ON THE SELECTION OF AN OPTIMAL OUTER APPROXIMATION OF A COHERENT LOWER PROBABILITY

ENRIQUE MIRANDA, IGNACIO MONTES, AND PAOLO VICIG

Abstract. Coherent lower probabilities are one of the most general tools within Imprecise Probability Theory, and can be used to model the available information about an unknown or partially known precise probability. In spite of their generality, coherent lower probabilities are sometimes difficult to deal with. For this reason, in previous papers we studied the problem of outer approximating a given coherent lower probability by a more tractable model, such as a 2- or completely monotone lower probability. Unfortunately, such an outer approximation is not unique in general, even if we restrict our attention to those that are undominated by other models from the same family. In this paper, we investigate whether a number of approaches may help in selecting a unique undominated outer approximation. These are based on minimising a distance with respect to the initial model, maximising the specificity, or preserving the same preferences as the original model. We apply them to 2- and completely monotone approximating lower probabilities, and also to the particular cases of possibility measures and p-boxes.

Keywords: Coherent lower probabilities, 2-monotonicity, belief functions, possibility measures, p-boxes, specificity.

1. Introduction

Probability measures are the standard mathematical tools used to model uncertainty in an experiment. However, due to a number of factors such as lack of information, unreliable sources or conflicting or noisy data, there are situations where it is arguably unreasonable to model uncertainty by means of a (precise) probability measure. In such cases, we can turn towards the Theory of Imprecise Probabilities [48], that encompasses different models that may be used as an alternative to probability theory in situations of imprecise or ambiguous information. Among them, we can find credal sets [22], coherent lower previsions [48], belief functions [40], possibility measures [50] or p-boxes [18].

One of the most general models within this theory is that of coherent lower and upper previsions [48], or their restriction to events: coherent lower and upper probabilities. These have been applied in different fields such as decision making [12, 21, 32, 43], finance [38], queuing theory [24], probabilistic graphical models [2, 7, 36], reliability [46] or game theory [27], among many others. However, their generality and flexibility to capture the available information in the experiment are counterbalanced by the difficulties that arise at times when using them in practice. For example, there is no simple procedure for computing the extreme points of the associated credal set and there is no unique coherent extension to gambles. On the other hand, these two issues are solved when the coherent lower probability satisfies the additional property of 2-monotonicity [9, 41], or that of complete monotonicity.

This led us [33, 34] to investigate the problem of replacing a coherent lower probability by a 2 or completely monotone one satisfying two reasonable properties: (i) it should not add information to the model; and (ii) it should be as close as possible to the initial coherent lower probability. This gave rise to the notion of undominated outer approximation, formerly introduced in [6]. In particular, in [33] we studied the properties of 2-monotone outer approximations as well as some approaches to compute them. Also, we considered the outer approximations in terms of particular 2-monotone models such as probability intervals [10] or the distortion models [30, 31] induced by the pari mutuel [29, 39] and linear-vacuous [48] models. In [34], we complemented the study considering completely monotone outer approximations and the particular cases of necessity measures [50] and p-boxes [18].

These outer approximation results can also be viewed as rules and properties for transformations of coherent imprecise probabilities into a less general formalism. As such, they are relevant to the operational problem of automated exchange of information among agents adopting different uncertainty representations, when information in terms of coherent imprecise probabilities has to be transformed into a more particular formalism [5].

One of the issues we encountered in [33, 34] is that, in general, there is no unique undominated outer approximation in terms of 2- or completely monotone lower probabilities; in fact their number could be infinite, and in addition their computation may be quite involved. The problem becomes somewhat simpler for necessity measures and p-boxes, where there are a finite number of undominated outer approximations and we have a procedure for determining them all, and it becomes trivial for probability intervals and distortion models, where the undominated outer approximation is unique and can be easily computed.

Since in general there is no unique undominated outer approximation in terms of 2- or completely monotone lower probabilities or even in terms of necessity measures and p-boxes, in this paper we explore a number of possibilities that may help single out a unique undominated outer approximation, that may be considered as optimal according to some criterion. Our approaches can be classified into two groups: those where we compare the outer approximation with the initial model, in terms of the distance between them [5, 23] or the preference relation they encompass; and those where we analyse some imprecision index of the new model, such as specificity [49]. Both approaches can be solved using common tools of operations research, such as linear or quadratic programming, and tools from graph theory.

The paper is organised as follows: after introducing some preliminary notions in Section 2, formalising the idea of outer approximation and summarising the main properties from [33, 34], in Sections 3, 4 and 5 we introduce and compare a number of different procedures to select an undominated outer approximation in terms of 2- and completely monotone lower probabilities, p-boxes and necessity measures, respectively. We conclude the paper in Section 6 summarising the main contributions of the paper and pointing out some future lines of research. In order to streamline the reading, a technical discussion has been relegated to an Appendix.

2. Preliminaries

Let us introduce the main concepts that we shall use in this paper. Throughout, we consider a finite possibility space $\mathcal{X} = \{x_1, \dots, x_n\}$, and denote by $\mathbb{P}(\mathcal{X})$ the set of all the probability measures defined on the power set $\mathcal{P}(\mathcal{X})$.

2.1. **Imprecise probability models.** Imprecise probability models can be given a variety of interpretations, such as the behavioural [26, 48] or the epistemic [20]. Under the latter, it is assumed that the uncertainty in a given experiment can be modelled by means of a probability measure P_0, but due to a number of reasons (conflicting or missing data, lack of resources, imprecise measurements, etc.) it is only possible to determine a set $\mathcal{M}$ of probability measures that is sure to include P_0.

2.1.1. *Coherent lower probabilities.* In those cases, we may also model the available information by means of a *lower probability*, which is 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 satisfies the normalisation properties $\underline{P}(\emptyset) = 0, \underline{P}(\mathcal{X}) = 1$. For every event $A \subseteq \mathcal{X}$, $\underline{P}(A)$ is understood as a lower bound for the true (but unknown) value of $P_0(A)$. Using this interpretation, any lower probability determines the set of probability measures that are compatible with it:

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq \underline{P}(A) \quad \forall A \subseteq \mathcal{X}\}.$$

We refer to this as the *credal set* associated with $\underline{P}$. It is then said that $\underline{P}$ *avoids sure loss* when $\mathcal{M}(\underline{P})$ is non-empty, and that it is *coherent* when it can be computed as:

$$\underline{P}(A) = \min_{P \in \mathcal{M}(\underline{P})} P(A) \quad \forall A \subseteq \mathcal{X},$$

meaning that the bounds determined by $\underline{P}$ are tight. From now on, all the lower probabilities we shall consider in this paper will be coherent.

It is sometimes of interest to consider the conjugate function of a lower probability, defined as $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \mathcal{X}$, and usually referred to as *upper probability.* The value $\overline{P}(A)$ may be interpreted as an upper bound for the unknown value $P_0(A)$, and for a given probability measure $P \in \mathbb{P}(\mathcal{X})$ it follows that

$$P(A) \geq \underline{P}(A) \quad \forall A \subseteq \mathcal{X} \iff P(A) \leq \overline{P}(A) \quad \forall A \subseteq \mathcal{X}.$$

This means that the probabilistic information of the lower probability and that of its conjugate upper probability are equivalent, and so it suffices to work with one of them. It also means that, for any coherent lower probability $\underline{P}$, its conjugate upper probability $\overline{P}$ satisfies

$$\overline{P}(A) = \max_{P \leq \overline{P}} P(A) \quad \forall A \subseteq \mathcal{X}.$$

2.1.2. *k-monotone lower probabilities.* A coherent lower probability $\underline{P}$ is said to be *k-monotone* when it satisfies

$$\begin{aligned}\underline{P}\big(\cup_{i=1}^{p} A_i\big) &\geq \sum_{i=1}^{p} \underline{P}(A_i) - \sum_{i \neq j} \underline{P}(A_i \cap A_j) + \cdots + (-1)^p \underline{P}\big(\cap_{i=1}^{p} A_i\big) \\ &= \sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \underline{P}\big(\cap_{i \in I} A_i\big)\end{aligned}$$

for every $1 \leq p \leq k$ and every $A_1, \ldots, A_p \subseteq \mathcal{X}$. In a similar manner, a coherent upper probability $\overline{P}$ is *k-alternating* if for every $1 \leq p \leq k$ and $A_1, \ldots, A_p \subseteq \mathcal{X}$:

$$\begin{aligned} \overline{P}\big(\cap_{i=1}^p A_i\big) &\leq \sum_{i=1}^p \overline{P}(A_i) - \sum_{i \neq j} \overline{P}(A_i \cup A_j) + \cdots + (-1)^p \overline{P}\big(\cup_{i=1}^p A_i\big) \\ &= \sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \overline{P}\big(\cup_{i \in I} A_i\big), \end{aligned}$$

meaning that a coherent lower probability $\underline{P}$ is k-monotone if and only if its conjugate $\overline{P}$ is k-alternating.

There are two particular cases of k-monotonicity of special interest. The first is 2-*monotonicity*, which corresponds to those coherent lower probabilities satisfying the inequality $\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B)$ for every $A, B \subseteq \mathcal{X}$; and the second is *complete monotonicity*, that refers to those coherent lower probabilities that are k-monotone for every k. Note also that any k-monotone lower probability is also k'-monotone for every $k' \leq k$.

Any lower probability can be represented in terms of a function called *Möbius inverse* of $\underline{P}$, which is denoted by $m_{\underline{P}} : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$, and defined by:

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

Reciprocally, given $m_{\underline{P}}$, we can retrieve the initial lower probability using the following expression:

$$\underline{P}(A) = \sum_{B \subseteq A} m_{\underline{P}}(B).$$

Moreover, $m_{\underline{P}}$ is the Möbius inverse associated with a:

- 2-monotone lower probability $\underline{P}$ if and only if [8] $m_{\underline{P}}$ satisfies:

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

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

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

- Completely monotone lower probability $\underline{P}$ if and only if [40] $m_{\underline{P}}$ satisfies:

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

$$m_{\underline{P}}(A) \geq 0 \quad \forall A \subseteq \mathcal{X}. \tag{Cmonot.2}$$

Note that conditions (2monot.1) and (Cmonot.1) correspond to the assessments $\underline{P}(\mathcal{X}) = 1, \underline{P}(\emptyset) = 0$, that are satisfied by any lower probability by definition.

Completely monotone lower probabilities are also connected with Dempster-Shafer Theory of Evidence [40], where they receive the name *belief functions*. In that case, the Möbius inverse is usually called *basic probability assignment*, and the events with strictly positive mass are called *focal events*.

2.1.3. *P-boxes and possibility measures.* P-boxes and necessity measures are two particular cases of completely monotone lower probabilities.

For any random variable X with values in $\mathcal{X} \subset \mathbb{R}$, its lower and upper distribution functions $\underline{F}, \overline{F} : \mathbb{R} \to [0,1]$ are defined by $\underline{F}(x) = \underline{P}(\{X \leq x\})$ and $\overline{F}(x) = \overline{P}(\{X \leq x\})$. When $\mathcal{X}$ is finite, $\underline{F}, \overline{F}$ are piece-wise constant, and may only be discontinuous at the points $x_1, \ldots, x_n$ of $\mathcal{X}$. If we assume that $x_1 < \ldots < x_n$, then we can shorten the notation and let $\underline{F}, \overline{F}$ be defined in $\mathcal{X}$, so that $\underline{F}(x_i) = \underline{P}(\{X \leq x_i\})$, $\overline{F}(x_i) = \overline{P}(\{X \leq x_i\})$. Then $(\underline{F}, \overline{F})$ constitutes a *p-box* in the sense of [18]. It holds moreover that $\underline{F}(x_n) = \overline{F}(x_n) = 1$, $\underline{F}(x_i) \leq \underline{F}(x_{i+1})$ and $\overline{F}(x_i) \leq \overline{F}(x_{i+1})$ for every $i = 1, \ldots, n-1$. Similarly to our comments about the epistemic interpretation of a lower probability, a p-box may be used to model the imprecise information about a cumulative distribution function F_{P_0}. A p-box defines a credal set by:

$$\mathcal{M}(\underline{F}, \overline{F}) = \{P \in \mathbb{P}(\mathcal{X}) \mid \underline{F}(x) \leq F_P(x) \leq \overline{F}(x) \quad \forall x \in \mathcal{X}\}. \tag{2}$$

Associated with this credal set, we can define a lower (and upper) probability as its lower (and upper) envelope:

$$\underline{P}_{(\underline{F}, \overline{F})}(A) = \inf_{P \in \mathcal{M}(\underline{F}, \overline{F})} P(A) = \inf\{P(A) \mid \underline{F}(x) \leq F_P(x) \leq \overline{F}(x) \quad \forall x \in \mathcal{X}\}. \tag{3}$$

This lower probability is not only coherent, but also completely monotone [44, Sec. 5.1]. The procedure for computing its focal events was determined in [13, Sec. 3.3].

Conversely, given the lower probability $\underline{P}_{(\underline{F}, \overline{F})}$, we can retrieve the p-box because:

$$\underline{F}(x_i) = \underline{P}_{(\underline{F}, \overline{F})}(\{x_1, \ldots, x_i\}), \ \overline{F}(x_i) = 1 - \underline{P}_{(\underline{F}, \overline{F})}(\{x_{i+1}, \ldots, x_n\}) \ \forall i = 1, \ldots, n-1.$$

This means that the probabilistic information gathered by $(\underline{F}, \overline{F})$ and $\underline{P}_{(\underline{F}, \overline{F})}$ is the same. Hence we will use the term *p-box* interchangeably to speak about a p-box $(\underline{F}, \overline{F})$ or its associated lower probability $\underline{P}_{(\underline{F}, \overline{F})}$.

When the possibility space $\mathcal{X}$ is not endowed with a total order, we can consider the notion of *generalised p-box*. A generalised p-box $(\underline{F}, \overline{F})$ is a pair of comonotone[1] mappings such that there exists $x \in \mathcal{X}$ with $\underline{F}(x) = \overline{F}(x) = 1$ and $\underline{F} \leq \overline{F}$.

From [13], a generalised p-box $(\underline{F}, \overline{F})$ defines an order $\leq_{(\underline{F}, \overline{F})}$ and a permutation σ of $\{1, \ldots, n\}$ such that:

$$\underline{F}(x_{\sigma(1)}) \leq_{(\underline{F}, \overline{F})} \cdots \leq_{(\underline{F}, \overline{F})} \underline{F}(x_{\sigma(n)}) = 1.$$
$$\overline{F}(x_{\sigma(1)}) \leq_{(\underline{F}, \overline{F})} \cdots \leq_{(\underline{F}, \overline{F})} \overline{F}(x_{\sigma(n)}) = 1.$$

Clearly, any generalised p-box defines a credal set using Equation (2) and a coherent lower probability using Equation (3). In those cases, we only need to consider the total order $\leq_{(\underline{F}, \overline{F})}$ in the possibility space induced by the generalised p-box.

On the other hand, a ***possibility measure*** [16, 50], usually denoted by Π, is a supremum-preserving function:

$$\Pi(\cup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i), \quad \forall A_i \subseteq \mathcal{X}, \ i \in I.$$

In our finite framework, the above condition can be equivalently expressed as

$$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\} \quad \forall A, B \subseteq \mathcal{X}, \tag{4}$$

[1]Two functions f, g are comonotone if for every $x, x' \in \mathcal{X}$, $f(x) < f(x')$ implies $g(x) \leq g(x')$.

or also as $\Pi(A) = \max_{x \in A} \Pi(\{x\})$ for every $A \subseteq \mathcal{X}$. It is thus clear that the values of Π can be determined using the values in the singletons; the restriction of Π to these, usually denoted by $\pi : \mathcal{X} \to [0, 1]$ and defined by $\pi(x) = \Pi(\{x\})$ for every $x \in \mathcal{X}$, is referred to as the *possibility distribution* of Π.

Every possibility measure is a coherent upper probability, and its conjugate lower probability, usually denoted by N, is called *necessity measure*. A necessity measure is a completely monotone lower probability whose focal events are nested. Moreover, any necessity measure can be obtained as the lower probability of a generalised p-box [45].

2.2. **Outer approximations of coherent lower probabilities.** Even if coherent lower probabilities are more general than 2-monotone lower probabilities, the latter have some practical advantages. Among them, we recall for instance the simplicity of the computation of the extreme points of the associated credal set [41], the fact that they can be extended to gambles using the Choquet integral [9] and that their Shapley value belongs to the credal set and can be obtained easily using the extreme points of $\mathcal{M}(\underline{P})$ [27, 41, 42]. Motivated by this, in [33] we proposed to replace a given coherent lower probability by a 2-monotone one satisfying two properties: (i) that it does not add information to the model; and (ii) that it is as close as possible to the initial coherent lower probability. The first of these conditions gives rise to the notion of *outer approximation*, and the second leads to the notion of *undominated* outer approximation. These two concepts were first formalised by Bronevich and Augustin:

Definition 1. [6] *Given a coherent lower probability $\underline{P}$ and a family $\mathcal{C}$ of coherent lower probabilities, $\underline{Q} \in \mathcal{C}$ is an* outer approximation *of $\underline{P}$ if $\underline{Q}(A) \leq \underline{P}(A)$ for every $A \subseteq \mathcal{X}$. Moreover, $\underline{Q}$ is* undominated *in $\mathcal{C}$ if there is no $\underline{Q}' \in \mathcal{C}$ such that $\underline{Q} \lneq \underline{Q}' \leq \underline{P}$.*

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

Similarly, if we consider a coherent upper probability $\overline{P}$ and a set $\mathcal{C}$ of coherent upper probabilities, we say that $\overline{Q} \in \mathcal{C}$ is an *outer approximation* of $\overline{P}$ if $\overline{Q}(A) \geq \overline{P}(A)$ for every $A \subseteq \mathcal{X}$. Moreover, $\overline{Q}$ is *non-dominating* in $\mathcal{C}$ if there is no $\overline{Q}' \in \mathcal{C}$ such that $\overline{Q} \gneq \overline{Q}' \geq \overline{P}$. It follows that $\overline{Q}$ is an outer approximation of $\overline{P}$ if and only if its conjugate $\underline{Q}$ is an outer approximation of the coherent lower probability $\underline{P}$ that is conjugate of $\overline{P}$, and also $\overline{Q}$ is non-dominating if and only if its conjugate $\underline{Q}$ is undominated.

Throughout the paper, and for the sake of simplicity, we denote by $\mathcal{C}_2$, $\mathcal{C}_\infty$, $\mathcal{C}_\Pi$ and $\mathcal{C}_{(\underline{F},\overline{F})}$ the families of 2-monotone lower probabilities, completely monotone lower probabilities, possibility measures[2] and generalised p-boxes.

