PARI-MUTUEL PROBABILITIES AS AN UNCERTAINTY MODEL

IGNACIO MONTES, ENRIQUE MIRANDA, AND SEBASTIEN DESTERCKE

Abstract. The pari-mutuel model is a betting scheme that has its origins in horse racing, and that has been applied in a number of contexts, mostly economics. In this paper, we consider the set of probability measures compatible with a pari-mutuel model, characterize its extreme points, and investigate the properties of the associated lower and upper probabilities. We show that the pari-mutuel model can be embedded within the theory of probability intervals, and prove necessary and sufficient conditions for it to be a belief function or a minitive measure. In addition, we also investigate the combination of different pari-mutuel models and their definition on product spaces.

Keywords: Pari-mutuel bets, credal sets, probability intervals, belief functions, information fusion.

1. Introduction

The *pari-mutuel model* (PMM) is a betting scheme originated in horse racing, and since then has often been employed in economics. If we consider a gambler betting on a event A and let $P_0(A)$ be the fair price for a bet that returns 1 if A happens, the gambler's gain is $I_A - P_0(A)$, while the house (a bookmaker, an insurance, ...) gains $P_0(A) - I_A$, with I_A the indicator function of event A. In order to assure a positive gain expectation, the house may increase the price of the bet by $1+\delta$, transforming its gain into $(1+\delta)P_0(A) - I_A$. This coefficient δ is then interpreted as some kind of taxation or commission coming from the house. We refer to [25, 31, 33] for some works on this model and to [14] for a critical study in the context of life insurance.

Beyond its use in economic problems, the pari-mutuel model has also been advocated as interesting within imprecise probability theory [37]. In this context, the discounted value $(1+\delta)P_0(A)$ can be interpreted as an upper bound of some probability, and one can consider the associated set $\mathcal{M}(P_0, \delta)$ of dominated probabilities. This induces a neighbourhood around an initial estimate P_0 that may be considered too precise. Such a probability set (or its associated expectation bounds) can then be used in different contexts, such as classification [34, 35] or risk analysis [24].

While working with sets of probabilities may be more realistic in a number of situations where the information is imprecise or ambiguous [2, 13, 15], its use also increases computational complexity, and the elicitation of the imprecise probability model is not always immediate. Because of this, it is interesting to consider models that can cope with scarce information while remaining simple of use. In this paper, we study the pari-mutuel model and investigate to which extent it satisfies these requirements.

Indeed, while some theoretically oriented studies for this model already exist [24, 37], many of its more numerical or practical aspects remain unstudied. Rectifying this issue is the task we set forth in this paper, where we explore practical aspects of the PMM as an imprecise probability model defined on a finite space. After recalling some preliminaries in Section 2, we study the following aspects of the PMM:

- The extreme points of the set of probability measures associated with a pari-mutuel model; in particular, we provide in Section 3 bounds on the number of such points, and characterize in which cases these bounds are attained. Also, we analyze their structure and we show how to compute them.
- Next in Section 4 we study the relationship between the PMM and other imprecise probability models. Although it is immediate to show that a PMM is always 2-monotone (and in particular coherent in the sense of Walley [37]), here we show that the PMM can be regarded as a particular instance of probability intervals. In addition, we also determine in which cases a PMM is equivalent to a belief and a plausibility function, and under which conditions it produces a minitive measure.
- Section 5 briefly discusses and illustrates the problem of outer-approximating a given probability set by a PMM model. This problem actually turns out to be quite simple and has a known solution, but is of practical importance, as approximate reasoning often plays a key role in complex applications involving imprecise probabilities [1].
- Finally, in Section 6 we tackle the problem of combining multiple PMMs, either defined on the same space, in which case the typical task is to merge these models into a single one, or on different spaces, in which case the aim is usually to build a joint model in the product space.

The paper concludes with some additional discussion in Section 7.

2. Preliminary concepts

We devote this section to the introduction of basic notions about imprecise probabilities and the PMM.

2.1. **Imprecise probabilities.**

The theory of imprecise probabilities [37] is an alternative to probability theory that is useful when the information about the experiment under study does not allow us to elicit a unique probability.

Given a universe $\mathcal{X}$, a *lower probability* on the power space $\mathcal{P}(\mathcal{X})$ of $\mathcal{X}$ is a functional $\underline{P} : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$, where $\underline{P}(A)$ can be understood as a lower bound for the unknown value of the probability of A. Any lower probability defines, by means of conjugacy, an upper probability $\overline{P} : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ by means of the formula $\overline{P}(A) = 1 - \underline{P}(A^c)$. Following the previous interpretation, $\overline{P}(A)$ can be interpreted as the upper bound of the unknown probability of A.

Any lower probability $\underline{P}$, and its conjugate upper probability $\overline{P}$, defines a convex and closed set of probabilities, usually called *credal set*, that contains all the probabilities compatible with the information given by $\underline{P}$ (and $\overline{P}$):

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

In this paper we are interested in a particular type of lower (and conjugate upper) probabilities, those satisfying the consistency requirement of *coherence*:

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

In other words, coherence means that $\underline{P}$ (resp., $\overline{P}$) is the lower (resp., upper) envelope of its credal set.

Coherent lower (and upper) probabilities satisfy the following properties (see [37, Section 2.7.4]):

Consistency of $\underline{P}$ and $\overline{P}$: $\underline{P}(A) \leq \overline{P}(A)$ for every A.
Monotonicity: $\underline{P}(A) \leq \underline{P}(B)$ and $\overline{P}(A) \leq \overline{P}(B)$ for every $A \subseteq B$.
Sub-additivity of $\overline{P}$: $\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B)$ for every A, B.
Super-additivity of $\underline{P}$: $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$ for every A, B.

Since $\mathcal{M}(\underline{P})$ is a closed and convex set of probabilities, it is characterized by its extreme points: a probability $P \in \mathcal{M}(\underline{P})$ is an extreme point if $P = \alpha P_1 + (1-\alpha)P_2$ for $\alpha \in (0,1)$, $P_1, P_2 \in \mathcal{M}(\underline{P})$ implies $P_1 = P_2 = P$.

2.2. **Pari-Mutuel Model.** In this paper we shall assume that our possibility space $\mathcal{X}$ is finite: $\mathcal{X} = \{x_1, \ldots, x_n\}$. Consider a precise probability measure P_0 defined on $\mathcal{P}(\mathcal{X})$. We shall assume throughout that for every $i = 1, \ldots, n$, $P_0(\{x_i\}) > 0$; our results can be extended straightforwardly to the general case[1].

Definition 1. *Let P_0 be a probability measure on $\mathcal{P}(\mathcal{X})$, and take $\delta > 0$. The* pari-mutuel model *(PMM) induced by P_0, δ, denoted by (P_0, δ), is given by the following lower and upper probabilities:*

$$\underline{P}(A) = \max\{(1+\delta)P_0(A) - \delta, 0\} \text{ and } \overline{P}(A) = \min\{(1+\delta)P_0(A), 1\} \ \forall A \subseteq \mathcal{X}. \quad (1)$$

Note that the assumption of $P_0(\{x_i\}) > 0 \ \forall i = 1, \ldots, n$ implies that $\overline{P}(A) \geq P_0(A) > 0$ for every $A \subseteq \mathcal{X}$. Moreover, Eq. (1) implies that $\overline{P}, \underline{P}$ are conjugate, meaning that $\overline{P}(A) = 1 - \underline{P}(A^c) \ \forall A \subseteq \mathcal{X}$.

It is also important to remark that in some works based on the PMM, the following definition is considered:

$$\underline{P}(A) = (1+\delta)P_0(A) - \delta \text{ and } \overline{P}(A) = (1+\delta)P_0(A) \ \forall A \subseteq \mathcal{X}. \quad (2)$$

However, as discussed by Walley [37, Sec. 2.9.3], for large values of δ this may produce lower and upper probabilities that are not coherent: specifically, if $P_0(A) > \frac{1}{1+\delta}$ we obtain $\overline{P}(A) > 1$ and $\underline{P}(A^c) < 0$, and as a consequence $\overline{P}, \underline{P}$ are not the upper and lower envelopes of the set of probability measures they bound. Since in this paper we are investigating the properties of a PMM from the point of view of imprecise probabilities, we have decided to follow Walley's suggestion and to work with the definition given by Eq. (1).

Remark 1. *In order to understand the meaning of the parameter δ in a PMM, note that* [37, Sec. 2.9.3] *$\overline{P}(A) - \underline{P}(A) \leq \delta$ for every $A \subseteq \mathcal{X}$, and that*

$$\begin{aligned}\overline{P}(A) - \underline{P}(A) = \delta &\iff (1+\delta)P_0(A) - \delta = \underline{P}(A) \text{ and } \overline{P}(A) = (1+\delta)P_0(A) \\ &\iff \frac{1}{1+\delta} \geq P_0(A) \geq \frac{\delta}{1+\delta}.\end{aligned}$$

In particular, $\overline{P}(A) - \underline{P}(A) = \delta$ whenever $0 < \underline{P}(A) < \overline{P}(A) < 1$. Therefore, δ may be understood in terms of the imprecision in the definition of $P(A)$. ♦

[1] Simply consider that if $P_0(\{x\}) = 0$, then $(1+\delta)P_0(\{x\}) = 0$ as well; thus, if $P_0(A) = 0$ we also obtain $\overline{P}(A) = 0$, and this allows to make a one-to-one correspondence between the elements of $\mathcal{M}(P_0, \delta)$ and those in $\mathcal{M}(P_0', \delta)$, where P_0' is the restriction of P_0 to $\mathcal{X}' := \{x \in \mathcal{X} : P_0(\{x\}) > 0\}$.

In this paper, we are going to study some properties of the PMM as an imprecise probability model [2]. Other works in this direction were carried out in [24, 37]. In particular, in [24] the authors studied the connection between the PMM and risk measures, the problem of updating a PMM and its extension to lower and upper expectation functionals, in the sense of Walley.

We begin by noting that, since the lower probability of a PMM can be obtained as a convex transformation of a probability measure, it follows [10] that $\underline{P}$ is 2-monotone, meaning that

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

for every $A, B \subseteq \mathcal{X}$. Since $\underline{P}, \overline{P}$ in Eq. (1) are conjugate, it follows that $\overline{P}$ is 2-alternating:

$$\overline{P}(A \cup B) + \overline{P}(A \cap B) \leq \overline{P}(A) + \overline{P}(B) \quad \forall A, B \subseteq \mathcal{X}.$$

As a consequence [37], $\underline{P}, \overline{P}$ are *coherent* lower and upper probabilities, that is, they are respectively the lower and upper envelopes of the *credal set* associated with the PMM, given by

$$\mathcal{M}(P_0, \delta) = \{P \text{ probability} \mid \underline{P}(A) \leq P(A) \leq \overline{P}(A) \ \forall A \subseteq \mathcal{X}\}. \tag{3}$$

Since the coherence of a PMM implies that it is uniquely determined by its (closed and convex) associated credal set, it becomes interesting to determine the extreme points of the set $\mathcal{M}(P_0, \delta)$ given by Eq. (3). This is what we set out to do in the following section, and it is particularly relevant if we want to use the PMM in some applied contexts, such as credal networks [1, 7]. Later on we shall study the connection of PMM with probability intervals and belief functions, as well as how to combine two different PMMs.

3. Extreme points induced by a PMM

In this section we are going to study the set $ext(\mathcal{M}(P_0, \delta))$ of extreme points of the credal set $\mathcal{M}(P_0, \delta)$ associated with a PMM.

Since the lower probability of a PMM is 2-monotone, the extreme points of $\mathcal{M}(P_0, \delta)$ are associated with permutations of $\mathcal{X}$ [5], in the following manner: if S^n denotes the permutations of $\{1, \ldots, n\}$, then for every $\sigma \in S^n$ we consider the probability measure P_σ given by

$$\begin{aligned} &P_\sigma(\{x_{\sigma(1)}\}) = \overline{P}(\{x_{\sigma(1)}\}), \\ &P_\sigma(\{x_{\sigma(k)}\}) = \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(k-1)}\}) \ \forall k = 2, \ldots, n. \end{aligned} \tag{4}$$

Then, the extreme points of $\mathcal{M}(P_0, \delta)$ are the probability measures P_σ defined as above: $ext(\mathcal{M}(P_0, \delta)) = \{P_\sigma : \sigma \in S^n\}$. As a consequence, the number of extreme points of $\mathcal{M}(P_0, \delta)$ is bounded above by $n!$, the number of permutations of a n-element space. In this section, we are going to study if this upper bound can be lowered for the credal sets associated with a PMM. Some results, for the particular case where P_0 is the uniform probability distribution, can be found in [34, Sect. 5.2] and [35, Sect. 4.2] [2].

[2] Although in [35] the authors define a PMM by means Eq. (2) instead of Eq. (1), both these definitions give rise to the same credal set, and therefore our results about the extreme points are also applicable in their context.

3.1. **Maximal number of extreme points.** We begin by establishing two preliminary but helpful properties of the PMM. As a coherent upper probability, $\overline{P}$ is sub-additive, but it is not necessarily additive. Our next result establishes additivity under some conditions.

Lemma 1. *Let $\overline{P}$ be the upper probability induced by a PMM (P_0, δ) by means of Eq. (1). If $\overline{P}(A) < 1$, then*

$$\overline{P}(A) = \sum_{x \in A} \overline{P}(\{x\}). \tag{5}$$

Proof. By Eq. (1), $\overline{P}(A) < 1$ implies that $\overline{P}(A) = (1+\delta)P_0(A)$. Furthermore, monotonicity of $\overline{P}$ implies that $\overline{P}(\{x\}) \leq \overline{P}(A) < 1$ for every $x \in A$, and therefore $\overline{P}(\{x\}) = (1+\delta)P(\{x\})$. Hence:

$$\overline{P}(A) = (1+\delta)P_0(A) = (1+\delta)\sum_{x \in A} P_0(\{x\}) = \sum_{x \in A} \overline{P}(\{x\}). \qquad \square$$

We deduce that if $\overline{P}(A \cup B) < 1$ and $A \cap B = \emptyset$, then

$$\overline{P}(A \cup B) = \sum_{x \in A \cup B} \overline{P}(\{x\}) = \sum_{x \in A} \overline{P}(\{x\}) + \sum_{x \in B} \overline{P}(\{x\}) = \overline{P}(A) + \overline{P}(B).$$

This lemma allows us to give an alternative expression for the set $\mathcal{M}(P_0, \delta)$.

