Proceedings of Machine Learning Research 215:367–378, 2023 ISIPTA 2023

Neighbourhood Models Induced by the Euclidean Distance and the Kullback-Leibler Divergence

Ignacio Montes IMONTES@UNIOVI.ES
Department of Statistics and Operations Research, University of Oviedo

Abstract

Neighbourhood or distortion models are particular imprecise probability models that appear by creating a neighbourhood around a probability measure given a distorting function and a distortion parameter. This paper investigates the distortion models obtained when considering the Euclidean distance or the Kullback-Leibler divergence as distorting function. We analyse the main properties of the credal sets induced by these two distorting functions as well as the main properties of the associated coherent lower previsions. To conclude the paper, we compare these two models with other well-known distortion models: the pari-mutuel, linear vacuous, constant odds ratio and total variation models.

Keywords: Euclidean distance, Kullback-Leibler divergence, distortion model, lower prevision, credal set

1. Introduction

An important goal in statistical or probabilistic inferences is that of being robust, meaning that small changes in the data do not produce big changes in the results. To achieve the goal of robustifying the probabilistic model some alternatives are to transform a probability measure by means of an increasing function [5, 6], to consider almost linear transformations of the probability measure [9], or more generally to consider distortion or neighbourhoood models, as suggested in [17]. These models create a neighbourhood around a probability measure and for this aim two tools are needed: a distorting function to compare probability measures and a distortion parameter measuring the amount of imprecision to be added to the model.

In our previous papers [12, 22, 23] we performed a detailed analysis of distortion or neighbourhood models. More in detail, (i) we showed that these models include as particular cases direct transformations of a probability measure by means of an increasing function, as was done for example by Bronevich [5, 6] (see [22, Prop.2]); (ii) we proved that some well-known models within the imprecise probability theory can be seen as neighbourhood or distortion models. This is the case of the linear vacuous [17, 34], pari-mutuel [21, 29, 34] or constant odds ratio models [34] (see [22, Secs.4-6]); (iii) we investigated the distortion models induced by some well-known distances such as the total variation, Kolmogorov or L_1 distances (see [23, Secs.2-4]); (iv) we performed a comparative analysis between the six models mentioned above (see [23, Sec.5]); and finally (v) we analysed the behaviour of these six models under processing [12]. Furthermore, in our last paper [1] we explained how to use distortion models as an alternative approach to that in [24, 25, 26, 27] to estimate human error probabilities.

A crucial point in the use of distortion or neighbourhood models is the choice of the distorting function. Apart from the distortion models studied in our previous papers, it may be intriguing to consider the use of the Euclidean distance or the Kullback-Leibler divergence as distorting functions. Their use is more than reasonable: on the one hand, when speaking about distances the Euclidean distance is of utmost relevance, and on the other hand it is unquestionable that the Kullback-Leibler divergence possesses very appropriate properties for comparing probability measures, and it has been widely used in statistics [13], information theory [4] or machine learning [14], among many other fields. Even more, the Kullback-Leibler divergence has already been used to create a distortion of a credal set under the terminology *discounting credal sets* [28, Sec.4]. This approach consists in distorting any probability measure in the credal set with the same factor and later taking the convex hull of the union of all the models. This approach is slightly different than ours, because we aim at distorting a probability measure creating a credal set and investigating the properties of the obtained model, instead of directly distorting a credal set.

Therefore, the goal of this paper is to perform a detailed analysis of the neighbourhood or distortion models induced by the Euclidean distance and Kullback-Leibler divergence. For this aim, after introducing some basic notions in Section 2, in Sections 3 and 4 we analyse the properties of the closed and convex sets of probability measures determined by both models. We also analyse the sets of extreme points and try to obtain an expression of the associated coherent lower previsions. Moreover, we study if these models are closed under conditioning with respect to the regular extension. Later, in Section 5 we compare the models induced by the Euclidean distance and Kullback-Leibler divergence with the linear vacuous, pari-mutuel and constant odds ratio models and the neighbourhood model induced by the total variation distance. This comparison is done in terms of the amount of imprecision and the properties satisfied by the associated coherent lower previsions. We conclude the paper with some final comments in Section 6. Due to space limitations, proofs have been omitted.

