



Unifying neighbourhood and distortion models: part I - new results on old models

Ignacio Montes, Enrique Miranda, Sébastien Destercke

► To cite this version:

Ignacio Montes, Enrique Miranda, Sébastien Destercke. Unifying neighbourhood and distortion models: part I - new results on old models. *International Journal of General Systems*, 2020, 49 (6), pp.602-635. 10.1080/03081079.2020.1778682 . hal-02944680

HAL Id: hal-02944680

<https://hal.science/hal-02944680>

Submitted on 21 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIFYING NEIGHBOURHOOD AND DISTORTION MODELS: PART I - NEW RESULTS ON OLD MODELS

IGNACIO MONTES, ENRIQUE MIRANDA, AND SÉBASTIEN DESTERCKE

ABSTRACT. Neighbourhoods of precise probabilities are instrumental to perform robustness analysis, as they rely on very few parameters. Many such models, sometimes referred to as *distortion* models, have been proposed in the literature, such as the pari mutuel model, the linear vacuous mixtures or the constant odds ratio model. This paper is the first part of a two paper series where we study the sets of probabilities induced by such models, regarding them as neighbourhoods defined over specific metrics or premetrics. We also compare them in terms of a number of properties: precision, number of extreme points, n-monotonicity, behaviour under conditioning, etc. This first part tackles this study on some of the most popular distortion models in the literature, while the second part studies less known neighbourhood models and summarises our findings.

Keywords: Neighbourhood models, distorted probabilities, pari mutuel model, linear vacuous mixtures, constant odds ratio.

1. INTRODUCTION

The most popular way to model uncertainty about some quantity or about the outcome of an experiment is by means of probability measures. However, their estimation is not always accurate nor reliable, as they may be obtained from data whose quality and quantity vary (one may think of different levels of noise, of imbalanced data sets, . . .), or from experts whose judgements we trust only partially. As some inferences and decisions drawn from such probabilities may be sensitive to small changes of the estimated probabilities, being able to perform robustness analysis is essential in critical applications. One way to perform such an analysis is to explore in a principled way neighbourhoods around the precise estimates, by defining what is usually referred to as a *distortion* model. These have been applied for example in the analysis of graphical models [9, 14], in reinforcement learning [12] or in regression problems [30]. This paper is the first of a two-paper study that offers a unified view of such models and performs a systematic study of some of their properties.

There are two basic procedures to determine a distortion model:

- The first one is to take the set of probabilities that are close to the original precise one under some criteria. We may for instance consider convex combinations with a set of probability measures where the weights represent the amount of conflicting data; assume that there is some bounded error when reporting the values of the probability distribution; or simply work with those measures at a given distance from the original model. In all those cases, we end up with a set of probability measures, usually called *neighbourhood model* [13, 16, 25, 26, 29]. Under some conditions, this set is

equivalent to the functional that we obtain by taking lower envelopes. This latter functional is a non-additive measure that is a coherent lower probability in the sense of Walley [32], and that is usually easier to handle than the dominating set of probabilities. This method also connects distortion models with the theory of imprecise probabilities.

- The second procedure consists in directly transforming the initial probability measure by means of a function on its set of values [3, 4, 5, 27, 33], [that satisfies some properties such as being increasing and/or convex](#). This procedure also produces a lower probability, but the interpretation of the resulting function may be more involved.

By means of these two procedures, a number of distortion models have been proposed in the literature [2, 7, 15, 32], such as linear vacuous mixtures, the pari mutuel model, etc. However, the problem of choosing an appropriate distortion model for a given situation is still somewhat unresolved. Our goal in this study is to contribute to the solution by analysing and comparing a number of distortion models from the literature. We shall first of all show that those distortion models obtained from the second procedure can be represented by means of the first procedure, thus unifying the two approaches. Then, we will compare a few models by means of: (a) the amount of imprecision they introduce in the original model for a fixed distortion factor; (b) the features of the distorting function that characterises the neighbourhood model; (c) the properties of the coherent lower probabilities they determine; (d) the complexity of their associated sets of probabilities, in terms of their number of extreme points; and (e) their behaviour under [conditioning](#). Throughout, we focus on neighbourhood models that induce polytopes in the space of probability distributions, meaning for instance that distortion models induced by the Euclidean norm or the Kullback-Leibler divergence [12] are out of the scope of the present study. [In particular, the Kullback-Leibler divergence is one the usual information measures in probability theory and satisfies some reasonable properties \(see for instance \[8\]\). It has already been used to create a neighbourhood around a fixed probability or a fixed credal set in \[23\], where the author refers to this model as the *discounting probability* or *discounting credal set*.](#)

This first paper will focus on models that are the most well-known in the literature: the pari mutuel, the linear vacuous mixture and the constant odds ratio models. We will see how they fit within our unified view, and establish new properties for them. The accompanying paper [22] will focus on models that have not been studied extensively, and that could benefit from a full systematic study: these are the models induced by the total variation, the Kolmogorov and the L_1 distances. This second paper will also wrap up the complete study in a synthetic comparison of the different models studied in both papers.

This first part of our study is organised as follows: in Section 2, we introduce some notions from imprecise probability theory that are necessary to follow the rest of the paper. This includes both their basic models as well as reminders [about how to update them under the light of new information](#). Section 3 introduces distortion models with a unified view, seeing them as neighbourhoods induced by a maximal distance from an initial probability distribution.

We then (re-)explore some well-known models from the literature under this light, and also establish new results for them. The pari mutuel, linear vacuous and constant odds ratio are respectively explored in Sections 4, 5 and 6. Some

preliminary conclusions are given in Section 7, but for a final synthesis and comparative study, we refer the reader to the second paper of this study [22], where we shall study models that have received less attention and provide the aforementioned synthesis. To ease the reading, proofs have been relegated to an Appendix.

A preliminary version of our study was presented in ISIPTA Conference in 2019 [20]. These two papers extend this preliminary version with additional comments, proofs, a study of the distortion model associated with the L_1 distance and a deeper comparison between the models.

2. PRELIMINARY CONCEPTS

This section introduces the main models within the theory of imprecise probabilities that we shall consider in this paper, as well as the operation of conditioning.

2.1. Imprecise probability models. Let $\mathcal{X} = \{x_1, \dots, x_n\}$ be a finite possibility space with cardinality n . We shall assume throughout that $n \geq 2$; otherwise the results are trivial. We shall use $A \subseteq \mathcal{X}$ to denote that A is a subset of \mathcal{X} and $A \subset \mathcal{X}$ to denote that A is a proper subset of \mathcal{X} . We denote by $\mathbb{P}(\mathcal{X})$ the set of all the probability measures defined on the power set of \mathcal{X} , $\mathcal{P}(\mathcal{X})$, and by $\mathbb{P}^*(\mathcal{X})$ the set of probabilities P satisfying $P(A) > 0$ for every non-empty $A \subseteq \mathcal{X}$. A set of probability measures defined on $\mathcal{P}(\mathcal{X})$ is called a *credal set* on \mathcal{X} [17]. All neighbourhood models explored in this paper will be representable in terms of credal sets.

By taking its lower envelope over events, a credal set \mathcal{M} determines a *lower probability* \underline{P} :

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

This lower probability satisfies the property of *coherence* [32], meaning that the lower probability of any event is attained by a probability measure that dominates \underline{P} . In addition to this interpretation, it can also be given a *behavioural* interpretation: $\underline{P}(A)$ is then regarded as the supremum acceptable buying price for the gamble that gives reward 1 if A happens and 0 otherwise, and coherence means that our supremum acceptable buying prices cannot be combined in order to make us subject to a sure loss and that they are as tight as possible; see [18, 32] for more details.

If instead of lower envelopes we take upper envelopes we obtain a *coherent upper probability*, that we shall denote \overline{P} . Both of them are related by conjugacy: $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \mathcal{X}$, and for this reason it suffices to focus on one of them; here we shall concentrate on lower probabilities.

Different credal sets $\mathcal{M}_1, \mathcal{M}_2$ may determine the same coherent lower probability \underline{P} by taking lower envelopes; the largest of them is given by

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

This set is usually referred to as the *credal set associated with \underline{P}* , and it is denoted by $\mathcal{M}(\underline{P})$. It is convex and closed in the weak-* topology, that coincides with the Euclidean one in the finite-dimensional case we are considering in this paper; as a consequence, we obtain $\underline{P}(A) = \min\{P(A) : P \in \mathcal{M}(\underline{P})\}$ for every $A \subseteq \mathcal{X}$. Since $\mathcal{M}(\underline{P})$ is closed and convex, it is characterised by its extreme points, i.e., by those probability measures $P \in \mathcal{M}(\underline{P})$ such that if there exist $\alpha \in (0, 1)$ and $P_1, P_2 \in \mathcal{M}(\underline{P})$ for which $P(A) = \alpha P_1(A) + (1 - \alpha)P_2(A)$ for every $A \subseteq \mathcal{X}$, then $P_1 = P_2 = P$.

Some coherent lower probabilities satisfy additional mathematical properties that make them more suitable for practical purposes. One such property is k -monotonicity. Given a natural number $k \geq 2$, a coherent lower probability is k -monotone [6] if

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

for every $1 \leq p \leq k$ and every $A_1, \dots, A_p \subseteq \mathcal{X}$.

We shall focus here on two extreme cases: if $k = 2$ we say that \underline{P} is a *2-monotone* lower probability; and if \underline{P} is k -monotone for every natural number k , we say that it is *completely monotone*, also called *belief function* [28] in the framework of Evidence Theory. One interesting property of 2-monotone lower probabilities is that the set $\text{ext}(\mathcal{M}(\underline{P}))$ of extreme points of $\mathcal{M}(\underline{P})$ is determined by the set S_n of permutations of the possibility space: for any such permutation $\sigma \in S_n$, define the probability measure P_σ by means of

$$\begin{aligned} P_\sigma(\{x_{\sigma(1)}\}) &= \underline{P}(\{x_{\sigma(1)}\}) \quad \text{and} \\ P_\sigma(\{x_{\sigma(k)}\}) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k)}\}) - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(k-1)}\}) \quad \forall k = 2, \dots, n. \end{aligned} \quad (2)$$

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

A particular instance of 2-monotone lower probabilities that we will use later on are *probability intervals* [10], that are uniquely determined by their restriction to singletons: \underline{P} is a probability interval when it satisfies

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x\}) \in [\underline{P}(\{x\}), \overline{P}(\{x\})] \quad \forall x \in \mathcal{X}\}.$$

Probability intervals are quite easy to handle, and the expression of their associated lower and upper probabilities can be easily given in terms of the lower and upper probabilities of the singletons (see for instance [10, Prop.4]).

On the other hand, a particular instance of completely monotone lower probability is the *vacuous lower probability*,

$$\underline{P}_v(A) = \begin{cases} 0, & \text{if } A \neq \mathcal{X}. \\ 1, & \text{if } A = \mathcal{X}. \end{cases} \quad (3)$$

The name vacuous comes from the fact that $\mathcal{M}(\underline{P}_v) = \mathbb{P}(\mathcal{X})$.

Since every probability measure P on $\mathcal{P}(\mathcal{X})$ is equivalent to its expectation operator, that for simplicity we denote also with the symbol P , we can use credal sets to obtain lower and upper expectation operators. Consider a function¹ $f : \mathcal{X} \rightarrow \mathbb{R}$. Then given a credal set \mathcal{M} , we get

$$\underline{P}(f) := \inf\{P(f) \mid P \in \mathcal{M}\} \quad \text{and} \quad \overline{P}(f) := \sup\{P(f) \mid P \in \mathcal{M}\}.$$

These two functionals are *coherent lower and upper previsions* in the theory of Walley, and are related by $\underline{P}(f) = -\overline{P}(-f)$ for any function f on \mathcal{X} . As in the

¹Walley's theory deals only with bounded random variables on \mathcal{X} , also called *gambles*; since in this paper we work with finite spaces, the restriction of boundedness is trivially satisfied.

case of lower probabilities, we will denote by $\mathcal{M}(\underline{P})$ the *credal set* associated with \underline{P} , defined by:

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}(f) \quad \forall f : \mathcal{X} \rightarrow \mathbb{R}\}. \quad (4)$$

We see then that a credal set \mathcal{M} can be used to determine both a coherent lower probability \underline{P} and a coherent lower prevision \underline{P}' . However, the sets

$$\begin{aligned} \mathcal{M}(\underline{P}) &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq \underline{P}(A) \quad \forall A \subseteq \mathcal{X}\} \text{ and} \\ \mathcal{M}(\underline{P}') &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}'(f) \quad \forall f : \mathcal{X} \rightarrow \mathbb{R}\} \end{aligned}$$

do not coincide in general, the second being smaller, or more informative: $\mathcal{M}(\underline{P}') \subseteq \mathcal{M}(\underline{P})$. A sufficient condition for their equality is that

$$\underline{P}'(f \wedge g) + \underline{P}'(f \vee g) \geq \underline{P}'(f) + \underline{P}'(g) \quad (5)$$

for every pair of gambles $f, g : \mathcal{X} \rightarrow \mathbb{R}$, where \wedge denotes the pointwise minimum and \vee the pointwise maximum. When a coherent lower prevision \underline{P}' satisfies Equation (5), it is called 2-monotone lower prevision [31]. In that case, its restriction to events $\underline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ given by $\underline{P}(A) := \underline{P}'(I_A)$ for every $A \subseteq \mathcal{X}$, is a 2-monotone lower probability. Moreover, both \underline{P} and \underline{P}' incude the same credal set: $\mathcal{M}(\underline{P}) = \mathcal{M}(\underline{P}')$. Indeed, if \underline{P} is a 2-monotone lower probability, it has a unique extension to gambles \underline{P}' satisfying the 2-monotonicity property in Equation (5): it corresponds to the Choquet integral with respect to \underline{P} [31].

Since, as we have mentioned, probability intervals are particular instances of 2-monotone lower probabilities, we deduce that they also have a unique extension to gambles that satisfies Equation (5). Its expression was determined in [10, Sec. 4].

Figure 1 summarises the relationships between the different imprecise probability models we have introduced; an arrow between two nodes means that the parent is a particular case of the child.