Corollary 1. *Let $\mathcal{M}(P_0, \delta)$ denote the credal set associated with a PMM (P_0, δ) by means of Eq. (3). Then, a probability measure P belongs to $\mathcal{M}(P_0, \delta)$ if and only if*

$$P(\{x\}) \leq (1+\delta)P_0(\{x\}) \quad \forall x \in \mathcal{X}. \tag{6}$$

Proof. To see that the condition is necessary, note that every element P of $\mathcal{M}(P_0, \delta)$ satisfies:

$$P(\{x\}) \leq \overline{P}(\{x\}) = \min\{(1+\delta)P_0(\{x\}), 1\} \leq (1+\delta)P_0(\{x\}).$$

To see that Eq. (6) implies that $P(A) \leq \overline{P}(A)$ for every A, we must consider two cases. On the one hand, if $\overline{P}(A) = 1$, then trivially $P(A) \leq \overline{P}(A)$. On the other hand, if $\overline{P}(A) < 1$, then from Lemma 1 :

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

where the inequality follows from Eq. (6). We conclude that $P(A) \leq \overline{P}(A)$ for every $A \subseteq \mathcal{X}$, and therefore $P \in \mathcal{M}(P_0, \delta)$. $\square$

Corollary 1 tells us that the set $\mathcal{M}(P_0, \delta)$ is entirely specified by upper bounds over the singletons of $\mathcal{X}$. Using the additivity property (5) from Lemma 1, we can prove the second preliminary result, which gives the form of the extreme points of $\mathcal{M}(P_0, \delta)$ in terms of $\underline{P}$ and $\overline{P}$.

Lemma 2. *Let P_0 be a probability on $\mathcal{X}$, $\delta > 0$ and $\underline{P}, \overline{P}$ be given by Eq. (1). The extreme point P_σ associated with the permutation σ by Eq. (4) is given by:*

$$\begin{aligned}
P(\{x_{\sigma(i)}\}) &= \overline{P}(\{x_i\}) \quad \forall i = \sigma(1), \ldots, \sigma(j-1), \\
P(\{x_{\sigma(j)}\}) &= \underline{P}(\{x_{\sigma(j)}, \ldots, x_{\sigma(n)}\}), \\
P(\{x_{\sigma(j+1)}\}) &= \ldots = P(\{x_{\sigma(n)}\}) = 0,
\end{aligned}$$

where $j \in \{1, \ldots, n\}$ satisfies $\overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j-1)}\}) < \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j)}\}) = 1$.

Note that if $j = 1$, the expression becomes $P(\{x_{\sigma(1)}\}) = 1$ and $P(\{x_{\sigma(2)}\}) = \ldots = P(\{x_{\sigma(n)}\}) = 0$.

Proof. By Lemma 1 and Eq. (4), the extreme point associated with σ is given by:

$$
\begin{aligned}
P(\{x_{\sigma(1)}\}) &= \overline{P}(\{x_{\sigma(1)}\}).\\
P(\{x_{\sigma(2)}\}) &= \overline{P}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \overline{P}(\{x_{\sigma(1)}\}) = \overline{P}(\{x_{\sigma(2)}\}).\\
&\ldots\\
P(\{x_{\sigma(j-1)}\}) &= \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j-1)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j-2)}\}) = \overline{P}(\{x_{\sigma(j-1)}\}).\\
P(\{x_{\sigma(j)}\}) &= \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j)}\}) - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j-1)}\})\\
&= 1 - \overline{P}(\{x_{\sigma(1)}, \ldots, x_{\sigma(j-1)}\}) = \underline{P}(\{x_{\sigma(j)}, \ldots, x_{\sigma(n)}\}),
\end{aligned}
$$

where the last equality follows by the conjugacy of $\underline{P}$ and $\overline{P}$; therefore, $P_\sigma(\{x_{\sigma(k)}\}) = 0$ for $k = j+1, \ldots, n$. □

The above result is illustrated in the following example:

Example 1. *Let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, P_0 the uniform probability distribution and $\delta = 0.5$. If we consider the permutation $\sigma = (1, 2, 3, 4)$, we obtain the extreme point P_σ given by:*

$$
\begin{aligned}
&P_\sigma(\{x_1\}) = \overline{P}(\{x_1\}) = 1.5 \cdot 0.25 = 0.375.\\
&P_\sigma(\{x_2\}) = \overline{P}(\{x_2\}) = 1.5 \cdot 0.25 = 0.375.\\
&P_\sigma(\{x_3\}) = \underline{P}(\{x_3, x_4\}) = 1.5 \cdot 0.5 - 0.5 = 0.25.\\
&P_\sigma(\{x_4\}) = 0.
\end{aligned}
$$

In fact, it can be proven that the extreme points of $\mathcal{M}(P_0, \delta)$ are given by

$$
\begin{aligned}
&P(\{x_i\}) = \overline{P}(\{x_i\}) = 0.375,\\
&P(\{x_j\}) = \overline{P}(\{x_j\}) = 0.375,\\
&P(\{x_k\}) = \underline{P}(\{x_k, x_l\}) = 0.25,\\
&P(\{x_l\}) = 0,
\end{aligned}
$$

for every possible combination of i, j, k, l in $\{1, 2, 3, 4\}$. ♦

Lemma 2 simplifies the computation of the extreme points of the credal set of a PMM. In this respect, we begin by determining the number of extreme points in a specific case:

Proposition 1. *Let P_0 denote the uniform distribution on $\{x_1, \ldots, x_n\}$ and consider $\delta > 0$.*

(1) If n is even and $\delta \in \left(\frac{n-2}{n+2}, 1\right)$, then $\mathcal{M}(P_0, \delta)$ has $\binom{n}{\frac{n}{2}} \dfrac{n}{2}$ different extreme points.

(2) If n is odd and $\delta \in \left(\frac{n-1}{n+1}, \frac{n+1}{n-1}\right)$, then $\mathcal{M}(P_0, \delta)$ has $\binom{n}{\frac{n+1}{2}} \dfrac{n+1}{2}$ different extreme points.

Proof. Let us prove the first statement; the proof of the second is analogous.

For $\delta \in \left(\frac{n-2}{n+2}, 1\right)$, given A with $|A| = \frac{n}{2}$ it holds that

$$\overline{P}(A) = \min\{(1+\delta)P(A), 1\} = \min\left\{(1+\delta)\frac{1}{2}, 1\right\} < 1,$$

because $\delta < 1$, while given A with $|A| = \frac{n}{2} + 1$:

$$\overline{P}(A) = \min\{(1+\delta)P(A), 1\} = \min\left\{(1+\delta)\frac{n+2}{2n}, 1\right\} = 1,$$

because $\delta > \frac{n-2}{n+2}$. Thus, all the sets with cardinality $\frac{n}{2}$ or less have an upper probability lower than 1, and all the sets with cardinality greater than $\frac{n}{2}$ have upper probability 1. Moreover, taking a permutation σ and its induced extreme point P_σ, since $(1+\delta)\frac{n+2}{2n} > 1$, we deduce that $P_\sigma(\{x_{\sigma(\frac{n}{2}+1)}\}) \neq P_\sigma(\{x_{\sigma(\frac{n}{2})}\})$.

As a consequence, for every two permutations σ_1, σ_2, it holds that $P_{\sigma_1} = P_{\sigma_2}$ if and only if $\sigma_1(\frac{n}{2}+1) = \sigma_2(\frac{n}{2}+1)$, and $\{\sigma_1(1), \ldots, \sigma_1(\frac{n}{2})\} = \{\sigma_2(1), \ldots, \sigma_2(\frac{n}{2})\}$. There are $\binom{n}{\frac{n}{2}}$ different ways of selecting the first $\frac{n}{2}$ elements, and for any of them there are $\frac{n}{2}$ different extreme points (as many as possibilities for choosing the element in the $\frac{n}{2}+1$-th position). Therefore, $\mathcal{M}(P_0, \delta)$ has $\frac{n}{2}\binom{n}{\frac{n}{2}}$ different extreme points. □

A similar result can be found in [34, Prop. 1 (2)], where the number of extreme points is written in terms of a parameter s: $s\binom{n}{s}$. Furthermore, taking [34, Prop. 1 (2)] and Lemma 1 into account, it can be seen that the extreme points of $\mathcal{M}(P_0, \delta)$ when P_0 is the uniform probability measure are given by:

$$P_E(\{x\}) = \begin{cases} \frac{1+\delta}{n} & \text{if } x \in E \setminus \{x^*\}, \\ \frac{1-\delta}{2} & \text{if } x = x^*, \\ 0 & \text{otherwise,} \end{cases}$$

where $E = \{x^*, x_{i_1}, \ldots, x_{i_{n/2}}\}$ is every set of $\frac{n}{2} + 1$ elements, if n is even, and:

$$P_O(\{x\}) = \begin{cases} \frac{1+\delta}{n} & \text{if } x \in O \setminus \{x^*\}, \\ 1 - \frac{(n-1)(1+\delta)}{2n} & \text{if } x = x^*, \\ 0 & \text{otherwise,} \end{cases}$$

where $O = \{x^*, x_{i_1}, \ldots, x_{i_{(n-1)/2}}\}$ is every set of $\frac{n-1}{2} + 1$ elements, if n is odd.

Next we show that the case depicted in Proposition 1 corresponds to the maximal number of extreme points associated with a PMM:

Proposition 2. *Consider a PMM (P_0, δ) on $\mathcal{X}$. The maximal number of extreme points of $\mathcal{M}(P_0, \delta)$ is:*

(1) $\frac{n}{2}\binom{n}{\frac{n}{2}}$*, if n is even.*
(2) $\frac{n+1}{2}\binom{n}{\frac{n+1}{2}}$*, if n is odd.*

Proof. Consider the case of n even; the proof for n odd is similar.

Given a permutation σ, denote $j = \min\{i = 1, \ldots, n \mid \overline{P}(\{\sigma(1), \ldots, \sigma(i)\}) = 1\}$. Then, the same extreme point P_σ is generated by $(j-1)! \cdot (n-j)!$ permutations: all those with $\{\sigma'(1), \ldots, \sigma'(j-1)\} = \{\sigma(1), \ldots, \sigma(j-1)\}$ and $\{\sigma'(j+1), \ldots, \sigma(n)\} = \{\sigma'(j+1), \ldots, \sigma'(n)\}$.

This value is lower bounded by

$$(j-1)! \cdot (n-j)! \geq \left(\frac{n}{2} - 1\right)! \cdot \left(\frac{n}{2}\right)!$$

and thus, the number of different extreme points of $\mathcal{M}(P_0,\delta)$ is bounded above by

$$\frac{n!}{\left(\frac{n}{2}-1\right)!\cdot\left(\frac{n}{2}\right)!}=\frac{n}{2}\binom{n}{\frac{n}{2}}.$$

Furthermore, by Proposition 1 this maximum is attained. □

Note that the maximal number of extreme points for n odd can equivalently be expressed by $\binom{n+1}{\frac{n+1}{2}}\frac{n+1}{4}$. As we shall see, the above formula of the maximal number of extreme points of the credal set of a PMM is related to that of probability intervals [32]. This is no coincidence: it is due to the connection between both models we shall establish in Section 4.1.

3.2. **A bound on the number of extreme points for an arbitrary PMM.** In this section, we shall establish a simple formula that provides an upper bound on the number of extreme points associated with a PMM. Let (P_0,δ) be a pari-mutuel model, and define

$$\mathcal{L}=\{A\subseteq\mathcal{X}\mid\overline{P}(A)=1\}. \tag{7}$$

This is a filter of subsets of $\mathcal{X}$, and as a consequence also a poset with respect to set inclusion.

Example 2. *Consider a four-element space $\mathcal{X}=\{x_1,x_2,x_3,x_4\}$ with probabilities 0.1, 0.1, 0.3 and 0.5, respectively, and let $\delta=0.3$. The poset $(\mathcal{L},\subseteq)$ is given by*

$$\mathcal{L}=\{\mathcal{X},\{x_2,x_3,x_4\},\{x_1,x_3,x_4\},\{x_3,x_4\}\}$$

and pictured, with the whole subset lattice, in Figure 1. ♦

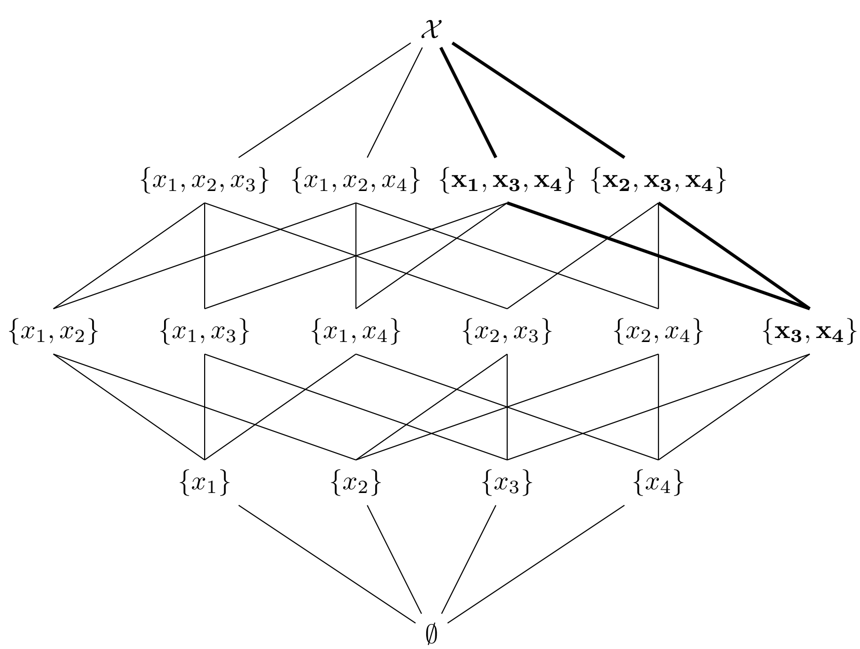


FIGURE 1. Set inclusion lattice and filter $\mathcal{L}$ (in bold) from Example 2.

We can use this filter to bound the number of extreme points of a PMM:

Proposition 3. *Consider a PMM (P_0, δ), and let $\mathcal{L}$ be given by Eq. (7). Then, the number of extreme points of $\mathcal{M}(P_0, \delta)$ is bounded above by:*

$$\sum_{A \in \mathcal{L}} \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right|. \tag{8}$$

Proof. For every $A \in \mathcal{L}$, define:

$$\mathcal{M}_A = \{P \in ext(\mathcal{M}(P_0, \delta)) \mid P(A) = 1, P(\{x\}) > 0 \ \forall x \in A\}.$$

Let us prove that $ext(\mathcal{M}(P_0, \delta)) = \cup_{A \in \mathcal{L}} \mathcal{M}_A$. On the one hand, it is obvious that $\mathcal{M}_A \subseteq ext(\mathcal{M}(P_0, \delta))$ for every $A \in \mathcal{L}$, and therefore $ext(\mathcal{M}(P_0, \delta)) \supseteq \cup_{A \in \mathcal{L}} \mathcal{M}_A$. Conversely, given $P \in ext(\mathcal{M}(P_0, \delta))$, if we define $A^* = \{x \in \mathcal{X} : P(\{x\}) > 0\}$, it holds that $P(A^*) = 1$, and $P \in \mathcal{M}_{A^*}$.

