

2-monotone outer approximations of coherent lower probabilities

Ignacio Montes¹, Enrique Miranda¹ and Paolo Vicig²

¹Department of Statistics and Operations Research, University of Oviedo (Spain), (imontes,mirandaenrique)@uniovi.es

²DEAMS “B. de Finetti”, University of Trieste (Italy), paolo.vicig@deams.units.it

September 12, 2018

Abstract

We investigate the problem of approximating a coherent lower probability on a finite space by a 2-monotone capacity that is at the same time as close as possible while not including additional information. We show that this can be tackled by means of a linear programming problem, and investigate the features of the set of undominated solutions. While our approach is based on a distance proposed by Baroni and Vicig, we also discuss a number of alternatives: quadratic programming, extensions of the total variation distance, and the Weber set from game theory. Finally, we show that our work applies to the more general problem of approximating coherent lower previsions.

Keywords: Coherent lower probabilities, 2-monotonicity, coherent lower previsions, probability intervals, distortion models, total variation distance.

Acknowledgement

*NOTICE: This is the authors' version of a work that was accepted for publication in the International Journal of Approximate Reasoning. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently

published in the International Journal of Approximate Reasoning, vol. 101, October 2018, pages 181–205, doi:10.1016/j.ijar.2018.07.004 © Copyright Elsevier

<https://www.sciencedirect.com/science/article/pii/S0888613X18303001>.

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



1 Introduction

Since they were thoroughly studied by Peter Walley in [29] as an extension of Bruno de Finetti’s work on subjective probability [9], coherent lower previsions have been considered one of the most general approaches to deal with imprecision and uncertainty.

They possess a number of advantages: first of all, they can be represented equivalently as closed and convex sets of probability measures, and as such they can be given an epistemic interpretation as a model for the imprecise knowledge of a probability measure. Moreover, the representation of these sets in terms of lower and upper envelopes allows for a number of mathematical advantages, for instance in the extension of the assessments to a greater domain.

Secondly, they can also be given a clear behavioural interpretation in terms of acceptable betting rates, thus extending de Finetti’s approach to be able to deal with indecision, something that arises frequently in cases where the available knowledge is imprecise.

Thirdly, they include as particular cases most of the models of non-additive measures that have been proposed in the literature, such as probability intervals [7], belief functions [25] or possibility measures [14]. It is thus possible to work with these particular models using all the machinery that has already been developed for coherent lower previsions.

In spite of these advantages, coherent lower previsions (or their restrictions to events, called *coherent lower probabilities*) also have a number of drawbacks that hinder their use in practice: for instance, they have no general easy representation in terms of the extreme points of their associated sets of probabilities, and they sometimes lack some attractive mathematical properties possessed by some more specific models.

One alternative that somewhat solves these issues while being sufficiently general is to work with 2-monotone capacities, which can be easily determined on finite spaces by means of a finite number of extreme points [26] and that still include as particular cases many of the imprecise probability models from the literature [1]. It is therefore interesting to determine if we can approximate a coherent lower probability by a 2-monotone one with a minimal loss of information. This problem was already considered by Bronevich and Augustin in [3], and they gave two algorithms that provide an outer approximation that is optimal in the sense we shall discuss later on.

Besides the above mentioned motivations, we stress that the approximation problem we are concerned with is relevant for practical purposes too. In particular, it is related to the issue of how to exchange information among agents adopting different uncertainty formalisms discussed e.g. in [2]. In fact, it solves this problem in the case of a sending agent, or more generally an uncertainty interchange format common to an open community of agents, operating with coherent imprecise probabilities, and a receiving agent adopting 2-monotone probabilities or some special cases of theirs.

After recalling some preliminary concepts in Section 2, in Section 3 we study the problem of finding undominated outer approximations that minimize the distance to the original model, in the sense proposed by Baroni and Vicig in [2]. In Section 4, we focus on outer approximations by means of some particular subfamilies of 2-monotone capacities and prove that this problem has a unique solution in those cases. A comparison with other approaches is given in Section 5. In Section 6 we show that our results allow to solve also the problem of outer approximating coherent lower previsions. Finally, In Section 7 we take a brief look at the problem of finding *inner* approximations of a given coherent lower probability. Some additional comments are provided in Section 8. In order to ease the reading, proofs, as well as auxiliary lemmas, have been gathered in an appendix.

2 Preliminary concepts

Let \mathcal{X} be a finite space with cardinality $|\mathcal{X}| = n$ and powerset $\mathcal{P}(\mathcal{X})$. A *lower probability* is a function $\underline{P} : \mathcal{K} \subseteq \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ defined on some subsets (events) of \mathcal{X} . Following Walley [29], a lower probability can be given a behavioural interpretation, so that $\underline{P}(A)$ is our supremum acceptable betting rate on event A (that is, the supremum amount of money we would pay for a gamble with reward 1 when A occurs and 0 when it does not). On the

other hand, lower probabilities can also be given an epistemic interpretation, in situations of imprecise knowledge about the probability of some events. Then the lower probability is understood as a lower bound of an ideal, but unknown, probability measure, and our information about this measure can be equivalently represented by means of the set of probability distributions that are compatible with the information given by \underline{P} . This set is called *credal set*, and it is defined by:

$$\mathcal{M}(\underline{P}) = \{P \text{ probability measure} \mid P(A) \geq \underline{P}(A) \quad \forall A \in \mathcal{K}\}. \quad (1)$$

Some usual consistency requirements are imposed to lower probabilities. One of the simplest is *avoiding sure loss*, which means that there is at least one probability compatible with \underline{P} , or equivalently, that $\mathcal{M}(\underline{P}) \neq \emptyset$. A stronger requirement is *coherence*, which means that the bounds \underline{P} gives for the probabilities of the different events are tight:

Definition 1. [29] *A lower probability $\underline{P} : \mathcal{K} \subseteq \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ is called coherent when $\underline{P}(A) = \min\{P(A) : P \in \mathcal{M}(\underline{P})\}$ for every $A \in \mathcal{K}$.*

The conjugate of a lower probability \underline{P} on \mathcal{K} is the function $\overline{P} : \mathcal{K}^c \subseteq \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ given by $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \in \mathcal{K}^c$, where $\mathcal{K}^c = \{A^c \mid A \in \mathcal{K}\}$. \overline{P} is called an *upper probability*, and $\overline{P}(A)$ can be understood as an upper bound for the unknown probability of A , or, under the behavioural interpretation, as the infimum betting rate against A . The conjugacy relation implies that the credal set given in Eq. (1) can be equivalently represented by:

$$\mathcal{M}(\underline{P}) = \{P \text{ probability} \mid P(A) \leq \overline{P}(A) \quad \forall A \in \mathcal{K}^c\}.$$

In addition, \underline{P} is coherent if and only if $\overline{P}(A) = \max_{P \in \mathcal{M}(\underline{P})} P(A)$ for every $A \in \mathcal{K}^c$.

Throughout this paper, we shall consider conjugate and coherent lower and upper probabilities $\underline{P}, \overline{P}$. Moreover, we shall assume that they are defined on the power set $\mathcal{P}(\mathcal{X})$. This assumption entails no loss of generality, since coherent models on a proper subset of $\mathcal{P}(\mathcal{X})$ can always be extended to the power set by means of the notion of *natural extension* (see [29] for more details):

$$\begin{aligned} \underline{P}(A) &= \min\{P(A) : P \in \mathcal{M}(\underline{P})\} \quad \forall A \subseteq \mathcal{X} \\ \overline{P}(A) &= \max\{P(A) : P \in \mathcal{M}(\underline{P})\} \quad \forall A \subseteq \mathcal{X}. \end{aligned}$$

If $\underline{P}, \overline{P}$ are coherent and conjugate on $\mathcal{P}(\mathcal{X})$, they satisfy the following properties [29, Section 2.7.4]:

- (C1) $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ for every $A \subseteq \mathcal{X}$.
- (C2) $\underline{P}(A) \leq \underline{P}(B)$ and $\overline{P}(A) \leq \overline{P}(B)$ for every $A, B \subseteq \mathcal{X}$ such that $A \subseteq B$.
- (C3) $\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B)$ for every $A, B \subseteq \mathcal{X}$.
- (C4) $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$ for every disjoint $A, B \subseteq \mathcal{X}$.
- (C5) $\underline{P}(A \cup B) \leq \underline{P}(A) + \overline{P}(B)$ for every $A, B \subseteq \mathcal{X}$.
- (C6) $\underline{P}(A) + \underline{P}(B) \leq 1 + \underline{P}(A \cap B)$ for every $A, B \subseteq \mathcal{X}$.

One particular case of coherent lower probabilities are the 2-monotone ones.

Definition 2. [5] A coherent lower probability $\underline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ is called 2-monotone if for every $A, B \subseteq \mathcal{X}$ it satisfies:

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \quad (2)$$

We denote by \mathcal{C}_2 the class of (coherent) 2-monotone lower probabilities defined on $\mathcal{P}(\mathcal{X})$.

The conjugate upper probability of a 2-monotone lower probability is called 2-alternating, and it satisfies Eq. (2) with the reverse inequality.

Coherent 2-monotone lower probabilities are also referred to as 2-monotone capacities, as we shall also often do, or sometimes called *convex* in the literature. They can be equivalently represented in terms of their Möbius inverse.

Definition 3. Consider a 2-monotone lower probability \underline{P} . Its Möbius inverse [25] $m_{\underline{P}} : \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R}$ is given by:

$$m_{\underline{P}}(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \underline{P}(B) \quad \forall A \subseteq \mathcal{X};$$

it holds that

$$\underline{P}(A) = \sum_{B \subseteq A} m_{\underline{P}}(B) \quad \forall A \subseteq \mathcal{X}. \quad (3)$$

Conversely [4], a function $m_{\underline{P}} : \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R}$ satisfying $\sum_{A \subseteq \mathcal{X}} m_{\underline{P}}(A) = 1$, $m_{\underline{P}}(\emptyset) = 0$, $m_{\underline{P}}(\{x_i\}) \geq 0$ for every $x_i \in \mathcal{X}$ and

$$\sum_{\{x_i, x_j\} \subseteq B \subseteq A} m_{\underline{P}}(B) \geq 0, \quad \forall A \subseteq \mathcal{X}, \forall x_i, x_j \in A, x_i \neq x_j,$$

determines a 2-monotone capacity by Eq. (3).

2-monotone capacities possess a number of interesting properties that are not satisfied in general by coherent lower probabilities. For instance, the extreme points of the credal set associated with a 2-monotone capacity can be easily determined using the permutations of the possibility space [26]; moreover, they have a unique extension as an expectation operator that preserves 2-monotonicity: their Choquet integral [8]. These properties shall be discussed more deeply in Sections 5.3 and 6, respectively. For more information on 2-monotonicity, we refer to [10, 11].

Taking this into account, and following an approach similar to the one in [3], in this paper we shall investigate how to approximate a coherent lower probability \underline{P} by a 2-monotone capacity \underline{Q} that at the same time (a) introduces no new information; and (b) is as close as possible to the original model.

The first constraint is modelled by requiring that the credal set determined by \underline{Q} includes that of \underline{P} , $\mathcal{M}(\underline{P}) \subseteq \mathcal{M}(\underline{Q})$, or in other words, that

$$\underline{Q}(E) \leq \underline{P}(E) \text{ for every } E \subseteq \mathcal{X}. \quad (4)$$

In that case, we shall say that \underline{Q} is an *outer approximation* of \underline{P} .

With respect to the second constraint, one preliminary idea would be to require \underline{Q} to be *undominated*, in the sense that there is no other 2-monotone capacity \underline{Q}' such that $\mathcal{M}(\underline{P}) \subseteq \mathcal{M}(\underline{Q}') \subsetneq \mathcal{M}(\underline{Q})$. However, this requirement alone does not determine a unique solution, as we will see later on in Example 1, nor does it provide us with a tool to determine any 2-monotone outer approximation, either.

The set of undominated outer approximations was studied by Bronevich and Augustin in [3], under the name *Pareto optimal* approximations. In particular, they showed [3, Lemma 1] that any undominated outer approximation of \underline{P} in \mathcal{C}_2 can be obtained as a convex combination of extreme undominated outer approximations, and provided a theoretical characterization of the undominated outer approximations. However, it is still an open problem to determine whether the set of undominated outer approximations is convex.

In the remainder of this paper, we focus on how to obtain undominated outer approximations of a coherent lower probability in the class \mathcal{C}_2 using different approaches, mainly through linear programming. Also, we study how to outer approximate a coherent lower probability using some subclasses of interest.

3 Approximations by linear programming

One way of determining an optimal outer approximation is to consider those outer approximations \underline{Q} of the coherent lower probability \underline{P} that minimize the distance proposed by Baroni and Vicig in [2], given by

$$d(\underline{P}, \underline{Q}) := \sum_{E \subseteq \mathcal{X}} (\underline{P}(E) - \underline{Q}(E)). \quad (5)$$

If we interpret $\underline{P}(E) - \underline{Q}(E)$ as the additional imprecision introduced on E when replacing $\underline{P}(E)$ with $\underline{Q}(E)$, then $d(\underline{P}, \underline{Q})$ has the meaning of total imprecision added by the outer approximation \underline{Q} . Measuring the additional imprecision by $d(\underline{P}, \underline{Q})$ appears natural, given that the imprecision of an uncertainty evaluation on an event E is usually measured in the same linear scale by $\overline{P}(E) - \underline{P}(E)$ [29, Section 2.7.2] (see also the final part of Section 5.1 for further comments).

To solve the minimization problem, we determine \underline{Q} through its *Möbius inverse* $m_{\underline{Q}}$, by means of Eq. (3), and consider thus the following linear programming problem:

$$\min d(\underline{P}, \underline{Q}) \quad (\text{LP-2monot})$$

subject to:

$$\sum_{E \subseteq \mathcal{X}} m_{\underline{Q}}(E) = 1, \quad m_{\underline{Q}}(\emptyset) = 0. \quad (\text{LP-2monot.1})$$

$$\sum_{\{x_i, x_j\} \subseteq B \subseteq E} m_{\underline{Q}}(B) \geq 0, \quad \forall E \subseteq \mathcal{X}, \forall x_i, x_j \in E, x_i \neq x_j. \quad (\text{LP-2monot.2})$$

$$m_{\underline{Q}}(\{x_i\}) \geq 0, \quad \forall x_i \in \mathcal{X}. \quad (\text{LP-2monot.3})$$

$$\sum_{B \subseteq E} m_{\underline{Q}}(B) \leq \underline{P}(E) \quad \forall E \neq \emptyset, \mathcal{X}. \quad (\text{LP-2monot.4})$$

In fact, (LP-2monot.2) characterizes 2-monotonicity of \underline{Q} via its Möbius inverse $m_{\underline{Q}}$ [4, Corollary 2], while (LP-2monot.3) ensures that \underline{Q} is also 1-monotone and non-negative [4, Proposition 1]; (LP-2monot.1) is only a normalization condition. Finally, (LP-2monot.4) implements condition (4) taking into account that $m_{\underline{Q}}$ determines \underline{Q} by means of Eq. (3)¹.