2. Preliminaries

In this paper we consider a finite possibility space $\mathcal{X} = \{x_1, \ldots, x_n\}$. We denote by $\mathbb{P}(\mathcal{X})$ the set of probability measures defined on $\mathcal{P}(\mathcal{X})$ and by $\mathbb{P}^*(\mathcal{X})$ the subset of probability measures assigning strictly positive probability to the non-empty events.

2.1. Credal Sets and (Coherent) Lower Previsions

In this subsection we take a quick look at credal sets and (coherent) lower previsions, even if we assume the reader to be familiar with these basic notions. We refer to [2, 34] for a detailed introduction.

A function $f : \mathcal{X} \to \mathbb{R}$ is called a *gamble*, and the set of all the gambles in $\mathcal{X}$ is denoted by $\mathcal{L}(\mathcal{X})$. A *lower prevision* is a function $\underline{P} : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$ and its conjugate *upper prevision* follows from the duality relation $\overline{P}(f) = -\underline{P}(-f)$ for any $f \in \mathcal{L}(\mathcal{X})$.

A lower prevision determines a closed and convex set of probability measures, usually called *credal set*, given by:

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}(f)\ \forall f \in \mathcal{L}(\mathcal{X})\}, \quad (1)$$

where $P(f)$ denotes the prevision or expectation of f with respect to the probability measure P. The lower prevision $\underline{P}$ is called *coherent* when

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

Credal sets (i.e., closed and convex sets of probability measures) and coherent lower previsions are in one-to-one correspondence: a credal set determines a coherent lower prevision by taking lower envelopes, and the initial credal set can be retrieved by using Equation (1).

Moreover, since the credal set is a closed and convex set, it is characterised by its extreme points. $P \in \mathcal{M}(\underline{P})$ is an *extreme point* if given $P_1, P_2 \in \mathcal{M}(\underline{P})$ and $\alpha \in (0, 1)$ such that $P = \alpha P_1 + (1 - \alpha)P_2$, then $P = P_1 = P_2$. The set of extreme points of $\mathcal{M}(\underline{P})$ is denoted by $\text{ext}\big(\mathcal{M}(\underline{P})\big)$, and it follows that $\mathcal{M}(\underline{P})$ is the convex hull of $\text{ext}\big(\mathcal{M}(\underline{P})\big)$.

When we restrict a coherent lower prevision to the domain $\{I_A \mid A \subseteq \mathcal{X}\}$ we obtain a *coherent lower probability*. In this case we use the notation $\underline{P}(A)$ instead of $\underline{P}(I_A)$ for any $A \subseteq \mathcal{X}$. That is, we identify the lower probability of an event with the lower prevision of its indicator. In a similar manner, we define the *coherent upper probability* by means of the conjugacy relation $\overline{P}(A) = 1 - \underline{P}(A^c)$ for any $A \subseteq \mathcal{X}$. In general, the set of probability measures compatible with the coherent lower probability is greater than the initial credal set:

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

Both sets of probability measures coincide when the coherent lower prevision satisfies in addition the property of *2-monotonicity*, meaning that:

$$\underline{P}(f \wedge g) + \underline{P}(f \vee g) \geq \underline{P}(f) + \underline{P}(g) \quad \forall f, g \in \mathcal{L}(\mathcal{X}),$$

where $\wedge$ and $\vee$ denote the pointwise minimum and maximum, respectively. This inequality restricted to indicators of events becomes:

$$\underline{P}(A \cap B) + \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) \quad \forall A, B \subseteq \mathcal{X},$$

and when the coherent lower probability satisfies it, $\underline{P}$ is called *2-monotone* on events.

