UNIFYING NEIGHBOURHOOD AND DISTORTION MODELS: PART II - NEW MODELS AND SYNTHESIS

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. In the first part of this study, we introduced a general, unified view encompassing such neighbourhoods, and revisited some well-known models (pari mutuel, linear vacuous, constant odds-ratio). In this second part, we study models that have received little to no attention, but are induced by classical distances between probabilities, such as the total variation, the Kolmogorov and the L_1 distances. We finish by comparing those models in terms of a number of properties: precision, number of extreme points, n-monotonicity, ... thus providing possible guidelines to select a neighbourhood rather than another.

Keywords: Neighbourhood models, distorted probabilities, total variation distance, Kolmogorov distance, L_1 distance.

1. Introduction and quick reminders

Our motivations for studying neighbourhood models have already been discussed in the first part of this study [19]. Here, let us simply recall that their interest is to provide simple models relying on very few parameters, which are:

- An initial probability P_0 defined on some finite space $\mathcal{X}$ with at least two elements and belonging to the set $\mathbb{P}(\mathcal{X})$ of probability distributions, that may be the result of a classic estimation procedure or of some expert elicitation.
- A function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \to [0, \infty)$ (in this second paper a distance) between probabilities, defining the shape of the neighbourhood. Hence, d satisfies positive definiteness (Ax.1), triangle inequality (Ax.2) and symmetry (Ax.3), as well as two weaker conditions than positive definiteness:
  Ax.1a: $d(P_1, P_2) = 0$ implies $P_1 = P_2$.
  Ax.1b: $d(P, P) = 0$.
  Also, it may satisfy other desirable properties: convexity (Ax.4) and continuity (Ax.5).
- A distortion factor δ defining the size of the neighbourhood around P_0.

From these three elements we define the associated neighbourhood as the credal set

$$B_d^{\delta}(P_0) := \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\} \tag{1}$$

whose associated *lower and upper previsions* are defined as

$$\underline{P}(f) := \inf \left\{P(f) \mid P \in B_d^{\delta}(P_0)\right\} \quad \text{and} \quad \overline{P}(f) := \sup \left\{P(f) \mid P \in B_d^{\delta}(P_0)\right\}$$

for any function $f : \mathcal{X} \to \mathbb{R}$, and where we also use P to denote the expectation operator. Lower and upper probabilities of an event $A \subseteq \mathcal{X}$ are analogously defined as

$$\underline{P}(A) = \inf \left\{P(A) \mid P \in B_d^{\delta}(P_0)\right\} \quad \text{and} \quad \overline{P}(A) = \sup \left\{P(A) \mid P \in B_d^{\delta}(P_0)\right\} \tag{2}$$

and simply correspond to taking expectations over the indicator function of A.

In our companion paper, we analysed some properties of the credal set $B_d^\delta(P_0)$ and its associated lower prevision/probability for three of the usual distortion models within the imprecise probability theory: the pari mutuel (PMM), linear vacuous (LV) and constant odds ratio (COR) models. In this second part of our study, we will investigate the features of the polytopes $B_d^\delta(P_0)$ induced by classic distances such as the total variation, the Kolmogorov and the L_1 distances. To avoid having too many technicalities and unnecessary details in an already long study, we will assume that $P_0(\{x\}) > 0$ for all $x \in \mathcal{X}$, and also that δ will be chosen small enough such that the lower probability $\underline{P}$ induced by $B_d^\delta(P_0)$ in Equation (2) satisfies $\underline{P}(\{x\}) > 0 \ \forall x \in \mathcal{X}$ (or, equivalently, that $B_d^\delta(P_0)$ is included in $\mathbb{P}^*(\mathcal{X}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(\{x\}) > 0 \ \forall x\}$) . Some details about the general case are given in Appendix B.

Our analysis of these models shall be made in terms of a number of features[1]:

- The properties of the distance d.
- The number of extreme points of $B_d^\delta(P_0)$, which is finite as the considered distances induce polytopes.
- The k-monotonicity of the associated lower probability, that is satisfied if the inequality

$$\underline{P}\big(\cup_{i=1}^{p} A_i\big) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,p\}} (-1)^{|I|+1} \underline{P}\big(\cap_{i\in I} A_i\big) \tag{3}$$

  holds for any collection of p events $A_i \subseteq \mathcal{X}$ where $p \leq k$. For lower previsions, we will look more specifically at 2-monotonicity, that is satisfied whenever

$$\underline{P}(f \wedge g) + \underline{P}(f \vee g) \geq \underline{P}(f) + \underline{P}(g),$$

  holds for any pair of real-valued functions f, g over $\mathcal{X}$. The restriction to events (indicator functions) of a 2-monotone lower previsions is a 2-monotone lower probability. Also, for a 2-monotone lower probability $\underline{P}$ there exists a unique 2-monotone lower prevision $\underline{P}'$ whose restriction to events coincide with $\underline{P}$, and it is given by the Choquet integral [29].

  A lower probability satisfying the property of k-monotonicity for every k is usually called complete monotone.
- The behaviour of a family of distortion models under a conditioning on an event B, where we will consider the conditional lower prevision resulting from regular extension, defined as

$$\underline{P}_B(f) = \inf\{P_B(f) \mid P \in \mathcal{M}(\underline{P})\}. \tag{4}$$

This second part of our study is organised as follows: the neighbourhood and distortion models induced by the total variation, Kolmogorov and L_1 distances are respectively studied in Sections 2, 3 and 4. Finally, we propose a synthetic comparative analysis in Section 5, before discussing some conclusions and perspectives. To ease the reading, proofs and some additional results have been relegated to Appendices A and B, respectively.

[1] We refer the reader to the first part for a more detailed introduction to these notions.

2. Distortion model based on the total variation distance

In our companion paper we have seen that the pari mutuel, linear vacuous and constant odds ratio are models that, even if they can be expressed as neighbourhoods as in Equation (1), are defined without an explicit link to a distorting function d.

In the current and two forthcoming sections we go the other way: we start with a known distance between probabilities and study its induced neighbourhood as well as the properties of the associated lower probability/prevision.

We begin with the total variation distance. Given two probabilities $P, Q \in \mathbb{P}(\mathcal{X})$, their *total variation* is defined by [15]:

$$d_{TV}(P,Q) = \max_{A \subset \mathcal{X}} |P(A) - Q(A)|. \tag{5}$$

2.1. **Properties of the TV model.** We begin by studying the axiomatic properties of this distance; they are summarised in the following proposition.

Proposition 1. *The total variation distance d_{TV} given by Equation (5) satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.3, Ax.4 and Ax.5.*

Note also that by definition d_{TV} is bounded above by 1, so we can restrict our attention to $\delta \in (0,1)$. This motivates the following:

Definition 1. *Let P_0 be a probability measure and consider a distortion factor $\delta \in (0,1)$. Its associated* total variation model *$\underline{P}_{TV}$ is the lower envelope of the credal set $B^{\delta}_{d_{TV}}(P_0)$.*

The following result gives a formula for the coherent lower probability $\underline{P}_{TV}$ and its conjugate upper probability $\overline{P}_{TV}$:

Theorem 2. [13] *Consider a probability measure P_0 and a distortion parameter $\delta \in (0,1)$. The lower envelope $\underline{P}_{TV}$ of the credal set $B^{\delta}_{d_{TV}}(P_0)$ is given, for any $A \neq \emptyset, \mathcal{X}$, by:*

$$\underline{P}_{TV}(A) = \max\{0, P_0(A) - \delta\}, \quad \overline{P}_{TV}(A) = \min\{1, P_0(A) + \delta\}, \tag{6}$$

and $\underline{P}_{TV}(\emptyset) = \overline{P}_{TV}(\emptyset) = 0$, $\underline{P}_{TV}(\mathcal{X}) = \overline{P}_{TV}(\mathcal{X}) = 1$.

This result was established [13] for an arbitrary $\delta \in (0,1)$; our assumption $\underline{P}_{TV}(A) > 0$ for every $A \neq \emptyset$ means that

$$\delta < \min_{x \in \mathcal{X}} P_0(\{x\}), \tag{7}$$

whence

$$\underline{P}_{TV}(A) = P_0(A) - \delta \text{ and } \overline{P}_{TV}(A) = P_0(A) + \delta \tag{8}$$

for every $A \neq \emptyset, \mathcal{X}$.

Let us give a behavioural interpretation of this model: assume that P_0 is the fair price for a bet on A. Then, the expected benefit from the house would be:

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

In order to guarantee a positive gain, the house may impose a fixed tax of $\delta > 0$ that the gambler has to pay for betting, regardless on the event A that is bet on. Then, if the gambler wants to buy a bet on A, he/she has to pay $P_0(A) + \delta$ units. In this way, the benefit of the house becomes

$$(P_0(A) + \delta) - I_A,$$

which assures a positive expected gain:

$$E(P_0 + \delta - I_A) = (P_0(A) + \delta - 1)P_0(A) + (P_0(A) + \delta)(1 - P_0(A)) = \delta > 0.$$

This gives rise to the upper probability $\overline{P}(A) = P_0(A) + \delta$. Hence, we can interpret the distortion parameter δ as the fixed tax that the gambler has to pay for betting, regardless on the non-trivial event A to be bet on.

Next we establish that whenever the distortion factor δ satisfies Equation (7), we can find a probability in $B^{\delta}_{d_{TV}}(P_0)$ whose distance to P_0 is exactly δ.

Proposition 3. *Consider the total variation model associated with a probability measure P_0 and a distortion parameter $\delta \in (0,1)$ satisfying Equation (7). Then:*

$$\max_{P \in B^{\delta}_{d_{TV}}(P_0)} d_{TV}(P, P_0) = \delta.$$

Let us study the properties of $\underline{P}_{TV}$ as a lower probability.

Proposition 4. *The coherent lower probability $\underline{P}_{TV}$ is 2-monotone.*

However, it is neither completely monotone nor a probability interval in general, as our next example shows:

Example 1. *Consider $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, let P_0 be the uniform probability distribution and consider $\delta = 0.1$. From Equation (8), $\underline{P}_{TV}, \overline{P}_{TV}$ are given by:*

$\lvert A\rvert$	$\underline{P}_{TV}$	$\overline{P}_{TV}$
1	0.15	0.35
2	0.40	0.60
3	0.65	0.85
4	1	1

Since $P = (0.15, 0.15, 0.35, 0.35)$ satisfies that $P(\{x_i\}) \in [\underline{P}_{TV}(\{x_i\}), \overline{P}_{TV}(\{x_i\})]$ for every $i = 1, \ldots, 4$ but $P(\{x_1, x_2\}) = 0.3 < 0.4 = \underline{P}_{TV}(\{x_1, x_2\})$, we conclude that $\underline{P}_{TV}$ is not a probability interval. Also, taking $A_1 = \{x_1, x_2\}$, $A_2 = \{x_1, x_3\}$ and $A_3 = \{x_2, x_3\}$, we obtain:

$$\begin{aligned} &\underline{P}_{TV}(A_1) + \underline{P}_{TV}(A_2) + \underline{P}_{TV}(A_3) - \underline{P}_{TV}(A_1 \cap A_2) - \underline{P}_{TV}(A_1 \cap A_3) \\ &\qquad - \underline{P}_{TV}(A_2 \cap A_3) + \underline{P}_{TV}(A_1 \cap A_2 \cap A_3) = 3 \cdot 0.40 - 3 \cdot 0.15 + 0 = 0.75, \end{aligned}$$

while $\underline{P}_{TV}(A_1 \cup A_2 \cup A_3) = 0.65$. Hence, Equation (3) is not satisfied, so $\underline{P}_{TV}$ is not 3-monotone, and as a consequence it is not completely monotone either. ♦

The fact that $\underline{P}_{TV}$ is not completely monotone can be easily deduced from the results in [3]. Specifically, [3, Prop.5] and [3, Cor.1] give necessary and sufficient conditions on f for the distortion model of the type $f(P_0)$ to be completely monotone. In particular, f must be differentiable in the interval $[0,1)$. It can be easily checked that $\underline{P}_{TV}(A) = f(P_0(A))$, where $f(t) = \max\{t - \delta, 0\}$. Since f is not differentiable, it follows that $\underline{P}_{TV}$ cannot be completely monotone. We refer to [3, App. B] for some interesting results on the k-monotonicity of the distortion models of the type $f(P_0)$.

Since $\underline{P}_{TV}$ is a 2-monotone lower probability, it has a unique 2-monotone extension to gambles, that can be determined by means of the Choquet integral [30, Sec. 3.2.4]:

$$\underline{P}_{TV}(f) = \sup f - \int_{\inf f}^{\sup f} \overline{F}_f(x)\mathrm{d}x = \inf f + \int_{\inf f}^{\sup f} 1 - \overline{F}_f(x)\mathrm{d}x,$$
$$\overline{P}_{TV}(f) = \sup f - \int_{\inf f}^{\sup f} \underline{F}_f(x)\mathrm{d}x = \inf f + \int_{\inf f}^{\sup f} 1 - \underline{F}_f(x)\mathrm{d}x,$$

where $\underline{F}_f$ and $\overline{F}_f$ denote the lower and upper distributions of f under $\underline{P}_{TV}, \overline{P}_{TV}$:

$$\overline{F}_f(x) = \overline{P}_{TV}(f \leq x), \qquad \underline{F}_f(x) = \underline{P}_{TV}(f \leq x) \quad \forall x \in \mathcal{X}.$$

An equivalent expression is given in [24, Sec. 3.2].

Next we establish the number of extreme points of $B^{\delta}_{d_{TV}}(P_0)$.

Proposition 5. *Let $B^{\delta}_{d_{TV}}(P_0)$ be the neighbourhood model associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$ satisfying Equation (7) by means of the total variation distance. Then the number of extreme points of $B^{\delta}_{d_{TV}}(P_0)$ is $n(n-1)$.*

Proposition 5 has been adapted to the general case in which δ does not satisfy Equation (7) in Appendix B.1.

2.2. **Conditioning the TV model.** Next proposition shows that when conditioning according to Equation (4), the resulting credal set is still a total variation model.

Proposition 6. *Consider the model $B^{\delta}_{d_{TV}}(P_0)$, where the distortion parameter δ satisfies Equation (7), and its induced lower probability $\underline{P}_{TV}$. Then, the credal set associated with the conditional model $\underline{P}_B$ ($B \neq \emptyset, \mathcal{X}$) coincides with the credal set $B^{\delta_B}_{d_{TV}}(P_{0|B})$, where*

$$P_{0|B}(A) = P_0(A|B) \quad \textit{and} \quad \delta_B = \frac{\delta}{P_0(B)}. \tag{9}$$

As with some of the models studied in the first part of this study (i.e., the PMM and LV models), we can see that the conditional model has an increased imprecision with respect to the original one because $\delta_B = \frac{\delta}{P_0(B)} > \delta$, this increase being between the one we observed in the PMM ($\delta_B = \frac{\delta}{\underline{P}_{PMM}(B)}$) and that of the LV model ($\delta_B = \frac{\delta}{\overline{P}_{LV}(B)}$); see Equations (16) and (19) of the first paper [19].

