\mu)$ and considering that $\mu = 0$ implies $\mu' = 0$, it suffices to show that the system's coefficient matrix $\mathbb{I} - \lambda \pi^T$ has a nonzero determinant. To do this, we create a well-chosen matrix, and rewrite it as two different products

$$\begin{pmatrix} \mathbb{I} & \lambda \\ \pi^T & 1 \end{pmatrix} = \begin{pmatrix} \mathbb{I} & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbb{I} - \lambda \pi^T & 0 \\ \pi^T & 1 \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ \pi^T & 1 \end{pmatrix} \begin{pmatrix} \mathbb{I} & \lambda \\ 0 & 1 - \pi^T \lambda \end{pmatrix}.$$

Taking the determinant gives

$$\begin{vmatrix} \mathbb{I} & \lambda \\ \pi^T & 1 \end{vmatrix} = |\mathbb{I} - \lambda \pi^T| = 1 - \pi^T \lambda.$$

So $|\mathbb{I} - \lambda \pi^T| \neq 0$ when $1 \neq \sum_{B \in \mathcal{B}} \lambda_B \pi_B$. Because of the assumption of the lemma, we should have $\sum_{B \in \mathcal{B}} \lambda_B I_B \neq \sum_{B \in \mathcal{B}} \lambda_B \pi_B$. This is satisfied because $\{I_B - \pi_B | B \in \mathcal{B}\}$ is linearly independent. □

3.3 Coherence

Coherent lower probabilities are at the core of Walley's theory [18, §2.5.4]. Weichselberger [19] uses the term F-probability for the very closely related concept (identical for finite spaces).

Definition 7 *A lower probability $\underline{P}$ is* coherent *if and only if it avoids sure loss and for all events A, all $\mathcal{B} \subseteq \wp(\Omega) \setminus \{A\}$, and all $\lambda \in (\mathbb{R}_{\geq 0})^{\{A\} \cup \mathcal{B}}$, it holds that*

$$\textstyle\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) - \lambda_A \underline{P}(A) \leq \sup(\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A).$$

Coherence implies avoiding sure loss and monotonicity, but it does not imply 2-monotonicity.

The above definition is similar enough to the one for avoiding sure loss to allow the same techniques for the removal of redundant constraints (implicit in Walley [18, §A]) to be applied, up to some technicalities. Because coherence implies positivity, there is no need to use binary masks. Working this out results in the following theorem.

Theorem 8 (Constraints for coherence) *A lower probability $\underline{P}$ is coherent if and only if $\underline{P}(\emptyset) = 0$, $\underline{P}(\Omega) = 1$, and*

(*a*) *for all events A it holds that $0 \leq \underline{P}(A) \leq 1$;*
(*b*) *for all*
  (*i*) *nonempty events A and $\mathcal{B} \subseteq \wp(A) \setminus \{A, \emptyset\}$ such that $\bigcup_{B \in \mathcal{B}} B = A$, $\{I_A\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly dependent set, and $\{I_B | B \in \mathcal{B}\}$ is a linearly independent set, and*
  (*ii*) *$\lambda \in (\mathbb{Q} \cap \,]0, 1])^{\mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \lambda_B I_B = I_A$,*
  it must hold that $\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) \leq \underline{P}(A)$;
(*c*) *for all*
  (*i*) *events A different from $\emptyset$ and Ω and $\mathcal{B} \subseteq \wp(\Omega) \setminus \{A, \emptyset, \Omega\}$ such that $\bigcup_{B \in \mathcal{B}} B = \Omega$, $\{I_A, I_\Omega\}\{I_A\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly dependent set, and $\{I_A\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly independent set, and*
  (*ii*) *$\lambda \in (\mathbb{Q}_{>0})^{\{A\} \cup \mathcal{B}}$ such that $\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A = 1$,*
  it must hold that $\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) - \lambda_A \underline{P}(A) \leq 1$.