When the coherent lower prevision satisfies the property of 2-monotonicity, the restriction to events (i.e., the coherent lower probability) determines the coherent lower prevision by using the Choquet integral [8]:

$$\underline{P}(f) = \min f + \int_{\min f}^{\max f} \underline{P}(\{x \in \mathcal{X} \mid f(x) \geq t\})\mathrm{d}t. \quad (2)$$

An additional property of 2-monotone lower previsions is that their credal set has at most $n!$ extreme points and that they can be easily computed [7, 30].

2.2. Updating a Coherent Lower Prevision

In the imprecise probability framework there is not a unique way of defining a conditional coherent lower prevision [19, 20]. One of these possibilities is to use the *regular extension* [34], defined by:

$$\underline{P}(f \mid B) = \inf \left\{P(f|B) \mid P \in \mathcal{M}(\underline{P}),\ P(B) > 0\right\}$$

whenever $\overline{P}(B) > 0$, and $\underline{P}(f \mid B) = \min_{x \in \mathcal{X}} f(x)$ when $\underline{P}(B) = 0$. Here, $P(f \mid B)$ denotes the prevision or expectation of f with respect to the probability measure $P(\cdot \mid B)$ given by $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ for any $A \subseteq \mathcal{X}$. When moreover $\underline{P}$ is 2-monotone, there is a simple expression for the restriction of $\underline{P}(\cdot \mid B)$ to events whenever $\overline{P}(B) > 0$, given by [33, Thm.7.2]:

$$\underline{P}(A \mid B) = \frac{\underline{P}(A \cap B)}{\underline{P}(A \cap B) + \overline{P}(A^c \cap B)} \quad (3)$$

whenever $\overline{P}(A^c \cap B) > 0$, and $\underline{P}(A \mid B) = 1$ otherwise.

2.3. Distortion Models

Distortion or *neighbourhood* models [22, 23] are obtained by doing a distortion of a probability measure or by creating a neighbourhood around a probability measure. Formally, given a probability measure P_0, a distortion factor $\delta > 0$ and a distorting function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \to \mathbb{R}^+$, we define the closed ball:

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

as well as its associated coherent lower prevision given by:

$$\underline{P}_d(f) = \inf \left\{P(f) \mid P \in B_d^\delta(P_0)\right\}.$$

The distorting function d may, or not, satisfy some properties such as ([22, Sec.3]):

Ax.1 Positive definiteness: $d(P_1, P_2) = 0$ if and only if $P_1 = P_2$.

Ax.2 Triangle inequality: $d(P_1, P_3) \leq d(P_1, P_2) + d(P_2, P_3)$ for every $P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.

Ax.3 Symmetry: $d(P_1, P_2) = d(P_2, P_1)$ for every $P_1, P_2 \in \mathbb{P}(\mathcal{X})$.

Ax.4 Convexity: $d(\alpha P_1 + (1 - \alpha)P_2, P_3) \leq \max\{d(P_1, P_3), d(P_2, P_3)\}$ for every $\alpha \in [0, 1]$ and $P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.

Ax.5 Continuity: for every $P, P_1, P_2 \in \mathbb{P}(\mathcal{X})$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that $\max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)| < \delta$ implies $|d(P_1, P) - d(P_2, P)| < \varepsilon$.

Even if none of these properties is imposed to the distorting function, satisfying (some of) them would guarantee that the ball $B_d^\delta(P_0)$ satisfies good topological or analytical properties. As an excellent example, we may mention that in general the probabilistic information given by $\underline{P}_d$ and $B_d^\delta(P_0)$ may be different in the sense that $B_d^\delta(P_0) \neq \mathcal{M}(\underline{P}_d)$; nevertheless, whenever the distorting function is continuous (Ax.5) and convex (Ax.4) both sets of probability measures coincide: $B_d^\delta(P_0) = \mathcal{M}(\underline{P}_d)$ [22, Prop.3.1].

In our previous papers [22, 23] we analysed some well-known examples of distortion models such as:

- The *Linear Vacuous* (LV) model [17, 34], given by:

$$\underline{P}_{LV}(A) = (1 - \delta)P_0(A) \quad \forall A \subset \mathcal{X}$$

  and $\underline{P}_{LV}(\mathcal{X}) = 1$.

