Inner approximations of coherent lower probabilities and their application to decision making problems

Enrique Miranda[1] · **Ignacio Montes[1]** · **Andrés Presa[2]**

Abstract

We consider a decision making problem under imprecision, where the probabilistic information is given in terms of a set of probability measures, and where finding the optimal alternative(s) may be difficult. To ease the computation, we propose to transform the initial model into another one that (1) belongs to some subclass with better mathematical properties, such as supermodularity or complete monotonicity; (2) is at least as informative as the original model, while being as close as possible to it. We show that the problem can be approached in terms of linear or quadratic programming and that it can be connected with the one of determining the incenter of a credal set. Finally, we compare the solutions of a decision making problem with the initial and the transformed models and illustrate how our approach can be applied in a decision making problem under severe uncertainty.

Keywords Coherent lower probabilities · 2-Monotone capacities · Belief functions · Distortion models · Decision making

1 Introduction

Since the pioneering work of Anscombe and Aumann (1963), Savage (1954) and Von Neumann and Morgestern (1947), probability measures are the most widespread tool in decision making problems under uncertainty. Nevertheless, due to a number of reasons such as the lack of information or the low quality of the data, eliciting a probability measure modelling the uncertainty may be at times a difficult problem. This has lead to the development of alter-

Enrique Miranda, Ignacio Montes and Andrés Presa have contributed equally to this work

✉ Enrique Miranda
  mirandaenrique@uniovi.es

  Ignacio Montes
  imontes@uniovi.es

  Andrés Presa
  presa@strw.leidenuniv.nl

[1] Department of Statistics and O.R., University of Oviedo, Federico García Lorca, 33007 Oviedo, Asturias, Spain

[2] Leiden Observatory, Leiden University, Niels Bohrweg 2, Leiden 2333, The Netherlands

natives that are better suited to deal with these situations. Indeed, quoting a recent publication in ANOR (Keith & Ahner, 2021, pp. 319–320),

> …Over the past several decades, various theories have been developed that generalize the theory of probability to address aspects of uncertainty that are difficult or impossible to model in standard probability theory.

These alternative theories are usually referred to as instances of *imprecise probability models* (Augustin et al., 2014), and include for instance belief functions (Shafer, 1976), possibility measures (Dubois & Prade, 1988), coherent lower probabilities (Walley, 1991) or submodular capacities (Choquet, 1953); they have also appeared under the name non-additive measures or games in coalitional game theory (Grabisch 2016).

Imprecise probability models have been applied extensively in the decision making context. According to (Grabisch 2016, p. 28),

> …The fields of decision theory and game theory seem to be the privileged area for the application of games and capacities.

In fact, several extensions of the expected utility paradigm that allow to model uncertainty with non-additive measures have been proposed [see for instance Gilboa and Schmeidler (1989), Klibanoff et al. (2005), Sarin and Wakker (1992) and the survey in Troffaes (2007)]. There have also been applications of imprecise probabilities in decision making problems within the context of machine learning (Mattei et al., 2020), environmental engineering (Sahlin et al., 2021) or signal processing (de Angelis et al., 2023), just to name a few. While these references illustrate the interest and generality of imprecise probability models, we should also signal that this greater generality also encompasses a greater complexity; thus, a balance must be found between the expressiveness of the model and its tractability.

Coherent lower probabilities are the starting point of this paper. In addition to having an epistemic interpretation as lower envelopes of a closed and convex set of probability measures, they have the advantage of including as particular cases most other models within imprecise probability theory; therefore, the properties established for coherent lower probabilities immediately apply to submodular capacities or belief functions, for instance. However, their generality comes with a price: for instance, there is not an easy procedure for determining the extreme points of the associated set of probabilities, nor is there a unique extension to expectation operators. This hampers the use of coherent lower probabilities in decision making problems (Troffaes, 2007), where the computation of the optimal alternatives could be involved.

To overcome this issue, it may be sensible to look for transformations of a given coherent lower probability into another one that is close and that at the same time belongs to a class with better mathematical properties. Indeed, in past works (Miranda et al., 2021; Montes et al., 2018, 2019) we considered *outer approximations* of a coherent lower probability, leading to a transformed model less informative than the original one. Here we move in the opposite direction, and look for transformations that shrink the credal set and where the associated lower probability belongs to a subfamily of interest. We shall call these *inner approximations*, since their associated set of probability measures will be included in the set of those that are compatible with the original lower probability.

Beyond decision making under uncertainty, there are several contexts where an inner approximation can be of interest: we may consider for instance the problem of selecting a representative element within the credal set associated with the coherent lower probability (Jaffray 1995; Weber 1988), or aim to reduce the imprecision inherent to the model so as

to make more informative inferences (Antonucci et al., 2015; Dubois et al., 1993). Recently, approximations of coherent lower probabilities in terms of belief functions have been used in statistical matching (Petturiti & Vantaggi, 2022), conditional coherent risk measures (Petturiti & Vantaggi, 2019) and for correcting incoherent beliefs (Petturiti & Vantaggi, 2022) [see also (Cinfrignini et al., 2023; Petturiti & Vantaggi, 2020)].

For these reasons, in this paper we shall investigate the problem of transforming a coherent lower probability into an inner approximation that belongs to some subfamily of interest. Specifically, for 2-monotone capacities and belief functions, we shall show in Sect. 3 that some interesting inner approximations may be obtained by means of linear and quadratic programming, and shall compare the properties of the transformed models with the ones obtained in Miranda et al. (2021) and Montes et al. (2018, 2019) as outer approximations. Next, in Sect. 4 we shall analyse the particular case of distortion models, where we shall characterise the existence of an inner approximation and the set of optimal ones according to some predetermined distance. In particular, in Sect. 4.4 we shall explore the connection between the problem at hand and that of determing the incenter of a credal set, following the ideas in Miranda and Montes (2023) and creating also a bridge with the problem of finding solutions of coalitional games. In Sect. 5 we shall compare the performance of the original and the transformed model with respect to different optimality criteria in the context of decision making with sets of probabilities (Troffaes, 2007). Finally, in Sect. 6 we apply these results on the example of decision making under severe uncertainty from Jansen et al. (2018). We conclude the paper with some additional comments in Sect. 7. To ease the reading, proofs as well as some supporting results have been gathered in an Appendix.

A preliminary version of this paper was presented at the the 19th International Conference on Information Processing and Management of Uncertainty (IPMU'2022) (Miranda et al., 2022). This expanded version includes the proofs of all the mathematical results, an extended discussion of the implications of using inner approximations in a decision making problem, additional examples, and an illustration on a decision making problem.

2 Preliminary concepts

Let $\mathcal{X}$ be a finite possibility space with cardinality n, and let $\mathcal{P}(\mathcal{X})$ denote its power set. We call *lower probability* a function $\underline{P} : \mathcal{P}(\mathcal{X}) \to [0, 1]$ that is monotone ($A \subseteq B \Rightarrow \underline{P}(A) \leq \underline{P}(B)$) and normalised ($\underline{P}(\emptyset) = 0$, $\underline{P}(\mathcal{X}) = 1$). Its conjugate *upper* probability is given by $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \mathcal{X}$.

For a lower probability $\underline{P}$, the associated set of dominating probabilities, or *credal set*, is given by:

$$\mathcal{M}(\underline{P}) = \{P \text{ probability measure } \mid P(A) \geq \underline{P}(A) \; \forall A \subseteq \mathcal{X}\}.$$

Following (Walley, 1991), we shall say that $\underline{P}$ *avoids sure loss* when $\mathcal{M}(\underline{P}) \neq \emptyset$, and that it is *coherent* when it is the lower envelope of $\mathcal{M}(\underline{P})$: $\underline{P}(A) = \min_{P \in \mathcal{M}(\underline{P})} P(A)$ for every $A \subseteq \mathcal{X}$.

As particular instances of coherent lower probabilities we have those that are 2-*monotone*, meaning that $\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B)$ for any $A, B \subseteq \mathcal{X}$. They are also referred to as *supermodular* or *convex* in the literature. On the other hand, a coherent lower probability is said to be *completely monotone*, or a *belief function*, when

$$\underline{P}\left(\cup_{i=1}^k A_i \right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \underline{P}\left(\cap_{i \in I} A_i \right)$$

for every $A_1, \ldots, A_k$ in $\mathcal{P}(\mathcal{X})$ and every $k \in \mathbb{N}$. We denote by $\mathcal{C}_2$ and $\mathcal{C}_\infty$ the families of 2-monotone lower probabilities and belief functions, respectively. The above definitions imply that $\mathcal{C}_\infty \subset \mathcal{C}_2$.

Any lower probability $\underline{P}$ can be alternatively expressed using the *Möbius* transformation, that is given by:

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

conversely, $m_{\underline{P}}$ allows to retrieve the initial lower probability by:

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

It is worth mentioning that the Möbius transformation is not only an equivalent representation of a lower probability, but it can be also used to characterise 2- or complete-monotonicity. Indeed, $\underline{P}$ is a 2-monotone lower probability if and only if its Möbius transformation $m_{\underline{P}}$ satisfies (Chateauneuf & Jaffray, 1989)

$$\sum_{A \subseteq \mathcal{X}} m_{\underline{P}}(A) = 1, \quad m_{\underline{P}}(\emptyset) = 0; \tag{2monot.1}$$

$$\sum_{\{x_i, x_j\} \subseteq B \subseteq A} m_{\underline{P}}(B) \geq 0, \quad \forall A \subseteq \mathcal{X}, \forall x_i, x_j \in A, x_i \neq x_j; \tag{2monot.2}$$

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

and it is completely monotone if and only if it satisfies (2monot.1) and

$$m_{\underline{P}}(A) \geq 0 \quad \forall A \subseteq \mathcal{X}. \tag{C-monot.}$$

3 Inner approximations of lower probabilities

3.1 Summary of the results on outer approximations

In previous papers (Miranda et al., 2021; Montes et al., 2018, 2019) we investigated the problem of outer approximating a coherent lower probability by means of a 2- or completely monotone lower probability. The definition of *outer approximation* goes back to Bronevich and Augustin (2009).

Definition 1 (Bronevich & Augustin, 2009) Let $\underline{P}$ be a coherent lower probability and let $\mathcal{C}$ be a class of coherent lower probabilities. $\underline{Q} \in \mathcal{C}$ is called an *outer approximation* of $\underline{P}$ in $\mathcal{C}$ if $\underline{Q}(A) \leq \underline{P}(A)$ for every $A \subseteq \mathcal{X}$. Moreover, $\underline{Q}$ is an *undominated* outer approximation if there is no other $\underline{Q}' \in \mathcal{C}$ such that $\underline{Q} \lneqq \underline{Q}' \leq \underline{P}$.

In terms of credal sets, $\underline{Q}$ is an outer approximation of $\underline{P}$ when $\mathcal{M}(\underline{P}) \subseteq \mathcal{M}(\underline{Q})$, and it is undominated if there is no other $\underline{Q}' \in \mathcal{C}$ such that $\mathcal{M}(\underline{P}) \subseteq \mathcal{M}(\underline{Q}') \subsetneq \mathcal{M}(\underline{Q})$.

The quest for computing outer approximations of a coherent lower probability $\underline{P}$ seeks to replace $\underline{P}$ with a model with better mathematical properties, such as 2-monotonicity, and

such that any element of $\mathcal{M}(\underline{P})$ is also compatible with the new model. This last requirement is sensible if we give $\underline{P}$ an epistemic interpretation, as a model for the imprecise knowledge of a probability measure P_0: if all we know about P_0 is that it belongs to $\mathcal{M}(\underline{P})$, we would like all the potential candidates to be also compatible with the transformed model. In addition, this new model should be as close as possible to the original one, so that their respective inferences are similar. A necessary condition in this regard is that the outer approximation is undominated.

To obtain undominated outer approximations, in Miranda et al. (2021) and Montes et al. (2018, 2019) we pursued a number of paths. The primal one was based on minimising the *Baroni and Vicig distance* (BV-distance, for short) (Baroni & Vicig, 2005) between the initial model and the outer approximation:

$$d_{\mathrm{BV}}\big(\underline{P}, \underline{Q}\big) = \sum_{E \subseteq \mathcal{X}} |\underline{P}(E) - \underline{Q}(E)| = \sum_{E \subseteq \mathcal{X}} \left| \underline{P}(E) - \sum_{B \subseteq E} m_{\underline{Q}}(B) \right|, \qquad (1)$$

which measures the amount of imprecision added to the model when replacing $\underline{P}$ by $\underline{Q}$. Another possibility is to consider the *quadratic* distance between the original and the transformed model:

$$d_q\big(\underline{P}, \underline{Q}\big) = \sum_{E \subseteq \mathcal{X}} \big(\underline{P}(E) - \underline{Q}(E)\big)^2 = \sum_{E \subseteq \mathcal{X}} \left(\underline{P}(E) - \sum_{B \subseteq E} m_{\underline{Q}}(B) \right)^2. \qquad (2)$$

Using either of these distances, we can set up an optimisation problem that gives us outer approximations.

Proposition 1 (Montes et al., 2018, 2019) *Let $\underline{P}$ be a coherent lower probability, and consider the condition*

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

(i) Let $\mathcal{C}_2^{oa}(\underline{P})$ be the set of coherent lower probabilities satisfying conditions (2monot.1)–(2monot.3) and (2monot.4). The linear programming problem of minimising Eq. (1) in $\mathcal{C}_2^{oa}(\underline{P})$ has optimal solutions that are undominated outer approximations of $\underline{P}$ in $\mathcal{C}_2$. Similarly, the quadratic problem of minimising Eq. (2) in $\mathcal{C}_2^{oa}(\underline{P})$ has a unique optimal solution that is an undominated outer approximation of $\underline{P}$ in $\mathcal{C}_2$.

(ii) Let $\mathcal{C}_\infty^{oa}(\underline{P})$ be the set of coherent lower probabilities satisfying conditions (2monot.1), (C-monot.) and (2monot.4). The linear programming problem of minimising Eq. (1) in $\mathcal{C}_\infty^{oa}(\underline{P})$ has optimal solutions that are undominated outer approximations of $\underline{P}$ in $\mathcal{C}_\infty$. Similarly, the quadratic problem of minimising Eq. (2) in $\mathcal{C}_\infty^{oa}(\underline{P})$ has a unique optimal solution that is an undominated outer approximation of $\underline{P}$ in $\mathcal{C}_\infty$.

Concerning the undominated outer approximations in $\mathcal{C}_2$, we have proven that the linear programming approach in Proposition 1(i) may have infinite different solutions (Montes et al., 2018, Ex.1), that the undominated outer approximations coincide with $\underline{P}$ on singletons and on events of cardinality $n-1$ (Montes et al., 2018, Prop.2), and that the optimal solution of the quadratic problem in Proposition 1(i) may not be an optimal solution of the linear problem.

With respect to $\mathcal{C}_\infty$, there exist undominated outer approximations that are not optimal solutions of the linear programming problem in Proposition 1(ii) (Montes et al., 2019, Ex.2) and the undominated outer approximations may not coincide with $\underline{P}$ on singletons or on events of cardinality $n-1$ (Montes et al., 2019, Ex.4).

While the linear programming approach has the advantage of using an in our view more natural distance between the initial and the transformed model, it also has the drawback of not providing a unique solution. The opposite holds for the quadratic approach: it gives a unique solution but the use of the quadratic distance is less natural in this context. This led us in Miranda et al. (2021) to combine the two approaches so as to get the best from both.

Proposition 2 (Miranda et al., 2021) *Let $\underline{P}$ be a coherent lower probability.*

(i) *The quadratic programming problem of minimising Eq. (2) in $\mathcal{C}_2^{oa}(\underline{P})$ subject also to:*

$$d_{\mathrm{BV}}(\underline{P}, \underline{Q}) = \min_{\underline{Q}' \in \mathcal{C}_2^{oa}} d_{\mathrm{BV}}(\underline{P}, \underline{Q}') \qquad \text{(2.monot-BV)}$$

has a unique optimal solution that is an undominated outer approximation in $\mathcal{C}_2$.

(ii) *The quadratic programming problem of minimising Eq. (2) in $\mathcal{C}_\infty^{oa}(\underline{P})$ subject also to:*

$$d_{\mathrm{BV}}(\underline{P}, \underline{Q}) = \min_{\underline{Q}' \in \mathcal{C}_\infty^{oa}} d_{\mathrm{BV}}(\underline{P}, \underline{Q}') \qquad \text{(C.monot-BV)}$$

has a unique optimal solution that is an undominated outer approximation in $\mathcal{C}_\infty$.

In other words, a possible approach to choose an outer approximation in $\mathcal{C}_2$ or $\mathcal{C}_\infty$ is to minimise the quadratic distance among those outer approximations minimising the BV-distance. Other possibilities were discussed in Miranda et al. (2021).

3.2 Inner approximations

The problem of inner approximating a coherent lower probability was superficially discussed in Montes et al. (2018, Sec. 7) as a sort of dual approach to that of outer approximations. In this subsection, we analyse the problem in detail and compare the features of both approaches.

Definition 2 (Montes et al., 2018, Sec. 7) Let $\underline{P}$ be a coherent lower probability and let $\mathcal{C}$ be a class of coherent lower probabilities. $\underline{Q} \in \mathcal{C}$ is called an *inner approximation* of $\underline{P}$ in $\mathcal{C}$ if $\underline{Q}(A) \geq \underline{P}(A)$ for every $A \subseteq \mathcal{X}$. It is said to be a *non-dominating* inner approximation if there is no other $\underline{Q}' \in \mathcal{C}$ such that $\underline{P} \leq \underline{Q}' \lneq \underline{Q}$.

In terms of credal sets, $\underline{Q}$ is an inner approximation of $\underline{P}$ if $\mathcal{M}(\underline{P}) \supseteq \mathcal{M}(\underline{Q})$ and $\underline{Q} \in \mathcal{C}$ is a non-dominating inner approximation of $\underline{P}$ in $\mathcal{C}$ if there is no other $\underline{Q}' \in \mathcal{C}$ such that $\mathcal{M}(\underline{P}) \supseteq \mathcal{M}(\underline{Q}') \supsetneq \mathcal{M}(\underline{Q})$.

Taking inspiration from the work summarised in Sect. 3.1 about the outer approximations, we can easily establish procedures for inner approximating a coherent lower probability $\underline{P}$ by another one $\underline{Q}$ that is 2- or completely monotone; we simply need to replace (2monot.4) by:

$$\sum_{B \subseteq E} m_{\underline{Q}}(B) \geq \underline{P}(E) \qquad \text{(2monot.4-inner)}$$

This leads at once to the following result:

Proposition 3 *Let $\underline{P}$ be a coherent lower probability.*

(i) *Let $\mathcal{C}_2^{ia}(\underline{P})$ be the set of coherent lower probabilities satisfying conditions (2monot.1)–(2monot.3) and (2monot.4-inner). The linear programming problem of minimising Eq. (1) in $\mathcal{C}_2^{ia}(\underline{P})$ has optimal solutions that are non-dominating inner approximations of $\underline{P}$ in $\mathcal{C}_2$. Similarly, the quadratic problem of minimising Eq. (2) in $\mathcal{C}_2^{ia}(\underline{P})$ has a unique optimal solution that is a non-dominating inner approximation of $\underline{P}$ in $\mathcal{C}_2$.*

(ii) Let $\mathcal{C}^{ia}_{\infty}(\underline{P})$ be the set of coherent lower probabilities satisfying conditions (2monot.1), (C-monot.) and (2monot.4-inner). The linear programming problem of minimising Eq. (1) in $\mathcal{C}^{ia}_{\infty}(\underline{P})$ has optimal solutions that are non-dominating inner approximations of $\underline{P}$ in $\mathcal{C}_{\infty}$. Similarly, the quadratic problem of minimising Eq. (2) in $\mathcal{C}^{ia}_{\infty}(\underline{P})$ has a unique optimal solution that is a non-dominating inner approximation of $\underline{P}$ in $\mathcal{C}_{\infty}$.