The feasible region of this linear programming problem is non-empty: it

¹The constraints (LP-2monot.1)–(LP-2monot.3) were also considered in [6, Thm. 1] when studying the problem of the consistency with 2-monotone coherent risk measures.

suffices to take into account that the *vacuous* lower probability

$$\underline{Q}_v(E) = \begin{cases} 0 & \text{if } E \neq \mathcal{X}, \\ 1 & \text{if } E = \mathcal{X}, \end{cases} \quad (6)$$

is a 2-monotone [23, Theorem 1] outer approximation of any coherent lower probability.

Moreover, the linear programming problem above has an optimal solution by means of Weierstrass' theorem. To see this, note that

- (i) $d(\underline{P}, \underline{Q}) = \sum_{E \subseteq \mathcal{X}} (\underline{P}(E) - \sum_{B \subseteq E} m_{\underline{Q}}(B))$ is continuous on the variables $m_{\underline{Q}}(B)$;
- (ii) the feasible region is bounded, since by [15, Theorem 1] the values of $m_{\underline{Q}}$ are bounded when \underline{Q} belongs to \mathcal{C}_2 ; and
- (iii) it is closed, being a polyhedral set in \mathbb{R}^{2^n} .

Operationally, the linear programming problem we are considering involves $2^n + n + 2^{n-2} \binom{n}{2}$ different constraints:

- 2 in (LP-2monot.1);
- $\sum_{\{x_i, x_j\} \subseteq \mathcal{X}: x_i \neq x_j} |\{A \supseteq \{x_i, x_j\}\}| = \sum_{x_i \neq x_j \in \mathcal{X}} 2^{n-2} = 2^{n-2} \binom{n}{2}$ in (LP-2monot.2);
- n in (LP-2monot.3); and
- $2^n - 2$ in (LP-2monot.4).

Thus, the problem becomes computationally hard to solve for large cardinalities of \mathcal{X} . This issue is intrinsic with the nature of imprecise probabilities: due to their non-additivity, fixing the values of \underline{Q} on any $x_i \in \mathcal{X}$ is not enough to determine \underline{Q} on $\mathcal{P}(\mathcal{X})$, and exponentially many further arguments may be needed. This will be one of the motivations behind our study of particular cases in Section 4.

Given this, our first result tells us that any solution of the linear programming problem is undominated:

Proposition 1. *Let \underline{P} be a coherent lower probability, and let \underline{Q} be an optimal solution of the linear programming problem (LP-2monot). Then, \underline{Q} is an undominated outer approximation of \underline{P} in \mathcal{C}_2 .*

Not surprisingly, the linear programming problem (LP-2monot) may not have a unique solution:

Example 1. Consider $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and let \underline{P} be the coherent lower probability that is the lower envelope of the probability measures with mass functions $P_1 = (0.5, 0.5, 0, 0)$, $P_2 = (0, 0, 0.5, 0.5)$. It is given by:

$$\underline{P}(A) = \begin{cases} 0 & \text{if } |A| = 1 \text{ or } A = \{x_1, x_2\}, \{x_3, x_4\} \\ 1 & \text{if } A = \mathcal{X} \\ 0.5 & \text{otherwise.} \end{cases}$$

To see that \underline{P} is not 2-monotone, note that, given $A = \{x_1, x_3\}$ and $B = \{x_2, x_3\}$,

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) = 0.5 < 1 = \underline{P}(A) + \underline{P}(B). \quad (7)$$

To see that (LP-2monot) may have more than one solution, note that, if \underline{Q} is a 2-monotone outer approximation of \underline{P} , it must satisfy

$$\underline{Q}(\{x_1, x_3\}) + \underline{Q}(\{x_2, x_3\}) \leq \underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_3\}).$$

Also, since \underline{Q} is an outer approximation of \underline{P} , it holds that:

$$\underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_3\}) \leq \underline{P}(\{x_1, x_2, x_3\}) + \underline{P}(\{x_3\}) = 0.5.$$

Taking Eq. (7) into account, the last equation can be written as:

$$\underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_3\}) \leq 0.5 = \underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_2, x_3\}) - 0.5.$$

Hence, any 2-monotone outer approximation \underline{Q} must satisfy

$$\underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_2, x_3\}) - \underline{Q}(\{x_1, x_2, x_3\}) - \underline{Q}(\{x_3\}) \geq 0.5,$$

whence

$$\underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_2, x_3\}) - \underline{Q}(\{x_1, x_3\}) - \underline{Q}(\{x_2, x_3\}) \geq 0.5. \quad (8)$$

Following a similar reasoning, \underline{Q} must also satisfy:

$$\underline{P}(\{x_1, x_4\}) + \underline{P}(\{x_2, x_4\}) - \underline{Q}(\{x_1, x_2, x_4\}) - \underline{Q}(\{x_4\}) \geq 0.5,$$

and therefore

$$\underline{P}(\{x_1, x_4\}) + \underline{P}(\{x_2, x_4\}) - \underline{Q}(\{x_1, x_4\}) - \underline{Q}(\{x_2, x_4\}) \geq 0.5. \quad (9)$$

If we now combine Eqs. (8) and (9) we deduce that $d(\underline{P}, \underline{Q}) \geq 1$ for any 2-monotone outer approximation of \underline{P} . This distance is attained, amongst others, by the 2-monotone capacities $\underline{Q}_1, \underline{Q}_2$ given by

$$\underline{Q}_1(A) = \begin{cases} 0 & \text{if } |A| = 1 \text{ or } A = \{x_1, x_2\}, \{x_3, x_4\} \\ 0.5 & \text{if } |A| = 3 \\ 1 & \text{if } A = \mathcal{X} \\ 0.25 & \text{otherwise,} \end{cases}$$

and

$$\underline{Q}_2(A) = \begin{cases} 0 & \text{if } |A| = 1 \text{ or } A = \{x_1, x_2\}, \{x_3, x_4\} \\ 0.5 & \text{if } |A| = 3 \\ 1 & \text{if } A = \mathcal{X} \\ 0.2 & \text{if } A = \{x_1, x_4\}, \{x_2, x_3\} \\ 0.3 & \text{otherwise.} \end{cases}$$

Their 2-monotonicity can easily be verified by means of Eq. (2). \blacklozenge

Since the feasible region of (LP-2monot) is bounded, the set of its optimal solutions is convex. Thus, every convex combination of two undominated optimal solutions of (LP-2monot) is still an undominated 2-monotone outer approximation. Note that this does not imply that the set of *all* undominated outer approximations is convex: whether this is true or not is still an open problem at present.

Obviously, if our initial model \underline{P} is not 2-monotone, any undominated 2-monotone capacity that outer approximates \underline{P} will not agree with \underline{P} on some event A . Interestingly, both models always agree on singletons, and the same applies to their conjugate upper probabilities:

Proposition 2. *Let \underline{P} be a coherent lower probability with conjugate upper probability \overline{P} . If \underline{Q} is an undominated 2-monotone capacity that outer approximates \underline{P} , and \overline{Q} denotes its conjugate upper probability, then $\underline{Q}(\{x\}) = \underline{P}(\{x\})$ and $\overline{Q}(\{x\}) = \overline{P}(\{x\})$ for every $x \in \mathcal{X}$. As a consequence, $\underline{Q}(A) = \underline{P}(A)$ and $\overline{Q}(A) = \overline{P}(A)$ for every event A with $|A| = n - 1$.*

As a consequence, both $\underline{P}, \overline{P}$ and its undominated outer approximation $\underline{Q}, \overline{Q}$ induce the same order on \mathcal{X} . This property may not extend to subfamilies of 2-monotone capacities: it does, for the probability intervals studied in Section 4.1; however, it is only partially satisfied by the ε -contamination model studied in Section 4.3, where the property is only satisfied by \underline{P} ,

and by the pari mutuel model we shall consider in Section 4.2, where the property is only satisfied by \bar{P} . More generally, the property ensured by Proposition 2 is a special case of the *(strict) preference preservation principle*, requiring the approximating uncertainty measure to preserve (strictly or at least weakly) the ordering induced on $\mathcal{P}(\mathcal{X})$ by the initial measure [2]. Preference preservation is known to be hard to apply and conflicting with other approximation criteria, in general. It is therefore notable that Proposition 2 guarantees preference preservation, at least on singletons and their complements, without ever requiring it a priori.

In general, it may be that no solution of the linear programming problem (LP-2monot) agrees with \underline{P} on a fixed event A , as we shall see in Example 3 later on. We may overcome this issue simply by adding a constraint to the linear programming problem (LP-2monot):

$$\min d(\underline{P}, \underline{Q}) \tag{LP-2monot-A}$$

subject to:

$$\sum_{E \subseteq \mathcal{X}} m_{\underline{Q}}(E) = 1, \quad m_{\underline{Q}}(\emptyset) = 0. \tag{LP-2monot.1}$$

$$\sum_{\{x_i, x_j\} \subseteq B \subseteq E} m_{\underline{Q}}(B) \geq 0, \quad \forall E \subseteq \mathcal{X}, \forall x_i, x_j \in E, x_i \neq x_j. \tag{LP-2monot.2}$$

$$m_{\underline{Q}}(\{x_i\}) \geq 0, \quad \forall x_i \in \mathcal{X}. \tag{LP-2monot.3}$$

$$\sum_{B \subseteq E} m_{\underline{Q}}(B) \leq \underline{P}(E) \quad \forall E \neq \emptyset, \mathcal{X}. \tag{LP-2monot.4}$$

$$\sum_{B \subseteq A} m_{\underline{Q}}(B) = \underline{P}(A). \tag{LP-2monot-A.5}$$

The only difference with respect to (LP-2monot) is the additional condition (LP-2monot-A.5), that guarantees that the lower probability \underline{Q} associated with $m_{\underline{Q}}$ satisfies $\underline{Q}(A) = \underline{P}(A)$.

The feasible region of the linear programming problem (LP-2monot-A) is not empty either, because it includes the *simple support function* on A , which is the 2-monotone lower probability \underline{Q} given by

$$\underline{Q}(E) = \begin{cases} \underline{P}(A) & \text{if } A \subseteq E \subsetneq \mathcal{X} \\ 1 & \text{if } A = \mathcal{X} \\ 0 & \text{otherwise.} \end{cases}$$

To see this, note that the mass function of \underline{Q} is $m_{\underline{Q}}(A) = \underline{P}(A)$, $m_{\underline{Q}}(\mathcal{X}) = 1 - \underline{P}(A)$, $m_{\underline{Q}}(E) = 0$ otherwise, and that it satisfies all the constraints

in the linear programming problem above. Moreover, the same reasoning as in (LP-2monot) allows us to conclude that (LP-2monot- A) has optimal solutions. This immediately lets us deduce the following:

Proposition 3. *Let $\underline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ be a coherent lower probability, and, for each $A \subseteq \mathcal{X}$ denote by \mathcal{M}_A the set of optimal solutions of the linear programming problem (LP-2monot- A). Then*

$$\underline{P} = \max\{\underline{Q} : \underline{Q} \in \cup_{A \subseteq \mathcal{X}} \mathcal{M}_A\}.$$

In other words, the solutions of the linear programming problems allow us to recover the initial coherent lower probability. In fact, a similar result can be derived just taking into account that the feasible region of the linear programming problem (LP-2monot- A) is non-empty for every event $A \subseteq \mathcal{X}$: the simple support functions mentioned above also allow us to recover the initial coherent lower probability.

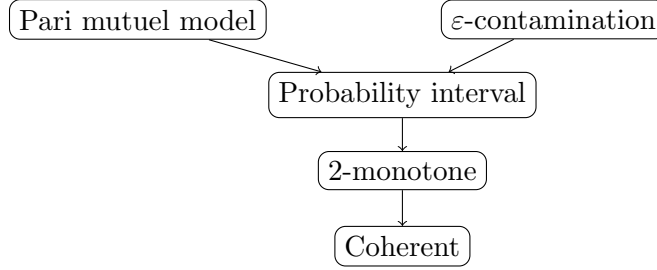
Next, we prove that any optimal solution of (LP-2monot- A) is also an undominated 2-monotone outer approximation of \underline{P} .

Proposition 4. *Let \underline{P} be a coherent lower probability, and let \underline{Q} be an optimal solution of the linear programming problem (LP-2monot- A). Then, \underline{Q} is an undominated outer approximation of \underline{P} in \mathcal{C}_2 and satisfies $\underline{Q}(A) = \underline{P}(A)$.*

At this point, one may think that all the undominated outer approximations of \underline{P} in \mathcal{C}_2 can be obtained as optimal solutions of either (LP-2monot) or (LP-2monot- A) for some A . However, we shall show in Section 5.1 that there are undominated outer approximations that are not optimal solutions of any of the linear programming problems.

4 Particular cases

In this section, we investigate the outer approximations of a coherent lower probability in some subfamilies of 2-monotone capacities: those associated with probability intervals and certain distortion models. With the term *distortion model* we refer to a model where an initial probability measure P_0 is modified in some sense. We summarize the relationship between the different models considered in this paper in the following figure, where an arrow between two nodes means that the model represented by the parent is a particular case of its descendant:



4.1 Probability intervals

We begin by considering probability intervals, which are imprecise probability models where only the lower and upper probabilities of the singletons are specified [7]. Thus, a *probability interval* is an n -tuple of intervals:

$$\mathcal{I} = \{[l_i, u_i] \mid i = 1, \dots, n\},$$

where $[l_i, u_i]$ is understood as the interval of possible values for the probability of x_i . This n -tuple determines a credal set by:

$$\mathcal{M}(\mathcal{I}) = \{P \text{ probability} \mid l_i \leq P(\{x_i\}) \leq u_i, i = 1, \dots, n\}.$$

This set contains all the probability measures that comply with the constraints on singletons given in \mathcal{I} . By taking lower and upper envelopes, this credal set determines a lower and an upper probability:

$$l(A) = \inf_{P \in \mathcal{M}(\mathcal{I})} P(A) \quad \text{and} \quad u(A) = \sup_{P \in \mathcal{M}(\mathcal{I})} P(A) \quad \forall A \subseteq \mathcal{X}, \quad (10)$$

and \mathcal{I} is said *coherent* when $l(\{x_i\}) = l_i$ and $u(\{x_i\}) = u_i$ for every $i = 1, \dots, n$. A necessary and sufficient condition for coherence is that

$$\sum_{j \neq i} l_j + u_i \leq 1 \quad \text{and} \quad \sum_{j \neq i} u_j + l_i \geq 1 \quad \forall i = 1, \dots, n.$$

In that case, $l(A), u(A)$ can be computed for every $A \subseteq \mathcal{X}$ by:

$$l(A) = \max \left\{ \sum_{x_i \in A} l_i, 1 - \sum_{x_i \notin A} u_i \right\} \quad \text{and} \quad u(A) = \min \left\{ \sum_{x_i \in A} u_i, 1 - \sum_{x_i \notin A} l_i \right\}. \quad (11)$$

Moreover, when a probability interval is coherent, its lower probability l is 2-monotone [7, Proposition 5].