- The *Pari Mutuel* (PMM) model [21, 29, 34], given by:

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

- The *Constant Odds Ratio* (COR) [3, 31, 34], given by the coherent lower prevision $\underline{P}_{COR}$ defined by means of the implicit equation:

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

  for any $f \in \mathcal{L}(\mathcal{X})$, where $g^+ = \max\{g, 0\}$ and $g^- = \max\{-g, 0\}$.

- The *Total Variation* (TV) model [16], given by:

$$\underline{P}_{TV}(A) = \max\{P_0(A) - \delta, 0\} \quad \forall A \subset \mathcal{X}$$

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

The LV, PMM and TV models are 2-monotone, hence their extension to gambles are obtained by applying the Choquet integral as in Equation (2). Moreover, the TV model considers as distorting function the TV-distance, hence it immediately satisfies axioms Ax.1–Ax.3, besides continuity and convexity. The five properties are also satisfied by the distorting function associated with the COR model. The distorting function associated with the LV do not satisfy symmetry, while that associated with the PMM satisfies neither symmetry nor the triangle inequality. We refer to [23, Sec.5.2] for a survey about these properties.

In our former studies we assumed that $P_0 \in \mathbb{P}^*(\mathcal{X})$ and that the distortion parameter is small enough so that $B_d^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, even if some general results were given in [23, Appendix B]. In what follows, we perform a detailed study of the distortion models induced by the Euclidean distance and the Kullback-Leibler divergence, and we make throughout this same technical assumption.

3. Distortion Model Induced by the Euclidean Distance

The Euclidean distance between two probability measures P_1, P_2 is given by:

$$d_E(P_1, P_2) = \sqrt{\sum_{i=1}^{n} \big(P_1(\{x_i\}) - P_2(\{x_i\})\big)^2}.$$

Being a distance it satisfies positive definiteness (Ax.1), the triangle inequality (Ax.2) and symmetry (Ax.3), but it is continuous (Ax.5) and convex (Ax.4) too, so

$$B_E^\delta(P_0) = \left\{P \in \mathbb{P}(\mathcal{X}) \mid d_E(P, P_0) \leq \delta\right\}$$

is a closed and convex set of probability measures (i.e, a credal set) [22, Prop.3.1] whose associated coherent lower prevision $\underline{P}_E$ is given by:

$$\underline{P}_E(f) = \min_{P \in B_E^\delta(P_0)} P(f) \quad \forall f \in \mathcal{L}(\mathcal{X}). \tag{4}$$

Moreover, the ball $B_E^\delta(P_0)$ can be expressed as:

$$B_E^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}_E(f)\ \forall f \in \mathcal{L}(\mathcal{X})\}.$$

Next, we perform a detailed analysis of this model.

3.1. Expression of the Lower Prevision $\underline{P}_E$

We start proving that the minimum value in Equation (4) is attained in the probability measures at distance δ from P_0.

Proposition 1 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Given a non-constant gamble f, if $\underline{P}_E(f) = P(f)$ or $\overline{P}_E(f) = P(f)$ for some $P \in B_E^\delta(P_0)$, then $d_E(P, P_0) = \delta$.*

This result helps us to find an explicit form of $\underline{P}_E(f)$, that can be obtained by solving an optimisation problem. For this aim, denote the probability measure $P \in B_E^\delta(P_0)$ as

$$P(\{x_i\}) = P_0(\{x_i\}) + \alpha_i \quad \forall i = 1, \ldots, n, \tag{5}$$

where the values $\alpha_1, \ldots, \alpha_n$ satisfy $\sum_{i=1}^n \alpha_i = 0$. Moreover, since $P \in B_E^\delta(P_0)$, it holds that:

$$d_E(P, P_0)^2 = \sum_{i=1}^n \big(P(\{x_i\}) - P_0(\{x_i\})\big)^2 = \sum_{i=1}^n \alpha_i^2 \leq \delta^2.$$