Example 1 Consider the possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and let $\underline{P}$ be the coherent lower probability given by:

A	$\underline{P}(A)$	$\underline{Q}$	Bel_1	Bel_2	$\underline{Q}'$	Bel_3
$\{x_1\}$	0	0	0.15	0	0.08	0.1
$\{x_2\}$	0	0	0	0	0	0
$\{x_3\}$	0	0	0.05	0	0	0.1
$\{x_4\}$	0	0.2	0.25	0.3	0.12	0.2
$\{x_1, x_2\}$	0	0	0.15	0.1	0.08	0.1
$\{x_1, x_3\}$	0.3	0.3	0.3	0.3	0.3	0.3
$\{x_1, x_4\}$	0.4	0.4	0.4	0.4	0.4	0.4
$\{x_2, x_3\}$	0	0	0.05	0.1	0	0.1
$\{x_2, x_4\}$	0.3	0.3	0.3	0.3	0.3	0.3
$\{x_3, x_4\}$	0.3	0.3	0.3	0.3	0.3	0.3
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5	0.5	0.5	0.5
$\{x_1, x_2, x_4\}$	0.5	0.5	0.5	0.5	0.58	0.5
$\{x_1, x_3, x_4\}$	0.5	0.7	0.55	0.7	0.62	0.6
$\{x_2, x_3, x_4\}$	0.5	0.5	0.5	0.5	0.5	0.5

$\underline{P}$ is coherent because it is the lower envelope of the probability mass functions $(0, 0, 0.5, 0.5)$, $(0.5, 0, 0, 0.5)$, $(0.4, 0.3, 0.3, 0)$ and $(0.2, 0.5, 0.1, 0.2)$. Solving the linear programming problem from Proposition 3 in $\mathcal{C}_2$, we get the optimal solution $\underline{Q}$. Note that $\underline{Q}$ satisfies $\underline{Q}(\{x_4\}) \neq \underline{P}(\{x_4\})$, showing that the non-dominating inner approximations do not necessarily coincide in the singletons with $\underline{P}$. On the other hand, Bel_1, Bel_2 and Bel_3 are different optimal solutions of the linear programming problem in the class $\mathcal{C}_{\infty}$. Observe that Bel_2 dominates $\underline{Q}$; thus non-dominating inner approximations in $\mathcal{C}_{\infty}$ may be dominating if we regard them as elements from $\mathcal{C}_2$.

In the quadratic approach, the non-dominating solution in $\mathcal{C}_2$ is $\underline{Q}'$, while that in $\mathcal{C}_{\infty}$ is Bel_3. The former is not an optimal solution of the linear problem in $\mathcal{C}_2$ while, as we have said Bel_3 is an optimal solution of the linear problem in $\mathcal{C}_{\infty}$. ♦

Example 1 shows that a coherent lower probability may have infinite non-dominating inner approximations in $\mathcal{C}_{\infty}$: in that example, any convex combination of Bel_1, Bel_2, Bel_3 will be a belief function that inner approximates $\underline{P}$ and is non-dominating (because it is at the minimum BV-distance with $\underline{P}$). Let us show that this may also be the case in $\mathcal{C}_2$.

Example 2 Consider the possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, the probability mass functions $P_1 = (0.25, 0.25, 0.25, 0.25)$ and $P_2 = (0.2, 0.2, 0.3, 0.3)$, and the coherent lower probability $\underline{P}$ that is the lower envelope of $\{P_1, P_2\}$. $\underline{P}$ is not 2-monotone, since:

$$0.5 + 0.5 = \underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_1, x_4\}) > \underline{P}(\{x_1, x_3, x_4\}) + \underline{P}(\{x_1\}) = 0.75 + 0.2.$$

Let us prove that P_1, P_2 are non-dominating inner approximations of $\underline{P}$ in $\mathcal{C}_2$; we shall establish it for P_1, the proof for P_2 being similar.

Assume that there exists a 2-monotone inner approximation $\underline{Q}$ of $\underline{P}$ such that $\underline{P} \leq \underline{Q} \lneqq P_1$. Then, there must be some event A such that $\underline{Q}(A) < P_1(A)$. Considering the events where P_1 and $\underline{P}$ (and consequently also $\underline{Q}$) agree on, A must be one of $\{x_1\}, \{x_2\}, \{x_1, x_2\}, \{x_1, x_2, x_3\}$ or $\{x_1, x_2, x_4\}$. By 2-monotonicity, we have that

$$\underline{Q}(\{x_1\}) \geq \underline{Q}(\{x_1, x_3\}) + \underline{Q}(\{x_1, x_4\}) - \underline{Q}(\{x_1, x_3, x_4\}) = 0.25, \text{ and similarly}$$

$$\underline{Q}(\{x_2\}) \geq \underline{Q}(\{x_2, x_3\}) + \underline{Q}(\{x_2, x_4\}) - \underline{Q}(\{x_2, x_3, x_4\}) = 0.25.$$

Since any coherent lower probability is super-additive (Walley, 1991, Sect. 2.7.4), we obtain

$$\underline{Q}(\{x_1, x_2, x_3\}) \geq \underline{Q}(\{x_1\}) + \underline{Q}(\{x_2\}) + \underline{Q}(\{x_3\}) = 0.75 = P_1(\{x_1, x_2, x_3\})$$

and similarly $\underline{Q}(\{x_1, x_2, x_4\}) = P_1(\{x_1, x_2, x_4\}) = 0.75$ and $\underline{Q}(\{x_1, x_2\}) = P_1(\{x_1, x_2\}) = 0.5$. Therefore, $\underline{Q} = P_1$, a contradiction. $\blacklozenge$

These two examples raise the need of some criteria to select a non-dominating inner approximation of the coherent lower probability. We may follow here the same approach as in Miranda et al. (2021): to choose the one minimising the quadratic distance among those that minimise the BV-distance.

Proposition 4 *Let $\underline{P}$ be a coherent lower probability on $\mathcal{P}(\mathcal{X})$.*

(i) *The quadratic programming problem of minimising Eq. (2) in $\mathcal{C}_2^{ia}(\underline{P})$ subject also to (2.monot-BV) has a unique solution that is a non-dominating inner approximation in $\mathcal{C}_2$.*
(ii) *The quadratic programming problem of minimising Eq. (2) in $\mathcal{C}_\infty^{ia}(\underline{P})$ subject also to (C.monot-BV) has a unique solution that is a non-dominating inner approximation in $\mathcal{C}_\infty$.*

Example 3 If we apply this idea to the coherent lower probability in Example 1, $\underline{Q}$ and Bel_3 are the optimal inner approximations minimising the quadratic distance among those minimising the BV-distance in $\mathcal{C}_2$ and $\mathcal{C}_\infty$, respectively. $\blacklozenge$

In what follows, we investigate if for some subfamilies of interest of $\mathcal{C}_2$ it is possible to characterise the inner approximations that minimise the BV-distance with respect to the original model. In this respect, the following result shows that the process of obtaining inner approximations can be made iterative. For this, given a family $\mathcal{C}$ of coherent lower probabilities and a coherent lower probability $\underline{P}$, we shall denote by $\tilde{\mathcal{C}}^{ia}(\underline{P})$ the class of non-dominating inner approximations of $\underline{P}$ in $\mathcal{C}$, and by $\mathcal{C}_{BV}^{ia}(\underline{P})$ the subclass of those that minimise the BV-distance with respect to $\underline{P}$. It follows that $\mathcal{C}_{BV}^{ia}(\underline{P}) \subseteq \tilde{\mathcal{C}}^{ia}(\underline{P})$.

Proposition 5 *Let $\underline{P}$ be a coherent lower probability, and consider two classes of coherent lower probabilities $\mathcal{C}$ and $\mathcal{C}'$ such that $\mathcal{C}' \subseteq \mathcal{C}$.*

(i) *If $\underline{Q} \in \tilde{\mathcal{C}}'^{ia}(\underline{P})$, then there exists some $\underline{P}' \in \tilde{\mathcal{C}}^{ia}(\underline{P})$ such that $\underline{Q} \in \tilde{\mathcal{C}}'^{ia}(\underline{P}')$.*
(ii) *If moreover $\underline{Q} \in \mathcal{C}_{BV}'^{ia}(\underline{P})$, then also $\underline{Q} \in \mathcal{C}_{BV}'^{ia}(\underline{P}')$ for some $\underline{P}' \in \mathcal{C}_{BV}^{ia}(\underline{P})$.*

4 Inner approximations with distortion models and incenters

In this section, we investigate the inner approximations of coherent lower probabilities by means of some *distortion model* (Destercke et al., 2022; Montes et al., 2020a, b). These are imprecise models determined by a probability measure P_0, a distorting function d and a

distortion parameter δ. These three elements allow to define a set of probability measures by means of $B_d^\delta(P_0) = \{P \mid d(P, P_0) \leq \delta\}$. The set $B_d^\delta(P_0)$ is closed and convex whenever d is continuous and convex (Montes et al., 2020a, Prop.1).

Several distortion models can be found in the literature, such as the constant odds ratio (Berger, 1990; Pericchi & Walley, 1991; Walley, 1991), the distortion models generated by the L_1 or Kolmogorov distances (Huber, 1981; Montes et al., 2020b), or those obtained through increasing transformations of a probability measure (Bronevich, 2007). In this paper, we focus on the linear vacuous (Walley, 1991), pari mutuel (Montes et al., 2019; Pelessoni et al., 2010; Walley, 1991) and total variation models (Seidenfeld & Wasserman, 1993). These classes will be denoted by $\mathcal{C}_{LV}$, $\mathcal{C}_{PMM}$ and $\mathcal{C}_{TV}$. Although there is no inclusion relationship between them, they have a connection with the classes $\mathcal{C}_2$ and $\mathcal{C}_\infty$ from Sect. 3: it holds that any pari-mutuel or total variation model is 2-monotone, but not necessarily completely monotone, while any linear vacuous model satisfies complete monotonicity; in other words, $\mathcal{C}_{LV} \subseteq \mathcal{C}_\infty$ and $\mathcal{C}_{PMM}, \mathcal{C}_{TV} \subseteq \mathcal{C}_2$, but $\mathcal{C}_{PMM}, \mathcal{C}_{TV} \not\subseteq \mathcal{C}_\infty$.

Throughout the section, and for the sake of simplicity, we assume that $\underline{P}(A) \in (0, 1)$ for any $A \neq \emptyset, \mathcal{X}$.

4.1 Linear vacuous model

Let P_0 be a probability measure and $\delta \in (0, 1)$ a distortion parameter. The *linear vacuous model* is given by the coherent lower probability

$$\underline{P}_{LV}(A) = (1 - \delta)P_0(A) \text{ if } A \subset \mathcal{X} \text{ and } \underline{P}_{LV}(\mathcal{X}) = 1;$$

its conjugate coherent upper probability is given by $\overline{P}_{LV}(A) = (1 - \delta)P_0(A) + \delta$ for any $A \neq \emptyset$. It holds that:

$$\overline{P}_{LV}(A) - \underline{P}_{LV}(A) = \delta \quad \forall A \neq \emptyset, \mathcal{X}. \tag{3}$$

The credal set $\mathcal{M}(\underline{P}_{LV})$ is formed by the convex combinations $(1 - \delta)P_0 + \delta P$ of P_0 with another probability measure P, with respective weights $(1-\delta)$ and δ. Thus, we may interpret this model by considering an experiment where the uncertainty model is the probability measure P_0, and where there is a proportion δ of contaminated data, coming from another probability measure P. We refer to Montes et al. (2020a) for a study of the properties of the linear vacuous as a distortion model.

In Montes et al. (2018, Prop. 8), we proved that for any coherent lower probability $\underline{P}$ satisfying $\sum_{i=1}^n \underline{P}(\{x_i\}) > 0$ there is a unique undominated outer approximation in $\mathcal{C}_{LV}$, where P_0 and δ are given by $\delta = 1 - \sum_{j=1}^n \underline{P}(\{x_j\})$ and $P_0(\{x_i\}) = \frac{\underline{P}(\{x_i\})}{1-\delta} \forall i = 1, \ldots, n$. Next, we investigate the inner approximations in $\mathcal{C}_{LV}$. We begin by establishing a necessary and sufficient condition for their existence.

Definition 3 (Miranda & Montes, 2023) A coherent lower probability $\underline{P}$ on $\mathcal{P}(\mathcal{X})$ is called *maximally imprecise* when $\underline{P}(A) < \overline{P}(A)$ for every $A \neq \emptyset, \mathcal{X}$.

While the existence of inner approximations in $\mathcal{C}_2$ or $\mathcal{C}_\infty$ is trivial because any element of the non-empty set $\mathcal{M}(\underline{P})$ is an inner approximation of $\underline{P}$, the same does not apply to particular subfamilies of $\mathcal{C}_2$, such as $\mathcal{C}_{LV}$.

Proposition 6 *Let $\underline{P}$ be a coherent lower probability. There exists a linear vacuous model $\underline{P}_{LV}$ that inner approximates $\underline{P}$ if and only if $\underline{P}$ is maximally imprecise.*

Consider now a maximally imprecise coherent lower probability $\underline{P}$, and let $\underline{P}_{\mathrm{LV}}$ be an inner approximation in $\mathcal{C}_{\mathrm{LV}}$ given by P_0 and δ. Their BV-distance is

$$d_{\mathrm{BV}}(\underline{P}, \underline{P}_{\mathrm{LV}}) = \sum_{A \subseteq \mathcal{X}} |\underline{P}_{\mathrm{LV}}(A) - \underline{P}(A)| = \sum_{A \subset \mathcal{X}} \underline{P}_{\mathrm{LV}}(A) - \sum_{A \subset \mathcal{X}} \underline{P}(A)$$

$$= \sum_{A \subset \mathcal{X}} (1 - \delta) P_0(A) - \sum_{A \subset \mathcal{X}} \underline{P}(A) = (1 - \delta) \sum_{A \subset \mathcal{X}} P_0(A) - \sum_{A \subset \mathcal{X}} \underline{P}(A), \qquad (4)$$

using that $\underline{P}(A) \in (0, 1)$ for every $A \neq \emptyset, \mathcal{X}$ and that $\underline{P}_{\mathrm{LV}}$ is an inner approximation of $\underline{P}$. Since $\sum_{A \subset \mathcal{X}} P_0(A)$ is constant for every probability measure P_0, this distance is minimised when $(1 - \delta)$ is minimised or, equivalently, when δ is maximised. With this idea in mind, we give an example showing that there may be more than one inner approximation in $\mathcal{C}_{\mathrm{LV}}$ minimising the BV-distance:

Example 4 Consider a three-element possibility space $\mathcal{X} = \{x_1, x_2, x_3\}$ and the coherent lower probability $\underline{P}$ given by:

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
$\underline{P}(A)$	0.2	0.05	0.1	0.4	0.4	0.5
$\underline{P}^1_{\mathrm{LV}}(A)$	0.2	0.2	0.3	0.4	0.5	0.5
$\underline{P}^2_{\mathrm{LV}}(A)$	0.2	0.3	0.2	0.5	0.4	0.5

It is coherent because it is the lower envelope of the probability mass functions $(0.2, 0.2, 0.6)$, $(0.2, 0.6, 0.2)$, $(0.35, 0.05, 0.6)$, $(0.5, 0.05, 0.45)$, $(0.3, 0.6, 0.1)$ and $(0.5, 0.4, 0.1)$. Any inner approximation $\underline{P}_{\mathrm{LV}}$ of $\underline{P}$ in $\mathcal{C}_{\mathrm{LV}}$ defined by (P_0, δ) satisfies

$$0.7 = 0.5 + 0.2 = \underline{P}(\{x_1\}) + \underline{P}(\{x_2, x_3\}) \leq \underline{P}_{\mathrm{LV}}(\{x_1\}) + \underline{P}_{\mathrm{LV}}(\{x_2, x_3\})$$
$$= (1 - \delta) P_0(\{x_1\}) + (1 - \delta) P_0(\{x_2, x_3\}) = 1 - \delta,$$

whence $\delta \leq 0.3$. Consider now $P^1_{\mathrm{LV}} = (2/7, 2/7, 3/7)$ and $P^2_{\mathrm{LV}} = (2/7, 3/7, 2/7)$. Together with $\delta = 0.3$, they give rise to $\underline{P}^1_{\mathrm{LV}}$ and $\underline{P}^2_{\mathrm{LV}}$ in the table above, which are then two different elements of $\mathcal{C}_{\mathrm{LV}}$ minimising the BV-distance. ◆

We look then for the largest $\delta > 0$ such that there is some probability measure P_0 such that $(1 - \delta) P_0(A) \geq \underline{P}(A)$, or equivalently, $P_0(A) \geq \frac{\underline{P}(A)}{1-\delta}$ for any $A \subset \mathcal{X}$. Thus, if for some fixed $\delta \in (0, 1)$ we define $\underline{Q}^\delta_{\mathrm{LV}}$ as

$$\underline{Q}^\delta_{\mathrm{LV}}(A) = \frac{\underline{P}(A)}{1 - \delta} \quad \text{if } A \neq \mathcal{X}, \quad \text{and} \quad \underline{Q}^\delta_{\mathrm{LV}}(\mathcal{X}) = 1 \qquad (5)$$

it is equivalent to look for the largest δ such that $\mathcal{M}(\underline{Q}^\delta_{\mathrm{LV}}) \neq \emptyset$, i.e., the largest δ such that $\underline{Q}^\delta_{\mathrm{LV}}$ avoids sure loss. Consider the following set:

$$\Lambda_{\mathrm{LV}} = \left\{ \delta \in (0, 1) \mid \mathcal{M}(\underline{Q}^\delta_{\mathrm{LV}}) \neq \emptyset \right\}. \qquad (6)$$

Λ_{LV} contains all the distortion parameters for which it is possible to find a linear vacuous model inner approximating $\underline{P}$. Proposition 6 tells us that Λ_{LV} is non-empty if and only if $\underline{P}$ is maximally imprecise. Λ_{LV} is also a directed set:

$$\delta_1 < \delta_2 \Rightarrow 1 - \delta_1 > 1 - \delta_2 \Rightarrow \frac{1}{1 - \delta_1} < \frac{1}{1 - \delta_2} \Rightarrow \underline{Q}^{\delta_1}_{\mathrm{LV}}(A) < \underline{Q}^{\delta_2}_{\mathrm{LV}}(A)$$

for any $A \neq \emptyset, \mathcal{X}$, meaning that $\mathcal{M}(\underline{Q}^{\delta_1}_{\mathrm{LV}}) \supset \mathcal{M}(\underline{Q}^{\delta_2}_{\mathrm{LV}})$. Our next result shows that Λ_{LV} has a maximum.

Proposition 7 *Let $\underline{P}$ be a maximally imprecise coherent lower probability. Then, the set Λ_{LV} defined in Eq. (6) has a maximum value δ_{LV}.*

It follows from Eq. (4) that any $P_0 \in \mathcal{M}\big(\underline{Q}_{LV}^{\delta_{LV}}\big)$ determines a LV model that is a non-dominating inner approximation of $\underline{P}$ in $\mathcal{C}_{LV}$.

Let us establish a more manageable expression for δ_{LV}, borrowing some notation from Miranda and Montes (2023). Let

$$\mathbb{A}(\mathcal{X}) = \left\{ \mathcal{A} = (A_i)_{i=1,\dots,k} \text{ for some } k \in \mathbb{N} \mid \exists \beta_{\mathcal{A}} \in \mathbb{N} : \sum_{i=1}^{k} I_{A_i} = \beta_{\mathcal{A}} \right\} \quad (7)$$

be the class of all finite families of subsets of $\mathcal{X}$ such that every $x \in \mathcal{X}$ belongs to the same number of elements in the family.

Theorem 8 *Let $\underline{P}$ be a maximally imprecise coherent lower probability. Then:*

$$\delta_{LV} = \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} \left(1 - \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \underline{P}(A) \right). \quad (8)$$

Next we prove that, under the assumption of 2-monotonicity, the expression above can be simplified. Let $\mathbb{A}^*(\mathcal{X})$ denote the set of partitions of $\mathcal{X}$.