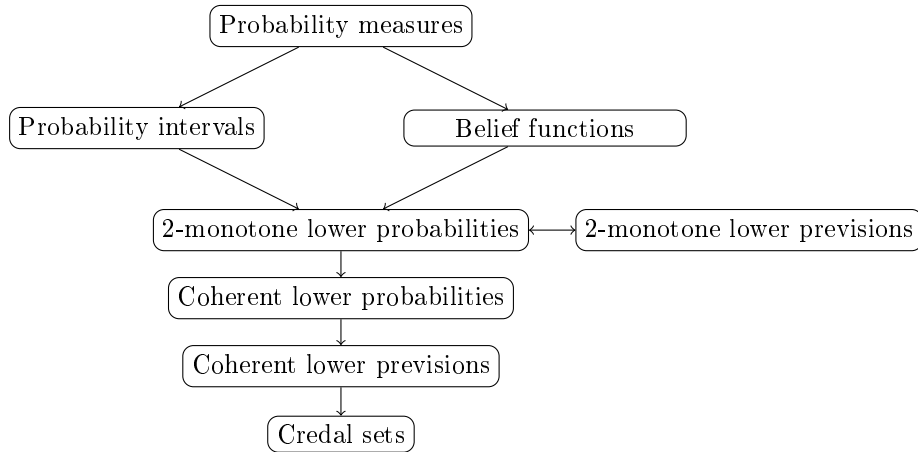


FIGURE 1. Relationships between different models.

2.2. Conditioning imprecise probability models. Once we get a credal set \mathcal{M} or its lower prevision \underline{P} over \mathcal{X} , we may wish to condition it upon the arrival of an observation B , obtaining thus an updated model \underline{P}_B . With precise probabilities and provided that $P(B) > 0$, there is essentially one conditioning rule: defining the updated model as $P_B(A) = P(A \cap B)/P(B)$, that also uniquely defines the associated expectation operator P_B .

Unfortunately, the extension of this conditioning operator is not unique in general when considering sets of probabilities [19, 32]. However, when the lower probability $\underline{P}(B)$ of the conditioning event is strictly positive (as will be the case here, see Equation (12) below), there is only one updating procedure that is consistent with Walley's behavioural notion of coherence: the *regular extension*, that consists in applying Bayes' rule to all the probability measures that dominate \underline{P} , and then take the lower envelope of the updated models. That is, the updated model \underline{P}_B is defined for any function f on \mathcal{X} as

$$\underline{P}_B(f) = \inf\{P_B(f) \mid P \in \mathcal{M}(\underline{P})\}. \quad (6)$$

This coincides with the *Generalised Bayes Rule* in [32] under the assumption of positive lower probabilities.

In general, updating through regular extension preserves many of the properties of the initial models, such as n-monotonicity. We will see that this is also the case for the distortion models considered in this first part of our study.

3. DISTORTION MODELS: GENERAL SETTING

The neighbourhood models we shall consider in this study originate by distorting a probability measure P_0 by means of some function d , with a fixed distorting factor $\delta > 0$. Part of the comparison between them shall be made in terms of the properties satisfied by the distorting function.

Given the set $\mathbb{P}(\mathcal{X})$, a function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow [0, \infty)$ will be called *distorting function*. We consider the following desirable properties that a distorting function may satisfy:

- Ax.1:** Positive definiteness: $d(P_1, P_2) = 0$ if and only if $P_1 = P_2$.
- Ax.1a:** $d(P_1, P_2) = 0$ implies that $P_1 = P_2$, for every $P_1, P_2 \in \mathbb{P}(\mathcal{X})$.
- Ax.1b:** Well-defined: $d(P, P) = 0$ for every $P \in \mathbb{P}(\mathcal{X})$.
- Ax.2:** Triangle inequality: $d(P_1, P_3) \leq d(P_1, P_2) + d(P_2, P_3)$ for every $P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.
- Ax.3:** Symmetry: $d(P_1, P_2) = d(P_2, P_1)$ for every $P_1, P_2 \in \mathbb{P}(\mathcal{X})$.
- Ax.4:** Convexity: $d(P_1, \alpha P_2 + (1 - \alpha)P_3) \leq \max\{d(P_1, P_2), d(P_1, P_3)\}$ for every $\alpha \in [0, 1]$ and every $P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.
- Ax.5:** Continuity: for every $P, P_1, P_2 \in \mathbb{P}$ and every $\epsilon > 0$, $\exists \delta > 0$ such that $\|P_1 - P_2\| < \delta$ implies that $|d(P_1, P) - d(P_2, P)| < \epsilon$, where $\|\cdot\|$ denotes the supremum norm.

Since the possibility space \mathcal{X} is finite, the topology associated with the supremum norm $\|\cdot\|$ is equivalent to that generated by the Euclidean distance. Thus, Ax.5 can be equivalently stated as

$$(P_n)_n \rightarrow P \Rightarrow d(P_n, P) \rightarrow 0,$$

where we are identifying probability measures with elements of the n-th dimensional Euclidean space, and convergence of probability measures is understood pointwise.

The function d is a *distance* when it satisfies Ax.1–Ax.3, and it is a *premetric* when it satisfies axiom Ax.1b. Also, note that Ax.1 holds if and only if both Ax.1a and Ax.1b are satisfied.

Given a distorting function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow [0, \infty)$, a probability measure $P_0 \in \mathbb{P}(\mathcal{X})$ and a parameter $\delta > 0$, we can consider the credal set

$$B_d^\delta(P_0) := \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\} \quad (7)$$

of those probability measures that differ at most in δ from P_0 , i.e., the closed ball centred in P_0 and with distortion parameter δ with respect to the distorting function d . We call it the *distortion* model on P_0 associated with the distorting function d and the factor $\delta > 0$. It can be used to determine the coherent lower prevision

$$\underline{P}(f) = \inf\{P(f) \mid P \in B_d^\delta(P_0)\} = \inf\{P(f) \mid d(P, P_0) \leq \delta\} \quad \forall f : \mathcal{X} \rightarrow \mathbb{R}, \quad (8)$$

whose associated credal set $\mathcal{M}(\underline{P})$ is given by Equation (4).

Our first result gives sufficient conditions for $B_d^\delta(P_0)$ to coincide with the credal set $\mathcal{M}(\underline{P})$ associated with the coherent lower prevision \underline{P} .

Proposition 1. *Consider a distorting function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow [0, \infty)$, a probability measure $P_0 \in \mathbb{P}(\mathcal{X})$ and a parameter $\delta > 0$, and let \underline{P} be the lower envelope of $B_d^\delta(P_0)$, given by Equation (8). If d satisfies Ax.4 and Ax.5, then $\mathcal{M}(\underline{P}) = B_d^\delta(P_0)$, where $\mathcal{M}(\underline{P})$ is the credal set induced by \underline{P} using Equation (4).*

In the remainder of this study, we shall investigate in detail the properties of a number of distortion models. Before we tackle this problem, we would like to study in some detail the second procedure for generating distortion models we mentioned in the introduction: sometimes in the literature [5] the term *distortion* model is referred to a non-additive measure of the type $\underline{P} = h(P_0)$, where $h : [0, 1] \rightarrow [0, 1]$ is a suitable distorting function that is increasing and satisfies $h(0) = 0, h(1) = 1$. In this sense, if h is convex, then $\underline{P} = h(P_0)$ is 2-monotone [6, 11], and in [3, Thm. 7] it is shown that the condition $h(t) \leq t$ for every $t \in [0, 1]$ guarantees that \underline{P} is a *lower probability*. Some other results in this direction can be found in [3, 4], in connection with the theory of aggregation operators; see also [33] and [5, Prop. 2]. Next we prove that such models can be incorporated into our formalism:

Proposition 2. *Let P_0 be a probability measure on $\mathbb{P}(\mathcal{X})$. Consider an increasing function $h : [0, 1] \rightarrow [0, 1]$ such that $h(0) = 0, h(1) = 1$ and $h(t) \leq t$ for every $t \in [0, 1]$, and let $\underline{P} = h(P_0)$ be the transformed model. Then there exist a premetric d and some parameter $\delta > 0$ such that $\mathcal{M}(\underline{P}) = B_d^\delta(P_0)$, where $B_d^\delta(P_0)$ is given by Equation (7) and $\mathcal{M}(\underline{P})$ is determined by Equation (4).*

The idea in the proof is to define, for the given h , the function $d_h^* : [0, 1] \times [0, 1] \rightarrow [0, 1]$ by

$$d_h^*(x, y) = \begin{cases} 0 & \text{if } x = y \\ 0.5 & \text{if } x \neq y, x \geq h(y) \\ 1 & \text{if } x \neq y, x < h(y), \end{cases} \quad (9)$$

and then let $d(P, Q) = \max_{A \subseteq \mathcal{X}} d_h^*(P(A), Q(A))$.

It can be checked that this function d is not necessarily a distance, because it may not satisfy axiom Ax.3 (symmetry). Furthermore, in addition to Ax.1b, such d also satisfies Ax.1a (hence also Ax.1), Ax.2 and Ax.4. The proof of this proposition

also shows that the implication

$$d(P, Q) = \max_{A \subseteq \mathcal{X}} d_h^*(P(A), Q(A)) \Rightarrow \underline{P} = \min B_d^\delta(P_0) \text{ is 2-monotone}$$

does not hold in general. To see that, it suffices to consider a function h and a probability P_0 such that $\underline{P} = h(P_0)$ is not 2-monotone and apply to them the distorting function d above.

As we said before, some earlier works in the literature have characterised those distortion models that are 2-monotone or a coherent lower probability [5, 33]. Our next result characterises those distortion models whose associated lower probabilities are probability intervals.

Proposition 3. *Let d be a distorting function satisfying Ax.4 and Ax.5. Then, the lower envelope \underline{P} of $B_d^\delta(P_0)$, given by Equation (8), is a probability interval for every $P_0 \in \mathbb{P}(\mathcal{X})$ and every $\delta > 0$ if and only if d satisfies:*

$$d(P, Q) = \max_{i=1, \dots, n} d_i(P(\{x_i\}), Q) \quad \forall P, Q \in \mathbb{P}(\mathcal{X}), \quad (10)$$

where for every $i = 1, \dots, n$

$$d_i(P(\{x_i\}), Q) = \inf\{d(P', Q) \mid P' \in \mathbb{P}(\mathcal{X}), P'(\{x_i\}) = P(\{x_i\})\}.$$

This result leads us to study those comparison measures between probability measures that are defined in terms of other comparison measures on real numbers. We have established the following:

Proposition 4. *Let $d^* : [0, 1] \times [0, 1] \rightarrow [0, \infty)$, and let us define $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow [0, \infty)$ by*

$$d(P, Q) = \max_{i=1, \dots, n} d^*(P(\{x_i\}), Q(\{x_i\})). \quad (11)$$

- (a) *If d^* satisfies Ax.j ($j=1a, 1b, 2, 3, 4, 5$) then so does d .*
- (b) *If d^* satisfies Ax.4 and Ax.5 and d is defined by Equation (11), then for any probability measure $P_0 \in \mathbb{P}(\mathcal{X})$ and any $\delta > 0$, the lower envelope on events of the set $B_d^\delta(P_0)$ given by Equation (7) is a probability interval.*

These results nicely characterise general distortion models, as they allow to connect generic properties of the used distance with corresponding features of the induced credal sets. They will also prove useful later on in this paper.

In the following sections, we shall study some of the most important distortion models from the literature, revisit them from our unifying perspective introduced in this section, and provide new results about them. We shall assume throughout that the probability measure P_0 belongs to the interior $\mathbb{P}^*(\mathcal{X})$ of $\mathbb{P}(\mathcal{X})$, or in other words, that $P_0(\{x\}) > 0$ for every $x \in \mathcal{X}$, and also that the distortion factor $\delta > 0$ is small enough so that also

$$Q(\{x\}) > 0 \quad \forall x \in \mathcal{X}, \forall Q \in B_d^\delta(P_0). \quad (12)$$

Hence, we will consider that the distorting function d is defined from $\mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X})$. Since in all cases d will satisfy Ax.4 and Ax.5, Proposition 1 implies that $B_d^\delta(P_0)$ is closed and that $\underline{P}(B) > 0$ for every $B \subseteq \mathcal{X}$.

4. PARI MUTUEL MODEL

Our analysis begins with the *pari mutuel model* (PMM, for short).

Definition 1. *Given a probability measure P_0 and a distortion parameter $\delta > 0$, the associated pari mutuel model is given by the conjugate coherent lower and upper probabilities:*

$\underline{P}_{PMM}(A) = \max\{0, (1 + \delta)P_0(A) - \delta\}$ and $\overline{P}_{PMM}(A) = \min\{1, (1 + \delta)P_0(A)\}$ for every $A \subseteq \mathcal{X}$.

In what follows, we investigate the properties of the PMM, we show that it can be expressed as a distortion model and we study its behaviour under conditioning.

4.1. Properties of the PMM. The pari mutuel model has its origins as a betting scheme in horse racing; if we interpret the distortion factor δ as a taxation from the house, then $\underline{P}_{PMM}, \overline{P}_{PMM}$ can be given a behavioural interpretation as betting rates for and against the event A , as discussed in some depth in [32, Sec. 2.9.3]. Assume that $P_0(A)$ denotes the fair price fixed by a house for event A . This means that if a gambler bets on A at the fair price $P_0(A)$, her expected gain would be:

$$E(I_A - P_0(A)) = (1 - P_0(A))P_0(A) - P_0(A)(1 - P_0(A)) = 0.$$

In a similar manner, the expected gain of the house would be $E(P_0 - I_A) = 0$.

In order to ensure a positive gain, the house can increase the fair price for A : $(1 + \delta)P_0(A)$ for some positive δ . In this way, the expected gain for a gambler would be:

$$\begin{aligned} E(I_A - (1 + \delta)P_0(A)) &= \\ (1 - (1 + \delta)P_0(A))P_0(A) + (-(1 + \delta)P_0(A))(1 - P_0(A)) &= -\delta P_0(A), \end{aligned}$$

whereas the expected gain of the house would be:

$$E((1 + \delta)P_0(A) - I_A) = \delta P_0(A).$$

This gives rise to the upper probability $\overline{P}(A) = (1 + \delta)P_0(A)$, that is corrected as $\overline{P}(A) = \min\{1, (1 + \delta)P_0(A)\}$ in order to guarantee coherence. This means that, for the pari mutuel model, the distortion parameter δ represents the selling price inflation rate for A .

Walley [32, Sec. 2.9.3] deduces this upper probability using a different reasoning related to racetracks. If $P_0(A)$ is the proportion of stakes betting for A , and τ (with $0 < \tau < 1$) denotes the proportion of stakes deducted by the house, a unit bet on A will return $\frac{1-\tau}{P_0(A)}$ if A occurs. Then, the selling price for A should be $\frac{P_0(A)}{1-\tau} = (1 + \delta)P_0(A)$, where $\delta = \frac{\tau}{1-\tau}$. We refer to [24] and [32, Sec. 2.9.3] for more details on the behavioural interpretation of this model.