Now, if $P \in \mathcal{M}_A \cap \mathcal{M}_B$ for two different $A, B \in \mathcal{L}$, then $P(A) = P(B) = 1$ and $P(\{x\}) > 0$ for every $x \in A \cup B$. Therefore, if there exists $x \in A \backslash B$, $P(B \cup \{x\}) = P(B) + P(\{x\}) > 1$, a contradiction. Therefore, $A = B$. This means that $\{\mathcal{M}_A : A \in \mathcal{L}\}$ is a partition of $ext(\mathcal{M}(P_0, \delta))$, whence

$$|ext(\mathcal{M}(P_0, \delta))| = \sum_{A \in \mathcal{L}} |\mathcal{M}_A|.$$

We prove next that $|\mathcal{M}_A| \leq \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right|$. We consider two cases:

Case 1: Assume that $\bigcap_{B \subseteq A, B \in \mathcal{L}} B = A$. This means that every $B \subset A$ satisfies $\overline{P}(B) < 1$, as no strict subset of A is in $\mathcal{L}$. From Lemma 2, for every $P \in \mathcal{M}_A$ there exists $x_P \in A$ such that:

$$P(\{x\}) = \begin{cases} \underline{P}(\{x_P\} \cup A^c) & \text{if } x = x_P. \\ \overline{P}(\{x\}) & \text{if } x \in A \backslash \{x_P\}. \\ 0 & \text{if } x \in A^c. \end{cases} \tag{9}$$

Therefore, $|\mathcal{M}_A|$ is at most equal to the cardinality of A, so:

$$|\mathcal{M}_A| \leq |A| = \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right|.$$

Case 2: Assume now that $B^* = \bigcap_{B \subseteq A, B \in \mathcal{L}} B \subset A$. Again, from Lemma 2 we know that for every $P \in \mathcal{M}_A$ exists $x_P \in A$ such that Eq. (9) holds. Let us see that x_P should belong to B^*. By contradiction, assume that $x_P \in A \backslash B^*$. This implies that there exists $B \in \mathcal{L}$ with $B \subset A$ such that $x_P \notin B$. In particular, we can take $B = A \backslash \{x_P\}$. Then, the probability P satisfies:

$$P(B) = \sum_{x \in B} \overline{P}(\{x\}) \geq \overline{P}(B) = 1,$$

where the inequality follows from the super-additivity of the coherent upper probability $\overline{P}$. Thus, $P(B) = 1$, and as a consequence $P(\{x_P\}) = 0$, a contradiction with $P \in \mathcal{M}_A$.

We conclude that $x_P \in B^*$. Therefore, the cardinality of $\mathcal{M}_A$ is at most the number of elements in B^*. Equivalently:

$$|\mathcal{M}_A| \leq |B^*| = \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right|. \qquad \square$$

Example 3. *Consider again Example 2. Using Proposition 3, the maximal number of extreme points of $\mathcal{M}(P_0, \delta)$ is bounded by Eq. (8). Let us compute this value; for every $A \in \mathcal{L}$, it holds:*

$$\left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right| = \left|\{x_3, x_4\}\right| = 2.$$

Therefore, the number of extreme points of $\mathcal{M}(P_0, \delta)$ is bounded by:

$$\sum_{A \in \mathcal{L}} \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right| = \left|\{x_3, x_4\}\right| + \left|\{x_3, x_4\}\right| + \left|\{x_3, x_4\}\right| + \left|\{x_3, x_4\}\right| = 2+2+2+2 = 8.$$

In fact, in this case this formula provides not only an upper bound but the exact number of extreme points, since

$$\begin{aligned}
\mathcal{M}_{\{x_3,x_4\}} &= \{(0, 0, 0.39, 0.61), (0, 0, 0.35, 0.65)\}.\\
\mathcal{M}_{\{x_1,x_3,x_4\}} &= \{(0.13, 0, 0.39, 0.48), (0.13, 0, 0.22, 0.65)\}.\\
\mathcal{M}_{\{x_2,x_3,x_4\}} &= \{(0, 0.13, 0.39, 0.48), (0, 0.13, 0.22, 0.65)\}.\\
\mathcal{M}_{\mathcal{X}} &= \{(0.13, 0.13, 0.39, 0.35), (0.13, 0.13, 0.09, 0.65)\}. \blacklozenge
\end{aligned}$$

Let us also show that the bound given in Proposition 3 is not always tight.

Example 4. *Consider $\mathcal{X} = \{x_1, x_2, x_3\}$, the uniform distribution P_0 on $\mathcal{X}$ and $\delta = 0.5$. It holds that:*

$$\mathcal{L} = \{\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, \mathcal{X}\}.$$

Using the previous result, the number of extreme points is bounded above by

$$\sum_{A \in \mathcal{L}} \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right| = \left|\{x_1, x_2\}\right| + \left|\{x_1, x_3\}\right| + \left|\{x_2, x_3\}\right| + \left|\emptyset\right| = 2 + 2 + 2 + 0 = 6.$$

However, the extreme points are:

$$\mathcal{M}_{\{x_1,x_2\}} = \{(0.5, 0.5, 0)\}, \quad \mathcal{M}_{\{x_1,x_3\}} = \{(0.5, 0, 0.5)\}, \quad \mathcal{M}_{\{x_2,x_3\}} = \{(0, 0.5, 0.5)\}.$$

Thus, there are only 3 different extreme points, half of the upper bound given in the proposition. $\blacklozenge$

Note that $A \in \mathcal{L}$ if and only if $P_0(A) \geq \frac{1}{1+\delta}$. We can characterize the tightness of the bound given in Eq. (8) by means of the strict inequality in this expression.

Proposition 4. *Consider a PMM (P_0, δ), and let $\mathcal{L}$ be given by Eq. (7). The number of extreme points of $\mathcal{M}(P_0, \delta)$ coincides with the bound determined by Eq. (8) if and only if $P_0(A) > \frac{1}{1+\delta}$ for every $A \in \mathcal{L}$.*

Proof. From the proof of Proposition 3, the bound given by Eq. (8) is tight if and only if, for every set $A \in \mathcal{L}$,

$$|\mathcal{M}_A| = \left| \bigcap_{B \subseteq A, B \in \mathcal{L}} B \right|. \tag{10}$$

Consider then $A \in \mathcal{L}$, and assume that $|A| > 1$ (the case of $|A| = 1$ is trivial). Let $B^* = \bigcap_{B \subseteq A, B \in \mathcal{L}} B \subseteq A$; then the proof of Proposition 3 shows that the elements of $\mathcal{M}_A$ are given by Eq. (9), for $x_P \in B^*$. Thus, Eq. (10) holds only if these extreme points are different for every $z \in B^*$. This is equivalent to $\underline{P}(\{z\} \cup A^c) \neq \overline{P}(\{z\})$ for every $z \in B^*$, because in that case $P(\{x_P\}) \neq \overline{P}(\{x_P\})$ for every $P \in \mathcal{M}_A$. If $|B^*| = 1$, then trivially $|\mathcal{M}_A| = 1 = |B^*|$. On the other hand, if $|B^*| > 1$, the fact that $\{z\}, A \backslash \{z\} \notin \mathcal{L}$ implies that $\overline{P}(\{z\}) < 1$ and $\underline{P}(\{z\} \cup A^c) = 1 - \overline{P}(A \backslash \{z\}) > 0$. Applying Eq. (1), we have

$$\underline{P}(\{z\} \cup A^c) = (1+\delta) P_0(\{z\} \cup A^c) - \delta$$

and

$$\overline{P}(\{z\}) = (1+\delta) P_0(\{z\}).$$

Thus, $\underline{P}(\{z\} \cup A^c) \neq \overline{P}(\{z\})$ if and only if $(1+\delta) P_0(A^c) - \delta \neq 0$. By conjugacy, $\underline{P}(A^c) = 1 - \overline{P}(A) = 0$, whence $(1+\delta) P_0(A^c) - \delta \leq 0$. Then, $\underline{P}(\{z\} \cup A^c) \neq \overline{P}(\{z\})$ is equivalent to $P_0(A^c) < \frac{\delta}{1+\delta}$, or, in other words, to $P_0(A) > \frac{1}{1+\delta}$. □

Indeed, if we go back to Example 4 we see that in that case it holds that $P_0(A) = \frac{2}{3} = \frac{1}{1+\delta}$ for $A = \{x_1, x_2\} \in \mathcal{L}$, meaning that the bound in Eq. (8) is not tight.

4. Connection with probability intervals and belief functions

In this section, we study the connection between the PMM and other relevant imprecise probability models. In particular, we show that PMMs in a finite setting are particular instances of probability intervals, and study the conditions for a PMM to induce a belief function and a minitive lower probability, respectively.

4.1. **Probability intervals.**

Given $\mathcal{X} = \{x_1, \ldots, x_n\}$, a *probability interval* [9, 32] on $\mathcal{P}(X)$ is a lower probability defined on the singletons and their complements. A probability interval can thus be represented by a n-tuple of intervals:

$$\mathcal{I} = \{[l_i, u_i] : i = 1, \ldots, n\}, \tag{11}$$

where it is assumed that $l_i \leq u_i$ and where $[l_i, u_i]$ means that the unknown or imprecisely specified probability of x_i belongs to the interval $[l_i, u_i]$. A probability interval determines a credal set by:

$$\mathcal{M}(\mathcal{I}) = \{P \text{ probability} \mid l_i \leq P(\{x_i\}) \leq u_i, \quad i = 1, \ldots, n\}, \tag{12}$$

and the lower and upper envelopes of $\mathcal{M}(\mathcal{I})$ determine coherent lower and upper probabilities by:

$$l(A) = \inf_{P \in \mathcal{M}(\mathcal{I})} P(A) \text{ and } u(A) = \sup_{P \in \mathcal{M}(\mathcal{I})} P(A) \quad \forall A \subseteq \mathcal{X}. \tag{13}$$

The probability interval $\mathcal{I}$ is called *proper* when its associated credal set $\mathcal{M}(\mathcal{I})$ is non-empty. This holds when:

$$\sum_{i=1}^{n} l_i \leq 1 \leq \sum_{i=1}^{n} u_i. \tag{14}$$

Furthermore, a probability interval $\mathcal{I}$ is called *reachable* (i.e., coherent in Walley's terminology [37], or an *atomic model* in the sense of [17]) whenever the functionals l, u determined by Eq. (13) determine the intervals from Eq. (11) when restricted to singletons, i.e., when $l(\{x_i\}) = l_i$ and $u(\{x_i\}) = u_i$ for all $i = 1, \dots, n$. This is equivalent to the following inequalities:

$$\sum_{j \neq i} l_j + u_i \leq 1 \text{ and } \sum_{j \neq i} u_j + l_i \geq 1 \ \forall i = 1, \dots, n. \tag{15}$$

When $\mathcal{I}$ is a reachable probability interval, $l(A)$ and $u(A)$ can be computed by:

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

for every $A \subseteq \mathcal{X}$. For a detailed study on probability intervals, we refer to [9]. See also [3, 16, 29, 30] for other relevant works on this topic.

Our next result shows that the PMM is a particular case of a reachable probability interval.

Theorem 1. *Consider a PMM (P_0, δ), and define the probability interval $\mathcal{I} = \{[l_i, u_i] : i = 1, \dots, n\}$ by:*

$$l_i = \underline{P}(\{x_i\}) \text{ and } u_i = \overline{P}(\{x_i\}),$$

where $\underline{P}, \overline{P}$ are given by Eq. (1)*. Then, if we denote by $\mathcal{M}(\mathcal{I})$ the credal set associated with $\mathcal{I}$ by means of Eq.* (12)*, it holds that:*

(1) *The probability interval $\mathcal{I} = \{[l_i, u_i] : i = 1, \dots, n\}$ is reachable.*
(2) *$\mathcal{M}(\mathcal{I}) = \mathcal{M}(P_0, \delta)$, or equivalently, $\underline{P}(A) = l(A)$ and $\overline{P}(A) = u(A)$ for every $A \subseteq \mathcal{X}$.*

Proof. First of all, let us see that $\mathcal{I} = \{[l_i, u_i] : i = 1, \dots, n\}$ satisfies Eq. (15):

$$\sum_{j \neq i} l_j + u_i = \sum_{j \neq i} \underline{P}(\{x_j\}) + \overline{P}(\{x_i\}) \leq \underline{P}(\{x_i\}^c) + \overline{P}(\{x_i\}) = 1;$$
$$\sum_{j \neq i} u_j + l_i = \sum_{j \neq i} \overline{P}(\{x_j\}) + \underline{P}(\{x_i\}) \geq \overline{P}(\{x_i\}^c) + \underline{P}(\{x_i\}) = 1,$$

taking into account that $\underline{P}$ is super-additive and $\overline{P}$ is sub-additive.

Let us now see that $\underline{P}(A) = l(A)$ for every $A \subseteq \mathcal{X}$. On the one hand, assume that $\underline{P}(A) = 0$, whence $\overline{P}(A^c) = 1$. Then $\underline{P}(\{x\}) = 0$ for every $x \in A$, and therefore $l(\{x\}) = 0$ for every $x \in A$. By definition:

$$\begin{aligned} l(A) &= \max \Big\{ \sum_{x \in A} l(\{x\}), 1 - \sum_{x \notin A} u(\{x\}) \Big\} \\ &= 1 - \sum_{x \notin A} u(\{x\}) = 1 - \sum_{x \notin A} \overline{P}(\{x\}) \leq 1 - \overline{P}(A^c) = 0. \end{aligned}$$

On the other hand, if $\underline{P}(A) = (1 + \delta) P_0(A) - \delta > 0$, then

$$\begin{aligned} l(A) = \max \Big\{ \sum_{x \in A} l(\{x\}), 1 - \sum_{x \notin A} u(\{x\}) \Big\} &= \max \Big\{ \sum_{x \in A} \underline{P}(\{x\}), 1 - \sum_{x \notin A} \overline{P}(\{x\}) \Big\} \\ = \max \Big\{ \sum_{x \in A} \underline{P}(\{x\}), 1 - \overline{P}(A^c) \Big\} &= \max \Big\{ \sum_{x \in A} \underline{P}(\{x\}), \underline{P}(A) \Big\} = \underline{P}(A), \end{aligned}$$

where the third and fifth equalities follow form Lemma 1 and the supper-additivity of $\underline{P}$, respectively.