Theorem 9 *Let $\underline{P}$ be a maximally imprecise 2-monotone lower probability with conjugate $\overline{P}$. Then:*

$$\delta_{LV} = \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} \left(1 - \sum_{A \in \mathcal{A}} \underline{P}(A), \frac{\sum_{A \in \mathcal{A}} \overline{P}(A) - 1}{|\mathcal{A}| - 1} \right). \quad (9)$$

Example 5 Let us continue with Example 4. There, we have shown that $\delta_{LV} = 0.3$. Since the lower probability $\underline{P}$ in that example is defined in a 3-element possibility space, it is also 2-monotone (Walley , 1981). Hence, δ_{LV} can be obtained using Theorem 9:

$\mathcal{A} \in \mathbb{A}^*(\mathcal{X})$	$1 - \sum_{A \in \mathcal{A}} \underline{P}(A)$	$\frac{1}{|\mathcal{A}|-1}\left(\sum_{A \in \mathcal{A}} \overline{P}(A) - 1\right)$
$\{x_1\}, \{x_2\}, \{x_3\}$	$1 - 0.2 - 0.05 - 0.1 = 0.65$	$0.7/2 = 0.35$
$\{x_1\}, \{x_2, x_3\}$	$1 - 0.2 - 0.5 = 0.3$	$0.3/1 = 0.3$
$\{x_2\}, \{x_1, x_3\}$	$1 - 0.05 - 0.4 = 0.55$	$0.55/1 = 0.55$
$\{x_3\}, \{x_1, x_2\}$	$1 - 0.1 - 0.4 = 0.5$	$0.5/1 = 0.5$

The minimum value is 0.3 (attained with the partition $\{x_1\}, \{x_2, x_3\}$), the same value we obtained in Example 4. ♦

Theorem 9 and Proposition 5 provide a simple procedure to determine a non-dominating linear vacuous model inner approximating $\underline{P}$: we first obtain a 2-monotone non-dominating inner approximation $\underline{Q}$ of $\underline{P}$ minimising the BV-distance (following the linear programming approach described in Sect. 3.2), and then apply Theorem 9 to $\underline{Q}$. This procedure is illustrated in Fig. 1.
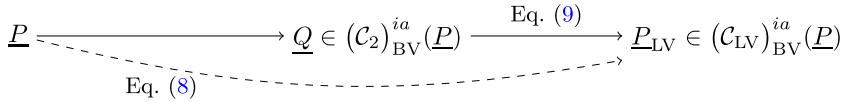
$$\underline{P} \xrightarrow{\hspace{3cm}} \underline{Q} \in (\mathcal{C}_2)^{ia}_{\mathrm{BV}}(\underline{P}) \xrightarrow{\text{Eq. (9)}} \underline{P}_{\mathrm{LV}} \in (\mathcal{C}_{\mathrm{LV}})^{ia}_{\mathrm{BV}}(\underline{P})$$

$$\text{Eq. (8)}$$

Fig. 1 Graphical description of the procedure for obtaining a non-dominating inner approximation in $\mathcal{C}_{\mathrm{LV}}$

4.2 Pari mutuel model

The second distortion model we consider is the *pari mutuel model*. Given a probability measure P_0 and a distortion parameter $\delta > 0$, the pari mutuel model is defined as the coherent lower probability:

$$\underline{P}_{\mathrm{PMM}}(A) = \max\{(1+\delta)P_0(A) - \delta, 0\} \,\forall A \subseteq \mathcal{X}$$

with conjugate coherent upper probability $\overline{P}_{\mathrm{PMM}}(A) = \min\{(1+\delta)P_0(A), 1\}$ for any $A \subseteq \mathcal{X}$. In Montes et al. (2018, Prop. 7), we proved that any coherent lower probability $\underline{P}$ has a unique undominated outer approximation in $\mathcal{C}_{\mathrm{PMM}}$ which is given by:

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

With respect to the inner approximations, we next show that a coherent lower probability $\underline{P}$ has an inner approximation in $\mathcal{C}_{\mathrm{PMM}}$ exactly under the same conditions as we saw in Proposition 6.

Proposition 10 *Let $\underline{P}$ be a coherent lower probability. There exists a pari mutuel model $\underline{P}_{\mathrm{PMM}}$ that inner approximates $\underline{P}$ if and only if $\underline{P}$ is maximally imprecise.*

Now, if $\underline{P}_{\mathrm{PMM}}$ is an inner approximation of $\underline{P}$ determined by (P_0, δ) and with conjugate $\overline{P}_{\mathrm{PMM}}$, it holds that:

$$d_{\mathrm{BV}}(\underline{P}, \underline{P}_{\mathrm{PMM}}) = \sum_{A \subseteq \mathcal{X}} |\underline{P}_{\mathrm{PMM}}(A) - \underline{P}(A)| = \sum_{A \subseteq \mathcal{X}} |\overline{P}(A) - \overline{P}_{\mathrm{PMM}}(A)|$$

$$= \sum_{A \subset \mathcal{X}} \overline{P}(A) - \sum_{A \subset \mathcal{X}} \overline{P}_{\mathrm{PMM}}(A) = \sum_{A \subset \mathcal{X}} \overline{P}(A) - (1+\delta) \sum_{A \subset \mathcal{X}} P_0(A),$$

where the fourth equality follows from the assumption $\underline{P}(A) \in (0, 1)$ for every $A \neq \emptyset, \mathcal{X}$ we are making throughout this section, which implies $\overline{P}(A) < 1$, and since $\underline{P}_{\mathrm{PMM}}$ is an inner approximation, $\overline{P}_{\mathrm{PMM}}(A) \leq \overline{P}(A) < 1$ whenever $A \neq \mathcal{X}$, hence $\overline{P}_{\mathrm{PMM}}(A) = (1+\delta)P_0(A)$ for $A \neq \mathcal{X}$.

Since $\sum_{A \subset \mathcal{X}} P_0(A)$ is constant for every probability measure P_0, the distance is minimised when $(1 + \delta)$ is maximised or, equivalently, when the distortion parameter δ is maximised. Therefore, we should look for the largest δ such that there is a probability measure P_0 satisfying $(1 + \delta)P_0(A) \leq \overline{P}(A)$ or equivalently $P_0(A) \leq \frac{\overline{P}(A)}{1+\delta}$ for any $A \subseteq \mathcal{X}$. This leads us to define, for some fixed $\delta > 0$, $\overline{Q}^{\delta}_{\mathrm{PMM}}$ as

$$\overline{Q}^{\delta}_{\mathrm{PMM}} = \frac{\overline{P}(A)}{1+\delta} \text{ if } A \neq \mathcal{X} \text{ and } \overline{Q}^{\delta}_{\mathrm{PMM}}(\mathcal{X}) = 1, \tag{10}$$

being $\underline{Q}^{\delta}_{\mathrm{PMM}}$ its conjugate lower probability. It follows that there is a PMM determined by (P_0, δ) inner approximating $\overline{P}$ if and only if the upper probability $\overline{Q}^{\delta}_{\mathrm{PMM}}$ in Eq. (10) avoids sure loss.

We also deduce that if there exists a PMM $\underline{P}_{\text{PMM}}$ defined by (P_0, δ) inner approximating $\underline{P}$, then for any $\delta' < \delta$ there exists another PMM with distortion parameter δ' inner approximating $\underline{P}$ as well. In other words, the set

$$\Lambda_{\text{PMM}} = \left\{ \delta \in (0, 1) \mid \mathcal{M}(\underline{Q}_{\text{PMM}}^{\delta}) \neq \emptyset \right\}, \tag{11}$$

is directed. It is not difficult to prove that it has a maximum.

Proposition 11 *Let $\underline{P}$ be a maximally imprecise coherent lower probability. Then, the set Λ_{PMM} defined in Eq. (11) has a maximum value δ_{PMM}.*

On the other hand, for any $P_0 \in \mathcal{M}(\underline{Q}_{\text{PMM}}^{\delta_{\text{PMM}}})$, the PMM determined by $(P_0, \delta_{\text{PMM}})$ is a non-dominating inner approximation of $\underline{P}$. This indicates that there may be more than one inner approximation in $\mathcal{C}_{\text{PMM}}$ minimising the BV-distance. The following example illustrates this fact:

Example 6 Consider the same coherent lower probability as in Example 4. The coherent lower probabilities $\underline{Q}_{\text{PMM}}^{1}$ and $\underline{Q}_{\text{PMM}}^{2}$ with conjugates given by:

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
$\overline{Q}_{\text{PMM}}^{1}(A)$	0.5	0.4	0.4	0.9	0.9	0.8
$\overline{Q}_{\text{PMM}}^{2}(A)$	0.5	0.35	0.45	0.85	0.95	0.8

are two different non-dominating inner approximations in $\mathcal{C}_{\text{PMM}}$ that minimise the BV-distance: $\overline{Q}_{\text{PMM}}^{1}$ is determined by $P_{\text{PMM}}^{1} = (0.5/1.3, 0.4/1.3, 0.4/1.3)$ and $\delta_1 = 0.3$, while $\overline{Q}_{\text{PMM}}^{2}$ is determined by $P_{\text{PMM}}^{2} = (0.5/1.3, 0.35/1.3, 0.45/1.3)$ and $\delta_2 = 0.3$. ◆

Let us give a more manageable expression of δ_{PMM}. Using the notation from Eq. (7), we obtain the following result.

Theorem 12 *Let $\underline{P}$ be a maximally imprecise coherent lower probability with conjugate $\overline{P}$. Then:*

$$\delta_{\text{PMM}} = \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} \left(\frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \overline{P}(A) - 1 \right). \tag{12}$$

When $\underline{P}$ is 2-monotone, Eq. (12) can be simplified further.

Theorem 13 *Let $\underline{P}$ be a maximally imprecise 2-monotone lower probability with conjugate $\overline{P}$. Then:*

$$\delta_{\text{PMM}} = \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - 1, \frac{1 - \sum_{A \in \mathcal{A}} \underline{P}(A)}{|\mathcal{A}| - 1} \right). \tag{13}$$

As for the LV model, Theorem 13 together with Proposition 5 gives a simple procedure for computing the value δ_{PMM}; it suffices to first inner approximate $\underline{P}$ by a 2-monotone lower probability $\underline{Q}$ and then apply Eq. (13) to $\underline{Q}$ and its conjugate $\overline{Q}$. This procedure is illustrated in Fig. 2.

Example 7 Let us continue with Example 4. Since $\underline{P}$ is 2-monotone, the value δ_{PMM} can be obtained by means of the computations in the following table:
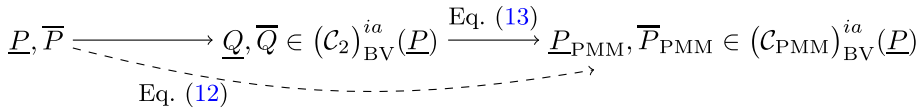
$$\underline{P}, \overline{P} \xrightarrow{\hspace{1cm}} Q, \overline{Q} \in \left(\mathcal{C}_2\right)^{ia}_{\mathrm{BV}}(\underline{P}) \xrightarrow{\text{Eq. (13)}} \underline{P}_{\mathrm{PMM}}, \overline{P}_{\mathrm{PMM}} \in \left(\mathcal{C}_{\mathrm{PMM}}\right)^{ia}_{\mathrm{BV}}(\underline{P})$$

$$\text{Eq. (12)}$$

Fig. 2 Graphical description of the procedure for obtaining a non-dominating inner approximation in $\mathcal{C}_{\mathrm{PMM}}$

$\mathcal{A} \in \mathbb{A}^*(\mathcal{X})$	$\sum_{A \in \mathcal{A}} \overline{P}(A) - 1$	$\frac{1}{|\mathcal{A}|-1}\left(1 - \sum_{A \in \mathcal{A}} \underline{P}(A)\right)$
$\{x_1\}, \{x_2\}, \{x_3\}$	$0.5+0.6+0.6-1=0.7$	$1/2(1-0.2-0.05-0.1)=0.325$
$\{x_1\}, \{x_2, x_3\}$	$0.5+0.8-1=0.3$	$1-0.2-0.5=0.3$
$\{x_2\}, \{x_1, x_3\}$	$0.6+0.95-1=0.55$	$1-0.05-0.4=0.55$
$\{x_3\}, \{x_1, x_2\}$	$0.6+0.9-1=0.5$	$1-0.1-0.4=0.5$

Thus, as we have already seen in Example 6, $\delta_{\mathrm{PMM}} = 0.3$. Two different inner approximations in $\mathcal{C}_{\mathrm{PMM}}$ associated with this value have been given in Example 6. ♦

4.3 Total variation model

The third and last distortion model we consider is the total variation model. Given a probability measure P_0 and a distortion parameter $\delta \in (0, 1)$, the total variation model is defined by the following coherent lower probability:

$$\underline{P}_{\mathrm{TV}}(A) = \max\{P_0(A) - \delta, 0\} \text{ if } A \neq \mathcal{X} \text{ and } \underline{P}_{\mathrm{TV}}(\mathcal{X}) = 1,$$

with conjugate coherent upper probability $\overline{P}_{\mathrm{TV}}(A) = \min\{P_0(A) + \delta, 1\}$ for any $A \neq \emptyset$. We showed in Destercke et al. (2022) that a coherent lower probability does not have a unique outer approximation in $\mathcal{C}_{\mathrm{TV}}$. With respect to the inner approximations, we prove next that there exists an inner approximation under the same conditions as for $\mathcal{C}_{\mathrm{LV}}$ and $\mathcal{C}_{\mathrm{PMM}}$.

Proposition 14 *Let $\underline{P}$ be a coherent lower probability. There exists a total variation model $\underline{P}_{\mathrm{TV}}$ that inner approximates $\underline{P}$ if and only if $\underline{P}$ is maximally imprecise.*

For any TV model $\underline{P}_{\mathrm{TV}}$ induced by P_0 and δ that inner approximates $\underline{P}$, their BV-distance is given by:

$$d_{\mathrm{BV}}(\underline{P}, \underline{P}_{\mathrm{TV}}) = \sum_{A \subseteq \mathcal{X}} |\underline{P}_{\mathrm{TV}}(A) - \underline{P}(A)| = \sum_{A \neq \emptyset, \mathcal{X}} |(P_0(A) - \delta) - \underline{P}(A)| =$$

$$\sum_{A \neq \emptyset, \mathcal{X}} (P_0(A) - \delta) - \sum_{A \neq \emptyset, \mathcal{X}} \underline{P}(A) = \sum_{A \neq \emptyset, \mathcal{X}} P_0(A) - \sum_{A \neq \emptyset, \mathcal{X}} \underline{P}(A) - \delta(2^n - 2),$$

where the second equality follows from our assumption $\underline{P}(A) \in (0, 1)$ for any $A \neq \emptyset, \mathcal{X}$, which implies that $\underline{P}_{\mathrm{TV}}(A) \geq \underline{P}(A) > 0$ for any $A \neq \emptyset$ because it is an inner approximation. Hence the BV-distance is minimised when δ is maximised.

In order to find a TV inner approximation of $\underline{P}$, we need to determine the existence of a probability measure P_0 such that $P_0(A) - \delta \geq \underline{P}(A)$ for any $A \neq \emptyset, \mathcal{X}$, which implies that $P_0(A) \geq \underline{P}(A) + \delta$ for every $A \neq \emptyset, \mathcal{X}$. This is equivalent to showing that

$$\underline{Q}^\delta_{\mathrm{TV}}(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ \underline{P}(A) + \delta, & \text{if } A \neq \emptyset, \mathcal{X}, \\ 1, & \text{if } A = \mathcal{X}, \end{cases}$$

is a lower probability that avoids sure loss, i.e., satisfying $\mathcal{M}\big(\underline{Q}_{\mathrm{TV}}^{\delta}\big) \neq \emptyset$. As we did for the LV and PMM models, we define the set

$$\Lambda_{\mathrm{TV}} = \left\{ \delta \in (0, 1) \mid \mathcal{M}\big(\underline{Q}_{\mathrm{TV}}^{\delta}\big) \neq \emptyset \right\}. \tag{14}$$

It is immediate that this is a directed set ($\delta_1 \in \Lambda_{\mathrm{TV}}$ implies that $\delta_2 \in \Lambda_{\mathrm{TV}}$ for any $\delta_2 < \delta_1$). It is also easy to prove that it has a maximum:

Proposition 15 *Let $\underline{P}$ be a maximally imprecise coherent lower probability. Then, the set Λ_{TV} defined in Eq. (14) has a maximum value δ_{TV}.*

Given the value δ_{TV}, any $P_0 \in \mathcal{M}\big(\underline{Q}_{\mathrm{TV}}^{\delta_{\mathrm{TV}}}\big)$ determines a non-dominating total variation model that inner approximates $\underline{P}$ and minimises the BV-distance.

On the other hand, the value δ_{TV} can be rewritten as follows:

$$\begin{aligned} \delta_{\mathrm{TV}} &= \max \left\{ \delta \in (0, 1) \mid \mathcal{M}\big(\underline{Q}_{\mathrm{TV}}^{\delta}\big) \neq \emptyset \right\} \\ &= \max \left\{ \delta \in (0, 1) \mid \exists P_0 \in \mathbb{P}(\mathcal{X}) \text{ s.t. } P_0(A) - \delta \geq \underline{Q}_{\mathrm{TV}}^{\delta}(A) \ \forall A \neq \emptyset, \mathcal{X} \right\} \\ &= \max \left\{ \delta \in (0, 1) \mid \exists P_0 \in \mathbb{P}(\mathcal{X}) \text{ s.t. } B_{\mathrm{TV}}^{\delta}(P_0) \subseteq \mathcal{M}\big(\underline{Q}_{\mathrm{TV}}^{\delta}\big) \right\}. \end{aligned} \tag{15}$$

Therefore, when $\underline{P}(A) \in (0, 1)$ for every $A \neq \emptyset, \mathcal{X}$, it coincides with what we called in Miranda and Montes (2023) the *incenter radius*,[1] (with respect to the TV distance) of the credal set $\mathcal{M}(\underline{P})$. Moreover, the probability measures P_0 such that $B_{\mathrm{TV}}^{\delta_{\mathrm{TV}}}(P_0) \subseteq \mathcal{M}(\underline{P})$ were called *incenters* of the credal set.

Hence, looking for the inner approximations of a coherent lower probability in $\mathcal{C}_{\mathrm{TV}}$ minimising the BV-distance is equivalent to looking for the incenter radius and the set of incenters (with respect to the TV distance). The results in Miranda and Montes (2023) provide then a simple formula for δ_{TV}.

Theorem 16 (Miranda & Montes, 2023, Thms. 4 and 5) *Let $\underline{P}$ be a maximally imprecise coherent lower probability with conjugate $\overline{P}$. Then*

$$\delta_{\mathrm{TV}} = \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} \frac{1}{|\mathcal{A}|} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A}} \underline{P}(A) \right). \tag{16}$$

If in addition $\underline{P}$ is 2-monotone, then

$$\delta_{\mathrm{TV}} = \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} \frac{1}{|\mathcal{A}|} \left\{ 1 - \sum_{A \in \mathcal{A}} \underline{P}(A), \ \sum_{A \in \mathcal{A}} \overline{P}(A) - 1 \right\}. \tag{17}$$

With this result, we obtain a simple procedure for computing a TV inner approximation: we first inner approximate the coherent lower probability $\underline{P}$ by means of a 2-monotone $\underline{Q}$ (using the procedures described in Sect. 3); next compute the value δ_{TV} using Eq. (17); and finally take any $P_0 \in \mathcal{M}(\underline{Q}_{\mathrm{TV}}^{\delta_{\mathrm{TV}}})$. These determine a TV model that inner approximates $\underline{P}$. This procedure is graphically illustrated in Fig. 3.