PROOF. A lower probability $\underline{P}$ is coherent if and only if (Definition 7) it avoids sure loss and

$$\forall A \subseteq \Omega, \quad \forall \mathcal{B} \subseteq \wp(\Omega) \setminus \{A\}, \quad \forall \lambda \in (\mathbb{R}_{\geq 0})^{\{A\} \cup \mathcal{B}},$$
$$\textstyle\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) - \lambda_A \underline{P}(A) \leq \sup(\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A).$$

First of all, we can restrict the $\mathcal{B}$-components of λ to $\mathbb{R}_{>0}$; for if $\lambda_B = 0$ for some $B \in \mathcal{B}$, the condition reduces to one with a set $\mathcal{B}$ of lower cardinality. The case $\lambda_A = 0$ is subsumed by avoiding sure loss. We will see that it can be recuperated later and now assume $\lambda_A > 0$.

Secondly, let us look at two simple cases:

(i) Taking $\mathcal{B} = \emptyset$, we find $\underline{P}(A) \geq \inf I_A$, so $\underline{P}(\Omega) \geq 1$ and for all $A \subset \Omega$, $\underline{P}(A) \geq 0$.

(ii) Taking $\lambda_A = 0$ and $\mathcal{B} = \{B\}$, we find $\underline{P}(B) \leq \sup I_B$, so $\underline{P}(\emptyset) \leq 0$ and for all nonempty $A \subseteq \Omega$, $\underline{P}(A) \leq 1$.

This proves all the conditions up to and including (a); among which $\underline{P}(\emptyset) = 0$ and $\underline{P}(\Omega) = 1$, together with the fact that $I_\emptyset$ and I_Ω are the constant functions 0 and 1, imply that any constraint for a $\mathcal{B}$ containing Ω or $\emptyset$ can be replaced by an equivalent constraint where Ω and $\emptyset$ have been removed from $\mathcal{B}$.

Each of the resulting constraints can now be replaced by a set of constraints written in some standard form. The right-hand side supremum is either 0, or – if this is not the case – equal to 1 or -1 (after normalizing λ). Even more, by increasing λ_B for some $B \in \mathcal{B}$, possibly decreasing λ_A (or even adding A to $\mathcal{B}$, which results in constraints subsumed by avoiding sure loss) and adding some well-chosen $C \in \wp(\Omega) \setminus \mathcal{B}$ to $\mathcal{B}$ (e.g., singletons not already in $\mathcal{B} \cup \{A\}$), we can obtain $\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A \in \{0, 1, -1\}$. As we have already shown that the components of $\underline{P}$ must be nonnegative, this procedure leads to either equivalent or more stringent constraints (the left-hand side never decreases). We investigate the three cases separately:

[0] This situation can only occur when $\bigcup_{B \in \mathcal{B}} B = A$. In this case we can normalize λ such that $\lambda_A = 1$ and the $\mathcal{B}$-components of λ are in $]0, 1]$ and such that $\sum_{B \in \mathcal{B}} \lambda_B I_B = I_A$ and thus that $\{I_A\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly dependent set. The constraint becomes $\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) \leq \underline{P}(A)$. Note that by replacing A by Ω, this case can be seen to encompass the constraints for avoiding sure loss; we therefore drop the separate requirement of avoiding sure loss and allow A to be Ω in this case.

We can use Lemma 5 – applied relative to A, i.e., with A taking the role of Ω – with $\pi = \underline{P}$ and $r = 1$ to find that we only need to consider those $\mathcal{B} \subseteq \wp(\Omega) \setminus \{A\}$ such that $\{I_B - \underline{P}(B) | B \in \mathcal{B}\}$ is linearly independent. As in the proof for avoiding sure loss, we can (after applying the lemma) maintain $r = 1$, as the strict positivity of the λ_B's allows us to renormalize λ if necessary. Lemma 6 – again applied relative to A – then shows that we only need to consider those $\mathcal{B}$ for which $\{I_B | B \in \mathcal{B}\}$ is linearly independent. So for these $\mathcal{B}$ we know that $\sum_{B \in \mathcal{B}} \lambda_B I_B = I_A$ has a unique (rational) solution λ. This case gives rise to the conditions in (b).