3. Distortion model based on the Kolmogorov distance

Let us now consider the neighbourhood model based on the Kolmogorov distance [14], that makes a comparison between the distribution functions associated with the probability measures. We shall assume in this section that the finite possibility space $\mathcal{X} = \{x_1, \ldots, x_n\}$ is totally ordered: $x_1 < x_2 < \ldots < x_n$. Given two probability measures P, Q defined on an ordered space $\mathcal{X}$, their *Kolmogorov distance* is defined by:

$$d_K(P,Q) = \max_{x \in \mathcal{X}} |F_P(x) - F_Q(x)|, \tag{10}$$

where F_P and F_Q denote the cumulative distribution functions associated with P and Q, respectively.

When dealing with this model, it will be useful to consider another of the usual model within the imprecise probability theory: p-boxes [11]. A *p-box* is a pair of cumulative distribution functions $\underline{F}, \overline{F} : \mathcal{X} \to [0,1]$ such that $\underline{F} \leq \overline{F}$. A p-box, usually denoted by $(\underline{F}, \overline{F})$, defines a closed and convex set of probabilities by:

$$\mathcal{M}(\underline{F}, \overline{F}) = \{P \in \mathbb{P}(\mathcal{X}) \mid \underline{F} \leq F_P \leq \overline{F}\}, \tag{11}$$

where F_P denotes the cumulative distribution function induced by P. In particular, a p-box $(\underline{F}, \overline{F})$ defines a coherent lower probability taking the lower envelope of its associated credal set in Equation (11):

$$\underline{P}_{(\underline{F}, \overline{F})}(A) = \inf\{P(A) \mid P \in \mathcal{M}(\underline{F}, \overline{F})\}. \tag{12}$$

This lower probability is not only coherent, but also completely monotone, and hence in particular it is 2-monotone. We refer to [28] for a mathematical study of p-boxes.

3.1. **Properties of the Kolmogorov model.** Our first result studies the properties satisfied by d_K.

Proposition 7. *The Kolmogorov distance given by Equation* (10) *satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.3, Ax.4 and Ax.5.*

Also, d_K is a distance that is bounded above by 1. This motivates the following definition:

Definition 2. *Let P_0 be a probability measure and consider a distortion factor $\delta \in (0,1)$. Its associated* Kolmogorov model *$\underline{P}_K$ is the lower envelope of the credal set $B^{\delta}_{d_K}(P_0)$.*

In other words, the coherent lower probability $\underline{P}_K$ obtained as the lower envelope of $B^{\delta}_{d_K}(P_0)$ is given by:

$$\underline{P}_K(A) = \min\{P(A) \mid P \in B^{\delta}_{d_K}(P_0)\} = \min\left\{P(A) \mid \max_{x \in \mathcal{X}} |F_P(x) - F_{P_0}(x)| \leq \delta\right\}$$

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

There exists an obvious connection between the credal sets $B^{\delta}_{d_{TV}}(P_0)$ and $B^{\delta}_{d_K}(P_0)$: since $d_K(P,Q) \leq d_{TV}(P,Q)$ for every $P, Q \in \mathbb{P}(\mathcal{X})$, we deduce that $B^{\delta}_{d_K}(P_0) \supseteq B^{\delta}_{d_{TV}}(P_0)$. However, the two sets do not coincide in general:

Example 2. *Consider a four-element space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, let P_0 be the uniform distribution and take $\delta = 0.1$. Consider also the probability P given by $P = (0.35, 0.05, 0.35, 0.25)$. Then P belongs to $B^{\delta}_{d_K}(P_0)$, because*

	x_1	x_2	x_3	x_4
F_P	0.35	0.4	0.75	1
F_{P_0}	0.25	0.5	0.75	1
$\lvert F_P - F_{P_0}\rvert$	0.1	0.1	0	0

However, P does not belong to $B^{\delta}_{d_{TV}}(P_0)$, since

$$\underline{P}_{TV}(\{x_2\}) = P_0(\{x_2\}) - \delta = 0.15 > 0.05 = P(\{x_2\}).$$

Thus, $B^{\delta}_{d_K}(P_0)$ is a strict superset of $B^{\delta}_{d_{TV}}(P_0)$. ♦

From this we also deduce that the behavioural interpretation of the Kolmogorov model is the same than that of the total variation but restricted to the cumulative events $\{x_1, \dots, x_k\}$ for $k = 1, \dots, n$.

The assumption of strictly positive lower probabilities implies that in particular $\underline{P}_K(\{x\}) > 0$ for every $x \in \mathcal{X}$, whence (see [28, Prop. 4]):

$$\underline{P}_K(\{x_1\}) > 0 \Rightarrow \delta < F_{P_0}(x_1),$$
$$\underline{P}_K(\{x_i\}) > 0 \Rightarrow \delta < \frac{1}{2}\left(F_{P_0}(x_i) - F_{P_0}(x_{i-1})\right) \quad \forall i = 2, \dots, n.$$

From this, we deduce that the assumption of strictly positive lower probabilities implies that:

$$\delta < \min_{1,\dots,n} \frac{1}{2}\left(F_{P_0}(x_i) - F_{P_0}(x_{i-1})\right), \tag{13}$$

where $F_{P_0}(x_0) := 0$.

Next we establish that any such distortion factor is attained in the neighbourhood model.

Proposition 8. *Consider the Kolmogorov model associated with a probability measure P_0 and a distortion parameter $\delta \in (0,1)$ satisfying Equation (13). Then:*

$$\max_{P \in B^{\delta}_{d_K}(P_0)} d_K(P, P_0) = \delta.$$

Since the Kolmogorov distance is defined over cumulative distributions, its associated credal set is equivalent to a p-box. To see this, note that we can rewrite

$$\begin{aligned} B^{\delta}_{d_K}(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid |F_P(x) - F_{P_0}(x)| \leq \delta,\ \forall x \in \mathcal{X}\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid F_{P_0}(x) - \delta \leq F_P(x) \leq F_{P_0}(x) + \delta,\ \forall x \in \mathcal{X}\}; \end{aligned}$$

if we define a p-box $(\underline{F}, \overline{F})$ by:

$$\underline{F}(x_i) = \max\{0, F_{P_0}(x_i) - \delta\}, \quad \overline{F}(x_i) = \min\{1, F_{P_0}(x_i) + \delta\} \quad \forall i = 1, \dots, n-1,$$
$$\underline{F}(x_n) = \overline{F}(x_n) = 1,$$

then $B^{\delta}_{d_K}(P_0) = \mathcal{M}(\underline{F}, \overline{F})$, where $\mathcal{M}(\underline{F}, \overline{F})$ is given in Equation (11). Also, since δ satisfies Equation (13), the expression of $(\underline{F}, \overline{F})$ simplifies to:

$$\begin{aligned} &\underline{F}(x_i) = F_{P_0}(x_i) - \delta, \quad \overline{F}(x_i) = F_{P_0}(x_i) + \delta \quad \forall i = 1, \dots, n-1, \\ &\underline{F}(x_n) = \overline{F}(x_n) = 1. \end{aligned} \tag{14}$$

This means that $B^{\delta}_{d_K}(P_0)$ is the credal set associated with a p-box, and as a consequence $\underline{P}_K$ is given by Equation (12), and it is completely monotone [28, Thm. 17]. Note that, although $\underline{P}_K$ and $\underline{P}_{TV}$ induce the same p-box, they are not the same coherent lower probability: $\underline{P}_K \lneq \underline{P}_{TV}$, as we have seen in Example 2. If we follow the notation in [4, 20], we deduce from [21, Prop. 16] that $\underline{P}_K$ is the unique undominated outer approximation of $\underline{P}_{TV}$ in terms of p-boxes.

The p-box induced by the Kolmogorov distance in Example 2 is represented in Figure 1.

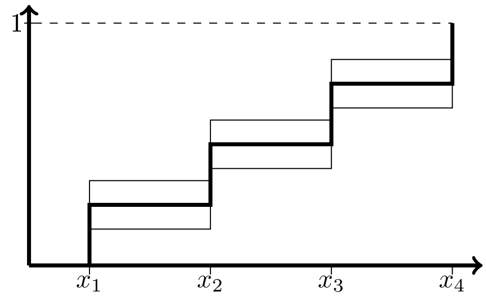

FIGURE 1. Cumulative distribution function F_{P_0} (bold line) and p-box $(\underline{F}, \overline{F})$ associated with the distortion model in Example 2.

Example 3. *Example 2 also shows that $\underline{P}_K$ is not a probability interval in general. If P_0 is the uniform distribution and $\delta = 0.1$, it holds that (see* [28, Prop. 4] *for details in how to compute $\underline{P}_K$ and $\overline{P}_K$):*

$$\underline{P}_K(\{x_1\}) = \underline{P}_K(\{x_4\}) = 0.15, \quad \underline{P}_K(\{x_2\}) = \underline{P}_K(\{x_3\}) = 0.05,$$
$$\overline{P}_K(\{x_1\}) = \overline{P}_K(\{x_4\}) = 0.35, \quad \overline{P}_K(\{x_2\}) = \overline{P}_K(\{x_3\}) = 0.45.$$

If we consider the probability measure P determined by the probability mass function $(0.35, 0.45, 0.05, 0.15)$, it satisfies $P(\{x_i\}) \in [\underline{P}(\{x_i\}), \overline{P}(\{x_i\})]$ for every $i = 1, \ldots, 4$. However, $F_P(x_2) - F_{P_0}(x_2) = 0.8 - 0.5 = 0.3 > 0.1$. Thus, $P \notin B^{\delta}_{d_K}(P_0)$ and therefore $\underline{P}_K$ is not a probability interval. ♦

Finally, we investigate the maximal number of extreme points in $B^{\delta}_{d_K}(P_0)$. It is known [17, Thm. 17] that the maximal number of extreme points induced by a p-box coincides with the n-th Pell number, where n is the cardinality of $\mathcal{X}$. The Pell numbers are recursively defined by:

$$\mathcal{P}_0 = 0, \quad \mathcal{P}_1 = 1, \quad \mathcal{P}_n = \mathcal{P}_{n-2} + 2\mathcal{P}_{n-1} \quad \forall n \geq 2.$$

Our next result shows that this maximal number is also attained in the p-boxes associated with the Kolmogorov models.

Proposition 9. *Let $B^{\delta}_{d_K}$ be the neighbourhood model associated with a probability measure P_0 and a distortion factor $\delta \in (0, 1)$ satisfying Equation* (13) *by means of the Kolmogorov distance. Then the number of extreme points of $B^{\delta}_{d_K}$ is $\mathcal{P}_n$.*

As we show in the proof in Appendix A, this maximal number of extreme points is attained for instance when P_0 is the uniform distribution and $\delta \in \left(\frac{1}{2n}, \frac{1}{n}\right)$.

Moreover, it is worth noting that when our assumption that $B^{\delta}_{d_k} \subseteq \mathbb{P}^*(\mathcal{X})$ does not hold, the maximal number of extreme points of the Kolmogorov model is difficult to compute, but it may be smaller than the Pell number $\mathcal{P}_n$. One trivial instance of this would be when δ is large enough so that $B^{\delta}_{d_k} = \mathbb{P}(\mathcal{X})$, in which case there would be n extreme points.

3.2. **Conditioning the Kolmogorov model.** For the Kolmogorov distance, known results [9] indicate that the conditional model obtained by applying the regular extension to the lower probability induced by a generalised p-box will not induce, in general, a generalised p-box. A simple adaptation of Example 3 in [9] (picking

$\underline{F}(x_2) = 0.3$ instead of 0.4) shows that it is also the case for p-boxes induced by the Kolmogorov distance.

4. Distortion model based on the L_1 distance

We conclude our analysis by considering the distortion model where the distorting function is the L_1-distance between probability measures, that for every $P, Q \in \mathbb{P}(\mathcal{X})$ is given by

$$d_{L_1}(P, Q) = \sum_{A \subset \mathcal{X}} |P(A) - Q(A)|.$$

This distance has been used for instance within robust statistics [25]. It seems therefore worthwhile to investigate the properties of the neighbourhood model it determines.

4.1. Properties of the L_1 model.

Our first result gives the properties satisfied by d_{L_1}.

Proposition 10. *The L_1-distance d_{L_1} satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2, Ax.3, Ax.4 and Ax.5.*

Using the L_1-distance, we can define the following distortion model.

Definition 3. *Let P_0 be a probability measure and consider a distortion parameter $\delta > 0$. Its associated L_1-*model *$\underline{P}_{L_1}$ is the lower prevision obtained as the lower envelope of the credal set $B^{\delta}_{d_{L_1}}(P_0)$.*

The restriction to events of the lower prevision $\underline{P}_{L_1}$ associated with this model is given in the next result. In the statement of this theorem, especially in Equations (15) and (16), recall that n denotes the cardinality of $\mathcal{X}$: $n = |\mathcal{X}|$.

Theorem 11. *Consider a probability measure P_0 and a distortion parameter $\delta > 0$. Let us define the lower probability $\underline{P}_{L_1}$ by*

$$\underline{P}_{L_1}(A) = P_0(A) - \frac{\delta}{\varphi(n, |A|)}, \quad \forall A \neq \mathcal{X} \tag{15}$$

and $\underline{P}_{L_1}(\mathcal{X}) = 1$, where, for any $k = 1, \ldots, n-1$, $\varphi(n, k)$ is given by:

$$\varphi(n, k) = \sum_{l=0}^{k} \binom{k}{l} \sum_{j=0}^{n-k} \binom{n-k}{j} \left| \frac{l}{k} - \frac{j}{n-k} \right|, \tag{16}$$

and $\varphi(n, n) = 0$.

If δ is small enough so that $\underline{P}_{L_1}(\{x\}) > 0$ for every $x \in \mathcal{X}$, then $\underline{P}_{L_1}$ is the restriction to events of the lower envelope of the credal set $B^{\delta}_{d_{L_1}}(P_0)$.

Let us comment on the meaning of the function $\varphi(n, k)$. Given events $A, B \subseteq \mathcal{X}$, denote by $\varphi^*_{\mathcal{X}}(A, B)$ the function given by:

$$\varphi^*_{\mathcal{X}}(A, B) = \left| \frac{|A^c \cap B|}{|A^c|} - \frac{|A \cap B|}{|A|} \right|.$$

This function measures the difference, in absolute value, between the common elements between A and B, relative to the size of A, and the common elements between A^c and B, relative to the size of A^c. Then, we can define $\varphi^*_{\mathcal{X}}(A)$ by:

$$\varphi^*_{\mathcal{X}}(A) = \sum_{B \subseteq \mathcal{X}} \varphi^*_{\mathcal{X}}(A, B).$$

If the cardinality of A is k, $|A| = k$, any $B \subseteq \mathcal{X}$ has l elements in common with A, for some $l = 0, \dots, k$, and j elements in common with A^c, for some $j = 0, \dots, n-k$. Hence, taking into account that such B could be chosen in $\binom{k}{l} \cdot \binom{n-k}{j}$ different ways, we obtain that $\varphi^*_{\mathcal{X}}(A) = \varphi(n, |A|)$.

Table 1 provides the values of φ up to cardinality $n = 12$.