---

[1] The definition of incenter radius given in Miranda and Montes (2023) is slightly different; however, as argued in Miranda and Montes (2023, Sec. 3.3) when $\underline{P}(A) \in (0, 1)$ for any $A \neq \emptyset, \mathcal{X}$, the definition coincides with Eq. (15).
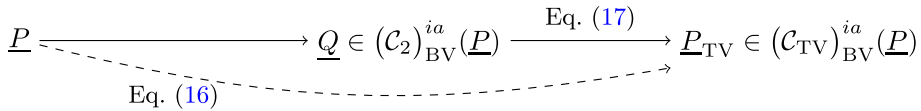
$$P \dashrightarrow Q \in \left(\mathcal{C}_2\right)^{ia}_{\mathrm{BV}}(\underline{P}) \xrightarrow{\text{Eq. (17)}} \underline{P}_{\mathrm{TV}} \in \left(\mathcal{C}_{\mathrm{TV}}\right)^{ia}_{\mathrm{BV}}(\underline{P})$$

Eq. (16)

Fig. 3 Graphical description of the procedure for obtaining a non-dominating inner approximation in $\mathcal{C}_{\mathrm{TV}}$

Example 8 Consider again our running Example 4. Using Eq. (17), we obtain:

$\mathcal{A} \in \mathbb{A}^*(\mathcal{X})$	$\frac{1}{|\mathcal{A}|}\left(1 - \sum_{A \in \mathcal{A}} \underline{P}(A)\right)$	$\frac{1}{|\mathcal{A}|}\left(\sum_{A \in \mathcal{A}} \overline{P}(A) - 1\right)$
$\{x_1\}, \{x_2\}, \{x_3\}$	$(1-0.2-0.05-0.1)/3 = 0.65/3$	$0.5+0.6+0.6/3 = 0.7/3$
$\{x_1\}, \{x_2, x_3\}$	$(1-0.2-0.5)/2 = 0.15$	$0.5+0.8-1/2 = 0.15$
$\{x_2\}, \{x_1, x_3\}$	$(1-0.05-0.4)/2 = 0.275$	$(0.6+0.95-1)/2 = 0.275$
$\{x_3\}, \{x_1, x_2\}$	$(1-0.1-0.4)/2 = 0.25$	$(0.6+0.9-1)/2 = 0.25$

Thus, the value δ_{TV} is given by 0.15. ♦

4.4 Inner approximations and incenters of credal sets

The last subsection shows that computing an inner approximation of a coherent lower probability in $\mathcal{C}_{\mathrm{TV}}$ is related to the computation of an incenter with respect to the TV-distance. This leads us to investigate the connection of the inner approximations in $\mathcal{C}_{\mathrm{LV}}$ and $\mathcal{C}_{\mathrm{PMM}}$ with the concept of incenter.

Recalling that throughout this section we are assuming that $\underline{P}(A) \in (0, 1)$ for any $A \neq \emptyset, \mathcal{X}$, we define

$$\delta_{\mathrm{LV}} = \max \left\{\delta \in (0, 1) \mid \exists P_0 \in \mathbb{P}(\mathcal{X}) \text{ such that } B^\delta_{\mathrm{LV}}(P_0) \subseteq \mathcal{M}(\underline{P})\right\}.$$
$$\delta_{\mathrm{PMM}} = \max \left\{\delta \in (0, 1) \mid \exists P_0 \in \mathbb{P}(\mathcal{X}) \text{ such that } B^\delta_{\mathrm{PMM}}(P_0) \subseteq \mathcal{M}(\underline{P})\right\}.$$

Here, $B^\delta_{\mathrm{LV}}(P_0)$ (resp., $B^\delta_{\mathrm{PMM}}(P_0)$) denotes the credal set associated with the LV (resp., PMM) distortion model determined by P_0 and δ.

Definition 4 Given a coherent lower probability $\underline{P}$ satisfying $\underline{P}(A) \in (0, 1)$ for any $A \neq \emptyset, \mathcal{X}$, δ_{LV} and δ_{PMM} are called *incenter radius* with respect to the LV or PMM model, respectively. Moreover, any P_0 such that $B^{\delta_{\mathrm{LV}}}_{\mathrm{LV}}(P_0) \subseteq \mathcal{M}(\underline{P})$ (respectively, $B^{\delta_{\mathrm{PMM}}}_{\mathrm{PMM}}(P_0) \subseteq \mathcal{M}(\underline{P})$) is called *incenter* with respect to the LV (resp., PMM) model.

Example 9 Let us continue with our running Example 4. As we have argued in Examples 4, 6 and 8, the LV, PMM and TV incenters radii are $\delta_{\mathrm{LV}} = \delta_{\mathrm{PMM}} = 0.3$ and $\delta_{\mathrm{TV}} = 0.15$. In addition, it can be easily seen that the LV incenters are $P^1_{\mathrm{LV}} = (2/7, 2/7, 3/7)$ and $P^2_{\mathrm{LV}} = (2/7, 3/7, 2/7)$, as well as their convex combinations. With respect to the PMM, the incenters are $P^1_{\mathrm{PMM}} = (5/13, 4/13, 4/13)$ and $P^2_{\mathrm{PMM}} = (5/13, 3.5/13, 4.5/13)$ as well as their convex combinations. And finally, with respect to the TV-distance, the incenters are $P^1_{\mathrm{TV}} = (0.35, 0.2, 0.45)$, $P^2_{\mathrm{TV}} = (0.35, 0.4, 0.25)$ and their convex combinations. Figure 4 shows a graphical representation of some of the incenters with respect to the LV (left), PMM (center) and TV (right). ♦

We next investigate the connection between these three radii:

Proposition 17 *In the conditions of Definition 4, it holds that $\delta_{\mathrm{TV}} \leq \min\{\delta_{\mathrm{LV}}, \delta_{\mathrm{PMM}}\}$.*
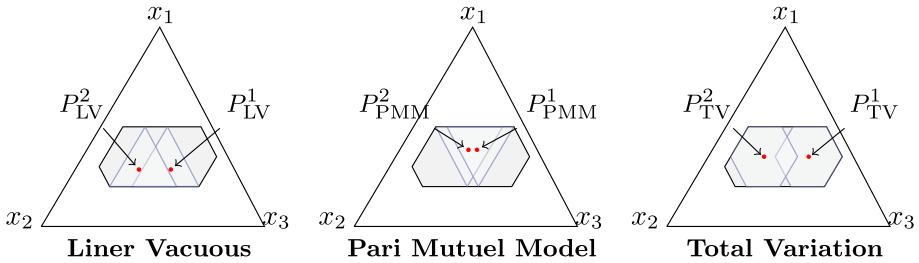
Fig. 4 Graphical representation of the incenters in Example 9

It follows from the running example that the inequality may be strict: $\delta_{LV} = \delta_{PMM} = 0.3 > \delta_{TV} = 0.15$.

Moreover, δ_{LV} and δ_{PMM} may not coincide.

Example 10 Consider the following coherent lower probabilities $\underline{P}_1$ and $\underline{P}_2$ with conjugate $\overline{P}_1$ and $\overline{P}_2$, respectively:

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
$\left[\underline{P}_1(A), \overline{P}_1(A)\right]$	[0.1,0.4]	[0.25,0.5]	[0.3,0.5]	[0.5,0.7]	[0.5,0.75]	[0.6,0.9]
$\left[\underline{P}_2(A), \overline{P}_2(A)\right]$	[0.1,0.4]	[0.2,0.4]	[0.3,0.5]	[0.5,0.7]	[0.6,0.8]	[0.6,0.9]

Since both $\underline{P}_1, \underline{P}_2$ are 2-monotone, we can apply Theorems 9 and 13, obtaining that in the case of $\underline{P}_1$, $\delta_{LV} = 0.2 > 0.175 = \delta_{PMM}$, while for $\underline{P}_2$ we obtain $\delta_{LV} = 0.15 < 0.2 = \delta_{PMM}$. On the other hand, by Theorem 16 it is $\delta_{TV} = 0.1$ in both cases. This shows that (i) $\delta_{LV}, \delta_{PMM}$ do not coincide in general; (ii) there is not a dominance relationship between them; and (iii) $\delta_{TV}, \delta_{LV}, \delta_{PMM}$ may all be different. ♦

We conclude this section by showing that the set of non-dominating inner approximations by a distortion model may strictly include those that minimise the BV-distance; in other words, there may be non-dominating inner approximations in $\mathcal{C}_{LV}$, $\mathcal{C}_{PMM}$ and $\mathcal{C}_{TV}$ with a parameter smaller than δ_{LV}, δ_{PMM} and δ_{TV}, respectively. However, it is those attaining these largest values that allow to make a connection with the notion of incenter.

Example 11 Considering again our running Example 4, we can easily check that:

- The LV model induced by $P_0 = (3/8, 1/2, 1/8)$ and $\delta = 0.2$ is a non-dominating inner approximation in $\mathcal{C}_{LV}$.
- The PMM determined by $P_0 = (7/23, 4/23, 12/23)$ and $\delta = 0.15$ is a non-dominating inner approximation in $\mathcal{C}_{PMM}$.
- The TV model associated with $P_0 = (0.3, 0.5, 0.2)$ and $\delta = 0.1$ defines a non-dominating inner approximation in $\mathcal{C}_{TV}$.

In all the cases, the parameter δ is smaller than δ_{LV}, δ_{PMM} and δ_{TV}, respectively. ♦

5 Decision making with inner and outer approximations

In this section we explain how inner and outer approximations can be used to obtain the optimal alternatives in decision making problems where the uncertainty is modelled by means of coherent lower probabilities.

Consider thus a finite set of alternatives D. For each $d \in D$, we assume that its utility depends on the outcome of an experiment taking values on $\mathcal{X}$, and we identify d with a variable $J_d : \mathcal{X} \to \mathbb{R}$. We aim at finding the optimal alternative(s) among those that are Pareto optimal:

$$\text{opt}_\geq = \{d \in D \mid \nexists e \in D \text{ such that } J_e \gneq J_d\}.$$

As we mentioned in the introduction, the expected utility paradigm has been extended in a number of ways to be able to deal with scenarios of imprecision or ambiguity about the probability measure that models the uncertainty. More specifically, we shall consider in this section five of these generalisations (we refer to Troffaes (2007) for a survey): Γ-maximin, Γ-maximax, maximality, interval dominance or E-admissibility. We analyse, for each of these criteria, whether there is a connection between the set of optimal alternatives under $\underline{P}$, and under an inner or outer approximation, $\underline{Q}_{in}, \underline{Q}_{ou}$.

Since these generalisations consider the lower and upper expectations of the different alternatives, we must recall here some basic facts from the theory of lower and upper *previsions* (Walley, 1991). Within this theory, any (bounded) mapping $f : \mathcal{X} \to \mathbb{R}$ is called a *gamble*, and the set of all gambles on $\mathcal{X}$ is denoted $\mathcal{L}(\mathcal{X})$. A *lower prevision* is a functional $\underline{P}$ defined on some subset $\mathcal{K}$ of $\mathcal{L}(\mathcal{X})$; its conjugate upper prevision is given by $\overline{P}(f) = -\underline{P}(-f)$ for every $f \in -\mathcal{K} := \{-g \mid g \in \mathcal{K}\}$. In particular, given a probability measure P on $\mathcal{X}$, its expectation operator $P : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$ given by $P(f) = \sum_{x \in \mathcal{X}} f(x) P(\{x\})$ is called a *coherent prevision* (de Finetti, 1974–1975).

A lower prevision on $\mathcal{L}(\mathcal{X})$ is called *coherent* if and only if there exists a closed and convex set $\mathcal{M}$ of coherent previsions such that $\underline{P}(f) = \min\{P(f) \mid P \in \mathcal{M}\}$; similarly, an upper prevision $\overline{P}$ is called coherent when $\overline{P}(f) = \max\{P(f) \mid P \in \mathcal{M}\}$ for every $f \in \mathcal{L}(\mathcal{X})$ for some closed and convex set of coherent previsions $\mathcal{M}$. In particular, a coherent lower probability $\underline{P}$ with associated credal set $\mathcal{M}(\underline{P})$ can be used to define a coherent lower and upper prevision: these are called the *natural extension* of $\underline{P}$ to $\mathcal{L}(\mathcal{X})$, and for any gamble $f : \mathcal{X} \to \mathbb{R}$, they are given by:

$$\underline{P}(f) := \min\{P(f) \mid P \in \mathcal{M}(\underline{P})\}, \quad \overline{P}(f) := \max\{P(f) \mid P \in \mathcal{M}(\underline{P})\}. \tag{18}$$

5.1 Γ-maximin

This criterion selects as optimal alternatives those maximising the lower prevision:

$$\text{opt}_{\underline{P}}(D) = \left\{ d \in \text{opt}_\geq \mid \underline{P}(J_d) = \max_{e \in D} \underline{P}(J_e) \right\}.$$

For this criterion, there is not an inclusion relationship between the optimal alternatives for $\underline{P}, \underline{Q}_{in}$ and $\underline{Q}_{ou}$, as we show in the next example.

Example 12 Consider the possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, the coherent lower probability $\underline{P}$, its undominated outer approximation $\underline{Q}_{ou}$ and its non-dominating inner approximation $\underline{Q}_{in}$ in $\mathcal{C}_2$ minimising the BV-distance given by:

A	$\underline{P}(A)$	$\underline{Q}_{in}(A)$	$\underline{Q}_{ou}(A)$	A	$\underline{P}(A)$	$\underline{Q}_{in}(A)$	$\underline{Q}_{ou}(A)$
$\{x_1\}$	0.1	0.1	0.1	$\{x_2, x_3\}$	0.3	0.3	0.2
$\{x_2\}$	0	0.1	0	$\{x_2, x_4\}$	0.4	0.4	0.4
$\{x_3\}$	0	0.1	0	$\{x_3, x_4\}$	0.4	0.4	0.4
$\{x_4\}$	0.3	0.3	0.3	$\{x_1, x_2, x_3\}$	0.5	0.5	0.5
$\{x_1, x_2\}$	0.1	0.2	0.1	$\{x_1, x_2, x_4\}$	0.6	0.7	0.6
$\{x_1, x_3\}$	0.3	0.3	0.3	$\{x_1, x_3, x_4\}$	0.7	0.8	0.7
$\{x_1, x_4\}$	0.6	0.6	0.5	$\{x_2, x_3, x_4\}$	0.6	0.6	0.6

Consider the set of alternatives $D = \{d_1, d_2, d_3\}$ whose utilities, as well as their lower previsions determined by $\underline{P}$, $\underline{Q}_{in}$ and $\underline{Q}_{ou}$ using natural extension, are given by:

	x_1	x_2	x_3	x_4	$\underline{P}(J_i)$	$\underline{Q}_{in}(J_i)$	$\underline{Q}_{ou}(J_i)$
J_1	3	2	$-9/10$	3	1.44	1.73	1.34
J_2	2	3	$2/3$	2	$1.4\overline{6}$	1.7	$1.4\overline{6}$
J_3	4	-2	-2	4	1.6	1.6	1

We obtain that $\mathrm{opt}_{\underline{P}}(D) = \{d_3\}$, $\mathrm{opt}_{\underline{Q}_{in}}(D) = \{d_1\}$ and $\mathrm{opt}_{\underline{Q}_{ou}}(D) = \{d_2\}$, so the three coherent lower probabilities give different results. ♦

5.2 Γ-maximax

This criterion selects as optimal alternatives those maximising the upper prevision:

$$\mathrm{opt}_{\overline{P}}(D) = \left\{ d \in \mathrm{opt}_{\geq} \mid \overline{P}(J_d) = \max_{e \in D} \overline{P}(J_e) \right\};$$

it can be seen as the dual of the Γ-maximin. Not surprisingly, for this criterion there is not a connection between $\mathrm{opt}_{\overline{P}}(D)$, $\mathrm{opt}_{\overline{Q}_{in}}(D)$ and $\mathrm{opt}_{\overline{Q}_{ou}}(D)$ either.

Example 13 Consider the setting in Example 12 and the set of alternatives $D = \{d_2, d_4, d_5\}$, where d_2 comes from Example 12, and d_4, d_5 are defined by:

	x_1	x_2	x_3	x_4
J_4	-1	-1	2.7	2.7
J_5	3	-2	-2	4

The upper previsions of the three alternatives for $\overline{P}$, $\overline{Q}_{in}$ and $\overline{Q}_{ou}$ are given by:

	J_2	J_4	J_5
$\overline{P}(J_i)$	2.3	2.33	2
$\overline{Q}_{in}(J_i)$	$2.0\overline{6}$	1.96	2
$\overline{Q}_{ou}(J_i)$	2.3	2.33	2.5

We observe that $\mathrm{opt}_{\overline{P}}(D) = \{d_4\}$, $\mathrm{opt}_{\overline{Q}_{in}}(D) = \{d_2\}$ and $\mathrm{opt}_{\overline{Q}_{ou}}(D) = \{d_5\}$, whence the three models give different solutions. ♦

5.3 Maximality

According to maximality, the optimal alternatives are those d satisfying $\underline{P}(J_e - J_d) \leq 0$ for any other alternative $e \in D$:

$$\text{opt}_{>\underline{P}} = \big\{d \in \text{opt}_{\geq} \mid \underline{P}(J_e - J_d) \leq 0 \ \forall e \in D\big\}.$$

We obtain the following result:

Proposition 18 *Let $\underline{P}$ and $\underline{Q}$ be two coherent lower probabilities such that $\underline{P} \leq \underline{Q}$. Then $\text{opt}_{>\underline{P}} \supseteq \text{opt}_{>\underline{Q}}$.*

Then, if $\underline{Q}_{in}$ and $\underline{Q}_{ou}$ are inner and outer approximations of $\underline{P}$, it holds that $\text{opt}_{>\underline{Q}_{in}} \supseteq \text{opt}_{>\underline{P}} \supseteq \text{opt}_{>\underline{Q}_{in}}$. The above inclusions may be strict:

Example 14 Consider the same setting as in Examples 12, 13, and the set of alternatives $D = \{d_1, d_2, d_6\}$, where d_1, d_2 were given in Example 12 and d_6 is:

	x_1	x_2	x_3	x_4
J_6	0	2	3.5	0

The following table gives the values of $\underline{P}$, $\underline{Q}_{in}$ and $\underline{Q}_{ou}$ for the differences between the gambles:

	$J_2 - J_1$	$J_6 - J_1$	$J_1 - J_2$	$J_6 - J_2$	$J_1 - J_6$	$J_2 - J_6$
$\underline{P}(J_i - J_j)$	-0.4	-2.1	$-0.02\overline{6}$	-1.7	0.04	$0.0\overline{6}$
$\underline{Q}_{in}(J_i - J_j)$	$-0.34\overline{3}$	-1.66	0.03	$-1.31\overline{6}$	0.48	0.45
$\underline{Q}_{ou}(J_i - J_j)$	-0.6	-2.4	$-0.22\overline{6}$	-1.8	-0.26	$-0.0\overline{3}$

We conclude that $\text{opt}_{>\underline{Q}_{in}} = \{d_1\}$, $\text{opt}_{>\underline{P}} = \{d_1, d_2\}$ and $\text{opt}_{>\underline{Q}_{ou}} = \{d_1, d_2, d_6\}$, and as a consequence the inclusions between these sets are strict. ◆

5.4 Interval dominance

This criterion computes $[\underline{P}(J_d), \overline{P}(J_d)]$ for each alternative d in D, and compares these intervals, giving rise to the following optimal alternatives:

$$\text{opt}_{\sqsupset \underline{P}} = \big\{d \in \text{opt}_{\geq} \mid \overline{P}(J_d) \geq \underline{P}(J_e) \ \forall e \in \text{opt}_{\geq}\big\}.$$

We obtain the following relationships:

Proposition 19 *Let $\underline{P}$ and $\underline{Q}$ be two coherent lower probabilities such that $\underline{P} \leq \underline{Q}$. Then $\text{opt}_{\sqsupset \underline{P}} \supseteq \text{opt}_{\sqsupset \underline{Q}}$.*

This implies that $\text{opt}_{\sqsupset \underline{Q}_{in}} \supseteq \text{opt}_{\sqsupset \underline{P}} \supseteq \text{opt}_{\sqsupset \underline{Q}_{ou}}$, and as we show in our next example, the inclusions may be strict.