By conjugacy of $\underline{P}, \overline{P}$ and l, u, we can simply see that $\overline{P}(A) = u(A)$ for every A:

$$\overline{P}(A) = 1 - \underline{P}(A^c) = 1 - l(A^c) = u(A).$$

Therefore, $l = \underline{P}$ and $u = \overline{P}$. □

On the other hand, the class of reachable probability intervals is larger than that of PMM, in the sense that not every reachable probability interval can also be expressed in terms of a PMM:

Example 5. *Consider the four-element space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and the probability interval $\mathcal{I} = \{[l_i, u_i] : i = 1, \ldots, 4\}$ given by:*

	x_1	x_2	x_3	x_4
l_i	0.2	0.1	0.3	0.2
u_i	0.4	0.2	0.5	0.4

which can be shown to be reachable using Eq. (15).

To see that $\mathcal{I}$ is not representable by a PMM (P_0, δ), note that by Remark 1, given a PMM every set A such that $0 < \underline{P}(A) < \overline{P}(A) < 1$ satisfies $\overline{P}(A) - \underline{P}(A) = \delta$. However, in this example we obtain:

$$0 < l(\{x_1\}) = l_1 = 0.2 < 0.4 = u_1 = u(\{x_1\}) < 1 \text{ and}$$
$$0 < l(\{x_2\}) = l_2 = 0.1 < 0.2 = u_2 = u(\{x_2\}) < 1,$$

while

$$u(\{x_1\}) - l(\{x_1\}) = 0.2 \text{ and } u(\{x_2\}) - l(\{x_2\}) = 0.1;$$

thus, the difference is not constant, and therefore l, u cannot be represented by means of a PMM. ♦

In particular, we deduce from Corollary 1 that a PMM will be a probability interval where one only specifies the upper probability bounds on the singletons (the lower bounds following then from (13), for instance). Thus, any property satisfied by probability intervals is also satisfied by the PMMs. Interestingly, the maximal number of extreme points for the credal set of a PMM, established in Theorem 2, coincides with the maximal number of extreme points for a probability interval, given in [32][3].

4.2. **Belief functions.** As we mentioned in Section 2, the lower probability of a PMM is 2-monotone. In this section we consider a stronger notion that extends 2-monotonicity, called *complete monotonicity.* A lower probability $\underline{P}$ is completely monotone if that for every $p \in \mathbb{N}$ and every sets $A_1, \ldots, A_p \subseteq \mathcal{X}$,

$$\underline{P}(\cup_{i=1}^{p} A_i) \geq \sum_{J \subseteq \{1,\ldots,n\}} (-1)^{|J|-1} \underline{P}(\cap_{i \in J} A_i). \tag{16}$$

A lower probability satisfying the property of complete monotonicity is also called a belief function in Dempster-Shafer Theory. Its conjugate upper probability is called a plausibility function.

[3]Note that there is a misprint when reporting this number in [9].

Belief functions [28] are determined by their Möbius inverse $m : \mathcal{P}(\mathcal{X}) \to [0,1]$, that is a mass function on the subsets of $\mathcal{X}$, by means of the formula

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

The sets $A \subseteq \mathcal{X}$ such that $m(A) > 0$ are called the *focal elements* of $\underline{P}$. A lower probability satisfying Eq. (16) for every $p \leq k$ is called *k-monotone*.

Conversely, the Möbius inverse m of a lower probability $\underline{P}$ is determined by the formula

$$m(B) = \sum_{A \subseteq B} (-1)^{|B \setminus A|} \underline{P}(A), \tag{17}$$

and $\underline{P}$ is a belief function if and only if the function m given by Eq. (17) satisfies $m(A) \geq 0$ for every $A \subseteq \mathcal{X}$.

In [3, Thm.3.1], a sufficient condition for a probability interval to be a completely monotone model was established; see also [9, Sect. 6] and [18]. In this section, we shall establish necessary and sufficient conditions for the particular types of probability intervals associated with PMMs. By doing so, we shall also show that the sufficient condition in [3] is not necessary.

We start with a simple result that implies that the PMM is not 3-monotone in general.

Proposition 5. *Let $\underline{P}$ be lower probability associated with a PMM (P_0, δ), with $|\mathcal{X}| \geq 3$. If there are different x_i, x_j, x_k such that $\underline{P}(\{x_i\}), \underline{P}(\{x_j\}), \underline{P}(\{x_k\}) > 0$, then $\underline{P}$ is not 3-monotone.*

Proof. Take $A_1 = \{x_i, x_j\}$, $A_2 = \{x_i, x_k\}$ and $A_3 = \{x_j, x_k\}$. The monotonicity of $\underline{P}$ implies that $\underline{P}(A_1), \underline{P}(A_2), \underline{P}(A_3) > 0$. On the one hand,

$$\underline{P}(A_1 \cup A_2 \cup A_3) = \underline{P}(\{x_i, x_j, x_k\}) = (1+\delta) P_0(\{x_i, x_j, x_k\}) - \delta,$$

while

$$\underline{P}(A_1) + \underline{P}(A_2) + \underline{P}(A_3) - \Big(\underline{P}(A_1 \cap A_2) + \underline{P}(A_1 \cap A_3) + \underline{P}(A_2 \cap A_3) \Big) + \underline{P}(A_1 \cap A_2 \cap A_3)$$

is equal to

$$\begin{aligned}
&\underline{P}(\{x_i, x_j\}) + \underline{P}(\{x_i, x_k\}) + \underline{P}(\{x_j, x_k\}) - \Big(\underline{P}(\{x_i\}) + \underline{P}(\{x_j\}) + \underline{P}(\{x_k\}) \Big) + \underline{P}(\emptyset) \\
&= (1+\delta) \big[P_0(\{x_i, x_j\}) + P_0(\{x_i, x_k\}) + P_0(\{x_j, x_k\}) \\
&\qquad - P_0(\{x_i\}) - P_0(\{x_j\}) - P_0(\{x_k\}) \big] \\
&= (1+\delta) P_0(\{x_i, x_j, x_k\}).
\end{aligned}$$

Now, if we compare both expressions, we obtain

$$(1+\delta) P_0(\{x_i, x_j, x_k\}) - \delta - (1+\delta) P_0(\{x_i, x_j, x_k\}) = -\delta < 0,$$

and therefore $\underline{P}$ is not 3-monotone. ☐

To see that the hypotheses of this proposition may be satisfied, let P_0 be the uniform distribution on $\{x_1, x_2, x_3\}$ and take $\delta = \frac{1}{3}$: then it follows from Eq. (1) that $\underline{P}(\{x_1\}) = \underline{P}(\{x_2\}) = \underline{P}(\{x_3\}) = \frac{1}{9}$.

In the remainder of this section we shall establish necessary and sufficient conditions for the lower probability of a PMM to be completely monotone. For this aim we define the *non-vacuity index* of a PMM as

$$k = \min\{|A| : \underline{P}(A) > 0\}. \tag{18}$$

We next give sufficient conditions to ensure completely monotonicity in terms of this index. Our first two results are quite simple and they correspond to the cases of $k = n$ or $k = n - 1$.

Proposition 6. *Let (P_0, δ) be a PMM, let $\underline{P}$ be its associated lower probability and let k be its non-vacuity index, given by Eq.(18). If $k = n$, then $\underline{P}$ is a belief function whose only focal set is $\mathcal{X}$ with mass 1.*

Proof. If $k = n$, this means that $\underline{P}(\mathcal{X}) = 1$ and $\underline{P}(A) = 0$ for every $A \subset \mathcal{X}$. This is the belief function associated with the basic probability assignment m given by $m(A) = 0$ for every $A \subset \mathcal{X}$ and $m(\mathcal{X}) = \underline{P}(\mathcal{X}) = 1$. □

In this case, by conjugacy, we obtain that for every non-empty $A \subset \mathcal{X}$, $\overline{P}(A) = 1 - \underline{P}(A^c) = 1$. This situation arises when $P_0(A) < 1$ and $\delta \geq \frac{P_0(A)}{P_0(A^c)}$ for every $A \subset \mathcal{X}$. It corresponds to the so-called *vacuous* model.

Next we consider the case of $k = n - 1$.

Proposition 7. *Let (P_0, δ) be a PMM, let $\underline{P}$ be its associated lower probability and let k be its non-vacuity index, given by Eq.(18). If $k = n - 1$ and*

$$\sum_{i=1}^{n} \underline{P}(\mathcal{X} \backslash \{x_i\}) \leq 1,$$

then $\underline{P}$ is a belief function.

Proof. By Eq. (17), we know that $m(A) = 0$ for every A such that $|A| < n - 1$. Now, for every $i = 1, \ldots, n$, $m(\mathcal{X} \backslash \{x_i\}) = \underline{P}(\mathcal{X} \backslash \{x_i\})$. Moreover,

$$m(\mathcal{X}) = \underline{P}(\mathcal{X}) - \sum_{i=1}^{n} \underline{P}(\mathcal{X} \backslash \{x_i\}) = 1 - \sum_{i=1}^{n} \underline{P}(\mathcal{X} \backslash \{x_i\}) \geq 0,$$

because by hypothesis $\sum_{i=1}^{n} \underline{P}(\mathcal{X} \backslash \{x_i\}) \leq 1$. Therefore, $m(A) \geq 0$ for every A, and as a consequence $\underline{P}$ is a belief function. □

Finally, we give two sufficient conditions for $\underline{P}$ to be a belief function when its non-vacuity index k is smaller than $n - 1$.

Proposition 8. *Let (P_0, δ) be PMM and let $\underline{P}$ be its associated lower probability. Assume there exists B such that $\underline{P}(A) > 0$ if and only if $B \subseteq A$, and assume that the non-vacuity index $k = |B|$ satisfies $k < n - 1$. Then, $\underline{P}$ is a belief function whose focal sets are B and $B \cup \{x\}$, for $x \notin B$, with respective masses:*

$$m(B) = \underline{P}(B) = (1 + \delta)P_0(B) - \delta \text{ and } m(B \cup \{x\}) = (1 + \delta)P_0(\{x\}).$$

Proof. First of all, use Eq. (17) to compute $m(B)$ and $m(B \cup \{x\})$ for $x \notin B$:

$$\begin{aligned} m(B) &= \underline{P}(B) = (1 + \delta)P_0(B) - \delta > 0. \\ m(B \cup \{x\}) &= \underline{P}(B \cup \{x\}) - \underline{P}(B) = (1 + \delta)P_0(B \cup \{x\}) - \delta - (1 + \delta)P_0(B) + \delta \\ &= (1 + \delta)P_0(\{x\}) > 0. \end{aligned}$$

Let us prove that $m(A) = 0$ for every set different from B, $B \cup \{x\}$. First of all, if $B \not\subseteq A$, it follows from Eq. (17) that $m(A) = 0$. Take now $A' \subseteq B^c$ such that $|A'| = j > 1$, and let us see that $m(A' \cup B) = 0$.

$$\begin{aligned} m(A' \cup B) &= \sum_{C \subseteq A' \cup B} (-1)^{|A' \cup B \setminus C|} \underline{P}(C) = \sum_{C' \subseteq A'} (-1)^{|A' \setminus C'|} \underline{P}(C' \cup B) \\ &= \sum_{i=0}^{j} \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i} \underline{P}(C' \cup B). \end{aligned}$$

Let us compute the value of the last term for every fixed $i \in \{0, \dots, j\}$. We start with the case $i = 0$:

$$\sum_{C' \subseteq A', |C'|=0} (-1)^j \underline{P}(C' \cup B) = (-1)^j \underline{P}(B) = (-1)^j \left((1+\delta)P_0(B) - \delta\right), \tag{19}$$

where the first equality follows because the only set $C' \subseteq A'$ with cardinality 0 is the empty set.

Consider now $i = 1, \dots, j$:

$$\begin{aligned} \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i} \underline{P}(C' \cup B) &= \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i} \left((1+\delta)P_0(B \cup C') - \delta\right) \\ &= \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i} \left((1+\delta)P_0(B) - \delta\right) + \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i}(1+\delta)P_0(C') \\ &= (-1)^{j-i} \binom{j}{i} \left((1+\delta)P_0(B) - \delta\right) + \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i}(1+\delta)P_0(C'). \end{aligned} \tag{20}$$

In the last equality we have taken into account that in the first sum, the term $(-1)^{j-i}((1+\delta)P_0(B) - \delta)$ does not depend on the sets C', so we sum this element as many times as sets C' of cardinality i are included in A', that is, the binomial coefficient $\binom{j}{i}$.

With respect to the second term, every element $x \in A'$ can be included in $\binom{j-1}{i-1}$ different sets $C' \subseteq A'$ of cardinality i. Therefore, Eq. (20) becomes:

$$\begin{aligned} &(-1)^{j-i} \binom{j}{i} \left((1+\delta)P_0(B) - \delta\right) + (-1)^{j-i}(1+\delta)\binom{j-1}{i-1} \sum_{x \in A'} P_0(\{x\}) \\ &\quad = (-1)^{j-i} \binom{j}{i} \left((1+\delta)P_0(B) - \delta\right) + (-1)^{j-i}(1+\delta)\binom{j-1}{i-1} P_0(A'). \end{aligned} \tag{21}$$

Using Eqs. (19) and (21) we obtain the value of $m(A' \cup B)$. It holds that:

$$
\begin{aligned}
m(A' \cup B) &= \sum_{i=0}^{j} \sum_{C' \subseteq A', |C'|=i} (-1)^{j-i} \underline{P}(C' \cup B) \\
&= (-1)^j \left((1+\delta)P_0(B) - \delta\right) + \sum_{i=1}^{j} \Bigg((-1)^{j-i} \binom{j}{i} \left((1+\delta)P_0(B) - \delta\right) \\
&\qquad + (-1)^{j-i}(1+\delta)\binom{j-1}{i-1} P_0(A') \Bigg) \\
&= \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \left((1+\delta)P_0(B) - \delta\right) + \sum_{i=1}^{j} (-1)^{j-i}(1+\delta)\binom{j-1}{i-1} P_0(A') \\
&= \left((1+\delta)P_0(B) - \delta\right) \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} + P_0(A')(1+\delta) \sum_{i=1}^{j} (-1)^{j-i} \binom{j-1}{i-1} \\
&= \left((1+\delta)P_0(B) - \delta\right) \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} + P_0(A')(1+\delta) \sum_{i=0}^{j-1} (-1)^{j-i-1} \binom{j-1}{i} \\
&= 0,
\end{aligned}
$$

where the last equality follows from the well known property of binomial coefficients:

$$
(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k, \tag{22}
$$

by taking $y = 1, x = -1, k = i, n = j$ and $y = 1, x = -1, k = i, n = j - 1$, respectively.