In our previous papers [33, 34], we investigated several properties of the undominated (non-dominating for $\mathcal{C}_\Pi$) outer approximations in these families. In particular, we showed that it is not immediate to determine the set of all the undominated outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$, and that these sets are infinite in

[2]Since possibility measures are particular cases of coherent upper probabilities, we shall say that Π is an outer approximation of a coherent upper probability $\overline{P}$ when its conjugate necessity measure N is an outer approximation of the conjugate lower probability $\underline{P}$.

general. The problem is somewhat simpler when the outer approximations are assumed to belong to either $\mathcal{C}_\Pi$ or $\mathcal{C}_{(\underline{F},\overline{F})}$. However, even in these cases there is no unique non-dominating outer approximation (in $\mathcal{C}_\Pi$) or undominated (in $\mathcal{C}_{(\underline{F},\overline{F})}$), and the problem of choosing among them arises. In this paper, we discuss different procedures to select an outer approximation. We first consider in Section 3 the problem of selecting an undominated outer approximation in $\mathcal{C}_2$ and $\mathcal{C}_\infty$, and later in Sections 4 and 5 we study the same problem in the classes $\mathcal{C}_{(\underline{F},\overline{F})}$ and $\mathcal{C}_\Pi$, respectively.

3. Selection of an outer approximation in $\mathcal{C}_2$ and $\mathcal{C}_\infty$

In this section, we investigate several approaches to select an undominated outer approximation in $\mathcal{C}_2$ and $\mathcal{C}_\infty$. As we showed in [33, Ex. 1] and [34, Ex. 1], the number of undominated outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$ is not finite in general. In [33, 34] we focused on those undominated outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$ that minimise the distance proposed in [5] with respect to the original coherent lower probability $\underline{P}$, given by:

$$d_{BV}(\underline{P},\underline{Q}) = \sum_{A\subseteq\mathcal{X}} |\underline{P}(A) - \underline{Q}(A)|. \tag{5}$$

This distance measures the amount of imprecision added to the model when replacing the initial $\underline{P}$ by the outer approximation $\underline{Q}$. Hence, it seems reasonable to select those outer approximations that minimise the imprecision added to the model. To see that this allows to remove some undominated outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$, we refer to [33, Ex. 3] and [34, Ex. 2].

The distance given in Equation (5) can also be regarded as a measure of the imprecision inherent to an outer approximation. Indeed, if we consider a coherent lower probability $\underline{P}$ and an outer approximation $\underline{Q} \leq \underline{P}$ with conjugate $\overline{Q}$, minimising the imprecision of $(\underline{Q},\overline{Q})$

$$\sum_{A\subseteq\mathcal{X}} \big(\overline{Q}(A) - \underline{Q}(A)\big) \tag{6}$$

is equivalent to minimising

$$\begin{aligned}\sum_{A\subseteq\mathcal{X}} \big(\overline{Q}(A) - \overline{P}(A) + \underline{P}(A) - \underline{Q}(A)\big) &= \sum_{A\subseteq\mathcal{X}} (\overline{Q}(A) - \overline{P}(A)) + \sum_{A\subseteq\mathcal{X}} (\underline{P}(A) - \underline{Q}(A)) \\ = \sum_{A\subseteq\mathcal{X}} (\underline{P}(A) - \underline{Q}(A)) + \sum_{A\subseteq\mathcal{X}} (\underline{P}(A) - \underline{Q}(A)) &= 2d_{BV}(\underline{P},\underline{Q}),\end{aligned} \tag{7}$$

applying conjugacy and taking into account that $\sum_{A\subseteq\mathcal{X}}(-\overline{P}(A) + \underline{P}(A))$ acts as a constant. Hence, if we consider a set of outer approximations, the closest one to the original model in the sense of Equation (5) will also be the one that minimises the imprecision of the approximation, when the latter is measured by means of Equation (6).

Let $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$ denote the set of undominated outer approximations of $\underline{P}$ in $\mathcal{C}_2$ and $\mathcal{C}_\infty$, respectively, that minimise the BV-distance with respect to $\underline{P}$. One extra advantage of restricting our selection to the sets $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$ is that they can be more easily determined than the general set of undominated outer approximations. To see this, note that d_{BV} can be equivalently expressed by:

$$d_{BV}(\underline{P}, \underline{Q}) = \sum_{A \subseteq \mathcal{X}} \left(\underline{P}(A) - \sum_{B \subseteq A} m_{\underline{Q}}(B) \right),$$

where $m_{\underline{Q}}$ denotes the Möbius inverse associated with $\underline{Q}$ by means of Equation (1).

Using this alternative expression, we can set up the linear programming problem of minimising $d_{BV}(\underline{P}, \underline{Q})$ subject to (2monot.1)÷(2monot.3), and also to:

$$\sum_{B \subseteq A} m_{\underline{Q}}(B) \leq \underline{P}(A) \quad \forall A \neq \emptyset, \mathcal{X}. \tag{OA}$$

As we have already argued in Section 2.1.2, from [8] we know that conditions (2monot.1)÷(2monot.3) characterise the 2-monotonicity of $\underline{Q}$, while condition (OA) assures that $\underline{Q}$ is an outer approximation of $\underline{P}$. It directly follows (see [33, Prop. 1]) that the set of optimal solutions of this linear programming problem coincides with $\mathcal{C}_2^{BV}(\underline{P})$.

In a similar manner, the set $\mathcal{C}_\infty^{BV}(\underline{P})$ can also be obtained solving a linear programming problem that minimises $d_{BV}(\underline{P}, \underline{Q})$ subject to (Cmonot.1)÷(Cmonot.2) and also to (OA). Again, the set of optimal solutions of this minimisation problem coincides with the set $\mathcal{C}_\infty^{BV}(\underline{P})$ [34, Prop. 3].

From the fact that the sets $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$ can be obtained solving a linear programming problem we deduce that: (i) both sets are non-empty and convex; (ii) the optimal solutions are infinitely many in general. In the rest of the section we discuss different approaches to select an undominated outer approximation within $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$.

3.1. **Approach based on the quadratic distance.** One possibility for obtaining a unique solution to our problem could be to use the quadratic distance. As we discussed in [33], this leads to consider the outer approximation minimising

$$d_q(\underline{P}, \underline{Q}) = \sum_{A \subseteq \mathcal{X}} \left(\underline{P}(A) - \underline{Q}(A) \right)^2. \tag{8}$$

If we set up the quadratic program based on minimising the quadratic distance in Equation (8) subject to conditions (2monot.1)÷(2monot.3) and (OA), it returns a unique undominated outer approximation in $\mathcal{C}_2$ (see [33, Sec. 5.1]). However, in spite of this advantage, the interpretation of this distance is in our opinion less intuitive than that of the BV-distance in Equation (5).

Our proposal is then to put together both approaches: within the sets $\mathcal{C}_2^{BV}(\underline{P})$ or $\mathcal{C}_\infty^{BV}(\underline{P})$ we choose the outer approximation that minimises the quadratic distance. This can be formalised as follows. Consider the following notation:

$$\delta_2^{BV} = \min_{\underline{Q} \in \mathcal{C}_2, \underline{Q} \leq \underline{P}} d_{BV}(\underline{P}, \underline{Q}), \qquad \delta_\infty^{BV} = \min_{\underline{Q} \in \mathcal{C}_\infty, \underline{Q} \leq \underline{P}} d_{BV}(\underline{P}, \underline{Q}).$$

Then, we set up the quadratic problem of minimising the quadratic distance in Equation (8) subject to (2monot.1)÷(2monot.3), (OA) and:

$$\sum_{A \subseteq \mathcal{X}} \left(\underline{P}(A) - \sum_{B \subseteq A} m_{\underline{Q}}(B) \right) = \delta_2^{BV}. \tag{2monot-δ}$$

Analogously, we can minimise the quadratic distance in Equation (8) subject to (Cmonot.1)÷(Cmonot.2), (OA) and:

$$\sum_{A \subseteq \mathcal{X}} \left(\underline{P}(A) - \sum_{B \subseteq A} m_{\underline{Q}}(B) \right) = \delta_{\infty}^{BV}. \qquad \text{(Cmonot-}\delta\text{)}$$

Proposition 1. *Let $\underline{P}$ be a coherent lower probability, and consider the quadratic program of minimising Equation* (8)*. The following properties hold:*

(1) The minimisation problem subject to (2monot.1)÷(2monot.3)*,* (OA) *and* (2monot-δ) *has a unique solution, which is an undominated outer approximation of $\underline{P}$ in $\mathcal{C}_2$.*
(2) The minimisation problem subject to (Cmonot.1)÷(Cmonot.2)*,* (OA) *and* (Cmonot-δ) *has a unique solution, which is an undominated outer approximation of $\underline{P}$ in $\mathcal{C}_\infty$.*

Proof. Let us consider the first case. Conditions (2monot.1)÷(2monot.3) assure that $\underline{Q}$ is a 2-monotone lower probability, while condition (OA) assures that $\underline{Q}$ is an outer approximation of $\underline{P}$ in $\mathcal{C}_2$. Finally, condition (2monot-δ) assures that $\underline{Q}$ minimises the BV-distance $d_{BV}(\underline{P}, \underline{Q})$. Hence, the feasible region of the minimisation problem coincides with $\mathcal{C}_2^{BV}(\underline{P})$. From [33, Prop. 1], this set is non-empty and convex. Since the associated matrix is semidefinite and positive [33, Sec. 5.1], there is an optimal solution to the quadratic problem, which is unique.

The same reasoning, using [34, Prop. 3] instead of [33, Prop. 1], proves the second item. □

The following example illustrates this result.

Example 1. *Consider the coherent lower probability given on $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$* [33, Ex. 1] *by:*

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

For this coherent lower probability, $\delta_2^{BV} = \delta_\infty^{BV} = 1$, and the outer approximations in $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$ coincide and are given by:

$$\mathcal{C}_2^{BV}(\underline{P}) = \mathcal{C}_\infty^{BV}(\underline{P}) = \left\{ \underline{Q}_\alpha : \alpha \in [0, 0.5] \right\},$$

where:

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

For a given $\underline{Q}_\alpha$, its Möbius inverse is:

$$m_{\underline{Q}_\alpha}(\{x_1, x_4\}) = m_{\underline{Q}_\alpha}(\{x_2, x_3\}) = \alpha, \quad m_{\underline{Q}_\alpha}(\{x_1, x_3\}) = m_{\underline{Q}_\alpha}(\{x_2, x_4\}) = 0.5 - \alpha,$$

and zero elsewhere. Therefore, if among these $\underline{Q}_\alpha$ we minimise the quadratic distance with respect to $\underline{P}$, we obtain that the optimal solution is $\underline{Q}_{0.25}$, both in $\mathcal{C}_2$ and $\mathcal{C}_\infty$. ♦

The use of the quadratic distance has the advantage that usually algorithms solving a linear programming problem only return one of the possible solutions, so it may not be immediate to determine if this solution is unique or not. By considering its distance with the original model and solving the associated quadratic problem, we end up with a unique outer approximation, and if it differs from the previous one, we are able to tell that the linear programming problem has an infinite number of solutions.

It is also interesting to stress that the solution we are obtaining here is *not* the outer approximation that minimises the quadratic distance, but the one that minimises this distance between the solutions to the linear programming problem; for a counterexample, we refer to [33, Ex. 3].

From our point of view, this is the preferable approach to select an undominated outer approximation in $\mathcal{C}_2$ and $\mathcal{C}_\infty$. In the rest of the section we explore other approaches to the problem.

3.2. **Approach based on the total variation distance.** Instead of selecting the outer approximation that minimises the quadratic distance, we may consider any other distance between lower probabilities. One interesting possibility is to use extensions of the *total variation distance* [23, Ch. 4.1]: given two probability measures $P_1, P_2 \in \mathbb{P}(\mathcal{X})$, their total variation distance is given by

$$d_{TV}(P_1, P_2) = \max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)| = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_1(\{x\}) - P_2(\{x\})|.$$

This distance may be extended in several non-equivalent ways to imprecise probabilities. Here we consider the extensions proposed in [33]:

$$d_1(\underline{P}_1, \underline{P}_2) = \max_{A \subseteq \mathcal{X}} |\underline{P}_1(A) - \underline{P}_2(A)|, \tag{9}$$

$$d_2(\underline{P}_1, \underline{P}_2) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\underline{P}_1(\{x\}) - \underline{P}_2(\{x\})|, \tag{10}$$

$$d_3(\underline{P}_1, \underline{P}_2) = \sup_{P_1 \geq \underline{P}_1, P_2 \geq \underline{P}_2} \left(\max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)| \right). \tag{11}$$

The second one is somewhat related to the distance of Baroni and Vicig given by Equation (5), but aggregating only the imprecision added on the singletons. On the other hand, the last one is the most compatible with the epistemic interpretation of lower probabilities mentioned at the beginning of the paper, as it considers the maximum distance between the probability measures that are compatible with $\underline{P}_1$ and $\underline{P}_2$, respectively.

In [33], we established that $d_3(\underline{P}_1, \underline{P}_2) \geq d_1(\underline{P}_1, \underline{P}_2)$ for every pair of coherent lower probabilities $\underline{P}_1$ and $\underline{P}_2$ [33, Prop. 9] and that no other dominance relationship between d_1, d_2 and d_3 holds in general [33, Ex. 4].

An alternative to the procedure in the previous section may be to consider the outer approximations in $\mathcal{C}_2^{BV}(\underline{P})$ or $\mathcal{C}_\infty^{BV}(\underline{P})$ that minimise one of $d_i(\underline{P}, \underline{Q})$, for $i = 1, 2, 3$.[3]

[3]Note that if we consider instead the outer approximations in $\mathcal{C}_2$ or $\mathcal{C}_\infty$ that minimise one of the distances d_1, d_2 or d_3, we may end up with outer approximations that are dominated, as shown in [33]. This is why in this section we apply these distances to the outer approximations minimising the BV-distance, that are necessarily undominated.

Since by [33, Prop. 2] any undominated outer approximation $\underline{Q}$ in $\mathcal{C}_2$ satisfies $\underline{Q}(\{x\}) = \underline{P}(\{x\})$ for every $x \in \mathcal{X}$, we always have $d_2(\underline{P}, \underline{Q}) = 0$ and therefore d_2 is not useful in this respect. Let us see in the following example that none of these extensions of the total variation allows to select a single undominated outer approximation in $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$.

Example 2. *Consider now the coherent lower probability $\underline{P}$ that is the lower envelope of the following probability mass functions:*

$$(0, 0.3, 0.3, 0.4), \quad (0.3, 0, 0.3, 0.4), \quad (0.3, 0.3, 0.4, 0), \quad (0.4, 0.2, 0.2, 0.2)$$
$$(0.3, 0.3, 0.1, 0.3), \quad (0.1, 0.4, 0.35, 0.15), \quad (1/6, 1/6, 1/6, 0.5).$$

It is given by:

A	$\underline{P}(A)$	$\underline{Q}_0(A)$	$\underline{Q}_1(A)$
$\{x_1\}$	0	0	0
$\{x_2\}$	0	0	0
$\{x_3\}$	0.1	0.1	0.1
$\{x_4\}$	0	0	0
$\{x_1, x_2\}$	0.3	0.2	0.2
$\{x_1, x_3\}$	0.3	0.3	0.3
$\{x_1, x_4\}$	0.25	0.25	0.25

A	$\underline{P}(A)$	$\underline{Q}_0(A)$	$\underline{Q}_1(A)$
$\{x_2, x_3\}$	0.3	0.3	0.3
$\{x_2, x_4\}$	0.3	0.25	0.3
$\{x_3, x_4\}$	0.4	0.35	0.3
$\{x_1, x_2, x_3\}$	0.5	0.5	0.5
$\{x_1, x_2, x_4\}$	0.6	0.6	0.6
$\{x_1, x_3, x_4\}$	0.6	0.6	0.6
$\{x_2, x_3, x_4\}$	0.6	0.6	0.6
$\mathcal{X}$	1	1	1

The undominated outer approximations of $\underline{P}$ in $\mathcal{C}_2$ minimising the BV-distance are $\underline{Q}_0$, $\underline{Q}_1$ and their convex combinations: $\underline{Q}_\alpha = \alpha \underline{Q}_0 + (1-\alpha)\underline{Q}_1$ for every $\alpha \in (0,1)$.

We obtain that:

$$d_1\big(\underline{P}, \underline{Q}_0\big) = d_1\big(\underline{P}, \underline{Q}_1\big) = d_1\big(\underline{P}, \underline{Q}_\alpha\big) = 0.1$$

and

$$d_3(\underline{P}, \underline{Q}_0) = d_3(\underline{P}, \underline{Q}_1) = d_3(\underline{P}, \underline{Q}_\alpha) = 0.5 = \max_{A \subseteq \mathcal{X}} |\overline{P}(A) - \underline{Q}_\alpha(A)| \quad \forall \alpha \in [0,1].$$

Thus, neither d_1 nor d_3 determines a unique undominated outer approximation among those in $\mathcal{C}_2^{BV}(\underline{P})$.