Example 15 Consider again the same setting as in Examples 12, 13 and 14. Consider also the set of alternatives $D = \{d_1, d_6, d_7\}$, where d_1 was defined in Example 12, d_6 was defined in Example 14 and d_7 is given by:

	x_1	x_2	x_3	x_4
J_7	3	3.5	-1	-1

We obtain that:

	J_1	J_6	J_7
$\left[\underline{P}(J_i), \overline{P}(J_i)\right]$	[1.44, 2.7]	[0.6, 1.4]	[$-$ 0.6, 1.55]
$\left[\underline{Q}_{in}(J_i), \overline{Q}_{in}(J_i)\right]$	[1.73, 2.41]	[0.75, 1.25]	[-0.15, 1.5]
$\left[\underline{Q}_{ou}(J_i), \overline{Q}_{ou}(J_i)\right]$	[1.34 , 2.8]	[0.4, 1.6]	[$-$0.6, 1.55]

Hence, we obtain the following sets of optimal alternatives: $\mathrm{opt}_{\neg \underline{Q}_{in}} = \{d_1\}$, $\mathrm{opt}_{\neg \underline{P}} = \{d_1, d_7\}$, and $\mathrm{opt}_{\neg \underline{Q}_{ou}} = \{d_1, d_6, d_7\}$, and therefore the inclusions are strict. ♦

5.5 E-admissibility

According to E-admissibility, we choose those alternatives that maximise the expected utility for at least one element of the credal set $\mathcal{M}(\underline{P})$:

$$\mathrm{opt}_{\mathcal{M}(\underline{P})} = \left\{ d \in \mathrm{opt}_{\geq} \mid \exists P \in \mathcal{M}(\underline{P}) \text{ such that } E_P(J_e) \leq E_P(J_d) \ \forall e \in \mathrm{opt}_{\geq} \right\}.$$

We next prove the following connection with respect to E-admissibility.

Proposition 20 *Let $\underline{P}$ and $\underline{Q}$ be two coherent lower previsions such that $\underline{P} \leq \underline{Q}$. Then $\mathrm{opt}_{\mathcal{M}(\underline{Q})} \subseteq \mathrm{opt}_{\mathcal{M}(\underline{P})}$.*

From this result we deduce that $\mathrm{opt}_{\mathcal{M}(\underline{Q}_{in})} \subseteq \mathrm{opt}_{\mathcal{M}(\underline{P})} \subseteq \mathrm{opt}_{\mathcal{M}(\underline{Q}_{ou})}$.

Example 16 Let us continue with Examples 12–15. If we consider the set of alternatives $D = \{d_1, d_5, d_8\}$, where d_8 is given by

	x_1	x_2	x_3	x_4
J_8	0.95	1.6	1.8	1

we obtain that $\mathrm{opt}_{\mathcal{M}(\underline{Q}_{in})} = \{d_1\}$, $\mathrm{opt}_{\mathcal{M}(\underline{P})} = \{d_1, d_5\}$ and $\mathrm{opt}_{\mathcal{M}(\underline{Q}_{ou})} = \{d_1, d_5, d_8\}$, showing that the inclusions are strict. ♦

5.6 Comparison between the decisions

Next we make a comparison between the optimal alternatives within a set D when we consider the initial coherent lower probability $\underline{P}$ and a 2-monotone inner approximation $\underline{Q}$, taking into account the distance $d_{BV}(\underline{P}, \underline{Q})$, and under any of the criteria considered previously in this section. In this respect, a first comment is that we may assume without loss of generality that for any alternative d its associated gamble is bounded between 0 and 1. Indeed, it follows by coherence that for any $a > 0, b \in \mathbb{R}$ and any gamble f, it holds that $\underline{P}(af + b) = a\underline{P}(f) + b$ and $\overline{P}(af + b) = a\overline{P}(f) + b$. As a consequence, given two gambles $f, g, a \neq 0, b \in \mathbb{R}$ and a coherent lower prevision $\underline{P}$, we obtain that:

- $\underline{P}(f) \geq \underline{P}(g) \Leftrightarrow \underline{P}(af + b) \geq \underline{P}(ag + b)$;
- $\overline{P}(f) \geq \overline{P}(g) \Leftrightarrow \overline{P}(af + b) \geq \overline{P}(ag + b)$;
- $\underline{\underline{P}}(f - g) \leq 0 \Leftrightarrow \underline{\underline{P}}((af + b) - (ag + b)) \leq 0$;
- $\overline{P}(f) \geq \underline{P}(g) \Leftrightarrow \overline{P}(af + b) \geq \underline{P}(ag + b)$.

This implies that the set of optimal decisions is invariant under affine transformations of the gambles associated with the alternatives. It is not difficult to establish the following[2]:

Proposition 21 *Let $\underline{P}$ be a coherent lower probability and let $\underline{Q}$ be an inner approximation in $\mathcal{C}_2$. We use the same notation $\underline{P}$ and $\underline{Q}$ to denote the natural extension to gambles defined in Eq. (18). If f is a gamble taking values in* $[0, 1]$*, then:*

$$d_{\mathrm{BV}}(\underline{P}, \underline{Q}) \leq \delta \Rightarrow |\underline{P}(f) - \underline{Q}(f)| \leq \delta.$$

As a consequence, we deduce that, if $d_{\mathrm{BV}}(\underline{P}, \underline{Q}) \leq \delta$, then:

- $\underline{P}(f) - \underline{P}(g) \geq \delta \Rightarrow \underline{Q}(f) - \underline{Q}(g) \geq 0$.
- $\overline{P}(f) - \overline{P}(g) \geq \delta \Rightarrow \overline{Q}(f) - \overline{Q}(g) \geq 0$.
- $\underline{P}(f - g) \leq -\delta \Rightarrow \underline{Q}(f - g) \leq 0$.
- $\overline{P}(f) - \underline{P}(g) \geq 2\delta \Rightarrow \overline{Q}(f) - \underline{Q}(g) \geq 0$.

These implications relate the optimal alternatives under Γ-maximin, Γ-maximax, maximality and interval dominance for the original and transformed models.

6 Illustration in a decision problem under severe uncertainty

After showing how inner and outer approximations can be used in decision making problems, we illustrate its applicability in a *real world toy example*, following the terminology in Jansen et al. (2018, Sec. 5). For this aim, we first summarise the context from Jansen et al. (2018).

6.1 Decision making under severe uncertainty: setup

Given a non-empty set of alternatives A, and two preorders $R_1 \subseteq A \times A$ and $R_2 \subseteq R_1 \times R_1$ on A and R_1, respectively, the triple $\mathcal{A} = [A, R_1, R_2]$ is called a *preference system* in A. R_1 and R_2 are interpreted as follows: $(a, b) \in R_1$ means that a is at least as preferable as b, while $((a, b), (c, d)) \in R_2$ means that exchanging b with a is at least as desirable as exchanging d with c.

Associated with R_1 and R_2 we can consider the indifference and strict preference relations I_{R_1}, I_{R_2} and P_{R_1}, P_{R_2}. Using them we can establish when the preference systems satisfies some sort of rationality.

Definition 5 (Jansen et al., 2018, Def. 2,3) Let $\mathcal{A} = [A, R_1, R_2]$ be a preference systems. $\mathcal{A}$ is *consistent* if there exists a function $u \colon A \to [0, 1]$ such that for any $a, b, c, d \in A$ the following properties hold:

i) If $(a, b) \in R_1$, then $u(a) \geq u(b)$, with equality if and only if $(a, b) \in I_{R_1}$.
ii) If $((a, b), (c, d)) \in R_2$, then $u(a) - u(b) \geq u(c) - u(d)$, with equality if and only if $((a, b)(c, d)) \in I_{R_2}$.

---

[2] A similar conclusion is obtained if instead of using d_{BV} to compare $\underline{P}$ and its transformation $\underline{Q}$ we considered the TV-distance: it is easy to establish that $\max_{A \subseteq \mathcal{X}} |\underline{P}(A) - \underline{Q}(A)| = \max_{f:0 \leq f \leq 1} |\underline{P}(f) - \underline{Q}(f)|$ when $\underline{Q}$ is 2-monotone. We acknowledge Jasper de Bock for this remark.

Each function u satisfying conditions (i) and (ii) above is said to *weakly represent* the preference system $\mathcal{A}$, and the set of all these functions is denoted as $\mathcal{U}_{\mathcal{A}}$. The subset of $\mathcal{U}_{\mathcal{A}}$ formed by the functions u satisfying in addition $\inf_{a \in A} u(a) = 0$ and $\sup_{a \in A} u(a) = 1$ is denoted by $\mathcal{N}_{\mathcal{A}}$. Moreover, given $\delta \in (0, 1)$, $\mathcal{N}_{\mathcal{A}}^{\delta}$ denotes the elements $u \in \mathcal{N}_{\mathcal{A}}$ satisfying $u(a) - u(b) \geq \delta$ for any $(a, b) \in P_{R_1}$ and $u(a) - u(b) - u(c) + u(d) \geq \delta$ for any $((a, b)(c, d)) \in P_{R_2}$. $\mathcal{N}_{\mathcal{A}}^{\delta}$ is called the *weak representation set of granularity at least δ*.

The granularity δ can be seen as a control parameter, in the sense that a given value δ guarantees that one decision is only considered preferred to another when the differences between their utilities are above a predetermined threshold.

Definition 6 (Jansen et al., 2018, Def. 4) Let $\mathcal{X}$ be the states of the nature, A the consequences and $D = \{X \mid X \colon \mathcal{X} \to A\}$ the set of alternatives. Each $\mathcal{G} \subseteq D$ is called *decision system*.

Assuming that the uncertainty about the states of the nature is given by means of a coherent lower prevision $\underline{P}$ with conjugate $\overline{P}$, the natural approach to determine the optimal decision is based in comparing the *generalised interval expectation with granularity δ* (Jansen et al., 2018, Def.5), given by:

$$E_{\mathcal{D}_{\delta}}(X) = \left[\inf_{u \in \mathcal{N}_{\mathcal{A}}^{\delta}} \underline{P}(u \circ X), \ \sup_{u \in \mathcal{N}_{\mathcal{A}}^{\delta}} \overline{P}(u \circ X) \right] = \left[\underline{P}_{\mathcal{D}_{\delta}}(X), \overline{P}_{\mathcal{D}_{\delta}}(X) \right].$$

Then, the following criteria can be considered:

$\mathcal{D}_{\delta}$-maximin: $\underline{\mathcal{G}}_{\delta} = \left\{ X \in \mathcal{G} \mid \forall Y \in \mathcal{G} \text{ it holds } \underline{P}_{\mathcal{D}_{\delta}}(X) \geq \underline{P}_{\mathcal{D}_{\delta}}(Y) \right\}$.

$\mathcal{D}_{\delta}$-maximax: $\overline{\mathcal{G}}_{\delta} = \left\{ X \in \mathcal{G} \mid \forall Y \in \mathcal{G} \text{ it holds } \overline{P}_{\mathcal{D}_{\delta}}(X) \geq \overline{P}_{\mathcal{D}_{\delta}}(Y) \right\}$.

$\mathcal{A}$-admissibility: $\mathcal{G}_{\mathcal{A}} = \left\{ X \in \mathcal{G} \mid \exists u \in \mathcal{U}_{\mathcal{A}} \colon \forall P \in \mathcal{M}(\underline{P}), \forall Y \in \mathcal{G} \text{ it holds } E_P(u \circ X) \geq E_P(u \circ Y) \right\}$.

The $\mathcal{D}_{\delta}$-maximin and $\mathcal{D}_{\delta}$-maximax criteria straightforwardly generalise Γ-maximin and Γ-maximax from Sect. 5, while $\mathcal{A}$-admissibility generalises E-admissibility. Computing the generalised interval expectations or finding the $\mathcal{A}$-admissible alternatives can be done by solving linear programming problems, as shown in Jansen et al. (2018, Prop. 3,4). However, this requires knowing the extreme points of the credal set, a task that simplifies considerably under 2-monotonicity.

6.2 Example setup (Jansen et al., 2018)

Consider a decision maker that must choose among three job offers, J_1, J_2 and J_3. Each job offer has a salary and several additional benefits, $\mathcal{B}$, which are: overtime premium (b_1), child care (b_2), advanced training (b_3), promotion prospects (b_4) and flexible hours (b_5). Moreover, the salary and benefits depend on the economic situation for which we envisage four scenarios: $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$. The situation is described in the following table (Jansen et al., 2018, p.127):

	x_1	x_2	x_3	x_4
J_1	$a_1 = (5000, \mathcal{B})$	$a_2 = (2700, \{b_1, b_2\})$	$a_3 = (2300, \{b_1, b_2, b_3\})$	$a_4 = (1000, \emptyset)$
J_2	$a_5 = (3500, \{b_1, b_5\})$	$a_6 = (2400, \{b_1, b_2\})$	$a_7 = (1700, \{b_1, b_2\})$	$a_8 = (2500, \{b_1\})$
J_3	$a_9 = (3000, \{b_1, b_2, b_3\})$	$a_{10} = (1000, \{b_1\})$	$a_{11} = (2000, \{b_1\}))$	$a_{12} = (3000, \{b_1, b_4, b_5\})$

Assuming incomparability among the benefits, the information is summarised by a preference system $\mathcal{A} = [A, R_1, R_2]$, where (i) $A = \{a_1, \ldots, a_{12}\}$ are the consequences, where each of them is a pair (y, B), where $y \in \mathbb{R}$ denotes the salary and $B \subseteq \mathcal{B}$ is the set of benefits; (ii) R_1 denotes a relation defined as:

$$R_1 = \{((y_1, B_1), (y_2, B_2)) \mid y_1 \geq y_2 \wedge B_2 \subseteq B_1\},$$

i.e., a_i is preferred to a_j with respect to R_1 when the salary of a_i is greater and all the benefits of a_j are also included in a_i; and (iii) R_2 is the relation:

$$R_2 = \Big\{(((y_1, B_1), (y_2, B_2)), ((y_3, B_3), (y_4, B_4))) \mid$$
$$y_1 - y_2 \geq y_3 - y_4 \wedge B_2 \subseteq B_4 \subseteq B_3 \subseteq B_1\Big\}.$$

In order to measure the uncertainty, the available information only allows to compare the probability of occurrence of each scenario:

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x_1\}) \geq P(\{x_2\}) \geq P(\{x_3\}) \geq P(\{x_4\})\}.$$

Using the results in Miranda and Destercke (2015), the lower probability $\underline{P}$ associated with this information is given in the following table:

A	$\underline{P}(A)$	$\underline{Q}_{in}$	$\underline{Q}_{ou}$	A	$\underline{P}(A)$	$\underline{Q}_{in}$	$\underline{Q}_{ou}$
$\{x_1\}$	1/4	7/24	1/4	$\{x_2, x_3\}$	0	0	0
$\{x_2\}$	0	0	0	$\{x_2, x_4\}$	0	0	0
$\{x_3\}$	0	0	0	$\{x_3, x_4\}$	0	0	0
$\{x_4\}$	0	0	0	$\{x_1, x_2, x_3\}$	3/4	3/4	3/4
$\{x_1, x_2\}$	1/2	1/2	1/2	$\{x_1, x_2, x_4\}$	2/3	2/3	2/3
$\{x_1, x_3\}$	1/2	1/2	11/24	$\{x_1, x_3, x_4\}$	1/2	13/24	1/2
$\{x_1, x_4\}$	1/3	1/3	7/24	$\{x_2, x_3, x_4\}$	0	0	0

This lower probability is not 2-monotone, as it can be easily seen taking the events $A = \{x_1, x_3\}$ and $B = \{x_1, x_4\}$. Hence, we may take a 2-monotone non-dominating inner approximation $\underline{Q}_{in}$ and a 2-monotone undominated outer approximation $\underline{Q}_{ou}$. We consider $\underline{Q}_{in}$ and $\underline{Q}_{ou}$ as we optimal solutions of the quadratic problem in Propositions 4 and 2, respectively, that are at a BV-distance $d_{BV}(\underline{P}, \underline{Q}_{in}) = 0.08\overline{3}$ and $d_{BV}(\underline{P}, \underline{Q}_{ou}) = 1.75$.

6.3 Results

Applying Propositions 1 and 2 in Jansen et al. (2018), we obtain that the preference system $\mathcal{A} = [A, R_1, R_2]$ is consistent, and that the maximum possible granularity degree[3] is $\delta = 0.053$. It can be easily seen that only the job offers J_1 and J_3 are $\mathcal{A}$-admissible for the three models: $\underline{P}$, $\underline{Q}_{in}$ and $\underline{Q}_{ou}$. The table shows the generalised interval expectations for different granularities, all of them smaller than 0.053, for the three models:

---

[3] In Jansen et al. (2018), the maximum granularity degree given is $\delta = 0.03\overline{7}$, but we believe that there is typo in the calculations. After a thorough analysis, we believe that in Jansen et al. (2018) it is considered that $((a_9, a_2), (a_2, a_6))$ belongs to I_{R_2}, which is incorrect, rather than to P_{R_2}.

		$\delta = 0$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.03$	$\delta = 0.04$	$\delta = 0.05$
$\underline{P}$	$E_{D_\delta}(J_1)$	[0.25, 1]	[0.2925, 1]	[0.335, 1]	[0.3775, 1]	[0.412, 1]	[0.4625, 1]
	$E_{D_\delta}(J_2)$	[0, 1]	[0.08, 0.93]	[0.16, 0.86]	[0.24, 0.79]	[0.32, 0.72]	[0.4, 0.65]
	$E_{D_\delta}(J_3)$	[0, 1]	[0.05$\overline{6}$, 0.93]	[0.11$\overline{3}$, 0.86]	[0.17, 0.79]	[0.22$\overline{6}$,0.72]	[0.28$\overline{3}$, 0.65]
$\underline{Q}_{in}$	$E_{D_\delta}(J_1)$	[7/24, 1]	[0.3304, 1]	[0.3692, 1]	[0.4079, 1]	[0.4467, 1]	[0.4854, 1]
	$E_{D_\delta}(J_2)$	[0, 1]	[0.08, 0.93]	[0.16, 0.86]	[0.24, 0.79]	[0.32, 0.72]	[0.4, 0.65]
	$E_{D_\delta}(J_3)$	[0, 1]	[0.052$\overline{6}$, 0.93]	[0.10$\overline{3}$, 0.86]	[0.155, 0.79]	[0.20$\overline{6}$,0.72]	[0.258$\overline{3}$, 0.65]
$\underline{Q}_{ou}$	$E_{D_\delta}(J_1)$	[0.25, 1]	[0.2925, 1]	[0.335, 1]	[0.3775, 1]	[0.412, 1]	[0.4625, 1]
	$E_{D_\delta}(J_2)$	[0, 1]	[0.078, 0.93]	[0.15$\overline{6}$, 0.86]	[0.235,0.79]	[0.31$\overline{3}$, 0.72]	[0.391$\overline{6}$, 0.65]
	$E_{D_\delta}(J_3)$	[0, 1]	[0.0475, 0.93]	[0.095, 0.86]	[0.1425, 0.79]	[0.19,0.72]	[0.2375$\overline{3}$, 0.65]