[1] This situation can only occur when $\bigcup_{B \in \mathcal{B}} B = \Omega$. The $\lambda \in (\mathbb{R}_{>0})^{\{A\} \cup \mathcal{B}}$ must be such that $\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A = 1$ and therefore $\{I_A, I_\Omega\} \cup \{I_B | B \in \mathcal{B}\}$ is a linearly dependent set. The constraint becomes $\sum_{B \in \mathcal{B}} \lambda_B \underline{P}(B) - \lambda_A \underline{P}(A) \leq 1$.

With some minor renaming, we can use Lemma 5 with $\pi = \underline{P}$ and $r = 1$ to find that we only need to consider those $A \subseteq \Omega$ and $\mathcal{B} \in \wp(\Omega) \setminus \{A\}$ such that $\{I_B - \underline{P}(B) | B \in \mathcal{B}\} \cup \{I_A - \underline{P}(A)\}$ is linearly independent. If,

after applying the lemma and after renormalizing λ, $r = 0$ or $r = -1$, we find ourselves in cases covered above and below. When, after renormalizing, $r = 1$, Lemma 6 shows, again with some minor renaming, that we only need to consider $\{I_A\} \cup \{I_B | B \in \mathcal{B}\}$ that are linearly independent. This implies that $\sum_{B \in \mathcal{B}} \lambda_B I_B - \lambda_A I_A = 1$ has a unique rational solution λ. This case gives rise to the conditions in (c).

[−1] This situation can only occur when $A = \Omega$, a choice which we have already excluded. □

3.4 Permutation invariance

As a last property, we look at (weak) *permutation invariance* (for detailed discussion, see [4]). It is the odd one out; whereas all the previous properties allow for lower probabilities that express a quite broad a range of uncertainty models, permutation invariance restricts them to some very specific ones. We mention it to show how easy it can be to add the constraints for another property.

A lower probability is invariant under permutations of the elements of the possibility space if and only if, for any event A and all events B resulting from some permutation, $\underline{P}(A) = \underline{P}(B)$. Let us give an example for $n = 3$: consider the permutation $(1 \to 3, 2 \to 1, 3 \to 2)$, then $A = \{1, 2\}$ becomes $B = \{1, 3\}$.

We can characterize permutation invariance, avoiding redundant constraints, as follows (see also [19, §4.3.1]):

Theorem 9 (Constraints for permutation invariance) *A lower probability $\underline{P}$ is permutation invariant if and only if for all $k = 1, \ldots, n-1$, any one event A such that $|A| = k$, and for all other events B with $|B| = k$, it holds that $\underline{P}(B) = \underline{P}(A)$.*

4 Results

We are not the first to hunt for extreme lower probabilities. However, as far as we know, we are as of yet the most systematic. ([13] contains a list of results.)

For $n = 4$, Shapley [16] gives a list with 37 of the 41 – he omits the 4 degenerate probabilities – 2-monotone extreme lower probabilities. For $n = 5$, we have found all 117 983 for this case. For $n \leq 4$, we have found the extreme (permutation invariant) k-monotone lower probabilities for all k.

In an example, Maaß [11] mentions the 8 extreme coherent lower probabilities for $n = 3$. We give a graphical representation of them in Figure 2 using their corresponding credal sets. Let us clarify: The set of all probabilities that dominate some lower probability is called its *credal set* (cfr. *core* in game theory [16]). All probabilities can be represented as a point of the $(n-1)$-dimensional unit simplex – a regular triangle for $n = 3$ – and so coherent lower probabilities can be represented by their credal set, which is a convex subset of this unit simplex [18, §3.6.1]. The vertices of the simplex correspond to the degenerate probabilities, so to the states – here a, b, and c.
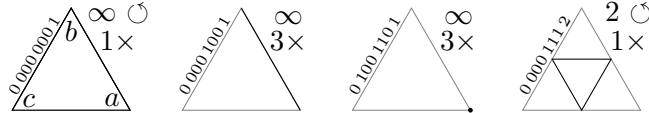

Fig. 2. The credal sets of the 8 extreme coherent lower probabilities for $n = 3$.