We conclude that $m(A) \geq 0$ for every A, whence $\underline{P}$ is a belief function; its focal sets are B and $B \cup \{x\}$ for every $x \notin B$. □

Proposition 9. *Let (P_0, δ) be PMM and $\underline{P}$ be the lower probability it induces by Eq. (1). Assume there is B such that $\underline{P}(A) > 0$ if and only if $B \subset A$ and that $\delta = \frac{P_0(B)}{1-P_0(B)}$, and assume that the non-vacuity index $k = |B|+1$ satisfies $k < n-1$. Then, $\underline{P}$ is a belief function whose focal sets are $B \cup \{x\}$ for every $x \notin B$ with masses:*

$$
m(B \cup \{x\}) = \frac{P_0(\{x\})}{1 - P_0(B)}.
$$

Proof. First of all, let us compute $m(B)$ and $m(B \cup \{x\})$ for $x \notin B$:

$$
\begin{aligned}
m(B) &= \underline{P}(B) = (1+\delta)P_0(B) - \delta = 0. \\
m(B \cup \{x\}) &= \underline{P}(B \cup \{x\}) - \underline{P}(B) = (1+\delta)P_0(B \cup \{x\}) - \delta = (1+\delta)P_0(\{x\}) \\
&= P_0(\{x\}) \left(1 + \frac{P_0(B)}{1 - P_0(B)}\right) = P_0(\{x\}) \frac{1 - P_0(B) + P_0(B)}{1 - P_0(B)} \\
&= \frac{P_0(\{x\})}{1 - P_0(B)}.
\end{aligned}
$$

Note that:

$$\sum_{x\notin B} m(B\cup\{x\}) = \sum_{x\notin B} \frac{P_0(\{x\})}{1-P_0(B)} = \frac{1}{1-P_0(B)}\sum_{x\notin B} P_0(\{x\})$$
$$= \frac{P_0(B^c)}{1-P_0(B)} = \frac{1-P_0(B)}{1-P_0(B)} = 1.$$

Let us see that $m(A)=0$ for every set different from B, $B\cup\{x\}$. First of all, if $B \not\subseteq A$, it follows from Eq. (17) that $m(B)=0$. Take now $A'\subseteq B^c$ with $|A'|=j>1$, and let us see that $m(A\cup B)=0$.

$$\begin{aligned}
m(A'\cup B) &= \sum_{C\subseteq A'\cup B} (-1)^{|A'\cup B\setminus C|}\underline{P}(C) = \sum_{C'\subseteq A'} (-1)^{|A\setminus C'|}\underline{P}(C'\cup B)\\
&= \sum_{i=0}^{j}\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}\underline{P}(C'\cup B).\\
&= \sum_{i=0}^{j}\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}\left((1+\delta)P_0(B\cup C')-\delta\right)\\
&= \sum_{i=0}^{j}\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}\left((1+\delta)P_0(B)-\delta\right)\\
&\quad + \sum_{i=0}^{j}\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}P_0(C')(1+\delta)\\
&= \sum_{i=0}^{j}\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}P_0(C')(1+\delta),
\end{aligned}$$

where last equality follows because $(1+\delta)P_0(B)-\delta=0$. Let us now analyze last expression for every $i=0,\dots,j$. First of all, for $i=0$, it holds that:

$$\sum_{C'\subseteq A,|C'|=0} (-1)^{j}P_0(C')(1+\delta) = P_0(\emptyset)(1+\delta) = 0.$$

For every $i=1,\dots,j$, we proceed as in the previous proof. Every $x\in A'$ can be included in exactly $\binom{j-1}{i-1}$ different sets $C'\subseteq A'$ of cardinality i, whence

$$\begin{aligned}
\sum_{C'\subseteq A',|C'|=i} (-1)^{j-i}P_0(C')(1+\delta) &= (-1)^{j-i}(1+\delta)\binom{j-1}{i-1}\sum_{x\in A'} P_0(\{x\})\\
&= (-1)^{j-i}(1+\delta)\binom{j-1}{i-1}P_0(A').
\end{aligned}$$

We deduce that:

$$\begin{aligned} m(A' \cup B) &= \sum_{i=1}^{j} (-1)^{j-i}(1+\delta)\binom{j-1}{i-1} P_0(A') \\ &= (1+\delta)P_0(A') \sum_{i=1}^{j} (-1)^{j-i} \binom{j-1}{i-1} \\ &= (1+\delta)P_0(A') \sum_{i=0}^{j-1} (-1)^{j-i-1} \binom{j-1}{i} = 0, \end{aligned}$$

where the last inequality follows again from the property of binomial coefficients described in Eq. (22) using $y=-1, x=1, k=i, n=j-1$.

We conclude that $m(A) \geq 0$ for every set A, whence $\underline{P}$ is a belief function. In addition, we have proven that the focal sets are $B \cup \{x\}$ for every $x \notin B$. □

We have established four sufficient conditions for the lower probability associated with a pari-mutuel model to be completely monotone. Next we show that these conditions are also necessary.

Theorem 2. *Consider a PMM (P_0, δ), and let $\underline{P}$ be its associated lower probability. $\underline{P}$ is a belief function if and only if one of the following conditions is satisfied:*

(B1) $k = n$.
(B2) $k = n-1$ *and* $\sum_{i=1}^{n} \underline{P}(\mathcal{X} \backslash \{x_i\}) \leq 1$.
(B3) $k < n-1$, *there exists a unique* B *with* $|B| = k$ *and* $\underline{P}(B) > 0$, *and* $\underline{P}(A) > 0$ *if and only if* $B \subseteq A$.
(B4) $k < n-1$, *there exists a unique* B *with* $|B| = k-1$ *and* $\delta = \frac{P_0(B)}{1-P_0(B)}$, *and* $\underline{P}(A) > 0$ *if and only if* $B \subset A$.

Proof. Sufficiency of these four conditions has been proven in Propositions 6–9. Conversely, let us see that if $\underline{P}$ is a belief function, one of the conditions must be satisfied.

First of all, if $k = n$, then $\underline{P}(A) = 0$ for every $A \subset \mathcal{X}$ and $\underline{P}(\mathcal{X}) = 1$. Therefore, $m(\mathcal{X}) = 1$ and $m(A) = 0$ for every $A \subset \mathcal{X}$, and as a consequence we are in case (B1).

Secondly, if $k = n-1$, this means that $\underline{P}(A) = 0$ for every A such that $|A| < n-1$. Furthermore, $m(\mathcal{X} \backslash \{x\}) = \underline{P}(\mathcal{X} \backslash \{x\})$. Since the sum of all the masses must be 1,

$$1 = \sum_{A \subseteq \mathcal{X}} m(A) = \sum_{x \in \mathcal{X}} m(\mathcal{X} \backslash \{x\}) + m(\mathcal{X}),$$

whence $\sum_{x \in \mathcal{X}} m(\mathcal{X} \backslash \{x\}) \leq 1$, so we are in case (B2).

In order to simplify the notation in the remainder of the proof, we shall assume without loss of generality that the elements in $\mathcal{X}$ are ordered so that

$$P_0(\{x_1\}) \geq P_0(\{x_2\}) \geq \ldots \geq P_0(\{x_n\}), \tag{23}$$

and denote $p_i = P_0(\{x_i\})$ for every $i = 1, \ldots, n$.

Assume that $k < n-1$ and that there is only one set B of cardinality k with $\underline{P}(B) > 0$. From Eq. (23), $B = \{x_1, \ldots, x_k\}$. By definition of $\underline{P}$, we obtain that:

- $m(B) = \underline{P}(B) = (1+\delta)(p_1 + \cdots + p_k) - \delta$.
- $\forall j = k+1, \ldots, n$, $m(B \cup \{x_j\}) = \underline{P}(B \cup \{x_j\}) - \underline{P}(B) = (1+\delta)p_j$.

As a consequence,

$$m(B)+\sum_{j=k+1}^{n} m(B\cup\{x_j\}) = (1+\delta)(p_1+\cdots+p_k)-\delta+\sum_{j=k+1}^{n}(1+\delta)p_j = (1+\delta)-\delta = 1.$$

Since $m(A) \geq 0$ because $\underline{P}$ is a belief function, we deduce that the only focal elements of $\underline{P}$ are the sets $B, B\cup\{x_j\}, j = k+1,\ldots,n$. As a consequence, we are in case (B3).

Finally, consider that $k < n-1$ and there are two different sets A_1, A_2 of cardinality k such that $\underline{P}(A_1), \underline{P}(A_2) > 0$. By Eq. (23), we can assume that $A_1 = \{x_1,\ldots,x_{k-1},x_k\}$ and $A_2 = \{x_1,\ldots,x_{k-1},x_{k+1}\}$.

Denote $C = \{x_1,\ldots,x_{k+2}\} = A_1\cup A_2\cup\{x_{k+2}\}$. Taking into account that $\underline{P}$ is a belief function and using Eq. (17),

$$0 \leq m(C\setminus\{x_i\}) = \underline{P}(C\setminus\{x_i\})-\sum_{j\neq i\in\{1,\ldots,k+2\}}\underline{P}(C\setminus\{x_i,x_j\})\ \forall i = 1,\ldots,k-1. \quad (24)$$

On the other hand, we also have that

$$\begin{aligned}0 \leq m(C) =&\underline{P}(C)-\sum_{i=1}^{k+2}\underline{P}(C\setminus\{x_i\})+\sum_{i\neq j\in\{1,\ldots,k+2\}}\underline{P}(C\setminus\{x_i,x_j\})\\ =&\underline{P}(C)-\underline{P}(C\setminus\{x_{k+2}\})-\underline{P}(C\setminus\{x_{k+1}\})-\underline{P}(C\setminus\{x_k\})\\ &+\underline{P}(\{x_1,\ldots,x_{k-1},x_k\})+\underline{P}(\{x_1,\ldots,x_{k-1},x_{k+1}\})\\ &-\sum_{i=1}^{k-1}\underline{P}(C\setminus\{x_i\})+\sum_{I}\underline{P}(C\setminus\{x_i,x_j\}),\end{aligned}$$

where $I = \{(i,j) : i < j, i,j \in \{1,\ldots,k+2\}\}\setminus\{(k+1,k+2),(k,k+2)\}$.

Applying the definition of $\underline{P}$, this is equal to

$$\begin{aligned}&-(1+\delta)p_{k+2}-\sum_{i=1}^{k-1}\underline{P}(C\setminus\{x_i\})+\sum_{I}\underline{P}(C\setminus\{x_i,x_j\})\\ &=-(1+\delta)p_{k+2}-\sum_{i=1}^{k-1}\left[\underline{P}(C\setminus\{x_i\})-\sum_{j\neq i}\underline{P}(C\setminus\{x_i,x_j\})\right]+\underline{P}(C\setminus\{x_k,x_{k+1}\})\\ &\leq-(1+\delta)p_{k+2}+\underline{P}(C\setminus\{x_k,x_{k+1}\}),\end{aligned}$$

where the inequality follows from Eq. (24).

From this we deduce that $\underline{P}(C\setminus\{x_k,x_{k+1}\}) > 0$, whence

$$0 \leq -(1+\delta)p_{k+2}+\underline{P}(C\setminus\{x_k,x_{k+1}\}) = (1+\delta)(p_1+\cdots+p_{k-1})-\delta.$$

Since on the other hand $\underline{P}(\{x_1,\ldots,x_{k-1}\}) = 0$ implies that $(1+\delta)(p_1+\cdots+p_{k-1})-\delta \leq 0$, we conclude that $(1+\delta)(p_1+\cdots+p_{k-1})-\delta = 0$, or, in other words,

$$\delta = \frac{p_1+\cdots+p_{k-1}}{1-p_1-\cdots-p_{k-1}}. \quad (25)$$

Now, for every $i = k,\ldots,n$, we have that

$$\begin{aligned}0 \leq m(\{x_1,\ldots,x_{k-1},x_i\}) &= \underline{P}(\{x_1,\ldots,x_{k-1},x_i\})\\ &= (1+\delta)(p_1+\cdots+p_{k-1}+p_i)-\delta = (1+\delta)p_i,\end{aligned}$$

and as a consequence

$$\sum_{i=k}^{n} m(\{x_1,\ldots,x_{k-1},x_i\}) = \sum_{i=k}^{n}(1+\delta)p_i = \sum_{i=k}^{n} p_i + \delta \sum_{i=k}^{n} p_i = \sum_{i=k}^{n} p_i + \sum_{j=1}^{k-1} p_j = 1,$$

taking into account the value of δ from Eq. (25). From this we deduce that the only focal elements of $\underline{P}$ are the sets $\{x_1,\ldots,x_{k-1},x_i\}$ for $i=k,\ldots,n$, and as a consequence we are in case (B4). □

We have already mentioned that in [3, Thm. 1], a sufficient condition for a probability interval to induce a completely monotone function is given. Since from Theorem 1 any PMM is in particular a probability interval, we now investigate the connection between [3, Thm. 1] and the characterization we have established in Theorem 2.

Theorem 3. [3, Thm. 1] *Consider a universe $\mathcal{X}=\{x_1,\ldots,x_n\}$ with $n\geq 3$, and consider the probability interval $\mathcal{I}=\{[l_i,u_i] : i=1,\ldots,n\}$. Denote by l and u the lower and upper probabilities induced from $\mathcal{I}$ by using Eq.* (13)*. If*

$$\left|\left\{i : u_i + \sum_{j\neq i} l_j < 1\right\}\right| \leq 2 \tag{26}$$

then l and u are a belief and a plausibility function.

Let us now show that this sufficient condition is not necessary in our framework. First of all, note that if we are in case (B1), the probability interval associated with a PMM trivially satisfies Eq. (26) because $l_i=\underline{P}(\{x_i\})=0$ and $u_i=\overline{P}(\{x_i\})=1$ for every $i=1,\ldots,n$. In the other three cases, we are going to see that the sufficient condition is not necessary.