Let \mathcal{C}_{PI} denote the set of all the probability intervals on \mathcal{X} . It is not hard to prove that for \mathcal{C}_{PI} , there is a unique undominated outer approximation:

Proposition 5. Let \underline{P} be a coherent lower probability on $\mathcal{P}(\mathcal{X})$ with conjugate upper probability \overline{P} . Define the probability interval \mathcal{I} by:

$$\mathcal{I} = \{[l_i, u_i] = [\underline{P}(\{x_i\}), \overline{P}(\{x_i\})] \mid i = 1, \dots, n\}, \quad (12)$$

and denote by l, u the lower and upper probability it induces by means of Eq. (10). Then, \mathcal{I} is a coherent probability interval and l is the unique undominated outer approximation of \underline{P} associated with \mathcal{C}_{PI} .

We can deduce from this result that the undominated outer approximation of \underline{P} associated with \mathcal{C}_{PI} also preserves the order that \underline{P} establishes on singletons.

4.2 Pari mutuel models

We next consider the *pari mutuel model* [19, 24, 29] (PMM, for short). This is a betting scheme originated in horse racing. It is determined by two elements: a probability measure P_0 and a distortion factor $\delta > 0$. For every event A of $\mathcal{P}(\mathcal{X})$, $P_0(A)$ is interpreted as a fair prize for a bet on A , and $\delta > 0$ denotes the loading of the house. They determine a coherent lower probability and its conjugate coherent upper probability by:

$$\underline{P}_\delta(A) = \max\{0, (1+\delta)P_0(A) - \delta\} \quad \text{and} \quad \overline{P}_\delta(A) = \min\{1, (1+\delta)P_0(A)\} \quad (13)$$

for every $A \subseteq \mathcal{X}$. The associated credal set is given by:

$$\mathcal{M}(P_0, \delta) = \{P \text{ probability} \mid \max\{0, (1+\delta)P_0(A) - \delta\} \leq P(A) \forall A \subseteq \mathcal{X}\}.$$

The following result from [19, Lemma 4] shall be useful later on.

Lemma 6. Let \overline{P}_δ be the upper probability induced by a PMM (P_0, δ) by means of Eq. (13). For every $A \subseteq \mathcal{X}$ such that $\overline{P}_\delta(A) < 1$, it holds that $\overline{P}_\delta(A) = \sum_{x \in A} \overline{P}_\delta(\{x\})$.

Thus, although a coherent upper probability is in general not additive, but only subadditive (C3), in the case of the PMM additivity holds on some subset of the domain.

The lower probability associated with a PMM is 2-monotone, as shown for instance in [24, Section 2]. Moreover, in [19] it is proven that PMMs correspond to particular instances of probability intervals.

We denote by \mathcal{C}_{PMM} all the pari mutuel models defined on $\mathcal{P}(\mathcal{X})$. Our next result gives the unique undominated outer approximation of a coherent lower probability in \mathcal{C}_{PMM} .

Proposition 7. Let \underline{P} be a coherent lower probability with conjugate upper probability \overline{P} . Define the constant value $\delta > 0$ and the probability P_0 by:

$$\delta = \sum_{i=1}^n \overline{P}(\{x_i\}) - 1, \quad P_0(\{x_i\}) = \frac{\overline{P}(\{x_i\})}{1 + \delta} \quad \forall i = 1, \dots, n.$$

Denote by \underline{P}_δ the coherent lower probability associated with the PMM (P_0, δ) by means of Eq. (13). Then, \underline{P}_δ is the unique outer approximation of \underline{P} in \mathcal{C}_{PMM} .

If we apply this result to the coherent lower probability from Example 1, we obtain that the unique undominated pari mutuel model that outer approximates \underline{P} is given by:

$$\underline{P}_\delta(A) = \begin{cases} 0 & \text{if } |A| = 1, 2 \\ 0.5 & \text{if } |A| = 3 \\ 1 & \text{if } |A| = \mathcal{X}; \end{cases}$$

in fact, this is the same optimal outer approximation we obtain in the family of probability intervals. However, the solutions in both subfamilies need not coincide in general: that obtained with probability intervals may be more informative than the one we obtain in the family of pari mutuel models, as we shall show in Example 2.

In contrast with the outer approximations in terms of 2-monotone lower probabilities and probability intervals, the outer approximations in terms of pari mutuel models may not preserve the order among singletons given by \underline{P} , as also Example 2 shall show. However, the upper probability associated with the outer approximation in Proposition 7 does coincide with the original upper probability \overline{P} on singletons, i.e. it holds that $\overline{P}_\delta(\{x_i\}) = \overline{P}(\{x_i\})$ for every $i = 1, \dots, n$. This is shown in the second part of the proof of Proposition 7 (Eq. (25)).

4.3 ϵ -contamination models

Another relevant distortion model is the ϵ -contamination model, also called *linear-vacuous mixture* in [29]. Given a probability measure P_0 and $\epsilon \in (0, 1)$, they determine the coherent lower probability

$$\underline{P}_\epsilon(A) = \begin{cases} (1 - \epsilon)P_0(A) & \text{if } A \neq \mathcal{X}. \\ 1 & \text{if } A = \mathcal{X}. \end{cases} \quad (14)$$

Equivalently, $\underline{P}_\varepsilon = (1-\varepsilon)P_0 + \varepsilon\underline{Q}_v$, where \underline{Q}_v is the vacuous lower probability defined in Eq. (6). The vacuous lower probability models the total absence of information, and the credal set $\mathcal{M}(\underline{Q}_v)$ is made up of all the probabilities in $\mathcal{P}(\mathcal{X})$. Also, the conjugate upper probability of $\underline{P}_\varepsilon$ is given by:

$$\overline{P}_\varepsilon(A) = \begin{cases} (1-\varepsilon)P_0(A) + \varepsilon & \text{if } A \neq \emptyset. \\ 0 & \text{if } A = \emptyset. \end{cases}$$

The lower probability induced by such a model is 2-monotone. This follows from the fact that it satisfies an even stronger property: complete monotonicity, as can be deduced for instance from [8, Theorems 5 and 11] and [16, Chapter 10.2].

Let \mathcal{C}_ε denote the set of all ε -contamination models in $\mathcal{P}(\mathcal{X})$. As with the PMM, there is only one undominated outer approximation for a coherent lower probability in \mathcal{C}_ε .

Proposition 8. *Let \underline{P} be a coherent lower probability satisfying the condition $\sum_{j=1}^n \underline{P}(\{x_j\}) > 0$. Define $\varepsilon \in (0, 1)$ and the probability P_0 by:*

$$\varepsilon = 1 - \sum_{j=1}^n \underline{P}(\{x_j\}), \quad P_0(\{x_i\}) = \frac{\underline{P}(\{x_i\})}{\sum_{j=1}^n \underline{P}(\{x_j\})} \quad \forall i = 1, \dots, n.$$

Denote by $\underline{P}_\varepsilon$ the ε -contamination model they determine by means of Eq. (14). Then, $\underline{P}_\varepsilon$ is the unique undominated outer approximation of \underline{P} in \mathcal{C}_ε .

Note that the assumption $\sum_{j=1}^n \underline{P}(\{x_j\}) > 0$ in this proposition is necessary for the existence of some outer approximation: if $\underline{P}(\{x_j\}) = 0$ for every $x_j \in \mathcal{X}$, any ε -contamination model that outer approximates $\underline{P}_\varepsilon$ should also satisfy $\underline{P}_\varepsilon(\{x_j\}) = 0$ for every $x_j \in \mathcal{X}$:

$$0 = \underline{P}_\varepsilon(\{x_j\}) = (1-\varepsilon)P_0(\{x_j\}) \quad \forall x_j \in \mathcal{X},$$

where P_0 is the precise probability in the ε -contamination model. However, since $\varepsilon \in (0, 1)$, it follows that $P_0(\{x_j\}) = 0$ for every $x_j \in \mathcal{X}$ and using additivity we obtain $P_0(\mathcal{X}) = \sum_{j=1}^n P_0(\{x_j\}) = 0$, a contradiction. This is what happens if we try to apply our result to the coherent lower probability from Example 1.

Moreover, from the above definition of P_0 and ε we deduce that $\underline{P}_\varepsilon$ coincides with \underline{P} on singletons:

$$\underline{P}_\varepsilon(\{x_i\}) = (1-\varepsilon)P_0(\{x_i\}) = \sum_{j=1}^n \underline{P}(\{x_j\}) \frac{\underline{P}(\{x_i\})}{\sum_{j=1}^n \underline{P}(\{x_j\})} = \underline{P}(\{x_i\}).$$

On the other hand, this is not the case for the upper probability: we obtain

$$\overline{P}_\varepsilon(\{x_i\}) = 1 - \underline{P}_\varepsilon(\mathcal{X} \setminus \{x_i\}) = 1 - \sum_{j \neq i} \underline{P}(\{x_j\}) \geq 1 - \underline{P}(\mathcal{X} \setminus \{x_i\}) = \overline{P}(\{x_i\}),$$

where the inequality follows from the super-additivity (C4) of \underline{P} and can be strict, as we shall show in Example 2. Thus, the optimal outer approximation of a coherent upper probability by an ε -contamination model will be in general more imprecise than the one in the family of probability intervals, given by Proposition 5.

To conclude this section, we show that the optimal outer approximations of a coherent lower probability in the family of PMM or ε -contamination models may indeed be more imprecise than the one in the family of probability intervals, and moreover that there is no dominance relationship between them:

Example 2. Let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and consider the coherent lower probability \underline{P} on $\mathcal{P}(\mathcal{X})$ given by

$$\underline{P}(A) = \begin{cases} 0.1 & \text{if } |A| = 1 \\ 0.4 & \text{if } |A| = 2 \\ 0.6 & \text{if } |A| = 3 \\ 1 & \text{if } |A| = 4. \end{cases}$$

To see that it is coherent, note that it can be obtained as the lower envelope of the probability measures associated with the mass functions

$$(0.2, 0.2, 0.2, 0.4), (0.2, 0.2, 0.4, 0.2), (0.2, 0.4, 0.2, 0.2), (0.4, 0.2, 0.2, 0.2), \\ (0.1, 0.3, 0.3, 0.3), (0.3, 0.1, 0.3, 0.3), (0.3, 0.3, 0.1, 0.3), (0.3, 0.3, 0.3, 0.1).$$

Moreover, it is not 2-monotone, since

$$\underline{P}(\{x_1, x_2, x_3\}) + \underline{P}(\{x_1\}) = 0.7 < 0.8 = \underline{P}(\{x_1, x_2\}) + \underline{P}(\{x_1, x_3\}).$$

From Proposition 5, the optimal outer approximation of \underline{P} in the family of probability intervals is given by

$$P(\{x_i\}) \in [\underline{P}(\{x_i\}), \overline{P}(\{x_i\})] = [0.1, 0.4] \quad \forall i = 1, \dots, 4.$$

On the other hand, Proposition 7 shows that the optimal outer approximation of \underline{P} by a *pari mutuel* model is given by (P_0, δ) , where P_0 is the uniform

probability distribution and $\delta = \sum_{i=1}^4 \bar{P}(\{x_i\}) - 1 = 0.6$. This gives

$$\underline{P}_\delta(A) = \begin{cases} 0 & \text{if } |A| = 1 \\ 0.2 & \text{if } |A| = 2 \\ 0.6 & \text{if } |A| = 3 \\ 1 & \text{if } |A| = 4. \end{cases}$$

Finally, from Proposition 8 the optimal outer approximation in the family of ϵ -contamination models is given by P_0 uniform and $\epsilon = 0.6$. This produces

$$\underline{P}_\epsilon(A) = \begin{cases} 0.1 & \text{if } |A| = 1 \\ 0.2 & \text{if } |A| = 2 \\ 0.3 & \text{if } |A| = 3 \\ 1 & \text{if } |A| = 4. \end{cases}$$

Thus, both \underline{P}_δ and \underline{P}_ϵ are more imprecise than the optimal outer approximation by a probability interval, and there is no dominance relation between them. \blacklozenge

5 Comparison with other approaches

In this section, we explore other alternatives to the linear programming approach we have considered so far, in order to justify better our choice. We first look for undominated 2-monotone outer approximations that can be obtained solving a quadratic problem. Next, we discuss the use of extensions of the total variation distance to determine the outer approximations, and finally we investigate whether the Weber set from coalitional game theory can be used to obtain undominated 2-monotone outer approximations.

5.1 Quadratic problems

As Example 1 shows, the linear programming problem (LP-2monot) may not have a unique solution. One way to overcome this issue is to consider, instead of the distance given by Eq. (5), the quadratic distance given by:

$$\tilde{d}(\underline{P}, \underline{Q}) := \sum_{E \subseteq \mathcal{X}} (\underline{P}(E) - \underline{Q}(E))^2.$$

This produces the following optimization problem:

$$\min \sum_{E \subseteq \mathcal{X}} (\underline{Q}(E)^2 - 2\underline{P}(E)\underline{Q}(E)) \quad (\text{QP-2monot})$$

subject to

$$\sum_{E \subseteq \mathcal{X}} m_{\underline{Q}}(E) = 1, \quad m_{\underline{Q}}(\emptyset) = 0. \quad (\text{LP-2monot.1})$$

$$\sum_{\{x_i, x_j\} \subseteq B \subseteq E} m_{\underline{Q}}(B) \geq 0, \quad \forall E \subseteq \mathcal{X}, \forall x_i, x_j \in E, x_i \neq x_j. \quad (\text{LP-2monot.2})$$

$$m_{\underline{Q}}(\{x_i\}) \geq 0, \quad \forall x_i \in \mathcal{X}. \quad (\text{LP-2monot.3})$$

$$\sum_{B \subseteq E} m_{\underline{Q}}(B) \leq \underline{P}(E) \quad \forall E \neq \emptyset, \mathcal{X}. \quad (\text{LP-2monot.4})$$

The quadratic programming problem in (QP-2monot) can be expressed in the usual matrix form:

$$\min \frac{1}{2} \underline{\vec{Q}}^t H \underline{\vec{Q}} + c^t \underline{\vec{Q}},$$

where $\underline{\vec{Q}}$ denotes a vector with the values of \underline{Q} , $H = 2I_{2^n-1}$ is twice the identity matrix of size $2^n - 1$ and $c = -2\underline{\vec{P}}$, where $\underline{\vec{P}}$ is used to denote the vector with the values of \underline{P} . Since H is a positive definite matrix, it follows that the quadratic programming problem (QP-2monot) subject to restrictions (LP-2monot.1)–(LP-2monot.4) has a unique optimal solution \underline{Q} .

From this it ensues, taking into account Lemma 14 in the Appendix, that the solution \underline{Q} is an undominated outer approximation of \underline{P} . Note that this outer approximation need not be one of the solutions of the linear programming problem (LP-2monot), as the next example shows.