With respect to the outer approximations in $\mathcal{C}_\infty$, it can be seen that $\delta_\infty^{BV} = 0.85$, and two belief functions Bel_1 and Bel_2 attaining this value are defined using the following Möbius inverses:

A	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_1, x_4\}$	$\{x_2, x_3\}$	$\{x_2, x_4\}$	$\{x_3, x_4\}$
m_{B_1}	0.1	0.2	0.1	$0.25 - \frac{0.35}{3}$	0.1	$0.3 - \frac{0.35}{3}$	$0.3 - \frac{0.35}{3}$
m_{B_2}	0.1	0.19	0.1	$0.25 - \frac{0.35}{3}$	0.11	$0.3 - \frac{0.35}{3}$	$0.3 - \frac{0.35}{3}$

It can be checked that both $Bel_1, Bel_2 \in \mathcal{C}_\infty^{BV}(\underline{P})$ and that

$$d_1(Bel_1, \underline{P}) = d_1(Bel_2, \underline{P}) = \frac{0.35}{3} = \min_{Bel \in \mathcal{C}_\infty^{BV}(\underline{P})} d_1(Bel, \underline{P}).$$
$$d_2(Bel_1, \underline{P}) = d_2(Bel_2, \underline{P}) = 0 = \min_{Bel \in \mathcal{C}_\infty^{BV}(\underline{P})} d_2(Bel, \underline{P}).$$
$$d_3(Bel_1, \underline{P}) = d_3(Bel_2, \underline{P}) = 0.5 = \min_{Bel \in \mathcal{C}_\infty^{BV}(\underline{P})} d_3(Bel, \underline{P}).$$

Therefore, none of d_1, d_2, d_3 determines a unique approximation in $\mathcal{C}_\infty^{BV}$ either. ♦

While in the previous examples there is not a unique undominated outer approximation minimizing the extensions of the total variation distance, this is not always the case, as we show in our next example. It also tells us that d_1, d_2, d_3 may produce different optimal outer approximations:

Example 3. *Consider $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and the lower probability $\underline{P}$ that is the lower envelope of*

$$(0, 0.2, 0.6, 0.2), \quad (0, 0.6, 0.3, 0.1), \quad (0.8, 0, 0.1, 0.1),$$
$$(0.35, 0.65, 0, 0), \quad (0.2, 0.2, 0.2, 0.4).$$

It is given by:

A	$\underline{P}(A)$	A	$\underline{P}(A)$
$\{x_1\}$	0	$\{x_2, x_3\}$	0.1
$\{x_2\}$	0	$\{x_2, x_4\}$	0.1
$\{x_3\}$	0	$\{x_3, x_4\}$	0
$\{x_4\}$	0	$\{x_1, x_2, x_3\}$	0.6
$\{x_1, x_2\}$	0.2	$\{x_1, x_2, x_4\}$	0.4
$\{x_1, x_3\}$	0.3	$\{x_1, x_3, x_4\}$	0.35
$\{x_1, x_4\}$	0.1	$\{x_2, x_3, x_4\}$	0.2
		$\mathcal{X}$	1

Since $\underline{P}(\{x_1, x_3, x_4\}) + \underline{P}(\{x_1\}) = 0.35 < 0.4 = \underline{P}(\{x_1, x_3\}) + \underline{P}(\{x_1, x_4\})$, we deduce that $\underline{P}$ is not 2-monotone, and as a consequence it is not completely monotone either. It can be checked that $\mathcal{C}_2^{BV}(\underline{P}) = \mathcal{C}_\infty^{BV}(\underline{P}) = \{\underline{Q}_\alpha \mid \alpha \in [0, 0.05]\}$, where

$$\underline{Q}_\alpha(A) = \begin{cases} 0.3 - \alpha & \text{if } A = \{x_1, x_3\} \\ 0.05 + \alpha & \text{if } A = \{x_1, x_4\} \\ \underline{P}(A) & \text{otherwise.} \end{cases}$$

From here it follows that

$$d_1(\underline{P}, \underline{Q}_\alpha) = \max\{\alpha, 0.05 - \alpha\},$$

whence the optimal outer approximation if we use d_1 is $\underline{Q}_{0.025}$. On the other hand,

$$d_3(\underline{P}, \underline{Q}_\alpha) = \max\{0.8, 0.85 - \alpha\} = 0.85 - \alpha,$$

whence the optimal outer approximation if we use d_3 is $\underline{Q}_{0.05}$. Thus, in this case both d_1 and d_3 determine a unique outer approximation, but they do not coincide. Note also that d_2 does not rule out any element from $\mathcal{C}_2^{BV}(\underline{P})$ because all of them satisfy $\underline{Q}_\alpha(\{x_i\}) = \underline{P}(\{x_i\})$ for $i = 1, 2, 3, 4$. ♦

3.3. **Approach based on preference preservation.** An interesting procedure to select an undominated outer approximation in $\mathcal{C}_2^{BV}(\underline{P})$ or $\mathcal{C}_\infty^{BV}(\underline{P})$ is to require some kind of preservation of the preferences given by $\underline{P}$. If $\underline{Q}$ denotes an undominated outer approximation in $\mathcal{C}_2^{BV}(\underline{P})$ or $\mathcal{C}_\infty^{BV}(\underline{P})$, we consider the following conditions[4]on preference preservation (on $\mathcal{P}(\mathcal{X})$ the first four, on $\mathcal{X}$ the last two):

C1: $\underline{P}(A) < \underline{P}(B) \Rightarrow \underline{Q}(A) < \underline{Q}(B)$.
C2: $\underline{P}(A) \leq \underline{P}(B) \Rightarrow \underline{Q}(A) \leq \underline{Q}(B)$.

[4]We may consider other possibilities based on notions such as interval dominance, meaning that $\overline{P}(A) < \underline{P}(B) \Rightarrow \overline{Q}(A) < \underline{Q}(B)$. However, we think that this condition will be in general too strong, as can be verified with the example in Remark 1.

C3: $\underline{P}(A) = \underline{P}(B) \Rightarrow \underline{Q}(A) = \underline{Q}(B)$.
C4: $\underline{P}(A) < \underline{P}(B) \Rightarrow \underline{Q}(A) \leq \underline{Q}(B)$.
C5: $\underline{P}(\{x\}) < \underline{P}(\{x'\}) \Rightarrow \underline{Q}(\{x\}) \leq \underline{Q}(\{x'\})$.
C6: $\underline{P}(\{x\}) = \underline{P}(\{x'\}) \Rightarrow \underline{Q}(\{x\}) = \underline{Q}(\{x'\})$.

Figure 1 summarises the relationships between these conditions:

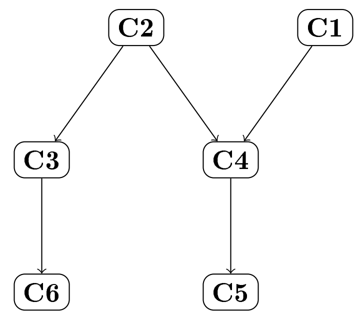


FIGURE 1. Relationship between the conditions.

The idea here is that the lower probability $\underline{P}$ induces a strict preference ($A \prec B \Leftrightarrow \underline{P}(A) < \underline{P}(B)$), a weak preference ($A \preceq B \Leftrightarrow \underline{P}(A) \leq \underline{P}(B)$) and an indifference relation ($A \sim B \Leftrightarrow \underline{P}(A) = \underline{P}(B)$). When comparing $\underline{P}$ and $\underline{Q}$ we may consider a *strong* preservation, in the sense that if an event A is strictly preferred (resp. weakly preferred, indifferent) to B in $\underline{P}$, then it is also strictly preferred (resp. weakly preferred, indifferent) in $\underline{Q}$; this is the idea behind conditions **C1**–**C3**. Or we may consider a *weak* preservation, in the sense that the strict preference between two events in $\underline{P}$ implies the weak preference in $\underline{Q}$ (condition **C4**). Finally, if we focus our attention on the preferences on singletons, this leads to conditions **C5** and **C6**. For earlier works using these or similar conditions, we refer to [5, 17, 47].

We may thus consider the possibility of choosing, among those outer approximations that minimise the BV-distance, the one that satisfies **C**i, for some $i \in \{1, \ldots, 6\}$; in this respect, we may argue that it does not make much sense to consider an outer approximation that satisfies **C3** only, but this condition may be of interest if it is required together with **C4** or **C1**.

However, this criterion is not valid because, as the next example shows, it could happen that either none of them satisfies **C**i or more than one satisfies it.

Example 4. *Consider again the coherent lower probability in Example 1. There, we have seen that* $\mathcal{C}_2^{BV}(\underline{P}) = \mathcal{C}_\infty^{BV}(\underline{P}) = \left\{\underline{Q}_\alpha \mid \alpha \in [0, 0.5]\right\}$. *Let us see which* $\underline{Q}_\alpha$ *satisfy each of the conditions* **C**i*:*

C1: *Neither* Q_0 *nor* $Q_{0.5}$ *satisfies* **C1***, since*

$$\underline{P}(\{x_1, x_2\}) < \underline{P}(\{x_1, x_3\}) = \underline{P}(\{x_1, x_4\})$$

but

$$\underline{Q}_0(\{x_1, x_4\}) = \underline{Q}_0(\{x_1, x_2\}) = \underline{Q}_{0.5}(\{x_1, x_2\}) = \underline{Q}_{0.5}(\{x_1, x_3\}) = 0.$$

However, every $\underline{Q}_\alpha$*, for* $\alpha \in (0, 0.5)$*, does satisfy* **C1***.*

C2,C3: *None of the* $\underline{Q}_\alpha$ *satisfies* **C3***, because:*

$$\underline{P}(\{x_1,x_3\}) = \underline{P}(\{x_1,x_4\}) = \underline{P}(\{x_1,x_2,x_3\}),$$

but there is no $\alpha \in [0,0.5]$ *satisfying*

$$\underline{Q}_\alpha(\{x_1,x_3\}) = \underline{Q}_\alpha(\{x_1,x_4\}) = \underline{Q}_\alpha(\{x_1,x_2,x_3\}).$$

Since **C2** *implies* **C3***, we conclude that no* $\underline{Q}_\alpha$ *satisfies* **C2***.*

C4,C5,C6: *All the outer approximations* $\underline{Q}_\alpha$ *satisfy* **C4, C5** *and* **C6***.*

We conclude from this example that none of the **C***i helps us selecting a unique undominated outer approximation in* $\mathcal{C}_2^{BV}(\underline{P})$ *or* $\mathcal{C}_\infty^{BV}(\underline{P})$*.* ♦

Remark 1. *One alternative to the approach considered in this paper would be to select the preference preservation as the primary criterion for selecting an outer approximation, and, if needed, compare the possible solutions in terms of the BV distance.*

While interesting, we believe that the first criterion to compare the outer approximation to the initial model should be a measure of their distance, since we consider that: (a) if we want to compare the inferences made by the initial and the transformed model, it is useful to have a measure of their distance; (b) the preference preservation conditions are in our view more suited in a qualitative context.

Another issue is that it may be impossible, in general, to require that an outer approximation is both undominated and preference preserving. To see an example, let $\mathcal{X} = \{x_1,x_2,x_3,x_4\}$*, and consider the lower probability* $\underline{P}$ *that is the lower envelope of the following probability measures:*

$$(0,0.05,0.05,0.9), \quad (0.03,0.02,0.02,0.93), \quad (0.95,0,0,0.05),$$
$$(0.06,0,0.94,0), \quad (0.09,0.875,0,0.035), \quad (0.04,0.02,0.94,0).$$

It is given by:

A	$\underline{P}(A)$	$\underline{Q}_\alpha$
$\{x_1\}$	0	0
$\{x_2\}$	0	0
$\{x_3\}$	0	0
$\{x_4\}$	0	0
$\{x_1,x_2\}$	0.05	$0.05-\alpha$
$\{x_1,x_3\}$	0.05	$0.02+\alpha$
$\{x_1,x_4\}$	0.04	$0.01+\alpha$

A	$\underline{P}(A)$	$\underline{Q}_\alpha$
$\{x_2,x_3\}$	0	0
$\{x_2,x_4\}$	0	0
$\{x_3,x_4\}$	0.035	0.035
$\{x_1,x_2,x_3\}$	0.07	0.07
$\{x_1,x_2,x_4\}$	0.06	0.06
$\{x_1,x_3,x_4\}$	0.125	0.125
$\{x_2,x_3,x_4\}$	0.05	0.05
$\mathcal{X}$	1	1

$\underline{P}$ *is not 2-monotone because the condition is not satisfied in the following two instances:*

$$0.07 = \underline{P}(\{x_1,x_2,x_3\}) + \underline{P}(\{x_1\}) \ngeq \underline{P}(\{x_1,x_2\}) + \underline{P}(\{x_1,x_3\}) = 0.10,$$
$$0.06 = \underline{P}(\{x_1,x_2,x_4\}) + \underline{P}(\{x_1\}) \ngeq \underline{P}(\{x_1,x_2\}) + \underline{P}(\{x_1,x_4\}) = 0.09.$$

We know from our results in [33] *that any undominated outer approximation in* $\mathcal{C}_2$ *coincides with* $\underline{P}$ *in the singletons and in the events of cardinality* $n-1=3$*. In fact, it can be seen that the set of undominated outer approximations in* $\mathcal{C}_2$ *is* $\left\{\underline{Q}_\alpha \mid \alpha \in [0,0.03]\right\}$*, and among them, the only one minimising the BV-distance is*

$\underline{Q}_{0.03}$. *It holds that none of the* $\underline{Q}_\alpha$ *satisfies* $\mathbf{C}i$ *for* $i = 1, 2, 4$*:*

$$\underline{P}(\{x_1,x_2\}) \geq \underline{P}(\{x_1,x_4\}) \Rightarrow \underline{Q}(\{x_1,x_2\}) \geq \underline{Q}(\{x_1,x_4\})$$
$$\Rightarrow 0.05-\alpha \geq 0.01+\alpha \Rightarrow \alpha \leq 0.02.$$
$$\underline{P}(\{x_1,x_4\}) \geq \underline{P}(\{x_3,x_4\}) \Rightarrow \underline{Q}(\{x_1,x_4\}) \geq \underline{Q}(\{x_3,x_4\})$$
$$\Rightarrow 0.01+\alpha \geq 0.035 \Rightarrow \alpha \geq 0.025.$$

These two conditions are incompatible, whence no undominated outer approximation in $\mathcal{C}_2$ *satisfies* $\mathbf{C}1$*,* $\mathbf{C}2$ *or* $\mathbf{C}4$*.*
In addition, $\mathbf{C}3$ *is not satisfied by any* $\underline{Q}_\alpha$ *either. The reason is that if* $\underline{Q}$ *satisfies* $\mathbf{C}3$*, then:*

$$0.05 = \underline{P}(\{x_2,x_3,x_4\}) = \underline{Q}(\{x_2,x_3,x_4\}) = \underline{Q}(\{x_1,x_2\}) = \underline{Q}(\{x_1,x_3\}),$$

a contradiction. We therefore conclude that the outer approximations in $\mathcal{C}_2$ *preserving preferences are not undominated. Since all the* $\underline{Q}_\alpha$ *above are belief functions, a similar comment applies to the outer approximations in* $\mathcal{C}_\infty$*.*

Note also that if we keep the condition of being undominated in order for an outer approximation to be considered acceptable, imposing some preference preservation condition is in general computationally heavier than minimising some distance. In fact, this can be done operationally by means of a linear programming problem, where the objective function $\min d_{BV}(\underline{P}, \underline{Q})$ *is replaced by some trivial condition of the type* $\min 0$*, and where, in addition to the constraints that guarantee that* $\underline{Q}$ *is undominated, we have additional constraints derived from the preference relations on the events. As our example above shows, the resulting problem may have no feasible solution even in a low dimension space, while this is not the case with the approach based on minimising the distance to the original model.* ♦

3.4. **Approach based on specificity measures.** In our last approach we consider a popular procedure for comparing two completely monotone lower probabilities: their *specificity*. A specificity measure is used to determine how imprecise a belief function is, in the sense that the greater the specificity, the smaller the imprecision. Among the many different proposals of specificity measure in the literature (see for example [11, 15]), we follow here the suggestion of Moral and de Campos [37] and consider the specificity measure defined by Yager [49]:

Definition 2. *Let* $\underline{Q}$ *be a completely monotone lower probability on* $\mathcal{P}(\mathcal{X})$ *with Möbius inverse* $m_{\underline{Q}}$*. Its* specificity *is given by*

$$S(\underline{Q}) = \sum_{\emptyset \neq A \subseteq \mathcal{X}} \frac{m_{\underline{Q}}(A)}{|A|}. \tag{12}$$

This function splits the mass of any focal event among its elements. Equation (12) can also be computed as follows:

$$S(\underline{Q}) = \sum_{i=1}^{n} \frac{1}{i} \sum_{A:|A|=i} m_{\underline{Q}}(A). \tag{13}$$

Yager established that for any belief function $\underline{Q}$ its specificity $S(\underline{Q})$ belongs to $[\frac{1}{n}, 1]$, and that the specificity measure is monotone: $\underline{Q} \leq \underline{Q}'$ implies $S(\underline{Q}) \leq S(\underline{Q}')$. Although Yager applied Equation (12) to belief functions only, we next show that similar properties hold when applying it on coherent lower probabilities.

Proposition 2. *Let $\underline{Q}$ be a coherent lower probability with Möbius inverse $m_{\underline{Q}}$, and let $S(\underline{Q})$ be given by Equation* (12)*. Then:*

(1) S is monotone: $\underline{Q} \leq \underline{Q}'$ implies $S(\underline{Q}) \leq S(\underline{Q}')$.
(2) $S(\underline{Q}) \in [\frac{1}{n}, 1]$.

Proof. First of all, note that $S(\underline{Q})$ can be expressed in terms of $\underline{Q}$ instead of $m_{\underline{Q}}$:

$$S(\underline{Q}) = \sum_{\emptyset \neq A \subseteq \mathcal{X}} \frac{m_{\underline{Q}}(A)}{|A|} = \sum_{\emptyset \neq A \subseteq \mathcal{X}} \sum_{B \subseteq A} (-1)^{|A \setminus B|} \frac{\underline{Q}(B)}{|A|}$$
$$= \sum_{B \subseteq \mathcal{X}} \underline{Q}(B) \sum_{\emptyset \neq A \supseteq B} (-1)^{|A \setminus B|} \frac{1}{|A|} = \sum_{B \subseteq \mathcal{X}} \underline{Q}(B) \sum_{k=|B|}^{n} \frac{(-1)^{k-|B|}}{k} \binom{n-|B|}{k-|B|}.$$

For every $j = 1, \ldots, n$, let us denote

$$f(j) = \sum_{k=j}^{n} \frac{(-1)^{k-j}}{k} \binom{n-j}{k-j}.$$

$f(j)$ can be rewritten as:

$$f(j) = \sum_{l=0}^{n-j} \frac{(-1)^l}{l+j} \binom{n-j}{l} = \frac{1}{j\binom{n-j+j}{n-j}} = \frac{1}{j\binom{n}{n-j}} \geq 0,$$

where the second equality follows from Melzak's formula (see, for instance, [25]). Let us proceed to establish the two statements.

(1) Given $\underline{Q} \leq \underline{Q}'$, the non-negativity of $f(j)$ implies that:

$$S(\underline{Q}) = \sum_{B \subseteq \mathcal{X}} \underline{Q}(B) f(|B|) \leq \sum_{B \subseteq \mathcal{X}} \underline{Q}'(B) f(|B|) = S(\underline{Q}'),$$

hence S is monotone.

(2) If $\underline{Q}$ is a precise probability measure, $m_{\underline{Q}}(A) > 0$ only if $|A| = 1$, whence

$$S(\underline{Q}) = \sum_{|A|=1} m_{\underline{Q}}(A) = 1.$$

Since any coherent lower probability is dominated by a precise probability measure, we deduce from the first item that $S(\underline{Q}) \leq 1$. Consider now the lower probability $\underline{Q}_v$ given by $\underline{Q}_v(\mathcal{X}) = 1$ and $\underline{Q}(A) = 0$ for any $A \neq \mathcal{X}$. It is a belief function whose only focal event $\mathcal{X}$ has mass 1, and it is dominated by any coherent lower probability on $\mathcal{P}(\mathcal{X})$. It satisfies:

$$S(\underline{Q}_v) = \frac{m_{\underline{Q}_v}(\mathcal{X})}{|\mathcal{X}|} = \frac{1}{n}.$$

Applying the first statement, we conclude that $S(\underline{Q}) \in [\frac{1}{n}, 1]$ for any coherent lower probability $\underline{Q}$. □

Therefore, we can choose an undominated outer approximation in $\mathcal{C}_2^{BV}(\underline{P})$ or $\mathcal{C}_\infty^{BV}(\underline{P})$ with the greatest specificity. Our next example shows that, as was the case with preference preservation, this criterion does not give rise to a unique undominated outer approximation.

Example 5. *Consider again Example 1. We have seen that the undominated outer approximations in* $\mathcal{C}_2$ *and* $\mathcal{C}_\infty$ *are* $\left\{\underline{Q}_\alpha \mid \alpha \in [0, 0.5]\right\}$ *in both cases, and that the Möbius inverse of each* $\underline{Q}_\alpha$ *is given by:*

$$m_{\underline{Q}_\alpha}(\{x_1, x_4\}) = m_{\underline{Q}_\alpha}(\{x_2, x_3\}) = \alpha, \quad m_{\underline{Q}_\alpha}(\{x_1, x_3\}) = m_{\underline{Q}_\alpha}(\{x_2, x_4\}) = 0.5 - \alpha,$$

and zero elsewhere. Hence, the specificity of $\underline{Q}_\alpha$ *is given by:*

$$S(\underline{Q}_\alpha) = \frac{1}{2}\left(\alpha + \alpha + 0.5 - \alpha + 0.5 - \alpha\right) = 0.5,$$

regardless of the value of $\alpha \in [0, 0.5]$*. We conclude that all the undominated outer approximations in* $\mathcal{C}_2$ *and* $\mathcal{C}_\infty$ *minimising the BV-distance have the same specificity. Thus, this criterion is not helpful in the selection process.* ♦

While in this section we have considered the specificity measure given by Equation (12), it is not the only possibility. We may for instance consider the notion of *non-specificity* proposed by Dubois and Prade in [15], given by

$$\sum_{\emptyset \neq A \subseteq \mathcal{X}} m(A) \log(|A|),$$

that was shown in [3] to be also applicable to 2-monotone lower probabilities. It can be verified using Example 5 above that this other definition does not help either to choose one among the undominated outer approximations.