We obtain the same conclusion for the three models: since the lower and upper limits for J_1 are greater than those of J_2 and J_3, J_1 is optimal with respect to $\mathcal{D}_\delta$-maximin and $\mathcal{D}_\delta$-maximax. For a better visualisation, we graphically show these results in Figs. 5, 6 and 7 for $\underline{P}$, $\underline{Q}_{in}$ and $\underline{Q}_{ou}$, respectively.
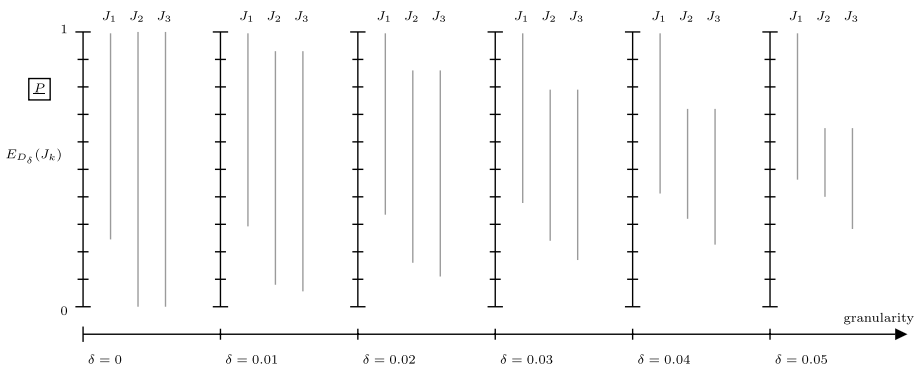
Fig. 5 Generalised interval expectation for different granularities with respect to the initial model $\underline{P}$
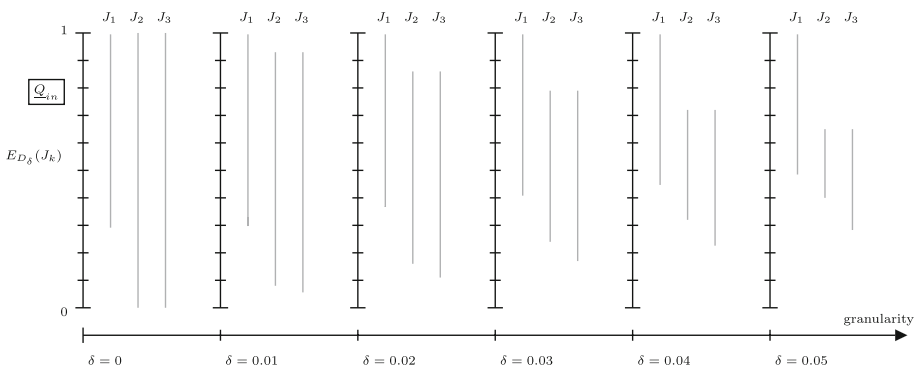
Fig. 6 Generalised interval expectation for different granularities with respect to the non-dominating inner approximation $\underline{Q}_{in}$
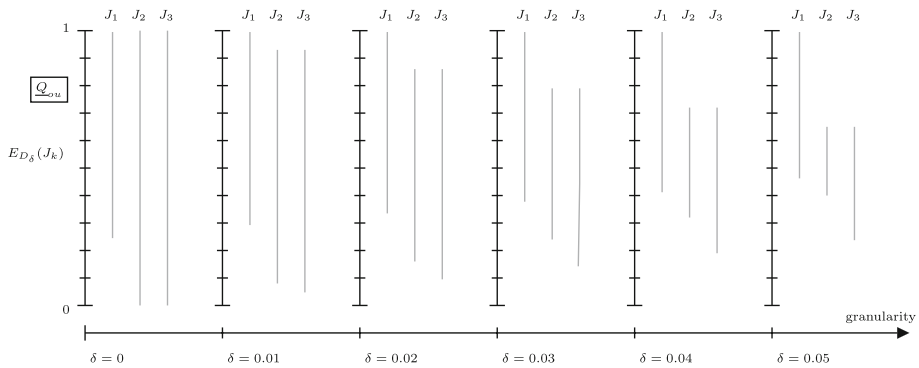
Fig. 7 Generalised interval expectation for different granularities with respect to the undominated outer approximation $\underline{Q}_{ou}$

6.4 Discussion

In this section we have presented a decision making problem to demonstrate that using the initial coherent lower probability, which is not 2-monotone, a non-dominating inner approximation $\underline{Q}_{in}$ or undominated outer approximation $\underline{Q}_{ou}$ yield the same results. One of the reasons is that the approximations are "very close" to the initial model $\underline{P}$, since for instance in the case of the inner approximation we have $d_{BV}(\underline{P}, \underline{Q}_{in}) = 0.08\overline{3}$. This aligns with our comments in Sect. 5.6: if the distance between the initial and transformed model is small enough, there will not be much difference between the optimal decisions with the two models.

In addition, the use of (inner or outer) has a number of benefits:

- First of all, following (Jansen et al., 2018, Props. 3,4,5), solving the decision making problem requires the knowledge of the extreme points of the credal set. The computation of the extreme points under the assumption of 2-monotonicity is a straightforward process and can be achieved using the procedure described in Shapley (1971). On the other hand, computing the extreme points of the credal set of an arbitrary coherent lower probability is far from trivial: while the maximum number of extreme points of the credal set of a coherent lower probability is upper bounded by $|\mathcal{X}|!$ (Derks & Kuipers, 2002; Wallner, 2007), their computation is not immediate except in some particular cases.
- Secondly, computing the generalised interval expectations requires solving a collection of linear programming problems (Jansen et al., 2018, Prop. 3), as many as the number of extreme points. In contrast, under the assumption of 2-monotonicity, these interval expectations coincide with the Choquet integral (Choquet, 1953), as explained in Jansen et al. (2018).
- Thirdly, some models of the imprecise probability theory induce credal sets with a non-finite number of extreme points, as for example if the starting point are coherent lower previsions (Walley, 1991). In that case, applying the procedure described in Jansen et al. (2018) would not be possible. This issue could be overcome by considering the restriction to events, which gives an outer approximation of the original model.

The spirit of these comments can be summarised by the following comment given in Jansen et al. (2018, p. 119):

Table 1 Properties to the inner and outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$

Property		$\mathcal{C}_2$	$\mathcal{C}_\infty$
Unique solution of the linear	Outer approximation	No (Montes et al., 2018, Ex.1)	No (Montes et al., 2019, Ex.1)
programming problem	Inner approximation	No (Ex. 2)	No (Ex. 1)
$\underline{P}(A) = \underline{Q}(A)$ for all A	Outer approximation	Yes (Montes et al., 2018, Prop.2)	No (Montes et al., 2019, Ex.4)
with cardinality 1 or $n-1$	Inner approximation	No (Ex. 1)	No (Ex. 1)
The optimal solution of the quadratic approach	Outer approximation	No (Montes et al., 2018, Ex.3)	No (Montes et al., 2019, Ex.6)
is an optimal solution of the linear approach	Inner approximation	No (Ex. 1)	No (Ex. 1)

"[This approach] …is ideal for situations where the number of extreme points is moderate and where closed formulas for computing the extreme points are available. For credal sets induced by 2-monotone lower/ 2-alternating upper probabilities such formulas exist."

7 Concluding remarks

7.1 Summary

The results in this paper show that it is possible to transform a coherent lower probability into a more manageable model with a minimal loss of information. While in our previous studies we considered approximations not adding new information to our model (that is, outer approximations), in this paper we have headed in the opposite direction and used inner approximations, that are more informative than the original model. We have considered transformations into the class of 2- or completely monotone lower probabilities (Sect. 3) and distortion models (Sect. 4). Our reasons for focusing on these models are that (i) 2-monotone lower probabilities overcome some of the shortcomings of coherent lower probabilities (Desercke, 2013) while being easier to handle; (ii) completely monotone lower probabilities (or belief functions) are connected to Dempster-Shafer theory, and the approximations by means of these model have proven to be quite powerful in statistical matching (Petturiti & Vantaggi, 2022) or in the correction of incoherent beliefs (Petturiti & Vantaggi, 2022); and (iii) the inner approximations in terms of distortion models are linked with the notion of incenter of a credal set, complementing in this way our analysis in Miranda and Montes (2023) and showing a connection with coalitional game theory.

Table 1 summarises some features of inner and outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$.

We observe that the properties satisfied by the inner approximation are, in most cases, similar to those of the outer approximations (Miranda et al., 2021; Montes et al., 2018, 2019).

7.2 Approximations of coherent lower probabilities in decision making problems

As argued in some references such as Grabisch (2016), Jansen et al. (2018), Keith and Ahner (2021), Troffaes (2007), decision making is an area where lower probabilities arise

naturally due to the difficulty that entails at times the elicitation of the probability measure that models the problem uncertainty. In Sect. 5 we have discussed how (inner and outer) approximations can be used within this framework to ease the computations. Our motivation is that the lack of 2-monotonicity hinders the computation of the optimal alternatives, because it renders more difficult determining the natural extension of the coherent lower and upper probabilities. We have shown that for some of the criteria (maximality, interval dominance and E-admissibility) it is possible to establish a connection between the optimal alternatives of the initial and transformed models, and that we can bound the error in terms of the BV-distance between them. This establishes a kind of continuity property: if the transformed model is close enough to the initial one, the change in the (lower or upper) expectations of the alternatives shall be small as well, and this can be used in the estimation of the set of optimal alternatives.

This has been exemplified in Sect. 6 where we have used inner and outer approximations in a decision making problem where the preferences depend on both cardinal and ordinal values and the uncertainty is given in terms of a set of probability measures. As we discussed in Sect. 6.4, our approach simplifies computations due to the practical advantages of 2-monotonicity.

7.3 Extension to infinite spaces

One critical assumption in this paper is that we are working with finite possibility spaces, and the sharp reader may wonder about the extent to which our work can be applied when the cardinality of $\mathcal{X}$ is infinite. While at a top level of generality the problem of approximating a coherent lower probability by a 2-monotone one can still be formulated, a number of technical difficulties are encountered quickly:

- One of the main advantages of using 2-monotone approximations on finite possibility spaces is that their credal set has at most $|\mathcal{X}|!$ different extreme points and that they can be easily obtained (Choquet, 1953). This is helpful because it makes computationally easier to determine the optimal solutions of a decision problem under the main criteria considered in the literature. If we move to infinite spaces, though, the number of extreme points need not be finite, and the benefits of using 2-monotonicity dilute somewhat.
- The connection with incenters established in Sect. 4.4 relies on the assumption that all proper subsets of the possibility space have strictly positive lower probability; this will not hold if the possibility space is uncountable. In addition, for the geometric interpretation we should first generalise the work in Miranda and Montes (2023).
- In order to determine the approximation that is "closest" to the original model, we have used the distance proposed by Baroni and Vicig as well as the quadratic distance. The expressions we have given for these distances are valid for the finite case only, and while it is possible to give extensions to arbitrary possibility spaces, the computation of the distance becomes more complex in that case.
- Related to the previous point, the computation of the (inner or outer) approximation had led us to solve linear or quadratic problems, that can be done efficiently for finite possibility spaces but becomes harder for arbitrary ones.

For all these reasons, we believe that extending our approach to infinite possibility spaces will be challenging and may not yield results as satisfactory as those presented in this paper.

7.4 Future research

Besides the extension to non-finite possibility spaces mentioned in the previous paragraph, it would be of interest to analyse the existence and computation of inner approximations in other families of imprecise models, such as probability intervals, p-boxes or possibility measures. For example, in this latter family it can be easily proved that an inner approximation exists if and only if there is an element $x \in \mathcal{X}$ satisfying $\overline{P}(\{x\}) = 1$, and that in that case there is a unique non-dominating inner approximation, given by $\overline{Q}(A) = \max_{x \in A} \overline{P}(\{x\})$.

It would be interesting as well to deepen in the comparison between the initial and the transformed models, along the lines of Proposition 21. Finally, it would be interesting to provide a geometric perspective on the transformations, along the lines of our comments in Sect. 4.4.

Appendix A. Proofs

Proof of Proposition 3 In all cases, the feasible region of the optimisation problem is non-empty because any probability $P \in \mathcal{M}(\underline{P})$ is an inner approximation of $\underline{P}$ that belongs to both $\mathcal{C}_2$ and $\mathcal{C}_\infty$. From here, the result follows with a proof analogous to those of Montes et al. (2018, Prop. 1, Sec. 5.1) and Montes et al. (2019, Prop. 2, Prop. 3). □

Proof of Proposition 4 The proof is analogous to that of Miranda et al. (2021, Prop. 1). □

Proof of Proposition 5 (i) First of all that $\underline{Q} \in \mathcal{C}$ because $\mathcal{C}' \subseteq \mathcal{C}$, and as a consequence it is an inner approximation of $\underline{P}$ in $\mathcal{C}$. If it is dominating, then there is some $\underline{P}' \in \mathcal{C}$ such that $\underline{Q} \gvertneqq \underline{P}' \geq \underline{P}$; moreover, we can assume without loss of generality that $\underline{P}'$ belongs to $\tilde{\mathcal{C}}^{ia}(\underline{P})$. Consequently, $\underline{Q}$ belongs to $\tilde{\mathcal{C}}'^{ia}(\underline{P}')$.

(ii) Similarly, if $\underline{Q} \notin \mathcal{C}_{\mathrm{BV}}^{ia}(\underline{P})$, then there exists $\underline{P}' \in \mathcal{C}_{\mathrm{BV}}^{ia}(\underline{P})$ such that $\underline{Q} \geq \underline{P}'$ and then, by construction, $\underline{Q} \in \mathcal{C}_{\mathrm{BV}}^{ia}(\underline{P}')$.

□

Lemma 22 *A coherent lower probability on a finite possibility space $\mathcal{X}$ is maximally imprecise if and only if there exists some $P \in \mathcal{M}(\underline{P})$ such that $P(A) \in (\underline{P}(A), \overline{P}(A))$ for every $A \neq \emptyset, \mathcal{X}$.*

Proof That the condition is sufficient is trivial. To see that it is also necessary, assume that $\underline{P}$ is maximally imprecise. By coherence, for any $A \neq \emptyset, \mathcal{X}$, there exist $Q_A^1, Q_A^2 \in \mathcal{M}(\underline{P})$ such that $Q_A^1(A) = \underline{P}(A)$ and $Q_A^2(A) = \overline{P}(A)$. Considering

$$P_0 = \frac{1}{2(2^n - 2)} \sum_{A \neq \emptyset, \mathcal{X}} (Q_A^1 + Q_A^2),$$

we obtain a probability measure P_0 that also belongs to $\mathcal{M}(\underline{P})$, because this set is convex, and such that $\underline{P}(B) < P_0(B) < \overline{P}(B)$ for every $B \neq \emptyset, \mathcal{X}$. $\qquad\square$

Proof of Proposition 6 To see that the condition is necessary, note that given an inner approximation $\underline{P}_{\mathrm{LV}}$ of $\underline{P}$ Eq. (3) implies $\min_{A \neq \emptyset, \mathcal{X}} \{\overline{P}_{\mathrm{LV}}(A) - \underline{P}_{\mathrm{LV}}(A)\} = \delta > 0$, hence a coherent lower probability that is not maximally imprecise cannot have an inner approximation in $\mathcal{C}_{\mathrm{LV}}$.

To see that the condition is also sufficient, let P_0 be a probability measure satisfying $\underline{P}(A) < P_0(A) < \overline{P}(A)$ for every $A \neq \emptyset, \mathcal{X}$, existing by Lemma 22. Then

$$\varepsilon = \min_{A \neq \emptyset, \mathcal{X}} \frac{P_0(A)}{\underline{P}(A)} > 1,$$

whence taking $\delta = 1 - 1/\varepsilon \in (0, 1)$, it holds that $P_0(A) \geq \frac{\underline{P}(A)}{1-\delta}$ for any $A \neq \emptyset, \mathcal{X}$, and as a consequence $(1 - \delta)P_0(A) \geq \underline{P}(A)$ for every $A \neq \emptyset, \mathcal{X}$. Hence, the LV model associated with (P_0, δ) is an inner approximation of $\underline{P}$. $\qquad\square$

Lemma 23 *Consider $\Lambda \subseteq \mathbb{R}$ and let $(\underline{Q}_\delta)_{\delta \in \Lambda}$ be a directed family of lower probabilities that avoid sure loss. Then the intersection $\cap_{\delta \in \Lambda} \mathcal{M}(\underline{Q}_\delta)$ is non-empty.*

Proof Since $\mathcal{X}$ is finite, the weak-* topology on the family of credal sets on $\mathcal{X}$ is equivalent to the Euclidean topology on the sets of mass functions. Thus, we obtain that $(\mathcal{M}(\underline{Q}_\delta))_{\delta \in \Lambda}$ is a directed family of non-empty compact sets. As a consequence, their intersection is non-empty. $\qquad\square$

Proof of Proposition 7 This follows applying Lemma 23 to $(\underline{Q}_{\mathrm{LV}}^\delta)_{\delta \in \Lambda_{\mathrm{LV}}}$. $\qquad\square$

Proof of Theorem 8 Let $\delta \in (0, 1)$, and let $\underline{Q}_{\mathrm{LV}}^\delta$ be the lower probability defined in Eq. (5). From Walley (1991), it avoids sure loss if and only if for every $k \in \mathbb{N}$ and $A_1, \ldots, A_k \subseteq \mathcal{X}$, it holds that

$$\sup_{x \in \mathcal{X}} \left(\sum_{i=1}^k I_{A_i}(x) - \sum_{i=1}^k \underline{Q}_{\mathrm{LV}}^\delta(A_i) \right) \geq 0. \tag{A.1}$$

Without loss of generality, we may assume that all the events $A_1, \ldots, A_k$ are proper subsets of $\mathcal{X}$ ($I_{A_i} - \underline{Q}_{\mathrm{LV}}^\delta(A_i)$ would be constant on zero otherwise). Taking into account that $\mathcal{X}$ is finite, Eq. (A.1) is equivalent to:

$$\max_{x \in \mathcal{X}} \sum_{i=1}^k I_{A_i}(x) - \frac{1}{1-\delta} \sum_{i=1}^k \underline{P}(A_i) \geq 0 \Leftrightarrow \max_{x \in \mathcal{X}} (1-\delta) \sum_{i=1}^k I_{A_i}(x) - \sum_{i=1}^k \underline{P}(A_i) \geq 0$$

$$\Leftrightarrow (\exists x \in \mathcal{X}) \left((1-\delta) \sum_{i=1}^k I_{A_i}(x) - \sum_{i=1}^k \underline{P}(A_i) \geq 0 \right)$$

$$\Leftrightarrow (\exists x \in \cup_{i=1}^k A_i) \left((1-\delta) \sum_{i=1}^k I_{A_i}(x) - \sum_{i=1}^k \underline{P}(A_i) \geq 0 \right)$$

$$\Leftrightarrow (\exists x \in \cup_{i=1}^{k} A_i) \left(\delta \sum_{i=1}^{k} I_{A_i}(x) \le \sum_{i=1}^{k} I_{A_i}(x) - \sum_{i=1}^{k} \underline{P}(A_i) \right)$$

$$\Leftrightarrow \delta \le \max_{x \in \cup_{i=1}^{k} A_i} \frac{\sum_{i=1}^{k} I_{A_i}(x) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{i=1}^{k} I_{A_i}(x)},$$

where the third equivalence follows from the assumption $\underline{P}(A_i) > 0$ for every $A_i \ne \emptyset$. If we consider now δ_{LV}, we deduce that for any family $A_1, \ldots, A_k$ of proper subsets of $\mathcal{X}$, it must be

$$\delta_{LV} \le \max_{x \in \cup_{i=1}^{k} A_i} \frac{\sum_{i=1}^{k} I_{A_i}(x) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{i=1}^{k} I_{A_i}(x)}.$$

In fact, we can express it as

$$\delta_{LV} = \min_{A_1, \ldots, A_k} \max_{x \in \cup_{i=1}^{k} A_i} \frac{\sum_{i=1}^{k} I_{A_i}(x) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{i=1}^{k} I_{A_i}(x)} \tag{A.2}$$

$$\le \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} \max_{x \in \cup_{A \in \mathcal{A}} A} \frac{\sum_{A \in \mathcal{A}} I_A(x) - \sum_{A \in \mathcal{A}} \underline{P}(A)}{\sum_{A \in \mathcal{A}} I_A(x)},$$

where the inequality follows because $\mathbb{A}(\mathcal{X})$ contains finite families of subsets of $\mathcal{X}$ with the additional constraint that the sum of their indicator functions is constant.