$n \backslash k$	1	2	3	4	5	6	7	8	9	10	11	12
2	2	-	-	-	-	-	-	-	-	-	-	-
3	4	4	-	-	-	-	-	-	-	-	-	-
4	8	6	8	-	-	-	-	-	-	-	-	-
5	16	12	12	16	-	-	-	-	-	-	-	-
6	32	22	20	22	32	-	-	-	-	-	-	-
7	64	44	40	40	44	64	-	-	-	-	-	-
8	128	84	76	70	76	84	128	-	-	-	-	-
9	256	168	144	140	140	144	168	256	-	-	-	-
10	512	326	288	264	252	264	288	326	512	-	-	-
11	1024	652	564	520	504	504	520	564	652	1024	-	-
12	2048	1276	1100	998	964	924	964	998	1100	1276	2048	-

TABLE 1. Values of $\varphi(n, k)$ for $n \leq 12$.

Throughout the remainder of this section, we will assume that δ is small enough so that $\underline{P}_{L_1}(A) > 0$ for every $A \neq \emptyset$. This restriction on δ can be equivalently expressed as

$$0 < \delta < \min_{A \subset \mathcal{X}} P_0(A)\varphi(n, |A|). \tag{17}$$

Comparing Equations (8) and (15), the only difference between them is that in $\underline{P}_{TV}$ we always subtract the distortion parameter δ, while in $\underline{P}_{L_1}$ we subtract $\frac{\delta}{\varphi(n,|A|)}$, taking into account the size of event A. Using this fact, the behavioural interpretation of the L_1-model is similar to that of the total variation. However, note that in the total variation model the gambler has to pay always the same tax δ for betting on any event, while in the case of the L_1-model, the tax depends on the size of the event, with the constraint that events of the same cardinality have the same tax.

Our next result shows that for the L_1 model, there is always a probability measure P in the ball $B^{\delta}_{d_{L_1}}(P_0)$ such that $d_{L_1}(P, P_0) = \delta$.

Proposition 12. *Consider the L_1 model associated with a probability measure P_0 and a distortion parameter $\delta > 0$ satisfying Equation (17). Then:*

$$\max_{P \in B^{\delta}_{d_{L_1}}(P_0)} d_{L_1}(P, P_0) = \delta.$$

Let us now study the properties of the lower envelope of the L_1-model. We begin by establishing that the lower prevision $\underline{P}_{L_1}$ associated with the L_1-distance is not a 2-monotone lower prevision.

Example 4. *Consider $\mathcal{X}=\{x_1,x_2,x_3,x_4\}$, P_0 the uniform probability measure and $\delta=1$. Let f be the gamble given by $f=4I_{\{x_1\}}+3I_{\{x_2\}}+2I_{\{x_3\}}+I_{\{x_4\}}$. If $\underline{P}_{L_1}$ was a 2-monotone lower prevision, then it would be given by the Choquet integral associated with its restriction to events, that produces*

$$\begin{aligned}(C)\int fd\underline{P}_{L_1} &= \underline{P}_{L_1}(\{x_1\})+\underline{P}_{L_1}(\{x_1,x_2\})+\underline{P}_{L_1}(\{x_1,x_2,x_3\})+\underline{P}_{L_1}(\{x_1,x_2,x_3,x_4\})\\ &= \frac{3}{24}+\frac{8}{24}+\frac{15}{24}+1=\frac{50}{24},\end{aligned}$$

using Equation (15). However, there is no $Q\in B^1_{d_{L_1}}(P_0)$ such that $Q(f)=(C)\int fd\underline{P}_{L_1}$: such a Q should satisfy

$$\begin{aligned}&Q(\{x_1\})=\underline{P}_{L_1}(\{x_1\}), \quad Q(\{x_1,x_2\})=\underline{P}_{L_1}(\{x_1,x_2\}),\\ &Q(\{x_1,x_2,x_3\})=\underline{P}_{L_1}(\{x_1,x_2,x_3\}), \quad Q(\{x_1,x_2,x_3,x_4\})=\underline{P}_{L_1}(\{x_1,x_2,x_3,x_4\}),\end{aligned}$$

meaning that it should be $Q=(\frac{3}{24},\frac{5}{24},\frac{7}{24},\frac{9}{24})$. This probability measure satisfies $d_{L_1}(P_0,Q)=\frac{28}{24}$, and as a consequence it does not belong to $B^1_{d_{L_1}}(P_0)$. ♦

The study of the k-monotonicity of the restriction of $\underline{P}_{L_1}$ to events can be given in terms of the function $\varphi(n,k)$, because:

$$\begin{aligned}\underline{P}\big(\cup_{i=1}^p A_i\big) - \sum_{\emptyset\neq I\subseteq\{1,\dots,p\}} (-1)^{|I|+1}\underline{P}\big(\cap_{i\in I}A_i\big) &= P_0\big(\cup_{i=1}^p A_i\big)-\frac{\delta}{\varphi(n,|\cup_{i=1}^p A_i|)}\\ &-\sum_{\emptyset\neq I\subseteq\{1,\dots,p\}}(-1)^{|I|+1}\left(P_0(\cap_{i\in I}A_i)-\frac{\delta}{\varphi(n,|\cap_{i\in I}A_i|)}\right)\\ &=-\frac{\delta}{\varphi(n,|\cup_{i=1}^p A_i|)}+\sum_{\emptyset\neq I\subseteq\{1,\dots,p\}}(-1)^{|I|+1}\frac{\delta}{\varphi(n,|\cap_{i\in I}A_i|)}.\end{aligned}$$

Hence, Equation (3) holds if and only if

$$\frac{1}{\varphi(n,|\cup_{i=1}^p A_i|)} \leq \sum_{\emptyset\neq I\subseteq\{1,\dots,p\}}(-1)^{|I|+1}\frac{1}{\varphi(n,|\cap_{i\in I}A_i|)} \tag{18}$$

for every $A_1,\dots,A_p\subset\mathcal{X}$ and $p\leq k$. In particular, taking $k=2$, $\underline{P}_{L_1}$ is 2-monotone in events if and only if:

$$\frac{1}{\varphi(n,|A\cup B|)}+\frac{1}{\varphi(n,|A\cap B|)}\leq\frac{1}{\varphi(n,|A|)}+\frac{1}{\varphi(n,|B|)}$$

for every $A,B\subset\mathcal{X}$. Also, this inequality is equivalent to:

$$\frac{1}{\varphi(n,k_1)}+\frac{1}{\varphi(n,k_4)}\leq\frac{1}{\varphi(n,k_2)}+\frac{1}{\varphi(n,k_3)} \tag{19}$$

for every $1\leq k_1\leq k_2\leq k_3\leq k_4\leq n$ such that $k_1+k_4=k_2+k_3$. These facts simplify the study of the properties of the restriction to events of $\underline{P}_{L_1}$.

Proposition 13. *Let $\underline{P}_{L_1}$ be the L_1-model generated by a probability measure P_0 and a distortion parameter $\delta>0$ satisfying Equation (17) in a n-element space $\mathcal{X}$.*

- *The restriction of $\underline{P}_{L_1}$ to events is a 2-monotone lower probability for $n\leq 11$.*
- *The restriction of $\underline{P}_{L_1}$ to events is a completely monotone lower probability for $n\leq 4$.*

Next example shows that the restriction of $\underline{P}_{L_1}$ to events is not completely monotone for cardinalities larger than $n = 4$.

Example 5. *Consider a 5-element space* $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$*, let* P_0 *be any probability measure and* $\delta > 0$ *be a distortion parameter satisfying Equation* (17)*. Consider now the events* $A_1 = \{x_1, x_2\}$*,* $A_2 = \{x_1, x_3\}$ *and* $A_3 = \{x_2, x_3\}$*. Applying Equation* (18) *to these events, we obtain the following inequality:*

$$\frac{1}{\varphi(5,3)} \leq \frac{3}{\varphi(5,2)} - \frac{3}{\varphi(5,1)}.$$

If we now replace the values of φ*, this is equivalent to:*

$$\frac{1}{12} \leq \frac{3}{12} - \frac{3}{16} \Leftrightarrow \frac{3}{16} \leq \frac{2}{12},$$

which does not hold. We conclude that Equation (18) *is not satisfied, so the restriction of* $\underline{P}_{L_1}$ *to events is not a 3-monotone lower probability, and as a consequence it is not completely monotone either.* ♦

Also, Proposition 13 shows that $\underline{P}_{L_1}$ is 2-monotone on events for $n \leq 11$. Somewhat surprisingly, it is not a 2-monotone lower probability in general for larger cardinalities:

Example 6. *Consider* $\mathcal{X} = \{x_1, \ldots, x_{12}\}$*, let* P_0 *be the uniform distribution and* $\delta = 1$*. Take* $A = \{x_1, x_2, x_3, x_4, x_5\}$ *and* $B = \{x_1, x_2, x_3, x_4, x_6\}$*. From Equation* (16)*,*

$$\varphi(12,4) = 998, \quad \varphi(12,5) = 964 \quad \textit{and} \quad \varphi(12,6) = 924.$$

Applying Proposition 11, we obtain

$$\underline{P}_{L_1}(A \cup B) + \underline{P}_{L_1}(A \cap B) = \frac{6}{12} - \frac{1}{924} + \frac{4}{12} - \frac{1}{998} < 2 \cdot \left(\frac{5}{12} - \frac{1}{964}\right) = \underline{P}_{L_1}(A) + \underline{P}_{L_1}(B),$$

whence the restriction to events of $\underline{P}_{L_1}$ *is not a 2-monotone lower probability. As a consequence, it is not a probability interval either.* ♦

To conclude the study of the L_1-model, it only remains to determine the extreme points of the closed ball $B^{\delta}_{d_{L_1}}(P_0)$. It is not difficult to see that these must lie on the boundary of the ball. However, its maximal number and an explicit formula is an open problem at this stage. As we have seen, $\underline{P}_{L_1}$ is neither 2-monotone in gambles (see Example 4) nor 2-monotone in events for cardinalities greater $n \geq 12$ (see Example 6). This means that for the general case there is not a simple procedure for determining the extreme points of $B^{\delta}_{d_{L_1}}(P_0)$, and in particular we cannot use the procedure based in the permutations, described in [19, Eq. (2)], for computing them. Also, the fact that complete monotonicity is not guaranteed for $n > 4$ suggests that geometrical intuitions we may get from a 3 state space may be misleading.

4.2. **Conditioning the L_1 model.** Let us first look at what happens when we condition a L_1-model over some event B. For the L_1-distance, we have to mention that since $\underline{P}_{L_1}$ is not 2-monotone (on gambles nor or events), we cannot use some of the useful results for conditioning from [16]. Furthermore, even in the case of a cardinality smaller than 12, where from Proposition 13 we know that $\underline{P}_{L_1}$ is 2-monotone in events, the conditional model is not necessarily a L_1-model, as next example shows:

Example 7. *Consider a five element space $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$, a probability measure P_0 with uniform distribution and a distortion parameter $\delta = 1$. The values of $\underline{P}_{L_1}$ and $\overline{P}_1$ are:*

$\lvert A\rvert$	1	2	3	4	5
$\underline{P}_{L_1}$	$11/80$	$19/60$	$31/60$	$59/80$	1
$\overline{P}_{L_1}$	$21/80$	$29/60$	$41/60$	$69/80$	1

Since $n = 5 < 12$, from Proposition 13 we know that $\underline{P}_{L_1}$ is 2-monotone. Hence, the conditional model $\underline{P}_B$, where $B = \{x_1, x_2, x_3, x_4\}$, can be computed using the formula [29, Thm. 7.2]*:*

$$\underline{P}_B(A) = \frac{\underline{P}_{L_1}(A \cap B)}{\underline{P}_{L_1}(A \cap B) + \overline{P}_{L_1}(A^c \cap B)} \quad \forall A \subseteq B,$$

that gives

$\lvert A\rvert$	1	2	3	4
$\underline{P}_B$	$33/197$	$19/48$	$124/187$	1

If we assume that $\underline{P}_B$ is a L_1-model, there must exist a probability measure $P_0' = (p_1, p_2, p_3, p_4)$ and a distortion parameter $\delta' > 0$ such that:

$$\underline{P}_B(A) = \begin{cases} P_0'(A) - \frac{\delta'}{8} & \text{if } |A| = 1. \\ P_0'(A) - \frac{\delta'}{6} & \text{if } |A| = 2. \\ P_0'(A) - \frac{\delta'}{8} & \text{if } |A| = 3. \end{cases}$$

Then:

$$4 \times \frac{33}{197} = \sum_{i=1}^{4} \underline{P}_B(\{x_i\}) = \sum_{i=1}^{4} \left(P_0'(\{x_i\}) - \frac{\delta'}{8} \right) = 1 - \frac{\delta'}{2} \Longrightarrow \delta' = \frac{130}{197}.$$

Also:

$$6 \times \frac{19}{48} = \sum_{A:|A|=2} \underline{P}_B(A) = \sum_{A:|A|=2} \left(P_0'(A) - \frac{\delta'}{6} \right) = 3 - \delta' \Longrightarrow \delta' = \frac{5}{8}.$$

This is a contradiction, meaning that $\underline{P}_B$ is not a L_1-model. ♦

5. Comparative and synthetic analysis of the distortion models

In our companion paper we have seen that some usual distortion models within the imprecise probability theory (the pari mutuel, lineal vacuous and constant odds ratio) can be expressed as neighbourhoods of probabilities. In this second paper, we have studied the neighbourhoods induced by the total variation, Kolmogorov and L_1 distances.

In this section we compare all these models from different points of view: the amount of imprecision, the properties of the distorting function, the properties of the lower probability associated with each model, the complexity of each model and their behaviour under conditioning.

5.1. **Amount of imprecision.** We first compare the amount of imprecision that is introduced by the different neighbourhood models we have considered in these two papers once the initial probability measure P_0 and the distortion factor $\delta \in (0,1)$ are fixed. Given two credal sets $\mathcal{M}_1, \mathcal{M}_2$, we shall say that $\mathcal{M}_1$ is *more informative* than $\mathcal{M}_2$ when $\mathcal{M}_1 \subseteq \mathcal{M}_2$; in terms of their lower envelopes $\underline{P}_1, \underline{P}_2$, this means that $\underline{P}_1(f) \geq \underline{P}_2(f)$ for every gamble f on $\mathcal{X}$.

We start with an example showing that some of the models are not related in general.

Example 8. *Consider $\mathcal{X} = \{x_1, x_2, x_3\}, P_0 = (0.5, 0.3, 0.2)$ and $\delta = 0.1$. Then the associated distortion models are:*

A	$\underline{P}_{PMM}(A)$	$\underline{P}_{LV}(A)$	$\underline{P}_{COR}(A)$	$\underline{P}_{TV}(A)$	$\underline{P}_K(A)$	$\underline{P}_{L_1}$
$\{x_1\}$	0.45	0.45	0.4737	0.4	0.4	0.475
$\{x_2\}$	0.23	0.27	0.2784	0.2	0.1	0.275
$\{x_3\}$	0.12	0.18	0.1837	0.1	0.1	0.175
$\{x_1, x_2\}$	0.78	0.72	0.7826	0.7	0.7	0.775
$\{x_1, x_3\}$	0.67	0.63	0.6774	0.6	0.5	0.675
$\{x_2, x_3\}$	0.45	0.45	0.4737	0.4	0.4	0.475

where we are assuming the order $x_1 < x_2 < x_3$ in the case of the Kolmogorov model. By considering the events $A = \{x_2\}$ and $B = \{x_1, x_2\}$, we observe that the pari mutuel model and the linear vacuous mixture are not comparable, in the sense that none of them is more imprecise than the other.