The lower probability \underline{P}_{PMM} determines a credal set by:

$$\mathcal{M}(\underline{P}_{PMM}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq \max\{0, (1 + \delta)P_0(A) - \delta\} \ \forall A \subseteq \mathcal{X}\}. \quad (13)$$

We refer to [21, 24, 32] for a study of the PMM from the point of view of imprecise probabilities. In particular, it was established in [21, Thm. 1] that the PMM is a particular case of probability interval, meaning that it is determined by its restriction to singletons:

$$\mathcal{M}(\underline{P}_{PMM}) = \{P \in \mathbb{P}(\mathcal{X}) \mid (1 + \delta)P_0(\{x\}) - \delta \leq P(\{x\}) \leq (1 + \delta)P_0(\{x\}), \ \forall x \in \mathcal{X}\}.$$

As every probability interval, the coherent lower probability \underline{P}_{PMM} is 2-monotone, and as a consequence, its extension to gambles, preserving 2-monotonicity, is unique and it is given by the Choquet integral: if f denotes a gamble and for any $\tau \in (0, 1)$ x_τ and $x_{1-\tau}$ denote the τ -quantile and $1 - \tau$ -quantile of f , given by:

$$x_\tau = \sup\{x \in \mathcal{X} \mid P_0(f \leq x) \leq \tau\}, \quad x_{1-\tau} = \sup\{x \in \mathcal{X} \mid P_0(f \leq x) \leq 1 - \tau\},$$

the lower and upper prevision of f for the PMM induced by P_0 and δ are [32, Sec. 3.2.5]:

$$\begin{aligned} \underline{P}_{PMM}(f) &= x_{\frac{1}{1+\delta}} - (1 + \delta)P_0((x_{\frac{1}{1+\delta}} - f)^+), \\ \overline{P}_{PMM}(f) &= x_{\frac{\delta}{1+\delta}} + (1 + \delta)P_0((f - x_{\frac{\delta}{1+\delta}})^+), \end{aligned}$$

where for any $g : \mathcal{X} \rightarrow \mathbb{R}$ we denote $g^+ = \max\{g, 0\}$.

It was also shown that the maximal number of extreme points of $B_{d_{PMM}}^\delta(P_0)$ is the same as for probability intervals [21, Prop. 1].

We are assuming that $P_0 \in \mathbb{P}(\mathcal{X})$ and $\underline{P}(\{x\}) > 0$ for every $x \in \mathcal{X}$, hence $P(A) > 0$ for every $A \neq \emptyset$. This means that:

$$\underline{P}_{PMM}(A) > 0 \Leftrightarrow \frac{P_0(A)}{1 - P_0(A)} > \delta \quad \forall A \neq \emptyset, \mathcal{X}$$

which in particular implies that

$$\delta < \min_{A \neq \emptyset, \mathcal{X}} \frac{P_0(A)}{1 - P_0(A)}. \quad (14)$$

Let us show that $\underline{P}_{PMM}, \overline{P}_{PMM}$ can be obtained as envelopes of the credal set determined by some distorting function, in the manner we introduced in Section 3.

Theorem 5. *Consider the pari mutuel model associated with a probability measure P_0 and a distortion factor $\delta > 0$. Then $\mathcal{M}(\underline{P}_{PMM}) = B_{d_{PMM}}^\delta(P_0)$, where the former credal set is given by Equation (13), $d_{PMM} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ is given by*

$$d_{PMM}(P, Q) = \max_{A \subset \mathcal{X}} \frac{Q(A) - P(A)}{1 - Q(A)} \quad (15)$$

and $B_{d_{PMM}}^\delta(P_0)$ is given by Equation (7).

Note that, from Equation (12), $B_{d_{PMM}}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, whence $Q(A) < 1$ for every $A \subset \mathcal{X}$ and every $Q \in B_{d_{PMM}}^\delta(P_0)$. Thus, Equation (15) is well-defined.

From this theorem we deduce that the pari mutuel model can be embedded into our framework of neighbourhood models. Also, this result motivates the study of the properties of the function d_{PMM} . These are summarised in the following proposition:

Proposition 6. *Let d_{PMM} be the function defined in Equation (15).*

- (a) d_{PMM} satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.4 and Ax.5.
- (b) d_{PMM} does not satisfy in general axioms Ax.2 nor Ax.3.
- (c) For every $P, Q \in \mathbb{P}^*(\mathcal{X})$, $d_{PMM}(P, Q) = \max_{x \in \mathcal{X}} \frac{P(\{x\}) - Q(\{x\})}{Q(\{x\})}$.

The fact that the pari mutuel model induces a probability interval can also be established showing that the distorting function d_{PMM} in Equation (15) satisfies Equation (10), using its equivalent expression in Proposition 6, or also using Proposition 4. For this aim, just consider item (c) in Proposition 6 and note that d_{PMM}

can be expressed as $d_{PMM}(P, Q) = \max_{x \in \mathcal{X}} d_{PMM}^*(P(\{x\}), Q(\{x\}))$, where d_{PMM}^* is given, for every $a, b \in (0, 1)$, by:

$$d_{PMM}^*(a, b) = \begin{cases} \frac{a-b}{b}, & \text{if } a \geq b, \\ 0 & \text{if } a < b. \end{cases}$$

In the case of the pari mutuel model, if the distortion factor δ satisfies Equation (14), then this maximum distortion is attained within the ball $B_{d_{PMM}}^\delta(P_0)$:

Proposition 7. *Consider the pari mutuel model associated with a probability measure P_0 and a distortion factor $\delta > 0$ satisfying Equation (14). Then:*

$$\max_{P \in B_{d_{PMM}}^\delta(P_0)} d(P, P_0) = \delta.$$

This does not hold for all values of δ , because given a probability measure P_0 , the set $\{d_{PMM}(P_0, Q) : Q \in \mathbb{P}(\mathcal{X})\}$ is bounded above. It is not difficult to see that if we take $\delta = \max_{A \subset \mathcal{X}} \frac{P_0(A)}{1-P_0(A)}$, we obtain that $B_{d_{PMM}}^\delta(P_0) = \mathbb{P}(\mathcal{X})$, or, in other words, that \underline{P}_{PMM} is the vacuous lower probability given by Equation (3). Therefore, larger values of δ will not give any additional information. Due to monotonicity of P_0 and Equation (12), it holds that

$$\max_{A \subset \mathcal{X}} \frac{P_0(A)}{1-P_0(A)} = \max_{x_i \in \mathcal{X}} \frac{P_0(\mathcal{X} \setminus \{x_i\})}{P_0(\{x_i\})}.$$

4.2. Conditioning the PMM. Next, we study the behaviour of the PMM under conditioning. We show that conditioning through regular extension using Equation (6) again yields a pari mutuel model.

Proposition 8. *Consider the model $B_{d_{PMM}}^\delta(P_0)$ and its induced lower probability \underline{P}_{PMM} . Then, for every event B the conditional model \underline{P}_B induces the credal set $B_{d_{PMM}}^{\delta_B}(P_{0|B})$ such that*

$$P_{0|B}(A) = P_0(A|B) \quad \text{and} \quad \delta_B = \frac{\delta}{(1+\delta)P_0(B) - \delta} = \frac{\delta}{\underline{P}_{PMM}(B)} \quad (16)$$

with $P_0(A|B)$ the conditional precise probability derived from P_0 .

We see that the imprecision of the pari mutuel model, related to the distortion factor, increases when conditioning.

5. LINEAR VACUOUS MODEL

Another popular neighbourhood model are *linear-vacuous* mixtures, also called *ϵ -contamination* models in the literature [15, 32] (for a related model starting from the distortion of belief functions, we refer to [23, 28]).

Definition 2. *Given a probability measure P_0 and a distortion parameter $\delta \in (0, 1)$, the associated linear vacuous mixture is given by the conjugate coherent lower and upper probabilities:*

$$\begin{aligned} \underline{P}_{LV}(A) &= \begin{cases} (1-\delta)P_0(A) & \text{if } A \neq \mathcal{X}. \\ 1 & \text{if } A = \mathcal{X}. \end{cases} \\ \overline{P}_{LV}(A) &= \begin{cases} (1-\delta)P_0(A) + \delta & \text{if } A \neq \emptyset. \\ 0 & \text{if } A = \emptyset. \end{cases} \end{aligned}$$

5.1. Properties of the LV model. The behavioural interpretation of linear vacuous mixtures is the following: if t is the price that the gambler pays for a bet on the occurrence of event A , her benefit is $I_A - t$. To guarantee a positive gain, the house can impose a tax δ that is proportional to the reward. Then, gambler's benefit becomes $I_A - t - \delta I_A = (1 - \delta)I_A - t$. If $P_0(A)$ denotes the probability of occurrence of A , the expected benefit becomes $(1 - \delta)P_0(A) - t$. Then the fair price is reduced to $(1 - \delta)P_0(A)$, that corresponds to the lower probability of the linear vacuous model. Hence, the distortion parameter δ can be understood as the deflation rate on the buying price on A . We refer to [32, Sec.2.9.2] for more details about the behavioural interpretation of this model.

The associated credal set is:

$$\mathcal{M}(\underline{P}_{LV}) = \{P \in \mathbb{P}(\mathcal{X}) \mid (1 - \delta)P_0(A) \leq P(A), \forall A \subseteq \mathcal{X}\}. \quad (17)$$

The lower probability \underline{P}_{LV} associated with a linear vacuous model is a belief function, because it is a convex combination of two completely monotone lower probabilities: the probability measure P_0 and the vacuous lower probability \underline{P}_v given by Equation (3). In addition, it is also a probability interval: if

$$P(\{x_i\}) \geq \underline{P}_{LV}(\{x_i\}) = (1 - \delta)P_0(\{x_i\}), \quad \forall x_i \in \mathcal{X}$$

then by additivity we conclude that $P(A) \geq (1 - \delta)P_0(A) = \underline{P}_{LV}(A)$. As a consequence, $\mathcal{M}(\underline{P}_{LV})$ can be expressed as

$$\mathcal{M}(\underline{P}_{LV}) = \{P \in \mathbb{P}(\mathcal{X}) \mid (1 - \delta)P_0(\{x\}) \leq P(\{x\}) \leq (1 - \delta)P_0(\{x\}) + \delta \forall x \in \mathcal{X}\}.$$

$\mathcal{M}(\underline{P}_{LV})$ being a convex mixture between the precise distribution P_0 and the set $\mathbb{P}(\mathcal{X})$, it has $|\mathcal{X}|$ extreme points corresponding for each element $x \in \mathcal{X}$ to the distribution

$$P_x(\{x\}) = (1 - \delta)P_0(\{x\}) + \delta, \quad P_x(\{y\}) = (1 - \delta)P_0(\{y\}) \quad \forall y \neq x.$$

Since the linear vacuous model is completely monotone, it is also 2-monotone, hence its extension to gambles, preserving complete monotonicity, is unique and is given by the Choquet integral [32, Sec. 3.2.4]:

$$\begin{aligned} \underline{P}_{LV}(f) &= (1 - \delta)P_0(f) + \delta \min(f), \\ \overline{P}_{LV}(f) &= (1 - \delta)P_0(f) + \delta \max(f) \quad \forall f : \mathcal{X} \rightarrow \mathbb{R}. \end{aligned}$$

Next we establish that linear-vacuous mixtures can be obtained by means of some appropriate distorting function.

Theorem 9. *Consider the linear vacuous mixture associated with a probability measure P_0 and a distortion factor $\delta > 0$. Then $\mathcal{M}(\underline{P}_{LV}) = B_{d_{LV}}^\delta(P_0)$, where the former credal set is given by Equation (17), $d_{LV} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ is given by:*

$$d_{LV}(P, Q) = \max_{A \neq \emptyset} \frac{Q(A) - P(A)}{Q(A)} \quad (18)$$

and $B_{d_{LV}}^\delta(P_0)$ is given by Equation (7).

As for the PMM, from this theorem we deduce that linear vacuous mixtures can be seen as a neighbourhood model. Also, note that the expression in Equation (18) is well-defined because, since $Q \in \mathbb{P}^*(\mathcal{X})$, $Q(A) = 0$ if and only if $A = \emptyset$.

Remark 1. *If we compare Equations (15) and (18), we observe that the expressions of the distorting functions for the pari mutuel and the linear vacuous, d_{PMM} and d_{LV} , are quite similar:*

$$d_{PMM}(P, Q) = \max_{A \subset \mathcal{X}} \frac{Q(A) - P(A)}{1 - Q(A)}, \quad d_{LV}(P, Q) = \max_{A \neq \emptyset} \frac{Q(A) - P(A)}{Q(A)}.$$

The only difference lies in the denominator, where d_{PMM} considers $1 - Q(A)$ (or $Q(A^c)$) while d_{LV} considers $Q(A)$. Indeed, if we consider the equivalent expression of $d_{PMM}(P, Q)$ as:

$$d_{PMM}(P, Q) = \max_{A \neq \emptyset} \frac{P(A) - Q(A)}{Q(A)},$$

then d_{PMM} and d_{LV} only differ in the sign of the numerator. This sort of dual behaviour has already been observed in the literature; for instance, it was proven in [21, Lem. 1] that the upper probability of the pari mutuel is additive for those events A, B such that $\overline{P}_{PMM}(A \cup B) < 1$, while this very same property holds for the lower probability of the linear vacuous whenever $\underline{P}_{LV}(A \cup B) < 1$. Other comments in this direction were made by Walley in [32, Sec. 4.6.5]. He explains that there is a geometrical relationship between the credal sets $B_{d_{PMM}}^\delta(P_0)$ and $B_{d_{LV}}^\delta(P_0)$, with the former being a reflection of the latter through P_0 . Also, he mentions that the elicitation of a PMM is determined by inequalities of the type $P(A) \leq (1 + \delta)P_0(A)$, while that of a LV is based on inequalities of the type $P(A) \geq (1 - \delta)P_0(A)$. ♦

Our next result summarises the properties of d_{LV} .

Proposition 10. *Let d_{LV} be the function defined in Equation (18).*

- (a) *d_{LV} satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.4 and Ax.5.*
- (b) *d_{LV} does not satisfy Ax.3 in general.*
- (c) *$d_{LV}(P, Q) = \max_{x \in \mathcal{X}} \frac{Q(\{x\}) - P(\{x\})}{Q(\{x\})}$ for every $P, Q \in \mathbb{P}^*(\mathcal{X})$.*

The fact that linear vacuous mixtures induce a probability interval can also be established showing that the distorting function d_{LV} in Equation (18) satisfies Equation (10), using its equivalent expression in Proposition 10, or also using Proposition 4. For this aim, just consider item (c) in Proposition 10 and note that d_{LV} can be expressed as $d_{LV}(P, Q) = \max_{x \in \mathcal{X}} d_{LV}^*(P(\{x\}), Q(\{x\}))$, where d_{LV}^* is given, for every $a, b \in (0, 1)$, by:

$$d_{LV}^*(a, b) = \begin{cases} \frac{b-a}{a}, & \text{if } b \geq a, \\ 0 & \text{if } b < a. \end{cases}$$

We conclude this section by establishing that for every $\delta \in (0, 1)$ there is always a probability in the credal set $\mathcal{M}(\underline{P}_{LV})$ such that $d_{LV}(P, P_0) = \delta$.