3.5. **Discussion.** We have seen in this section several approaches for selecting one undominated outer approximation in $\mathcal{C}_2$ and $\mathcal{C}_\infty$. We have focused on the undominated outer approximations minimising the BV-distance, i.e., on the sets $\mathcal{C}_2^{BV}(\underline{P})$ and $\mathcal{C}_\infty^{BV}(\underline{P})$. In these two sets, we have proposed to minimise the quadratic distance, a total variation distance, to preserve the preferences encompassed by the initial model and an approach based on maximising the specificity. Among all these approaches, we have seen that the only approach that selects one single undominated outer approximation is the one based on minimising the quadratic distance among the outer approximations minimising the BV-distance. The other approaches, albeit interesting, are not useful in general for the purposes of this paper, since either they produce more than one or no optimal solutions.

4. Selection of an outer approximation in $\mathcal{C}_{(\underline{F},\overline{F})}$

We consider now the case of undominated outer approximations in the set of generalised p-boxes, $\mathcal{C}_{(\underline{F},\overline{F})}$. As we shall see, the number of undominated outer approximations in $\mathcal{C}_{(\underline{F},\overline{F})}$ is finite, and there is a simple procedure for determining them.

In order to see this, let us remark that the lower probability in Equation (3) can be computed using the values of the lower and upper distribution functions $\underline{F}$ and $\overline{F}$ [44, Prop. 4]. To see how this comes about, note that any $A \subseteq \{x_1, \ldots, x_n\}$ can be expressed as a finite union of events of consecutive elements, where these events are as large as possible: for instance if $n = 4$ the event $A = \{x_1, x_2, x_4\}$ would be expressed as $A = \{x_1, x_2\} \cup \{x_4\}$.

Since without loss of generality we can add an element y_0^* to $\mathcal{X}$ to denote the impossible event, with $\overline{F}(y_0^*) = 0$ and $y_0^* < x_1$, we can express $A = (y_0^*, y_1^*] \cup$

$(y_2^*, y_3^*] \cup \dots \cup (y_{2m}^*, y_{2m+1}^*]$ for some $m \geq 0$, with $y_0^* \leq y_1^* < y_2^* < \ldots < y_{2m}^* < y_{2m+1}^* \in \mathcal{X}$. For instance, under this notation, and again when $n = 4$,

$$\{x_2, x_3\} = (x_1, x_3] \text{ and } \{x_1, x_3, x_4\} = (y_0^*, x_1] \cup (x_2, x_4].$$

It then holds that [44, Prop. 4] that:

$$\underline{P}_{(\underline{F},\overline{F})}(A) = \sum_{l=0}^{m} \max\left\{0, \underline{F}(y_{2l+1}^*) - \overline{F}(y_{2l}^*)\right\}. \tag{14}$$

Let us denote by S_n the set of permutations of $\{1, \ldots, n\}$. For each $\sigma \in S_n$, consider the total order $\leq_\sigma$ given by $x_{\sigma(1)} \leq_\sigma x_{\sigma(2)} \leq_\sigma \ldots \leq_\sigma x_{\sigma(n)}$, and define the p-box $(\underline{F}_\sigma, \overline{F}_\sigma)$ by:

$$\underline{F}_\sigma(x_{\sigma(i)}) = \underline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}), \quad \overline{F}_\sigma(x_{\sigma(i)}) = \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) \tag{15}$$

for every $i = 1, \ldots, n$. From [34, Thm. 17], $\mathcal{C}_{(\underline{F},\overline{F})} = ((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$. This means that the number of undominated outer approximations is bounded above by $n!$. Our next result lowers this bound and shows that the number of different undominated outer approximations is at most $\frac{n!}{2}$. For this aim, given a permutation $\sigma \in S_n$, we denote by $\overline{\sigma}$ the permutation given by $\overline{\sigma}(i) = \sigma(n - i + 1)$ for every $i = 1, \ldots, n$.

Proposition 3. *Let $\underline{P}$ be a coherent lower probability with conjugate upper probability $\overline{P}$, and let $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$ be the family of undominated outer approximations in $\mathcal{C}_{(\underline{F},\overline{F})}$. If we denote by $\underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)}$ the coherent lower probability associated with $(\underline{F}_\sigma, \overline{F}_\sigma)$, then $\underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)} = \underline{P}_{(\underline{F}_{\bar{\sigma}},\overline{F}_{\bar{\sigma}})}$.*

Proof. Given $x_{\sigma(i)}$ and applying Equation (14):

$$\begin{aligned}
\underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)}(\{x_{\sigma(i)}\}) &= \max\left\{0, \underline{F}_\sigma(x_{\sigma(i)}) - \overline{F}_\sigma(x_{\sigma(i-1)})\right\} \\
&= \max\left\{0, \underline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\})\right\}.
\end{aligned}$$

On the other hand:

$$\begin{aligned}
\underline{P}_{(\underline{F}_{\bar{\sigma}},\overline{F}_{\bar{\sigma}})}(\{x_{\sigma(i)}\}) &= \max\left\{0, \underline{F}_{\bar{\sigma}}(x_{\sigma(i)}) - \overline{F}_{\bar{\sigma}}(x_{\sigma(i+1)})\right\} \\
&= \max\left\{0, \underline{P}(\{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\}) - \overline{P}(\{x_{\sigma(i+1)}, \ldots, x_{\sigma(n)}\})\right\} \\
&= \max\left\{0, 1 - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}) - 1 + \underline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\})\right\} \\
&= \max\left\{0, \underline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\})\right\} \\
&= \max\left\{0, \underline{F}_\sigma(x_{\sigma(i)}) - \overline{F}_\sigma(x_{\sigma(i-1)})\right\} = \underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)}(\{x_{\sigma(i)}\}).
\end{aligned}$$

Next, if we consider a set of consecutive elements $A = \{x_{\sigma(i)}, \ldots, x_{\sigma(i+l)}\}$, for some $l \geq 1$ and some $i = 1, \ldots, n-1$, we deduce from Equation (14) that

$$\begin{aligned}
\underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)}(A) &= \max\{0, \underline{F}_\sigma(x_{\sigma(i+l)}) - \overline{F}_\sigma(x_{\sigma(i-1)})\} \\
&= \max\{0, \underline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i+l)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\})\} \\
&= \max\{0, 1 - \overline{P}(\{x_{\sigma(i+l+1)}, \ldots, x_{\sigma(n)}\}) - 1 + \underline{P}(\{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\})\} \\
&= \max\{0, \underline{P}(\{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\}) - \overline{P}(\{x_{\sigma(i+l+1)}, \ldots, x_{\sigma(n)}\})\} \\
&= \max\{0, \underline{P}(\{x_{\bar{\sigma}(n-i+1)}, \ldots, x_{\bar{\sigma}(1)}\}) - \overline{P}(\{x_{\bar{\sigma}(n-i-l)}, \ldots, x_{\bar{\sigma}(1)}\})\} \\
&= \max\{0, \underline{F}_{\bar{\sigma}}(x_{\bar{\sigma}(n-i+1)}) - \overline{F}_{\bar{\sigma}}(x_{\bar{\sigma}(n-i-l)})\} \\
&= \underline{P}_{(\underline{F}_{\bar{\sigma}},\overline{F}_{\bar{\sigma}})}(\{x_{\bar{\sigma}(n-i-l+1)}, \ldots, x_{\bar{\sigma}(n-i+1)}\}) = \underline{P}_{(\underline{F}_{\bar{\sigma}},\overline{F}_{\bar{\sigma}})}(A).
\end{aligned}$$

If we now apply Equation (14) we conclude that $\underline{P}_{(\underline{F}_\sigma,\overline{F}_\sigma)} = \underline{P}_{(\underline{F}_{\overline{\sigma}},\overline{F}_{\overline{\sigma}})}$. □

Thus, each permutation and its opposite induce a p-box with the same associated coherent lower probability. As a consequence, there are at most $\frac{n!}{2}$ different undominated outer approximations in $\mathcal{C}_{(\underline{F},\overline{F})}$. In the remainder of this section we explore different approaches for selecting one of them.

4.1. **Approach based on minimising imprecision.** In the case of p-boxes, we can propose a different approach than the ones considered so far, based on minimising the associated imprecision. Since any p-box is an ordered pair of distribution functions that are determined by their values on $\mathcal{X}$, we may measure their inherent imprecision by computing the distance between the lower and the upper distribution functions on $\mathcal{X}$. This produces the following measure of imprecision Imp:

$$Imp(\underline{F}_\sigma,\overline{F}_\sigma) = \sum_{x\in\mathcal{X}} \left(\overline{F}_\sigma(x) - \underline{F}_\sigma(x)\right). \tag{16}$$

Thus, our goal will be to determine the p-box minimising the imprecision in this equation.

Taking into account the definition of the p-box $(\underline{F}_\sigma,\overline{F}_\sigma)$ once the permutation σ has been fixed (Equation (15)), we can also express Equation (16) in terms of the lower and upper probabilities $\underline{P},\overline{P}$. Indeed, given a permutation σ inducing the order $x_{\sigma(1)} \leq_\sigma \ldots \leq_\sigma x_{\sigma(n)}$ and its associated p-box $(\underline{F}_\sigma,\overline{F}_\sigma)$, Equation (16) becomes:

$$\sum_{x\in\mathcal{X}}(\overline{F}_\sigma(x) - \underline{F}_\sigma(x)) = \sum_{i=1}^{n} \left(\overline{P}(\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\}) - \underline{P}(\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\})\right)$$

In other words, we consider a chain of events from a singleton to $\mathcal{X}$ and compute the differences between $\overline{P}$ and $\underline{P}$ for those events[5]. Note however that, because of Equation (15), this chain of events will vary with the p-box considered, because it depends on the order associated with the permutation σ; in other words, the key events taken into account when measuring the imprecision of the p-box $(\underline{F}_\sigma,\overline{F}_\sigma)$ are not always the same.

The above correspondence means that we can find the p-box(es) minimising Equation (16) by solving a shortest path problem. For this aim, consider the Hasse diagram of $\mathcal{P}(\mathcal{X})$, and we assign the following weights: for every $A \neq \mathcal{X}$ and $x_i \notin A$, we assign the weight $\overline{P}(A\cup\{x_i\}) - \underline{P}(A\cup\{x_i\})$ to the edge $A \to A\cup\{x_i\}$. In this way for every permutation σ, the sum of absolute differences between $\overline{F}_\sigma$ and $\underline{F}_\sigma$ corresponds to the sum of the values of the arcs that move from the event $\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\}$ to the event $\{x_{\sigma(1)},\ldots,x_{\sigma(i)},x_{\sigma(i+1)}\}$, for every $i = 0,\ldots,n-1$. Therefore, minimising the total distance corresponds to calculating the *shortest path* from $\emptyset$ to $\mathcal{X}$. Proposition 3 assures that the optimal solution to this shortest path problem will never be unique, because if the optimal solution is attained in the permutation σ, it will also be attained in the reverse permutation $\overline{\sigma}$. Nevertheless, if the only optimal solutions to the shortest path problem are determined by a permutation σ and its reverse $\overline{\sigma}$, then the undominated outer approximation minimising the imprecision is unique.

[5]The idea of comparing two non-additive measures in terms of their distance on some class of events that may be strictly included in $\mathcal{P}(\mathcal{X})$ is already present in the paper by Baroni and Vicig in [5].

We illustrate this procedure in the following example:

Example 6. *Consider the coherent conjugate lower and upper probabilities given by:*

A	$\underline{P}(A)$	$\overline{P}(A)$
$\{x_1\}$	0.25	0.4
$\{x_2\}$	0.2	0.5
$\{x_3\}$	0.2	0.5
$\{x_1, x_2\}$	0.5	0.8
$\{x_1, x_3\}$	0.5	0.8
$\{x_2, x_3\}$	0.6	0.75
$\mathcal{X}$	1	1

These lower and upper probabilities are coherent because they are the lower and upper envelope of the following probability mass functions:

$$(0.4, 0.4, 0.2), \quad (0.25, 0.5, 0.25), \quad (0.3, 0.2, 0.5)$$

The Hasse diagram with the weights defined above is represented in Figure 2.

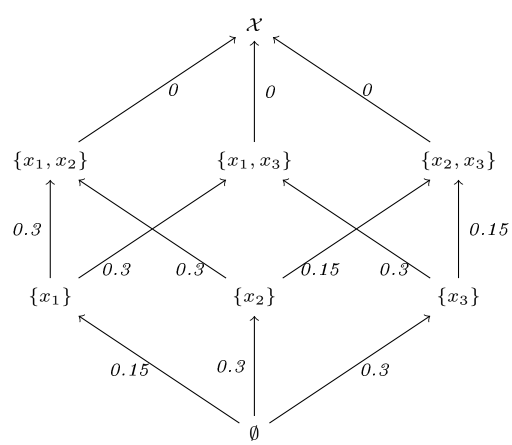


FIGURE 2. Hasse diagram with the weights in Example 6.

Solving the shortest path problem, we obtain four different optimal solutions, those associated with the paths:

$$\emptyset \to \{x_1\} \to \{x_1, x_2\} \to \mathcal{X}, \qquad \emptyset \to \{x_1\} \to \{x_1, x_3\} \to \mathcal{X},$$
$$\emptyset \to \{x_2\} \to \{x_2, x_3\} \to \mathcal{X}, \qquad \emptyset \to \{x_3\} \to \{x_2, x_3\} \to \mathcal{X}.$$

They correspond to the permutations $(1, 2, 3)$, $(1, 3, 2)$, $(2, 3, 1)$ *and* $(3, 2, 1)$, *and have an imprecision of 0.45. From Proposition 3 we know that the permutations* $(1, 2, 3)$ *and* $(3, 2, 1)$ *give rise to the same p-box, and the same applies to* $(1, 3, 2)$ *and* $(2, 3, 1)$, *so we have two different optimal solutions. Although this allows to discard the other two permutations, which have an imprecision of 0.6, it does not single out a unique p-box.* ♦

4.2. **Approach based on specificity measures.** As we have already said, p-boxes are connected with completely monotone lower probabilities ([44, Sec. 5.1]). Indeed, a p-box is equivalent to a completely monotone lower probability whose focal events are ordered intervals (see [13, Sec. 3.3]). Hence, we could also use the approach based on maximising the specificity. For each p-box $(\underline{F}_\sigma, \overline{F}_\sigma)$ we can consider its associated completely monotone lower probability, given by Equation (14), and compute its specificity, as we did in Section 3.4. However, the same drawback as in Section 3.4 appears: this procedure does not produce a unique solution, as we show in the next example.

Example 7. *Consider again Example 6. From Proposition 3 we know that the six different permutations give rise to only three different p-boxes. Hence, we can restrict ourselves to the permutations $\sigma_1 = (1,2,3)$, $\sigma_2 = (1,3,2)$ and $\sigma_3 = (2,1,3)$. In the next table, we show the values of the completely monotone lower probabilities induced by those p-boxes, computed using Equation* (14).[6]

			$\sigma_1 = (1,2,3)$		$\sigma_2 = (1,3,2)$		$\sigma_3 = (2,1,3)$	
A	$\underline{P}(A)$	$\overline{P}(A)$	$\underline{P}_{\sigma_1}(A)$	$m_{\sigma_1}(A)$	$\underline{P}_{\sigma_2}(A)$	$m_{\sigma_2}(A)$	$\underline{P}_{\sigma_3}(A)$	$m_{\sigma_3}(A)$
$\{x_1\}$	0.25	0.4	0.25	0.25	0.25	0.25	0	0
$\{x_2\}$	0.2	0.5	0.1	0.1	0.2	0.2	0.2	0.2
$\{x_3\}$	0.2	0.5	0.2	0.2	0.1	0.1	0.2	0.2
$\{x_1, x_2\}$	0.5	0.8	0.5	0.15	0.45	0	0.5	0.3
$\{x_1, x_3\}$	0.5	0.8	0.45	0	0.5	0.15	0.5	0.3
$\{x_2, x_3\}$	0.6	0.75	0.6	0.3	0.6	0.3	0.4	0
$\mathcal{X}$	1	1	1	0	1	0	1	0

Using Equation (13) *we obtain the following specificities:*

$$S(\underline{P}_{\sigma_1}) = S(\underline{P}_{\sigma_2}) = 0.775, \quad S(\underline{P}_{\sigma_3}) = 0.7.$$

From these values, we can discard the p-box induced by the permutation σ_3, but we are not able to choose between σ_1 and σ_2. Note that this is the same result as in Example 6, where we minimised the imprecision. ♦

This example may lead us to think that the approach based on minimising the imprecision and the one based on maximising the specificity always give rise to the same solutions. However, we shall see in Example 8 later on that this is not the case.

4.3. **Approach based on the BV-distance.** Since the lower probability associated with a p-box is completely monotone, we could also apply the criterion based on minimising the BV-distance between each of the coherent lower probabilities associated with the p-boxes $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$ and $\underline{P}$.

Our next example shows that this criterion does not give rise to a unique solution, and in fact that it does not produce the same solution as the criteria based on minimising the imprecision or maximising the specificity.