Finally, if we express a gamble f as $f = \sum_{i=1}^n a_i I_{\{x_i\}}$, its prevision is given by:

$$P(f) = \sum_{i=1}^n a_i \big(P_0(\{x_i\}) + \alpha_i\big) = P_0(f) + \sum_{i=1}^n a_i \alpha_i.$$

Hence, in order to obtain $\underline{P}_E$ and $\overline{P}_E$ we set up the optimisation problem:

$$\min_{\alpha_1,\ldots,\alpha_n} / \max_{\alpha_1,\ldots,\alpha_n}\ P_0(f) + \sum_{i=1}^n a_i \alpha_i$$

subject to:

$$\sum_{i=1}^n \alpha_i = 0 \qquad \sum_{i=1}^n \alpha_i^2 = \delta^2. \tag{6}$$

The former condition in Equation (6) guarantees that we obtain a probability measure P as in Equation (5), while the latter guarantees that $d_E(P, P_0) = \delta$.

Theorem 2 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. For any gamble $f \in \mathcal{L}(\mathcal{X})$ given by $f = \sum_{i=1}^n a_i I_{\{x_i\}}$, it holds that:*

$$\underline{P}_E(f) = P_0(f) - \delta\sqrt{n} S_f, \quad \overline{P}_E(f) = P_0(f) + \delta\sqrt{n} S_f$$

where $\bar{f} = \frac{1}{n}\sum_{i=1}^n a_i$ and $S_f^2 = \frac{1}{n}\sum_{i=1}^n \big(a_i - \bar{f}\big)^2$.

Also, by considering indicator of events we obtain the associated coherent lower and upper probabilities.

Corollary 3 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. For any $A \subseteq \mathcal{X}$, it holds that:*

$$\underline{P}_E(A) = P_0(A) - \delta\sqrt{\frac{|A|(n - |A|)}{n}}. \tag{7}$$

$$\overline{P}_E(A) = P_0(A) + \delta\sqrt{\frac{|A|(n - |A|)}{n}}. \tag{8}$$

3.2. Properties of $B_E^\delta(P_0)$ and $\underline{P}_E, \overline{P}_E$

In this subsection we investigate some properties of the ball $B_E^\delta(P_0)$ and its associated coherent lower and upper prevision. We start characterising the set of extreme points of $B_E^\delta(P_0)$. We already know from Proposition 1 that the extreme points are in the boundary of $B_E^\delta(P_0)$. We next prove the converse: all the probability measures on the boundary are extreme points.

Proposition 4 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. It holds that:*

$$ext\big(B_E^\delta(P_0)\big) = \{P \in \mathbb{P}^*(\mathcal{X}) \mid d_E(P, P_0) = \delta\}.$$

From this result we deduce that $B_E^\delta(P_0)$ has infinitely many extreme points. Following our comments at the end of Section 2.1, this means that $\underline{P}_E$ is not 2-monotone because 2-monotone lower previsions have at most $n!$ different extreme points. In spite of these comments, $\underline{P}_E$ satisfies the property of 2-monotonicity on events.

Proposition 5 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Then, $\underline{P}_E$ is 2-monotone on events.*

We conclude this subsection showing an example of an Euclidean model.

Example 1 *Consider the probability measure $P_0 = (0.5, 0.3, 0.2)$ and $\delta = 0.1$. The graphical representation of the credal set $B_E^\delta(P_0)$, given in Figure 1, shows that the ball is a circle around P_0. There, we can also see the credal set determined by the restriction to events, given by:*

$$\{P \in \mathbb{P}^*(\mathcal{X}) \mid P(A) \geq \underline{P}_E(A)\ \forall A \subseteq \mathcal{X}\},$$

where the values $\underline{P}_E(A)$ are given in Equations (7-8), respectively. Since $\underline{P}$ is 2-monotone on events but not on gambles, both credal sets do not coincide. ♦

3.3. Conditioning $\underline{P}_E$

We now want to update an Euclidean model $B_E^\delta(P_0)$, and we wonder whether the updated model $\underline{P}_E(\cdot \mid B)$ obtained by applying regular extension corresponds to an Euclidean model with respect to $P_0(\cdot \mid B)$ and a distortion parameter δ^*. We show in the next example that the answer is negative.