Also, this example tells us that $\underline{P}_{L_1}$ neither dominates nor is dominated by $\underline{P}_{PMM}$, considering the events $A = \{x_1\}$ and $B = \{x_1, x_2\}$. By considering the events $A = \{x_1\}$ and $B = \{x_3\}$, we also observe that $\underline{P}_{L_1}$ neither dominates nor is dominated by $\underline{P}_{LV}$ or $\underline{P}_{COR}$. ♦

It was already stated in [30, Sec. 2.9.4] that the restriction to events of $\underline{P}_{COR}$, denoted by $\underline{Q}_{COR}$, dominates the lower probabilities of both the linear vacuous, $\underline{P}_{LV}$, and the pari mutuel $\underline{P}_{PMM}$. In terms of credal sets,

$$B^{\delta}_{d'_{COR}}(P_0) \subseteq B^{\delta}_{d_{LV}}(P_0) \cap B^{\delta}_{d_{PMM}}(P_0).$$

Since $\underline{Q}_{COR}$ is the restriction to events of $\underline{P}_{COR}$, it holds that $B^{\delta}_{d_{COR}}(P_0) \subseteq B^{\delta}_{d'_{COR}}(P_0)$.

Next we compare these models with the one associated with the total variation:

Proposition 14. *Consider the coherent lower probabilities $\underline{P}_{PMM}$, $\underline{P}_{LV}$ and $\underline{P}_{TV}$ induced by the probability measure P_0 and the distortion factor $\delta \in (0,1)$. It holds that*

$$B^{\delta}_{d_{PMM}}(P_0) \cup B^{\delta}_{d_{LV}}(P_0) \subseteq B^{\delta}_{d_{TV}}(P_0).$$

In terms of their associated coherent lower probabilities, we can equivalently state that $\underline{P}_{TV} \leq \min\{\underline{P}_{PMM}, \underline{P}_{LV}\}$.

In other words, the total variation model is more imprecise than both the pari mutuel and linear vacuous, and as a consequence it is also more imprecise than the constant odds ratio.

On the other hand, taking into account the comments given in Section 3, we observe that the model based on the Kolmogorov distance is more imprecise than the total variation distance: $B^{\delta}_{d_{TV}}(P_0) \subseteq B^{\delta}_{d_K}(P_0)$.

Finally, since $d_{TV}(P,Q) \leq d_{L_1}(P,Q)$ for any pair of probability measures P, Q, we deduce that $B^{\delta}_{d_{L_1}}(P_0) \subseteq B^{\delta}_{d_{TV}}(P_0) \subseteq B^{\delta}_{d_K}(P_0)$. Thus, $\underline{P}_{L_1}$ is more precise than the distortion models associated with the total variation and the Kolmogorov distance.

These relationships are summarised in Figure 2.

$$\begin{tikzcd}
 & & B^{\delta}_{d_K}(P_0) \arrow[d] & & \\
 & & B^{\delta}_{d_{TV}}(P_0) \arrow[dll] \arrow[d] \arrow[drr] & & \\
B^{\delta}_{d_{PMM}}(P_0) \arrow[dr] & & B^{\delta}_{d_{LV}}(P_0) \arrow[dl] & & B^{\delta}_{d_{L_1}}(P_0) \\
 & B^{\delta}_{d'_{COR}}(P_0) \arrow[d] & & & \\
 & B^{\delta}_{d_{COR}}(P_0) & & &
\end{tikzcd}$$

FIGURE 2. Relationships between the different models. An arrow between two nodes means that parent includes the child.

The credal sets of the neighbourhood models from Example 8 are represented in Figure 3.

5.2. **Properties of the distorting function.** We may also compare the different models by means of the properties of the associated distorting function, as introduced in [19, Sec. 3]. Our models satisfy the ones in Table 2.

Model	Ax.1	Ax.1a	Ax.1b	Ax.2	Ax.3	Ax.4	Ax.5	Result
d_{PMM}	✓	✓	✓			✓	✓	[19, Prop. 6]
d_{LV}	✓	✓	✓	✓		✓	✓	[19, Prop. 10]
d_{COR}	✓	✓	✓	✓	✓	✓	✓	[19, Prop. 15]
d'_{COR}	✓	✓	✓	✓	✓	✓	✓	[19, Prop. 19]
d_{TV}	✓	✓	✓	✓	✓	✓	✓	Prop. 1
d_K	✓	✓	✓	✓	✓	✓	✓	Prop. 7
d_{L_1}	✓	✓	✓	✓	✓	✓	✓	Prop. 10

TABLE 2. Summary of the properties satisfied by the distorting functions d_{PMM}, d_{LV}, d_{COR}, d'_{COR}, d_{TV}, d_K and d_{L_1}.

Since d_{TV}, d_K and d_{L_1} are distances, they satisfy Ax.1–Ax.3 (and hence also Ax.1a and Ax.1b). So does the distance associated with the constant odds ratio model. On the other hand, neither d_{LV} nor d_{PMM} are symmetrical, meaning that Q may belong to the neighbourhood model of P with radius δ while P does not belong to the neighbourhood model of Q with radius δ; in addition, d_{PMM} does not satisfy Ax.2 either: this means that two different probability measures may be unidentifiable with respect to d_{PMM}, and also that d_{PMM} does not satisfy

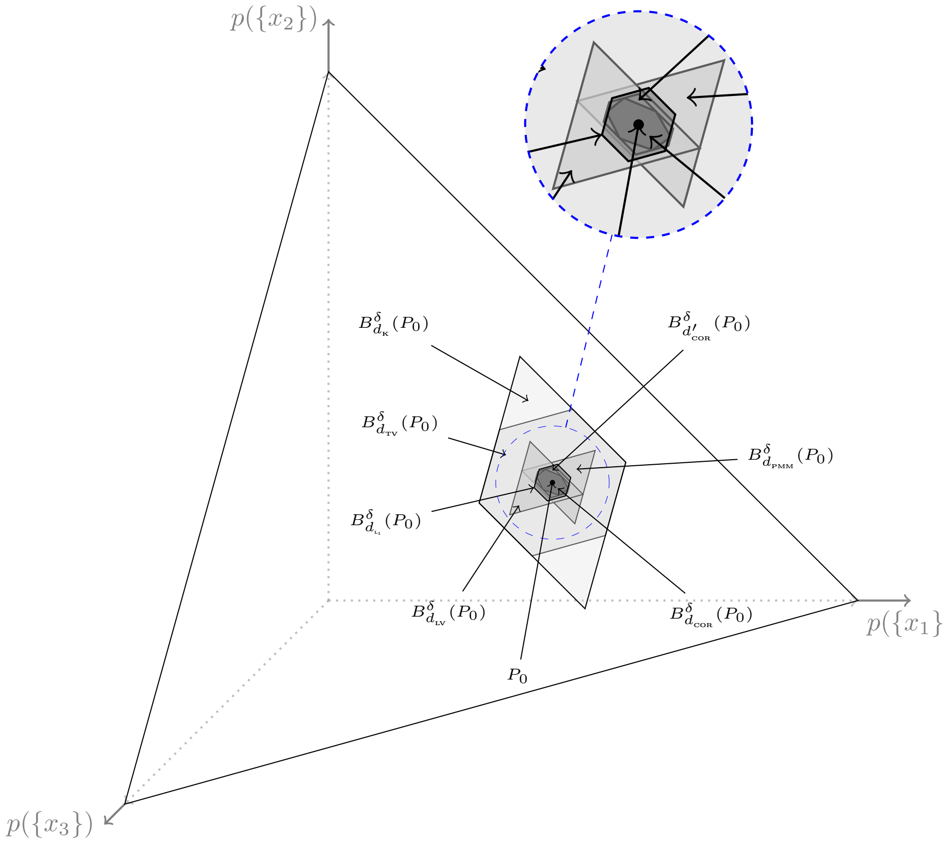

FIGURE 3. Graphical representation of the different credal sets in Example 8.

the triangle inequality in general. It follows that d_{PMM}, d_{LV} are only premetrics instead of distances. Nevertheless, one may argue for instance that the property of symmetry is less natural than other axioms we have considered in these papers, because the roles of the original model and the distorted one are not the same.

Finally, all distorting functions satisfy Ax.4 and Ax.5, meaning that the open balls are convex and continuous. From [19, Prop. 1], this implies that the closed ball coincides with the credal set associated with the coherent lower probability.

5.3. **Properties of the associated coherent lower probability.** We may also compare the different distortion models in terms of the properties of the coherent lower probability they determine. As we recalled in the Introduction, there are a number of particular cases of coherent lower probabilities that may be of interest in practice. The first of them is 2-monotonicity: not only it guarantees that the distortion model has a unique extension to gambles as a lower prevision, but also it allows us to use a simple formula [19, Eq. (2)] to compute the extreme points of the credal set. It turns out that all the models we have considered in this paper are 2-monotone, except for the constant odds ratio, that only satisfies 2-monotonicity once we consider its restriction to events, and the L_1-model, which is neither 2-monotone on gambles nor on events.

Two particular cases of 2-monotone lower probabilities are probability intervals and belief functions. The first correspond to those that are uniquely determined by their restrictions to singletons. Thus, for them there is a simpler representation of the associated credal set. With respect to the examples considered in these papers, only the pari mutuel model and the linear vacuous mixture satisfy this property.

Belief functions are completely monotone lower probabilities. They connect the model with Shafer's Evidential Theory [27], and allow to represent the lower probability by means of a basic probability assignment. In this respect, both the linear vacuous mixture and Kolmogorov's model are belief functions: the former, because it is a convex combination of two completely monotone models, and the latter because every coherent lower probability associated with a p-box is (see [28, Sec.5.1.]). On the other hand, the pari mutuel is not 3-monotone in general ([18, Prop. 5]) and the total variation is not completely monotone (see Example 1). Finally, both the constant odds ration on events and the neighbourhood model based on Kolmogorov's distance induce a completely monotone lower probability that is not a probability interval.

Table 3 summarises the results mentioned in this subsection.

Model	2-monotone	Complete monotone	Probability interval
$\underline{P}_{PMM}$	YES	NO ([18, Prop. 5])	YES ([18, Thm. 1])
$\underline{P}_{LV}$	YES	YES ([19, Sec. 5])	YES ([19, Sec. 5])
$\underline{P}_{COR}$	NO ([19, Ex. 1])	NO	NO ([19, Ex. 2])
$\underline{Q}_{COR}$	YES ([19, Prop. 17])	YES ([19, Prop. 17])	NO ([19, Ex. 2])
$\underline{P}_{TV}$	YES (Prop. 4)	NO (Ex. 1)	NO (Ex. 1)
$\underline{P}_{K}$	YES	YES	NO (Ex. 3)
$\underline{P}_{L_1}$	NO (Ex. 4, Ex. 6)	NO (Ex. 5)	NO (Ex.6)

TABLE 3. Properties satisfied by the coherent lower probabilities induced by the neighbourhood models.

We therefore conclude that, from the point of view of these properties, the most adequate model is the linear vacuous model. The only models that do not satisfy 2-monotonicity are the constant odds ratio (on gambles) and the L_1-model.

5.4. **Complexity.** One important feature of a neighbourhood model is that it has a simple representation in terms of a finite number of extreme points. As we said before, when its lower envelope is 2-monotone there are at most $n!$ different extreme points, that are related to the permutations of the possibility space (see Eq. (2) in [19]). On the other hand, the credal set associated with a coherent lower probability also has at most $n!$ different extreme points, but their representation is not as simple [31]; and a general credal set may have an infinite number of extreme points.

In the case of the pari mutuel and the linear vacuous models, the extreme points were studied in [18] and [30], respectively. In these papers, we have computed the maximum number of extreme points also for the neighbourhood models $B^{\delta}_{d_{TV}}(P_0)$, $B^{\delta}_{d_K}(P_0)$, $B^{\delta}_{d_{COR}}(P_0)$ and $B^{\delta}_{d'_{COR}}(P_0)$. Table 4 summarises the results[2].

[2]In [18, Prop. 2], it is stated that the maximal number of extreme points induced by a PMM is $\frac{n}{2}\binom{n}{\frac{n}{2}}$, when n even, and $\frac{n+1}{2}\binom{n}{\frac{n+1}{2}}$, when n is odd. The expression given in Table 4 is equivalent.

Model	Maximal number of extreme points	Result
$\underline{P}_{PMM}$	$\frac{n!}{\lfloor \frac{n}{2} \rfloor (\lfloor \frac{n}{2} \rfloor -1)!(n-\lfloor \frac{n}{2} \rfloor -1)!}$	[18, Prop. 2]
$\underline{P}_{LV}$	n	[30, Sec.3.6.3(b)]
$\underline{P}_{COR}$	2^n-2	[19, Prop. 13]
$\underline{Q}_{COR}$	$n!$	[19, Prop. 17]
$\underline{P}_{TV}$	$n(n-1)$	Prop. 5
$\underline{P}_{K}$	$\mathcal{P}_n$	Prop. 9

TABLE 4. Maximal number of extreme points in the different neighbourhood models.

We observe that the the simplest model is the linear vacuous, followed by the total variation distance, the constant odds ratio, the Kolmogorov model, the pari mutuel, and, finally, the constant odds ratio restricted to events. We see also that the bound is usually much smaller than the general bound of $n!$ that holds for arbitrary coherent or 2-monotone lower probabilities.

For illustrative purposes, next table gives the maximal number of extreme points for small values of n:

$\lvert\mathcal{X}\rvert$	2	3	4	5	6	7	8	9	10
$\underline{P}_{PMM}$	2	6	12	30	60	140	280	630	1260
$\underline{P}_{LV}$	2	3	4	5	6	7	8	9	10
$\underline{P}_{COR}$	2	4	6	14	30	62	126	254	510
$\underline{Q}_{COR}$	2	6	24	120	720	5040	40320	362880	3628800
$\underline{P}_{TV}$	2	6	12	20	30	42	56	72	90
$\underline{P}_{K}$	2	5	12	29	70	169	408	985	2378

Finally, let us recall that the extreme points in $B^{\delta}_{d_{L_1}}(P_0)$ have not been determined yet. As we have seen in Examples 4 and 6, $\underline{P}_{L_1}$ is neither 2-monotone in gambles nor in events, so the procedure described in [19, Eq. (2)] cannot be applied.

5.5. **Conditioning.** For each of the models, we have studied how to update the model when new information is obtained, and we have checked whether the conditional model obtained by regular extension belongs to the same family of distortion models. Table 5 summarises our results. From this table, it is clear that all the models present a correct behaviour except the Kolmogorov and L_1-models.

	Conditioning	Parameter δ
PMM	YES [19, Prop. 8]	Increases [19, Eq.(16)]
LV	YES [30, Sec. 6.6.2]	Increases [19, Eq.(19)]
COR	YES [30, Sec. 6.6.3]	Does not change [19, Eq.(29)]
TV	YES (Prop. 6)	Increases (Eq. (9))
K	NO [9, Ex. 3]	-
L_1	NO (Ex. 7)	-

TABLE 5. Behaviour of the neighbourhood models when applying the regular extension.

We observe that among the models that are preserved under conditioning (PMM, LV, COR and TV), the constant odds ratio model is the only one that does not increase its imprecision, since its parameter δ remains unchanged after conditioning. In other words, it is the only one that does not produce the phenomenon of *dilation* [26]. In the other three cases, PMM, LV and TV, dilation happens, being the PMM the model where the parameter increases more, followed by the TV model and the LV.