Proposition 11. *Consider the linear-vacuous mixture associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$. Then*

$$\max_{P \in B_{d_{LV}}^\delta(P_0)} d_{LV}(P, P_0) = \delta.$$

5.2. Conditioning the LV model. In the case of the LV model, Walley [32, Sec. 6.6.2] showed that the conditional model \underline{P}_B induced by $B_{d_{LV}}^\delta(P_0)$ is again a linear vacuous model $B_{d_{LV}}^{\delta_B}(P_{0|B})$ such that

$$P_{0|B}(A) = P_0(A|B) \text{ and } \delta_B = \frac{\delta}{(1-\delta)P_0(B) + \delta} = \frac{\delta}{\bar{P}_{LV}(B)}. \quad (19)$$

Again, we can see that the imprecision of the model increases when conditioning; yet this increase is smaller than the one observed for the PMM (see Equation (16)).

6. CONSTANT ODDS RATIO

Next we consider the constant odds ratio model. It was given a behavioural interpretation by Peter Walley in [32, Sec. 2.9.4], and studied in [1, 2, 25, 29].

6.1. Properties of the COR model. We start explaining the behavioural interpretation of this model. [Assume that any profit made by the gambler is subject to a fixed taxation rate \$\delta\$.](#) This means that if a gambler pays a price t for a bet for the gamble f , her net gain is:

$$(1-\delta)(f-t)^+ - (f-t)^-,$$

where for any $g : \mathcal{X} \rightarrow \mathbb{R}$, $g^+ = \max\{g, 0\}$ and $g^- = \max\{-g, 0\}$. If P_0 is the fair price for the gambles, then [the expected net gain would be 0, which means that:](#)

$$(1-\delta)P_0((f-t)^+) - P_0((f-t)^-) = 0,$$

[or equivalently, that](#)

$$(1-\delta)P_0((f-t)^+) = P_0((f-t)^-).$$

As a consequence, if $\underline{P}(f)$ denotes the supremum acceptable buying price for the gamble f , then it must hold that:

$$(1-\delta)P_0((f - \underline{P}(f))^+) = P_0((f - \underline{P}(f))^-). \quad (20)$$

This equation cannot be explicitly solved for arbitrary gambles f ; however, in the particular case when f is the indicator of an event A , [Equation \(20\) becomes:](#)

$$(1-\delta)P_0((I_A - \underline{P}(I_A))^+) = P_0((I_A - \underline{P}(I_A))^-). \quad (21)$$

[It is easy to see that:](#)

$$P_0((I_A - \underline{P}(I_A))^+) = (1 - \underline{P}(A))P_0(A), \quad P_0((I_A - \underline{P}(I_A))^-) = (1 - P_0(A))\underline{P}(A).$$

[Hence, Equation \(21\) simplifies to:](#)

$$(1-\delta)(1 - \underline{P}(A))P_0(A) = (1 - P_0(A))\underline{P}(A),$$

[whence](#) we obtain the lower probability

$$\underline{P}(A) = \frac{(1-\delta)P_0(A)}{1 - \delta P_0(A)} \quad \forall A \subseteq \mathcal{X}. \quad (22)$$

In the proof of some of the forthcoming results we will use the common short notation $G(f) = f - \underline{P}(f)$ for any gamble. This notation simplifies Equation (20):

$$1 - \delta P_0(G(f)^+) = P_0(G(f)^-).$$

6.1.1. Constant odds ratio on gambles.

Definition 3. Given a probability measure P_0 and a distortion parameter $\delta \in (0, 1)$, the associated constant odds ratio model is the coherent lower prevision that is the unique solution to the equation:

$$(1 - \delta)P_0((f - \underline{P}_{COR}(f))^+) = P_0((f - \underline{P}_{COR}(f))^-). \quad (23)$$

Its conjugate coherent upper prevision $\overline{P}_{COR}(f)$ is the unique solution of the equation:

$$P_0((f - \overline{P}_{COR}(f))^+) = (1 - \delta)P_0((f - \overline{P}_{COR}(f))^-),$$

which is equivalent to:

$$\delta P_0((f - \overline{P}_{COR}(f))^+) + (1 - \delta)P_0(f - \overline{P}_{COR}(f)) = 0.$$

The reason why in this case we are first considering the coherent lower prevision instead of the coherent lower probability that is its restriction to events is that, unlike the previous two examples, the constant odds ratio model is not 2-monotone in general, as we will show in the forthcoming Example 1, while its restriction to events is. As a consequence, the credal sets determined by these two models do not coincide in general.

We will therefore make a separate study of the constant odds ratio model \underline{P}_{COR} and its restriction to events, that will be denoted by \underline{Q}_{COR} . We will start with \underline{P}_{COR} in this section, and will study \underline{Q}_{COR} in Section 6.1.2. The following proposition summarises some first properties of this model:

Proposition 12. [32] Let \underline{P}_{COR} be the constant odds ratio determined by a probability measure P_0 and a distortion factor $\delta \in (0, 1)$.

(a) The credal set associated with \underline{P}_{COR} is

$$\mathcal{M}(\underline{P}_{COR}) = \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \frac{P(A)}{P(B)} \geq (1 - \delta) \frac{P_0(A)}{P_0(B)} \quad \forall A, B \neq \emptyset \right\}. \quad (24)$$

(b) The extreme points of this credal set are $\{P_A \mid \emptyset \neq A \subseteq \mathcal{X}\}$, where P_A is given by:

$$P_A(B) = \frac{(1 - \delta)P_0(A \cap B) + P_0(A^c \cap B)}{1 - \delta P_0(A)}, \quad \forall B \subseteq \mathcal{X}. \quad (25)$$

Proposition 12 allows us to establish the following result:

Proposition 13. $\mathcal{M}(\underline{P}_{COR})$ has $2^n - 2$ different extreme points.

Let us now show that \underline{P}_{COR} can be expressed as a distortion model by means of a suitable function.

Theorem 14. Consider the constant odds ratio model (on gambles) associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$. Then $\mathcal{M}(\underline{P}_{COR}) = B_{d_{COR}}^\delta(P_0)$, where the former is given by Equation (24), $d_{COR} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ is given by:

$$d_{COR}(P, Q) = \max_{A, B \neq \emptyset} \left\{ 1 - \frac{P(A) \cdot Q(B)}{P(B) \cdot Q(A)} \right\} \quad (26)$$

and $B_{d_{COR}}^\delta(P_0)$ is given by Equation (7).

Note that the expression in Equation (26) is well-defined, because the assumption $P, Q \in \mathbb{P}^*(\mathcal{X})$ implies that $P(B) \cdot P(A) = 0$ if and only if either $A = \emptyset$ or $B = \emptyset$, and these two cases have been excluded in the expression of the maximum.

The following result summarises the properties of d_{COR} :

Proposition 15. *Let d_{COR} be the function defined in Equation (26). It satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.3, Ax.4 and Ax.5. As a consequence, it is a distance.*

Finally, we prove that we can always find a probability in $B_{d_{COR}}^\delta(P_0)$ that attains the distance δ .

Proposition 16. *Consider the constant odds ratio model (on gambles) associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$. Then:*

$$\max_{P \in B_{d_{COR}}^\delta(P_0)} d_{COR}(P, P_0) = \delta.$$

6.1.2. Constant odds ratio on events. We turn now to the restriction to events of the constant odds ratio model, \underline{Q}_{COR} . Let us show by the means of an example that \underline{P}_{COR} and \underline{Q}_{COR} do not coincide in general. First notice that, from Equation (22), the restriction over events is given by:

$$\underline{Q}_{COR}(A) = \frac{(1 - \delta)P_0(A)}{1 - \delta P_0(A)}, \quad \overline{Q}_{COR}(A) = \frac{P_0(A)}{1 - \delta P_0(A^c)}$$

for every $A \subseteq \mathcal{X}$.

It is easy to see that \underline{Q}_{COR} is a convex transformation of the probability measure P_0 . Thus, it is a 2-monotone lower probability (in fact, in Proposition 17 below we shall show that it is completely monotone). However, \underline{P}_{COR} is not a 2-monotone lower prevision:

Example 1. *To see that \underline{P}_{COR} is not a 2-monotone lower prevision, it suffices to show that the credal sets $\mathcal{M}(\underline{P}_{COR})$ and $\mathcal{M}(\underline{Q}_{COR})$ do not coincide.*

Consider a three element space \mathcal{X} , the probability measure $P_0 = (0.5, 0.3, 0.2)$ and the distortion factor $\delta = 0.1$. Then \underline{Q}_{COR} is given by:

| A | P_0 | \underline{Q}_{COR} |
|----------------|-------|-----------------------|
| $\{x_1\}$ | 0.5 | 0.4737 |
| $\{x_2\}$ | 0.3 | 0.2784 |
| $\{x_3\}$ | 0.2 | 0.1837 |
| $\{x_1, x_2\}$ | 0.8 | 0.7826 |
| $\{x_1, x_3\}$ | 0.7 | 0.6774 |
| $\{x_2, x_3\}$ | 0.5 | 0.4737 |

Since the coherent lower probability \underline{Q}_{COR} is 2-monotone, using the procedure described in Equation (2), the extreme points of $\mathcal{M}(\underline{Q}_{COR})$ are associated with the permutations of $\{1, 2, 3\}$. This produces the following extreme points:

| σ | P_σ |
|-----------|--------------------------|
| (1, 2, 3) | (0.4737, 0.3089, 0.2174) |
| (1, 3, 2) | (0.4737, 0.3226, 0.2037) |
| (2, 1, 3) | (0.5042, 0.2784, 0.2174) |
| (2, 3, 1) | (0.5263, 0.2784, 0.1953) |
| (3, 1, 2) | (0.4937, 0.3226, 0.1837) |
| (3, 2, 1) | (0.5263, 0.2900, 0.1837) |

However, for $\sigma = (1, 2, 3)$, the extreme point P_σ does not belong to $\mathcal{M}(\underline{P}_{COR})$: if we take $A = \{x_1\}$ and $B = \{x_3\}$, we obtain

$$P_\sigma(A)P_0(B) = 0.09474 < 0.09783 = (1 - \delta)P_0(A)P_\sigma(B),$$

whence applying Equation (24) we deduce that $P_\sigma \notin \mathcal{M}(\underline{P}_{COR})$. Thus, $\mathcal{M}(\underline{P}_{COR})$ is a proper subset of $\mathcal{M}(\underline{Q}_{COR})$ or, in other words, that \underline{P}_{COR} and \underline{Q}_{COR} do not coincide. \blacklozenge

The main properties of the lower probability \underline{Q}_{COR} are given in the following proposition:

Proposition 17. *Let \underline{Q}_{COR} be the restriction of the constant odds ratio to events determined by a probability measure P_0 and a distortion factor $\delta \in (0, 1)$.*

- (a) \underline{Q}_{COR} is completely monotone.
- (b) The credal set determined by \underline{Q}_{COR} is given by

$$\mathcal{M}(\underline{Q}_{COR}) = \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \frac{P(A)}{P(A^c)} \geq (1 - \delta) \frac{P_0(A)}{P_0(A^c)} \quad \forall A \subset \mathcal{X} \right\}. \quad (27)$$

- (c) The maximum number of extreme points of $\mathcal{M}(\underline{Q}_{COR})$ is $n!$.

On the other hand, unlike the lower probability induced by the pari mutuel model or the linear vacuous, \underline{Q}_{COR} is not a probability interval, meaning that it is not determined by its restriction to singletons:

Example 2. Consider $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, let P_0 be the uniform probability distribution and consider $\delta = 0.1$. Then $\underline{Q}_{COR}, \overline{Q}_{COR}$ are given by:

| $ A $ | P_0 | \underline{Q}_{COR} | \overline{Q}_{COR} |
|-------|-------|-----------------------|----------------------|
| 1 | 1/4 | 9/39 | 10/37 |
| 2 | 1/2 | 9/19 | 10/19 |
| 3 | 3/4 | 27/37 | 30/39 |
| 4 | 1 | 1 | 1 |

Since the probability measure $P = (0.235, 0.235, 0.265, 0.265)$ satisfies that $P(\{x_i\}) \in [\underline{Q}_{COR}(\{x_i\}), \overline{Q}_{COR}(\{x_i\})]$ for every $i = 1, \dots, 4$ but $P(\{x_1, x_2\}) = 0.47 < \frac{9}{19} = \underline{Q}_{COR}(\{x_1, x_2\})$, we deduce that \underline{Q}_{COR} is not a probability interval. \blacklozenge

From Proposition 17 we can prove that the credal set $\mathcal{M}(\underline{Q}_{COR})$ can be expressed as the credal set induced by a distortion model.

Theorem 18. *Consider the constant odds ratio restricted to events associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$. Then $\mathcal{M}(\underline{Q}_{COR}) = B_{d'_{COR}}^\delta(P_0)$, where $\mathcal{M}(\underline{Q}_{COR})$ is given by Equation (27), $d'_{COR} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ is given by*

$$d'_{COR}(P, Q) = \max_{\emptyset \neq A \neq \mathcal{X}} \left\{ 1 - \frac{P(A)}{P(A^c)} \frac{Q(A^c)}{Q(A)} \right\} \quad (28)$$

and $B_{d'_{COR}}^\delta(P_0)$ is given by Equation (7).

Again, note that the expression in Equation (28) is well-defined, because since $P, Q \in \mathbb{P}^*(\mathcal{X})$, the product $P(A^c)Q(A)$ can only be 0 for $A = \emptyset$ or $A = \mathcal{X}$, and these two possibilities are excluded.

We next study the properties of the function d'_{COR} .

Proposition 19. *Let d'_{COR} be the function defined in Equation (28). It satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.3, Ax.4 and Ax.5. As a consequence, it is a distance.*

Finally, we show that for the constant odds ratio restricted to events, we can always find a probability in $B_{d_{COR}}^\delta(P_0)$ at distance δ .