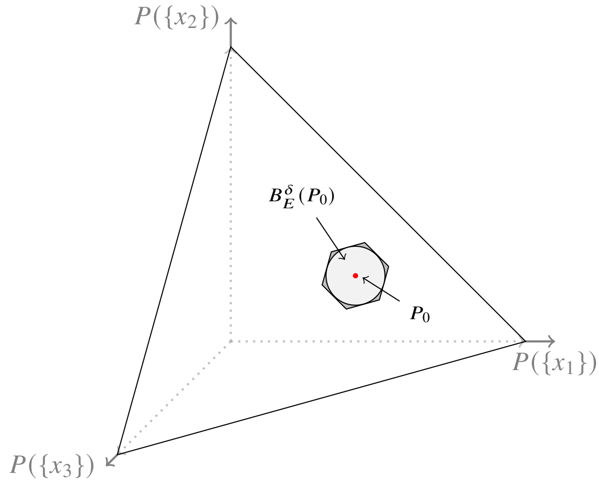

Figure 1: Graphical representation of the probability measure P_0 (in red), the credal set $B_E^\delta(P_0)$ (filled in light gray) and the credal set determined by the restriction to events (filled in dark gray).

Example 2 *Consider a four element possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and the probability measure P_0 given by $P_0(\{x_i\}) = 1/4$ for $i = 1, \ldots, 4$. Taking $\delta = \frac{1}{5\sqrt{3}}$, we obtain the following values of $\underline{P}_E$ on events:*

$\lvert A\rvert$	1	2	3	4
$\underline{P}_E(A)$	0.15	$0.5 - \delta$	0.65	1
$\overline{P}_E(A)$	0.35	$0.5 + \delta$	0.85	1

Since $\underline{P}_E$ is 2-monotone on events, we compute $\underline{P}_E(A \mid B)$ (for $B \neq \emptyset$ and $A \subseteq B$) by applying Equation (3):

$$\underline{P}_E(A \mid B) = \frac{P_0(A \cap B) - \delta\sqrt{\frac{|A\cap B|(4-|A\cap B|)}{4}}}{P_0(B) - \delta\sqrt{\frac{|A\cap B|(4-|A\cap B|)}{4}} + \delta\sqrt{\frac{|A^c\cap B|(4-|A^c\cap B|)}{4}}}. \quad (9)$$

Taking $B = \{x_1, x_2, x_3\}$, if $\underline{P}_E$ was an Euclidean model determined by $P_0^* \in \mathbb{P}(B)$ and δ^* we would have:

$$\underline{P}_E(A \mid B) = P_0^*(A) - \delta^*\sqrt{\frac{|A|(3-|A|)}{3}} \quad \forall A \subseteq B. \quad (10)$$

Moreover, since P_0 is uniform, P_0^* would be uniform too because $\underline{P}_E(\{x_1\}|B) = \underline{P}_E(\{x_2\}|B) = \underline{P}_E(\{x_3\}|B)$. From Equations (9) and (10), given $A \subseteq B$ we obtain:

$$|A| = 1 \Rightarrow \frac{\frac{1}{4} - \delta\frac{\sqrt{3}}{2}}{\frac{3}{4} - \delta\frac{\sqrt{3}}{2} + \delta} = \frac{1}{3} - \delta^*\sqrt{\frac{2}{3}},$$

$$|A| = 2 \Rightarrow \frac{\frac{1}{2} - \delta}{\frac{3}{4} - \delta + \delta\frac{\sqrt{3}}{2}} = \frac{2}{3} - \delta^*\sqrt{\frac{2}{3}},$$

but δ^* cannot satisfy both equalities simultaneously. This contradicts our initial assumption, hence $\underline{P}_E(\cdot \mid B)$ is not an Euclidean model. ♦