6. Final synthesis and conclusions

In these papers, we have made a unified study 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 the extent to which the estimation of P_0 is reliable. This produces a number of credal sets, that can most often[3] 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.

	Old models				**New models**		
	PMM	LV	COR (gambles)	COR(events)	TV	K	L_1
Imprecision	$+$	$+$	$++$	$++$	$-$	$--$	$+$
Properties of d	$-$	$+$	$++$	$++$	$++$	$++$	$++$
Properties of $\underline{P}$	$+$	$++$	$--$	$+$	$-$	$+$	$--$
# extreme points	$+$	$++$	$+$	$--$	$++$	$+$	Open
Conditioning	$+$	$+$	$++$	N.A.	$+$	$--$	$--$

Table 6. Qualitative assessment of the model properties evaluated in this study (Very Good $(++)$, Good $(+)$, Bad (-), Very Bad $(--)$, Not Applicable (N.A.).

Table 6 tries to synthesise in an evaluation table the various aspects of the models we have studied here. It shows that the linear vacuous model tends to result in a highly tractable and stable model, at least for those aspects we considered here. It is shortly followed by the pari mutuel and odds ratio model, that may explain why these models are the one that have received the most attention. The COR seems to be the best model in terms of the amount of imprecision introduced, in the sense that for a fixed distortion parameter δ, it induces the smallest credal set among the six models. However, note that the credal sets may also be compared in terms of other properties, such as their geometry. Some relevant works in this context are [2, 7]. The total variation model and its restriction to cumulative events also satisfy a number of interesting properties, but can provide quite imprecise models. Finally, our study of the L_1-model tends to suggest that it is quite impractical to handle, with properties that are quite difficult to characterise and sometimes not very intuitive. Nevertheless, as the L_1 distance is a quite common choice, we still

[3]As for some models, we need to consider lower previsions.

think it is interesting to have a better understanding of the neighbourhood model it induces.

Overall, it seems that the old models, those analysed in the first part of this study, have in general better properties than the new ones. Nevertheless, there is not a model that is uniformly better than all the others, and this leads naturally to the question of which model should we use in each situation. In this respect, if our priority is to have a distorted model where we introduce as little imprecision as possible, we should choose the COR. Also, the extension to gambles of this model has the advantage than the imprecision is preserved when conditioning. Secondly, if we look for a model that is easy to handle, we believe that the LV has a quite good behaviour. Thirdly, if we have an experiment whose probabilistic information is given in terms of cumulative events, it seems reasonable to consider a Kolmogorov model. Fourthly, if we are following a behavioural interpretation of the distortion model, the decision should be made in terms of the interpretation of the parameter δ. Remember that δ could be interpreted as *the inflation rate for the selling price for A* (PMM), *the deflation rate for the buying price for A* (LV), *the constant rate on the investments* (COR) or *the fixed tax to be paid for betting, regardless on the event* (TV). The behavioural interpretation of the Kolmogorov model is equivalent to that of the TV-model but restricted to cumulative events, while in the L_1-model δ has a similar interpretation as in the TV-model, but considering the size of the event.

While our study encompassed several aspects, it is not exhaustive and it provides only some guidelines to pick a neighbourhood model. Indeed, there may be other criteria that may help choose a given neighbourhood model. For instance, when two models are incomparable in terms of the inclusion of their respective neighbourhood models, we may compare them by means of imprecision indices (see for instance [1, 5]); or we may consider other properties of the associated lower probability, such as being k-additive or minitive.

One crucial assumption we have done throughout is that the support of the original probability measure coincides with the possibility space, i.e., that no singleton has zero probability, and that the same applies to all probability measures in the neighbourhood. While most of the results in the paper also hold in the more general case where zero probabilities arise, this is not always the case. Moreover, the treatment of the problem of conditioning becomes more involved, the expressions of the distance should be suitably modified to avoid zeros in the denominators, and the number of extreme points of the credal set may also be reduced considering the size of the support. For all these reasons, we have preferred to consider this simplifying assumption in this already long and involved study. Some more general results can be found in Appendix B.

Another assumption we have made in these two papers is that the distorting functions induce polytopes in the space of probability measures, meaning for example the Euclidean distance or the Kullback-Leibler divergence are out of the scope. Of course, our study could be extended to any type of distorting function. For instance, a similar study could be made using divergence measures [23], that include as an example the total variation distance. Out of them, the Kullback-Leibler is one of the most prominent, and was already considered in [12, 22] as a means to build neighbourhoods around a probability or a set of probabilities. A preliminary study of the distortion model associated with the Kullback-Leibler distance leads us to believe that the neighbourhood model will be more involved than the ones we have considered in our analysis, due to (i) the difficulty in giving an explicit formula for its associated lower probability; (ii) that it is not clear whether the lower probability satisfies helpful properties such as 2-monotonicity; and (iii) that from this some problems ensue for determining the conditional models. Nevertheless, a detailed study of this approach would be one of the main future lines of research.

Also, while we have characterised under which cases the distortion model is a probability interval, it could be interesting to give necessary and sufficient conditions, in terms of the distance, for the associated distortion model to be 2 or completely monotone. A preliminary analysis of this problem has not been successful, and we conjecture that such conditions would turn out to be somewhat artificial. The study of this problem is left as a future line of research. It would also be of interest to deepen in the comparison between the models in this paper as well as the study of other neighbourhood models, such as those in [6].

Finally, a natural next step would be to look at the distortion of *imprecise* probability models. This could be done in two manners: distorting each probability measure compatible with the imprecise model, and then taking the lower envelope of the union of the credal sets that result [22]; or to consider directly a distance between imprecise probability models, as in [20, 21].

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 want to thank the reviewers for the careful reading of the paper. Their many valuable suggestions and comments helped us to improve the quality of the manuscript.

Appendix A. Proofs

Proof of Proposition 1. Since d_{TV} is a distance, it immediately satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2 and Ax.3. Ax.4 holds because d_{TV} is the metric induced by the supremum norm, and Ax.5 follows by definition. □

Proof of Proposition 3. If $\delta \leq P_0(A)$ for a fixed event A, we can consider the probability measure P determined by:

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

This function P satisfies the following properties:

(1) $\sum_{x\in\mathcal{X}} P(\{x\}) = 1$:

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

(2) $P(\{x\}) \geq 0$ for every $x \in \mathcal{X}$. If $x \notin A$, $P(\{x\}) \geq P_0(\{x\}) \geq 0$. If $x \in A$,

$$P(\{x\}) \geq 0 \Leftrightarrow P_0(\{x\}) \geq \delta\frac{P_0(\{x\})}{P_0(A)} \Leftrightarrow P_0(A) \geq \delta,$$

which holds by hypothesis. Taking items (1) and (2) into account, P is a probability measure.

(3) $P \in B^{\delta}_{d_{TV}}(P_0)$. Let us see that $d_{TV}(P, P_0) \leq \delta$:

(a) Assume that $B \subseteq A$:

$$|P_0(B) - P(B)| = \left|P_0(B) - \sum_{x\in B}\left(P_0(\{x\}) - \delta\frac{P_0(\{x\})}{P_0(A)}\right)\right|$$
$$= \left|P_0(B) - P_0(B) + \delta\frac{P_0(B)}{P_0(A)}\right| = \delta\frac{P_0(B)}{P_0(A)} \leq \delta.$$

(b) Assume that $B \subseteq A^c$:

$$|P_0(B) - P(B)| = \left|P_0(B) - \sum_{x\in B}\left(P_0(\{x\}) + \delta\frac{P_0(\{x\})}{P_0(A)}\right)\right|$$
$$= \left|P_0(B) - P_0(B) - \delta\frac{P_0(B)}{P_0(A^c)}\right| = \delta\frac{P_0(B)}{P_0(A^c)} \leq \delta.$$

(c) Assume that $B \not\subseteq A$ and $B \not\subseteq A^c$:

$$|P_0(B) - P(B)| = |P_0(A\cap B) + P_0(A^c\cap B) - P(A\cap B) - P(A^c\cap B)|$$
$$= \left|P_0(A\cap B) + P_0(A^c\cap B) - \sum_{x\in A^c\cap B}\left(P_0(\{x\}) + \delta\frac{P_0(\{x\})}{P_0(A^c)}\right)\right.$$
$$\left. - \sum_{x\in A\cap B}\left(P_0(\{x\}) - \delta\frac{P_0(\{x\})}{P_0(A)}\right)\right|$$
$$= \left|\delta\frac{P_0(A\cap B)}{P_0(A)} - \delta\frac{P_0(A^c\cap B)}{P_0(A^c)}\right| \leq \delta.$$

We therefore conclude that $d_{TV}(P, P_0) \leq \delta$, hence $P \in B^{\delta}_{d_{TV}}(P_0)$.

(4) Finally:

$$P(A) = \sum_{x\in A}\left(P_0(\{x\}) - \delta\frac{P_0(\{x\})}{P_0(A)}\right) = P_0(A) - \delta.$$

Therefore, we conclude that:

$$\underline{P}_{TV}(A) = P(A) = \max\{0, P_0(A) - \delta\} = P_0(A) - \delta.$$

Thus, $\delta \geq d_{TV}(P, P_0) \geq |P(A) - P_0(A)| = \delta$. Since moreover by assumption $\delta < \min_x P_0(\{x\}) \leq P_0(A)\ \forall A \neq \emptyset$, we deduce that $\max_{P \in B^{\delta}_{d_{TV}}(P_0)} d_{TV}(P, P_0) = \delta$. $\square$

Proof of Proposition 4. From [8, Ex. 2.1], every convex transformation of a probability measure is 2-monotone. Thus, the proof directly follows just noting that for every $A \subseteq \mathcal{X}$ it holds that $\underline{P}_{TV}(A) = f(P_0(A))$, where $f(t) = \max\{0, t - \delta\}$ $\forall t \in [0, 1)$ is a convex function satisfying $f(1) = 1$. $\square$

Proof of Proposition 5. The 2-monotonicity of $\underline{P}_{TV}$ implies that the extreme points of $B^{\delta}_{d_{TV}}(P_0)$ are in correspondence with the permutations of the possibility space, as mentioned in [19, Eq. (2)]. Let us use this to determine the maximum number of extreme points. Recall that the assumption of positive lower probabilities implies that $\delta < \min_{x \in \mathcal{X}} P_0(\{x\})$. In that case, the extreme point P_σ is determined by:

$$
\begin{aligned}
&P_\sigma(\{x_{\sigma(1)}\}) = \underline{P}_{TV}(\{x_{\sigma(1)}\}) = P_0(\{x_{\sigma(1)}\}) - \delta. \\
&P_\sigma(\{x_{\sigma(1)}, x_{\sigma(2)}\}) = \underline{P}_{TV}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) = P_0(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \delta. \\
&\ldots \\
&P_\sigma(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) = \underline{P}_{TV}(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) = P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) - \delta. \\
&P_\sigma(\{x_{\sigma(1)}, \ldots, x_{\sigma(n)}\}) = 1.
\end{aligned}
$$

We deduce that:

$$
\begin{aligned}
&P_\sigma(\{x_{\sigma(1)}\}) = P_0(\{x_{\sigma(1)}\}) - \delta. \\
&P_\sigma(\{x_{\sigma(2)}\}) = P_0(\{x_{\sigma(2)}\}). \\
&\ldots \\
&P_\sigma(\{x_{\sigma(n-1)}\}) = P_0(\{x_{\sigma(n-1)}\}). \\
&P_\sigma(\{x_{\sigma(n)}\}) = 1 - \big(P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(n-1)}\}) - \delta\big) = P_0(\{x_{\sigma(n}\}) + \delta.
\end{aligned}
$$

As a consequence, the extreme points are determined by those elements in the first and last positions, whence there are $n(n-1)$ different extreme points. $\square$

Proof of Proposition 6. We need to prove that

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

Since $\underline{P}$ is a 2-monotone probability by Proposition 4, it follows from [29, Thm. 7.2] that:

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

Since by assumption $P_0(A \cap B) - \delta > 0$, the equality becomes

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

and since $P_0(A^c \cap B) + P_0(A \cap B) < 1$ (as $P_0(\{x\}) > 0$ for all $x \in \mathcal{X}$), we also have $P_0(A^c \cap B) + \delta < 1$, as $P_0(A \cap B)$ is an upper bound of δ. The equality then becomes

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

which finishes our proof. □

Proof of Proposition 7. Since d_K is a distance, it immediately satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2 and Ax.3. To see that it satisfies Ax.4, note that, given $Q = \alpha Q_1 + (1-\alpha)Q_2$,

$$\begin{aligned}|F_P(x) - F_Q(x)| &= |\alpha(F_P(x) - F_{Q_1}(x)) + (1-\alpha)(F_P(x) - F_{Q_2}(x))| \\ &\leq \alpha|F_P(x) - F_{Q_1}(x)| + (1-\alpha)|F_P(x) - F_{Q_2}(x)|,\end{aligned}$$

whence Ax.4 follows.

Finally, since $d_K(P,Q) \leq d_{TV}(P,Q)$ and the latter satisfies Ax.5, we deduce that so does d_K. □

Proof of Proposition 8. By assumption, $\underline{P}_K(\{x_1\}) = F_{P_0}(x_1) - \delta > 0$. Define the function F by:

$$F(x) = \begin{cases} F_{P_0}(x_1) - \delta & \text{if } x = x_1. \\ F_{P_0}(x) & \text{otherwise.} \end{cases}$$

F is a cumulative distribution function, and its associated probability P belongs to $B^{\delta}_{d_K}(P_0)$:

- If $x = x_1$, $|F(x) - F_{P_0}(x)| = |F_{P_0}(x_1) - \delta - F_{P_0}(x_1)| = \delta$.
- If $x \neq x_1$, $|F(x) - F_{P_0}(x)| = |F_{P_0}(x) - F_{P_0}(x)| = 0$.

Thus, $P \in B^{\delta}_{d_K}(P_0)$, and also $d_K(P, P_0) = \delta$. □

Proof of Proposition 9. Let us show that there exist P_0 and δ such that the p-box they induce using Equation (14) has exactly $\mathcal{P}_n$ different extreme points.

For this aim, let P_0 be the uniform distribution on the n-element space $\mathcal{X}$, and take $\delta = \frac{3}{4n}$ (indeed, the proof still holds for $\delta \in \left(\frac{1}{2n}, \frac{1}{n}\right)$). These give rise to the following p-box:

$$\underline{F}(x_i) = \frac{i}{n} - \delta = \frac{4i-3}{4n}, \quad \overline{F}(x_i) = \frac{i}{n} + \delta = \frac{4i+3}{4n}, \quad \forall i = 1, \ldots, n-1$$

and $\underline{F}(x_n) = \overline{F}(x_n) = 1$. This means that:

$$\begin{aligned}\underline{F}(x_1) < \underline{F}(x_2) < \overline{F}(x_1) < \underline{F}(x_3) < \ldots < \underline{F}(x_{i+1}) < \overline{F}(x_i) < \underline{F}(x_{i+2}) \\ < \ldots < \underline{F}(x_{n-1}) < \overline{F}(x_{n-2}) < \overline{F}(x_{n-1}) < \underline{F}(x_n) = \overline{F}(x_n) = 1.\end{aligned}$$

A graphical representation of this p-box for the case $n = 5$ can be seen in Figure 4.