Let $A_1, \ldots, A_k$ be a family where the minimum in Eq. (A.2) is attained, and take x^* such that

$$\sum_{i=1}^{k} I_{A_i}(x^*) = \max_{x \in \mathcal{X}} \sum_{i=1}^{k} I_{A_i}(x).$$

Then:

$$\delta_{LV} = \frac{\sum_{i=1}^{k} I_{A_i}(x^*) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{i=1}^{k} I_{A_i}(x^*)}.$$

Complete now $A_1, \ldots, A_k$ with $B_1, \ldots, B_l$ so that

$$\sum_{i=1}^{k} I_{A_i}(x) + \sum_{j=1}^{l} I_{B_j}(x) = \sum_{i=1}^{k} I_{A_i}(x^*) \quad \forall x \in \mathcal{X}.$$

To see that this can be done, simply observe that if we express

$$\sum_{i=1}^{k} I_{A_i} = c_1 I_{C_1} + c_2 I_{C_2} + \cdots + c_m I_{C_m}$$

for $\sum_{i=1}^{k} I_{A_i}(x^*) = c_1 > c_2 > \cdots > c_m = 0$ and pairwise disjoint sets $C_1, \ldots, C_m$ then we can consider the family

$$\{B_1, \ldots, B_l\} := \{\underbrace{C_2, \ldots, C_2}_{c_1 - c_2}, \underbrace{C_3, \ldots, C_3}_{c_1 - c_3}, \ldots, \underbrace{C_m, \ldots, C_m}_{c_1}\},$$

and then we obtain

$$\sum_{i=1}^{k} I_{A_i}(x) + \sum_{j=1}^{l} I_{B_j}(x) = c_1 = \sum_{i=1}^{k} I_{A_i}(x^*) \quad \forall x \in \mathcal{X}.$$

Now,

$$\max_{x \in \mathcal{X}} \frac{\sum_{i=1}^{k} I_{A_i}(x) + \sum_{j=1}^{l} I_{B_j}(x) - \sum_{i=1}^{k} \underline{P}(A_i) - \sum_{j=1}^{l} \underline{P}(B_j)}{\sum_{i=1}^{k} I_{A_i}(x) + \sum_{j=1}^{l} I_{B_j}(x)}$$

$$= \frac{\sum_{i=1}^{k} I_{A_i}(x^*) + \sum_{j=1}^{l} I_{B_j}(x^*) - \sum_{i=1}^{k} \underline{P}(A_i) - \sum_{j=1}^{l} \underline{P}(B_j)}{\sum_{i=1}^{k} I_{A_i}(x^*) + \sum_{j=1}^{l} I_{B_j}(x^*)}$$

$$= \frac{\sum_{i=1}^{k} I_{A_i}(x^*) - \sum_{i=1}^{k} \underline{P}(A_i) - \sum_{j=1}^{l} \underline{P}(B_j)}{\sum_{i=1}^{k} I_{A_i}(x^*)}$$

$$\leq \frac{\sum_{i=1}^{k} I_{A_i}(x^*) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{i=1}^{k} I_{A_i}(x^*)} = \delta_{\mathrm{LV}}.$$

Therefore, the minimum is attained with families whose sum is constant. Moreover, given a family $\mathcal{A} \in \mathbb{A}(\mathcal{X})$ where the minimum is attained, it holds that:

$$\delta_{\mathrm{LV}} = \max_{x \in \mathcal{X}} \frac{\sum_{A \in \mathcal{A}} I_A(x) - \sum_{i=1}^{k} \underline{P}(A_i)}{\sum_{A \in \mathcal{A}} I_A(x)} = \frac{\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A}} \underline{P}(A)}{\beta_{\mathcal{A}}} = 1 - \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \underline{P}(A)$$

whence Eq. (8) holds. □

In order to prove Theorem 9, we need first to recall a couple of lemmas.

Lemma 24 *Let $\underline{P}$ be a maximally imprecise coherent lower probability, and consider $\mathcal{A} \in \mathbb{A}(\mathcal{X})$ such that there exists $\mathcal{A}_1 \in \mathbb{A}(\mathcal{X})$ with $\mathcal{A}_1 \subset \mathcal{A}$. Then:*

(a) $\mathcal{A}_2 = \mathcal{A} \backslash \mathcal{A}_1 \in \mathbb{A}(\mathcal{X})$ and $\beta_{\mathcal{A}} = \beta_{\mathcal{A}_1} + \beta_{\mathcal{A}_2}$.
(b) For any function $h : \mathbb{A}(\mathcal{X}) \to \mathbb{R}$ satisfying

$$h(\mathcal{A}_1 \dot{\cup} \mathcal{A}_2) = \frac{\beta_{\mathcal{A}_1}}{\beta_{\mathcal{A}_1} + \beta_{\mathcal{A}_2}} h(\mathcal{A}_1) + \frac{\beta_{\mathcal{A}_2}}{\beta_{\mathcal{A}_1} + \beta_{\mathcal{A}_2}} h(\mathcal{A}_2), \tag{A.3}$$

it holds that $h(\mathcal{A}_1 \dot{\cup} \mathcal{A}_2) \geq \min\{h(\mathcal{A}_1), h(\mathcal{A}_2)\}$. Here, $\mathcal{A}_1 \dot{\cup} \mathcal{A}_2$ denotes the element of $\mathbb{A}(\mathcal{X})$ obtained by putting together the events in $\mathcal{A}_1$ and in $\mathcal{A}_2$.

Proof The proof is an extension of Miranda and Montes (2023, Lem. 18).

(a) Let $\mathcal{A} \in \mathbb{A}(\mathcal{X})$ and assume that there exists $\mathcal{A}_1 \in \mathbb{A}(\mathcal{X})$ such that $\mathcal{A}_1 \subset \mathcal{A}$. This means that $\sum_{A \in \mathcal{A}_1} I_A = \beta_{\mathcal{A}_1} < \beta_{\mathcal{A}}$. Take $\mathcal{A}_2 = \mathcal{A} \setminus \mathcal{A}_1 \subset \mathcal{A}$. It holds that:

$$\sum_{A \in \mathcal{A}_2} I_A = \sum_{A \in \mathcal{A}} I_A - \sum_{A \in \mathcal{A}_1} I_A = \beta_{\mathcal{A}} - \beta_{\mathcal{A}_1} \in \mathbb{N},$$

whence $\mathcal{A}_2 \in \mathbb{A}(\mathcal{X})$ and also $\beta_{\mathcal{A}} = \beta_{\mathcal{A}_1} + \beta_{\mathcal{A}_2}$.
(b) Trivial.

□

Lemma 25 (Miranda & Montes, 2023, Lem. 19) *Let $\underline{P}$ be a maximally imprecise coherent lower probability, and let $\mathcal{A} = (A_i)_{i \in I} \in \mathbb{A}(\mathcal{X})$.*

(i) If $\beta_{\mathcal{A}} = 1$, then $\mathcal{A}$ is a partition of $\mathcal{X}$.
(ii) If $\beta_{\mathcal{A}} = |\mathcal{A}| - 1$, then $\mathcal{A}^c = (A_i^c)_{i \in I}$ is a partition of $\mathcal{X}$.

(iii) *If $1 < \beta_{\mathcal{A}} < |\mathcal{A}| - 1$ and for every $A, B \in \mathcal{A}$ at least one of $A \cap B$, $A \backslash B$ and $B \backslash A$ is empty, then there exists $\mathcal{A}_1 \in \mathbb{A}^*(\mathcal{X})$ such that $\mathcal{A}_1 \subset \mathcal{A}$.*

For any $\mathcal{A} \in \mathbb{A}(\mathcal{X})$, let

$$h_{\mathcal{A}}^{\text{LV}} = 1 - \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \underline{P}(A) = \frac{1}{\beta_{\mathcal{A}}} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A}} \underline{P}(A) \right).$$

It is easy to show that $h_{\mathcal{A}}^{\text{LV}}$ satisfies Eq. (A.3).

Proof of Theorem 9 Consider an element $\mathcal{A} \in \mathbb{A}(\mathcal{X})$.

If $\beta_{\mathcal{A}} = 1$, then $\mathcal{A}$ is partition so $\mathcal{A} \in \mathbb{A}^*(\mathcal{X})$, and moreover

$$h_{\mathcal{A}}^{\text{LV}} = 1 - \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \underline{P}(A) = 1 - \sum_{A \in \mathcal{A}} \underline{P}(A).$$

If $\beta_{\mathcal{A}} = |\mathcal{A}| - 1$, then $\mathcal{A}^c = (A_i^c)_{i=1,\dots,k}$ is a partition, so it belongs to $\mathbb{A}^*(\mathcal{X})$, and

$$
\begin{aligned}
h_{\mathcal{A}}^{\text{LV}} &= \frac{1}{\beta_{\mathcal{A}}} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A}} \underline{P}(A) \right) = \frac{1}{|\mathcal{A}| - 1} \left(|\mathcal{A}| - 1 - \sum_{A \in \mathcal{A}} \underline{P}(A) \right) \\
&= \frac{1}{|\mathcal{A}| - 1} \left(|\mathcal{A}| - 1 - \sum_{A \in \mathcal{A}} (1 - \overline{P}(A^c)) \right) = \frac{1}{|\mathcal{A}| - 1} \left(\sum_{A \in \mathcal{A}} \overline{P}(A^c) - 1 \right) \\
&= \frac{1}{|\mathcal{A}| - 1} \left(\sum_{A \in \mathcal{A}^c} \overline{P}(A) - 1 \right) = \frac{1}{|\mathcal{A}^c| - 1} \left(\sum_{A \in \mathcal{A}^c} \overline{P}(A) - 1 \right)
\end{aligned}
$$

where the last equality holds because the number of elements in $\mathcal{A}$ and $\mathcal{A}^c$ is the same, hence $|\mathcal{A}| = |\mathcal{A}^c|$. Therefore, the value $\delta_{\text{LV}} = \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} h_{\mathcal{A}}^{\text{LV}}$ satisfies:

$$\delta_{\text{LV}} \leq \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X}) \text{ or } \mathcal{A}^c \in \mathbb{A}^*(\mathcal{X})} h_{\mathcal{A}}^{\text{LV}} = \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} \left(1 - \sum_{A \in \mathcal{A}} \underline{P}(A), \frac{\sum_{A \in \mathcal{A}} \overline{P}(A) - 1}{|\mathcal{A}| - 1} \right).$$

To see that we have the equality, let $\mathcal{A} = (A_i)_{i=1,\dots,k}$ be an element in $\mathbb{A}(\mathcal{X})$ where the minimum in Eq. (8) is attained. Assume now that $1 < \beta_{\mathcal{A}} < |\mathcal{A}| - 1$, and let us prove that it is possible to find some $\mathcal{A}^* \in \mathbb{A}^*(\mathcal{X})$ such that $\beta_{\mathcal{A}^*} < \beta_{\mathcal{A}}$ and where the value in Eq. (8) is attained.

From item (iii) in Lemma 25 we deduce that either there is $\mathcal{A}^* \in \mathbb{A}(\mathcal{X})$ with $\mathcal{A}^* \subset \mathcal{A}$ or there are two different $A_i, A_j \in \mathcal{A}$ with $A_i \cap A_j \neq \emptyset$, $A_i \backslash A_j \neq \emptyset$ and $A_j \backslash A_i \neq \emptyset$. In this second case, applying 2-monotonicity with the sets A_i and A_j above we deduce that:

$$
\begin{aligned}
h_{\mathcal{A}}^{\text{LV}} &= \frac{1}{\beta_{\mathcal{A}}} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A}} \underline{P}(A) \right) \\
&= \frac{1}{\beta_{\mathcal{A}}} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A} \backslash \{A_i, A_j\}} \underline{P}(A) - \underline{P}(A_i) - \underline{P}(A_j) \right) \\
&\geq \frac{1}{\beta_{\mathcal{A}}} \left(\beta_{\mathcal{A}} - \sum_{A \in \mathcal{A} \backslash \{A_i, A_j\}} \underline{P}(A) - \underline{P}(A_i \cap A_j) - \underline{P}(A_i \cup A_j) \right) = h_{\mathcal{A}_1}^{\text{LV}},
\end{aligned}
$$

where $\mathcal{A}_1 = \left(\mathcal{A}\backslash\{A_i, A_j\}\right) \cup (A_i \cap A_j, A_i \cup A_j)$, using that $\beta_{\mathcal{A}_1} = \beta_{\mathcal{A}}$. Thus, the minimum in Eq. (8) is also attained in $\mathcal{A}_1$.

Now, if in $\mathcal{A}_1$ it is possible to find two different events B_i, B_j with $B_i \cap B_j \neq \emptyset$, $B_i \backslash B_j \neq \emptyset$ and $B_j \backslash B_i \neq \emptyset$ a similar reasoning shows that $\mathcal{A}_2 = \mathcal{A}_1 \cup \left(B_i \cup B_j, B_i \cap B_j\right)\backslash(B_i, B_j)$ also satisfies $\beta_{\mathcal{A}_2} = \beta_{\mathcal{A}_1}$ and $h_{\mathcal{A}}^{\mathrm{LV}} = h_{\mathcal{A}_1}^{\mathrm{LV}} = h_{\mathcal{A}_2}^{\mathrm{LV}}$. Iterating the procedure, we find after a finite number of steps that there are no different events C and D in the family $\mathcal{A}_k$ such that $C \cap D \neq \emptyset$, $C\backslash D \neq \emptyset$ and $D\backslash C \neq \emptyset$. But then, applying Lemma 25 we deduce that there is $\mathcal{A}^* \in \mathbb{A}^*(\mathcal{X})$ with $\mathcal{A}^* \subset \mathcal{A}_k$.

Since $\mathcal{A}_k = \mathcal{A}^* \dot{\cup} (\mathcal{A}_k \backslash \mathcal{A}^*)$, we deduce from Lemma 24 that either $h_{\mathcal{A}_k}^{\mathrm{LV}} = h_{\mathcal{A}^*}^{\mathrm{LV}}$ or $h_{\mathcal{A}_k}^{\mathrm{LV}} = h_{\mathcal{A}_k \backslash \mathcal{A}^*}^{\mathrm{LV}}$. Since both $\beta_{\mathcal{A}^*}$ and $\beta_{\mathcal{A}_k \backslash \mathcal{A}^*}$ are strictly smaller than $\beta_{\mathcal{A}_k}$, we deduce that we can find another element of $\mathbb{A}(\mathcal{X})$ where the value δ_{LV} is attained and with a smaller value of $\beta_{\mathcal{A}}$. If we repeat this process we end up with a family $\mathcal{A}' \in \mathbb{A}(\mathcal{X})$ such that $\beta_{\mathcal{A}'} = 1$, and where the minimum value in Eq. (8) is attained, at which point we apply the first part of the proof. $\qquad\square$

Proof of Proposition 10 To see that the condition is necessary, assume ex-absurdo that there exists $A \neq \emptyset, \mathcal{X}$ such that $\overline{P}(A) = \underline{P}(A)$ and let $\overline{P}_{\mathrm{PMM}}$ be the upper probability of a PMM defined using (P_0, δ) such that $\overline{P}_{\mathrm{PMM}} \leq \overline{P}$. Then, it should be $\overline{P}_{\mathrm{PMM}}(A) = \overline{P}(A) = \underline{P}(A) = \underline{P}_{\mathrm{PMM}}(A)$. From the definition of the PMM, this can only hold if $\overline{P}_{\mathrm{PMM}}(A) = \underline{P}_{\mathrm{PMM}}(A) \in \{0, 1\}$, meaning that it should be $\underline{P}(A) \in \{0, 1\}$, a contradiction with our assumption of $\underline{P}(A) \in (0, 1)$ for every $A \neq \emptyset, \mathcal{X}$.

Conversely, if $\overline{P}(A) > \underline{P}(A)$ for any $A \neq \emptyset, \mathcal{X}$ then by Lemma 22 there is some $P_0 \in \mathcal{M}(\underline{P})$ such that $P_0(A) \in (\underline{P}(A), \overline{P}(A))$ for every event $A \neq \emptyset, \mathcal{X}$. If we now consider $\delta > 0$ small enough such that $(1 + \delta)P_0(A) < \overline{P}(A)$ for any $A \neq \emptyset, \mathcal{X}$, we obtain that the PMM determined by (P_0, δ) is an inner approximation of $\overline{P}$. $\qquad\square$

Proof of Proposition 11 This follows applying Lemma 23 to $(\underline{Q}_{\mathrm{PMM}}^{\delta})_{\delta \in \Lambda_{\mathrm{PMM}}}$. $\qquad\square$

Proof of Theorem 12 Let $\delta > 0$, and let $\overline{P}_{\delta}$ be the upper probability defined in Eq. (10). From Walley (1991), it avoids sure loss if and only if for every $k \in \mathbb{N}$ and every $A_1, \ldots, A_k \subseteq \mathcal{X}$ it holds that:

$$\max_{x \in \mathcal{X}} \left(\sum_{i=1}^{k} \overline{Q}_{\mathrm{PMM}}^{\delta}(A_i) - \sum_{i=1}^{k} I_{A_i}(x) \right) \geq 0. \tag{A.4}$$

Without loss of generality, we may assume that all the events $A_1, \ldots, A_k$ are proper subsets of $\mathcal{X}$ ($\overline{Q}_{\mathrm{PMM}}^{\delta}(A_i) - I_{A_i}$ would be constant on zero otherwise) and also that $\cup_{i=1}^{k} A_i = \mathcal{X}$ (otherwise (A.4) holds trivially taking $x \in (\cup_{i=1}^{k} A_i)^c$). Taking into account that $\mathcal{X}$ is finite, Eq. (A.4) is equivalent to:

$$\max_{x \in \mathcal{X}} \sum_{i=1}^{k} \left(\frac{\overline{P}(A_i)}{1+\delta} - I_{A_i}(x) \right) \geq 0 \Leftrightarrow \max_{x \in \mathcal{X}} \sum_{i=1}^{k} \left(\overline{P}(A_i) - (1+\delta)I_{A_i}(x) \right) \geq 0$$

$$\Leftrightarrow (\exists x \in \mathcal{X}) \left(\sum_{i=1}^{k} \left(\overline{P}(A_i) - (1+\delta)I_{A_i}(x) \right) \geq 0 \right)$$

$$\Leftrightarrow \min_{x \in \mathcal{X}} (1+\delta) \sum_{i=1}^{k} I_{A_i}(x) \leq \sum_{i=1}^{k} \overline{P}(A_i) \Leftrightarrow \delta \leq \frac{\sum_{i=1}^{k} \overline{P}(A_i)}{\min_{x \in \mathcal{X}} \sum_{i=1}^{k} I_{A_i}(x)} - 1$$

for any $A_1, \ldots, A_k \neq \emptyset, \mathcal{X}$. In particular, given $\mathcal{A} = \{A_1, \ldots, A_k\} \in \mathbb{A}(\mathcal{X})$ with $\sum_{i=1}^k I_{A_i} = \beta_{\mathcal{A}}$, it should be $\delta \leq \frac{1}{\beta_{\mathcal{A}}} \sum_{i=1}^k \overline{P}(A_i) - 1$, whence:

$$\delta_{\text{PMM}} \leq \min_{\mathcal{A}} \frac{1}{\beta_{\mathcal{A}}} \sum_{i=1}^k \overline{P}(A_i) - 1.$$

To see that we have the equality, consider an arbitrary family $A_1, \ldots, A_k$, and take $x^* \in \mathcal{X}$ so that $\min_{x \in \mathcal{X}} \sum_{i=1}^k I_{A_i}(x) = \sum_{i=1}^k I_{A_i}(x^*)$. For each $i = 1, \ldots, k$ take $B_i \subseteq A_i$ such that $\sum_{i=1}^k I_{B_i}(x) = \sum_{i=1}^k I_{A_i}(x^*)$ for every $x \in \mathcal{X}$. Then

$$\frac{\sum_{i=1}^k \overline{P}(B_i)}{\min_{x \in \mathcal{X}} \sum_{i=1}^k I_{B_i}(x)} - 1 = \frac{\sum_{i=1}^k \overline{P}(B_i)}{\min_{x \in \mathcal{X}} \sum_{i=1}^k I_{A_i}(x)} - 1 \leq \frac{\sum_{i=1}^k \overline{P}(A_i)}{\min_{x \in \mathcal{X}} \sum_{i=1}^k I_{A_i}(x)} - 1.$$

Thus:

$$\delta_{\text{PMM}} = \min_{\mathcal{A}} \left\{ \frac{1}{\beta_{\mathcal{A}}} \sum_{i=1}^k \overline{P}(A_i) - 1 \right\}.$$

and as a consequence Eq. (12) holds. $\square$

In order to prove Theorem 13, let us denote, for any $\mathcal{A} \in \mathbb{A}(\mathcal{X})$:

$$h_{\mathcal{A}}^{\text{PMM}} = \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \overline{P}(A) - 1 = \frac{1}{\beta_{\mathcal{A}}} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - \beta_{\mathcal{A}} \right).$$

It is easy to see that $h_{\mathcal{A}}^{\text{PMM}}$ satisfies Eq. (A.3).

Proof of Theorem 13 Consider an element $\mathcal{A} \in \mathbb{A}(\mathcal{X})$.