4. Distortion Model Induced by the Kullback-Leibler Divergence

We consider now the Kullback-Leibler divergence between two probability measures P_1, P_2, given by:

$$D_{KL}(P_1, P_2) = \sum_{i=1}^{n} P_1(\{x_i\}) \log\left(\frac{P_1(\{x_i\})}{P_2(\{x_i\})}\right),$$

assuming that if $P_2(\{x_i\}) = 0$ then $P_1(\{x_i\}) = 0$ and that $0 \cdot \infty = 0$. The Kullback-Leibler divergence satisfies positive definiteness (Ax.1), but it does not satisfy the triangle inequality (Ax.2) and symmetry (Ax.3). Nevertheless, it is convex (Ax.4) and continuous (Ax.5), hence

$$B_{KL}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid D_{KL}(P, P_0) \leq \delta\}$$

is closed and convex. Its associated lower prevision $\underline{P}_{KL}$ is given by:

$$\underline{P}_{KL}(f) = \min_{P \in B_{KL}^\delta(P_0)} P_0(f) \quad \forall f \in \mathcal{L}(\mathcal{X}), \quad (11)$$

and the ball can be equivalently expressed as:

$$B_{KL}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}_{KL}(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\}.$$

Our goal now is to investigate the properties of the KL-distortion model. We follow the same steps as for the Euclidean model: we look for an expression for the coherent lower prevision in Equation (11), we investigate the main properties of the model and its behaviour under conditioning.

4.1. Expression of the Lower Prevision

First of all, we show that the minimum value in Equation (11) is attained in the probability measures at a divergence δ from P_0.

Proposition 6 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Given a non-constant gamble f, if $\underline{P}_{KL}(f) = P(f)$ or $\overline{P}_{KL}(f) = P(f)$ for some $P \in B_{KL}^\delta(P_0)$, then $D_{KL}(P, P_0) = \delta$.*

Next we try to obtain an expression for $\underline{P}_{KL}$. For this aim, again we express the gamble f as $f = \sum_{i=1}^n a_i I_{\{x_i\}}$. Let us express the probability measure $P \in B_{KL}^\delta(P_0)$ as

$$P(\{x_i\}) = \alpha_i P_0(\{x_i\}) \quad \forall i = 1, \ldots, n, \quad (12)$$

where the values $\alpha_1, \dots, \alpha_n$ satisfy $\sum_{i=1}^n P_0(\{x_i\})\alpha_i = 1$ and $\alpha_i \geq 0$ for $i = 1, \dots, n$. Hence:

$$P(f) = \sum_{i=1}^n a_i P(\{x_i\}) = \sum_{i=1}^n a_i \alpha_i P_0(\{x_i\})$$

and

$$D_{KL}(P, P_0) = \sum_{i=1}^n \alpha_i P_0(\{x_i\}) \log(\alpha_i).$$

In this way, for obtaining $\underline{P}_{KL}(f)$ and $\overline{P}_{KL}(f)$ we need to solve the optimisation problem

$$\min_{\alpha_1,\dots,\alpha_n} / \max_{\alpha_1,\dots,\alpha_n} \sum_{i=1}^n a_i \alpha_i P_0(\{x_i\})$$

subject to

$$\sum_{i=1}^n \alpha_i P_0(\{x_i\}) = 1, \quad \sum_{i=1}^n \alpha_i P_0(\{x_i\}) \log(\alpha_i) = \delta.$$

The former condition guarantees that we obtain a probability measure P while the latter guarantees that $D_{KL}(P, P_0) = \delta$. This is a convex optimisation problem, where the feasible region is closed, convex and bounded, hence there exists a unique solution. However, giving an explicit formula for $\underline{P}_{KL}(f)$ seems a challenging problem.

In spite of the complexity of the problem, we can give an easier expression for $\underline{P}_{KL}$ in indicators of events (i.e., for the restriction to events).