In order to determine the focal events associated with this p-box, we apply the results in [10, Sec. 3.3]. There, it is explained that if the values taken by $\underline{F}$ and $\overline{F}$ are denoted by:

$$0 = \gamma_0 < \gamma_1 < \ldots < \gamma_M, \tag{20}$$

there are M focal events, given by:

$$E_j = \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge (1 - g_{\underline{F}}(x_i) < \gamma_j)\},$$

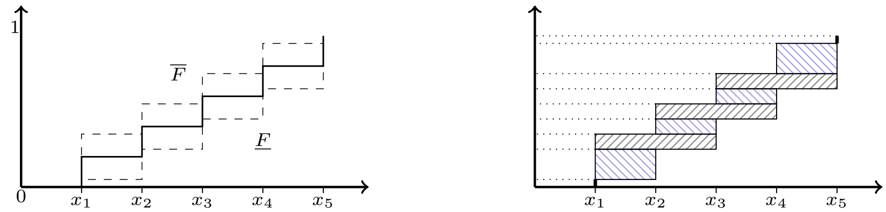

FIGURE 4. P-box (left) and focal sets of its associated belief function (right), given by $E_1 = \{x_1\}$, $E_2 = \{x_1, x_2\}$, $E_3 = \{x_1, x_2, x_3\}$, $E_4 = \{x_2, x_3\}$, $E_5 = \{x_2, x_3, x_4\}$, $E_6 = \{x_3, x_4\}$, $E_7 = \{x_3, x_4, x_5\}$, $E_8 = \{x_4, x_5\}$ and $E_9 = \{x_5\}$.

where $g_{\underline{F}}(x_i) = 1 - \max\{\underline{F}(x_j) \mid \underline{F}(x_j) < \underline{F}(x_i),\ j = 0, 1, \ldots, i\}$. In our case, where $\underline{F}(x_i) < \underline{F}(x_{i+1})$ for every $i = 1, \ldots, n-1$, $g_{\underline{F}}(x_i) = 1 - \underline{F}(x_{i-1})$ (here we assume that $\underline{F}(x_0) = 0$). Thus, from Equation (20) there are $2n - 1$ different focal events. It holds that:

- For $j = 1$, $\gamma_1 = \underline{F}(x_1)$ and:
$$\begin{aligned} E_1 &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_1 \wedge \underline{F}(x_{i-1}) < \gamma_1\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \underline{F}(x_1) \wedge \underline{F}(x_{i-1}) < \underline{F}(x_1)\} = \{x_1\}. \end{aligned}$$
- For $j = 2$, $\gamma_2 = \underline{F}(x_2)$ and:
$$\begin{aligned} E_2 &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_2 \wedge \underline{F}(x_{i-1}) < \gamma_2\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \underline{F}(x_2) \wedge \underline{F}(x_{i-1}) < \underline{F}(x_2)\} = \{x_1, x_2\}. \end{aligned}$$
- For $j = 2k + 1$ ($k = 1, \ldots, n-3$), $\gamma_j = \overline{F}(x_k)$ and:
$$\begin{aligned} E_j &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge \underline{F}(x_{i-1}) < \gamma_j\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \overline{F}(x_k) \wedge \underline{F}(x_{i-1}) < \overline{F}(x_k)\} = \{x_k, x_{k+1}, x_{k+2}\}. \end{aligned}$$
- For $j = 2k$ ($k = 2, \ldots, n-2$), $\gamma_j = \underline{F}(x_{k+1})$ and:
$$\begin{aligned} E_j &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge \underline{F}(x_{i-1}) < \gamma_j\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \underline{F}(x_{k+1}) \wedge \underline{F}(x_{i-1}) < \underline{F}(x_{k+1})\} = \{x_k, x_{k+1}\}. \end{aligned}$$
- For $j = 2n - 3$, $\gamma_j = \overline{F}(x_{n-2})$ and:
$$\begin{aligned} E_j &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge \underline{F}(x_{i-1}) < \gamma_j\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \overline{F}(x_{n-2}) \wedge \underline{F}(x_{i-1}) < \overline{F}(x_{n-2})\} = \{x_{n-2}, x_{n-1}, x_n\}. \end{aligned}$$
- For $j = 2n - 2$, $\gamma_j = \overline{F}(x_{n-1})$ and:
$$\begin{aligned} E_j &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge \underline{F}(x_{i-1}) < \gamma_j\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \overline{F}(x_{n-1}) \wedge \underline{F}(x_{i-1}) < \overline{F}(x_{n-1})\} = \{x_{n-1}, x_n\}. \end{aligned}$$
- For $j = 2n - 1$, $\gamma_j = \overline{F}(x_n) = \underline{F}(x_n) = 1$ and:
$$\begin{aligned} E_j &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq \gamma_j \wedge \underline{F}(x_{i-1}) < \gamma_j\} \\ &= \{x_i \in \mathcal{X} \mid \overline{F}(x_i) \geq 1 \wedge \underline{F}(x_{i-1}) < 1\} = \{x_n\}. \end{aligned}$$

This means that the focal sets of the p-box $(\underline{F}, \overline{F})$ are:

$$\begin{aligned}
&E_1 = \{x_1\}, \\
&E_{2k} = \{x_k, x_{k+1}\}, \text{ for } k = 1, \dots, n-1, \\
&E_{2k+1} = \{x_k, x_{k+1}, x_{k+1}\}, \text{ for } k = 1, \dots, n-2, \\
&E_{2n-1} = \{x_n\}.
\end{aligned}$$

Hence $(\underline{F}, \overline{F})$ is a p-box of the *Pell family*, and according to the results in [17], its number of extreme points is exactly $\mathcal{P}_n$ (see [17, Prop. 16]). □

Proof of Proposition 10. Since d_{L_1} is a distance, it immediately satisfies Ax.1 (hence also Ax.1a and Ax.1b), Ax.2 and Ax.3. To see that it satisfies Ax.4, note that given P, Q_1, Q_2 and $\alpha \in (0,1)$,

$$|(\alpha Q_1 + (1-\alpha)Q_2)(B) - P(B)| \leq \alpha |Q_1(B) - P(B)| + (1-\alpha)|Q_2(B) - P(B)|,$$

whence

$$\begin{aligned}
d_{L_1}(\alpha Q_1 + (1-\alpha)Q_2, P) &\leq \alpha d_{L_1}(Q_1, P) + (1-\alpha) d_{L_1}(Q_2, P) \\
&\leq \max\{d_{L_1}(Q_1, P), d_{L_1}(Q_2, P)\},
\end{aligned}$$

and as a consequence Ax.4 holds. Finally, since d_{L_1} is the metric induced by the L_1-norm, it also satisfies Ax.5. □

Proof of Theorem 11. Let us prove first of all that $\underline{P}_{L_1}(A) \leq P_0(A) - \frac{\delta}{\varphi(n,|A|)}$. For simplicity throughout this proof, we will use $\varphi(|A|)$ to denote $\varphi(n, |A|)$ and assume that $|A| = k$. Consider P given by:

$$P(\{x_j\}) = \begin{cases} P_0(\{x_j\}) - \frac{\delta}{|A| \cdot \varphi(|A|)}, & \text{if } x_j \in A. \\ P_0(\{x_j\}) + \frac{\delta}{|A^c| \cdot \varphi(|A|)}, & \text{if } x_j \notin A. \end{cases} \tag{21}$$

It satisfies the following properties:

(1) P is non-negative. On the one hand, if $x_j \notin A$, $P(\{x_j\}) \geq P_0(\{x_j\}) > 0$. On the other hand, for $x_j \in A$ it holds that:

$$P(\{x_j\}) = P_0(\{x_j\}) - \frac{\delta}{|A| \cdot \varphi(|A|)} = \underline{P}_{L_1}(\{x\}) > 0,$$

where the inequality follows by assumption.

(2) $\sum_{j=1}^n P(\{x_j\}) = 1$ (and therefore P is a probability measure):

$$\begin{aligned}
\sum_{j=1}^n P(\{x_j\}) &= \sum_{x_j \in A} P(\{x_j\}) + \sum_{x_i \notin A} P(\{x_i\}) \\
&= \sum_{x_j \in A} \left(P_0(\{x_j\}) - \frac{\delta}{|A| \cdot \varphi(|A|)} \right) + \sum_{x_j \notin A} \left(P_0(\{x_j\}) + \frac{\delta}{|A^c| \cdot \varphi(|A|)} \right) \\
&= 1 + \frac{\delta}{\varphi(|A|)} - \frac{\delta}{\varphi(|A|)} = 1.
\end{aligned}$$

(3) $P(A) = P_0(A) - \frac{\delta}{\varphi(|A|)}$:

$$P(A) = \sum_{x_j \in A} \left(P_0(\{x_j\}) - \frac{\delta}{|A| \cdot \varphi(|A|)} \right) = P_0(A) - \frac{|A|\delta}{|A| \cdot \varphi(|A|)} = P_0(A) - \frac{\delta}{\varphi(|A|)}.$$

(4) For any $B \subseteq \mathcal{X}$:

$$
\begin{aligned}
|P(B) - P_0(B)| &= |P(A \cap B) + P(A^c \cap B) - P_0(A \cap B) - P_0(A^c \cap B)| \\
&= \left| \sum_{x_j \in A \cap B} \frac{\delta}{|A| \cdot \varphi(|A|)} - \sum_{x_i \in A^c \cap B} \frac{\delta}{|A^c| \cdot \varphi(|A|)} \right| \\
&= \frac{\delta}{\varphi(|A|)} \left| \frac{|A \cap B|}{|A|} - \frac{|A^c \cap B|}{|A^c|} \right|.
\end{aligned}
$$

(5) $P \in \mathcal{M}(\underline{P}_{L_1})$, i.e., $d(P, P_0) \leq \delta$:

$$
d(P, P_0) = \sum_{B \subseteq \mathcal{X}} |P(B) - P_0(B)| = \sum_{B \subseteq \mathcal{X}} \frac{\delta}{\varphi(|A|)} \left| \frac{|A \cap B|}{|A|} - \frac{|A^c \cap B|}{|A^c|} \right|.
$$

If l is the cardinality of $|A \cap B|$ and j that of $|A^c \cap B|$, the above expression can be rewritten as:

$$
\begin{aligned}
d(P, P_0) &= \frac{\delta}{\varphi(|A|)} \sum_{l=0}^{k} \binom{k}{l} \sum_{j=0}^{n-k} \binom{n-k}{j} \left| \frac{l}{k} - \frac{j}{n-k} \right| \\
&= \frac{\delta}{\varphi(|A|)} \sum_{l=0}^{k} \binom{k}{l} \sum_{j=0}^{n-k} \binom{n-k}{j} \left| \frac{l(n-k) - jk}{k(n-k)} \right|,
\end{aligned}
$$

where $\binom{k}{l}$ and $\binom{n-k}{j}$ denote the number of events of B that have l elements from A and j from A^c. Finally, applying Equation (16):

$$
d(P, P_0) = \frac{\delta}{\varphi(|A|)} \varphi(|A|) = \delta.
$$

Thus, $\underline{P}_{L_1}(A) \leq P_0(A) - \frac{\delta}{|A| \cdot \varphi(|A|)}$.

In order to prove that the inequality is indeed an equality, let us prove an equivalent expression for φ. Note that:

$$
\varphi(|A|) = \sum_{B \subseteq \mathcal{X}} \left| \frac{|A^c \cap B|}{|A^c|} - \frac{|A \cap B|}{|A|} \right|.
$$

If we separate the events B in terms of their cardinality m:

$$
\varphi(|A|) = \sum_{m=0}^{n} \sum_{\substack{B \subseteq \mathcal{X} \\ |B| = m}} \left| \frac{|A^c \cap B|}{|A^c|} - \frac{|A \cap B|}{|A|} \right|.
$$

For every $B \subseteq \mathcal{X}$, let l be the cardinality of $A \cap B$:

$$
\begin{aligned}
\varphi(|A|) &= \sum_{m=0}^{n} \sum_{l=0}^{\min\{m,k\}} \sum_{\substack{B \subseteq \mathcal{X} \\ |B| = m, |A \cap B| = l}} \left| \frac{|A^c \cap B|}{|A^c|} - \frac{|A \cap B|}{|A|} \right| \\
&= \sum_{m=0}^{n} \sum_{l=0}^{\min\{m,k\}} \sum_{\substack{B \subseteq \mathcal{X} \\ |B| = m, |A \cap B| = l}} \left| \frac{m-l}{n-k} - \frac{l}{k} \right|.
\end{aligned}
$$

Now, $\frac{m-l}{n-k} - \frac{l}{k}$ is decreasing in l; moreover, for $l = 0$ it gives $\frac{m}{n-k} > 0$ while for $l = m$ it is $-\frac{l}{k} < 0$ and for $l = k$ it is $\frac{m-k}{n-k} - 1 \leq 0$. Thus, given m fixed, we denote m^* the maximum l such that $\frac{m-l}{n-k} - \frac{l}{k} \geq 0$. This allows us to rewrite $\varphi(|A|)$ as:

$$\sum_{m=0}^{n} \left(\sum_{l=0}^{m^*} \sum_{\substack{B \subseteq \mathcal{X} \\ |B|=m, |A\cap B|=l}} \left(\frac{m-l}{n-k} - \frac{l}{k} \right) + \sum_{l=m^*+1}^{\min\{m,k\}} \sum_{\substack{B \subseteq \mathcal{X} \\ |B|=m, |A\cap B|=l}} \left(\frac{l}{k} - \frac{m-l}{n-k} \right) \right).$$

Finally, given $|B| = m$ and $|A \cap B| = l$, there are $\binom{k}{l} \cdot \binom{n-k}{m-l}$ different events B whose value $\left| \frac{m-l}{n-k} - \frac{l}{k} \right|$ coincides. This gives:

$$\varphi(|A|) = \sum_{m=0}^{n} \left(\sum_{l=0}^{m^*} \binom{k}{l} \cdot \binom{n-k}{m-l} \left(\frac{m-l}{n-k} - \frac{l}{k} \right) + \sum_{l=m^*+1}^{\min\{m,k\}} \binom{k}{l} \cdot \binom{n-k}{m-l} \left(\frac{l}{k} - \frac{m-l}{n-k} \right) \right).$$

Let us prove next that $\underline{P}_{L_1}(A) = P_0(A) - \frac{\delta}{|A|\cdot\varphi(|A|)}$. Let Q be given by:

$$Q(\{x_j\}) = \begin{cases} P_0(\{x_j\}) - \frac{\delta}{|A|\cdot\varphi(|A|)} - \varepsilon_j, & \text{if } x_j \in A. \\ P_0(\{x_j\}) + \frac{\delta}{|A^c|\cdot\varphi(|A|)} + \varepsilon_j, & \text{if } x_j \notin A, \end{cases}$$

where:

- $\sum_{x_j \in A} \varepsilon_j = \varepsilon > 0$.
- $\sum_{x_i \notin A} \varepsilon_i = \varepsilon > 0$.
- The values ε_j are such that Q is a probability measure.
- We impose no restriction on the sign of ε_j.