Example 6. *Consider the four-element universe $\mathcal{X}=\{x_1,x_2,x_3,x_4\}$ and the initial probability P_0 =(0.1,0.1,0.2,0.6). In the next table we summarize the values of $\underline{P},\overline{P}$ for $\delta=1$ and $\delta=1.5$:*

	$\delta=1$			$\delta=1.5$		
A	$\underline{P}(A)$	$\overline{P}(A)$	$m(A)$	$\underline{P}(A)$	$\overline{P}(A)$	$m(A)$
$\{x_1\}$	0	0.2	0	0	0.25	0
$\{x_2\}$	0	0.2	0	0	0.25	0
$\{x_3\}$	0	0.4	0	0	0.5	0
$\{x_4\}$	0.2	1	0.2	0	1	0
$\{x_1,x_2\}$	0	0.4	0	0	0.5	0
$\{x_1,x_3\}$	0	0.6	0	0	0.75	0
$\{x_1,x_4\}$	0.4	1	0.2	0.25	1	0.25
$\{x_2,x_3\}$	0	0.6	0	0	0.75	0
$\{x_2,x_4\}$	0.4	1	0.2	0.25	1	0.25
$\{x_3,x_4\}$	0.6	1	0.4	0.5	1	0.5
$\{x_1,x_2,x_3\}$	0	0.8	0	0	1	0
$\{x_1,x_2,x_4\}$	0.6	1	0	0.5	1	0
$\{x_1,x_3,x_4\}$	0.8	1	0	0.75	1	0
$\{x_2,x_3,x_4\}$	0.8	1	0	0.75	1	0
$\mathcal{X}$	1	1	0	1	1	0

We see that for $\delta = 1$, we are in case (B3) and for $\delta = 1.5$ we are in case (B4) of Theorem 2, so $\underline{P}$ is a belief function. However, in none of these cases the probability interval associated with the PMM satisfies Eq. (26)*, as $|\{i : u_i + \sum_{j\neq i} l_j < 1\}| = 3$ in both cases.*

On the other hand, take $n = 3$, the initial probability $P_0 = (0.2, 0.4, 0.4)$ and $\delta = 1$. We are in case (B2) of Theorem 2, so $\underline{P}$ is a belief function; its focal sets are $\{x_1, x_2\}, \{x_1, x_3\}$ and $\{x_2, x_3\}$ with masses 0.2, 0.2 and 0.6, respectively. However, Eq. (26) *is not satisfied, because every x_i satisfies $0 = l_i < u_i < 1$, and therefore $|\{i : u_i + \sum_{j\neq i} l_j < 1\}| = 3$.* ♦

Therefore, the sufficient condition in terms of probability intervals given in Theorem 3 is only necessary for the PMM in the very particular case of (B1). On the other hand, Example 6 also shows that conditions (B2), (B3), (B4) may indeed be fulfilled by a PMM; to see that (B1) may also be satisfied, it suffices to consider $P_0 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\delta > 2$, taking into account the comments after Proposition 6.

4.3. **Minitive functions.** One particular family of belief functions is that of minitive functions. A function $\underline{P} : \mathcal{P}(\mathcal{X}) \to [0, 1]$ is called *minitive* if $\underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}$ for every $A, B \subseteq \mathcal{X}$. Any minitive function is in particular a belief function, and it corresponds to the particular case of *nested* focal elements. This means that they can be totally ordered by means of set inclusion.

Taking Theorem 2 into account, we can characterize the PMMs that induce a minitive function.

Corollary 2. *Let (P_0, δ) be PMM and $\underline{P}$ be the lower probability it induces by Eq.* (1)*. $\underline{P}$ is a minitive function if and only if one of the following conditions is satisfied:*

(B1) $k = n$.

(B2)* $k = n - 1$ *and there exists only one $x \in \mathcal{X}$ such that $m(\mathcal{X}\backslash\{x\}) > 0$.*

Proof. First of all, if condition (B1) is satisfied, from Theorem 2 we know that $\underline{P}$ is a belief function whose only focal set is $\mathcal{X}$. Therefore, $\underline{P}$ is in particular minitive.

If (B2*) is satisfied, in particular condition (B2) is satisfied in Theorem 2, so $\underline{P}$ is a belief function with only two focal sets: $\mathcal{X}$ and $\mathcal{X}\backslash\{x\}$, which are nested. Therefore $\underline{P}$ is not only a belief function but also minitive.

Let us now see that if $\underline{P}$ is minitive, no other situation is possible. Since any minitive function is a belief function, from Theorem 2 it must satisfy one of (B1), (B2), (B3) or (B4). If (B3) or (B4) are satisfied, this means that there is a set B with cardinality smaller than $n - 1$ such that $B \cup \{x\}$ is focal for every $x \notin B$. Therefore, there are $x_1, x_2 \in B^c$ $(x_1 \neq x_2)$ such that $B \cup \{x_1\}$ and $B \cup \{x_2\}$ are focal. But these two sets are not nested, and therefore $\underline{P}$ is not a minitive function.

Similarly, if (B2) holds, then $\sum_{i=1}^{n} \underline{P}(\mathcal{X}\backslash\{x_i\}) \leq 1$ and the focal sets are $\mathcal{X}$ and $\mathcal{X}\backslash\{x_i\}$ for every $i = 1, \ldots, n$. Therefore, since the focal sets must be nested, there can only be one $x \in \mathcal{X}$ such that $\underline{P}(\mathcal{X}\backslash\{x\}) > 0$, so (B2*) must be satisfied. □

4.4. **Extreme points of completely monotone PMM.** Our previous results allow us to compute the maximum number of extreme points for the credal set associated with a completely monotone PMM:

Proposition 10. *Let (P_0, δ) be a completely monotone PMM, and consider its associated credal set $\mathcal{M}(P_0, \delta)$. Then the number of extreme points of $\mathcal{M}(P_0, \delta)$ is:*

(a) *n when (P_0, δ) satisfies condition (B1) in Theorem 2.*
(b) *At most $\binom{n}{2}$ when (P_0, δ) satisfies condition (B2).*
(c) *At most $k2^{n-k}$ when (P_0, δ) satisfies condition (B3).*
(d) *At most $(k-1)2^{n-k+1} - (k-2)$ when (P_0, δ) satisfies condition (B4).*

Proof. (a) If we are in case (B1), then $\mathcal{M}(P_0, \delta)$ is the set of all probability measures. This set has n different extreme points: the degenerate probability measures.

(b) If we are in case (B2), we deduce from Proposition 7 that the focal elements of the belief function $\underline{P}$ are sets of the type $\{x_i\}^c$ for every $i = 1, \dots, n$ together with $\mathcal{X}$. Therefore, given a permutation σ, the extreme point P_σ it induces by Eq. (4) assigns positive mass to at most $x_{\sigma(1)}$ and $x_{\sigma(2)}$. Since there are at most $\binom{n}{2}$ different pairs, we have at most $\binom{n}{2}$ different extreme points.

(c) If we are in case (B3), we deduce from Proposition 8 that the focal elements of the belief function $\underline{P}$ are the sets $B, B \cup \{x\}$ for every $x \notin B$. Thus, given a permutation σ, if i is the smallest index in $\{1, \dots, n\}$ such that $x_{\sigma(i)} \in B$, then the associated extreme point P_σ gives positive mass to the elements $x_{\sigma(1)}, \dots, x_{\sigma(i)}$. Moreover, for every other permutation σ' such that $\sigma(i) = \sigma'(i), \{\sigma(1), \dots, \sigma(i-1)\} = \{\sigma'(1), \dots, \sigma'(i-1)\}$, it follows that $P_\sigma = P_{\sigma'}$. Thus, the number of different extreme points results from combining the element of B that comes first (and there are k possibilities for that) with the groups of $0, 1, 2, ..., n-k$ elements of B^c, for which there are

$$\binom{n-k}{0} + \binom{n-k}{1} + \cdots + \binom{n-k}{n-k} = 2^{n-k}.$$

Thus, we have $k2^{n-k}$ different extreme points.

(d) Finally, if we are in case (B4), we deduce from Proposition 9 that the focal elements of the belief function $\underline{P}$ are the sets $B \cup \{x\}$ for every $x \notin B$ for some given B. If $B = \emptyset$ (k=1), we obtain that the focal elements are the singletons, and then $\underline{P}$ is a probability measure, meaning that there is only one extreme point. Assume next that $k > 1$. Then, given a permutation σ, if i is the smallest index in $\{1, \dots, n\}$ such that $x_{\sigma(i)} \in B$, then the associated extreme point P_σ gives positive mass to the elements $x_{\sigma(1)}, \dots, x_{\sigma(i)}$, when $i < n-k+2$, and to $x_{\sigma(1)}, \dots, x_{\sigma(i-1)}$, when $i = n-k+2$ (that is, when all the $n-k+1$ elements of B^c go first).

When $i < n-k+2$, for every other permutation σ' such that $\sigma'(i) = \sigma(i)$ and $\{\sigma'(1), \dots, \sigma'(i-1)\} = \{\sigma(1), \dots, \sigma(i-1)\}$ it holds that $P_{\sigma'} = P_\sigma$. On the other hand, if $i = n+k+2$, every other permutation σ' with $\{\sigma'(1), \dots, \sigma'(i-1)\} = \{\sigma(1), \dots, \sigma(i-1)\}$ satisfies $P_{\sigma'} = P_\sigma$. Thus, the number of extreme points is

$$(k-1)\binom{n-k+1}{0} + (k-1)\binom{n-k+1}{1} + \cdots + (k-1)\binom{n-k+1}{n-k} + \binom{n-k+1}{n-k+1} = (k-1)2^{n-k+1} - (k-2). \quad \square$$

Thus, the maximum number of extreme points of $\mathcal{M}(P_0, \delta)$ is 2^{n-1} when $n \geq 3$, and it corresponds to $k = 1, 2$ in case (B3), and to $k = 2$ in case (B4). This number is significantly lower than the bound we have established in Theorem 2 for general PMM, and also that for the number of extreme points for the credal set of a belief function, known to be $n!$.

The number of extreme points is even lower for PMM inducing a minitive function: taking into account Corollary 2, we are either in case (B1), and then we have n different extreme points, or in case (B2*), where there two focal elements $\mathcal{X} \setminus \{x\}$ and $\mathcal{X}$. In this second case there are two possibilities:

- If $x_{\sigma(1)} \neq x$, then $P_\sigma(\{x_{\sigma(1)}\}) = 1$, $P_\sigma(\{x_{\sigma(i)}\}) = 0 \ \forall i = 2, \ldots, n$. This determines $n-1$ different extreme points.
- If $x_{\sigma(1)} = x$, then $P_\sigma(\{x_{\sigma(1)}\}) = m(\mathcal{X}), P_\sigma(\{x_{\sigma(2)}\}) = 1 - m(\mathcal{X})$ and $P_\sigma(\{x_{\sigma(i)}\}) = 0 \ \forall i = 3, \ldots, n$. This determines $n-1$ different extreme points.

Thus, the maximum number of extreme points of a minitive PMM is $2(n-1)$, much lower that the maximum number of extreme points of the credal set associated with a minitive function, known to be 2^{n-1} [19].

5. Approximation by PMM

As we said in the introduction, one of the main reasons to study the properties of simple probability sets is that they can be instrumental to approximate more complex models, therefore reducing the computational burden of inferences in exchange for being less-committal. Such approximations are commonly used in graphical models [1, 8] or in risk analysis [4].

In the case of a PMM, the question is to know whether an initial credal set $\mathcal{M}$ can be easily outer approximated by a PMM, so that $\mathcal{M} \subseteq \mathcal{M}(P_0, \delta)$ and, among the outer-approximating PMMs, if there is a minimal one with this property, in the sense that there is no other PMM (P_0', δ) with $\mathcal{M} \subseteq \mathcal{M}(P_0', \delta) \subsetneq \mathcal{M}(P_0, \delta)$. For other imprecise probability models such an outer-approximation is not unique, but in the case of PMMs it turns out that there is a unique such outer approximation [21], that is moreover straightforward to compute.

Proposition 11. [21, Sec. 4.1.] *Let $\mathcal{M}$ be a convex set of probabilities and $\overline{P}$ be its upper probability. The PMM (P_0, δ) such that*

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

is the unique minimal PMM model that outer approximates $\mathcal{M}$.

This property means that approximating a given probability set with a PMM is actually quite easy. In practice, such approximations may make the model much easier to handle, as our next examples show.

Example 7. *Consider an experiment with four possible outcomes, and where their respective probabilities have been estimated to be equal to 0.1, 0.2, 0.3 and 0.4, but where there has been some problems in the transmission so that we cannot tell which probability corresponds to which value (frequencies were communicated, but not which one corresponded to which element). If we denote the possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, this means that the available information is given by the probability measure associated with the mass function* $(0.1, 0.2, 0.3, 0.4)$ *and its* *permutations. We obtain then a set $\mathcal{M}$ of 24 probability measures, whose lower envelope is the belief function given by*

$$Bel(A) = \begin{cases} 0.1 & \text{if } |A| = 1 \\ 0.3 & \text{if } |A| = 2 \\ 0.6 & \text{if } |A| = 3 \\ 1 & \text{if } |A| = 4. \end{cases}$$

Working with a belief function on a space of cardinality four entails specifying its Möbius inverse, for which we must specify the basic probability assignment on the $2^4 = 16$ *subsets of* $\mathcal{X}$*, or considering the* $4! = 24$ *different extreme points of its associated credal set* $\mathcal{M}(\underline{P})$*. One alternative is then to consider a PMM that is as close as possible to the belief function while not introducing any additional information. Using the results in* [21] *and recalled by Proposition 11, it can be verified that the closest PMM to Bel is given by* $P_0 = (0.25, 0.25, 0.25, 0.25)$ *and* $\delta = 0.6$*, that produces*

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

The uniformity of P_0 *models the fact that we have symmetric information about the different values in* $\mathcal{X}$*, while the distortion factor* δ *measures the imprecision introduced by the lack of information about the correct order.*

The lower probability $\underline{P}_\delta$ *is determined by five values only: the probability mass function* P_0 *and the distortion factor* δ*. Moreover, its associated credal set* $\mathcal{M}(\underline{P})$ *has 12 extreme points: the probability measures associated with* $(0, 0.2, 0.2, 0.4)$ *and its permutations. Thus, the representation in terms of extreme points is also simpler (the initial* $\mathcal{M}(\underline{P})$ *having 24 of them).*

Finally, for any function $f : \mathcal{X} \to \mathbb{R}$ *it follows that the difference between its Choquet integral with respect to the original belief function and its outer approximation is bounded above by:*

$$\left| (C) \int f dBel - (C) \int f d\underline{P}_\delta \right| \leq \max_{A \subseteq \mathcal{X}} |Bel(A) - P_\delta(A)| \sup |f| = 0.1 \sup |f|,$$

meaning that the loss of information entailed by the PMM is not too large. ♦

Example 8. *Consider an expert that must provide some conditional probability bounds in a credal network* [26]*, with the space* $\mathcal{X} = \{x_1, x_2, x_3\}$ *of interest counting three elements (these could be categories that a given object can take, such as kind of movies in recommender systems, allele in genetic codes, ...). The expert has some knowledge about probabilities, so is confident enough to provide numbers, but prefers to give the following intervals:*

$$p(x_1) \in [0.1, 0.24], \quad p(x_2) \in [0.45, 0.6], \quad p(x_3) \in [0.2, 0.36]$$

that correspond to a probability interval $\mathcal{I}$*. It can be checked that this model has the maximal number of extreme points a probability interval can have, i.e., six in our case. Using Proposition 11, the outer-approximating PMM of* $\mathcal{I}$ *is given by* $P_0 = (0.2, 0.5, 0.3)$ *and* $\delta = 0.2$*, that has only three extreme points. Figure 2 pictures the two models, in term of probability sets over the simplex. This example* *also illustrates the fact that, while probability intervals and PMMs have potentially the same maximal number of extreme points, it can be the case that a probability interval reaching this number can be approximated by a PMM with a much lower number.* ♦

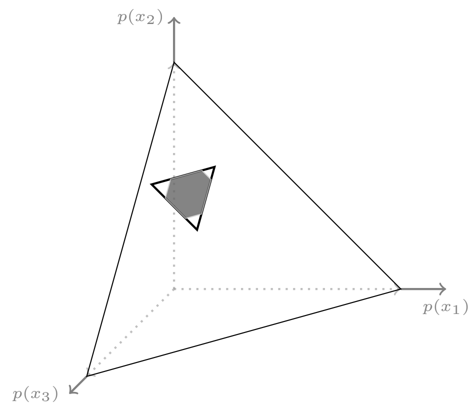


FIGURE 2. Example of aproximation.