Example 8. *Let $\underline{P}, \overline{P}$ be the lower and upper envelope of the following probability mass functions:*

$$(\epsilon, 0.45 - \epsilon, 0.55), \quad (0.4, 0.1, 0.5), \quad (0.3, 0.5, 0.2),$$

[6] In the examples of this section, and for the sake of simplicity, we use the short notation $\underline{P}_{\sigma_i}$ for $\underline{P}_{(\underline{F}_{\sigma_i}, \overline{F}_{\sigma_i})}$ and m_{σ_i} for its associated Möbius inverse.

for some fixed $\epsilon \in (0, 0.03)$*. Their values, as well as the values of the completely monotone lower probabilities associated with the p-boxes* $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$ *are given by:*

A	$\underline{P}(A)$	$\overline{P}(A)$	$\sigma_1 = (1,2,3)$ $\underline{P}_{\sigma_1}(A)$	$\sigma_2 = (1,3,2)$ $\underline{P}_{\sigma_2}(A)$	$\sigma_3 = (2,1,3)$ $\underline{P}_{\sigma_3}(A)$
$\{x_1\}$	ϵ	0.4	ϵ	ϵ	0
$\{x_2\}$	0.1	0.5	0.05	0.1	0.1
$\{x_3\}$	0.2	0.55	0.2	0.1	0.2
$\{x_1, x_2\}$	0.45	0.8	0.45	$0.1 + \epsilon$	0.45
$\{x_1, x_3\}$	0.5	0.9	$0.2 + \epsilon$	0.5	0.5
$\{x_2, x_3\}$	0.6	$1 - \epsilon$	0.6	0.6	0.3
$\mathcal{X}$	1	1	1	1	1

It can be easily seen that:

	Imp	$S(\underline{P}_{\sigma_i})$	$d_{BV}(\underline{P}, \underline{P}_{\sigma_i})$
$\underline{P}_{\sigma_1}$	$0.75 - \epsilon$	$0.625 + \epsilon/2$	$0.35 - \epsilon$
$\underline{P}_{\sigma_2}$	$0.8 - \epsilon$	$0.6 + \epsilon/2$	$0.45 - \epsilon$
$\underline{P}_{\sigma_3}$	0.75	$0.641\overline{6}$	$0.3 + \epsilon$

We see first of all that for $\epsilon = 0.025$ *the criterion based on minimising the BV-distance does not give a unique solution, because both* σ_1 *and* σ_3 *minimise the distance. Moreover the three approaches do not agree, because the criterion based on minimising the imprecision chooses the p-box induced by the permutation* $(1,2,3)$*, the criterion based on maximising the specificity selects the p-box induced by the permutation* $(2,1,3)$*, while the criterion based on minimising the BV-distance selects the p-box induced by the permutation* $(2,1,3)$*, if* $\epsilon \in (0, 0.025)$*, and the p-box induced by the permutation* $(1,2,3)$*, if* $\epsilon \in (0.025, 0.03)$*.* ♦

4.4. **Approach based on the quadratic distance.** We consider again the approach based on minimising the quadratic distance. Unfortunately, in the case of p-boxes this approach is not very useful for several reasons: first of all, the set of undominated outer approximations in $\mathcal{C}_{(\underline{F}, \overline{F})}$ is not convex, hence we cannot use the good properties of the quadratic programming problems; secondly, the interpretation of the quadratic distance is not clear; and thirdly, the solution in this case is not unique, as we show in the next example.

Example 9. *Let us continue with Example 7. There, we have seen that there are three different outer approximations in* $\mathcal{C}_{(\underline{F}, \overline{F})}$*, those associated with the permutations* $\sigma_1 = (1,2,3)$*,* $\sigma_2 = (1,3,2)$ *and* $\sigma_3 = (2,1,3)$*. If we compute their quadratic distance with respect to* $\underline{P}$*, we obtain the following values:*

	$\underline{P}_{\sigma_1}$	$\underline{P}_{\sigma_2}$	$\underline{P}_{\sigma_3}$
$d_q(\underline{P}, \underline{P}_{\sigma_i})$	0.0125	0.0125	0.1025

We see that both $\underline{P}_{\sigma_1}$ *and* $\underline{P}_{\sigma_2}$ *minimise the quadratic distance, and as a consequence the optimal solution is not unique.* ♦

4.5. **Approach based on preference preservation.** In Section 3.3 we considered several properties about preference preservation. For selecting a p-box among those in $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$, we may choose the p-box satisfying one of those conditions. However, as we show in the following example, for each condition it may be that none of the p-boxes or more than one satisfies it.

Example 10. *Let $\underline{P}, \overline{P}$ be the coherent lower and upper probability obtained as the lower and upper envelopes of*

$$(0.2, 0.3, 0.5), \quad (0.5, 0.2, 0.3), \quad (0.3, 0.5, 0.2).$$

Their values, as well as those of the completely monotone lower probabilities associated with the p-boxes $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$, are given by:

A	$\underline{P}(A)$	$\overline{P}(A)$	$\sigma_1 = (1,2,3)$ $\underline{P}_{\sigma_1}(A)$	$\sigma_2 = (1,3,2)$ $\underline{P}_{\sigma_2}(A)$	$\sigma_3 = (2,1,3)$ $\underline{P}_{\sigma_3}(A)$
$\{x_1\}$	0.2	0.5	0.2	0.2	0
$\{x_2\}$	0.2	0.5	0	0.2	0.2
$\{x_3\}$	0.2	0.5	0.2	0	0.2
$\{x_1, x_2\}$	0.5	0.8	0.5	0.4	0.5
$\{x_1, x_3\}$	0.5	0.8	0.4	0.5	0.5
$\{x_2, x_3\}$	0.5	0.8	0.5	0.5	0.4
$\mathcal{X}$	1	1	1	1	1

Next table shows which conditions **C***i are satisfied by these p-boxes:*

	$\underline{P}_{\sigma_1}$	$\underline{P}_{\sigma_2}$	$\underline{P}_{\sigma_3}$
C1	Yes	Yes	Yes
C2	No	No	No
C3	No	No	No
C4	Yes	Yes	Yes
C5	Yes	Yes	Yes
C6	No	No	No

Therefore, none of the conditions allows us to distinguish between these p-boxes. ♦

4.6. **Approach based on the total variation distance.** As we did in Section 3.2, one possibility to choose among the p-boxes in $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$ is to consider those p-boxes minimising one of the extensions of the total variation distance in Equations (9)÷(11). Unfortunately, none of d_1, d_2 and d_3 allows to select a single p-box, as we show in the following example.

Example 11. *Let us continue with Example 6. Example 7 gives the completely monotone lower probabilities associated with the p-boxes in $((\underline{F}_\sigma, \overline{F}_\sigma))_{\sigma \in S_n}$. For them, it holds that:*

	$\sigma_1 = (1,2,3)$ $\underline{P}_{\sigma_1}$	$\sigma_2 = (1,3,2)$ $\underline{P}_{\sigma_2}$	$\sigma_3 = (2,1,3)$ $\underline{P}_{\sigma_3}$
$d_1(\underline{P}, \underline{P}_{\sigma_i})$	0.1	0.1	0.25
$d_2(\underline{P}, \underline{P}_{\sigma_i})$	0.05	0.05	0.125
$d_3(\underline{P}, \underline{P}_{\sigma_i})$	0.4	0.4	0.4

We see that none of the extensions of the total variation distance allows to select a single p-box. ♦

4.7. **Discussion.** There are at most $\frac{n!}{2}$ undominated outer approximations in the class $\mathcal{C}_{(\underline{F},\overline{F})}$. There exist several procedures we can use to discard some of them, being the most reasonable, in our view (i) minimising imprecision; (ii) minimising the BV-distance; and (iii) maximising specificity. Two important drawbacks of these three approaches are that in general they do not select a unique p-box, and that when they do so, there does not seem to be a relationship between the options selected by each of them. Hence, we cannot use all three approaches simultaneously. Using the approach based on minimising the imprecision has the advantage of having a simple procedure for finding the optimal p-box(es), while for the other two approaches we would need to compute the lower probability associated with the p-box by means of Equation (14).

Taking these comments into account, it seems reasonable to (i) consider the p-box(es) minimising the imprecision, albeit with the reservations mentioned in Section 4.1 about Equation (16); if there is more than one, we could (ii) select among them the ones minimising the BV-distance; if there are still more than one p-box, we (iii) select the one maximising the specificity; if necessary, we could (iv) compute their quadratic distance, and if all these methods do not select a single p-box, all of the remaining ones would be equally preferred.

5. Selection of an outer approximation in $\mathcal{C}_\Pi$

In Section 3 we explained that the sets of undominated outer approximations in $\mathcal{C}_2$ and $\mathcal{C}_\infty$ are not finite in general, and indeed we do not have a procedure for determining all of them. In contrast, the set of non-dominating outer approximations in $\mathcal{C}_\Pi$ is finite and can be easily determined (see [34, Sec. 6]), as in the case of p-boxes.

Throughout this section, we shall assume that all non-impossible events have strictly positive upper probability, so that $\overline{P}(\{x\}) > 0$ for every $x \in \mathcal{X}$. This assumption shall be useful in some of the proofs later on. Moreover, as we shall detail in Appendix A, in the general case we can always restrict our attention to $\mathcal{X}^* = \{x \in \mathcal{X} \mid \overline{P}(\{x\}) > 0\}$, determine the outer approximations $\Pi^* \in \mathcal{C}_\Pi$ of the restriction of $\overline{P}$ to $\mathcal{P}(\mathcal{X}^*)$, and then extend these to $\mathcal{P}(\mathcal{X})$ by $\Pi(A) = \Pi^*(A \cap \mathcal{X}^*)$, or equivalently taking $\pi(x) = 0$ for every x satisfying $\overline{P}(\{x\}) = 0$. We refer to Appendix A for detailed explanations and proofs.

Given the conjugate coherent lower and upper probabilities $\underline{P}$ and $\overline{P}$, each permutation $\sigma \in S_n$ defines the following possibility measure[7]:

$$\Pi_\sigma(\{x_{\sigma(1)}\}) = \overline{P}(\{x_{\sigma(1)}\}), \text{ and} \tag{17}$$

$$\Pi_\sigma(\{x_{\sigma(i)}\}) = \max_{A \in \mathcal{A}_{\sigma(i)}} \overline{P}(A \cup \{x_{\sigma(i)}\}), \text{ where for every } i > 1: \tag{18}$$

$$\mathcal{A}_{\sigma(i)} = \left\{ A \subseteq \{x_{\sigma(1)}, \dots, x_{\sigma(i-1)}\} \mid \overline{P}(A \cup \{x_{\sigma(i)}\}) > \max_{x \in A} \Pi_\sigma(\{x\}) \right\}, \tag{19}$$

and $\Pi_\sigma(B) = \max_{x \in B} \Pi_\sigma(\{x\})$ for every $B \subseteq \mathcal{X}$. Then, the family of non-dominating outer approximations of $\overline{P}$ is $(\Pi_\sigma)_{\sigma \in S_n}$ (see [34, Prop. 11, Cor. 13]).

Note that the procedure above is well-defined because, as we have mentioned before, we are assuming that $\overline{P}(\{x\}) > 0$ for every $x \in \mathcal{X}$. Hence $\emptyset \in \mathcal{A}_{\sigma(i)}$, which guarantees that $\mathcal{A}_{\sigma(i)}$ is non-empty.

[7]Here we are assuming that $\max_{x \in \emptyset} \Pi_\sigma(\{x\}) := 0$.

In this section we propose a number of approaches to select a unique outer approximation of $\overline{P}$ among those determined by Equations (17)÷(19). The procedure above may determine the same possibility measure more than once, using different permutations $\sigma \in S_n$. The next result is concerned with such cases, and will be helpful later on for reducing the candidate possibilities.

Proposition 4. *Let $(\Pi_\sigma)_{\sigma \in S_n}$ be the family of non-dominating outer approximations of $\overline{P}$ in $\mathcal{C}_\Pi$. Consider a permutation $\sigma \in S_n$ and its associated possibility measure Π_σ. Assume that there exists $i \in \{2, \ldots, n\}$ such that $\Pi_\sigma(\{x_{\sigma(i)}\}) \neq \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\})$. Then, there exists $\sigma' \in S_n$ such that*

$$\Pi_\sigma(A) = \Pi_{\sigma'}(A) \quad \forall A \in \mathcal{P}(\mathcal{X}) \quad \text{and} \quad \Pi_{\sigma'}(\{x_{\sigma'(j)}\}) = \overline{P}(\{x_{\sigma'(1)}, \ldots, x_{\sigma'(j)}\}) \quad \forall j.$$

Proof. In order to ease the notation, assume that $\sigma = (1, 2, \ldots, n)$, and denote its associated possibility measure by Π.

Take the smallest i such that $\Pi(\{x_i\}) \neq \overline{P}(\{x_1, \ldots, x_i\})$. This means that $\Pi(\{x_1\}) = \overline{P}(\{x_1\})$ and, for any $k = 2, \ldots, i-1$:

$$\begin{aligned}\Pi(\{x_k\}) &= \max\left\{\overline{P}(A \cup \{x_k\}) \mid \overline{P}(A \cup \{x_k\}) > \max_{x_j \in A} \Pi(\{x_j\}),\ A \subseteq \{x_1, \ldots, x_{k-1}\}\right\} \\ &= \overline{P}(\{x_1, \ldots, x_k\}).\end{aligned}$$

Moreover, applying monotonicity we deduce that $\Pi(\{x_k\}) = \overline{P}(\{x_1, \ldots, x_k\}) \geq \overline{P}(\{x_1, \ldots, x_{k-1}\}) = \Pi(\{x_{k-1}\})$ for $k = 2, \ldots, i-1$.

Equations (17)÷(19) also imply that $\Pi(\{x_j\}) \leq \overline{P}(\{x_1, \ldots, x_j\})$ for every $j = 1, \ldots, n$. Thus, if

$$\begin{aligned}\Pi(\{x_i\}) &= \max\left\{\overline{P}(A \cup \{x_i\}) \mid \overline{P}(A \cup \{x_i\}) > \max_{x \in A} \Pi(\{x\}),\ A \subseteq \{x_1, \ldots, x_{i-1}\}\right\} \\ &\neq \overline{P}(\{x_1, \ldots, x_i\}),\end{aligned}$$

then it must be $\Pi(\{x_i\}) < \overline{P}(\{x_1, \ldots, x_i\})$. We deduce that $\{x_1, \ldots, x_{i-1}\} \notin \mathcal{A}_i$, and as a consequence

$$\overline{P}(\{x_1, \ldots, x_i\}) \leq \max_{j=1,\ldots,i-1} \Pi(\{x_j\}) = \Pi(\{x_{i-1}\}) = \overline{P}(\{x_1, \ldots, x_{i-1}\}), \tag{20}$$

so by monotonicity the inequality in (20) becomes an equality. For every $A \subseteq \{x_1, \ldots, x_{i-1}\}$ such that $x_{i-1} \in A$, it holds that:

$$\overline{P}(A \cup \{x_i\}) \leq \overline{P}(\{x_1, \ldots, x_i\}) = \overline{P}(\{x_1, \ldots, x_{i-1}\}) = \Pi(\{x_{i-1}\}) = \max_{x \in A} \Pi(\{x\}),$$

where the first equality follows from Equation (20). Thus, such an A does not belong to $\mathcal{A}_i$, and as a consequence it is not valid for defining $\Pi(\{x_i\})$. If we denote $A_j = \{x_1, \ldots, x_j\}$ for $j = 1, \ldots, n$, this means that $A_{i-1} \notin \mathcal{A}_i$. Consider now $A_{i-2} = \{x_1, \ldots, x_{i-2}\}$:

$$\overline{P}(A_{i-2} \cup \{x_i\}) \geq \overline{P}(\{x_1, \ldots, x_{i-2}\}) = \Pi(\{x_{i-2}\}) = \max_{x \in A_{i-2}} \Pi(\{x\}).$$

Here, we have two options, either $\overline{P}(A_{i-2} \cup \{x_i\}) = \overline{P}(\{x_1, \ldots, x_{i-2}\})$ or $\overline{P}(A_{i-2} \cup \{x_i\}) > \overline{P}(\{x_1, \ldots, x_{i-2}\})$. If the former condition holds, we iterate the procedure. At the end we have two cases:

(1) For every $j \in \{1,\dots,i-1\}$, $\overline{P}(A_j \cup \{x_i\}) = \overline{P}(A_j) = \Pi(\{x_j\})$. In that case, the maximum in Equation (18) is attained for $A = \emptyset$ because it is the only event in $\mathcal{A}_i$, whence $\Pi(\{x_i\}) = \overline{P}(\{x_i\})$ and also $\Pi(\{x_i\}) \leq \Pi(\{x_1\}) \leq \ldots \leq \Pi(\{x_{i-1}\})$. In this case we consider the permutation $\sigma' = (i, 1, \dots, i-1, i+1, \dots, n)$, i.e., the permutation that moves i to the first position. It holds that:

$$\Pi_{\sigma'}(\{x_i\}) = \overline{P}(\{x_i\}) = \Pi(\{x_i\}).$$

$$\begin{aligned}\Pi_{\sigma'}(\{x_1\}) &= \max\left\{\overline{P}(A \cup \{x_1\}) \mid \overline{P}(A \cup \{x_1\}) > \max_{x\in A} \Pi_{\sigma'}(\{x\}),\ A \subseteq \{x_i\}\right\} \\ &= \overline{P}(\{x_1\}) = \Pi(\{x_1\});\end{aligned}$$

to see the second equality note that either $\mathcal{A}_{\sigma'(i)} = \{\emptyset\}$ or $\mathcal{A}_{\sigma'(i)} = \{\emptyset, \{x_i\}\}$. By assumption, $\overline{P}(\{x_1, x_i\}) = \overline{P}(\{x_1\})$. Hence, applying the procedure in Equations (17)÷(19) with $A = \{x_i\}$ or $A = \emptyset$ we obtain the same value $\overline{P}(\{x_1\})$ as candidate assignment for $\Pi(\{x_1\})$.

Suppose now that $\Pi_{\sigma'}(\{x_k\}) = \Pi(\{x_k\})$ for every $k \leq j-1 < j < i$, and let us prove that also $\Pi_{\sigma'}(\{x_j\}) = \Pi(\{x_j\})$. Since $\Pi(\{x_j\}) = \overline{P}(\{x_1,\dots,x_j\})$, there exists some event $B \in \mathcal{A}_j$ such that $\Pi(\{x_j\}) = \overline{P}(B \cup \{x_j\}) = \overline{P}(\{x_1,\dots,x_j\})$. By Equation (19), $B \subseteq \{x_1,\dots,x_{j-1}\}$ and $\overline{P}(B \cup \{x_j\}) > \max_{x\in B} \Pi(\{x\}) = \max_{x\in B} \Pi_{\sigma'}(\{x\})$. This implies that $B \in \mathcal{A}_{\sigma'(j)}$ and as a consequence that

$$\Pi_{\sigma'}(\{x_j\}) \geq \overline{P}(B \cup \{x_j\}) = \overline{P}(\{x_1,\dots,x_j\}).$$

On the other hand, for every $A \in \mathcal{A}_{\sigma'(j)}$, it holds that

$$\overline{P}(A \cup \{x_j\}) \leq \overline{P}(A_j \cup \{x_i\}) = \overline{P}(\{x_1,\dots,x_j\}),$$

whence

$$\Pi_{\sigma'}(\{x_j\}) = \max_{A\in\mathcal{A}_{\sigma'(j)}} \overline{P}(A \cup \{x_j\}) \leq \overline{P}(\{x_1,\dots,x_j\}),$$

and therefore that

$$\Pi_{\sigma'}(\{x_j\}) = \overline{P}(\{x_1,\dots,x_j\}) = \Pi(\{x_j\}).$$

Finally, for every $j > i$, $\mathcal{A}_{\sigma'(j)} = \mathcal{A}_j$, which implies that $\Pi_{\sigma'}(\{x_j\}) = \Pi(\{x_j\})$ for $j = i+1,\dots,n$.