We can also express Q in terms of P:

$$Q(\{x_j\}) = \begin{cases} P(\{x_j\}) - \varepsilon_j, & \text{if } x_j \in A. \\ P(\{x_j\}) + \varepsilon_j, & \text{if } x_j \notin A. \end{cases}$$

Let us study the properties of Q:

(1) $Q(A) < P_0(A) - \frac{\delta}{\varphi(|A|)}$:

$$\begin{aligned} Q(A) &= \sum_{x_j \in A} Q(\{x_j\}) = \sum_{x_j \in A} (P(\{x_j\}) - \varepsilon_j) \\ &= P(A) + \sum_{x_j \in A} \varepsilon_j = P_0(A) - \frac{\delta}{\varphi(|A|)} - \varepsilon < P_0(A) - \frac{\delta}{\varphi(|A|)}. \end{aligned}$$

(2) Let us compute $d(Q, P_0)$:

$$
\begin{aligned}
d(Q,P_0) &= \sum_{B\subseteq\mathcal{X}} |Q(B)-P_0(B)| \\
&= \sum_{B\subseteq\mathcal{X}} |Q(A\cap B)+Q(A^c\cap B)-P_0(A\cap B)-P_0(A^c\cap B)| \\
&= \sum_{B\subseteq\mathcal{X}} \left| \frac{\delta}{\varphi(A)}\left(\frac{|A^c\cap B|}{n-k}-\frac{|A\cap B|}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right| \\
&= \sum_{m=0}^{n} \sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}} \left| \frac{\delta}{\varphi(A)}\left(\frac{|A^c\cap B|}{n-k}-\frac{|A\cap B|}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right| \\
&= \sum_{m=0}^{n} \sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}} \sum_{l=0}^{\min\{m,k\}} \left| \frac{\delta}{\varphi(A)}\left(\frac{m-l}{n-k}-\frac{l}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right| \\
&= \sum_{m=0}^{n} \sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}} \left(\sum_{l=0}^{m^*} \left| \frac{\delta}{\varphi(A)}\left(\frac{m-l}{n-k}-\frac{l}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right| \right. \\
&\qquad \left. + \sum_{l=m^*+1}^{\min\{m,k\}} \left| \frac{\delta}{\varphi(A)}\left(\frac{m-l}{n-k}-\frac{l}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right| \right) \\
&\geq \sum_{m=0}^{n} \sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}} \left(\sum_{l=0}^{m^*} \left(\frac{\delta}{\varphi(A)}\left(\frac{m-l}{n-k}-\frac{l}{k}\right)+\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j \right) \right. \\
&\qquad \left. + \sum_{l=m^*+1}^{\min\{m,k\}} \left(\frac{\delta}{\varphi(A)}\left(\frac{l}{k}-\frac{m-l}{n-k}\right)-\sum_{x_i\in A^c\cap B}\varepsilon_i+\sum_{x_j\in A\cap B}\varepsilon_j \right) \right) \\
&= \sum_{m=0}^{n} \sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}} \sum_{l=0}^{\min\{m,k\}} \frac{\delta}{\varphi(A)} \left| \frac{m-l}{n-k}-\frac{l}{k} \right|
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{m=0}^{n}\sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}}\left(\sum_{l=0}^{m^*}\left(\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j\right)\right.\\
&+\sum_{m=0}^{n}\sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}}\left.\left(\sum_{l=m^*+1}^{\min\{m,k\}}\left(-\sum_{x_i\in A^c\cap B}\varepsilon_i+\sum_{x_j\in A\cap B}\varepsilon_j\right)\right)\right)\\
=&\frac{\delta}{\varphi(|A|)}\varphi(|A|)+\sum_{m=0}^{n}\sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}}\left(\sum_{l=0}^{m^*}\left(\sum_{x_i\in A^c\cap B}\varepsilon_i-\sum_{x_j\in A\cap B}\varepsilon_j\right)\right.\\
&+\sum_{m=0}^{n}\sum_{\substack{B\subseteq\mathcal{X}\\|B|=m}}\left.\left(\sum_{l=m^*+1}^{\min\{m,k\}}\left(-\sum_{x_i\in A^c\cap B}\varepsilon_i+\sum_{x_j\in A\cap B}\varepsilon_j\right)\right)\right)
\end{aligned}
$$

In order to simplify this expression, it suffices to check the number of times we add each ε_j and its sign. For a fixed event B with $|B|=m$ and $|A\cap B|=l$, it holds that :

- if $x_j\in A\cap B$, ε_j appears $\binom{k-1}{l-1}\binom{n-k}{m-l}$ times (once fixed $x_j\in A\cap B$, we must select $l-1$ elements among the remaining $k-1$ in $A\setminus\{x_j\}$, and $m-l$ elements among the $n-k$ in A^c);
- if $x_i\in A^c\cap B$, ε_i appears $\binom{k}{l}\binom{n-k-1}{m-l-1}$ times (once fixed $x_i\in A^c\cap B$, we must select l elements from the k elements in A, and $m-l-1$ elements from the remaining $n-k-1$ in $A^c\setminus\{x_i\}$).

This allows to simplify the expression above into:

$$
\begin{aligned}
d(Q,P_0)\geq{}&\delta+\sum_{m=0}^{n}\left(\sum_{l=0}^{m^*}\left(\binom{k}{l}\binom{n-k-1}{m-l-1}\sum_{x_i\in A^c}\varepsilon_i-\binom{k-1}{l-1}\binom{n-k}{m-l}\sum_{x_j\in A}\varepsilon_j\right)\right.\\
&+\left.\sum_{l=m^*+1}^{\min\{m,k\}}\left(-\binom{k}{l}\binom{n-k-1}{m-l-1}\sum_{x_i\in A^c}\varepsilon_i+\binom{k-1}{l-1}\binom{n-k}{m-l}\sum_{x_j\in A}\varepsilon_j\right)\right)\\
={}&\delta+\sum_{m=0}^{n}\left(\sum_{l=0}^{m^*}\left(\binom{k}{l}\binom{n-k-1}{m-l-1}\varepsilon-\binom{k-1}{l-1}\binom{n-k}{m-l}\varepsilon\right)\right.\\
&+\left.\sum_{l=m^*+1}^{\min\{m,k\}}\left(-\binom{k}{l}\binom{n-k-1}{m-l-1}\varepsilon+\binom{k-1}{l-1}\binom{n-k}{m-l}\varepsilon\right)\right)\\
={}&\delta+\varepsilon\sum_{m=0}^{n}\left(\sum_{l=0}^{m^*}\left(\binom{k}{l}\binom{n-k-1}{m-l-1}-\binom{k-1}{l-1}\binom{n-k}{m-l}\right)\right.
\end{aligned}
$$

$$+\sum_{l=m^*+1}^{\min\{m,k\}}\left(-\binom{k}{l}\binom{n-k-1}{m-l-1}+\binom{k-1}{l-1}\binom{n-k}{m-l}\right)\Bigg)$$
$$=\delta+\varepsilon\sum_{m=0}^{n}\left(\sum_{l=0}^{m^*}\binom{k}{l}\binom{n-k}{m-l}\left(\frac{m-l}{n-k}-\frac{l}{k}\right)+\sum_{l=m^*+1}^{\min\{m,k\}}\binom{k}{l}\binom{n-k}{m-l}\left(\frac{l}{k}-\frac{m-l}{n-k}\right)\right)$$
$$=\delta+\varphi(|A|)\varepsilon.$$

As a consequence, for any such Q it can only be $d(Q,P_0)\leq\delta$ if $\varepsilon=0$, or, in other words, if $Q=P$. We conclude that $\underline{P}_{L_1}(A)=P_0(A)-\frac{\delta}{|A|\cdot\varphi(|A|)}$. □

Proof of Proposition 12. Consider a fixed event $A\neq\emptyset,\mathcal{X}$, and the probability measures given in Equation (21) in the proof of Theorem 11 by:

$$P(\{x_j\})=\begin{cases}P_0(\{x_j\})-\frac{\delta}{|A|\cdot\varphi(|A|)}, & \text{if } x_j\in A.\\ P_0(\{x_j\})+\frac{\delta}{|A^c|\cdot\varphi(|A|)}, & \text{if } x_j\notin A.\end{cases}$$

It follows from the proof of Theorem 11 that $d_{L_1}(P,P_0)=\delta$. □

Proof of Proposition 13. The proof for the complete monotonicity of $\underline{P}_{L_1}$ for $n\leq 4$ follows straightforwardly by verifying Equation (18) for the different cardinalities for $\cup_{i=1}^{p}A_i$ and $\cap_{I\subseteq\{1,\ldots,p\}}A_i$.

Analogously, it can be seen that $\underline{P}_{L_1}$ is 2-monotone in events if and only if Equation (19) is satisfied for every $1\leq k_1\leq k_2\leq k_3\leq k_4\leq n$ such that $k_1+k_4=k_2+k_3$. It can be easily verified that this is the case for all the possible combinations of k_1,k_2,k_3,k_4 and $n\leq 11$. □

Proof of Proposition 14. The inequality $\underline{P}_{TV}(A)\leq\underline{P}_{LV}(A)$ is equivalent to $P_0(A)-\delta\leq(1-\delta)P_0(A)$, and this is equivalent to $P_0(A)\leq 1$. On the other hand, $\underline{P}_{TV}(A)\leq\underline{P}_{PMM}(A)$ if and only if $P_0(A)-\delta\leq(1+\delta)P_0(A)-\delta$, which is equivalent to $\delta P_0(A)\geq 0$. □

Appendix B. Additional results without the assumption of positive lower probabilities

Throughout our study, we have assumed that the original probability measure P_0 satisfies $P_0(\{x\})>0$ for every $x\in\mathcal{X}$, and also that the distortion parameter $\delta>0$ is small enough so that the ball $B_d^{\delta}(P_0)$ is included in the interior of the set of probability measures over $\mathcal{X}$. This assumption is crucial for many of the results we have established in this paper. In this appendix, we establish some results for arbitrary δ.

B.1. **The total variation model.** We begin by considering the total variation model for arbitrary $\delta>0$. The expression of the lower envelope $\underline{P}_{TV}$ of the neighbourhood $B_{d_{TV}}^{\delta}(P_0)$ can be found in Theorem 2. From its proof, we see that if $\delta\leq P_0(A)$ for a fixed event A, the probability measure P determined by

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

satisfies $P \in B^{\delta}_{d_{TV}}(P_0)$ and

$$\underline{P}_{TV}(A) = P(A) = \max\{0, P_0(A) - \delta\} = P_0(A) - \delta.$$

Thus,

$$\delta \geq d_{TV}(P, P_0) \geq |P(A) - P_0(A)| = \delta.$$

On the other hand, if $\delta > P_0(A)$ for every $A \subset \mathcal{X}$, we obtain that $\underline{P}_{TV}$ is vacuous. As a consequence,

$$\max_{P \in B^{\delta}_{d_{TV}}(P_0)} d_{TV}(P, P_0) = \delta \Leftrightarrow \delta \leq \max_{A \subset \mathcal{X}} P_0(A).$$

On the other hand, an analogous proof to that of Proposition 4 shows that $\underline{P}_{TV}$ is a 2-monotone lower probability for an arbitrary δ. This is instrumental in our next result, where we establish the number of extreme points induced by the total variation model. Denote by

$$\mathcal{L} := \{A \subseteq \mathcal{X} \mid \underline{P}_{TV}(A) = 0\}, \tag{22}$$

the set of events with null lower probability, and we define for every $A \in \mathcal{L}$ the number s_A as

$$s_A = \big(n - |A^{\uparrow}|\big)\big(n - |A| - 1\big), \text{ where } A^{\uparrow} = \bigcup_{B \supseteq A, B \in \mathcal{L}} B.$$

Using this notation, we give the exact number of extreme points of $B^{\delta}_{d_{TV}}(P_0)$.

Proposition 15. *Let $B^{\delta}_{d_{TV}}(P_0)$ be the neighbourhood model associated with a probability measure P_0 and a distortion factor $\delta > 0$ by means of the total variation distance. Then the number of extreme of extreme points of $B^{\delta}_{d_{TV}}(P_0)$ is $\sum_{A \in \mathcal{L}} s_A$. Therefore, if $|\mathcal{X}| = n$, the maximal number of extreme points of $B^{\delta}_{d_{TV}}(P_0)$ is*

$$\frac{n!}{\big(\lfloor \frac{n}{2} \rfloor - 1\big)! \cdot \big(n - \lfloor \frac{n}{2} \rfloor - 1\big)!},$$

where $\lfloor \frac{n}{2} \rfloor$ denotes the largest natural number that is smaller than or equal to $\frac{n}{2}$.

Proof. The 2-monotonicity of $\underline{P}_{TV}$ implies that the extreme points of $B^{\delta}_{d_{TV}}(P_0)$ are in correspondence with the permutations of the possibility space, as mentioned in [19, Eq. (2)]. Let us use this to determine the maximum number of extreme points. The case $\delta < \min_{x \in \mathcal{X}} P_0(\{x\})$ has already been considered in the proof of Proposition 5. On the other hand, when $\delta \geq \min_{x \in \mathcal{X}} P_0(\{x\})$, it holds that:

$$P_{\sigma}(\{x_{\sigma(1)}\}) = \max\{0, P_0(\{x_{\sigma(1)}\}) - \delta\}.$$
$$P_{\sigma}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) = \max\{0, P_0(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \delta\}.$$
$$\ldots$$
$$P_{\sigma}(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) = \max\{0, P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}) - \delta\}.$$
$$P_{\sigma}(\{x_{\sigma(1)}, \ldots, x_{\sigma(n)}\}) = 1.$$

We deduce that:

$$
\begin{aligned}
&P_\sigma(\{x_{\sigma(1)}\}) = \max\{0, P_0(\{x_{\sigma(1)}\} - \delta\}.\\
&P_\sigma(\{x_{\sigma(2)}\}) = \max\{0, P_0(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \delta\} - \max\{0, P_0(\{x_{\sigma(1)}\}) - \delta\}\\
&\quad = \begin{cases} 0 & \text{if } \underline{P}_{TV}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) = 0 \\ P_0(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \delta - \max\{0, P_0(\{x_{\sigma(1)}\}) - \delta\} & \text{if } \underline{P}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) \neq 0 \end{cases}\\
&\quad = \begin{cases} 0 & \text{if } \underline{P}_{TV}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) = 0 \\ \min\big\{P_0(\{x_{\sigma(1)}, x_{\sigma(2)}\} - \delta, P_0(\{x_{\sigma(2)}\})\big\}) & \text{if } \underline{P}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) \neq 0. \end{cases}
\end{aligned}
$$

If we denote by k_σ the number in $\{0, 1, \ldots, n\}$ satisfying:

$$P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k_\sigma)}\}) \leq \delta < P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k_\sigma)}, x_{\sigma(k_\sigma+1)}\}),$$

then

$$P_\sigma(\{x_{\sigma(i)}\}) = \begin{cases} 0 & \text{if } i = 1, \ldots, k_\sigma. \\ P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k_\sigma+1)}\}) - \delta & \text{if } i = k_\sigma + 1. \\ P_0(\{x_{\sigma(i)}\}) & \text{if } i = k_\sigma + 2, \ldots, n-1. \\ P_0(\{x_{\sigma(n)}\}) + \delta & \text{if } i = n. \end{cases} \tag{23}$$