Example 3. Consider a four element space \mathcal{X} and a lower probability \underline{P}

whose values are given in the next table:

A	$\underline{P}(A)$	\underline{Q}	\underline{Q}_1	\underline{Q}_2	\underline{Q}_3
$\{x_1\}$	0.1	0.1	0.1	0.1	0.1
$\{x_2\}$	0	0	0	0	0
$\{x_3\}$	0	0	0	0	0
$\{x_4\}$	0.3	0.3	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0.1	0.1	0.1	0.1
$\{x_1, x_3\}$	0.3	0.26	0.3	0.2	0.2
$\{x_1, x_4\}$	0.6	0.54	0.5	0.6	0.6
$\{x_2, x_3\}$	0.3	0.24	0.2	0.2	0.3
$\{x_2, x_4\}$	0.4	0.36	0.4	0.3	0.3
$\{x_3, x_4\}$	0.4	0.36	0.4	0.4	0.3
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5	0.5	0.5
$\{x_1, x_2, x_4\}$	0.6	0.6	0.6	0.6	0.6
$\{x_1, x_3, x_4\}$	0.7	0.7	0.7	0.7	0.7
$\{x_2, x_3, x_4\}$	0.6	0.6	0.6	0.6	0.6
\mathcal{X}	1	1	1	1	1

Note that \underline{P} is a coherent lower probability because it is the lower envelope of the probability measures with mass functions

$$(0.1, 0, 0.4, 0.5), (0.4, 0.1, 0.2, 0.3), (0.3, 0.3, 0, 0.4).$$

Moreover, \underline{P} is not 2-monotone, because taking for example $A = \{x_1, x_3\}$ and $B = \{x_1, x_4\}$, it holds that:

$$\underline{P}(A \cap B) + \underline{P}(A \cup B) = 0.1 + 0.7 < 0.3 + 0.6 = \underline{P}(A) + \underline{P}(B).$$

Solving the linear programming problem (LP-2monot), we obtain a unique solution, the 2-monotone lower probability \underline{Q}_1 . Note that this 2-monotone lower probability satisfies $\underline{Q}_1(\{x_1, x_4\}) < \underline{P}(\{x_1, x_4\})$ and $\underline{Q}_1(\{x_2, x_3\}) < \underline{P}(\{x_2, x_3\})$. Solving the linear programming problem (LP-2monot-A) with $A = \{x_1, x_4\}$ we obtain as optimal solutions $\underline{Q}_2, \underline{Q}_3$, as well as their convex combinations; taking $A = \{x_2, x_3\}$, we obtain a unique optimal solution, again \underline{Q}_3 .

On the other hand, the quadratic programming problem (QP-2monot) has the unique solution \underline{Q} . This solution is not dominated by \underline{Q}_1 , as we can see on the event $A = \{x_1, x_4\}$; moreover, it is not dominated by any solution of (LP-2monot-A) for $A = \{x_1, x_4\}$ or $A = \{x_2, x_3\}$ because the optimal distance (5) to \underline{P} is 0.3, while the distance (5) between \underline{Q} and \underline{P} is 0.24.

We therefore conclude that the optimal solution of the quadratic programming problem is not an optimal solution of the linear programming problems (LP-2monot) and (LP-2monot-A). ♦

Although this example shows that quadratic programming may reach undominated outer approximations that cannot be obtained via linear programming, this is not always the case. For instance, it may be checked that the quadratic programming solution for approximating \underline{P} in Example 1 is \underline{Q}_1 , one of the linear programming solutions. Note further that $d(\underline{P}, \underline{Q}_1) = 1 > \tilde{d}(\underline{P}, \underline{Q}_1) = \frac{1}{4}$. This fact may be generalized: being applied to numbers in $[0, 1]$, \tilde{d} tends to be smaller than d , i.e. the same outer approximation is perceived as closer with \tilde{d} .

In our view, however, the interpretation of the quadratic solution is less clear than those of the linear programming problem (LP-2monot). In particular, if the imprecision of an event E is measured [29, Section 2.7.2] by $\overline{P}(E) - \underline{P}(E)$, \tilde{d} involves a different scaling for quantifying the additional imprecision introduced by \underline{Q} (and its conjugate \overline{Q}), unlike d . This can be seen from the simple relation among final, additional and initial imprecision of the evaluation on E that arises from the linear scaling guaranteed by d :

$$\overline{Q}(E) - \underline{Q}(E) = (\overline{Q}(E) - \overline{P}(E)) + (\overline{P}(E) - \underline{P}(E)) + (\underline{P}(E) - \underline{Q}(E)).$$

In words, the final imprecision $\overline{Q}(E) - \underline{Q}(E)$ is the sum of the initial one and of the additional imprecision due to \underline{Q} and its conjugate \overline{Q} . This relation does not hold using \tilde{d} , nor generally with further alternative distances.

5.2 The total variation distance

Another possibility for defining a distance between lower probabilities would be to consider an extension of the *total variation distance* [18, Chapter 4.1] to the imprecise case. Recall that given two probability measures P_1 and P_2 , their *total variation* is defined as

$$\|P_1 - P_2\| = \max_{E \subseteq \mathcal{X}} |P_1(E) - P_2(E)|.$$

This definition can be equivalently expressed as [18, Prop. 4.2]:

$$\|P_1 - P_2\| = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_1(\{x\}) - P_2(\{x\})|.$$

In an imprecise framework, given two coherent lower probabilities $\underline{P}_1, \underline{P}_2$, we can extend the definition above in a number of (not necessarily equivalent) ways:

$$d_1(\underline{P}_1, \underline{P}_2) = \max_{E \subseteq \mathcal{X}} |\underline{P}_1(E) - \underline{P}_2(E)|, \quad (15)$$

$$d_2(\underline{P}_1, \underline{P}_2) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\underline{P}_1(\{x\}) - \underline{P}_2(\{x\})|, \quad (16)$$

$$d_3(\underline{P}_1, \underline{P}_2) = \sup_{P_1 \in \mathcal{M}(\underline{P}_1), P_2 \in \mathcal{M}(\underline{P}_2)} \|P_1 - P_2\|, \quad (17)$$

and we refer to [1, Section 11.4] for some comments on d_1 in the context of imprecise Markov chains. It is not difficult to show that d_3 is always greater than or equal to d_1 .

Proposition 9. *For every two coherent lower probabilities \underline{P}_1 and \underline{P}_2 it holds that $d_3(\underline{P}_1, \underline{P}_2) \geq d_1(\underline{P}_1, \underline{P}_2)$.*

One may wonder whether there are some other connections among d_1, d_2 and d_3 . The next examples show that there is no other dominance relationship among them.

Example 4. *Consider a three element space $\mathcal{X} = \{x_1, x_2, x_3\}$ and the coherent lower probabilities $\underline{P}_1, \underline{P}_2, \underline{P}_3$ given by:*

	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	\mathcal{X}
\underline{P}_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	1
\underline{P}_2	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
\underline{P}_3	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

It follows from Eq. (15) that $d_1(\underline{P}_1, \underline{P}_2) = \frac{1}{3}$ and $d_1(\underline{P}_2, \underline{P}_3) = \frac{1}{6}$, while Eq. (16) implies that $d_2(\underline{P}_1, \underline{P}_2) = 0.5$ and $d_2(\underline{P}_2, \underline{P}_3) = 0$. Thus, there is no dominance relationship between d_1 and d_2 . Moreover, any element $P_2 \in \mathcal{M}(\underline{P}_2)$ will be of the type (α, β, γ) , with $0 \leq \alpha, \beta, \gamma \leq 0.5$ and $\alpha + \beta + \gamma = 1$. Then we deduce from Eq. (17) that:

$$\begin{aligned} d_3(\underline{P}_1, \underline{P}_2) &= \max_{(\alpha, \beta) \in [0, 0.5]^2: 0.5 \leq \alpha + \beta \leq 1} \left\{ \frac{1}{3} - \alpha, \frac{1}{3} - \alpha + \frac{1}{3} - \beta \right\} \\ &\leq \max \left\{ \frac{1}{3}, \frac{1}{6} \right\} = \frac{1}{3} < 0.5. \end{aligned}$$

Taking Proposition 9 into account, we conclude that

$$d_2(\underline{P}_1, \underline{P}_2) > d_3(\underline{P}_1, \underline{P}_2) = d_1(\underline{P}_1, \underline{P}_2) = \frac{1}{3}.$$

On the other hand, if we take $P_2 = (0, 0.5, 0.5) \in \mathcal{M}(P_2)$ and $P_3 = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6}) \in \mathcal{M}(P_3)$, we obtain that

$$d_3(\underline{P}_2, \underline{P}_3) \geq \|P_2 - P_3\| = \frac{1}{2} \left(\frac{2}{3} + \frac{1}{3} + \frac{1}{3} \right) = \frac{2}{3},$$

whence

$$d_3(\underline{P}_2, \underline{P}_3) > d_1(\underline{P}_2, \underline{P}_3) > d_2(\underline{P}_2, \underline{P}_3).$$

Thus, there is no dominance relation between d_2 and d_3 either, and moreover d_1 and d_3 are not equivalent in general. \blacklozenge

At this point, one may think that these distances could be used to obtain undominated outer approximations. However, all these extensions may lead to outer approximations that are dominated, and therefore cannot be considered adequate for our problem, as the next examples show.

Example 5. Consider a four element space and the lower probability \underline{P} defined in the following table:

A	$\underline{P}(A)$	$\underline{Q}'_1(A)$	$\underline{Q}'_2(A)$
$\{x_1\}$	0.1	0.1	0.1
$\{x_2\}$	0	0	0
$\{x_3\}$	0	0	0
$\{x_4\}$	0.1	0.1	0.1
$\{x_1, x_2\}$	0.4	0.3	0.3
$\{x_1, x_3\}$	0.4	0.3	0.3
$\{x_1, x_4\}$	0.4	0.4	0.3
$\{x_2, x_3\}$	0.2	0.2	0.2
$\{x_2, x_4\}$	0.4	0.3	0.3
$\{x_3, x_4\}$	0.4	0.3	0.3
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5
$\{x_1, x_2, x_4\}$	0.6	0.6	0.6
$\{x_1, x_3, x_4\}$	0.6	0.6	0.6
$\{x_2, x_3, x_4\}$	0.5	0.5	0.5
\mathcal{X}	1	1	1

Note that \underline{P} is a coherent lower probability because it is the lower envelope of the probability measures with mass functions

$$(0.4, 0, 0.2, 0.4), (0.3, 0.1, 0.1, 0.5), (0.3, 0.3, 0.3, 0.1), (0.1, 0.3, 0.3, 0.3), \\ (0.4, 0.2, 0, 0.4), (0.2, 0.2, 0.4, 0.2), (0.2, 0.4, 0.2, 0.2), (0.5, 0.1, 0.1, 0.3).$$

To see that it is not 2-monotone, note that, taking $A = \{x_1, x_2\}$ and $B = \{x_1, x_3\}$ it holds that:

$$\underline{P}(\{x_1, x_2, x_3\}) + \underline{P}(\{x_1\}) = 0.6 < 0.8 = \underline{P}(\{x_1, x_2\}) + \underline{P}(\{x_1, x_3\}). \quad (18)$$

Any outer approximation \underline{Q} in the class of 2-monotone lower probabilities must satisfy:

$$\underline{Q}(\{x_1, x_2\}) + \underline{Q}(\{x_1, x_3\}) \leq \underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_1\}).$$

Also, since \underline{Q} is an outer approximation of \underline{P} , and using Eq. (18), it holds that:

$$\underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_1\}) \leq \underline{P}(\{x_1, x_2, x_3\}) + \underline{P}(\{x_1\}) \\ = \underline{P}(\{x_1, x_2\}) + \underline{P}(\{x_1, x_3\}) - 0.2.$$

Hence, if \underline{Q} is a 2-monotone outer approximation of \underline{P} , it must satisfy:

$$\underline{Q}(\{x_1, x_2\}) + \underline{Q}(\{x_1, x_3\}) \leq \underline{P}(\{x_1, x_2\}) + \underline{P}(\{x_1, x_3\}) - 0.2.$$

Hence, $d_1(\underline{P}, \underline{Q}) \geq 0.1$. To see that this is indeed the minimum distance among the 2-monotone outer approximations, note that it is attained by the 2-monotone capacities $\underline{Q}'_1, \underline{Q}'_2$ in the table above. Thus, both $\underline{Q}'_1, \underline{Q}'_2$ are optimal outer approximations with respect to the distance d_1 , even if \underline{Q}'_2 is dominated by \underline{Q}'_1 . \blacklozenge

Example 6. Consider again the coherent lower probability from Example 1. Any 2-monotone outer approximation \underline{Q} of \underline{P} , undominated or not, shall satisfy $\underline{Q}(\{x_j\}) = 0$ for every j , and as a consequence $d_2(\underline{P}, \underline{Q}) = 0$. As for d_3 , $\|P_1 - P_2\| = 1$ for the probability measures $P_1 = (0.5, 0.5, 0, 0)$ and $P_2 = (0, 0, 0.5, 0.5)$ from $\mathcal{M}(\underline{P})$, and by definition this is the maximum value of the total variation $\|P_1 - P_2\|$. Because $\mathcal{M}(\underline{P}) \subset \mathcal{M}(\underline{Q})$, also $d_3(\underline{P}, \underline{Q}) = 1$ for any 2-monotone outer approximation \underline{Q} of \underline{P} , even for the ‘most dominated’ vacuous lower probability \underline{Q}_v . Thus, \bar{d}_2, \bar{d}_3 do not rule out the undominated solutions, either. \blacklozenge

5.3 The Weber set

A coherent lower probability \underline{P} is uniquely determined by its credal set $\mathcal{M}(\underline{P})$ using Eq. (1). This set is closed and convex, and can thus be determined by its extreme points. We say that a probability measure P is an extreme point of $\mathcal{M}(\underline{P})$ if there are no $P_1, P_2 \in \mathcal{M}(\underline{P})$ and $\alpha \in (0, 1)$ such that $P = \alpha P_1 + (1 - \alpha)P_2$, and shall denote by $ext(\mathcal{M}(\underline{P}))$ the set of extreme points of $\mathcal{M}(\underline{P})$.

In general, there is no easy procedure to determine the extreme points of $\mathcal{M}(\underline{P})$ when \underline{P} is a coherent lower probability. A nice feature of a 2-monotone capacity is that the extreme points of its credal set are in a one-to-one correspondence with the permutations of the possibility space, in the following manner: let \underline{P} be a 2-monotone capacity, and for any permutation σ of $\{1, \dots, n\}$, define the precise probability P_σ by

$$\begin{aligned} P_\sigma(\{x_{\sigma(1)}\}) &= \underline{P}(\{x_{\sigma(1)}\}), \\ P_\sigma(\{x_{\sigma(k)}\}) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k)}\}) - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k-1)}\}) \end{aligned} \quad (19)$$

for every $k = 2, \dots, n$. It was first proven by Shapley [26] that, if S_n denotes the set of permutations of $\{1, \dots, n\}$, then

$$ext(\mathcal{M}(\underline{P})) = \{P_\sigma \mid \sigma \in S_n\}.$$

In general, even when \underline{P} is not 2-monotone but only coherent, we can define the set of probabilities:

$$W(\underline{P}) = \{P_\sigma \mid \sigma \in S_n\}, \quad (20)$$

where P_σ is defined as in Eq. (19). This is called the *Weber set* of \underline{P} , and it holds that [17] \underline{P} is 2-monotone if and only if $ext(\mathcal{M}(\underline{P})) = W(\underline{P})$. Otherwise, $\mathcal{M}(\underline{P})$ is a proper subset of $conv(W(\underline{P}))$, the set of convex combinations of elements in $W(\underline{P})$. This implies that the lower envelope \underline{Q} of $conv(W(\underline{P}))$ is a coherent lower probability that outer approximates \underline{P} . Computationally, the procedure for obtaining \underline{Q} is nimbler than the linear programming problem in Section 3, thus looking for \underline{Q} seems to be a tempting alternative in the outer approximation problem.