(2) There exists $j \in \{1,\dots,i-1\}$ such that

$$\overline{P}(A_j \cup \{x_i\}) > \overline{P}(\{x_1,\dots,x_j\}) = \Pi(\{x_j\}).$$

In that case, we consider:

$$k = \max\left\{j \in \{1,\dots,i-1\} \mid \Pi(\{x_j\}) < \overline{P}(A_j \cup \{x_i\}) \leq \Pi(\{x_i\})\right\}. \tag{21}$$

In this case, the maximum in Equation (18) is attained in the event $A_k = \{x_1,\dots,x_k\}$, which belongs to $\mathcal{A}_i$ by definition of k. Then, it holds that

$$\Pi(\{x_i\}) = \overline{P}(A_k \cup \{x_i\}) = \overline{P}(\{x_1,\dots,x_k,x_i\}).$$

Now, consider the permutation $\sigma' = (1, \dots, k, i, k+1, \dots, i-1, i+1, \dots, n)$, i.e., the permutation that moves i just after element k. It holds that:

$$\Pi_{\sigma'}(\{x_1\}) = \overline{P}(\{x_1\}) = \Pi(\{x_1\}).$$
$$\dots$$
$$\Pi_{\sigma'}(\{x_k\}) = \overline{P}(\{x_1, \dots, x_k\}) = \Pi(\{x_k\}).$$
$$\begin{aligned}\Pi_{\sigma'}(\{x_i\}) &= \max\left\{\overline{P}(A \cup \{x_i\}) \mid \overline{P}(A \cup \{x_i\}) > \max_{x \in A} \Pi_{\sigma'}(\{x\}),\ A \subseteq \{x_1, \dots, x_k\}\right\} \\ &= \max\left\{\overline{P}(A \cup \{x_i\}) \mid \overline{P}(A \cup \{x_i\}) > \max_{x \in A} \Pi(\{x\}),\ A \subseteq \{x_1, \dots, x_k\}\right\} \\ &= \overline{P}(\{x_1, \dots, x_k, x_i\}) = \Pi(\{x_i\}),\end{aligned}$$

where the last two equalities hold taking $A = A_k = \{x_1, \dots, x_k\}$, because from Equation (21) we have that

$$\Pi(\{x_i\}) = \overline{P}(A_k \cup \{x_i\}) > \overline{P}(A_k) = \Pi(\{x_k\}).$$

For the element x_{k+1}, it holds that:

- $\Pi_{\sigma'}(\{x_{k+1}\}) \geq \Pi(\{x_{k+1}\})$, as a consequence of the inclusion $\mathcal{A}_{k+1} \subseteq \mathcal{A}_{\sigma'(k+1)}$;
- Conversely,

$$\begin{aligned}&\Pi_{\sigma'}(\{x_{k+1}\}) \\ &= \max\{\overline{P}(A \cup \{x_{k+1}\}) \mid \overline{P}(A \cup \{x_{k+1}\}) > \max_{x_j \in A} \Pi_{\sigma'}(\{x_j\}), A \subseteq \{x_1, \dots, x_k, x_i\}\} \\ &\leq \overline{P}(\{x_1, \dots, x_{k+1}, x_i\}) = \Pi(\{x_{k+1}\}),\end{aligned}$$

where the equality follows from Equation (21).

Therefore, $\Pi_{\sigma'}(\{x_{k+1}\}) = \Pi(\{x_{k+1}\})$.

With an analogous reasoning, we obtain that:

$$\Pi_{\sigma'}(\{x_{k+2}\}) = \Pi(\{x_{k+2}\}),\ \dots,\ \Pi_{\sigma'}(\{x_{i-1}\}) = \Pi(\{x_{i-1}\}).$$

Finally, it trivially holds that for any $j = i+1, \dots, n$, $\Pi_{\sigma'}(\{x_j\}) = \Pi(\{x_k\})$, so we conclude that $\Pi_{\sigma'} = \Pi$.

In both cases, $\Pi_{\sigma'}$ satisfies:

$$\Pi_{\sigma'}(\{x_{\sigma'(j)}\}) = \overline{P}(\{x_{\sigma'(1)}, \dots, x_{\sigma'(j)}\}) \quad \forall j = 1, \dots, i.$$

Now, if for $\Pi_{\sigma'}$ there exists $j > i$ such that $\Pi_{\sigma'}(\{x_{\sigma'(j)}\}) \neq \overline{P}(\{x_{\sigma'(1)}, \dots, x_{\sigma'(j)}\})$, we just need to iterate the procedure. □

5.1. **Approach based on the BV-distance.** Our first approach consists in looking for a possibility measure, among $(\Pi_\sigma)_{\sigma \in S_n}$, that minimises the BV-distance with respect to the original model. If we denote by N_σ the conjugate necessity measure of Π_σ, the BV-distance can be expressed by:

$$\begin{aligned}d_{BV}(\underline{P}, N_\sigma) &= \sum_{A \subseteq \mathcal{X}} (\underline{P}(A) - N_\sigma(A)) \\ &= \sum_{A \subseteq \mathcal{X}} (\Pi_\sigma(A) - \overline{P}(A)) = \sum_{A \subseteq \mathcal{X}} \Pi_\sigma(A) - \sum_{A \subseteq \mathcal{X}} \overline{P}(A).\end{aligned}$$

To ease the notation, from now on for each permutation $\sigma \in S_n$, we denote by $\vec{\beta}_\sigma$ the ordered vector determined by the values $\Pi_\sigma(\{x_{\sigma(i)}\})$, $i = 1, \ldots, n$, so that $\beta_{\sigma,1} \leq \ldots \leq \beta_{\sigma,n} = 1$. Using this notation:

$$\sum_{A \subseteq \mathcal{X}} \Pi_\sigma(A) = \beta_{\sigma,1} + 2\beta_{\sigma,2} + \ldots + 2^{n-1}\beta_{\sigma,n} = \sum_{i=1}^{n} 2^{i-1}\beta_{\sigma,i}. \tag{22}$$

This means that, in order to minimise $d_{BV}(\underline{P}, N_\sigma)$, we must minimise Equation (22), or, equivalently, given that the last term is 2^{n-1} for any $\vec{\beta}_\sigma$,

$$\sum_{i=1}^{n-1} 2^{i-1}\beta_{\sigma,i}. \tag{23}$$

It is easy to show that if a dominance relation exists between $\vec{\beta}_\sigma$ and $\vec{\beta}_{\sigma'}$, this induces an order between the values in Equation (22).

Lemma 5. *Let $\vec{\beta}_\sigma$ and $\vec{\beta}_{\sigma'}$ be the vectors associated with the possibility measures Π_σ and $\Pi_{\sigma'}$. If $\beta_{\sigma,i} \leq \beta_{\sigma',i}$ for every $i = 1, \ldots, n$, then $d_{BV}(\underline{P}, N_\sigma) \leq d_{BV}(\underline{P}, N_{\sigma'})$. Furthermore, if $\beta_{\sigma,j} < \beta_{\sigma',j}$ for some $j = 1, \ldots, n$, then $d_{BV}(\underline{P}, N_\sigma) < d_{BV}(\underline{P}, N_{\sigma'})$.*

Proof. The proof follows easily from Equation (22):

$$\begin{aligned}\sum_{A \subseteq \mathcal{X}} \Pi_\sigma(A) &= \beta_{\sigma,1} + 2\beta_{\sigma,2} + \ldots + 2^{n-1}\beta_{\sigma,n} \\ &\leq \beta_{\sigma',1} + 2\beta_{\sigma',2} + \ldots + 2^{n-1}\beta_{\sigma',n} = \sum_{A \subseteq \mathcal{X}} \Pi_{\sigma'}(A).\end{aligned} \tag{24}$$

Then, we conclude that $d_{BV}(\underline{P}, N_\sigma) \leq d_{BV}(\underline{P}, N_{\sigma'})$. If in addition $\beta_{\sigma,j} < \beta_{\sigma',j}$ for some $j = 1, \ldots, n$, the inequality in Equation (24) is strict. □

This result may contribute to rule out some of the permutations in S_n, as illustrated in the next example.

Example 12. *Consider the coherent conjugate lower and upper probabilities from Example 6. The possibility measures Π_σ and their associated vectors $\vec{\beta}_\sigma$ for every $\sigma \in S_n$ are given in Table 1.*

σ	$\Pi_\sigma(\{x_1\})$	$\Pi_\sigma(\{x_2\})$	$\Pi_\sigma(\{x_3\})$	$\vec{\beta}_\sigma$
$\sigma_1 = (1,2,3)$	0.4	0.8	1	$(0.4, 0.8, 1)$
$\sigma_2 = (1,3,2)$	0.4	1	0.8	$(0.4, 0.8, 1)$
$\sigma_3 = (2,1,3)$	0.8	0.5	1	$(0.5, 0.8, 1)$
$\sigma_4 = (2,3,1)$	1	0.5	0.75	$(0.5, 0.75, 1)$
$\sigma_5 = (3,1,2)$	0.8	1	0.5	$(0.5, 0.8, 1)$
$\sigma_6 = (3,2,1)$	1	0.75	0.5	$(0.5, 0.75, 1)$

TABLE 1. Possibility measures $(\Pi_\sigma)_{\sigma \in S_n}$ for the coherent lower and upper probabilities in Example 12, as well as their associated vectors $(\vec{\beta}_\sigma)_{\sigma \in S_n}$.

Taking $\sigma_1 = (1,2,3)$ *and* $\sigma_3 = (2,1,3)$*, we can see that*

$$\vec{\beta}_{\sigma_1} = (0.4, 0.8, 1), \quad \vec{\beta}_{\sigma_3} = (0.5, 0.8, 1).$$

Since $\beta_{\sigma_1} \lneq \beta_{\sigma_3}$*, Lemma 5 implies* $d_{BV}(\underline{P}, N_{\sigma_1}) < d_{BV}(\underline{P}, N_{\sigma_3})$*. Hence, we can discard* Π_{σ_3}*. The same happens with the vectors* $\vec{\beta}_{\sigma_1}$ *and* $\vec{\beta}_{\sigma_5}$*, so we deduce that* $d_{BV}(\underline{P}, N_{\sigma_1}) < d_{BV}(\underline{P}, N_{\sigma_5})$*.* ♦

In the general case, the family $(\vec{\beta}_\sigma)_{\sigma\in S_n}$ is not totally ordered. Then, the problem of minimising the BV-distance is solved by casting it into a shortest path problem, similarly as we did in Section 4.1, as we shall now illustrate.

As we said before, the possibility measure(s) in $(\Pi_\sigma)_{\sigma\in S_n}$ that minimise the BV-distance to the original model will be the one(s) for which the sum $\sum_{A\subseteq\mathcal{X}} \Pi_\sigma(A)$ is minimised. In turn, this sum can be computed by means of Equation (22), once we order the values $\Pi_\sigma(\{x_{\sigma(i)}\})$, for $i = 1, \dots, n$. As a consequence, if Π_σ satisfies

$$\Pi_\sigma(\{x_{\sigma(i)}\}) = \overline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}) \quad \forall i = 1, \dots, n, \tag{25}$$

then the monotonicity of $\overline{P}$ will imply that $\Pi_\sigma(\{x_{\sigma(1)}\}) \leq \Pi_\sigma(\{x_{\sigma(2)}\}) \leq \cdots \leq \Pi_\sigma(\{x_{\sigma(n)}\})$, and then by Equation (22), that

$$\sum_{A\subseteq\mathcal{X}} \Pi_\sigma(A) = \sum_{i=1}^{n} 2^{i-1}\overline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}).$$

On the other hand, if Π_σ does not satisfy Equation (25), then by construction (see Equations (17)÷(19)) we have that $\Pi_\sigma(\{x_{\sigma(i)}\}) \leq \overline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\})$ for every $i = 1, \dots, n$, with strict inequality on some i. This means that

$$\sum_{A\subseteq\mathcal{X}} \Pi_\sigma(A) < \sum_{i=1}^{n} 2^{i-1}\overline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}).$$

Further, from Proposition 4 we know that if Π_σ does not satisfy Equation (25) then it is possible to find another permutation σ' that does so and such that $\Pi_\sigma(A) = \Pi_{\sigma'}(A)$ for every $A \subseteq \mathcal{X}$.

This means that we can find a Π_σ minimising the BV-distance by solving a shortest path problem. For this aim, we consider the Hasse diagram of $\mathcal{P}(\mathcal{X})$, and we assign the following weights: if $x_i \notin A$, we assign the weight $2^{|A|}\overline{P}(A \cup \{x_i\})$ to the edge $A \to A \cup \{x_i\}$, and the fictitious weight 0 to $\mathcal{X} \setminus \{x_i\} \to \mathcal{X}$. Since these weights are non-negative, we can find the shortest path efficiently by means of Dijkstra's algorithm [14, 19].

In this diagram, there are two types of paths:

(a) Paths whose associated possibility measure Π_σ satisfies Equation (25); then the value of Equation (23) for Π_σ coincides with the value of the path, with the weights established above.

(b) Paths whose associated possibility measure Π_σ does not satisfy the equality in Equation (25); then the value of Equation (23) for Π_σ shall be strictly smaller than the value of the path, and shall moreover coincide with the value of the path determined by some other permutation σ', as established in Proposition 4. This means that the shortest path can never be found among these ones.

As a consequence, if we find the shortest path we shall determine a permutation σ whose associated possibility measure Π_σ satisfies Equation (25), i.e., it coincides with the upper probability that we are outer approximating in the chain of events determined by the path. Moreover, this possibility measure shall minimise the BV-distance with respect to the original model among all the non-dominating outer approximations in $\mathcal{C}_\infty$. In this manner we shall obtain all such possibility measures; although we may not be able to identify the set of all permutations that generate them, this allows us to skip the procedure in Equations (17)÷(19).

Example 13. *Consider the coherent conjugate lower and upper probabilities from Example 12. Figure 3 pictures the Hasse diagram with the weights on the edges we discussed before.*

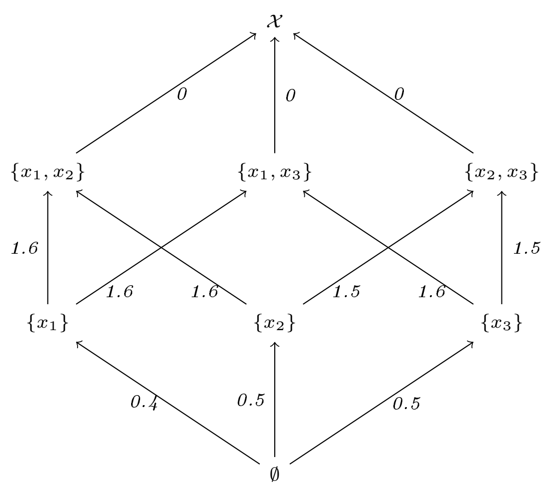


FIGURE 3. Hasse diagram with the associated weights for Example 13.

Solving the shortest path problem from $\emptyset$ to $\mathcal{X}$ using Dijkstra's algorithm, we obtain an optimal value of 2 that is attained with the following paths:

$$\emptyset \to \{x_1\} \to \{x_1, x_2\} \to \mathcal{X}, \qquad \emptyset \to \{x_1\} \to \{x_1, x_3\} \to \mathcal{X}.$$
$$\emptyset \to \{x_2\} \to \{x_2, x_3\} \to \mathcal{X}, \qquad \emptyset \to \{x_3\} \to \{x_2, x_3\} \to \mathcal{X}.$$

These four paths correspond to the permutations $\sigma_1 = (1,2,3)$, $\sigma_2 = (1,3,2)$, $\sigma_4 = (2,3,1)$ and $\sigma_6 = (3,2,1)$. Even if they induce four different possibility measures, all of them are at the same distance (with respect to d_{BV}) from $\overline{P}$. The other two possibility measures are those that were discarded in Example 12 using Lemma 5. ♦

This example shows that with this approach we obtain the possibility measure Π_σ at a minimum BV-distance. It also shows that the solution is not unique, and that the vectors $\vec{\beta}_\sigma$ and $\vec{\beta}_{\sigma'}$ that are not pointwisely ordered may be associated with two different possibility measures Π_σ and $\Pi_{\sigma'}$ minimising the BV-distance (such as σ_1 and σ_6 in the example). Nevertheless, we can determine situations in which the BV-distance selects one single Π_σ, using the following result.

Proposition 6. *Let $\underline{P}$ and $\overline{P}$ be coherent conjugate lower and upper probabilities. If there is a permutation $\sigma \in S_n$ satisfying*

$$\overline{P}\left(\{x_{\sigma(1)},\dots,x_{\sigma(j)}\}\right) = \min_{|A|=j} \overline{P}(A) \quad \forall j = 1,\dots,n, \tag{26}$$

then Π_σ minimises the BV-distance.

Proof. A possibility measure Π_σ minimises the BV-distance with respect to $\overline{P}$ when it minimises Equation (22). Moreover, naming $\sigma \in S_n$ a permutation satisfying Equation (26), we obtain its associated vector $\vec{\beta}_\sigma$ given by:

$$\vec{\beta}_\sigma = \left(\min_{|A|=1} \overline{P}(A), \min_{|A|=2} \overline{P}(A), \dots, \min_{|A|=n-1} \overline{P}(A), 1\right).$$

Since any permutation σ' that is not discarded in Proposition 4 will satisfy

$$\Pi_{\sigma'}(\{x_{\sigma'(1)},\dots,x_{\sigma'(i)}\}) \geq \min_{|A|=i} \overline{P}(A) = \Pi_\sigma(\{x_{\sigma(1)},\dots,x_{\sigma(i)}\}),$$

we deduce that $\vec{\beta}_\sigma$ is pointwisely dominated by any other $\vec{\beta}_{\sigma'}$, hence using Lemma 5 we obtain that $d_{BV}(\underline{P}, N_\sigma) \leq d_{BV}(\underline{P}, N_{\sigma'})$. □

As a consequence of this result, if there is only one permutation satisfying Equation (26), this approach allows to select a unique undominated outer approximation. The next example illustrates this:

Example 14. *Consider again the conjugate coherent lower and upper probabilities $\underline{P}$ and $\overline{P}$ in Example 3. It holds that:*

$$\overline{P}(\{x_4\}) = 0.4 = \min_{|A|=1} \overline{P}(A).$$

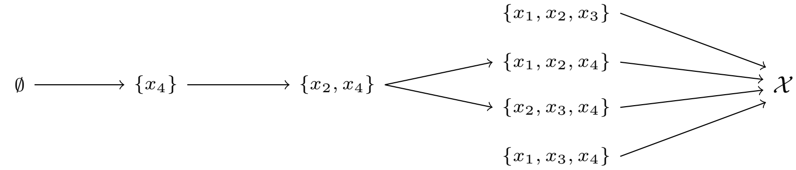
$$\overline{P}(\{x_2, x_4\}) = 0.7 = \min_{|A|=2} \overline{P}(A).$$

$$\overline{P}(\{x_1, x_2, x_3\}) = 1 = \overline{P}(\{x_1, x_2, x_4\}) = \overline{P}(\{x_1, x_3, x_4\}) = \overline{P}(\{x_2, x_3, x_4\}).$$

There are two chains of events satisfying Equation (26), namely $\{x_4\} \subseteq \{x_2, x_4\} \subseteq \{x_1, x_2, x_4\} \subseteq \mathcal{X}$ and $\{x_4\} \subseteq \{x_2, x_4\} \subseteq \{x_2, x_3, x_4\} \subseteq \mathcal{X}$. They are associated with the permutations $\sigma = (4, 2, 1, 3)$ and $\sigma' = (4, 2, 3, 1)$. From Proposition 6, the possibility measure $\Pi_\sigma = \Pi_{\sigma'}$ they determine, that is given by the possibility distribution $\pi = (1, 0.7, 1, 0.4)$, is the unique undominated outer approximation in $\mathcal{C}_\Pi$ minimising the BV-distance. ♦

The verification of Equation (26) can be done quite simply using tools from graph theory. For any $k \in \{0,\dots,n\}$, let $\mathcal{A}_k = \{A \subseteq \mathcal{X}$ such that $|A| = k, \overline{P}(A) = \min_{|B|=k} \overline{P}(B)\}$. We obtain in particular that $\mathcal{A}_0 = \{\emptyset\}$ and $\mathcal{A}_n = \{\mathcal{X}\}$. Consider now a graph where the nodes are the sets in $\mathcal{A}_k$ for $k = 0,\dots,n$, and where we add an arrow from $A \in \mathcal{A}_k$ to $B \in \mathcal{A}_{k+1}$ if and only if $A \subset B$. It follows that Equation (26) holds if and only if in the resulting graph there exists a path from $\emptyset$ to $\mathcal{X}$.