In Figure 2, we only give the border of the credal sets and show only one element of each permutation class. At the top right, we indicate the following: the total number of elements of the permutation class (e.g., $4\times$), if it is permutation invariant ($\circlearrowleft$), and k-monotonicity for $k \in \mathbb{N}_{>1}$ (∞ for complete monotonicity, 2 for 2-monotonicity). Along the simplex's left edge, we give a vector that is proportional to the extreme coherent lower probability (component order: $\emptyset$ $\{a\}\{b\}\{c\}$ $\{a,b\}\{a,c\}\{b,c\}$ Ω). Remember that $\underline{P}(\Omega) = 1$.

As convex combinations preserve monotonicity, we can immediately see from Figure 2 that for $n = 3$, all coherent lower probabilities are 2-monotone. This was already known [17, bottom of p. 58], but it illustrates how these computational results can help in finding theoretical results. (For $n = 2$, all are completely monotone.)

Once we implemented our program, finding all 402 extreme coherent lower probabilities for $n = 4$ was easy.[8] Figure 3 shows the corresponding credal sets, as well as those for the 16 extreme 3-monotone lower probabilities, in the same way we did for $n = 3$ in Figure 2. The unit simplex is now a regular tetrahedron, so we had to use a projection from three dimensions to two.

Take $\Omega = \{a, b, c, d\}$. In this case, the component order we use is $\emptyset$ $\{a\}\{b\}\{c\}\{d\}$ $\{a,b\}\{a,c\}\{a,d\}\{b,c\}\{b,d\}\{c,d\}$ $\{a,b,c\}\{a,b,d\}\{a,c,d\}\{b,c,d\}$ Ω.

With Figure 3 as a guide, we can give some observations and results.