In fact, in the case of cardinality four, the lower envelope of $conv(W(\underline{P}))$ is indeed 2-monotone:

Proposition 10. *Let $\underline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ be a coherent lower probability, where $|\mathcal{X}| \leq 4$, and denote by \underline{Q} the coherent lower probability defined by $\underline{Q}(E) = \min\{P(E) \mid P \in conv(W(\underline{P}))\}$ for every $E \subseteq \mathcal{X}$, where $W(\underline{P})$ is given by Eq. (20). Then, \underline{Q} is a 2-monotone outer approximation of \underline{P} .*

Yet, and perhaps surprisingly, the lower envelope of the Weber set is not 2-monotone in general for cardinalities greater than 4, as our next example shows:

Example 7. Consider a five element space \mathcal{X} and the lower probability defined as the lower envelope of the probability measures with mass functions:

$$\begin{aligned} &(0.2080, 0.0113, 0.2693, 0.2962, 0.2152), \\ &(0.2786, 0.2732, 0.1442, 0.2410, 0.0630), \\ &(0.6097, 0.0275, 0.2191, 0.0399, 0.0839). \end{aligned}$$

If we denote by \underline{Q} the lower probability $\underline{Q}(E) = \min\{P(E) \mid P \in \text{conv}(W(\underline{P}))\}$ for every $E \subseteq \mathcal{X}$, it can be verified that:

$$\begin{aligned} \underline{Q}(\{x_2, x_3\}) &= P_{\sigma_1}(\{x_2, x_3\}) = 0.1717, & \text{where } \sigma_1 &= (3, 4, 2, 1, 5). \\ \underline{Q}(\{x_2, x_5\}) &= P_{\sigma_2}(\{x_2, x_5\}) = 0.0905, & \text{where } \sigma_2 &= (5, 3, 4, 2, 1). \\ \underline{Q}(\{x_2, x_3, x_5\}) &= P_{\sigma_3}(\{x_2, x_3, x_5\}) = 0.2347, & \text{where } \sigma_3 &= (5, 3, 4, 2, 1). \end{aligned}$$

Also, from Lemma 15 in the Appendix, \underline{P} and \underline{Q} coincide on singletons, whence $\underline{Q}(\{x_2\}) = \underline{P}(\{x_2\}) = 0.0113$.

Then, we have that:

$$\begin{aligned} \underline{Q}(\{x_2, x_3\}) + \underline{Q}(\{x_2, x_5\}) &= 0.1717 + 0.09105 = 0.2622 \\ &> 0.2460 = 0.0113 + 0.2347 = \underline{Q}(\{x_2\}) + \underline{Q}(\{x_2, x_3, x_5\}), \end{aligned}$$

and as a consequence \underline{Q} is not 2-monotone. \blacklozenge

Thus, we cannot use the lower envelope of the Weber set as a 2-monotone outer approximation for cardinalities greater than four. Further, even for this case the outer approximation we obtain may be dominated:

Example 8. Consider a four-element space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, and the

coherent lower probability \underline{P} from Example 3, reported in the following table:

A	$\underline{P}(A)$	$\underline{Q}(A)$	$\underline{Q}_1(A)$
$\{x_1\}$	0.1	0.1	0.1
$\{x_2\}$	0	0	0
$\{x_3\}$	0	0	0
$\{x_4\}$	0.3	0.3	0.3
$\{x_1, x_2\}$	0.1	0.1	0.1
$\{x_1, x_3\}$	0.3	0.2	0.3
$\{x_1, x_4\}$	0.6	0.5	0.5
$\{x_2, x_3\}$	0.3	0.2	0.2
$\{x_2, x_4\}$	0.4	0.3	0.4
$\{x_3, x_4\}$	0.4	0.3	0.4
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5
$\{x_1, x_2, x_4\}$	0.6	0.6	0.6
$\{x_1, x_3, x_4\}$	0.7	0.7	0.7
$\{x_2, x_3, x_4\}$	0.6	0.6	0.6
\mathcal{X}	1	1	1

If we compute $\underline{Q} = \min\{P \mid P \in \text{conv}(W(\underline{P}))\}$, we obtain the values depicted in the table above. However, this 2-monotone capacity is dominated by the 2-monotone outer approximation \underline{Q}_1 given in the same table, which is the solution of the linear programming problem (LP-2monot). \blacklozenge

6 Approximations of coherent lower previsions

The problem considered in this paper can be generalized from coherent lower probabilities to the richer framework of coherent lower *previsions* [29]: if we denote by $\mathcal{L}(\mathcal{X})$ the set of bounded real-valued functions on \mathcal{X} , a coherent lower prevision is a function $\underline{P} : \mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R}$ that satisfies

- $\underline{P}(f) \geq \inf f$
- $\underline{P}(\lambda f) = \lambda \underline{P}(f)$
- $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$

for every $f, g \in \mathcal{L}(\mathcal{X})$ and every $\lambda > 0$. A coherent lower prevision \underline{P} can be equivalently expressed in the following way: \underline{P} is coherent when it is the lower envelope of a set of expectation operators with respect to a family of probability measures on \mathcal{X} :

$$\mathcal{M}(\underline{P}) = \{P \text{ probability} \mid E_P(f) \geq \underline{P}(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\}.$$

Then, \underline{P} is a coherent lower prevision if and only if $\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} E_P(f)$ for every $f \in \mathcal{L}(\mathcal{X})$.

The notion of 2-monotonicity has also been extended to lower previsions [8, 28]: \underline{P} is a 2-monotone lower prevision if and only if

$$\underline{P}(f \wedge g) + \underline{P}(f \vee g) \geq \underline{P}(f) + \underline{P}(g) \quad \forall f, g \in \mathcal{L}(\mathcal{X}),$$

where \wedge and \vee denote the pointwise minimum and maximum, respectively.

The restriction of a coherent lower prevision to $\{I_A \mid A \subseteq \mathcal{X}\}$ is a coherent lower probability, and the convention $\underline{P}(I_A) = \underline{P}(A)$ is usually adopted. In general, a coherent lower probability \underline{P} on $\mathcal{P}(\mathcal{X})$ may have more than one extension to $\mathcal{L}(\mathcal{X})$ as a coherent lower prevision; however, if \underline{P} is 2-monotone, then it has a unique extension to $\mathcal{L}(\mathcal{X})$ as a 2-monotone lower prevision: its Choquet integral.

Theorem 11. [8] *Let $\underline{P}' : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ be a 2-monotone lower probability. Then, there exists a unique 2-monotone lower prevision $\underline{P} : \mathcal{L}(\mathcal{X}) \rightarrow [0, 1]$ such that $\underline{P}'(A) = \underline{P}(I_A)$ for every $A \in \mathcal{P}(\mathcal{X})$, and it is given by:*

$$\underline{P}(f) = (C) \int f d\underline{P}' \quad \forall f \in \mathcal{L}(\mathcal{X}),$$

where $(C) \int f d\underline{P}'$ denotes the Choquet integral of f with respect to \underline{P}' , given by:

$$(C) \int f d\underline{P}' = \inf f + \int_{\inf f}^{\sup f} \underline{P}'(\{f \geq t\}) dt.$$

The lower prevision \underline{P} in this theorem is the *natural extension* of \underline{P}' to $\mathcal{L}(\mathcal{X})$: if \underline{P}' is a 2-monotone lower probability, its unique 2-monotone extension to $\mathcal{L}(\mathcal{X})$ is given by $\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P}')} E_P(f)$.

Similarly to what we have done in the rest of the paper, we could study the problem of outer approximating a coherent lower prevision by a 2-monotone one. Interestingly, this problem turns out to be equivalent to the one we have analyzed for finite spaces \mathcal{X} , and anyway the equivalence extends to infinite spaces \mathcal{X} too, as our next result shows:

Proposition 12. *Let $\underline{P} : \mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R}$ be a coherent lower prevision, and let \underline{P}' be its restriction to events. Then, there is a one-to-one correspondence between the sets*

$$\{\underline{Q} : \mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R} \text{ 2-monotone undominated outer approximation of } \underline{P}\}$$

and

$$\{\underline{Q}' : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1] \text{ 2-monotone undominated outer approximation of } \underline{P}'\}.$$

The key in this result is that if we want to outer approximate a coherent lower prevision, we can simply consider its restriction to events, outer approximate it and then apply the procedure of natural extension. Figure 1 provides an illustration.

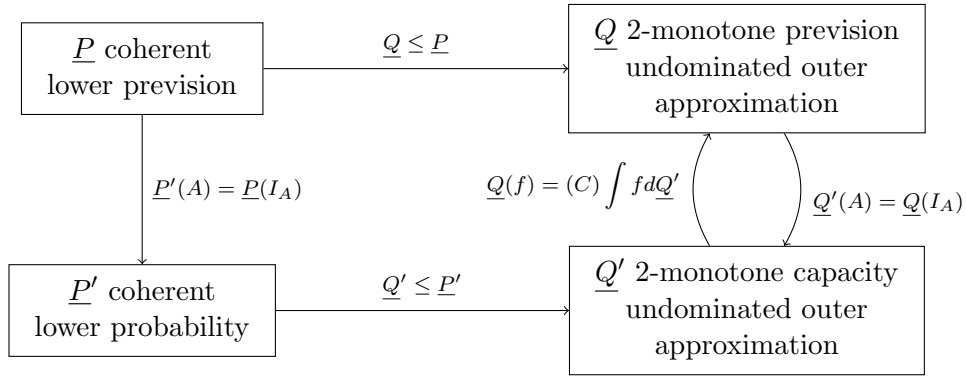


Figure 1: Correspondence between the 2-monotone outer approximations.

Therefore, it suffices to focus on outer approximations of coherent lower probabilities instead of lower previsions.

Remark 1. *Using a proof similar to that of Proposition 12, that we can find in the Appendix, we can establish a similar result for the particular cases of 2-monotone lower probabilities that we have considered in Section 4: it suffices to take into account that a 2-monotone lower probability has a unique extension to $\mathcal{L}(\mathcal{X})$ that preserves 2-monotonicity, as we have mentioned. For the extension of the pari mutuel model to gambles, we refer the reader to [24, Section 3]; ϵ -contamination models on gambles were considered by Walley in [29, Section 2.9.2].*

Note that the extension of the pari mutuel model to gambles can be interpreted in risk measurement as a coherent and 2-monotone risk measure, termed Tail Value-at-Risk or TVaR [24, Section 3], while coherent upper previsions correspond to coherent risk measures. Thus the present approximation problem also has the (risk measurement) meaning of approximating a generic coherent risk measure with the closest (2-monotone) TVaR. ♦

7 Inner approximations of coherent lower probabilities

So far, we have studied how to outer approximate a coherent lower probability. Similarly, we may also consider the problem of finding inner approximations². Following the reasoning in Section 2, given a coherent lower probability \underline{P} , a 2-monotone lower probability \underline{Q} is an inner approximation of \underline{P} when $\underline{P} \leq \underline{Q}$; it is also called *non-dominating* when any other 2-monotone inner approximation \underline{Q}' such that $\underline{P} \leq \underline{Q}' \leq \underline{Q}$ must satisfy $\underline{Q}' = \underline{Q}$.

The problem of looking for non-dominating inner approximations can be tackled in a similar way as we did in Section 3: we can set up the minimization problem

$$\min \sum_{E \subseteq \mathcal{X}} (\underline{Q}(E) - \underline{P}(E)), \quad (\text{LP-inner})$$

which can be equivalently expressed in terms of the Möbius inverse $m_{\underline{Q}}$:

$$\min \sum_{E \subseteq \mathcal{X}} \left(\left(\sum_{B \subseteq E} m_{\underline{Q}}(B) \right) - \underline{P}(E) \right).$$

In order to guarantee that $m_{\underline{Q}}$ defines a 2-monotone lower probability \underline{Q} , we need to consider the restrictions (LP-2monot.1)–(LP-2monot.3), and to guarantee that \underline{Q} is an inner approximation, we consider the following constraint:

$$\sum_{B \subseteq E} m_{\underline{Q}}(B) \geq \underline{P}(E) \quad \forall E \neq \emptyset, \mathcal{X}. \quad (\text{LP-2monot.4}^*)$$

The feasible region of the linear programming problem (LP-inner) with constraints (LP-2monot.1)–(LP-2monot.3) and (LP-2monot.4*) is non-empty: since \underline{P} is coherent, its credal set is non-empty, so every (precise) probability in $\mathcal{M}(\underline{P})$ is in particular 2-monotone and satisfies all the constraints. Also, the same reasoning as in Section 3 ensures that the linear programming problem has an optimal solution. Not surprisingly, next result says that every optimal solution is a non-dominating inner approximation. Its proof is analogous to that of Prop. 1 and therefore omitted in the Appendix.

²Some results in this respect were established by Dubois and Prade in [13, Section 2.4].

Proposition 13. *Let \underline{P} be a coherent lower probability, and let \underline{Q} be an optimal solution of the linear programming problem (LP-inner) with constraints (LP-2monot.1)–(LP-2monot.3) and (LP-2monot.4*). Then \underline{Q} is a non-dominating inner approximation of \underline{P} in \mathcal{C}_2 .*

Although from a formal point of view there are mathematical similarities between the existence of outer and inner approximations, and some of the results in Section 3 can be easily adapted to the setting of inner approximations, this second problem is not as interesting, in our view. The main reason is that the use of inner approximations entails breaking the first of our two initial conditions: that *the approximation introduces no new information*. Indeed, if $\mathcal{M}(\underline{P})$ represents a set of possible candidates for an unknown or partially defined precise probability P_0 , an inner approximation \underline{Q} reduces the set of possible candidates: $\mathcal{M}(\underline{Q}) \subset \mathcal{M}(\underline{P})$; thus, the elicitation of an inner approximation is always adding (unwarranted) information to the model.

Moreover, the benefits ensued by using particular models for approximating the coherent lower probability that we have seen in Section 4 are not preserved in general when looking for inner approximations. Of course, an inner approximation in \mathcal{C}_{PI} always exists, taking for example a precise probability in $\mathcal{M}(\underline{P})$. However, the existence of inner approximations in \mathcal{C}_{PMM} and \mathcal{C}_ϵ is not guaranteed. For example, if there is an element $x \in \mathcal{X}$ such that $\underline{P}(\{x\}) = \overline{P}(\{x\}) \in (0, 1)$, any inner approximation must satisfy the condition $\underline{P}(\{x\}) = \underline{Q}(\{x\}) = \overline{Q}(\{x\}) = \overline{P}(\{x\}) \in (0, 1)$, thus having imprecision $\overline{Q} - \underline{Q} = 0$ at x . This is impossible for both the PMM and for the ϵ -contamination models: on events with non-extreme evaluation, they have constant positive imprecision, $\delta > 0$ and $\epsilon > 0$, respectively.