Proposition 20. *Consider the constant odds ratio model restricted to event associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$. Then:*

$$\sup_{P \in B_{d_{COR}}^\delta(P_0)} d'_{COR}(P, P_0) = \delta.$$

The proofs of Propositions 19 and 20 have been omitted in the Appendix because they are analogous to those of Propositions 15 and 16, d'_{COR} being the same maximum as d_{COR} but over a smaller domain.

6.2. Conditioning the COR model. In this part, we will concentrate on the odds-ratio model defined on gambles, and not restricted to events. The reason is that processing should primarily be done on the model itself, and not on an approximation of it.

In the case of the odds ratio model, Walley [32, Sec. 6.6.3] proved that the conditional model \underline{P}_B induced by $B_{d_{COR}}^\delta(P_0)$ is again a odds ratio model $B_{d_{COR}}^{\delta_B}(P_{0|B})$ such that

$$P_{0|B}(A) = P_0(A|B) \text{ and } \delta_B = \delta. \quad (29)$$

In contrast with the previous models, we can notice that the amount of imprecision in the updated model is the same as in the original one.

7. CONCLUSIONS

In this paper, we offer a unified view of a number of imprecise probability models, where a probability measure P_0 is distorted by means of a suitable function, and also taking into account some distorting factor representing to which extent the estimation of P_0 is reliable. This produces a number of credal sets, that can be equivalently represented in terms of the coherent lower probability that is obtained by taking lower envelopes. As we have seen, this framework includes in particular the models where the probability measure is directly transformed by means of some monotone function.

Following the presentation of our general model, here we have revisited some well-known distortion models under the light of this unified view. These are the pari mutuel, the linear vacuous and the constant odds ratio models. In the process, we have also established a number of additional results for those models, e.g., characterising their extreme points or the results of different operations such as conditioning.

One crucial assumption throughout is that all the probability measures take strictly positive values in the singletons. This property is assumed for the sake of technical simplicity. While most of the results in this paper also hold without this assumption, this is not always the case: for instance, the decomposition of d_{PMM} and d_{LV} in terms of singletons (see Prop. 6(c) and Prop. 10(c)) no longer holds, and the maximum distortion is not always attained within the ball $B_{d_{PMM}}^\delta(P_0)$ (i.e., Proposition 7 does not necessarily hold). Moreover, in the general case the

expressions of d_{PMM} , d_{LV} , d_{COR} and d'_{COR} should be suitably modified to avoid zeros in their denominator, and their continuity is no longer trivial.

In addition from being a simplifying assumption from the technical point of view, we think that it is also reasonable from the practical point of view: on the one hand, if we start with a probability P_0 including zero probabilities, we could always consider its restriction to $\mathcal{X}' = \{x \in \mathcal{X} \mid P_0(\{x\}) > 0\}$. On the other hand, it seems reasonable to consider a ‘small’ distortion parameter, so that the positivity of probabilities is preserved. If this parameter is related to the reliability of the information source or the quality of the data, taking a ‘large’ distortion parameter would mean that the source is not very reliable or that the data quality is low. For all these reasons, we have preferred to consider this simplifying assumption in this already long and involved paper. Some additional comments will be given in [22, App. B]

In the next part of this study, we will pursue our exploration of distortion models induced by less explored distances, namely the total variation distance, the Kolmogorov distance and the L_1 distance. We will finish by providing an overview of the different models, comparing their features.

ACKNOWLEDGEMENTS

This work was carried out in the framework of the Labex MS2T, funded by the French Government, through the National Agency for Research (Reference ANR-11-IDEX-0004-02). We also acknowledge the financial support of project PGC2018-098623-B-I00 from the Ministry of Science, Innovation and Universities of Spain and project IDI/2018/000176 from the Department of Employment, Industry and Tourism of Asturias. [We would also like to thank the reviewers for their interesting and insightful comments.](#)

APPENDIX A. PROOFS

Proof of Proposition 1. First of all, Ax.4 trivially implies that $B_d^\delta(P_0)$ is a convex set. Similarly, Ax.5 implies that it is closed: if $d(P_n, P_0) \leq \delta$ for every n and $(P_n)_n \rightarrow P$, then

$$d(P, P_0) = d(\lim_n P_n, P_0) = \lim_n d(P_n, P_0) \leq \delta.$$

Thus, $B_d^\delta(P_0)$ is a closed and convex set. Now, since in our finitary setting the weak- $*$ topology on $\mathbb{P}(\mathcal{X})$ coincides with the one associated to the Euclidean distance between the mass functions and any closed set is compact, we can use the one-to-one correspondence between coherent lower previsions and closed convex sets of probability measures [32, Thm. 3.6.1] to deduce that $B_d^\delta(P_0)$ coincides with the credal set generated by \underline{P} . \square

Proof of Proposition 2. Let us define the function d_h^* by Equation (9), and let $d(P, Q) = \max_{A \subseteq \mathcal{X}} d^*(P(A), Q(A))$.

By definition,

$$P \in \mathcal{M}(\underline{P}) \Leftrightarrow P(A) \geq \underline{P}(A) = h(P_0(A)) \quad \forall A \subseteq \mathcal{X}.$$

Let us see that this is also equivalent to $d_h^*(P(A), P_0(A)) \leq 0.5$ for every $A \subseteq \mathcal{X}$. On the one hand, if $P(A) \geq h(P_0(A))$ for every $A \subseteq \mathcal{X}$, then $d_h^*(P(A), P_0(A))$ cannot

be 1, hence $d_h^*(P(A), P_0(A)) \leq 0.5$. On the other hand, if $d_h^*(P(A), P_0(A)) \leq 0.5$ for every $A \subseteq \mathcal{X}$, then either $P(A) = P_0(A)$, hence property $h(t) \leq t$ implies that:

$$P(A) \geq h(P(A)) = h(P_0(A)) = \underline{P}(A);$$

or $P(A) \neq P_0(A)$ and $P(A) \geq h(P_0(A))$, which immediately implies $P(A) \geq \underline{P}(A)$.

We conclude that $P \in \mathcal{M}(\underline{P})$ if and only if $d_h^*(P(A), P_0(A)) \leq 0.5$ for every $A \subseteq \mathcal{X}$, or equivalently that $d(P, P_0) \leq 0.5$, whence $\mathcal{M}(\underline{P}) = B_d^{0.5}(P_0)$.

Finally, d is a premetric because d_h^* is. \square

Proof of Proposition 3. From Proposition 1, since d satisfies Ax.4 and Ax.5, $\mathcal{M}(\underline{P}) = B_d^\delta(P_0)$ for every $P_0 \in \mathbb{P}(\mathcal{X})$ and every $\delta > 0$. Moreover, if we fix P_0 and δ , and define the credal set $\mathcal{M}_i := \{P \in \mathbb{P}(\mathcal{X}) : P(\{x_i\}) \in [\underline{P}(\{x_i\}), \bar{P}(\{x_i\})]\}$, it immediately follows that $\mathcal{M}(\underline{P}) \subseteq \cap_{i=1}^n \mathcal{M}_i$, and that

$$\underline{P} \text{ is a probability interval} \Leftrightarrow \mathcal{M}(\underline{P}) = \cap_{i=1}^n \mathcal{M}_i.$$

Now, if we define the function $f_{i,P_0} : [0, 1] \rightarrow [0, \infty)$ by $f_{i,P_0}(t) = \min\{d(P, P_0) : P(\{x_i\}) = t\}$, from the continuity of d it holds that $\{P(\{x_i\}) \mid d(P, P_0) \leq \delta\} = f_{i,P_0}^{-1}([0, \delta])$. In addition,

$$\begin{aligned} \mathcal{M}_i &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x_i\}) \in [\underline{P}(\{x_i\}), \bar{P}(\{x_i\})]\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x_i\}) \in \{R(\{x_i\}) : R \in B_d^\delta(P_0)\}\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x_i\}) \in f_{i,P_0}^{-1}([0, \delta])\} = \{P \in \mathbb{P}(\mathcal{X}) \mid f_{i,P_0}(P(\{x_i\})) \leq \delta\}, \end{aligned}$$

whence $\mathcal{M}(\underline{P}) \subseteq \{P \in \mathbb{P}(\mathcal{X}) : f_{i,P_0}(P(\{x_i\})) \leq \delta \ \forall i = 1, \dots, n\}$, and \underline{P} is a probability interval if and only if

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) : f_{i,P_0}(P(\{x_i\})) \leq \delta \ \forall i = 1, \dots, n\}. \quad (30)$$

Let us prove that this is equivalent to Equation (10).

Assume that Equation (10) does not hold. Since for any $P, Q \in \mathbb{P}(\mathcal{X})$

$$d(P, Q) \geq d_i(P(\{x_i\}), Q) \quad \forall i = 1, \dots, n$$

and as a consequence

$$d(P, Q) \geq \max_{i=1, \dots, n} d_i(P(\{x_i\}), Q),$$

this means that there are $P, P_0 \in \mathbb{P}(\mathcal{X})$ and $\epsilon > 0$ such that

$$d(P, P_0) - \epsilon > d_i(P(\{x_i\}), P_0) = f_{i,P_0}(P(\{x_i\})) \quad \forall i = 1, \dots, n.$$

Given $\lambda = d(P, P_0)$, this means that $f_{i,P_0}(P(\{x_i\})) = d_i(P(\{x_i\}), P_0) \leq \lambda - \epsilon$ for every $i = 1, \dots, n$, whence $P \in \cap_{i=1}^n \{P' \in \mathbb{P}(\mathcal{X}) \mid d_i(P'(\{x_i\})) \leq \lambda - \epsilon\}$ but it does not belong to $B_d^{\lambda-\epsilon}(P_0)$. Thus, if we take this P_0 and $\delta = \lambda - \epsilon$, we see that Equation (30) does not hold and therefore the lower envelope of $B_d^\delta(P_0)$ is not a probability interval.

Conversely, if Equation (10) holds then given $P_0 \in \mathbb{P}(\mathcal{X})$ and $\delta > 0$, if for $P \in \mathbb{P}(\mathcal{X})$ it holds that $d_i(P(\{x_i\})) \leq \delta$ for every $i = 1, \dots, n$, then

$$\min\{d(P', P_0) \mid P'(\{x_i\}) = P(\{x_i\})\} \leq \delta \quad \forall i = 1, \dots, n,$$

whence by Equation (10)

$$\max_{i=1, \dots, n} \min\{d(P', P_0) \mid P'(\{x_i\}) = P(\{x_i\})\} = d(P, P_0) \leq \delta$$

and as a consequence $P \in B_d^\delta(P_0)$. Therefore, the lower envelope of $B_d^\delta(P_0)$ is a probability interval. \square

Proof of Proposition 4. (a) Consider $P, Q, R \in \mathbb{P}(\mathcal{X})$ with respective mass functions $(p_1, \dots, p_n), (q_1, \dots, q_n)$ and (r_1, \dots, r_n) . Since $d^*(p_i, q_i) \geq 0 \ \forall i = 1, \dots, n$ because $d(P, Q) \geq 0$, we deduce that d^* takes values in $[0, \infty)$. Let us now analyse the properties satisfied by d :

Ax.1a: $d(P, Q) = 0$ implies that $d^*(p_i, q_i) = 0 \ \forall i = 1, \dots, n$. By hypothesis, this implies that $p_i = q_i \ \forall i = 1, \dots, n$, whence $P = Q$.

Ax.1b: $d^*(p_i, p_i) = 0 \ \forall i = 1, \dots, n$ implies that $d(P, P) = 0$.

Ax.1: d^* satisfies Ax.1 if and only if it satisfies Ax.1a and Ax.1b, whence so does d .

Ax.2: If $d^*(p_i, r_i) \leq d^*(p_i, q_i) + d_i(q_i, r_i) \ \forall i = 1, \dots, n$, then given P, Q, R there must be some $j \in \{1, \dots, n\}$ such that $d(P, R) = d^*(p_j, r_j)$, whence $d(P, R) \leq d^*(p_j, q_j) + d^*(q_j, r_j) \leq d(P, Q) + d(Q, R)$.

Ax.3: $d^*(p_i, q_i) = d^*(q_i, p_i) \ \forall i = 1, \dots, n$ implies that $d(P, Q) = d(Q, P)$.

Ax.4: $d^*(p_i, \alpha q_i + (1 - \alpha)r_i) \leq \max\{d^*(p_i, q_i), d^*(p_i, r_i)\} \ \forall i = 1, \dots, n$ implies that $d(P, \alpha Q + (1 - \alpha)R) \leq \max\{d(P, Q), d(P, R)\}$.

Ax.5: If d^* is continuous, then if a sequence $(a_n, b_n)_n$ converges pointwise to (a, b) it holds that $\lim_n d^*(a_n, b_n) = d^*(a, b)$. Therefore, if $(P_n)_n$ converges pointwisely to P and $(Q_n)_n$ converges pointwise to Q , then $\lim_n d(P_n, Q_n) = d(P, Q)$.