Let us now define a partition of S_n, the set of all permutations of $\{1, \ldots, n\}$. For each $A \in \mathcal{L} = \{A \subseteq \mathcal{X} | \underline{P}_{TV}(A) = 0\}$, we define:

$$S_{n,A} = \big\{\sigma \in S_n \mid \{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\} = A, \text{ and } \{x_{\sigma(1)}, \ldots, x_{\sigma(k+1)}\} \notin \mathcal{L}\big\}.$$

It is immediate that $\{S_{n,A}\}_{A \in \mathcal{L}}$ is a partition of S_n. Let us prove that the number of different extreme points induced by $\sigma \in S_{n,A}$ is exactly s_A. As we discussed in Equation (23), given $\sigma \in S_{n,A}$ with $A = \{x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}$, it holds that:

$$
\begin{aligned}
&P_\sigma(\{x_{\sigma(1)}\}) = \ldots = P_\sigma(\{x_{\sigma(k)}\}) = 0.\\
&P_\sigma(\{x_{\sigma(k+1)}\}) = P_0(\{x_{\sigma(1)}, \ldots, x_{\sigma(k+1)}\}) - \delta.\\
&P_\sigma(\{x_{\sigma(i)}\}) = P_0(\{x_{\sigma(i)}\}) \quad \forall i = k+2, \ldots, n-1.\\
&P_\sigma(\{x_{\sigma(n)}\}) = P_0(\{x_{\sigma(n)}\}) + \delta.
\end{aligned}
$$

As we can see, we only need to focus on the elements in the positions $k+1$ and n. In addition, the element $k+1$ must be such that $\{x_{\sigma(1)}, \ldots, x_{\sigma(k+1)}\} \notin \mathcal{L}$. Hence, we can choose $\big(n - |A^{\uparrow}|\big)$ elements for the position $k+1$ and then $\big(n - |A| - 1\big)$ for the position n. Thus, there are $s_A = \big(n - |A^{\uparrow}|\big)\big(n - |A| - 1\big)$ different extreme points. We conclude that:

$$\big|ext(B^{\delta}_{d_{TV}}(P_0))\big| \leq \sum_{A \in \mathcal{L}} s_A.$$

To see that $\sigma_1 \in S_{A,n}$ and $\sigma_2 \in S_{B,n}$ induce different extreme points for $A \neq B$, we just need to realize that $P_{\sigma_1}(A) = 0$ and $P_{\sigma_2}(B) = 0$. Assume that there exists $x \in A \setminus B$, whence

$$P_{\sigma_1}\big((A \cap B) \cup \{x\}\big) \leq P_{\sigma_1}(A) = 0, \quad P_{\sigma_2}\big((A \cap B) \cup \{x\}\big) \geq P_{\sigma_2}(\{x\}) > 0.$$

Hence, $P_{\sigma_1} \neq P_{\sigma_2}$. We therefore conclude that:

$$\big|ext(B^{\delta}_{d_{TV}}(P_0))\big| = \sum_{A \in \mathcal{L}} s_A.$$

Let us derive now the formula for the maximum number of extreme points. Taking Equation (23) into account, two different permutations σ and σ' give rise to the same extreme point, i.e. $P_\sigma = P_{\sigma'}$, if and only if:

$$\begin{aligned} k_\sigma &= k_{\sigma'}, \\ \{\sigma(1), \dots, \sigma(k_\sigma)\} &= \{\sigma'(1), \dots, \sigma'(k_{\sigma'})\}, \\ \sigma(k_\sigma) &= \sigma'(k_{\sigma'}), \\ \{\sigma(k_\sigma + 2), \dots, \sigma(n-1)\} &= \{\sigma'(k_{\sigma'} + 2), \dots, \sigma'(n-1)\}, \\ \sigma(n) &= \sigma'(n). \end{aligned}$$

This means for any permutation σ, there are $k_\sigma! \cdot (n-2-k_\sigma)!$ different permutations σ' such that $P_\sigma = P_{\sigma'}$. This number corresponds to the possible ways of combining $\sigma(1), \dots, \sigma(k_\sigma)$ in the first k_σ positions (that is, $k_\sigma!$), and the possible ways of combining $\sigma(k_\sigma + 2)$, $\dots$, $\sigma(n-1)$ into the positions $k_\sigma + 2, \dots, n-1$ (that is, $(n-2-k_\sigma)!$).

In order to maximise the number of extreme points, we need to minimise the number of permutations giving rise to the same extreme points, so we need to minimise $k_\sigma! \cdot (n-2-k_\sigma)!$. This value can be seen as the denominator of the combinatorial number $\binom{n-2}{k_\sigma}$. When n is even, it is minimised for $k_\sigma = \frac{n-2}{2}$, while for an odd n, it is minimised both for $\frac{n-3}{2}$ and $\frac{n-1}{2}$. In what follows, we consider the value $\frac{n-3}{2}$ for n odd; if we consider the other value, we obtain the same result. We therefore let:

$$k_\sigma = \left\lfloor \frac{n-2}{2} \right\rfloor = \left\lfloor \frac{n}{2} \right\rfloor - 1.$$

Therefore, the number of extreme points is bounded above by:

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

Let us now see that this bound is attained. Let P_0 be the uniform distribution over the n-element space $\mathcal{X}$, and let δ be a distortion parameter such that:

$$P_0\Big(\Big\{x_1, \dots, x_{\lfloor \frac{n}{2} \rfloor - 1}\Big\}\Big) = \frac{1}{n} \cdot \Big(\Big\lfloor \frac{n}{2} \Big\rfloor - 1\Big) \leq \delta < \frac{1}{n} \cdot \Big(\Big\lfloor \frac{n}{2} \Big\rfloor\Big) = P_0\Big(\Big\{x_1, \dots, x_{\lfloor \frac{n}{2} \rfloor}\Big\}\Big).$$

This means that the set $\mathcal{L}$ defined in Equation (22) is:

$$\mathcal{L} = \{A \subseteq \mathcal{X} \mid \underline{P}_{TV}(A) = 0\} = \Big\{A \subseteq \mathcal{X} \Big| \, |A| \leq \Big\lfloor \frac{n}{2} \Big\rfloor - 1 \Big\}.$$

Thus, $s_A = 0$, for any A with cardinality $|A| < \lfloor \frac{n}{2} \rfloor - 1$ because $A^\uparrow = \mathcal{X}$, while for those events A with cardinality $|A| = \lfloor \frac{n}{2} \rfloor - 1$,

$$s_A = (n - |A^\uparrow|)(n - |A| - 1) = \Big(n - \lfloor \frac{n}{2} \rfloor + 1\Big) \cdot \Big(n - \lfloor \frac{n}{2} \rfloor\Big).$$

Furthermore, the number of events A of cardinality $|A| = \lfloor \frac{n}{2} \rfloor - 1$ is:

$$\binom{n}{\lfloor \frac{n}{2} \rfloor - 1}.$$

Then the number of extreme points of $B^{\delta}_{d_{TV}}(P_0)$ is:

$$\binom{n}{\lfloor \frac{n}{2} \rfloor - 1} \cdot \Big(n - \Big\lfloor \frac{n}{2} \Big\rfloor + 1\Big) \cdot \Big(n - \Big\lfloor \frac{n}{2} \Big\rfloor\Big).$$

The proof concludes once we realize that the previous expression can be equivalently expressed as:

$$\frac{n!}{\left(\lfloor \frac{n}{2} \rfloor - 1\right)! \cdot \left(n - \lfloor \frac{n}{2} \rfloor + 1\right)} \cdot \left(n - \left\lfloor \frac{n}{2} \right\rfloor + 1\right) \cdot \left(n - \left\lfloor \frac{n}{2} \right\rfloor\right)$$
$$= \frac{n!}{\left(\lfloor \frac{n}{2} \rfloor - 1\right)! \cdot \left(n - \lfloor \frac{n}{2} \rfloor - 1\right)}. \quad \square$$

B.2. **Kolmogorov model.** With respect to the Kolmogorov model, from Proposition 9 the maximum number of extreme points of $B^{\delta}_{d_K}(P_0)$ cannot be increased when we consider larger δ, and also that the associated lower probability is completely monotone but not a probability interval. With respect to which values of δ are informative, note that, if there exists $x^* \in \mathcal{X}$ such that $F_{P_0}(x^*) + \delta \leq 1$, then we can define the function F by:

$$F(x) = \begin{cases} \min\{F_{P_0}(x), F_{P_0}(x^*) - \delta\}, & \text{if } x < x^*. \\ F_{P_0}(x^*) - \delta, & \text{if } x = x^*. \\ F_{P_0}(x), & \text{if } x > x^*. \end{cases}$$

F is a cumulative distribution function. Its associated probability, P, belongs to $B^{\delta}_{d_K}(P_0)$:

- If $x < x^*$, there are two possibilities:
  (1) If $F_{P_0}(x^*) - \delta \leq F_{P_0}(x)$, then

$$\begin{aligned} |F(x) - F_{P_0}(x)| &= |\min\{F_{P_0}(x), F_{P_0}(x^*) - \delta\} - F_{P_0}(x)| \\ &= |F_{P_0}(x^*) - \delta - F_{P_0}(x)| = -F_{P_0}(x^*) + \delta + F_{P_0}(x) \leq \delta. \end{aligned}$$

  (2) If $F_{P_0}(x^*) - \delta > F_{P_0}(x)$, then

$$|F(x) - F_{P_0}(x)| = |F_{P_0}(x) - F_{P_0}(x)| = 0.$$

- If $x > x^*$, it holds that:

$$|F(x) - F_{P_0}(x)| = |F_{P_0}(x) - F_{P_0}(x)| = 0.$$

- Finally, if $x = x^*$,

$$|F(x) - F_{P_0}(x)| = |F_{P_0}(x^*) - \delta - F_{P_0}(x^*)| = \delta.$$

Thus, $P \in B^{\delta}_{d_K}(P_0)$, and also $d_K(P, P_0) = \delta$. We conclude that

$$\max_{P \in B^{\delta}_{d_K}(P_0)} d_K(P, P_0) = \delta.$$

On the other hand, if there exists $P \in B^{\delta}_{d_K}(P_0)$ such that $d_K(P, P_0) = \delta$. Then, there exists $x \in \mathcal{X}$ such that $|F_P(x) - F_{P_0}(x)| = \delta$. This implies that either $F_{P_0}(x) + \delta \leq 1$ or $F_{P_0}(x) - \delta \geq 0$. As a consequence,

$$\max_{P \in B^{\delta}_{d_K}(P_0)} d_K(P, P_0) = \delta \Leftrightarrow \delta \leq \max\left\{\max_{x \in \mathcal{X}}(1 - F_{P_0}(x)), \max_{x \in \mathcal{X}} F_{P_0}(x)\right\}.$$

B.3. **The L_1 model.** We have established in Theorem 11 that, when $\delta > 0$ is small enough, the lower envelope $\underline{P}_{L_1}$ of $B^{\delta}_{d_{L_1}}(P_0)$ is $\underline{P}_{L_1}(A) = P_0(A) - \frac{\delta}{\varphi(n,|A|)}$. One may think that for general values of δ (that is, including those for which $\underline{P}_{L_1}(A) = 0$ for some event $A \neq \emptyset$), the lower envelope $\underline{P}_{L_1}$ of $B^{\delta}_{d_{L_1}}(P_0)$ is given by $\underline{P}_{L_1}(A) = \max\left\{P_0(A) - \frac{\delta}{\varphi(n,|A|)}, 0\right\}$. However, as our next example shows, this is not true.

Example 9. *Consider a four-element space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, the probability measure $P_0 = (0.15, 0.2, 0.3, 0.35)$ and the distortion parameter $\delta = 2.1$. The probability measure $P = (0, 0, 0.5, 0.5)$ satisfies $P(A) \geq \max\left\{P_0(A) - \frac{\delta}{\varphi(|A|)}, 0\right\}$ for every $A \subseteq \mathcal{X}$, but $d_{L_1}(P, P_0) = 2.2 > \delta$, hence $P \notin B^{\delta}_{d_{L_1}}(P_0)$. This means that $\underline{P}_{L_1}(A)$ is not given by $\max\left\{\underline{P}_0(A) - \frac{\delta}{\varphi(|A|)}, 0\right\}$.* ♦

We have not succeeded in finding a close expression of the lower envelope $\underline{P}_{L_1}$ of $B^{\delta}_{d_{L_1}}(P_0)$ for large values of δ, and this is left as an open problem. It is not difficult to establish, though, that for $\delta \geq 2^{n-1}$ $\underline{P}_{L_1}$ is the vacuous lower probability. Therefore, the model is completely uninformative for such large distortions.

References

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

[2] P.K. Black. *Random sets*, chapter Geometry structure of lower probabilities, pages 361–383. Springer, 1997.

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

[5] A. Bronevich and A. Lepskiy. Imprecision indices: axiomatic, properties and applications. *International Journal of General Systems*, 44(7-8):812–832, 2015.

[6] C. Corsato, R. Pelessoni, and P. Vicig. Nearly-linear uncertainty measures. *International Journal of Approximate Reasoning*, 114:1–28, 2019.

[7] F. Cuzzolin. Geometry of upper probabilities. In *Proceedings of the 3rd International Symposium on Imprecise Probability: Theories and Applications*, pages 188–203. Carleton Scientific, 2003.

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

[9] S. Destercke and D. Dubois. The role of generalised p-boxes in imprecise probability models. In *Proceedings of the 6th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'2009)*, pages 179–188, 2009.

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

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

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

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

[16] E. Miranda and I. Montes. Coherent updating of non-additive measures. *International Journal of Approximate Reasoning*, 56(B):159–177, 2015.
[17] I. Montes and S. Destercke. On extreme points of p-boxes and belief functions. *Annals of Mathematics and Artificial Intelligence*, 81:405–428, 2017.
[18] I. Montes, E. Miranda, and S. Destercke. Pari-mutuel probabilities as an uncertainty model. *Information Sciences*, 481:550 – 573, 2019.
[19] I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models: Part I- New results on old models. *International Journal of General Systems*, 2020. Submitted for publication.
[20] I. Montes, E. Miranda, and P. Vicig. 2-monotone outer approximations of coherent lower probabilities. *International Journal of Approximate Reasoning*, 101:181–205, 2018.
[21] I. Montes, E. Miranda, and P. Vicig. Outer approximations of coherent lower probabilities with belief functions. *International Journal of Approximate Reasoning*, 110:1–30, 2019.
[22] 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.
[23] F. Österreicher. Csiszárs f-divergences: basic properties. Talk at the Workshop of the research group in mathematical inequalities and applications, 2002. Victoria University, Melbourne.
[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] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*. John Wiley & Sons, 2005.
[26] T. Seidenfeld and L. Wasserman. Dilation for sets of probabilities. *The Annals of Statistics*, 21:1139–54, 1993.
[27] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
[28] M. C. M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767–791, 2011.
[29] P. Walley. Coherent lower (and upper) probabilities. Technical Report 22, University of Warwick, Coventry, 1981.
[30] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
[31] A. Wallner. Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning*, 44(3):339–357, 2007.

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

UMR CNRS 7253 Heudiasyc, Sorbonne universités, Université de technologie de Compiègne CS 60319 - 60203 Compiègne cedex, France
Email address: sebastien.destercke@hds.utc.fr