8 Conclusions

In this paper we have investigated the problem of approximating a lower probability by means of a 2-monotone one that satisfies two properties: firstly, it must not add new information, and secondly, it should be as close as possible to the original lower probability, so as to minimize the increase in imprecision. This has led to the quest for undominated 2-monotone capacities that outer approximate the original model, which was firstly tackled in [3].

We have first considered the closest outer approximation under the distance proposed by Baroni and Vicig in [2], and have obtained undominated

outer approximations by means of linear programming. Although the approximation with this procedure is not unique in general, it becomes unique if we focus on some particular subfamilies of 2-monotone capacities, such as those associated with notable distortion models or probability intervals.

We have also discussed alternative approaches based on other distances, such as the quadratic or the total variation distance. Although they display some interesting features, they also possess a number of drawbacks that make them less suitable in our view for the problem at hand than the distance of Baroni and Vicig.

We have also investigated a more general problem: the outer approximation of coherent lower previsions by 2-monotone ones. Our results show that this problem is actually equivalent to the one we have studied in the main bulk of the paper: in order to outer approximate a lower prevision, we just need to outer approximate its restriction to events and then take the natural extension to $\mathcal{L}(\mathcal{X})$.

We think that this paper opens a line of research that can be pursued in different ways. Firstly, in this paper we are obtaining undominated outer approximations by minimising the distance proposed by Baroni and Vicig. An alternative procedure would be to minimise a divergence instead of a distance [22] or to consider linear imprecision indices [3]. Secondly, although here we have considered the problem of outer approximating a coherent lower probability by means of a 2-monotone capacity, we could also consider outer approximations by means of belief functions, or maxitive capacities. Some advances have already been done in this setting in [2, 12, 13, 21]. Thirdly, we would also like to study in greater detail the loss of information entailed by the outer approximations, as well as the elicitation among them when there is more than one undominated outer approximation. We think that the results in [27] could be helpful in this regard. Fourthly, we shall mention that the enumeration of *all* the undominated outer approximations is not a simple task. We have seen that one way of obtaining undominated outer approximations in \mathcal{C}_2 is by solving a linear programming problem, but Example 3 shows that this procedure does not give all of them. Although the undominated outer approximations in \mathcal{C}_{PI} , \mathcal{C}_{PMM} and \mathcal{C}_ε are unique, it remains as an open problem to enumerate all of them in \mathcal{C}_2 . In this sense, the theoretical characterization given in [3, Proposition 4] may be useful.

Acknowledgements

The research in this paper has been supported by project TIN2014-59543-P. We would also like to thank Sébastien Destercke and Serafín Moral for some helpful suggestions and Massimiliano Kaucic for his valuable computational assistance. Also, we want to thank the reviewers for their suggestions which helped us to improve the quality of the paper. A preliminary version of this paper was presented at the IPMU'2018 conference [20]; the current version includes additional results, proofs and discussions.

Appendix: Proofs

Lemma 14. *Let \underline{P} be a coherent lower probability and denote by \underline{P}' an outer approximation in a class \mathcal{C} . If \underline{P}' is a solution to the problem:*

$$\min_{\underline{P}' \in \mathcal{C}, \underline{P}' \leq \underline{P}} \sum_{E \subseteq \mathcal{X}} g(\underline{P}(E) - \underline{P}'(E)), \quad (21)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing, then \underline{P}' is an undominated outer approximation of \underline{P} in \mathcal{C} .

Proof. Assume ex-absurdo that \underline{P}' is dominated in \mathcal{C} , so there exists $\underline{P}'' \in \mathcal{C}$ such that $\underline{P}' \leq \underline{P}'' \leq \underline{P}$. Then, $g(\underline{P}(E) - \underline{P}'(E)) \geq g(\underline{P}(E) - \underline{P}''(E))$ for every event $E \subseteq \mathcal{X}$, with strict inequality in some event E_0 . Hence:

$$\sum_{E \subseteq \mathcal{X}} g(\underline{P}(E) - \underline{P}'(E)) > \sum_{E \subseteq \mathcal{X}} g(\underline{P}(E) - \underline{P}''(E)).$$

This contradicts that \underline{P}' is a solution of Eq. (21). Thus, \underline{P}' is undominated. \square

Proof of Proposition 1. First of all, as we have already argued, \underline{Q} is a 2-monotone lower probability by conditions (LP-2monot.1), (LP-2monot.2) and (LP-2monot.3). Also, condition (LP-2monot.4) implies that \underline{Q} outer approximates \underline{P} : $\underline{Q} \leq \underline{P}$. Finally, from Lemma 14 we deduce that it is undominated in \mathcal{C}_2 . \square

Proof of Proposition 2. Let us first prove that \underline{P} and \underline{Q} coincide on singletons. Assume that there exists $x \in \mathcal{X}$ such that $\underline{Q}(\{x\}) < \underline{P}(\{x\})$. Let us define \underline{Q}' recursively by:

$$\underline{Q}'(\{x\}) \in (\underline{Q}(\{x\}), \underline{P}(\{x\})], \quad \underline{Q}'(\{x'\}) = \underline{Q}(\{x'\}) \quad \forall x' \neq x,$$

and for every event A of cardinality $|A| > 1$:

$$\underline{Q}'(A) = \max \left\{ \underline{Q}(A), \max_{C_1, \dots, C_j} \left\{ \sum_{i=1}^j \underline{Q}'(C_i) \right\} \right\}, \quad (22)$$

where C_1, \dots, C_j form a partition of A .

Let us now establish some properties of \underline{Q}' :

- (a) $\underline{Q} \leq \underline{Q}' \leq \underline{P}$: for every event A , it follows by definition that $\underline{Q}'(A) \geq \underline{Q}(A)$. Moreover, $\underline{Q}'(\{x\}) > \underline{Q}(\{x\})$. Thus, $\underline{Q} \leq \underline{Q}'$.

To see that $\underline{Q}' \leq \underline{P}$, we proceed iteratively on the cardinalities of the events on \mathcal{X} . First, recall that the inequality follows on singletons. Next, given a partition C_1, \dots, C_j of A , if $\underline{Q}'(C_i) \leq \underline{P}(C_i)$ for every $i = 1, \dots, j$, we get that:

$$\sum_{i=1}^j \underline{Q}'(C_i) \leq \sum_{i=1}^j \underline{P}(C_i) \leq \underline{P}(A),$$

where the last inequality follows from the superadditivity (C4) of \underline{P} . Since also $\underline{Q} \leq \underline{P}$, we deduce from Eq. (22) that $\underline{Q}' \leq \underline{P}$.

- (b) $\underline{Q}'(A) = \underline{Q}(A)$ if $x \notin A$: we apply induction on $|A|$.
If $|A| = 1$, this holds by definition.
If $|A| = k$, take any partition C_1, \dots, C_j of A . Then:

$$\sum_{i=1}^j \underline{Q}'(C_i) = \sum_{i=1}^j \underline{Q}(C_i) \leq \underline{Q}(A),$$

where the equality follows by the induction hypothesis and the inequality follows by the super-additivity (C4) of \underline{Q} .

- (c) \underline{Q}' can be equivalently expressed by:

$$\underline{Q}'(A) = \begin{cases} \max\{\underline{Q}(A), \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\})\} & \text{if } x \in A. \\ \underline{Q}(A) & \text{if } x \notin A. \end{cases}$$

On the one hand, we already know that $\underline{Q}'(A) = \underline{Q}(A)$ when $x \notin A$.

Let us now assume that $x \in A$ and apply induction on $|A|$:

If $|A| = 2$, if $x \in A$, then $A = \{x, y\}$ ($x \neq y$) and it holds that:

$$\underline{Q}'(A) = \max \{ \underline{Q}(A), \underline{Q}'(\{x\}) + \underline{Q}'(\{y\}) \} = \max \{ \underline{Q}(A), \underline{Q}'(\{x\}) + \underline{Q}(\{y\}) \},$$

where the last equality follows from the previous item.

Assume that the alternative expression holds until $|A| = k-1$ and take A with cardinality k . Consider a partition C_1, \dots, C_j of A . Then, there will be exactly one event including x ; assume for instance that it is C_1 . By item (b) and superadditivity (C4) of \underline{Q} , it holds that:

$$\begin{aligned} \sum_{i=1}^j \underline{Q}'(C_i) &= \underline{Q}'(C_1) + \underline{Q}(C_2) + \dots + \underline{Q}(C_j) \\ &\leq \underline{Q}'(C_1) + \underline{Q}(C_2 \cup \dots \cup C_j) = \underline{Q}'(C_1) + \underline{Q}'(C_2 \cup \dots \cup C_j). \end{aligned}$$

It ensues that the inner maximum in Eq. (22) is achieved by some two-element partition of A . Take any such partition, whose elements are C_1 , including x , and C_2 . It holds that:

$$\begin{aligned} \underline{Q}'(C_1) + \underline{Q}'(C_2) &= \underline{Q}'(C_1) + \underline{Q}(C_2) \\ &= \max \{ \underline{Q}(C_1) + \underline{Q}(C_2), \underline{Q}'(\{x\}) + \underline{Q}(C_1 \setminus \{x\}) + \underline{Q}(C_2) \} \\ &\leq \max \{ \underline{Q}(A), \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\}) \}, \end{aligned}$$

where the second equality follows applying the induction hypothesis and the inequality from the super-additivity (C4) of \underline{Q} . Since we can do this for every two-element partition of A , we conclude that:

$$\underline{Q}'(A) \leq \max \{ \underline{Q}(A), \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\}) \}.$$

The converse inequality is trivial.

- (d) If \underline{Q} is 2-monotone, so is \underline{Q}' : consider two events A, B and let us prove that

$$\underline{Q}'(A \cup B) + \underline{Q}'(A \cap B) \geq \underline{Q}'(A) + \underline{Q}'(B).$$

There are a number of possibilities:

1. If either $x \notin A \cup B$ or $x \in A \setminus B$ and $\underline{Q}'(A) = \underline{Q}(A)$, then:

$$\underline{Q}'(A) + \underline{Q}'(B) = \underline{Q}(A) + \underline{Q}(B) \leq \underline{Q}(A \cup B) + \underline{Q}(A \cap B) \leq \underline{Q}'(A \cup B) + \underline{Q}'(A \cap B),$$

where the equality follows from (b).

2. If $x \in A \setminus B$, and $\underline{Q}'(A) = \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\})$, then:

$$\begin{aligned} \underline{Q}'(A) + \underline{Q}'(B) &= \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\}) + \underline{Q}(B) \\ &\leq \underline{Q}'(\{x\}) + \underline{Q}((A \setminus \{x\}) \cap B) + \underline{Q}((A \setminus \{x\}) \cup B) \\ &\leq \underline{Q}((A \setminus \{x\}) \cap B) + \underline{Q}'(A \cup B) \\ &= \underline{Q}(A \cap B) + \underline{Q}'(A \cup B) = \underline{Q}'(A \cap B) + \underline{Q}'(A \cup B). \end{aligned}$$

3. If $x \in B \setminus A$, the proof is analogous to either case (1) or case (2).
4. Finally, assume that $x \in A \cap B$. There are four cases:
 - If $\underline{Q}'(A) = \underline{Q}(A)$, $\underline{Q}'(B) = \underline{Q}(B)$, we continue as in case (1).
 - $\underline{Q}'(A) = \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\})$, $\underline{Q}'(B) = \underline{Q}(B)$. Then:

$$\begin{aligned} \underline{Q}'(A) + \underline{Q}'(B) &= \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\}) + \underline{Q}(B) \\ &\leq \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\} \cap B) + \underline{Q}(A \setminus \{x\} \cup B) \\ &\leq \underline{Q}'(A \cap B) + \underline{Q}'(A \setminus \{x\} \cup B) \leq \underline{Q}'(A \cap B) + \underline{Q}'(A \cup B). \end{aligned}$$

- If $\underline{Q}'(A) = \underline{Q}(A)$ and $\underline{Q}'(B) = \underline{Q}'(\{x\}) + \underline{Q}(B \setminus \{x\})$, the proof is analogous to the one in the previous bullet.
- Finally, if $\underline{Q}'(A) = \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\})$ and $\underline{Q}'(B) = \underline{Q}'(\{x\}) + \underline{Q}(B \setminus \{x\})$, then:

$$\begin{aligned} \underline{Q}'(A) + \underline{Q}'(B) &= \underline{Q}'(\{x\}) + \underline{Q}'(\{x\}) + \underline{Q}(A \setminus \{x\}) + \underline{Q}(B \setminus \{x\}) \\ &\leq \underline{Q}'(\{x\}) + \underline{Q}'(\{x\}) + \underline{Q}((A \cap B) \setminus \{x\}) + \underline{Q}((A \cup B) \setminus \{x\}) \\ &\leq \underline{Q}'(A \cap B) + \underline{Q}'(A \cup B). \end{aligned}$$

We conclude that \underline{Q}' is a 2-monotone outer approximation of \underline{P} that dominates \underline{Q} , so \underline{Q} cannot be undominated.

Now, we prove that \overline{P} and \overline{Q} also coincide in singletons; the proof is analogous to the equality between the lower probabilities. Assume that there exists $x \in \mathcal{X}$ such that $\overline{Q}(\{x\}) > \overline{P}(\{x\})$. Let us define \overline{Q}' recursively by:

$$\overline{Q}'(\{x\}) \in [\overline{P}(\{x\}), \overline{Q}(\{x\})], \quad \overline{Q}'(\{x'\}) = \overline{Q}(\{x'\}) \quad \forall x' \neq x,$$

and for every event A of cardinality $|A| > 1$:

$$\overline{Q}'(A) = \min \left\{ \overline{Q}(A), \min_{C_1, \dots, C_j} \left\{ \sum_{i=1}^j \overline{Q}'(C_i) \right\} \right\}, \quad (23)$$

where C_1, \dots, C_j form a partition of A .

Let us now establish some properties of \overline{Q}' :

- (a) $\overline{Q} \succeq \overline{Q}' \geq \overline{P}$: for every event A , it follows by definition that $\overline{Q}'(A) \leq \overline{Q}(A)$. Moreover, $\overline{Q}'(\{x\}) < \overline{Q}(\{x\})$. Thus, $\overline{Q} \succeq \overline{Q}'$.

To see that $\overline{Q}' \geq \overline{P}$, we proceed iteratively on the cardinalities of the events on \mathcal{X} . First, recall that the inequality follows on singletons.

Next, given a partition C_1, \dots, C_j of A , if $\overline{Q}'(C_i) \geq \overline{P}(C_i)$ for every $i = 1, \dots, j$, we get that:

$$\sum_{i=1}^j \overline{Q}'(C_i) \geq \sum_{i=1}^j \overline{P}(C_i) \geq \overline{P}(A),$$

where the last inequality follows from the sub-additivity (C3) of \overline{P} . Since also $\overline{Q} \geq \overline{P}$, we deduce from Eq. (23) that $\overline{Q}' \geq \overline{P}$.