Using results from Choquet [3, Ch. 7], it can be proven that the extreme

---

[8] For $n = 5$, we know 1 743 093 of the extreme coherent lower probabilities. These have been found by a computer in our lab, after some months of vertex enumeration. A hardware failure cut this gargantuan task short.
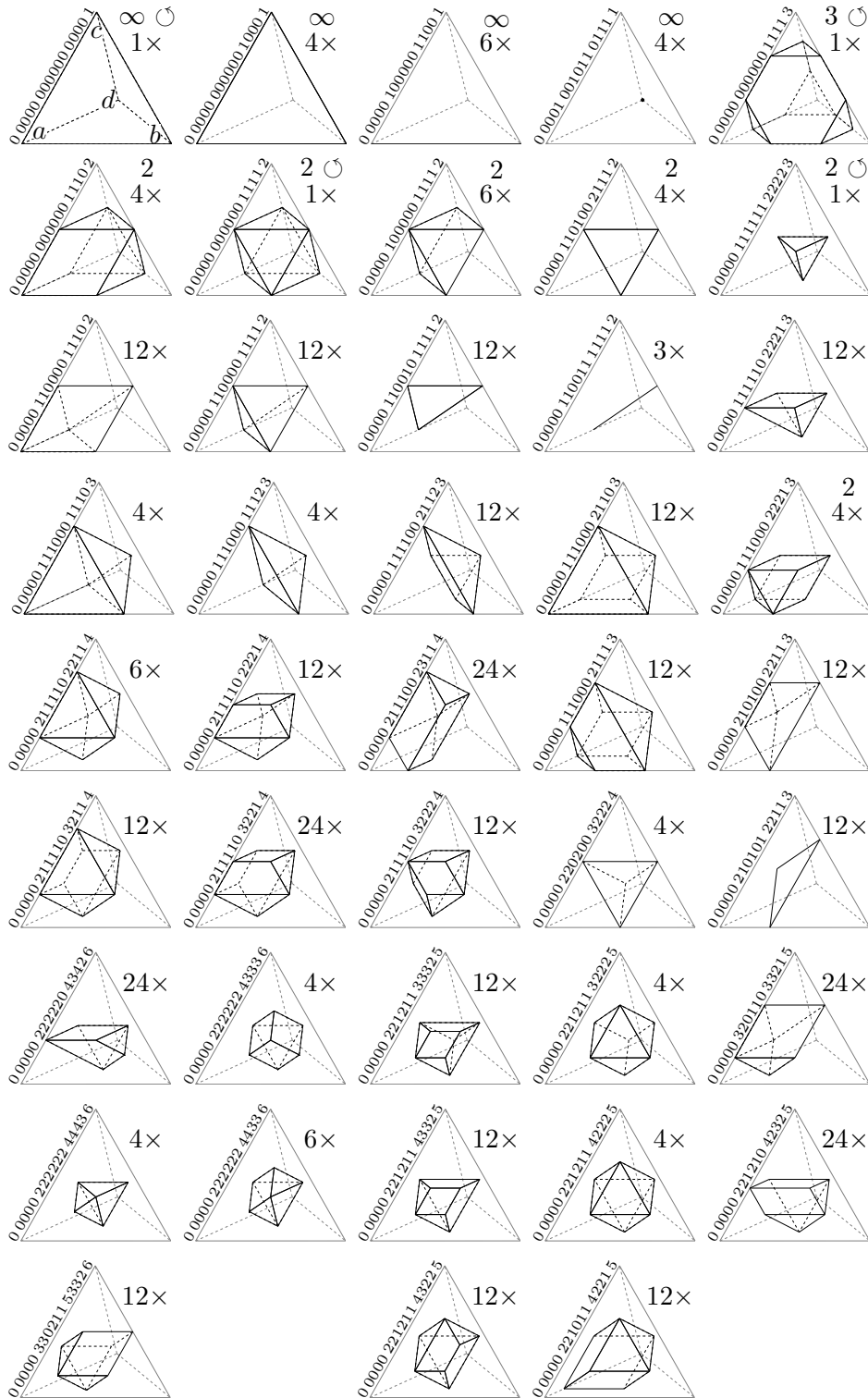
Fig. 3. The credal sets for $n = 4$ of the extreme coherent lower probabilities (all except last of top row) and extreme 3-monotone lower probabilities (top row).

completely monotone lower probabilities are the *vacuous* lower probabilities (cfr. *unanimity games* in game theory) with respect to events: $\underline{P}$ is vacuous with respect to A if $\underline{P}(B) = 1$ for $A \subseteq B$ and 0 otherwise. [9] These correspond to the first three (Figure 2) and first four (Figure 3) permutation classes shown. The last of these classes corresponds to the degenerate probabilities, which are all the extreme (classical) probabilities.

We have observed that the extreme completely monotone probabilities are always included in the extreme coherent probabilities. This is not the case for all of the extreme k-monotone and permutation invariant lower probabilities. An example for $n = 4$ is the only non-completely monotone lower probability of the extreme 3-monotone lower probabilities (shown in Figure 3).

Notice that, except for the degenerate probabilities, all credal sets touch all tetrahedron faces. This is so because it can be shown that the degenerate probabilities are the only extreme coherent lower probabilities that can be nonzero in singletons.

5 Conclusions

Obtaining sets of extreme lower probabilities for many cases and formulating a systematic approach to calculating them are the main results of the research presented in this paper.

Our results also give a partial answer to an open problem formulated by Maaß [11,12]. He showed, in a much more general setting than ours, so, roughly speaking, that lower probabilities satisfying whatever property expressible using linear inequality constraints, can also be expressed as a convex combination of extreme lower probabilities, which are to be determined. His open problem was: what are these extreme lower probabilities for coherence and k-monotonicity? We obtained them, for small possibility spaces.