If $\beta_{\mathcal{A}} = 1$, then $\mathcal{A}$ is partition so $\mathcal{A} \in \mathbb{A}^*(\mathcal{X})$, and moreover

$$h_{\mathcal{A}}^{\text{PMM}} = \frac{1}{\beta_{\mathcal{A}}} \sum_{A \in \mathcal{A}} \overline{P}(A) - 1 = \sum_{A \in \mathcal{A}} \overline{P}(A) - 1.$$

If $\beta_{\mathcal{A}} = |\mathcal{A}| - 1$, then $\mathcal{A}^c = (A_i^c)_{i=1,\ldots,k}$ is a partition, so it belongs to $\mathbb{A}^*(\mathcal{X})$. Also:

$$h_{\mathcal{A}}^{\text{PMM}} = \frac{1}{\beta_{\mathcal{A}}} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - \beta_{\mathcal{A}} \right) = \frac{1}{|\mathcal{A}| - 1} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - |\mathcal{A}| + 1 \right)$$

$$= \frac{1}{|\mathcal{A}| - 1} \left(\sum_{A \in \mathcal{A}} \left(1 - \underline{P}(A^c) \right) - |\mathcal{A}| + 1 \right) = \frac{1}{|\mathcal{A}| - 1} \left(1 - \sum_{A \in \mathcal{A}} \underline{P}(A^c) \right)$$

$$= \frac{1}{|\mathcal{A}| - 1} \left(1 - \sum_{A \in \mathcal{A}^c} \underline{P}(A) \right) = \frac{1}{|\mathcal{A}^c| - 1} \left(\sum_{A \in \mathcal{A}^c} \underline{P}(A) - 1 \right)$$

where the last equality holds because the number of elements in $\mathcal{A}$ and $\mathcal{A}^c$ is the same, hence $|\mathcal{A}| = |\mathcal{A}^c|$. Therefore, the value $\delta_{\text{PMM}} = \min_{\mathcal{A} \in \mathbb{A}(\mathcal{X})} h_{\mathcal{A}}^{\text{PMM}}$ satisfies

$$\delta_{\text{PMM}} \leq \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} h_{\mathcal{A}}^{\text{PMM}} = \min_{\mathcal{A} \in \mathbb{A}^*(\mathcal{X})} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - 1, \frac{1 - \sum_{A \in \mathcal{A}} \underline{P}(A)}{|\mathcal{A}| - 1} \right).$$

To see that we have the equality, let $\mathcal{A} = (A_i)_{i=1,\ldots,k}$ be an element in $\mathbb{A}(\mathcal{X})$ where the minimum in Eq. (12) is attained. Assume now that $1 < \beta_{\mathcal{A}} < |\mathcal{A}| - 1$, and let us prove that

it is possible to find some $\mathcal{A}^* \in \mathbb{A}^*(\mathcal{X})$ such that $\beta_{\mathcal{A}^*} < \beta_{\mathcal{A}}$ and where the value in Eq. (12) is attained.

From item (iii) in Lemma 25 we deduce that either there is $\mathcal{A}^* \in \mathbb{A}(\mathcal{X})$ with $\mathcal{A}^* \subset \mathcal{A}$ or there are two different $A_i, A_j \in \mathcal{A}$ with $A_i \cap A_j \neq \emptyset$, $A_i \backslash A_j \neq \emptyset$ and $A_j \backslash A_i \neq \emptyset$. In this second case, applying 2-monotonicity with the sets A_i and A_j above we deduce that:

$$
\begin{aligned}
h_{\mathcal{A}}^{\mathrm{PMM}} &= \frac{1}{\beta_{\mathcal{A}}} \left(\sum_{A \in \mathcal{A}} \overline{P}(A) - \beta_{\mathcal{A}} \right) \\
&= \frac{1}{\beta_{\mathcal{A}}} \left(\sum_{A \in \mathcal{A} \backslash \{A_i, A_j\}} \overline{P}(A) + \overline{P}(A_i) + \overline{P}(A_j) - \beta_{\mathcal{A}} \right) \\
&\geq \frac{1}{\beta_{\mathcal{A}}} \left(\sum_{A \in \mathcal{A} \backslash \{A_i, A_j\}} \overline{P}(A) + \overline{P}(A_i \cap A_j) + \overline{P}(A_i \cup A_j) - \beta_{\mathcal{A}} \right) = h_{\mathcal{A}_1}^{\mathrm{PMM}},
\end{aligned}
$$

where $\mathcal{A}_1 = \left(\mathcal{A} \backslash \{A_i, A_j\} \right) \cup (A_i \cap A_j, A_i \cup A_j)$, using that $\beta_{\mathcal{A}_1} = \beta_{\mathcal{A}}$. Thus, the minimum in Eq. (12) is also attained in $\mathcal{A}_1$.

Now, if in $\mathcal{A}_1$ it is possible to find two different events B_i, B_j with $B_i \cap B_j \neq \emptyset$, $B_i \backslash B_j \neq \emptyset$ and $B_j \backslash B_i \neq \emptyset$ a similar reasoning shows that $\mathcal{A}_2 = \mathcal{A}_1 \cup \left(B_i \cup B_j, B_i \cap B_j \right) \backslash (B_i, B_j)$ also satisfies $\beta_{\mathcal{A}_2} = \beta_{\mathcal{A}_1}$ and $h_{\mathcal{A}}^{\mathrm{PMM}} = h_{\mathcal{A}_1}^{\mathrm{PMM}} = h_{\mathcal{A}_2}^{\mathrm{PMM}}$. Iterating the procedure, we find after a finite number of steps that there are no different events C, D in the family $\mathcal{A}_k$ such that $C \cap D \neq \emptyset$, $C \backslash D \neq \emptyset$ and $D \backslash C \neq \emptyset$. But then, applying Lemma 25 we deduce that there is $\mathcal{A}^* \in \mathbb{A}^*(\mathcal{X})$ with $\mathcal{A}^* \subset \mathcal{A}_k$.

Since $\mathcal{A}_k = \mathcal{A}^* \dot{\cup} (\mathcal{A}_k \backslash \mathcal{A}^*)$, we deduce from Lemma 24 that either $h_{\mathcal{A}_k}^{\mathrm{PMM}} = h_{\mathcal{A}^*}^{\mathrm{PMM}}$ or $h_{\mathcal{A}_k}^{\mathrm{PMM}} = h_{\mathcal{A}_k \backslash \mathcal{A}^*}^{\mathrm{PMM}}$. Since both $\beta_{\mathcal{A}^*}$ and $\beta_{\mathcal{A}_k \backslash \mathcal{A}^*}$ are strictly smaller than $\beta_{\mathcal{A}_k}$, we deduce that we can find another element of $\mathbb{A}(\mathcal{X})$ where the value δ_{PMM} is attained and with a smaller value of $\beta_{\mathcal{A}}$. If we repeat this process we end up with a family $\mathcal{A}' \in \mathbb{A}(\mathcal{X})$ such that $\beta_{\mathcal{A}'} = 1$, and where the minimum value in Eq. (8) is attained, at which point we apply the first part of the proof. $\qquad \square$

Proof of Proposition 14 To see that the condition is necessary, note that since by assumption $\underline{P}(A) > 0$ for every $A \neq \emptyset$, any inner approximation $\underline{P}_{\mathrm{TV}}$ in $\mathcal{C}_{\mathrm{TV}}$ should satisfy $\underline{P}_{\mathrm{TV}}(A) = P_0(A) - \delta$ and $\overline{P}_{\mathrm{TV}}(A) = P_0(A) + \delta$ for any $A \neq \emptyset, \mathcal{X}$. As a consequence, it will be $\overline{P}_{\mathrm{TV}}(A) - \underline{P}_{\mathrm{TV}}(A) = 2\delta > 0$ for any $A \neq \emptyset, \mathcal{X}$, meaning that there cannot be any inner approximation of $\underline{P}$ if it is not maximally imprecise.

Conversely, assume that $\underline{P}(A) < \overline{P}(A)$ for any $A \neq \emptyset, \mathcal{X}$. Applying Lemma 22, there exists $P_0 \in \mathcal{M}(\underline{P})$ such that $\underline{P}(A) < P_0(A) < \overline{P}(A)$ for any $A \neq \emptyset, \mathcal{X}$. Taking δ such that $0 < \delta < \min_{A \neq \emptyset, \mathcal{X}} \left(P_0(A) - \underline{P}(A) \right)$, we obtain that $\delta < P_0(A) - \underline{P}(A)$, so $\underline{P}(A) < P_0(A) - \delta$ for any $A \neq \emptyset, \mathcal{X}$. Hence, P_0 and δ determine a TV model that inner approximates $\underline{P}$. $\qquad \square$

Proof of Proposition 15 This follows applying Lemma 23 to $(\underline{Q}_{\mathrm{TV}}^{\delta})_{\delta \in \Lambda_{\mathrm{TV}}}$. $\qquad \square$

Proof of Proposition 17 For any probability measure P_0 and any $\delta > 0$ it holds that $B_{\mathrm{LV}}^{\delta}(P_0) \cup B_{\mathrm{PMM}}^{\delta}(P_0) \subseteq B_{\mathrm{TV}}^{\delta}(P_0)$ (Montes et al., 2020b, Prop.5.1). Hence, $B_{\mathrm{TV}}^{\delta}(P_0) \subseteq \mathcal{M}(\underline{P})$ implies that $B_{\mathrm{LV}}^{\delta}(P_0) \subseteq \mathcal{M}(\underline{P})$ and $B_{\mathrm{PMM}}^{\delta}(P_0) \subseteq \mathcal{M}(\underline{P})$, and therefore $\delta_{\mathrm{TV}} \leq \min\{\delta_{\mathrm{LV}}, \delta_{\mathrm{PMM}}\}$. $\qquad \square$

Proof of Proposition 18 $\underline{P} \leq \underline{Q}$ implies that $\underline{P}(J_e - J_d) \leq \underline{Q}(J_e - J_d)$ for any $e, d \in D$. Hence, $\underline{Q}(J_e - J_d) \leq 0$ implies $\underline{P}(J_e - J_d) \leq 0$, so any optimal alternative for $\underline{Q}$ under maximality is also optimal for $\underline{P}$. $\qquad \square$

Proof of Proposition 19 $\underline{P} \leq \underline{Q}$ is equivalent to $\overline{P} \geq \overline{Q}$. If $d \in D$ is optimal for $\underline{Q}$ under interval dominance, then $\overline{Q}(J_d) \geq \underline{Q}(J_e)$ for every $e \in D$, but this implies that:

$$\overline{P}(J_d) \geq \overline{Q}(J_d) \geq \underline{Q}(J_e) \geq \underline{P}(J_e) \quad \forall e \in D,$$

meaning that d is optimal too for $\underline{P}$. This implies that $\text{opt}_{\sqsupset_{\underline{P}}} \supseteq \text{opt}_{\sqsupset_{\underline{Q}}}$. □

Proof of Proposition 19 (Proof of Proposition 20) Consider $d \in \text{opt}_{\mathcal{M}(\underline{Q})}$ under E-admissibility. This means that there exists $P \in \mathcal{M}(\underline{Q})$ such that $E_P(J_d) = \max_{d' \in D} E_P(J_{d'})$. Since $\underline{P} \leq \underline{Q}$, then $\mathcal{M}(\underline{P}) \supseteq \mathcal{M}(\underline{Q})$, whence P belongs to $\mathcal{M}(\underline{P})$, and as a consequence d also belongs to $\text{opt}_{\mathcal{M}(\underline{Q})}$ under E-admissibility. □

Proof of Proposition 21 Since $\mathcal{M}(\underline{Q}) \subseteq \mathcal{M}(\underline{P})$, it follows that $\underline{P}(f) \leq \underline{Q}(f)$, whence $| \underline{Q}(f) - \underline{P}(f) | = \underline{Q}(f) - \underline{P}(f)$.

If $\underline{Q}$ is 2-monotone, then $\underline{Q}(f)$ coincides with the Choquet integral of f with respect to $\underline{Q}$, $(C) \int f d\underline{Q}$, while by coherence we have that $\underline{P}(f) \geq (C) \int f d\underline{P}$ (Walley, 1981).

Assume that $f = \sum_{i=1}^{n} x_i I_{A_i}$, for $x_1 \geq x_2 \geq \cdots \geq x_n$ in $[0,1]$ and a partition $\{A_1, \ldots, A_n\}$ of $\mathcal{X}$. Then

$$\underline{Q}(f) - \underline{P}(f) \leq (C) \int f d\underline{Q} - (C) \int f d\underline{P} = \sum_{i=1}^{n} (x_i - x_{i+1})(\underline{Q}(A_i) - \underline{P}(A_i))$$

$$\leq \sum_{i=1}^{n} (\underline{Q}(A_i) - \underline{P}(A_i)) \leq \sum_{A \subseteq \mathcal{X}} \underline{Q}(A) - \underline{P}(A) \leq \delta.$$

By conjugacy, we deduce that also $|\overline{P}(f) - \overline{Q}(f)| = \overline{P}(f) - \overline{Q}(f) \leq \delta$. □

References

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics, 34*, 199–205.

Antonucci, A., de Campos, C., Huber, D., & Zaffalon, M. (2015). Approximate credal network updating by linear programming with applications to decision making. *International Journal of Approximate Reasoning, 58*, 25–38.

Augustin, T., Coolen, F., de Cooman, G., & Troffaes, M. (Eds.). (2014). *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics Wiley.

Baroni, P., & Vicig, P. (2005). An uncertainty interchange format with imprecise probabilities. *International Journal of Approximate Reasoning, 40*, 147–180.

Berger, J. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference, 25*, 303–328.

Bronevich, A., & Augustin, T. (2009) Approximation of coherent lower probabilities by 2-monotone measures. In T. Augustin, F. P. A. Coolen, S. Moral, & M. C. M. Troffaes (Eds.), *ISIPTA'09—Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications* (pp. 61–70).

Bronevich, A. (2007). Necessary and sufficient consensus conditions for the eventwise aggregation of lower probabilities. *Fuzzy Sets and Systems, 158*, 881–894.

Chateauneuf, A., & Jaffray, J.-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences, 17*(3), 263–283.

Choquet, G. (1953). Theory of capacities. *Annales de l'Institut Fourier, 5*, 131–295.

Cinfrignini, A., Petturiti, D., & Vantaggi, B. (2023). Envelopes of equivalent martingale measures and a generalized no-arbitrage principle in a finite setting. *Annals of Operations Research, 321*, 103–137.

de Angelis, M., Gray, A., Ferson, S., & Patelli, E. (2023). Robust online updating of a digital twin with imprecise probability. *Mechanical Systems and Signal Processing, 186*, 109877.

de Finetti, B. (1974–1975). *Theory of probability: A critical introductory treatment*. Wiley.

Derks, J., & Kuipers, J. (2002). On the number of extreme points of the core of a transferable utility game. In *Chapters in game theory. Theory and decision library C* (Vol. 31, pp. 83–97). Springer.

Destercke, S. (2013). Independence and 2-monotonicity: Nice to have, hard to keep. *International Journal of Approximate Reasoning, 54*(4), 478–490.

Destercke, S., Montes, I., & Miranda, E. (2022). Processing multiple distortion models: A comparative study. *International Journal of Approximate Reasoning, 145C*, 91–120.

Dubois, D., Prade, H., & Sandri, S. (1993). On possibility/probability transformations. Theory and decision library. In R. Lower & M. Roubens (Eds.), *Fuzzy logic* (Vol. 12, pp. 103–112). Springer.

Dubois, D., & Prade, H. (1988). *Possibility theory*. Plenum Press.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics, 18*, 141–153.

Grabisch, M. (2016). *Set functions, games and capacities in decision making*. Springer.

Huber, P. J. (1981). *Robust statistics*. Wiley.

Jaffray, J. (1995). On the maximum-entropy probability which is consistent with a convex capacity. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 3*(1), 27–33.

Jansen, C., Schollmeyer, G., & Augustin, T. (2018). Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. *International Journal of Approximate Reasoning, 98*, 112–131.

Keith, A., & Ahner, D. (2021). A survey of decision making and optimization under uncertainty. *Annals of Operations Research, 300*, 319–353.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica, 73*(6), 1849–1892.

Mattei, L., Antonucci, A., Mauá, D., Facchini, A., & Villanueva Serena, J. (2020). Tractable inference in credal sentential decision diagrams. *International Journal of Approximate Reasoning, 125*, 26–48.

Miranda, E., Montes, I., & Presa, A. (2022). Inner approximations of credal sets by non-additive measures. In *19th International conference on information processing and management of uncertainty in knowledge-based systems, IPMU 2022. Communications in computer and information science* (Vol. 1601, pp. 743–756). Springer.

Miranda, E., & Destercke, S. (2015). Extreme points of the credal sets generated by comparative probabilities. *Journal of Mathematical Psychology, 64*(65), 44–57.

Miranda, E., & Montes, I. (2023). Centroids of the core of exact capacities: A comparative study. *Annals of Operations Research, 321*, 409–449.

Miranda, E., Montes, I., & Vicig, P. (2021). On the selection of an optimal outer approximation of a coherent lower probability. *Fuzzy Sets and Systems, 424C*, 1–36.

Montes, I., Miranda, E., & Destercke, S. (2019). Pari-mutuel probabilities as an uncertainty model. *Information Sciences, 481*, 550–573.

Montes, I., Miranda, E., & Destercke, S. (2020). Unifying neighbourhood and distortion models: Part I-New results on old models. *International Journal of General Systems, 49*(6), 602–635.

Montes, I., Miranda, E., & Destercke, S. (2020). Unifying neighbourhood and distortion models: Part II-New models and synthesis. *International Journal of General Systems, 49*(6), 636–674.

Montes, I., Miranda, E., & Vicig, P. (2018). 2-Monotone outer approximations of coherent lower probabilities. *International Journal of Approximate Reasoning, 101*, 181–205.

Montes, I., Miranda, E., & Vicig, P. (2019). Outer approximating coherent lower probabilities with belief functions. *International Journal of Approximate Reasoning, 110*, 1–30.

Pelessoni, R., Vicig, P., & Zaffalon, M. (2010). Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning, 51*, 1145–1158.

Pericchi, L. R., & Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *International Statistical Review, 59*, 1–23.

Petturiti, D., & Vantaggi, B. (2022). How to assess coherent beliefs: A comparison of different notions of coherence in Dempster-Shafer theory of evidence. In T. Augustin, F. G. Cozman, & G. Wheeler (Eds.), *Reflections on the foundations of probability and statistics: Essays in honor of Teddy Seidenfeld. Theory and decision library A* (Vol. 54, pp. 161–185). Springer.

Petturiti, D., & Vantaggi, B. (2019). Conditional submodular Choquet expected values and conditional coherent risk measures. *International Journal of Approximate Reasoning, 113*, 14–38.

Petturiti, D., & Vantaggi, B. (2020). Modeling agent's conditional preferences under objective ambiguity in Dempster–Shafer theory. *International Journal of Approximate Reasoning, 119*, 151–176.

Petturiti, D., & Vantaggi, B. (2022). Probability envelopes and their Dempster-Shafer approximations in statistical matching. *International Journal of Approximate Reasoning, 150*, 199–222.

Sahlin, U., Troffaes, M., & Edsman, L. (2021). Robust decision analysis under severe uncertainty and ambiguous tradeoffs: An invasive species case study. *Risk Analysis, 41*(11), 2140–2153.

Sarin, R., & Wakker, P. (1992). A simple axiomatization of nonadditive expected utility. *Econometrica, 60*(6), 1255–1272.

Savage, L. J. (1954). *The foundations of statistics*. Wiley.

Seidenfeld, T., & Wasserman, L. (1993). Dilation for sets of probabilities. *The Annals of Statistics, 21*, 1139–54.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Shapley, L. S. (1971). Cores of convex games. *International Journal of Game Theory, 1*, 11–26.

Troffaes, M. C. M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning, 45*(1), 17–29.

Von Neumann, J., & Morgestern, O. (1947). *Theory of games and economic behaviour*. Princeton University Press.

Walley, P. (1981). Coherent lower (and upper) probabilities. Statistics research report.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall.

Wallner, A. (2007). Extreme points of coherent lower probabilities in finite spaces. *International Journal of Approximate Reasoning, 44*, 339–357.

Weber, R. J. (1988). Probabilistic values for games. In A. E. Roth (Ed.), *The Shapley value. Essays in honour of L.S. Shapley* (pp. 101–119). Cambridge University Press.