6. Combining multiple PMMs

In this section, we study what happens when we consider multiple PMMs, first characterizing our uncertainty of a common variable (they could be two precise assessments provided by experts with different types of expertise), and then characterizing our uncertainty over multiple variables (they could be two assessments coming from the same expert about different variables). For simplicity, we focus on binary cases where either two sources communicate their uncertainty or two variables are concerned. Most of the conclusions extend straightforwardly to the multivariate case.

6.1. **Information fusion of PMMs.** When two sets $\mathcal{M}(P_0^1, \delta_1)$ and $\mathcal{M}(P_0^2, \delta_2)$ are provided to describe our uncertainty over $\mathcal{X}$, one often needs to combine them into a single model [11]. Three classical ways to achieve such a combination are to consider the conjunction (intersection), the disjunction (union) or the average (convex mixture) of the models. Using the commutativity and (quasi-)associativity of these operators, extensions to an arbitrary number of sources is straightforward.

Before starting our study of such models, recall that from Corollary 1 $\mathcal{M}(P_0, \delta)$ is the set of probability measures satisfying the constraints in Eq. (6): every probability $P \in \mathcal{M}(P_0, \delta)$ must satisfy $(1+\delta)P_0(\{x\}) \geq P(\{x\}) \ \ \forall x \in \mathcal{X}$. As already indicated in Section 4, this corresponds to a probability interval where only upper bounds are provided (since the lower bounds can be derived from them).

Conjunction. Let us denote by

$$\mathcal{M}(P_0^{\cap}, \delta^{\cap}) := \mathcal{M}(P_0^1, \delta_1) \cap \mathcal{M}(P_0^2, \delta_2)$$

the probability set obtained by conjunctively combining $\mathcal{M}(P_0^1, \delta_1)$ and $\mathcal{M}(P_0^2, \delta_2)$.

Proposition 12. *The set $\mathcal{M}(P_0^{\cap}, \delta^{\cap})$ is non-empty if and only if*

$$\sum_{x \in \mathcal{X}} \min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\}), 1\right\} \geq 1. \tag{27}$$

In that case, it is induced by the PMM $(P_0^{\cap}, \delta^{\cap})$ such that

$$\delta^{\cap} = \left(\sum_{x \in \mathcal{X}} \min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\}\right) - 1 \tag{28}$$

$$P_0^{\cap}(\{x\}) = \frac{\min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\}}{1+\delta^{\cap}}. \tag{29}$$

Proof. Given the two sets of constraints given by Eq. (6) applied to (P_0^1, δ_1) and (P_0^2, δ_2), their intersection is the set of probability measures satisfying

$$\min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\} \geq P(\{x\}) \ \ \forall x \in \mathcal{X}.$$

This corresponds to a specific probability interval where

$$l_i = 0, u_i = \min\left\{(1+\delta_1)P_0^1(\{x_i\}), (1+\delta_2)P_0^2(\{x_i\}), 1\right\} \ \forall i;$$

by Eq. (14), the associated credal set is non-empty if and only if

$$\sum_{x \in \mathcal{X}} \min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\}), 1\right\} \geq 1.$$

Assume now that this intersection is non-empty, and let us prove that in that case it can be induced by the PMM $(P_0^{\delta}, \delta^{\cap})$. Taking into account Corollary 1, if $\mathcal{M}(\mathcal{I}) = \mathcal{M}(P_0^{\cap}, \delta^{\cap})$, then for every probability measure P it should hold that

$$P(\{x_i\}) \leq u_i \iff P(\{x_i\}) \leq (1+\delta^{\cap})P_0^{\cap}(\{x_i\}) \ \forall x_i \in \mathcal{X}. \tag{30}$$

If we make $u_i = (1+\delta^{\cap})P_0^{\cap}(\{x_i\})$ and take into account that

$$\sum_{x \in \mathcal{X}} (1+\delta^{\cap})P_0^{\cap}(\{x\}) = \sum_{x \in \mathcal{X}} \min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\} = 1+\delta^{\cap},$$

we obtain that Eq. (30) is satisfied for the values of $P_0^{\cap}$ and $\delta^{\cap}$ given in Eqs. (28) and (29), and that moreover with those definitions $P_0^{\cap}(\{x\}) \in [0,1]$ for every $x \in \mathcal{X}$. $\square$

In the particular case where $P_0^1 = P_0^2$, Eq. (27) is always satisfied because:

$$\begin{aligned}\sum_{x \in \mathcal{X}} &\min\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\}), 1\right\} \\ &= \sum_{x \in \mathcal{X}} \min\left\{(1+\min\{\delta_1, \delta_2\})P_0(\{x\}), 1\right\} \geq \sum_{x \in \mathcal{X}} P_0(\{x\}) = 1,\end{aligned}$$

and the values of $\delta^{\cap}$ and $P_0^{\cap}$ given in Eqs. (28) and (29) become:

$$\delta^{\cap} = \min\{\delta_1, \delta_2\} \text{ and } P_0^{\cap} = P_0.$$

Indeed, if $P_0^1 = P_0^2$ and $\delta_1 \leq \delta_2$, we have $\mathcal{M}(P_0^1, \delta_1) \subseteq \mathcal{M}(P_0^2, \delta_2)$, hence $\mathcal{M}(P_0^1, \delta_1) \cap \mathcal{M}(P_0^2, \delta_2) = \mathcal{M}(P_0^1, \delta_1)$. In the more general case, Proposition 12 provides us with a simple procedure to verify the non-emptiness of $\mathcal{M}(P_0^{\cap}, \delta^{\cap})$, as well as efficient formulae to compute the conjunction of two (or more) PMMs.

Example 9. *Consider the space $\mathcal{X} = \{x_1, x_2, x_3\}$ and the two following models*

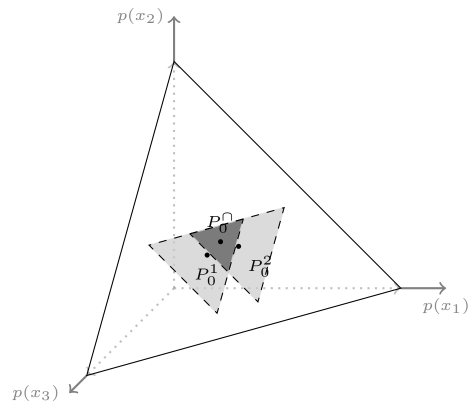
$$P_0^1 = (0.3, 0.3, 0.4), \quad \delta_1 = 0.3,$$

$$P_0^2 = (0.4, 0.3, 0.3), \quad \delta_2 = 0.3,$$

that are such that $\mathcal{M}(P_0^1, \delta_1) \cap \mathcal{M}(P_0^2, \delta_2) \neq \emptyset$. Their conjunction is given by

$$P_0^{\cap} = (1/3, 1/3, 1/3), \quad \delta^{\cap} = 0.17.$$

The result is illustrated on Figure 3, where the initial two PMMs are in light gray, and the resulting conjunction is in dark gray. ♦



$\mathcal{M}(P_0^1, \delta_1) \cap \mathcal{M}(P_0^2, \delta_2)$

FIGURE 3. Example of conjunction.

Disjunction. When the intersection of two credal sets is empty (they are conflicting), an alternative is to consider their union, that is to consider $\mathcal{M}(P_0^1, \delta_1) \cup \mathcal{M}(P_0^2, \delta_2)$ or its convex hull, since $\mathcal{M}(P_0^1, \delta_1) \cup \mathcal{M}(P_0^2, \delta_2)$ will not be convex in general.

The convex hull $\mathrm{conv}(\mathcal{M}(\mathrm{P}_0^1, \delta_1) \cup \mathcal{M}(\mathrm{P}_0^2, \delta_2))$ will not be induced by a PMM in general, either. However, we can easily provide a best outer-approximating PMM $(P_0^{\cup}, \delta^{\cup})$ using the fact that any outer-approximation of $\mathcal{M}(P_0^1, \delta_1) \cup \mathcal{M}(P_0^2, \delta_2)$ must satisfy the constraint

$$\max\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\} \geq P(\{x\}) \ \ \forall x \in \mathcal{X}.$$

Indeed, using the same arguments as in the proof of Proposition 12, we can define

$$\delta^{\cup} = \left(\sum_{x \in \mathcal{X}} \max\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\} \right) - 1 \tag{31}$$

and

$$P_0^{\cup}(\{x\}) = \frac{\max\left\{(1+\delta_1)P_0^1(\{x\}), (1+\delta_2)P_0^2(\{x\})\right\}}{1+\delta^{\cup}} \tag{32}$$

so that $\mathcal{M}(P_0^{\cup}, \delta^{\cup}) \supseteq \mathcal{M}(P_0^1, \delta_1) \cup \mathcal{M}(P_0^2, \delta_2)$. To see that this inclusion holds, note that for every event A, we have

$$\sum_{x \in A} \max\left\{\overline{P}^1(\{x\}), \overline{P}^2(\{x\})\right\} \geq \max\left\{\sum_{x \in A} \overline{P}^1(\{x\}), \sum_{x \in A} \overline{P}^2(\{x\})\right\}$$

where $\overline{P}^1, \overline{P}^2$ are the upper probabilities induced by (P_0^1, δ_1) and (P_0^2, δ_2), respectively.

Example 10. *Consider the space $\mathcal{X} = \{x_1, x_2, x_3\}$ and the two following models*

$$P_0^1 = (0.3, 0.4, 0.3), \quad \delta_1 = 0.2,$$

$$P_0^2 = (0.2, 0.2, 0.6), \quad \delta_2 = 0.3,$$

that satisfy $\mathcal{M}(P_0^1, \delta_1) \cap \mathcal{M}(P_0^2, \delta_2) = \emptyset$. Their outer-approximation is given by

$$P_0^{\cup} = (0.222, 0.297, 0.481), \quad \delta^{\cup} = 0.62.$$

The result is illustrated on Figure 4, where the initial two PMMs are in light gray, and the resulting outer-approximation of the disjunction is in dark gray. ♦

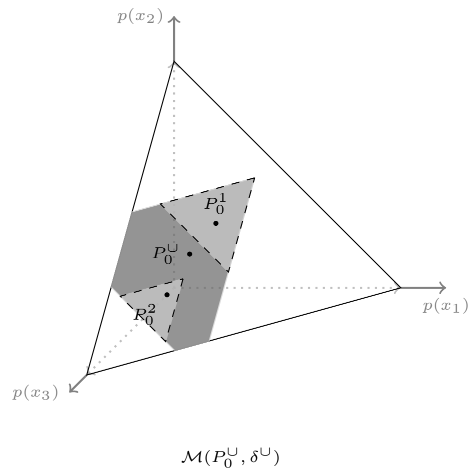


FIGURE 4. Example of approximated disjunction.

Mixture. The mixture of two PMMs, that is, the computation of

$$\mathcal{M}(P_0^{\epsilon}, \delta_{\epsilon}) := \epsilon \mathcal{M}(P_0^1, \delta_1) + (1 - \epsilon)\mathcal{M}(P_0^2, \delta_2)$$

for a given $\epsilon \in (0, 1)$ is straightforward when the the PMMs to be mixed do not involve zero lower probabilities. In that case, we deduce that the upper probability of the mixture $\mathcal{M}(P_0^{\epsilon}, \delta_{\epsilon})$, that is equal to the mixtures of the upper probabilities of $\mathcal{M}(P_0^1, \delta_1)$ and $\mathcal{M}(P_0^2, \delta_2)$, is given by $1 + \delta_1)P_0^1(A) + (1 - \epsilon)(1 + \delta_2)P_0^2(A)$ for any $A \neq \mathcal{X}$; as a consequence, the model $\mathcal{M}(P_0^{\epsilon}, \delta_{\epsilon})$ is described by the constraints

$$\epsilon(1 + \delta_1)P_0^1(\{x\}) + (1 - \epsilon)(1 + \delta_2)P_0^2(\{x\}) \geq P(\{x\}) \ \ \forall x \in \mathcal{X}$$

on a probability measure P. From this, we deduce that

$$\begin{aligned}1+\delta_\epsilon &= \sum_{x\in\mathcal{X}} \epsilon(1+\delta_1)P_0^1(\{x\}) + (1-\epsilon)(1+\delta_2)P_0^2(\{x\})\\ &= \epsilon(1+\delta_1)\sum_{x\in\mathcal{X}} P_0^1(\{x\}) + (1-\epsilon)(1+\delta_2)\sum_{x\in\mathcal{X}} P_0^2(\{x\})\\ &= \epsilon(1+\delta_1) + (1-\epsilon)(1+\delta_2),\end{aligned}$$

and

$$P_0^\epsilon(\{x\}) = \frac{\epsilon(1+\delta_1)P_0^1(\{x\}) + (1-\epsilon)(1+\delta_2)P_0^2(\{x\})}{1+\delta_\epsilon}.$$

Example 11. *Consider the initial models of Example 10 with $\epsilon = 0.5$ to have an arithmetic average, we obtain the model*

$$p_0^\epsilon = (0.248, 0.296, 0.456), \quad \delta^\epsilon = 0.25.$$

The result is illustrated on Figure 5, where the initial two PMMs are in light gray, and the resulting average is in dark gray. ♦

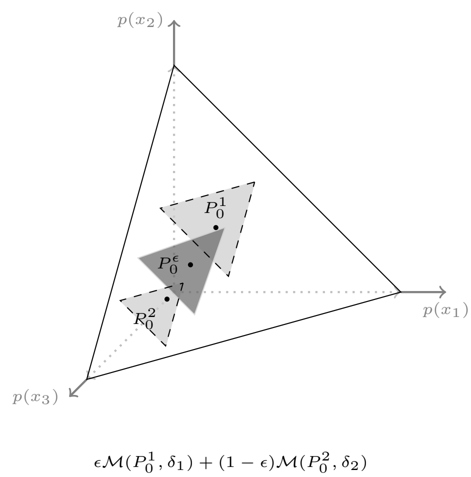


FIGURE 5. Example of PMM mixture.