- (b) $\overline{Q}'(A) = \overline{Q}(A)$ if $x \notin A$: we apply induction on $|A|$.
 If $|A| = 1$, this holds by definition.
 If $|A| = k$, take any partition C_1, \dots, C_j of A . Then:

$$\sum_{i=1}^j \overline{Q}'(C_i) = \sum_{i=1}^j \overline{Q}(C_i) \geq \overline{Q}(A),$$

where the equality follows by the induction hypothesis and the inequality follows by the sub-additivity (C3) of \overline{Q} .

- (c) \overline{Q}' can be equivalently expressed by:

$$\overline{Q}'(A) = \begin{cases} \min\{\overline{Q}(A), \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\})\} & \text{if } x \in A. \\ \overline{Q}(A) & \text{if } x \notin A. \end{cases}$$

On the one hand, we already know that $\overline{Q}'(A) = \overline{Q}(A)$ when $x \notin A$. Let us now assume that $x \in A$ and apply induction on $|A|$:
 If $|A| = 2$, if $x \in A$, then $A = \{x, y\}$ ($x \neq y$) and it holds that:

$$\overline{Q}'(A) = \min\{\overline{Q}(A), \overline{Q}'(\{x\}) + \overline{Q}'(\{y\})\} = \min\{\overline{Q}(A), \overline{Q}'(\{x\}) + \overline{Q}(\{y\})\},$$

where the last equality follows from the previous item.

Assume that the alternative expression holds until $|A| = k-1$ and take A with cardinality k . Consider a partition C_1, \dots, C_j of A . Then, there will be exactly one event including x ; assume for instance that it is C_1 . By item (b) and sub-additivity (C3) of \overline{Q} , it holds that:

$$\begin{aligned} \sum_{i=1}^j \overline{Q}'(C_i) &= \overline{Q}'(C_1) + \overline{Q}(C_2) + \dots + \overline{Q}(C_j) \\ &\geq \overline{Q}'(C_1) + \overline{Q}(C_2 \cup \dots \cup C_j) = \overline{Q}'(C_1) + \overline{Q}'(C_2 \cup \dots \cup C_j). \end{aligned}$$

It ensues that the inner minimum in Eq. (23) is achieved by some two-element partition of A . Take any such partition, whose elements are C_1 , including x , and C_2 . It holds that:

$$\begin{aligned}\overline{Q}'(C_1) + \overline{Q}'(C_2) &= \overline{Q}'(C_1) + \overline{Q}(C_2) \\ &= \min \{ \overline{Q}(C_1) + \overline{Q}(C_2), \overline{Q}'(\{x\}) + \overline{Q}(C_1 \setminus \{x\}) + \overline{Q}(C_2) \} \\ &\geq \min \{ \overline{Q}(A), \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\}) \},\end{aligned}$$

where the second equality follows applying the induction hypothesis and the inequality from the sub-additivity (C3) of \overline{Q} . Since we can do this for every two-element partition of A , we conclude that:

$$\overline{Q}'(A) \geq \min \{ \overline{Q}(A), \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\}) \}.$$

The converse inequality is trivial.

- (d) If \overline{Q} is 2-alternating, so is \overline{Q}' : consider two events A, B and let us prove that

$$\overline{Q}'(A \cup B) + \overline{Q}'(A \cap B) \leq \overline{Q}'(A) + \overline{Q}'(B).$$

There are a number of possibilities:

1. If either $x \notin A \cup B$ or $x \in A \setminus B$ and $\overline{Q}'(A) = \overline{Q}(A)$, then

$$\overline{Q}'(A) + \overline{Q}'(B) = \overline{Q}(A) + \overline{Q}(B) \geq \overline{Q}(A \cup B) + \overline{Q}(A \cap B) \geq \overline{Q}'(A \cup B) + \overline{Q}'(A \cap B),$$

where the equality follows from (b).

2. If $x \in A \setminus B$ and $\overline{Q}'(A) = \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\})$, then:

$$\begin{aligned}\overline{Q}'(A) + \overline{Q}'(B) &= \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\}) + \overline{Q}(B) \\ &\geq \overline{Q}'(\{x\}) + \overline{Q}((A \setminus \{x\}) \cap B) + \overline{Q}((A \setminus \{x\}) \cup B) \\ &\geq \overline{Q}((A \setminus \{x\}) \cap B) + \overline{Q}'(A \cup B) \\ &= \overline{Q}(A \cap B) + \overline{Q}'(A \cup B) = \overline{Q}'(A \cap B) + \overline{Q}'(A \cup B).\end{aligned}$$

3. If $x \in B \setminus A$, the proof is analogous to either case (1) or case (2).
4. Finally, assume that $x \in A \cap B$. There are four cases:
 - If $\overline{Q}'(A) = \overline{Q}(A)$, $\overline{Q}'(B) = \overline{Q}(B)$, continue as in case (1).

◦ $\overline{Q}'(A) = \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\})$, $\overline{Q}'(B) = \overline{Q}(B)$. Then:

$$\begin{aligned} \overline{Q}'(A) + \overline{Q}'(B) &= \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\}) + \overline{Q}(B) \\ &\geq \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\} \cap B) + \overline{Q}(A \setminus \{x\} \cup B) \\ &\geq \overline{Q}'(A \cap B) + \overline{Q}'(A \setminus \{x\} \cup B) \geq \overline{Q}'(A \cap B) + \overline{Q}'(A \cup B). \end{aligned}$$

◦ If $\overline{Q}'(A) = \overline{Q}(A)$ and $\overline{Q}'(B) = \overline{Q}'(\{x\}) + \overline{Q}(B \setminus \{x\})$, the proof is analogous to the one in the previous bullet.

◦ Finally, if $\overline{Q}'(A) = \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\})$ and $\overline{Q}'(B) = \overline{Q}'(\{x\}) + \overline{Q}(B \setminus \{x\})$, then:

$$\begin{aligned} \overline{Q}'(A) + \overline{Q}'(B) &= \overline{Q}'(\{x\}) + \overline{Q}'(\{x\}) + \overline{Q}(A \setminus \{x\}) + \overline{Q}(B \setminus \{x\}) \\ &\geq \overline{Q}'(\{x\}) + \overline{Q}'(\{x\}) + \overline{Q}((A \cap B) \setminus \{x\}) + \overline{Q}((A \cup B) \setminus \{x\}) \\ &\geq \overline{Q}'(A \cap B) + \overline{Q}'(A \cup B). \end{aligned}$$

We conclude that \overline{Q}' is a 2-alternating outer approximation of \overline{P} that is dominated by \overline{Q} , so \overline{Q} cannot be an optimal outer approximation of \overline{P} .

From the equalities $\overline{Q}(\{x\}) = \overline{P}(\{x\})$ and $\underline{Q}(\{x\}) = \underline{P}(\{x\})$ for every $x \in \mathcal{X}$ we deduce by conjugacy that $\underline{Q}(A) = \underline{P}(A)$ and $\overline{Q}(A) = \overline{P}(A)$ for every event $A \subset \mathcal{X}$ with cardinality $n - 1$. \square

Proof of Proposition 3. It suffices to take into account that for any event A , any \underline{Q} in the non-empty set \mathcal{M}_A satisfies $\underline{Q}(A) = \underline{P}(A)$, and that any element in $\cup_{A \subseteq \mathcal{X}} \mathcal{M}_A$ is an outer approximation of \underline{P} . \square

Proof of Proposition 4. This result follows immediately from Lemma 14. \square

Proof of Proposition 5. Let us denote by l the lower probability associated with \mathcal{I} by means of Eq. (10). In order to prove that \mathcal{I} is a coherent probability interval, we just need to prove that l is a coherent lower probability. But this holds because it is the natural extension of the restriction to events of cardinality 1 or $n - 1$ of the coherent lower probability \underline{P} .

Let us now see that l is an outer approximation of \underline{P} , that is, $l(A) \leq \underline{P}(A)$ for every $A \subseteq \mathcal{X}$. Since \mathcal{I} is a coherent probability interval, l can be

computed using Eq. (11). Then:

$$\begin{aligned}
l(A) &= \max \left\{ \sum_{x_i \in A} l_i, 1 - \sum_{x_i \notin A} u_i \right\} = \max \left\{ \sum_{x_i \in A} \underline{P}(\{x_i\}), 1 - \sum_{x_i \notin A} \overline{P}(\{x_i\}) \right\} \\
&\leq \max \left\{ \underline{P}(A), 1 - \sum_{x_i \notin A} \overline{P}(\{x_i\}) \right\} \leq \max \{ \underline{P}(A), 1 - \overline{P}(A^c) \} \\
&= \max \{ \underline{P}(A), \underline{P}(A) \} = \underline{P}(A),
\end{aligned}$$

where the first and second inequalities follow from the super- and sub-additivity (conditions (C4) and (C3)) of \underline{P} and \overline{P} , respectively, and the last equality follows because \underline{P} and \overline{P} are conjugate.

Finally, let us see that l is the unique undominated outer approximation of \underline{P} using a probability interval. For this aim, assume that there exists a coherent probability interval $\mathcal{I}' = \{[l'_i, u'_i] \mid i = 1, \dots, n\}$ with associated coherent lower and upper probabilities l', u' such that $l'(A) \leq \underline{P}(A) \leq \overline{P}(A) \leq u'(A)$. From Eq. (12), we just need to realize that:

$$l'_i \leq \underline{P}(\{x_i\}) = l_i \text{ and } u'_i \geq \overline{P}(\{x_i\}) = u_i,$$

so $[l_i, u_i] \subseteq [l'_i, u'_i]$ for every $i = 1, \dots, n$. From [7, Proposition 6], this implies that $l'(A) \leq l(A)$ and $u'(A) \geq u(A)$ for every $A \subseteq \mathcal{X}$. \square

Proof of Proposition 7. First of all, note that P_0 is indeed a probability measure since $P_0(\{x_i\}) \geq 0$ for every $i = 1, \dots, n$ and:

$$\sum_{i=1}^n P_0(\{x_i\}) = \sum_{i=1}^n \frac{\overline{P}(\{x_i\})}{1 + \delta} = \frac{\sum_{i=1}^n \overline{P}(\{x_i\})}{1 + \sum_{i=1}^n \overline{P}(\{x_i\}) - 1} = 1.$$

As a consequence,

$$P_0(A) = \sum_{x_i \in A} P_0(\{x_i\}) = \frac{1}{1 + \delta} \sum_{x_i \in A} \overline{P}(\{x_i\}) \quad \forall A \subseteq \mathcal{X}. \quad (24)$$

Let us first show that \underline{P}_δ is an outer approximation of \underline{P} . For this aim, let us see that $\overline{P}_\delta \geq \overline{P}$. If $\overline{P}_\delta = 1$, the inequality trivially holds. If $\overline{P}_\delta(A) < 1$, Eq. (24) and Lemma 6 imply that:

$$\overline{P}_\delta(A) = \min\{(1 + \delta)P_0(A), 1\} = (1 + \delta)P_0(A) = \sum_{x_i \in A} \overline{P}(\{x_i\}) \geq \overline{P}(A),$$

where the inequality follows from the sub-additivity (C3) of \bar{P} . We conclude that $\bar{P}(A) \leq \bar{P}_\delta(A)$ for every $A \subset \mathcal{X}$, and therefore, by conjugacy:

$$\underline{P}_\delta(A) = 1 - \bar{P}_\delta(A^c) \leq 1 - \bar{P}(A^c) = \underline{P}(A).$$

Let us now prove that \underline{P}_δ is the (unique) undominated outer approximation of \underline{P} in \mathcal{C}_{PMM} . For this aim, let \underline{P}'_δ be the upper probability induced by another pari mutuel model that outer approximates \underline{P} .

First of all, note that for every $x \in \mathcal{X}$:

$$\bar{P}_\delta(\{x\}) = \min\{1, (1 + \delta)P_0(\{x\})\} = \min\{1, \bar{P}(\{x\})\} = \bar{P}(\{x\}), \quad (25)$$

whence $\bar{P}(\{x\}) = \bar{P}_\delta(\{x\}) \leq \bar{P}'_\delta(\{x\})$. Let us see that this inequality also holds for every event A . On the one hand, if $\bar{P}'_\delta(A) = 1$, trivially $\bar{P}_\delta(A) \leq \bar{P}'_\delta(A) = 1$. On the other hand, if $\bar{P}'_\delta(A) < 1$, Lemma 6 implies that

$$\bar{P}'_\delta(A) = \sum_{x_i \in A} \bar{P}'_\delta(\{x_i\}) \geq \sum_{x_i \in A} \bar{P}(\{x_i\}) = \sum_{x_i \in A} \bar{P}_\delta(\{x_i\}) \geq \bar{P}_\delta(A),$$

where the first inequality follows because \bar{P}'_δ is an outer approximation of \bar{P} , and the second inequality follows from the sub-additivity (C3) of \bar{P}_δ . We conclude that $\bar{P}_\delta(A) \leq \bar{P}'_\delta(A)$ for every $A \subseteq \mathcal{X}$. \square

Proof of Proposition 8. First of all, note that P_0 is indeed a probability because $P_0(\{x_i\}) \geq 0$ and:

$$\sum_{i=1}^n P_0(\{x_i\}) = \sum_{i=1}^n \frac{P(\{x_i\})}{\sum_{j=1}^n P(\{x_j\})} = 1.$$

Let us now see that $\underline{P}_\varepsilon$ is an outer approximation of \underline{P} , that is, $\underline{P}_\varepsilon(A) \leq \underline{P}(A)$ for every $A \subseteq \mathcal{X}$. For $A = \mathcal{X}$, it trivially holds that $\underline{P}_\varepsilon(\mathcal{X}) = \underline{P}(\mathcal{X}) = 1$. Take $A \subset \mathcal{X}$. On the one hand, the super-additivity (C4) of \underline{P} implies:

$$\underline{P}(A) \geq \sum_{x_i \in A} \underline{P}(\{x_i\}). \quad (26)$$

On the other hand:

$$\underline{P}_\varepsilon(A) = (1 - \varepsilon)P_0(A) = \left(\sum_{j=1}^n \underline{P}(\{x_j\}) \right) \frac{\sum_{x_i \in A} \underline{P}(\{x_i\})}{\sum_{j=1}^n \underline{P}(\{x_j\})} = \sum_{x_i \in A} \underline{P}(\{x_i\}). \quad (27)$$

Combining Eqs. (26) and (27), we conclude that $\underline{P}_\varepsilon(A) \leq \underline{P}(A)$.

Let us now see that $\underline{P}_\varepsilon$ is the only undominated outer approximation of \underline{P} in \mathcal{C}_ε . Assume that there exists $\varepsilon' \in (0, 1)$ and P'_0 inducing an ε' -contamination model \underline{Q} that outer approximates \underline{P} . Then:

$$\underline{P}_\varepsilon(\{x_i\}) = (1 - \varepsilon)P_0(\{x_i\}) = \underline{P}(\{x_i\}) \geq \underline{Q}(\{x_i\}),$$

where the second equality follows from the definition of ε and P_0 , the inequality because \underline{Q} is an outer approximation of \underline{P} .