(b) Denote $\mathcal{M}_i = \{P \in \mathbb{P}(\mathcal{X}) \mid d^*(P(\{x_i\}), P_0(\{x_i\})) \leq \delta\}$. Then

$$\begin{aligned} B_d^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid \max_{i=1, \dots, n} d^*(P(\{x_i\}), P_0(\{x_i\})) \leq \delta\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid d^*(P(\{x_i\}), P_0(\{x_i\})) \leq \delta \ \forall i = 1, \dots, n\} = \cap_{i=1}^n \mathcal{M}_i, \end{aligned}$$

whence the lower envelope of $B_d^\delta(P_0)$ is a probability interval. \square

Proof of Theorem 5. First of all, note that $d_{PMM}(P, Q) \geq 0$ for every $P, Q \in \mathbb{P}^*(\mathcal{X})$, since applying Equation (15) with $A = \emptyset$, we get:

$$d_{PMM}(P, Q) \geq \frac{Q(\emptyset) - P(\emptyset)}{1 - Q(\emptyset)} = 0.$$

From Equation (7),

$$\begin{aligned} B_{d_{PMM}}^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid d_{PMM}(P, P_0) \leq \delta\} \\ &= \left\{P \in \mathbb{P}(\mathcal{X}) \mid \max_{A \subseteq \mathcal{X}} \frac{P_0(A) - P(A)}{1 - P_0(A)} \leq \delta\right\}. \end{aligned}$$

For every $A \subseteq \mathcal{X}$,

$$\begin{aligned} \frac{P_0(A) - P(A)}{1 - P_0(A)} \leq \delta &\Leftrightarrow P_0(A) - P(A) \leq \delta(1 - P_0(A)) \\ &\Leftrightarrow (1 + \delta)P_0(A) - \delta \leq P(A) \Leftrightarrow P(A^c) \leq (1 + \delta)P_0(A^c). \end{aligned}$$

We deduce that $P \in B_{d_{PMM}}^\delta(P_0)$ if and only if

$$P(A) \leq (1 + \delta)P_0(A) \ \forall A \subseteq \mathcal{X}. \quad (31)$$

Let us now see that $\mathcal{M}(\underline{P}_{PMM})$ equals $B_{d_{PMM}}^\delta(P_0)$. From [21, Cor.1],

$$\mathcal{M}(\underline{P}_{PMM}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x\}) \leq (1 + \delta)P_0(\{x\}), \ \forall x \in \mathcal{X}\}.$$

Applying Equation (31), if $P \in B_{d_{PMM}}^\delta(P_0)$ then in particular it holds that $P(\{x\}) \leq (1 + \delta)P_0(\{x\})$, so $P \in \mathcal{M}(\underline{P}_{PMM})$. Conversely, if $P \in \mathcal{M}(\underline{P}_{PMM})$

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

whence $P \in B_{d_{PMM}}^\delta(P_0)$. Thus, the distortion model induced d_{PMM} , P_0 and $\delta > 0$ coincides with the pari mutuel model induced by P_0 and δ . \square

Lemma 21. *Let $\alpha_1, \dots, \alpha_k$, be strictly positive numbers and let β_1, \dots, β_k be non-negative numbers. Then*

$$\frac{\sum_{i=1}^k (\alpha_i - \beta_i)}{\sum_{i=1}^k \alpha_i} \leq \max_{i=1, \dots, k} \frac{\alpha_i - \beta_i}{\alpha_i} \quad \text{and} \quad \frac{\sum_{i=1}^k (\beta_i - \alpha_i)}{\sum_{i=1}^k \alpha_i} \leq \max_{i=1, \dots, k} \frac{\beta_i - \alpha_i}{\alpha_i}.$$

Proof. Let us prove the first inequality by induction; the proof of the second inequality is similar. The result trivially holds for $k = 1$. Let us consider the general case:

$$\begin{aligned} \frac{\sum_{i=1}^k (\alpha_i - \beta_i)}{\sum_{i=1}^k \alpha_i} &= \frac{(\alpha_1 + \alpha_2 - \beta_1 - \beta_2) + \sum_{i=3}^k (\alpha_i - \beta_i)}{(\alpha_1 + \alpha_2) + \sum_{i=3}^k \alpha_i} \\ &\leq \max \left\{ \frac{\alpha_1 + \alpha_2 - \beta_1 - \beta_2}{\alpha_1 + \alpha_2}, \frac{\alpha_3 - \beta_3}{\alpha_3}, \dots, \frac{\alpha_k - \beta_k}{\alpha_k} \right\} \\ &\leq \max \left\{ \frac{\alpha_1 - \beta_1}{\alpha_1}, \frac{\alpha_2 - \beta_2}{\alpha_2}, \frac{\alpha_3 - \beta_3}{\alpha_3}, \dots, \frac{\alpha_k - \beta_k}{\alpha_k} \right\}, \end{aligned}$$

where we have used twice the induction hypothesis. \square

Proof of Proposition 6. (a) We already know that d_{PMM} takes non-negative values from the proof of Theorem 5. In order to prove that d_{PMM} satisfies Ax.1, note that

$$d_{PMM}(P, Q) = 0 \Leftrightarrow Q(A) \leq P(A) \forall A \subset \mathcal{X} \Leftrightarrow Q(A) = P(A) \forall A \subset \mathcal{X} \Leftrightarrow Q = P.$$

To see that d_{PMM} satisfies Ax.4, take P_1, P_2, P_3 such that $d_{PMM}(P_1, P_2) \leq \delta$ and $d_{PMM}(P_1, P_3) \leq \delta$, and take $\alpha \in (0, 1)$. Then for every event $A \subset \mathcal{X}$ it holds that

$$\begin{aligned} &\frac{(\alpha P_2 + (1 - \alpha)P_3)(A) - P_1(A)}{(\alpha P_2 + (1 - \alpha)P_3)(A^c)} \\ &= \frac{\alpha(P_2(A) - P_1(A)) + (1 - \alpha)(P_3(A) - P_1(A))}{\alpha P_2(A^c) + (1 - \alpha)P_3(A^c)} \\ &= \frac{\alpha(P_1(A^c) - P_2(A^c)) + (1 - \alpha)(P_1(A^c) - P_3(A^c))}{\alpha P_2(A^c) + (1 - \alpha)P_3(A^c)} \leq \delta, \end{aligned}$$

using the second inequality in Lemma 21.

To see that d_{PMM} satisfies Ax.5 it suffices to note that the maximum is a continuous function and that the ratio in Equation (15) is bounded above once $Q \in \mathbb{P}^*(\mathcal{X})$ is fixed.

- (b) To see that d_{PMM} does not satisfy the triangle inequality (Ax.2), consider $\mathcal{X} = \{x_1, x_2, x_3\}$ and the probabilities P_1, P_2, P_3 given by:

| | P_1 | P_2 | P_3 |
|----------------|-------|-------|--------|
| $\{x_1\}$ | 0.439 | 0.400 | 0.0279 |
| $\{x_2\}$ | 0.236 | 0.424 | 0.5922 |
| $\{x_3\}$ | 0.325 | 0.176 | 0.3799 |
| $\{x_1, x_2\}$ | 0.675 | 0.824 | 0.6201 |
| $\{x_1, x_3\}$ | 0.764 | 0.576 | 0.4078 |
| $\{x_2, x_3\}$ | 0.561 | 0.600 | 0.9721 |

We obtain

$$d_{PMM}(P_1, P_3) = 14.735, \quad d_{PMM}(P_2, P_3) = 13.337, \quad d_{PMM}(P_1, P_2) = 0.8466.$$

We can see that the triangle inequality is not satisfied because:

$$d_{PMM}(P_1, P_3) = 14.735 > 0.8466 + 13.337 = d_{PMM}(P_1, P_2) + d_{PMM}(P_2, P_3).$$

The fact that d_{PMM} is not symmetric (Ax.3) is trivial, since in the denominator the term $1 - Q$ appears. For an explicit example, take $\mathcal{X} = \{x_1, x_2\}$, Q the uniform distribution, and $P(\{x_1\}) = \alpha < \frac{1}{2}$. By Equation (15),

$$d_{PMM}(P, Q) = \frac{0.5 - \alpha}{1 - 0.5} = \frac{0.5 - \alpha}{0.5} \text{ and}$$

$$d_{PMM}(Q, P) = \frac{1 - \alpha - 0.5}{\alpha} = \frac{0.5 - \alpha}{\alpha}$$

and therefore $d_{PMM}(P, Q) \neq d_{PMM}(Q, P)$.

- (c) For every $P, Q \in \mathbb{P}^*(\mathcal{X})$, we can equivalently express d_{PMM} as

$$d_{PMM}(P, Q) = \max_{Q(A) \neq 1} \frac{Q(A) - P(A)}{1 - Q(A)} = \max_{B \subset \mathcal{X}} \frac{P(B) - Q(B)}{Q(B)}.$$

Since by Equation (12) $Q(\{x\}) > 0 \forall x \in \mathcal{X}$,

$$\frac{P(B) - Q(B)}{Q(B)} = \frac{\sum_{x \in B} (P(\{x\}) - Q(\{x\}))}{\sum_{x \in B} Q(\{x\})} \leq \max_{x \in B} \frac{P(\{x\}) - Q(\{x\})}{Q(\{x\})}$$

for every non-empty event B , using the second inequality in Lemma 21. This implies that

$$d_{PMM}(P, Q) \leq \max_{x \in \mathcal{X}} \frac{P(\{x\}) - Q(\{x\})}{Q(\{x\})}.$$

The other inequality trivially holds, just taking $B = \mathcal{X} \setminus \{x\}$, where x is an element where the maximum is attained. \square

Proof of Proposition 7. On the one hand, from Equation (14) we know that $\delta < \frac{P_0(A)}{1 - P_0(A)}$ for every $A \subset \mathcal{X}$. Let us prove that there exists $P \in B_{d_{PMM}}^\delta(P_0)$ such that $d_{PMM}(P, P_0) = \delta$. Take $x \in \mathcal{X}$, and consider P determined by:

$$P(\{x'\}) = \begin{cases} (1 + \delta)P_0(\{x\}) - \delta & \text{if } x' = x, \\ (1 + \delta)P_0(\{x\}) & \text{if } x' \neq x. \end{cases}$$

It satisfies the following properties:

- P is a probability measure. First of all, $P(\{x'\}) > 0$ for every $x' \neq x$ by definition. On the other hand, $P(\{x\}) = \underline{P}_{PMM}(\{x\}) > 0$. Also:

$$\sum_{x' \in \mathcal{X}} P(\{x'\}) = (1 + \delta)P_0(\{x\}) - \delta + \sum_{x' \neq x} (1 + \delta)P_0(\{x'\}) = 1 + \delta - \delta = 1.$$

We conclude that P is a probability measure.

- $d_{PMM}(P, P_0) = \delta$. Let us compute $\max_{B \subset \mathcal{X}} \frac{P_0(B) - P(B)}{1 - P_0(B)}$. There are two cases:

– If $x \notin B$,

$$\frac{P_0(B) - P(B)}{1 - P_0(B)} = \frac{P_0(B) - (1 + \delta)P_0(B)}{1 - P_0(B)} = \frac{-\delta P_0(B)}{1 - P_0(B)} \leq 0 \leq \delta.$$

– If $x \in B$,

$$\frac{P_0(B) - P(B)}{1 - P_0(B)} = \frac{P_0(B) - (1 + \delta)P_0(B) + \delta}{1 - P_0(B)} = \frac{\delta(1 - P_0(B))}{1 - P_0(B)} = \delta.$$

Therefore, $d_{PMM}(P, P_0) = \delta$. \square

Proof of Proposition 8. We need to prove that

$$\underline{P}_B(A) = \max \left\{ 0, (1 + \delta) \frac{P_0(A \cap B)}{P_0(B)} - \delta_B \right\}$$

with $\delta_B = \delta / ((1 + \delta)P_0(B) - \delta)$. Since by [21, Thm. 1] the PMM induces a probability interval, and hence a 2-monotone lower probability, we have that [31, Thm. 7.2]

$$\begin{aligned} \underline{P}_B(A) &= \frac{\underline{P}_{PMM}(A \cap B)}{\underline{P}_{PMM}(A \cap B) + \overline{P}_{PMM}(A^c \cap B)} \\ &= \frac{\max\{0, (1 + \delta)P_0(A \cap B) - \delta\}}{\max\{0, (1 + \delta)P_0(A \cap B) - \delta\} + \min\{1, (1 + \delta)P_0(A^c \cap B)\}}, \end{aligned}$$

that is zero as soon as $(1 + \delta)P_0(A \cap B) \leq \delta$. Let us now look at the case $(1 + \delta)P_0(A \cap B) > \delta$, in which case the equality becomes

$$\underline{P}_B(A) = \frac{(1 + \delta)P_0(A \cap B) - \delta}{(1 + \delta)P_0(A \cap B) - \delta + \min\{1, (1 + \delta)P_0(A^c \cap B)\}}$$

since we have $P_0(A \cap B) + P_0(A^c \cap B) < 1$, we can deduce $(1 + \delta)P_0(A \cap B) + (1 + \delta)P_0(A^c \cap B) < (1 + \delta)$. Furthermore, as δ is a lower bound of $(1 + \delta)P_0(A \cap B)$, we also have $(1 + \delta)P_0(A^c \cap B) < 1$, hence the equality becomes

$$\begin{aligned} \underline{P}_B(A) &= \frac{(1 + \delta)P_0(A \cap B) - \delta}{(1 + \delta)P_0(A \cap B) - \delta + (1 + \delta)P_0(A^c \cap B)} \\ &= \frac{(1 + \delta)P_0(A \cap B) - \delta}{(1 + \delta)P_0(B) - \delta} = \frac{(1 + \delta)P_0(B)}{\underline{P}_{PMM}(B)} \frac{P_0(A \cap B)}{P_0(B)} - \frac{\delta}{\underline{P}_{PMM}(B)}, \end{aligned}$$

which finishes the proof because $(1 + \delta)P_0(B) / \underline{P}_{PMM}(B) = 1 + \delta_B$. \square

Proof of Theorem 9. First of all, note that $d_{LV}(P, Q) \geq 0$ for every $P, Q \in \mathbb{P}^*(\mathcal{X})$. The reason is that taking the sure event \mathcal{X} we obtain that:

$$d_{LV}(P, Q) \geq \frac{Q(\mathcal{X}) - P(\mathcal{X})}{Q(\mathcal{X})} = 0,$$

whence d_{LV} .

Using the function d_{LV} , the probability P_0 and the parameter $\delta \in (0, 1)$, we can define the following credal set using Equation (7):

$$\begin{aligned} B_{d_{LV}}^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid d_{LV}(P, P_0) \leq \delta\} \\ &= \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \max_{A \neq \emptyset} \frac{P_0(A) - P(A)}{P_0(A)} \leq \delta \right\}. \end{aligned}$$

This means that $P \in B_{d_{LV}}^\delta(P_0)$ if and only if for every $A \neq \emptyset$, it holds that:

$$\frac{P_0(A) - P(A)}{P_0(A)} \leq \delta \Leftrightarrow P_0(A) - P(A) \leq \delta P_0(A) \Leftrightarrow P(A) \geq (1 - \delta)P_0(A).$$

Hence, $B_{d_{LV}}^\delta(P_0)$ can be equivalently expressed as

$$B_{d_{LV}}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq (1 - \delta)P_0(A), \forall A \subseteq \mathcal{X}\};$$

by Equation (17), this is the credal set associated with a linear vacuous model induced by P_0 and δ . \square

Proof of Proposition 10. (a) We already know from Theorem 9 that d_{LV} is non-negative.

Consider now $P \in \mathbb{P}(\mathcal{X})$ and $A \neq \emptyset$, it holds that:

$$\frac{P(A) - P(A)}{P(A)} = 0.$$

This means that $d_{LV}(P, P) = 0$ and therefore d_{LV} satisfies Ax.1b.

Thirdly, assume that $d_{LV}(P, Q) = 0$. This means that for every $A \subseteq \mathcal{X}$ with $Q(A) > 0$, it holds that:

$$\frac{Q(A) - P(A)}{Q(A)} \leq 0 \Rightarrow Q(A) = P(A).$$

Since we are assuming that $Q(\{x\}) > 0$ for every $x \in \mathcal{X}$, this means that $Q(\{x\}) \leq P(\{x\}) \forall x \in \mathcal{X}$, whence $Q = P$. Thus, d_{LV} satisfies Ax.1a. Since we have also proved that $d(P, P) = 0$ for every P , we conclude that it also satisfies Ax.1.