Even though our initial interest in extreme lower probabilities was kindled by the hope of using them in a more practical application – finding conservative approximations for lower and upper probabilities, to be precise –, at this moment we know of no concrete way to use them in a practical application. The main reason for this is that there are so many of them even for small possibility spaces: For a possibility space of cardinality 5 there are already too many extreme coherent lower previsions to practically enumerate, so the

---

[9] In general (when Ω can be infinite), the extreme completely monotone lower probabilities are those that take the value 1 on a proper filter of sets and are 0 elsewhere.

almost self-evident idea of using them for interpolation purposes also becomes impractical. For approximation purposes, we would need an efficient algorithm to find the most important terms in the (not necessarily unique) decomposition of a lower prevision in terms of the many extreme lower previsions. We are not aware of any such algorithm.

So, at the moment of writing, the relevance of our results is mainly theoretical, although they also have their pedagogical importance: This topic also allows one to become familiar with different ways of looking at lower probabilities and their properties. A lower probability can be seen as a set function satisfying some properties, as a convex combination of some special set functions (extreme lower probabilities), as a point of a convex subset of a vector space, and – for coherent ones – as a credal set. They can be k-monotone, avoid sure loss, be coherent, be permutation invariant, etc.

And last but not least, this topic can lead to beautiful figures.

Acknowledgments

References

[1] D. Avis, `lrs` homepage, `http://cgm.cs.mcgill.ca/~avis/C/lrs.html`; version of June 1, 2005 retrieved (2005).

[2] A. Chateauneuf, J.-Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Math. Social Sci. 17 (1989) 263–283.

[3] G. Choquet, Theory of Capacities, Ann. Inst. Fourier 5 (1954) 131–295.

[4] G. de Cooman, E. Miranda, Symmetry of models versus models of symmetry, in: W. Harper, G. Wheeler (eds.), Probability and Inference: Essays in Honor of Henry E. Keyburg, Jr., King's College Publications, London, 2007, pp. 67–149.

[5] G. de Cooman, M. Troffaes, E. Miranda, n-Monotone lower previsions and lower integrals, in: F. G. Cozman, R. Nau, T. Seidenfeld (eds.), ISIPTA '05: Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, 2005.

[6] G. de Cooman, M. C. M. Troffaes, E. Miranda, n-Monotone lower previsions, Journal of Intelligent and Fuzzy Systems 16 (4) (2005) 253–263.

[7] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.

[8] T. L. Fine, Theories of probability: An examination of foundations, Academic Press, 1973.

[9] K. Fukuda, Frequently Asked Questions in Polyhedral Computation, `http://www.ifor.math.ethz.ch/~fukuda/polyfaq/`; version of June 18, 2004 used (2004).

[10] K. Fukuda, `cdd` and `cddplus` homepage, `http://www.ifor.math.ethz.ch/~fukuda/cdd_home/cdd.html`; version of February 12, 2007 retrieved (2007).

[11] S. Maaß, Continuous Linear Representations of Coherent Lower Previsions, in: J.-M. Bernard, T. Seidenfeld, M. Zaffalon (eds.), ISIPTA '03: Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, 2003.

[12] S. Maaß, Exact functionals, functionals preserving linear inequalities, Lévy's metric, Ph.D. thesis, Universität Bremen (2003).

[13] E. Quaeghebeur, `constraints` homepage, `http://users.UGent.be/~equaeghe/constraints.php` (2007).

[14] R. T. Rockafellar, Convex Analysis, Princeton Univ. Press, 1970.

[15] G. Shafer, A mathematical theory of evidence, Princeton University Press, 1976.

[16] L. S. Shapley, Cores of convex games, Int. J. Game Theory 1 (1971) 11–26.

[17] P. Walley, Coherent lower (and upper) probabilities, Tech. rep., Department of statistics, University of Warwick (1981).

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

[19] K. Weichselberger, Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, vol. I: Intervallwarscheinlichkeit als umfassendes Konzept, Physica-Verlag, 2001.