Applying this on Example 14, we obtain the following graph:



Trivially, there is a path from $\emptyset$ to $\mathcal{X}$. This path determines a chain of events satisfying Equation (26).

Nevertheless, the condition in Proposition 6 is only sufficient, but not necessary. See for instance Examples 12 and 13: there, neither $\sigma_1 = (1,2,3)$ nor $\sigma_2 = (1,3,2)$ satisfy Equation (26) because the sequence of events with minimal upper probability does not form a chain, but both Π_{σ_1} and Π_{σ_2} minimise the BV-distance. To see that in those examples Equation (26) does not hold, note that

$$\min_{|A|=1} \overline{P}(A) = 0.4 = \overline{P}(\{x_1\}), \quad \min_{|A|=2} \overline{P}(A) = 0.75 = \overline{P}(\{x_2, x_3\}).$$

We obtain thus a graph with only two arrows: $\emptyset \to \{x_1\}$ and $\{x_2, x_3\} \to \mathcal{X}$, and where it is therefore impossible to build a path from $\emptyset$ to $\mathcal{X}$.

5.2. **Approach based on the quadratic distance.** We consider now the approach based on minimising the quadratic distance d_q considered in Equation (8). In this case, since we are dealing with possibility measures instead of their conjugate necessity measures, we can rewrite the quadratic distance as:

$$d_q(\overline{P}, \Pi) = \sum_{A \subseteq \mathcal{X}} \big(\Pi(A) - \overline{P}(A)\big)^2. \tag{27}$$

Unfortunately, selecting the possibility measures among those in $(\Pi_\sigma)_{\sigma \in S_n}$ by this approach fails for several reasons. Firstly, as we have already mentioned, the interpretation of this distance is not clear; secondly, the main advantage of using the quadratic distance in Section 3.1 is that the feasible region of the minimisation problem is convex, but this is not the case with our family $(\Pi_\sigma)_{\sigma \in S_n}$; and thirdly, since the feasible region is not convex, we cannot guarantee the uniqueness of a solution, as we show in the next example.

Example 15. *Consider the same conjugate coherent lower and upper probabilities of Example 12. In Table 1 we have specified the possibility measures* $(\Pi_\sigma)_{\sigma \in S_n}$*; their quadratic distance with respect to* $\overline{P}$ *are given by next table:*

	Π_{σ_1}	Π_{σ_2}	Π_{σ_3}	Π_{σ_4}	Π_{σ_5}	Π_{σ_6}
$d_q\big(\overline{P}, \Pi_{\sigma_i}\big)$	0.4425	0.4425	0.5125	0.5025	0.5125	0.5025

We can see that there are two possibility measures minimising the quadratic distance, those associated with the permutations $\sigma_1 = (1,2,3)$ *and* $\sigma_2 = (1,3,2)$. ♦

5.3. **Approach based on measuring specificity.** Since any possibility measure is in particular a plausibility measure (that is, the conjugate of a belief function), it makes sense to compare them by means of specificity measures. In this section, we investigate which possibility measure(s) among $(\Pi_\sigma)_{\sigma \in S_n}$ are the most specific.

With each possibility measure in $(\Pi_\sigma)_{\sigma \in S_n}$, we consider again its associated vector $\vec{\beta}_\sigma$. In the case of possibility measures, we know that the focal events are nested: they are given by $A_i := \{x_{\sigma(n-i+1)}, \ldots, x_{\sigma(n)}\}$, with $m(A_i) = \beta_{\sigma, n-i+1} -$

$\beta_{\sigma,n-i}$. Hence the specificity measure in Equation (13) simplifies to:

$$\begin{aligned} S(\Pi_\sigma) &= \frac{\beta_{\sigma,n}-\beta_{\sigma,n-1}}{1}+\frac{\beta_{\sigma,n-1}-\beta_{\sigma,n-2}}{2}+\ldots+\frac{\beta_{\sigma,2}-\beta_{\sigma,1}}{n-1}+\frac{\beta_{\sigma,1}}{n} \\ &= \beta_{\sigma,n}-\beta_{\sigma,n-1}\left(1-\frac{1}{2}\right)-\beta_{\sigma,n-2}\left(\frac{1}{2}-\frac{1}{3}\right)-\ldots-\beta_{\sigma,1}\left(\frac{1}{n-1}-\frac{1}{n}\right) \\ &= 1-\frac{\beta_{\sigma,n-1}}{2}-\frac{\beta_{\sigma,n-2}}{2\cdot 3}-\ldots-\frac{\beta_{\sigma,1}}{n(n-1)}. \end{aligned}$$

Thus, a most specific possibility measure will minimise

$$\frac{\beta_{\sigma,1}}{n(n-1)}+\frac{\beta_{\sigma,2}}{(n-1)(n-2)}+\ldots+\frac{\beta_{\sigma,n-1}}{2}. \tag{28}$$

Our first result is similar to Lemma 5, and allows to discard some of the possibility measures Π_σ. For this aim we use again the vectors $\vec{\beta}_\sigma$.

Lemma 7. *Let $\vec{\beta}_\sigma$ and $\vec{\beta}_{\sigma'}$ be the vectors associated with the possibility measures Π_σ and $\Pi_{\sigma'}$. If $\beta_{\sigma,i} \le \beta_{\sigma',i}$ for every $i = 1,\ldots,n$, then $S(\Pi_\sigma) \ge S(\Pi_{\sigma'})$. Furthermore, if $\beta_{\sigma,j} < \beta_{\sigma',j}$ for some $j = 1,\ldots,n$, then $S(\Pi_\sigma) > S(\Pi_{\sigma'})$.*

Proof. For the permutations σ and σ', the specificities of Π_σ and $\Pi_{\sigma'}$ are related as follows:

$$\begin{aligned} S(\Pi_\sigma) &= 1-\frac{\beta_{\sigma,n-1}}{2}-\frac{\beta_{\sigma,n-2}}{2\cdot 3}-\ldots-\frac{\beta_{\sigma,1}}{n(n-1)} \\ &\ge 1-\frac{\beta_{\sigma',n-1}}{2}-\frac{\beta_{\sigma',n-2}}{2\cdot 3}-\ldots-\frac{\beta_{\sigma',1}}{n(n-1)} = S(\Pi_{\sigma'}). \end{aligned} \tag{29}$$

Then, $S(\Pi_\sigma) \ge S(\Pi_\sigma)$. If in addition $\beta_{\sigma,j} < \beta_{\sigma,j}$ for some $j = 1,\ldots,n$, the inequality in Equation (29) is strict. □

Example 16. *Consider again Examples 12 and 13. In Table 1 we can see the possibility measures $(\Pi_\sigma)_{\sigma\in S_n}$ and their associated vectors $(\vec{\beta}_\sigma)_{\sigma\in S_n}$. As we argued in Example 12, $\vec{\beta}_{\sigma_1} \lneq \vec{\beta}_{\sigma_3}$, where $\sigma_1 = (1,2,3)$ and $\sigma_3 = (2,1,3)$. Hence according to Lemma 7, $S(\Pi_{\sigma_1}) > S(\Pi_{\sigma_3})$. This means that we can discard Π_{σ_3}. A similar reasoning allows us to discard Π_{σ_5}.* ♦

If we want to find those possibility measures maximising the specificity, we have to minimise Equation (28). Here we can make the same considerations as in the previous section: if the possibility measure Π_σ associated with a permutation σ satisfies Equation (25) then the monotonicity of $\overline{P}$ implies that $\Pi_\sigma(\{x_{\sigma(1)}\}) \le \Pi_\sigma(\{x_{\sigma(2)}\}) \le \cdots \le \Pi_\sigma(\{x_{\sigma(n)}\})$, and then Equation (28) becomes

$$\sum_{i=1}^{n-1}\frac{\overline{P}(\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\})}{(n-i)(n-i+1)}.$$

On the other hand, if Π_σ does not satisfy Equation (25), then by construction $\Pi_\sigma(\{x_{\sigma(i)}\}) \le \overline{P}(\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\})$ for every $i = 1,\ldots,n$, with strict inequality on some i. This means that

$$S(\Pi_\sigma) > 1-\sum_{i=1}^{n-1}\frac{\overline{P}(\{x_{\sigma(1)},\ldots,x_{\sigma(i)}\})}{(n-i)(n-i+1)},$$

or, equivalently, that the value of Equation (28) for Π_σ is strictly smaller than

$$\sum_{i=1}^{n-1} \frac{\overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\})}{(n-i)(n-i+1)}.$$

Moreover, from Proposition 4 we know that if Π_σ does not satisfy Equation (25) then it is possible to find another permutation σ' that does so and such that $\Pi_\sigma = \Pi_{\sigma'}$.

This means that we can find a Π_σ maximising the specificity by solving a shortest path problem, similarly to what we did in the case of the BV-distance. For this aim, we consider the Hasse diagram of $\mathcal{P}(\mathcal{X})$, and for every $A \neq \mathcal{X}$ and $x_i \notin A$ we assign the weight

$$\frac{\overline{P}(\{A \cup \{x_i\}\})}{(n-|A|-1)(n-|A|)} \tag{30}$$

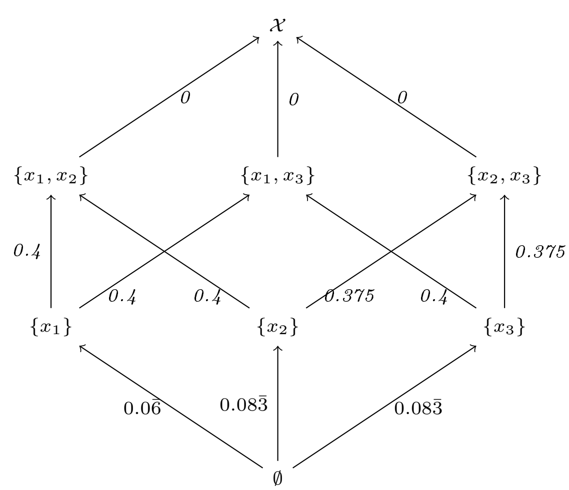
to the edge $A \to A \cup \{x_i\}$, and we give the weight 0 to $\mathcal{X} \setminus \{x_i\} \to \mathcal{X}$.

In this diagram, there are two types of paths:

(a) Paths whose associated possibility measure Π_σ satisfies Equation (25); then the value of Equation (28) for Π_σ coincides with the value of the path.
(b) Paths whose associated possibility measure Π_σ does not satisfy the equality in Equation (25); then the value of Equation (28) for Π_σ shall be strictly smaller than the value of the path, and shall moreover coincide with the value of the path determined by some other permutation σ', as established in Proposition 4. This means that the shortest path can never be found among these ones.

As a consequence, if we find the shortest path we shall determine a permutation σ whose associated possibility measure Π_σ satisfies Equation (25), and therefore that maximises the specificity. In this manner we shall obtain all such possibility measures; although we may not be able to identify the set of all permutations that generate them, this allows us to skip the computations in Equations (17)÷(19).

Example 17. *Consider again the running Examples 12, 13 and 16. In the next figure we can see the Hasse diagram of $\mathcal{P}(\mathcal{X})$ with the weights from Equation (30).*



Solving the shortest path problem from $\emptyset$ to $\mathcal{X}$, we obtain two optimal solutions:

$$\emptyset \to \{x_2\} \to \{x_2, x_3\} \to \mathcal{X}, \qquad \emptyset \to \{x_3\} \to \{x_2, x_3\} \to \mathcal{X}.$$

They correspond to the permutations $\sigma_4 = (2, 3, 1)$ and $\sigma_6 = (3, 2, 1)$. ♦

These examples also illustrate that the approach based on minimising the BV-distance and the approach based on maximising the specificity are not equivalent: in Example 12 we have seen that the possibility measures minimising the BV-distance are the ones associated with the permutations $(3, 1, 2)$ and $(3, 2, 1)$, while those maximising the specificity are the ones associated with $(2, 3, 1)$ and $(3, 2, 1)$. Then, the possibility measure associated with the permutation $(2, 3, 1)$ maximises the specificity but does not minimise the BV-distance with respect to $\underline{P}$, while the possibility measure associated with the permutation $(3, 1, 2)$ minimises the BV-distance but does not maximise the specificity. Hence, it seems that the best solution in this case is the possibility measure associated with the permutation $(3, 2, 1)$, which is the only one minimising the BV-distance and maximising the specificity at the same time.

To conclude this subsection, we prove that in the same conditions of Proposition 6, the specificity measure may allow to select a unique possibility measure Π_σ.

Proposition 8. *Let $\underline{P}$ and $\overline{P}$ be coherent conjugate lower and upper probabilities. If there is a permutation $\sigma \in S_n$ satisfying Equation* (26)*, then Π_σ maximises the specificity.*

Proof. For maximising the specificity, a possibility measure must minimise Equation (13). If $\sigma \in S_n$ is a permutation satisfying Equation (26), its associated vector $\vec{\beta}_\sigma$ is given by:

$$\vec{\beta}_\sigma = \left(\min_{|A|=1} \overline{P}(A), \min_{|A|=2} \overline{P}(A), \ldots, \min_{|A|=n-1} \overline{P}(A), 1 \right).$$

By definition, $\vec{\beta}_\sigma$ is pointwisely dominated by any other $\vec{\beta}_{\sigma'}$, hence using Lemma 7 we obtain that $S(\Pi_\sigma) \geq S(\Pi_{\sigma'})$ for every $\sigma' \in S_n$. □

We arrive at the same conclusion of Proposition 6: if there is a unique permutation satisfying Equation (26), then there is a unique possibility measure maximising the specificity; and in that case the chosen possibility measure maximises the specificity and at the same time minimises the BV-distance.

5.4. **Approach based on preference preservation.** In order to choose among the possibility measures $(\Pi_\sigma)_{\sigma \in S_n}$, we could try again the approach in Section 3.3 based on different conditions about preference preservation. In this case we shall consider the versions of **C1–C6** in terms of upper probabilities: for instance, **C1** now becomes

$$\overline{P}(A) < \overline{P}(B) \Rightarrow \overline{Q}(A) < \overline{Q}(B).$$

Note that this is no loss of generality with respect to conditions **C1–C4**: for instance, it can easily be shown using conjugacy that

$$\begin{aligned} &\overline{P}(A) < \overline{P}(B) \Rightarrow \overline{Q}(A) < \overline{Q}(B) \quad \forall A, B \subseteq \mathcal{X} \\ \Leftrightarrow\ &\underline{P}(A) < \underline{P}(B) \Rightarrow \underline{Q}(A) < \underline{Q}(B) \quad \forall A, B \subseteq \mathcal{X}, \end{aligned}$$

meaning that it does not matter if we formulate condition **C1** in terms of lower or in terms of upper probabilities. A similar reasoning can be made for **C2**, **C3** and **C4**. It does make a difference, however, in the case of conditions **C5** and **C6**. For instance, in the case of **C6** the conditions

$$\underline{P}(\{x\}) = \underline{P}(\{x'\}) \Rightarrow \underline{Q}(\{x\}) = \underline{Q}(\{x'\}) \quad \forall x, x' \in \mathcal{X}$$

and

$$\overline{P}(\{x\}) = \overline{P}(\{x'\}) \Rightarrow \overline{Q}(\{x\}) = \overline{Q}(\{x'\}) \quad \forall x, x' \in \mathcal{X}$$

will not be equivalent in general. Since in this section our uncertainty is expressed in terms of coherent upper probabilities and possibility measures, we shall consider the versions of **C5** and **C6** in terms of upper probabilities. To avoid misunderstandings, we will denote the conditions in terms of upper probabilities as $\overline{\mathbf{C}i}$ instead of $\mathbf{C}i$.

Nevertheless, the properties of preference preservation are not really helpful in the selection problem, as we show in the next example.

Example 18. *Let $\overline{P}$ be the upper envelope of the probability mass functions:*

$$(0.1, 0.2, 0.7), \quad (0.1, 0.6, 0.3), \quad (0.5, 0.5, 0), \quad (0.4, 0.1, 0.5).$$

This upper probability, as well as the possibility measures in $(\Pi_\sigma)_{\sigma \in S_n}$, are given by:

σ		$(1,2,3)$	$(1,3,2)$	$(2,1,3)$	$(2,3,1)$	$(3,1,2)$	$(3,2,1)$
A	$\overline{P}$	Π_{σ_1}	Π_{σ_2}	Π_{σ_3}	Π_{σ_4}	Π_{σ_5}	Π_{σ_6}
$\{x_1\}$	0.5	0.5	0.5	1	1	0.9	1
$\{x_2\}$	0.6	1	1	0.6	0.6	1	0.9
$\{x_3\}$	0.7	0.9	0.9	0.9	0.9	0.7	0.7
$\{x_1, x_2\}$	1	1	1	1	1	1	1
$\{x_1, x_3\}$	0.9	0.9	0.9	1	1	0.9	1
$\{x_2, x_3\}$	0.9	1	1	0.9	0.9	1	0.9
$\mathcal{X}$	1	1	1	1	1	1	1

It can be easily seen that none of the Π_{σ_i}, for $i = 1, \dots, n$, satisfies $\overline{\mathbf{C5}}$, and as a consequence they do not satisfy the stronger conditions $\overline{\mathbf{C1}}$, $\overline{\mathbf{C2}}$ and $\overline{\mathbf{C4}}$. Also, property $\overline{\mathbf{C3}}$ is neither satisfied by the Π_{σ_i}, for $i = 1, \dots, n$. Finally, condition $\overline{\mathbf{C6}}$ holds trivially in this example because all the values of $\overline{P}$ in the singletons are different.