Next, let us show that $d_{LV}(P, Q)$ satisfies the triangular inequality (Ax.2), in the sense that for any P_1, P_2, P_3 , we have

$$d_{LV}(P_3, P_1) \leq d_{LV}(P_3, P_2) + d_{LV}(P_2, P_1),$$

or equivalently:

$$\max_{A \neq \emptyset} \frac{P_1(A) - P_3(A)}{P_1(A)} \leq \max_{A \neq \emptyset} \frac{P_2(A) - P_3(A)}{P_2(A)} + \max_{A \neq \emptyset} \frac{P_1(A) - P_2(A)}{P_1(A)}. \quad (32)$$

Let us first assume that the maximal values are obtained on A_1, A_2, A_3 , hence Equation (32) becomes

$$\frac{P_1(A_1) - P_3(A_1)}{P_1(A_1)} \leq \frac{P_2(A_2) - P_3(A_2)}{P_2(A_2)} + \frac{P_1(A_3) - P_2(A_3)}{P_1(A_3)}.$$

There are two possible cases:

– If $P_2(A_1) \geq P_1(A_1)$, we have that

$$\begin{aligned} \frac{P_1(A_1) - P_3(A_1)}{P_1(A_1)} &\leq \frac{P_2(A_1) - P_3(A_1)}{P_2(A_1)} + \frac{P_1(A_3) - P_2(A_3)}{P_1(A_3)} \\ &\leq \frac{P_2(A_2) - P_3(A_2)}{P_2(A_2)} + \frac{P_1(A_3) - P_2(A_3)}{P_1(A_3)} \end{aligned}$$

where the first inequality holds because

$$\frac{P_1(A_1) - P_3(A_1)}{P_1(A_1)} \leq \frac{P_2(A_1) - P_3(A_1)}{P_2(A_1)} \text{ and } \frac{P_1(A_3) - P_2(A_3)}{P_1(A_3)} \geq 0.$$

– If $P_1(A_1) \geq P_2(A_1)$,

$$\begin{aligned} \frac{P_1(A_1) - P_3(A_1)}{P_1(A_1)} &= \frac{P_2(A_1) - P_3(A_1)}{P_1(A_1)} + \frac{P_1(A_1) - P_2(A_1)}{P_1(A_1)} \\ &\leq \frac{P_2(A_1) - P_3(A_1)}{P_2(A_1)} + \frac{P_1(A_1) - P_2(A_1)}{P_1(A_1)} \\ &\leq \frac{P_2(A_2) - P_3(A_2)}{P_2(A_2)} + \frac{P_1(A_3) - P_2(A_3)}{P_1(A_3)}. \end{aligned}$$

We conclude that d_{LV} satisfies Ax.2.

To see that d_{LV} also satisfies Ax.4, take $Q_1, Q_2, P \in \mathbb{P}^*(\mathcal{X})$ such that $d_{LV}(P, Q_1) \leq \delta$ and $d_{LV}(P, Q_2) \leq \delta$, and take $\alpha \in (0, 1)$. Then for every event $A \neq \emptyset$ it holds that

$$\begin{aligned} &\frac{(\alpha Q_1 + (1 - \alpha)Q_2)(A) - P(A)}{(\alpha Q_1 + (1 - \alpha)Q_2)(A)} \\ &= \frac{\alpha(Q_1(A) - P(A)) + (1 - \alpha)(Q_2(A) - P(A))}{\alpha Q_1(A) + (1 - \alpha)Q_2(A)} \\ &\leq \max \left\{ \frac{\alpha(Q_1(A) - P(A))}{\alpha Q_1(A)}, \frac{(1 - \alpha)(Q_2(A) - P(A))}{\alpha Q_2(A)} \right\} \leq \delta, \end{aligned}$$

applying the first inequality in Lemma 21.

Finally, if $\|P - Q\| < \delta$, it follows that

$$\frac{Q(A) - P(A)}{Q(A)} \leq \frac{\delta}{\min_{A \neq \emptyset} \underline{P}_{LV}(A)},$$

and $\min_{A \neq \emptyset} \underline{P}(A) > 0$ by Equation (12). Thus, d_{LV} satisfies Ax.6.

- (b) Let us show that d_{LV} is not symmetric (Ax.3). Take $\mathcal{X} = \{x_1, x_2\}$, and let Q be the uniform distribution on \mathcal{X} . Take also $P \in \mathbb{P}^*(\mathcal{X})$ given by $P(\{x_1\}) = \alpha < 0.5$. By Equation (18),

$$d_{LV}(P, Q) = \frac{0.5 - \alpha}{0.5}, \quad d_{LV}(Q, P) = \frac{1 - \alpha - 0.5}{1 - \alpha} = \frac{0.5 - \alpha}{1 - \alpha},$$

so $d_{LV}(P, Q) \neq d_{LV}(Q, P)$.

- (c) Consider any $A \neq \emptyset$. Then

$$\frac{Q(A) - P(A)}{Q(A)} = \frac{\sum_{x \in A} (Q(\{x\}) - P(\{x\}))}{\sum_{x \in A} Q(\{x\})} \leq \max_{x \in A} \frac{Q(\{x\}) - P(\{x\})}{Q(\{x\})},$$

where the last inequality follows from Lemma 21. This implies that:

$$d_{LV}(P, Q) \leq \max_{x \in A} \frac{Q(\{x\}) - P(\{x\})}{Q(\{x\})}.$$

The reverse inequality follows just taking the element x where the maximum is attained, and considering $B = \mathcal{X} \setminus \{x\}$. \square

Proof of Proposition 11. Let us prove that there exists $P \in B_{d_{LV}}^\delta(P_0)$ such that $d_{LV}(P, P_0) = \delta$. Consider a fixed event $A \subset \mathcal{X}$ and P given by:

$$P(\{x\}) = \begin{cases} (1 - \delta)P_0(\{x\}) & \text{if } x \in A, \\ (1 - \delta)P_0(\{x\}) + \delta \frac{P_0(\{x\})}{P_0(A^c)} & \text{if } x \notin A, \end{cases}$$

and $P(B) = \sum_{x \in B} P(\{x\})$. P satisfies the following properties:

- P is a probability measure. First of all, $P(\{x\}) \geq 0$ trivially holds for every $x \in \mathcal{X}$. Also:

$$\begin{aligned} \sum_{x \in \mathcal{X}} P(\{x\}) &= \sum_{x \in A} P(\{x\}) + \sum_{x \notin A} P(\{x\}) \\ &= \sum_{x \in A} (1 - \delta)P_0(\{x\}) + \sum_{x \notin A} \left((1 - \delta)P_0(\{x\}) + \delta \frac{P_0(\{x\})}{P_0(A^c)} \right) \\ &= (1 - \delta)P_0(A) + \delta + (1 - \delta)P_0(A^c) = 1. \end{aligned}$$

We conclude that P is a probability measure.

- $P \in B_{d_{LV}}^\delta(P_0)$. Let us prove that $\frac{P_0(\{x\}) - P(\{x\})}{P_0(\{x\})} \leq \delta$ for every $B \subseteq \mathcal{X}$. We consider the following cases:

– If $x \in A$,

$$\frac{P_0(\{x\}) - P(\{x\})}{P_0(\{x\})} = \frac{P_0(\{x\}) - (1 - \delta)P_0(\{x\})}{P_0(\{x\})} = \delta.$$

– If $x \in A^c$,

$$\begin{aligned} \frac{P_0(\{x\}) - P(\{x\})}{P_0(\{x\})} &= \frac{P_0(x) - (1 - \delta)P_0(x) - \delta \frac{P_0(x)}{P_0(A^c)}}{P_0(x)} \\ &= \delta \frac{P_0(x) - \frac{P_0(x)}{P_0(A^c)}}{P_0(x)} = \delta \left(1 - \frac{1}{P_0(A^c)} \right) \leq 0 \leq \delta. \end{aligned}$$

- Finally, $d_{LV}(P, P_0) = \delta$, because we have seen that for every $x \in A$ it holds that:

$$\frac{P_0(\{x\}) - P(\{x\})}{P_0(\{x\})} = \delta$$

and applying Proposition 10(c) we get:

$$d_{LV}(P, P_0) = \max_{x \in \mathcal{X}} \frac{P_0(\{x\}) - P(\{x\})}{P_0(\{x\})} = \delta.$$

We conclude that $\max_{P \in B_{d_{LV}}^\delta(P_0)} d_{LV}(P, P_0) = \delta$. \square

Proof of Proposition 13. Recall that we are assuming throughout that $n \geq 2$. Let P_A be the extreme point associated with $A \subseteq \mathcal{X}$ by means of Equation (25). For any $x \in A$, it holds that

$$\begin{aligned} P_A(\{x\}) &= \frac{(1 - \delta)P_0(\{x\})}{1 - \delta P_0(A)} < P_0(\{x\}) \\ \Leftrightarrow P_0(\{x\}) - \delta P_0(\{x\}) &< P_0(\{x\}) - \delta P_0(\{x\})P_0(A) \Leftrightarrow \delta P_0(\{x\})[1 - P_0(A)] > 0, \end{aligned}$$

that holds if A is a proper subset of \mathcal{X} by Equation (12). On the other hand, if $x \notin A$, it holds that

$$P_A(\{x\}) = \frac{P_0(\{x\})}{1 - \delta P_0(A)} > P_0(\{x\}).$$

From this we deduce that if A_1, A_2 are proper subsets of \mathcal{X} and $A_1 \neq A_2$, then there is some $x \in A_1 \triangle A_2$, and in that case $P_{A_1}(\{x\}) \neq P_{A_2}(\{x\})$.

Finally, both $A = \emptyset$ and $A = \mathcal{X}$ produce $P_A = P_0$. Therefore, the number of extreme points is at most $2^n - 1$, and at least $2^n - 2$. To see that P_0 is not an extreme point, note that, from Equation (24) for $\epsilon > 0$ small enough, the ball centered in P_0 and with radius ϵ in the Euclidean distance is included in the credal set $B_{d_{COR}}^\epsilon(P_0)$. Thus, $B_{d_{COR}}^\delta(P_0)$ has $2^n - 2$ different extreme points. \square

Proof of Theorem 14. First of all, let us see that d_{COR} is non-negative. For this aim, note that if we consider $A = B = \mathcal{X}$, we obtain:

$$d_{COR}(P, Q) \geq 1 - \frac{P(\mathcal{X}) \cdot P(\mathcal{X})}{P(\mathcal{X}) \cdot P(\mathcal{X})} = 0.$$

Using the function d_{COR} , the probability P_0 and the parameter $\delta \in (0, 1)$ we can define the following credal set by Equation (7):

$$\begin{aligned} B_{d_{COR}}^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid d_{COR}(P, P_0) \leq \delta\} \\ &= \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \max_{A, B \subseteq \mathcal{X}} \left(1 - \frac{P(A) \cdot P_0(B)}{P(B) \cdot P_0(A)} \right) \leq \delta \right\}. \end{aligned}$$

This means that $P \in B_{d_{COR}}^\delta(P_0)$ if and only if for every non-empty $A, B \subseteq \mathcal{X}$, it holds that:

$$\begin{aligned} 1 - \frac{P(A) \cdot P_0(B)}{P(B) \cdot P_0(A)} \leq \delta &\Leftrightarrow (1 - \delta)P(B)P_0(A) \leq P(A)P_0(B) \\ &\Leftrightarrow (1 - \delta)\frac{P_0(A)}{P_0(B)} \leq \frac{P(A)}{P(B)}. \end{aligned}$$

This means that $P \in B_{d_{COR}}^\delta(P_0)$ if and only if $P \in \mathcal{M}(\underline{P}_{COR})$. \square

Proof of Proposition 15. From the proof of Theorem 14 we already know that d_{COR} is non-negative.

Take $P \in \mathbb{P}^*(\mathcal{X})$, and let us see that $d(P, P) = 0$. For every non-empty A, B ,

$$1 - \frac{P(A) \cdot P(B)}{P(B) \cdot P(A)} = 0,$$

whence $d(P, P) = 0$. This means that d_{COR} satisfies Ax.1b.

Let us see that d_{COR} also satisfies property Ax.1a. Take $P, Q \in \mathbb{P}^*(\mathcal{X})$ such that $d_{COR}(P, Q) = 0$. This means that for each non-empty $A, B \subseteq \mathcal{X}$, it holds that:

$$1 - \frac{P(A) \cdot Q(B)}{P(B) \cdot Q(A)} \leq 0 \Leftrightarrow P(B) \cdot Q(A) \geq P(A) \cdot Q(B),$$

and this can only hold for every non-empty A, B if $P = Q$. From this we deduce that d_{COR} satisfies Ax.1.