Now, given $A \subseteq \mathcal{X}$, it holds that:

$$\begin{aligned} \underline{P}_\varepsilon(A) &= (1 - \varepsilon)P_0(A) = (1 - \varepsilon) \sum_{x_i \in A} P_0(\{x_i\}) = \sum_{x_i \in A} \underline{P}(\{x_i\}) \\ &\geq \sum_{x_i \in A} \underline{Q}(\{x_i\}) = \sum_{x_i \in A} (1 - \varepsilon')P'_0(\{x_i\}) = (1 - \varepsilon')P'_0(A) = \underline{Q}(A). \end{aligned}$$

We therefore conclude that $\underline{P}_\varepsilon(A) \geq \underline{Q}(A)$ for every $A \subseteq \mathcal{X}$, so $\underline{P}_\varepsilon$ is the unique optimal outer approximation of \underline{P} in \mathcal{C}_ε . \square

Proof of Proposition 9. Let E be an event such that $d_1(\underline{P}_1, \underline{P}_2) = |\underline{P}_1(E) - \underline{P}_2(E)|$. Since both $\underline{P}_1, \underline{P}_2$ are coherent, there exist $P_1 \in \mathcal{M}(\underline{P}_1), P_2 \in \mathcal{M}(\underline{P}_2)$ such that $P_1(E) = \underline{P}_1(E)$ and $P_2(E) = \underline{P}_2(E)$. Then, it holds that:

$$d_3(\underline{P}_1, \underline{P}_2) \geq \|P_1 - P_2\| \geq |P_1(E) - P_2(E)| = |\underline{P}_1(E) - \underline{P}_2(E)| = d_1(\underline{P}_1, \underline{P}_2),$$

whence $d_3 \geq d_1$. \square

The following lemma establishes a property of the lower envelope \underline{Q} of the Weber set of a coherent lower probability \underline{P} , that we shall use in the proof of Proposition 10.

Lemma 15. *Let $|\mathcal{X}| = n \geq 2$, \underline{P} be a coherent lower probability on $\mathcal{P}(\mathcal{X})$ and $A \in \mathcal{P}(\mathcal{X})$. If $|A| \in \{1, n - 1\}$ then:*

$$\underline{Q}(A) = \min_{\sigma \in S_n} P_\sigma(A) = \underline{P}(A).$$

Proof. We consider first of all the case $|A| = 1$. Assume then that $A = \{x_i\}$, and take a generic permutation $\sigma \in S_n$. Using Eq. (19) and the superadditivity (C4) of \underline{P} , we have:

- If $i = \sigma(1)$, then $P_\sigma(\{x_i\}) = \underline{P}(\{x_i\})$.

- If $i = \sigma(k)$ for some $k > 1$, then:

$$\begin{aligned} P_\sigma(\{x_i\}) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k)}\}) - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k-1)}\}) \\ &\geq \underline{P}(\{x_{\sigma(k)}\}) = \underline{P}(\{x_i\}). \end{aligned}$$

Thus, $\underline{Q}(\{x_i\}) = \underline{P}(\{x_i\})$.

Consider now the case $|A| = n - 1$. To simplify the notation, suppose $A = \{x_2, \dots, x_n\}$. Applying Eq. (19) and with some simple computations, we obtain that:

- If $1 = \sigma(n)$, then

$$P_\sigma(A) = P_\sigma(\{x_2\}) + \dots + P_\sigma(\{x_n\}) = \underline{P}(\{x_2, \dots, x_n\}) = \underline{P}(A).$$

- If $1 = \sigma(1)$, it holds that:

$$P_\sigma(A) = 1 - \underline{P}(\{x_{\sigma(1)}\}) = \overline{P}(A) \geq \underline{P}(A).$$

- If $1 = \sigma(k)$, for some $1 < k < n$, it holds that:

$$\begin{aligned} P_\sigma(A) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k-1)}\}) + 1 - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k)}\}) \\ &\geq \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k-1)}, x_{\sigma(k+1)}, \dots, x_{\sigma(n)}\}) = \underline{P}(A), \end{aligned}$$

where the inequality holds from (C6).

We conclude that $\underline{Q}(A) = \underline{P}(A)$. □

Proof of Proposition 10. It suffices to consider the case of $|\mathcal{X}| = 4$, since for cardinalities smaller than four any coherent lower probability is 2-monotone.

By construction, \underline{Q} is an outer approximation of \underline{P} . From Lemma 15, we also know that $\underline{Q}(A) = \underline{P}(A)$ for $|A| = 1$ or $|A| = 3$. Let us see what happens with events of cardinality 2. Assume without loss of generality that $A = \{x_1, x_2\}$; the other cases follow by analogy.

We shall determine the value of $P_\sigma(A)$ for the different permutations of \mathcal{X} . It is not difficult to see that the only differentiating element within all permutations is the relative position of $\{x_1, x_2\}$ in $(\sigma(1), \sigma(2), \sigma(3), \sigma(4))$, in the sense that if, for instance, they are in the second and fourth positions, the conclusions we draw on $\sigma = (3, 1, 4, 2)$ are analogous to those on $(3, 2, 4, 1)$, $(4, 1, 3, 2)$ and $(4, 2, 3, 1)$. For this reason, we depict a few cases only, each representative of one of the possible alternatives.

- If σ satisfies $\{\sigma(1), \sigma(2)\} = \{1, 2\}$, then $P_\sigma(A) = \underline{P}(A)$.

- If σ satisfies $\{\sigma(3), \sigma(4)\} = \{1, 2\}$, then $P_\sigma(A) = \bar{P}(A) \geq \underline{P}(A)$.
- If $\sigma = (1, 3, 2, 4)$, then:

$$P_\sigma(A) = \underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}) - \underline{P}(\{x_1, x_3\}).$$

This means that $P_\sigma(A) < \underline{P}(A)$ if and only if

$$\underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}) < \underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_1, x_2\}),$$

or in other words if and only if \underline{P} violates the 2-monotonicity for the events $A = \{x_1, x_2\}$ and $B = \{x_1, x_3\}$.

- If $\sigma = (1, 3, 4, 2)$, then $P_\sigma(A) = \underline{P}(\{x_1\}) + \bar{P}(\{x_2\}) \geq \underline{P}(A)$, where the inequality follows from property (C5).
- If $\sigma = (3, 1, 2, 4)$, then:

$$\begin{aligned} P_\sigma(A) &= \underline{P}(\{x_1, x_3\}) - \underline{P}(\{x_3\}) + \underline{P}(\{x_1, x_2, x_3\}) - \underline{P}(\{x_1, x_3\}) \\ &= \underline{P}(\{x_1, x_2, x_3\}) - \underline{P}(\{x_3\}) \geq \underline{P}(\{x_1, x_2\}), \end{aligned}$$

where the inequality follows from the super-additivity (C4) of \underline{P} .

- If $\sigma = (3, 1, 4, 2)$, then:

$$P_\sigma(A) = \underline{P}(\{x_1, x_3\}) - \underline{P}(\{x_3\}) + \bar{P}(\{x_2\}).$$

This means that $P_\sigma(A) < \underline{P}(A)$ if and only if:

$$\underline{P}(\{x_1, x_3\}) + \bar{P}(\{x_2\}) < \underline{P}(\{x_3\}) + \underline{P}(\{x_1, x_2\}),$$

but from the properties (C4) and (C5) of a coherent lower probability it holds that:

$$\underline{P}(\{x_1, x_2\}) + \underline{P}(\{x_3\}) \leq \underline{P}(\{x_1, x_2, x_3\}) \leq \underline{P}(\{x_1, x_3\}) + \bar{P}(\{x_2\}),$$

so the inequality cannot occur in the reverse order.

We conclude that $\underline{Q}(A) < \underline{P}(A)$ in those events of cardinality 2 where \underline{P} violates the 2-monotonicity with some other event B such that $|A \cap B| = 1$.

It only remains to see that \underline{Q} is 2-monotone. Let $W(\underline{Q})$ the Weber set of \underline{Q} . Then \underline{Q} is 2-monotone if and only if $\underline{Q} = \min \text{conv}(W(\underline{P})) := \underline{Q}'$.

Reasoning as before, $\underline{Q}'(A) = \underline{Q}(A)$ for events A of cardinality 1 or 3. Assume that there exists an event A of cardinality 2, $A = \{x_1, x_2\}$, such

that $\underline{Q}'(A) < \underline{Q}(A)$. This means that \underline{Q} violates the 2-monotonicity for A and another event B of cardinality 2. Assume without loss of generality that $B = \{x_1, x_3\}$. Then:

$$\begin{aligned} & \underline{Q}(\{x_1, x_2\}) + \underline{Q}(\{x_1, x_3\}) \\ & > \underline{Q}(\{x_1, x_2, x_3\}) + \underline{Q}(\{x_1\}) = \underline{P}(\{x_1, x_2, x_3\}) + \underline{P}(\{x_1\}). \end{aligned} \quad (28)$$

On the other hand, by definition $\underline{Q} \leq P_\sigma$ for any permutation σ . Using this fact with $\sigma = (1, 3, 2, 4)$:

$$\begin{aligned} & \underline{Q}(\{x_1, x_2\}) + \underline{Q}(\{x_1, x_3\}) \leq P_\sigma(\{x_1, x_2\}) + \underline{P}(\{x_1, x_3\}) \\ & = \underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}) - \underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_1, x_3\}) \\ & = \underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}). \end{aligned}$$

From this inequality and Eq. (28), we obtain:

$$\underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}) > \underline{P}(\{x_1\}) + \underline{P}(\{x_1, x_2, x_3\}),$$

a contradiction.

We therefore conclude that $\underline{Q}(A) = \underline{Q}'(A)$ also for events of cardinality 2, and as a consequence that \underline{Q} is 2-monotone. \square

Proof of Proposition 12. Let \underline{Q} be a 2-monotone and undominated lower prevision that outer approximates \underline{P} . Then, its restriction to events \underline{Q}' satisfies:

$$\underline{Q}'(A) = \underline{Q}(I_A) \leq \underline{P}(I_A) = \underline{P}'(A) \quad \forall A \subseteq \mathcal{X}.$$

Also, \underline{Q}' is 2-monotone because \underline{Q} is 2-monotone. Therefore, \underline{Q}' is a 2-monotone outer approximation of \underline{P}' .

If there is some other 2-monotone outer approximation \underline{Q}'_1 of \underline{P}' such that $\underline{Q}' \preceq \underline{Q}'_1 \leq \underline{P}'$, then its natural extension \underline{E}_1 , that is its only 2-monotone extension, will satisfy $\underline{E}' = \underline{Q} \leq \underline{E}_1 \leq \underline{E} \leq \underline{P}$, since \underline{P} is a coherent extension of \underline{P}' and so it dominates its natural extension \underline{E} , and also taking into account that the procedure of natural extension preserves dominance. This means that $\underline{Q} \preceq \underline{E}_1 \leq \underline{P}$, a contradiction with \underline{Q} being undominated.

Conversely, if \underline{Q}' is a 2-monotone and undominated lower probability that outer approximates \underline{P}' , its natural extension \underline{E}' will satisfy:

- a) \underline{E}' is 2-monotone, from [29].
- b) $\underline{E}' \leq \underline{E} \leq \underline{P}$, since \underline{P} dominates the natural extension of its restriction to events and natural extension preserves dominance.

- c) If $\underline{E}' \leq \underline{E}_1 \leq \underline{P}$ for some other 2-monotone outer approximation of \underline{P} , then $\underline{Q}' \leq \underline{Q}'_1 \leq \underline{P}'$, where \underline{Q}'_1 is the restriction of \underline{E}_1 to events. Thus, $\underline{Q}' = \underline{Q}'_1$, whence $\underline{E}' = \underline{E}_1$ because a 2-monotone lower probability on $\mathcal{P}(\mathcal{X})$ has a unique 2-monotone extension to $\mathcal{L}(\mathcal{X})$.

We therefore conclude that \underline{E}' is also a 2-monotone and undominated lower prevision that outer approximates \underline{P} . \square

References

- [1] T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Wiley, 2014.
- [2] P. Baroni and P. Vicig. An uncertainty interchange format with imprecise probabilities. *International Journal of Approximate Reasoning*, 40:147–180, 2005.
- [3] A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In T. Augustin, F. Coolen, S. Moral, and M. Troffaes, editors, *ISIPTA '09 – Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 61–70, Durham, United Kingdom, 2009. SIPTA.
- [4] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17(3):263–283, 1989.
- [5] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.
- [6] G. Coletti, D. Petturiti, and B. Vantaggi. Conditional submodular coherent risk measures. In J. Medina, M. Ojeda-Aciego, J.L. Verdegay, D.A. Pelta, I.P. Cabrera, B. Bouchon-Meunier, and R.R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*, pages 239–250, Cham, 2018. Springer International Publishing.
- [7] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.

- [8] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n -Monotone exact functionals. *Journal of Mathematical Analysis and Applications*, 347:143–156, 2008.
- [9] B. de Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Chichester, 1974–1975.
- [10] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.
- [11] S. Destercke. Independence and 2-monotonicity: Nice to have, hard to keep. *International Journal of Approximate Reasoning*, 54(4):478–490, 2013.
- [12] D. Dubois and H. Prade. Consonant approximations of belief functions. *International Journal of Approximate Reasoning*, 4(5-6):419–449, 1990.
- [13] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.
- [14] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In Ph. Smets, editor, *Handbook on Defeasible Reasoning and Uncertainty Management Systems*, volume 1: Quantified Representation of Uncertainty and Imprecision, pages 169–226. Kluwer Academic Publishers, Dordrecht, 1998.
- [15] M. Grabisch and P. Miranda. Exact bounds of the Möbius inverse of monotone set functions. *Discrete Applied Mathematics*, 186:7–12, 2015.
- [16] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [17] T. Ichiishi. Supermodularity: Applications to convex games and to the greedy algorithm for LP. *Journal of Economic Theory*, 25:283–286, 1981.
- [18] D. A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.
- [19] I. Montes, E. Miranda, and S. Destercke. A study of the Pari-Mutuel Model from the point of view of imprecise probabilities. In A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, vol 62, pages 229–240, 2017.

- [20] I. Montes, E. Miranda, and P. Vicig. Approximations of coherent lower probabilities by 2-monotone capacities. In J. Medina, M. Ojeda-Aciego, J.L. Verdegay, D.A. Pelta, I.P. Cabrera, B. Bouchon-Meunier, and R.R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*, pages 214–225, Cham, 2018. Springer International Publishing.
- [21] I. Montes, E. Miranda, and P. Vicig. Outer approximations of coherent lower probabilities using belief functions. In *Proceedings of the 5th International Conference on Belief Functions*, 2018.
- [22] S. Moral. Divergence measures and approximate algorithms for valuation based systems. In J. Medina, M. Ojeda-Aciego, J.L. Verdegay, I. Perfilieva, B. Bouchon-Meunier, and R.R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, pages 591–602, Cham, 2018. Springer International Publishing.
- [23] H. T. Nguyen, N. T. Nguyen, and T. Wang. On capacity functionals in interval probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:359–377, 1997.
- [24] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51:1145–1158, 2010.
- [25] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
- [26] L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.
- [27] D. Škulj. Errors bounds for finite approximations of coherent lower previsions. In A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 289–300, 2017.
- [28] P. Walley. Coherent lower (and upper) probabilities. Technical Report Statistics Research Report 22, University of Warwick, Coventry, 1981.
- [29] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.