This means that the none of the conditions $\overline{\mathbf{C}i}$ is useful in this example. ♦

5.5. **Approach based on the total variation distance.** Our last approach is based on selecting the possibility measure among $(\Pi_\sigma)_{\sigma \in S_n}$ by minimising one of the extensions of the total variation distance with respect to $\overline{P}$. In that case, the distances d_1, d_2 and d_3 given in Equations (9)÷(11) must be rewritten in terms of the upper probabilities, giving rise to the distances $\overline{d_1}$, $\overline{d_2}$ and $\overline{d_3}$ given by:

$$\overline{d_1}(\overline{P}_1, \overline{P}_2) = \max_{A \subseteq \mathcal{X}} |\overline{P}_1(A) - \overline{P}_2(A)|,$$

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

$$\overline{d_3}(\overline{P}_1, \overline{P}_2) = \sup_{P_1 \leq \overline{P}_1, P_2 \leq \overline{P}_2} \left(\max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)| \right).$$

As in the case of $\mathcal{C}_2$ and $\mathcal{C}_\infty$, this approach is not fruitful:

Example 19. *Consider again Example 12. If we compute* $\overline{d_i}(\overline{P}, \Pi_j)$*, for* $i = 1, 2, 3$ *and* $j = 1, \dots, 6$*, we obtain the following values:*

	Π_{σ_1}	Π_{σ_2}	Π_{σ_3}	Π_{σ_4}	Π_{σ_5}	Π_{σ_6}
$\overline{d_1}(\overline{P}, \Pi_{\sigma_i})$	0.5	0.5	0.5	0.6	0.5	0.6
$\overline{d_2}(\overline{P}, \Pi_{\sigma_i})$	0.4	0.4	0.45	0.425	0.45	0.425
$\overline{d_3}(\overline{P}, \Pi_{\sigma_i})$	0.8	0.8	0.8	0.75	0.8	0.75

Thus, none of $\overline{d_1}, \overline{d_2}$ *or* $\overline{d_3}$ *allows to select a single possibility measure. On the other hand, the intersection of the sets of optimal outer approximations with respect to* $\overline{d_1}, \overline{d_2}$ *and* $\overline{d_3}$ *is empty: in other words, there is no possibility measure among* $(\Pi_\sigma)_{\sigma \in S_n}$ *minimising the three distances simultaneously.* ♦

5.6. **Discussion.** In this section we have explored five different approaches for selecting a non-dominating outer approximation in $\mathcal{C}_\Pi$, i.e., among the possibility measures $(\Pi_\sigma)_{\sigma \in S_n}$, where Π_σ is defined following Equations (17)÷(19). These approaches are based on minimising the quadratic or the BV-distance, maximising the specificity, some preference preservation and minimising the total variation distance.

With respect to the idea of minimising the quadratic distance, it has the drawback in this case of not producing a unique solution, while being also less intuitive than the other approaches, in our opinion.

Concerning the BV-distance and specificity measures, we have seen a simple procedure for finding the possibility measures minimising the BV-distance or maximising the specificity; we have also showed that these two approaches yield the same solution in the cases depicted in Propositions 6 and 8. The procedure is based on solving a shortest path problem. The drawback in both cases is again that there could be more than one optimal solution. In that situation, and using the same ideas as in Section 3, we propose: (i) to look for the possibility measures minimising the BV-distance; if this procedure does not give a unique solution, (ii) choose among them the possibility measure(s) with greatest specificity; if again there is no unique possibility measure, (iii) compute their quadratic distance (see Equation (27)) with respect to the initial model and select the one minimising it; if again there is no single solution, (iv) all of them are equally preferred, hence we can select any of them.

6. Conclusions

In our previous work [33, 34], we have considered the problem of approximating a coherent lower probability by a more tractable model that satisfies some interesting additional property, such as 2-monotonicity. In determining an optimal outer approximation, we considered two criteria in [33, 34]: first of all, that the outer approximation is undominated, meaning that it is not possible to find another outer approximation of the original model that is more precise; since the set of such outer approximations is difficult to determine, and nonetheless infinite, we have reduced it by focusing on those that are closest to the original model, in terms of the distance between imprecise probability models proposed by Baroni and Vicig. This has the advantage that the set can be determined by means of linear programming, and that the distance has a clear interpretation as a measure of imprecision. However, in most cases the set of solutions to this problem will have more than one element and for this reason we have considered in this paper a number of comparison criteria between its elements.

The criteria we have considered can be grouped into two categories: on the one hand, we can analyse how similar are the initial model and its outer approximation. Here, we have considered other distances, such as the quadratic one or generalisations of the total variation distance, and we have also compared the two models in terms of the preferences they encompass. In the first line of work, we should remark that the quadratic distance allows us to single out a unique outer approximation, while this is not the case for the total variation distance. With respect to the second criterion, we have explored a number of possibilities, but all of them can be shown to be too weak or too restrictive for our purposes. In this respect, it may be interesting to consider the preference modelling in terms of the sets of desirable gambles associated with the coherent lower probabilities [48, 51], and to define comparison measures between them.

Our second approach has been to consider how tight is the outer approximation. Here, we have measured this in terms of the specificity measure, but this is not the only possible approach; we may instead consider other measures of imprecision or non-specificity, such as those proposed in [3, 4]. Note also that our two approaches are somewhat related, as we have discussed in Section 3 (see Equations (6) and (7)).

As future work, we would like to point out (i) the study of the selection of outer approximations within some distortion models, complementing the work in [33] and extending it to the imprecise case; (ii) a deeper study on the preservation of preferences from the point of view of optimality criteria in decision making with imprecise probabilities [43], or by characterising the set of outer approximations that encompass the same preferences as the original model; (iii) to see if some additional properties can be established when the original model is 2-monotone and we want its outer approximation to satisfy a stronger condition, such as complete monotonicity; (iv) to deepen the analysis of the computational complexity of the different approaches, to see the extent to which they can be employed in large possibility spaces; and (v) finally, while we believe that minimising the BV-distance is the best approach for obtaining undominated outer approximations, it would be interesting to compare our outer approximations with those that minimise the Kullback-Leibler divergence, considering that the latter is a widespread tool within probability theory and that it has already been used in the framework of imprecise probabilities (see for example [1, 35]).

Acknowledgements

The research reported in this paper has been supported by project PGC2018-098623-B-I00 from the Ministry of Science, Innovation and Universities of Spain. This work is an expanded version, with proofs and additional results, of a paper presented at the IPMU'2020 conference [28]. We would also like to thank the reviewers for their comments, that led to corrections and improvements on the first version of the manuscript.

References

[1] J. Abellán and M. Gómez. Measures of divergence on credal sets. *Fuzzy Sets and Systems*, 157:1514–1531, 2005.

[2] J. Abellán and A.R. Masegosa. An ensemble method using credal decision trees. *European Journal of Operational Research*, 205(1):218–226, 2010.

[3] J. Abellán and S. Moral. A non-specificity measure for convex sets of probability distributions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(3):357–367, 2000.

[4] J. Abellán and S. Moral. Maximum entropy for credal sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:587–597, 2003.

[5] P. Baroni and P. Vicig. An uncertainty interchange format with imprecise probabilities. *International Journal of Approximate Reasoning*, 40:147–180, 2005.

[6] A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors, *ISIPTA '09 – Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 61–70, Durham, United Kingdom, 2009. SIPTA.

[7] G. Busanello and D. Petturiti. Dag representation of asymmetric independence models arisingin coherent conditional possibility theory. *Fuzzy Sets and Systems*, 250:1–24, 2014.

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

[9] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.

[10] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.

[11] L. M. de Campos, M. T. Lamata, and S. Moral. Distances between fuzzy measures through associated probabilities: some applications. *Fuzzy Sets and Systems*, 35:57–68, 1990.

[12] T. Denoeux. Decision-making with belief functions: A review. *International Journal of Approximate Reasoning*, 109:87–110, 2019.

[13] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I. Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[14] E. Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematik*, pages 269–271, 1959.

[15] D. Dubois and H. Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24:161–182, 1987.

[16] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.

[17] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In R. Lowen and M. Roubens, editors, *Fuzzy Logic- State of the Art*, pages 103–112. Kluwer, 1993.

[18] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.

[19] J. Gross and J. Yellen. *Graph theory and its applications*. Chapman and Hall, 2006.

[20] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

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

[22] I. Levi. *The enterprise of knowledge*. MIT Press, Cambridge, 1980.

[23] D. A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

[24] S. Lopatazidis, J. De Bock, G. de Cooman, S. De Vuyst, and J. Walraevens. Robust queueing theory: an initial study using imprecise probabilities. *Queueing Systems*, 82:75–101, 2016.

[25] Z. A. Melzak. Problem 4458. *Am. Math. Monthly*, 58, 1951.

[26] E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[27] E. Miranda and I. Montes. Shapley and Banzhaf values as probability transformations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 26(6):917–947, 2018.

[28] E. Miranda, I. Montes, and P. Vicig. On the elicitation of an optimal outer approximation of a coherent lower probability. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2020. Communications in Computer and Information Science*, volume 1238, pages 67–81. Springer, 2020.

[29] I. Montes, E. Miranda, and S. Destercke. Pari-mutuel probabilities as an uncertainty model. *Information Sciences*, 481:550 – 573, 2019.
[30] I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models: Part I- New results on old models. *International Journal of General Systems*, 49(6):602–635, 2020.
[31] I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models: Part II- New models and synthesis. *International Journal of General Systems*, 49(6):636–674, 2020.
[32] I. Montes, E. Miranda, and S. Montes. Decision making with imprecise utilities and beliefs by means of statistical preference and stochastic dominance. *European Journal of Operational Research*, 234(1):209–220, 2014.
[33] I. Montes, E. Miranda, and P. Vicig. 2-monotone outer approximations of coherent lower probabilities. *International Journal of Approximate Reasoning*, 101:181–205, 2018.
[34] I. Montes, E. Miranda, and P. Vicig. Outer approximating coherent lower probabilities with belief functions. *International Journal of Approximate Reasoning*, 110:1–30, 2019.
[35] S. Moral. Discounting imprecise probabilities. In E. Gil, J. Gil, and M. Gil, editors, *The Mathematics of the Uncertain*, volume 142 of *Studies in Systems, Decision and Control*. Springer, 2018.
[36] S. Moral. Learning with imprecise probabilities as model selection and averaging. *International Journal of Approximate Reasoning*, 109:111–124, 2019.
[37] S. Moral and L.M. de Campos. Partially specified belief functions. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 492–499. Morgan and Kaufmann, 1993.
[38] R. Pelessoni and P. Vicig. Imprecise previsions for risk measurement. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 11(4):393–412, 2003.
[39] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51:1145–1158, 2010.
[40] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
[41] L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.
[42] L. S. Shapley. A value for n-person games. *Annals of Mathematics Studies*, 28:307–317, 1953.
[43] M. C. M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29, 2007.
[44] M. C. M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767–791, 2011.
[45] M. C. M. Troffaes, E. Miranda, and S. Destercke. On the connection between probability boxes and possibility measures. *Information Sciences*, 224:88–108, 2013.
[46] L. V. Utkin and F.P.A. Coolen. Imprecise reliability: an introductory overview. In *Computational intelligence in reliability engineering*, pages 261–306. Springer, 2007.
[47] F. Voorbraak. A computationally efficient approximation of Dempster Shafer theory. *International Journal of Man-Machine Studies*, 30:525–536, 1989.
[48] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
[49] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9:249–260, 1983.
[50] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
[51] M. Zaffalon and E. Miranda. Axiomatising incomplete preferences through sets of desirable gambles. *Journal of Artificial Intelligence Research*, 60:1057–1126, 2017.

Appendix A. On the strict positivity assumption for upper probabilities

Throughout Section 5, we have assumed that all (non-impossible) events have strictly positive upper probability. To see that we can assume this without loss of generality, we establish first a property of coherent upper probabilities.

Lemma 9. *Let $\overline{P} : \mathcal{P}(\mathcal{X}) \to [0,1]$ be a coherent upper probability. If $\overline{P}(A_0) = 0$ for a given $A_0 \subset \mathcal{X}$, then $\overline{P}(A \cup A_0) = \overline{P}(A)$ for every $A \subseteq \mathcal{X}$.*

Proof. Since any coherent upper probability is subadditive [48, Sec. 2.7.4(d)], we have that

$$\overline{P}(A \cup A_0) \leq \overline{P}(A) + \overline{P}(A_0) = \overline{P}(A) \leq \overline{P}(A \cup A_0),$$

where last inequality follows by monotonicity [48, Sec. 2.7.4(c)]. Thus, $\overline{P}(A \cup A_0) = \overline{P}(A)$. □

This property allows us to establish the following:

Proposition 10. *Let $\overline{P}, \overline{Q} : \mathcal{P}(\mathcal{X}) \to [0,1]$ be two coherent upper probabilities such that $\overline{P} \leq \overline{Q}$. Assume that $\overline{P}(A_0) = 0 < \overline{Q}(A_0)$ for a given $A_0 \subset \mathcal{X}$, and let us define $\overline{Q}' : \mathcal{P}(\mathcal{X}) \to [0,1]$ by*

$$\overline{Q}'(A) = \overline{Q}(A \setminus A_0) \quad \forall A \subseteq \mathcal{X}. \tag{31}$$

Then,

(1) $\overline{P} \leq \overline{Q}' \lneq \overline{Q}$.
(2) If $\overline{Q}$ is k-alternating, so is $\overline{Q}'$.
(3) If $\overline{Q}$ is a possibility measure, so is $\overline{Q}'$.

Proof. To see that $\overline{Q}'$ is normalised, i.e., that $\overline{Q}'(\mathcal{X}) = 1$, note that

$$\overline{Q}'(\mathcal{X}) = \overline{Q}(\mathcal{X} \setminus A_0) \geq \overline{P}(\mathcal{X} \setminus A_0) = \overline{P}(\mathcal{X}) = 1,$$

where the one-but last equality follows from Lemma 9.

(1) By definition and coherence of $\overline{Q}$, it holds that $\overline{Q}' \leq \overline{Q}$. The inequality is strict because $\overline{Q}'(A_0) = \overline{Q}(\emptyset) = 0 < \overline{Q}(A_0)$.
To see that $\overline{P} \leq \overline{Q}'$, note that for any event A it holds that $\overline{Q}'(A) = \overline{Q}(A \setminus A_0) \geq \overline{P}(A \setminus A_0) = \overline{P}(A)$, where last equality follows applying Lemma 9.

(2) Consider events $A_1, \ldots, A_p$, with $p \leq k$ and let us establish that

$$\overline{Q}'\big(\cap_{i=1}^{p} A_i\big) \leq \sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \overline{Q}'\big(\cup_{i \in I} A_i\big). \tag{32}$$

Since $\overline{Q}'(A) = \overline{Q}(A \setminus A_0)$ for every $A \subseteq \mathcal{X}$, it follows that

$$\begin{aligned}
\sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \overline{Q}'\big(\cup_{i \in I} A_i\big) &= \sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \overline{Q}\big((\cup_{i \in I} A_i) \setminus A_0\big) \\
&= \sum_{\emptyset \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \overline{Q}\big(\cup_{i \in I} (A_i \setminus A_0)\big) \\
&\geq \overline{Q}\big(\cap_{i=1}^{p} (A_i \setminus A_0)\big) = \overline{Q}\big((\cap_{i=1}^{p} A_i) \setminus A_0\big) \\
&= \overline{Q}'\big(\cap_{i=1}^{p} A_i\big),
\end{aligned}$$

where the inequality holds because $\overline{Q}$ is k-alternating. Thus, Equation (32) follows.

(3) Let us prove that $\overline{Q}'$ satisfies Equation (4). Consider two events A, B.

$$\overline{Q}'(A \cup B) = \overline{Q}\big((A \cup B) \setminus A_0\big) = \overline{Q}\big((A \setminus A_0) \cup (B \setminus A_0)\big)$$
$$= \max\big\{\overline{Q}(A \setminus A_0), \overline{Q}(B \setminus A_0)\big\} = \max\big\{\overline{Q}'(A), \overline{Q}'(B)\big\}.$$

As a consequence $\overline{Q}'$ is a possibility measure. □

Note that the coherence of $\overline{Q}$ does not imply in general that of $\overline{Q}'$:

Example 20. *Let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, $P \in \mathbb{P}(\mathcal{X})$ the probability measure associated with the mass function $(0.5, 0.5, 0, 0)$ and $\overline{Q}$ the upper envelope of the probability mass functions $(0.5, 0.5, 0, 0)$ and $(0.25, 0.25, 0.25, 0.25)$. Then, taking into account that a probability measure is in particular a coherent upper probability, $P \leq \overline{Q}$ and $P(\{x_4\}) = 0 < 0.25 = \overline{Q}(\{x_4\})$. If we consider the upper probability $\overline{Q}'$ determined by $\overline{Q}$ and $A_0 = \{x_4\}$ in Equation (31), we obtain*

$$\overline{Q}'(\{x_1, x_3\}) = \overline{Q}'(\{x_2, x_3\}) = 0.5,\ \overline{Q}'(\{x_3\}) = 0.25,\ \overline{Q}'(\{x_4\}) = 0.$$

This implies that $\overline{Q}'$ is not coherent, since any probability $Q \in \mathbb{P}(\mathcal{X})$ satisfying $Q \leq \overline{Q}'$ and $Q(\{x_3\}) = 0.25$ should satisfy $Q(\{x_1\}) \leq 0.25, Q(\{x_2\}) \leq 0.25, Q(\{x_4\}) = 0$, and as a consequence it would not be normalised. ♦

Proposition 10 allows us to deduce the following:

Corollary 11. *Let $\overline{P} : \mathcal{P}(\mathcal{X}) \to [0,1]$ be a coherent upper probability and let $\overline{Q}$ be a non-dominating outer approximation of $\overline{P}$ in $\mathcal{C}_2$, $\mathcal{C}_\infty$ or $\mathcal{C}_\Pi$. If $\overline{P}(\{x\}) = 0$, then also $\overline{Q}(\{x\}) = 0$.*

Proof. Ex-absurdo, if $\overline{Q}(\{x\}) > 0$ then we can define the coherent upper probability $\overline{Q}'$ from Equation (31) and it will be an outer approximation of $\overline{P}$ that is dominated by $\overline{Q}$, a contradiction. □

As a consequence, we may assume without loss of generality that $\overline{P}(\{x\}) > 0$ for every $x \in \mathcal{X}$; otherwise, we consider the event $\mathcal{X}_0 := \{x \in \mathcal{X} \mid \overline{P}(\{x\}) = 0\}$. Note that by subadditivity of the coherent $\overline{P}$, also $\overline{P}(\mathcal{X}_0) = 0$. Thus, we work with the restriction of $\overline{P}$ to $\mathcal{P}(\mathcal{X} \setminus \mathcal{X}_0)$ and its outer approximations, and then extend each $\overline{Q}^*$ of these to $\mathcal{P}(\mathcal{X})$ simply by making $\overline{Q}(A) = \overline{Q}^*(A \setminus \mathcal{X}_0)$ for every $A \subseteq \mathcal{X}$. In the particular case of possibility measures, taking $\Pi(A) = \Pi^*(A \setminus \mathcal{X}_0)$ is equivalent to assign the value $\pi(x) = 0$ to any element x in $\mathcal{X}_0$.

If we restrict to the non-dominating outer approximations of $\overline{P}$ in $\mathcal{P}(\mathcal{X} \setminus \mathcal{X}_0)$ and apply the extension above to each of them, we obtain that $\overline{Q}$ is non-dominating also on $\mathcal{P}(\mathcal{X})$. In fact, for any other outer approximation $\overline{Q}_1$ on $\mathcal{P}(\mathcal{X})$ there must be some A such that $\overline{Q}_1(A) \geq \overline{Q}_1(A \setminus \mathcal{X}_0) \gneq \overline{Q}^*(A \setminus \mathcal{X}_0) = \overline{Q}(A)$ by monotonicity.

University of Oviedo, Spain
Email address: mirandaenrique@uniovi.es

University of Oviedo, Spain
Email address: imontes@uniovi.es

University of Trieste, Italy
Email address: paolo.vicig@deams.units.it