Let us now prove that d_{COR} satisfies Ax.2. For this let A_1, B_1, A_2, B_2 and A_3, B_3 be the pairs of events where the maxima of $d_{COR}(P_1, P_3)$, $d_{COR}(P_2, P_3)$ and $d_{COR}(P_1, P_2)$ are respectively attained, i.e.:

$$\begin{aligned} d_{COR}(P_1, P_3) &= 1 - \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)}, & d_{COR}(P_2, P_3) &= 1 - \frac{P_2(A_2)P_3(B_2)}{P_2(B_2)P_3(A_2)} \\ d_{COR}(P_1, P_2) &= 1 - \frac{P_1(A_3)P_2(B_3)}{P_1(B_3)P_2(A_3)}. \end{aligned}$$

We then have to prove that $d_{COR}(P_1, P_3) \leq d_{COR}(P_1, P_2) + d_{COR}(P_2, P_3)$, or equivalently:

$$1 + \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \geq \frac{P_1(A_3)P_2(B_3)}{P_1(B_3)P_2(A_3)} + \frac{P_2(A_2)P_3(B_2)}{P_2(B_2)P_3(A_2)}. \quad (33)$$

We consider the following cases:

Case 1: Assume that $\frac{P_1(A_1)}{P_1(B_1)} \geq \frac{P_2(A_1)}{P_2(B_1)}$. In that case:

$$\frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \geq \frac{P_2(A_1)P_3(B_1)}{P_2(B_1)P_3(A_1)} \geq \frac{P_2(A_2)P_3(B_2)}{P_2(B_2)P_3(A_2)},$$

whence

$$d_{COR}(P_1, P_3) \leq d_{COR}(P_2, P_3) \leq d_{COR}(P_1, P_2) + d_{COR}(P_2, P_3).$$

Case 2: Assume that $\frac{P_3(B_1)}{P_3(A_1)} \geq \frac{P_2(B_1)}{P_2(A_1)}$. Then:

$$\frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \geq \frac{P_1(A_1)P_2(B_1)}{P_1(B_1)P_2(A_1)} \geq \frac{P_1(A_3)P_2(B_3)}{P_1(B_3)P_2(A_3)},$$

whence

$$d_{COR}(P_1, P_3) \leq d_{COR}(P_1, P_2) \leq d_{COR}(P_1, P_2) + d_{COR}(P_2, P_3).$$

Case 3: Assume that $\frac{P_1(A_1)}{P_1(B_1)} \not\geq \frac{P_2(A_1)}{P_2(B_1)}$ and $\frac{P_3(B_1)}{P_3(A_1)} \not\geq \frac{P_2(B_1)}{P_2(A_1)}$. It holds that:

$$\begin{aligned} \frac{P_1(A_3)P_2(B_3)}{P_1(B_3)P_2(A_3)} + \frac{P_2(A_2)P_3(B_2)}{P_2(B_2)P_3(A_2)} &\leq \frac{P_1(A_1)P_2(B_1)}{P_1(B_1)P_2(A_1)} + \frac{P_2(A_1)P_3(B_1)}{P_2(B_1)P_3(A_1)} \\ &= \frac{P_1(A_1)P_2(B_1)P_3(A_1)P_3(B_1)}{P_1(B_1)P_2(A_1)P_3(B_1)P_3(A_1)} + \frac{P_2(A_1)P_3(B_1)}{P_2(B_1)P_3(A_1)} \\ &= \left(\frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \right) \left(\frac{P_2(B_1)P_3(A_1)}{P_2(A_1)P_3(B_1)} \right) + \frac{P_2(A_1)P_3(B_1)}{P_2(B_1)P_3(A_1)}. \end{aligned} \quad (34)$$

We need to prove that this quantity is not greater than $1 + \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)}$. If we use the following notation:

$$t := \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \quad \alpha = \frac{P_2(B_1)P_3(A_1)}{P_2(A_1)P_3(B_1)},$$

then, Equation (34) is smaller than or equal to $1 + \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)}$ if and only if $\alpha t + \frac{1}{\alpha} \leq 1 + t$. Since $\alpha \geq 1$ because $\frac{P_3(B_1)}{P_3(A_1)} \leq \frac{P_2(B_1)}{P_2(A_1)}$, it follows that

$$\alpha t + \frac{1}{\alpha} \leq 1 + t \Leftrightarrow t \leq \frac{1}{\alpha} \Leftrightarrow \frac{P_1(A_1)P_3(B_1)}{P_1(B_1)P_3(A_1)} \leq \frac{P_2(A_1)P_3(B_1)}{P_2(B_1)P_3(A_1)} \Leftrightarrow \frac{P_1(A_1)}{P_1(B_1)} \leq \frac{P_2(A_1)}{P_2(B_1)},$$

equality that holds by assumption. Then, Equation (33) holds, and so is the triangle inequality.

That d_{COR} is symmetric (Ax.3) is trivial, because if we exchange the role of P and Q , we just need to also exchange the role of A and B .

We prove next that it satisfies Ax.4. From Equation (26),

$$d_{COR}(P_0, Q) \leq \delta \Leftrightarrow (1 - \delta) \frac{P_0(A)}{P_0(B)} \leq \frac{Q(A)}{Q(B)},$$

and trivially this inequality is preserved under convex combinations.

Finally, to see that d_{COR} is continuous, note that, for any $P, Q \in \mathbb{P}(\mathcal{X})$ such that $\|P - Q\| < \delta$, it holds that

$$\frac{P(A)Q(B)}{P(B)Q(A)} \geq \left(\max_{x \in \mathcal{X}} \frac{P(\{x\}) - \delta}{P(\{x\})} \right)^2,$$

because for any event A it holds that

$$\frac{P(A)}{Q(A)} \geq \frac{P(A)}{P(A) + \delta} \geq \max_{x \in \mathcal{X}} \frac{P(\{x\})}{P(\{x\}) + \delta} \geq \max_{x \in \mathcal{X}} \frac{P(\{x\}) - \delta}{P(\{x\})},$$

and similarly given $B \neq \emptyset$

$$\frac{Q(B)}{P(B)} \geq \frac{P(B) - \delta}{P(B)} \geq \max_{x \in \mathcal{X}} \frac{P(\{x\}) - \delta}{P(\{x\})}.$$

Using Equation (12),

$$d_{COR}(P, Q) \leq 1 - \left(\max_{x \in \mathcal{X}} \frac{P(\{x\}) - \delta}{P(\{x\})} \right)^2$$

and as a consequence d_{COR} satisfies Ax.5. \square

Proof of Proposition 16. Let us fix an event $A \subseteq \mathcal{X}$, and consider the associated extreme point P_A of $B_{d_{COR}}^\delta(P_0)$ given by Equation (25). Let us now consider the events A, A^c . We have

$$\frac{P_A(A)}{P_A(A^c)} = \frac{(1 - \delta)P_0(A \cap A) + P_0(A^c \cap A)}{(1 - \delta)P_0(A \cap A^c) + P_0(A^c \cap A^c)} = \frac{(1 - \delta)P_0(A)}{P_0(A^c)}.$$

Since

$$\frac{(1 - \delta)P_0(A)}{P_0(A^c)} \frac{P_0(A^c)}{P_0(A)} = 1 - \delta,$$

we obtain that $d_{COR}(P_A, P_0) \geq \delta$ and as a consequence $d_{COR}(P_A, P_0) = \delta$. \square

Proof of Proposition 17. (a) The lower probability \underline{Q}_{COR} can be expressed as $\underline{Q}_{COR} = f(P_0)$, where $f : [0, 1] \rightarrow [0, 1]$ is given by:

$$f(t) = \frac{(1 - \delta)t}{1 - \delta t} = (1 - \delta)t \sum_{n=0}^{\infty} (\delta t)^n \quad \forall t \in [0, 1].$$

Since $f(0) = 0$, $f(1) = 1$ and $f^{(n)} \geq 0$, it follows that f is completely monotone [4, Thm.3]. Then, since $f(P_0)$ is the composition of two completely monotone functions we deduce that it is also completely monotone [4, Thm.2].

(b) By definition of the credal set,

$$\begin{aligned} P \in \mathcal{M}(\underline{Q}_{COR}) &\Leftrightarrow P(A) \geq \underline{Q}_{COR}(A) \quad \forall A \Leftrightarrow P(G(I_A)) \geq 0 \\ &\Leftrightarrow P(G(I_A)^+) - P(G(I_A)^-) \geq 0 \Leftrightarrow \frac{P(G(I_A)^+)}{P(G(I_A)^-)} \geq 1 = \frac{(1 - \delta)P_0(G(I_A)^+)}{P_0(G(I_A)^-)}. \end{aligned}$$

Since $G(I_A) = I_A - \underline{Q}_{COR}(A)$, it follows that $G(I_A)^+ = (1 - \underline{Q}_{COR}(A))I_A$ and $G(I_A)^- = -(-\underline{Q}_{COR}(A))I_{A^c} = \underline{Q}_{COR}(A)I_{A^c}$. Therefore,

$$\begin{aligned} \frac{P(G(I_A)^+)}{P(G(I_A)^-)} &\geq \frac{(1 - \delta)P_0(G(I_A)^+)}{P_0(G(I_A)^-)} \\ &\Leftrightarrow \frac{(1 - \underline{Q}_{COR}(A))P(A)}{\underline{Q}_{COR}(A)P(A^c)} \geq (1 - \delta) \frac{(1 - \underline{Q}_{COR}(A))P_0(A)}{\underline{Q}_{COR}(A)P_0(A^c)} \\ &\Leftrightarrow \frac{P(A)}{P(A^c)} \geq (1 - \delta) \frac{P_0(A)}{P_0(A^c)}. \end{aligned}$$

- (c) Since \underline{Q}_{COR} is a belief function, in particular it is 2-monotone, thus its associated credal set has at most $n!$ extreme points, and these are in correspondence with the permutations (see the procedure described in Equation (2)). Let us prove that this upper bound is attained.

Let P_0 be the uniform distribution on $\mathcal{X} = \{1, \dots, n\}$, and take $\delta < \frac{1}{n}$. For any permutation σ , applying Equation (2) we obtain that

$$\begin{aligned} \underline{Q}_{COR}(\{x_{\sigma(1)}\}) &= \frac{(1 - \delta)/n}{1 - \delta/n} = \frac{1 - \delta}{n - \delta} \\ \underline{Q}_{COR}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) &= \frac{2(1 - \delta)/n}{1 - 2\delta/n} = \frac{2(1 - \delta)}{n - 2\delta}, \end{aligned}$$

and, more generally, $\underline{Q}_{COR}(\{x_{\sigma(1)}, \dots, x_{\sigma(k)}\}) = \frac{k(1 - \delta)}{n - k\delta}$. This implies that the extreme point associated with σ satisfies

$$P_\sigma(\{x_{\sigma(k)}\}) = \frac{k(1 - \delta)}{n - k\delta} - \frac{(k - 1)(1 - \delta)}{n - (k - 1)\delta} = \frac{n(1 - \delta)}{(n - k\delta)(n - (k - 1)\delta)}.$$

This implies that $P_\sigma(\{x_{\sigma(k)}\}) \neq P_{\sigma'}(\{x_{\sigma'(k')}\})$ for any $k \neq k'$, and as a consequence the $n!$ different permutations produce $n!$ different extreme points. \square

Proof of Theorem 18. First of all, d'_{COR} is non negative because if $\frac{P(A)Q(A^c)}{P(A^c)Q(A)} > 1$ for some $A \neq \emptyset, \mathcal{X}$, then taking A^c we obtain $\frac{P(A^c)Q(A)}{P(A)Q(A^c)} < 1$, hence the maximum is always non-negative.

To prove that $\mathcal{M}(\underline{Q}_{COR})$ coincides with $B_{d'_{COR}}^\delta(P_0)$, it suffices to note that

$$\begin{aligned} 1 - \frac{P(A)}{P(A^c)} \frac{Q(A^c)}{Q(A)} \leq \delta \quad \forall A \neq \emptyset, \mathcal{X} &\Leftrightarrow 1 - \delta \leq \frac{P(A)}{P(A^c)} \frac{Q(A^c)}{Q(A)} \quad \forall A \neq \emptyset, \mathcal{X} \\ &\Leftrightarrow \frac{P(A)}{P(A^c)} \geq (1 - \delta) \frac{Q(A)}{Q(A^c)} \quad \forall A \subseteq \mathcal{X}, \end{aligned}$$

and to replace Q by P_0 . \square

REFERENCES

- [1] A. Benavoli and M. Zaffalon. Density-ratio robustness in dynamic state estimation. *Mechanical Systems and Signal Processing*, 37(1-2):54-75, 2013.
- [2] J. Berger. Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25:303-328, 1990.
- [3] A. Bronevich. On the closure of families of fuzzy measures under eventwise aggregations. *Fuzzy Sets and Systems*, 153:45-70, 2005.
- [4] A. Bronevich. Necessary and sufficient consensus conditions for the eventwise aggregation of lower probabilities. *Fuzzy Sets and Systems*, 158:881-894, 2007.

- [5] A. Chateauneuf. Decomposable capacities, distorted probabilities and concave capacities. *Mathematical Social Sciences*, 31:19–37, 1996.
- [6] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.
- [7] C. Corsato, R. Pelessoni, and P. Vicig. Nearly-linear uncertainty measures. *International Journal of Approximate Reasoning*, 114:1–28, 2019.
- [8] I. Csiszár. Axiomatic characterization of information measures. *Entropy*, 10:261–273, 2008.
- [9] 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.
- [10] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
- [11] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.
- [12] S. Filippi, O. Cappé, and A. Garivier. Optimism in reinforcement learning and Kullback-Leibler divergence. In *48th Annual Allerton Conference on Communication, Control, and Computing*, pages 115–122. IEEE, 2010.
- [13] T. Herron, T. Seidenfeld, and L. Wasserman. Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444, 1997.
- [14] M. Hourbracq, C. Baudrit, P.-H. Willemin, and S. Destercke. Dynamic credal networks: introduction and use in robustness analysis. In *Eighth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2013)*, pages 159–169, 2013.
- [15] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [16] P. J. Huber and V. Strassen. Minimax tests and the Neyman–Pearson lemma for capacities. *The Annals of Statistics*, 1:251–263, 1973.
- [17] I. Levi. *The enterprise of knowledge*. MIT Press, Cambridge, 1980.
- [18] E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.
- [19] E. Miranda and I. Montes. Coherent updating of non-additive measures. *International Journal of Approximate Reasoning*, 56(B):159–177, 2015.
- [20] E. Miranda, I. Montes, and S. Destercke. A unifying frame for neighbourhood and distortion models. *Proceedings of Machine Learning Research*, 103:304–313, 2019. Proceedings of ISIPTA Conference.
- [21] I. Montes, E. Miranda, and S. Destercke. Pari-mutuel probabilities as an uncertainty model. *Information Sciences*, 481:550 – 573, 2019.
- [22] I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models: Part II- New models and synthesis. *International Journal of General Systems*, 2020. Submitted for publication.
- [23] S. Moral. Discounting imprecise probabilities. In E. Gil, J. Gil, and M. Gil, editors, *The Mathematics of the Uncertain*, volume 142 of *Studies in Systems, Decision and Control*. Springer, 2018.
- [24] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51:1145–1158, 2010.
- [25] L. R. Pericchi and P. Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 59:1–23, 1991.
- [26] H. Rieder. Least favourable pairs for special capacities. *Annals of Statistics*, 5(5):909–921, 1997.
- [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, New Jersey, 1976.
- [29] L. Utkin. A framework for imprecise robust one-class classification models. *Journal of Machine Learning Research and Cybernetics*, 5(3):379–393, 2014.
- [30] L. Utkin and A. Wiencierz. Improving over-fitting in ensemble regression by imprecise probabilities. *Information Sciences*, 317:315–328, 2015.
- [31] P. Walley. Coherent lower (and upper) probabilities. Technical Report 22, University of Warwick, Coventry, 1981.
- [32] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

- [33] A. Wallner. Bi-elastic neighbourhood models. In J.-M-Bernard, T. Seidenfeld, and M. Zafalon, editors, *Proceedings of the 3rd International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'2003)*, volume 18, pages 593–607. Carleton Scientific, 2003.

UNIVERSITY OF OVIEDO, DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
Email address: `imontes@uniovi.es`

UNIVERSITY OF OVIEDO, DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
Email address: `mirandaenrique@uniovi.es`

TECHNOLOGICAL UNIVERSITY OF COMPIÈGNE
Email address: `sebastien.destercke@hds.utc.fr`