Other, more elaborate combinations can be derived from these basic ones (see for instance [22, 36]). While we will not discuss them in details in this paper, the next example provides an illustrative practical case where they would be useful.

Example 12. *Let us consider three different classifiers, each producing output estimates in the form of probabilities over four classes $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ (to be interpreted as P_0), and a reliability index of their estimates in addition (to be interpreted as δ). Let us now consider the three following outputs:*

$$\begin{aligned}P_0^1 &= (0.7, 0.1, 0.1, 0.1), \quad \delta^1 = 0.4;\\ P_0^2 &= (0.4, 0.2, 0.2, 0.2), \quad \delta^2 = 0.5;\\ P_0^3 &= (0.25, 0.25, 0.25, 0.25), \quad \delta^3 = 0.8.\end{aligned}$$

In this scenario, the first two classifiers would choose class x_1, with the second being more uncertain (in every way) than the first, and the last classifier would consider all classes equally likely, but with a very low reliability level (maybe due to lack of data). Using Proposition 12, it is straightforward to check that $\mathcal{M}(P_0^1,\delta^1)\cap\mathcal{M}(P_0^2,\delta^2)$ and $\mathcal{M}(P_0^3,\delta^3)\cap\mathcal{M}(P_0^2,\delta^2)$ are non-empty, but that on the other hand $\mathcal{M}(P_0^1,\delta^1)\cap\mathcal{M}(P_0^2,\delta^2)\cap\mathcal{M}(P_0^3,\delta^3)=\emptyset$. In this scenario, full conjunction is impossible, and disjunction would provide a very imprecise result, losing the information that x_1 is likely to be the most probable class. One alternative solution is to combine conjunction and disjunctions, by taking the conjunction of every maximal subset of sources for which it is non-empty, and then considering the disjunction between them. In our case, this would come down to take

$$\left(\mathcal{M}(P_0^1,\delta^1)\cap\mathcal{M}(P_0^2,\delta^2)\right)\cup\left(\mathcal{M}(P_0^3,\delta^3)\cap\mathcal{M}(P_0^2,\delta^2)\right)$$

or the best PMM approximating it, obtained using Equations (31)-(32). *This gives the PMM*

$$P_0^{(1\cap 2)\cup(2\cap 3)}=(0.4,0.2,0.2,0.2),\quad \delta^{(1\cap 2)\cup(2\cap 3)}=0.5,$$

which still encodes the fact that x_1 is highly likely compared to other classes. ♦

6.2. **Marginal and joint PMM.** A second related problem would be the study of PMM on product spaces, as an uncertainty model about two different variables X,Y taking respective values on $\mathcal{X}=\{x_1,\dots,x_n\}$ and $\mathcal{Y}=\{y_1,\dots,y_m\}$. These can arise as a combination of two marginal PMMs into a joint one, or we may instead be interested in deriving the marginal models from a given joint PMM. We discuss the two possibilities in this section.

From marginals to joint. Assuming we have a discounting factor δ and two probabilities P_0^X and P_0^Y given on $\mathcal{X}$ and $\mathcal{Y}$, respectively, it seems reasonable to wonder whether we should first:

- combine P_0^X and P_0^Y into a joint probability $P_0^{X,Y}$ over $\mathcal{X}\times\mathcal{Y}$ and consider the discounted set $\mathcal{M}(P_0^{X,Y},\delta)$ or;
- compute the discounted sets into $\mathcal{M}(P_0^X,\delta)$ and $\mathcal{M}(P_0^Y,\delta)$ and then combine the two resulting sets into a joint set $\mathcal{M}^{X,Y}$.

Under an assumption of independence, the first approach has the advantage that $P_0^{X,Y}$ has a unique formal definition corresponding to stochastic independence, while the joint $\mathcal{M}^{XY}$ will strongly depend on the chosen extension of the classical notion of probabilistic independence, since many of them exist [6]. To facilitate the discussion, we will only consider the set $\mathcal{M}^{XY}$ resulting from the assumption of *strong independence*, that corresponds to the convex hull of all stochastic products of probabilities within $\mathcal{M}(P_0^X,\delta)$ and $\mathcal{M}(P_0^Y,\delta)$. One question that arises is what is the relation between the set $\mathcal{M}(P_0^{X,Y},\delta)$ resulting from the first approach and the set $\mathcal{M}^{XY}$ resulting from the second approach. To ease the notation, let $\overline{P}^1$ and $\overline{P}^2$ denote their respective upper probabilities over $\mathcal{X}\times\mathcal{Y}$. An immediate remark is that the two joint sets will not be equal in general: we have

$$\overline{P}^1(A\times B)=(1+\delta)P_0^X(A)P_0^Y(B)\le(1+\delta)P_0^X(A)(1+\delta)P_0^Y(A)=\overline{P}^2(A\times B)\quad(33)$$

whenever $\overline{P}^2(A\times B)<1$. The second value is obtained from the fact that under the strong independence assumption [6], we have the factorisation property

$$\overline{P}^2(A\times B)=\overline{P}^X(A)\overline{P}^Y(B),\quad(34)$$

where $\overline{P}^X$, $\overline{P}^Y$ are the marginal upper probabilities of $\overline{P}$. Note that the inequality in (33) is strict as soon as $\delta, P_0^X(A)$ and $P_0^Y(B)$ are all strictly positive. From (33), a natural question is whether $\mathcal{M}(P_0^{X,Y}, \delta) \subseteq \mathcal{M}^{XY}$. The next example shows that it will not be the case in general.

Example 13. *Consider the spaces $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2\}$ with the two probabilities P_0^X, P_0^Y given by*

$$P_0^X(x_1) = 0.3, \quad P_0^X(x_2) = 0.7, \quad P_0^Y(y_1) = 0.5, \quad P_0^Y(y_2) = 0.5,$$

and let $\delta = 0.1$. Given the event $E = \{(x_2, y_2)\}^c$, we obtain

$$\begin{aligned}\overline{P}^1(E) &= 1 - \underline{P}^1(\{(x_2, y_2)\}) = 1 - 0.285 = 0.715\\ &> \overline{P}^2(E) = 1 - \underline{P}^2(\{(x_2, y_2)\}) = 1 - 0.67 \cdot 0.45 = 1 - 0.3015 = 0.6985,\end{aligned}$$

and therefore it cannot be $\mathcal{M}(P_0^{X,Y}, \delta) \subseteq \mathcal{M}^{XY}$. ♦

Most other independence notions (including epistemic independence and random set independence, for instance) used within imprecise probability theory [6] also satisfy Eq. (34), hence the inequality concerning events of the kind $A \times B$ remains true for them. As this factorisation property is also true for lower probabilities, Example 13 also applies to them. This may be an important issue when having to choose whether one should first combine then discount, or discount then combine. We can nevertheless notice that there is essentially one way to apply the first option (using stochastic independence), and many to apply the second (as one has to choose an adequate notion of independence).

From joint to marginals. Let us now start from a PMM $(P_0^{X,Y}, \delta)$ on a product space $\mathcal{X} \times \mathcal{Y}$. Its associated credal set is the set of probability measures satisfying

$$(1 + \delta)P_0(\{x_i, y_j\}) \geq P(\{x_i, y_j\}) \quad \forall (x_i, y_j) \in \mathcal{X} \times \mathcal{Y}.$$

The question is then to know, if we want to marginalize all probabilities contained in $\mathcal{M}(P_0^{X,Y}, \delta)$ over $\mathcal{X}$ or $\mathcal{Y}$, what is the shape of the resulting credal sets $\mathcal{M}^X$ and $\mathcal{M}^Y$? The answer is pretty straightforward as soon as we realize that the marginal model on X is described by the constraints

$$P(\{x_i\}) = \sum_{y_j \in \mathcal{Y}} P(\{x_i, y_j\}) \leq \sum_{y_j \in \mathcal{Y}} (1 + \delta)P_0(\{x_i, y_j\}) \leq (1 + \delta)P_0(\{x_i\})$$

$\forall x_i \in \mathcal{X}$, which correspond to the set $\mathcal{M}(P_0^X, \delta)$ where P_0^X is the marginal of P_0 over $\mathcal{X}$. The same holds for $\mathcal{M}^Y$.

7. Conclusions

Our results show that the pari-mutuel model is a computationally simple model within imprecise probability theory that at the same time keeps enough generality to be useful in a number of practical situations. On the one hand, we have proven that it can be embedded within the theory of probability intervals, and as such can be used quite easily within graphical models. In this respect, it is interesting to note that, even if not all probability intervals can be represented as a pari-mutuel model, the tightest bound on the number of extreme points of their associated set of probabilities is the same for both of them.

In addition, we have also determined in which cases a pari-mutuel model is equivalent to a pair of conjugate belief and plausibility functions. Such a representation is interesting because it allows for instance to use pari-mutuel models in the context of random sets [23]. Belief functions and 2-monotone lower probabilities have also been considered in some of the extensions of the expected utility paradigm that deal with imprecise information [12, 27]. In this respect, now that we have clarified the connection between the pari-mutuel model and other models within imprecise probability theory, it would be interesting to study the preferences that can be modelled by means of pari-mutuel probabilities. Nevertheless, we should stress that our necessary and sufficient conditions for the pari-mutuel model to be embedded in the theory of belief and plausibility functions, that improve upon earlier results from the literature, show that the inclusion within the theory of belief functions only holds in quite restrictive scenarios. This phenomenon is even more acute if we consider the particular case of minitive measures.

With respect to the processes of combination we have considered in Section 6, we have shown that this model is closed under conjunction, marginalization and average, while it is not under disjunction or when building a joint model from marginal ones. Most of our observations extend directly to the case of more than two models, due to the associativity and commutativity of the operations involved.

Acknowledgements

The research reported in this paper has been supported by project TIN2014-59543-P, and by the project Labex MS2T, financed by the French Government through the program "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02). This paper is an updated version with proofs and additional results of a paper presented at ISIPTA'2017 [20].

References

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

[2] T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Wiley, 2014.

[3] P. Boodgumarn, P. Thipwiwatpotjana, and W. Lodwick. When a probability interval is a random set. *Science Asia*, 39:319–327, 2013.

[4] O. Bouissou, E. Goubault, J. Goubault-Larrecq, and S. Putot. A generalization of p-boxes to affine arithmetic. *Computing*, 94(2-4):189–201, 2012.

[5] 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.

[6] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

[7] F. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39(2-3):167–184, 2005.

[8] J. De Bock, C. De Campos, and A. Antonucci. Global sensitivity analysis for map inference in graphical models. In *Advances in Neural Information Processing Systems*, pages 2690–2698, 2014.

[9] 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.

[10] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.

[11] D. Dubois, W. Liu, J. Ma, and H. Prade. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. *Information Fusion*, 32:12–39, 2016.

[12] B. Sloth T. Tranaes E. Hendon, H.J. Jacobsen. Expected utility with lower probabilities. *Journal of Risk and Uncertainty*, 8(2):197–216, 1994.
[13] T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.
[14] H. Gerber. *An introduction to mathematical risk theory.* Huebner Foundation, 1979.
[15] M. Grabisch. *Set functions, games and capacities in decision making.* Springer, 2016.
[16] P. Guo and H. Tanaka. Decision making with interval probabilities. *European Journal of Operational Research*, 203(2):444–454, 2010.
[17] T. Herron, T. Seidenfeld, and L. Wasserman. Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444, 1997.
[18] J. F. Lemmer and H. E. Kyburg Jr. Conditions for the existence of belief functions corresponding to intervals of belief. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 488–493, 1991.
[19] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning*, 33(1):95–115, 2003.
[20] I. Montes, E. Miranda, and S. Destercke. A study of the Pari-Mutuel Model from the point of view of imprecise probabilities. In *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'2017)*, 2017.
[21] I. Montes, E. Miranda, and P. Vicig. Approximations of coherent lower probabilities by 2-monotone capacities. In J. Medina, M. Ojeda-Aciego, J.L. Verdegay, D.A. Pelta, I.P. Cabrera, B. Bouchon-Meunier, and R.R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*, pages 214–225, Cham, 2018. Springer International Publishing.
[22] S. Moral and J. del Sagrado. Aggregation of imprecise probabilities. In *Aggregation and fusion of imperfect information*, pages 162–188. Springer, 1998.
[23] H. T. Nguyen. *An introduction to random sets.* Chapman and Hall, 2006.
[24] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51(9):1145–1158, 2010.
[25] M. Peters, A. So, and Y. Ye. Pari-mutuel markets: mechanisms and performance. *Lecture Notes in Computer Science*, 4858:82–95, 2007.
[26] A. Piatti, A. Antonucci, and M. Zaffalon. Building knowledge-based systems by credal networks: a tutorial. *Advances in Mathematics Research*, 11:227–279, 2010.
[27] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.
[28] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.
[29] D. Skulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009.
[30] H. Tanaka, K. Sugihara, and Y. Maeda. Non-additive measures by interval probability functions. *Information Sciences*, 164:209–227, 2004.
[31] D. Terrell. A test on the gambler's fallacy evidence from pari-mutuel gambles. *Journal of Risk and Uncertainty*, 8(3):309–317, 1994.
[32] B. Tessem. Interval probability propagation. *International Journal of Approximate Reasoning*, 7(3–4):95–120, 1992.
[33] R.H. Thaler and W.T. Ziemba. Parimutuel betting markets: racetracks and lotteries. *Journal of Economic Perspectives*, 2:161–174, 1988.
[34] L. Utkin. A framework for imprecise robust one-class classification models. *Journal of Machine Learning Research and Cybernetics*, 5(3):379–393, 2014.
[35] L. Utkin and A. Wiencierz. Improving overfitting in ensemble regression by imprecise probabilities. *Information Sciences*, 317:315–328, 2015.
[36] P. Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, Coventry, 1982. Statistics Research Report 23.
[37] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

(I.Montes) University of Oviedo, Dep. of Statistics and Operations Research. ORCID:0000-0001-6534-1613

Email address: `imontes@uniovi.es`

(E. Miranda) University of Oviedo, Dep. of Statistics and Operations Research. ORCID:0000-0001-7763-3779

Email address: `mirandaenrique@uniovi.es`

(S. Destercke) Sorbonnes Universite, Universite de Technologie de Compiègne. ORCID: 0000-0003-2026-468X

Email address: `sebastien.destercke@hds.utc.fr`