Theorem 7 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Let $A \subseteq \mathcal{X}$ and $P \in B^\delta_{KL}(P_0)$ such that $\underline{P}_{KL}(A) = P(A)$. If we express P as in Equation (12), then it holds that:*

1. *there exists $\alpha < 1$ such that $\alpha_i = \alpha$ for any $x_i \in A$;*
2. *there exists $\beta > 1$ such that $\alpha_i = \beta$ for any $x_i \notin A$;*
3. *α and β satisfy the relation:*
$$\beta\big(1 - P_0(A)\big) = 1 - \alpha P_0(A); \tag{13}$$
4. *letting $p = P_0(A)$, the value α is the solution in the interval $[0, 1]$ of the implicit equation*
$$\alpha p \log(\alpha) + (1 - \alpha p) \log\left(\frac{1 - \alpha p}{1 - p}\right) = \delta; \tag{14}$$
5. *$\underline{P}_{KL}(A) = \alpha P_0(A)$, where α is given in the previous item;*
6. *there exists a convex function g such that $\underline{P}_{KL}(A) = g\big(P_0(A)\big)$ for any $A \subseteq \mathcal{X}$.*

This results helps in computing the restriction to events of $\underline{P}_{KL}$. In particular, we deduce that for computing $\underline{P}_{KL}(A)$ it suffices with finding the value $\alpha < 1$ solving Equation (14).

4.2. Properties of $B^\delta_{KL}(P_0)$, $\underline{P}_{KL}$ and $\overline{P}_{KL}$

We now investigate the main properties of the distortion model induced by the Kullback-Leibler divergence. We start proving that the extreme points are all the probability measures in the boundary of the ball.

Proposition 8 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. It holds that*

$$ext\big(B^\delta_{KL}(P_0)\big) = \{P \in \mathbb{P}^*(\mathcal{X}) \mid D_{KL}(P, P_0) = \delta\}.$$

Following the same argument as for the Euclidean model, this result allows us to deduce that $B^\delta_{KL}(P_0)$ has infinitely many extreme points meaning that $\underline{P}_{KL}$ is not 2-monotone. Nevertheless, even if $\underline{P}_{KL}$ is not 2-monotone, its restriction to events is a 2-monotone lower probability.

Proposition 9 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Then, $\underline{P}_{KL}$ is 2-monotone on events.*

We next show an example to see the shape of $B^\delta_{KL}(P_0)$.

Example 3 *Let us continue with Example 1 considering the probability measure $P_0 = (0.5, 0.3, 0.2)$ and the distortion parameter $\delta = 0.1$. As it can be seen in Figure 2, the shape of $B^\delta_{KL}(P_0)$ is different than the shape of the Euclidean model, which is a circle. This figure also shows the credal set determined by the restriction to events. Since $\underline{P}_{KL}$ is 2-monotone on events but not on gambles, both credal sets do not coincide.* ♦

The last result of this subsection shows a useful property of the Kullback-Leibler model that will be used in the next subsection.

Lemma 10 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. If there are two different $x_i, x_j \in \mathcal{X}$ such that $\underline{P}_{KL}(\{x_i\}) = \underline{P}_{KL}(\{x_j\})$, then $P_0(\{x_i\}) = P_0(\{x_j\})$.*

4.3. Conditioning a Kullback-Leibler Model

We now investigate if conditioning a Kullback-Leibler model by means of regular extension we obtain again a Kullback-Leibler model. According to Proposition 9, $\underline{P}_{KL}$ is 2-monotone on events, meaning that the regular extension on events can be computed using Equation (3). Nevertheless, the next example shows that the Kullback-Leibler model is not preserved under conditioning.

Example 4 *Consider a four element possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, the uniform probability distribution $P_0 = (0.25, 0.25, 0.25, 0.25)$ and the distortion factor $\delta = 0.1$. The values of $\underline{P}_{KL}$ on events are given by:*

$\lvert A\rvert$	1	2	3
$\underline{P}_{KL}(A)$	0.077	0.2802	0.5434

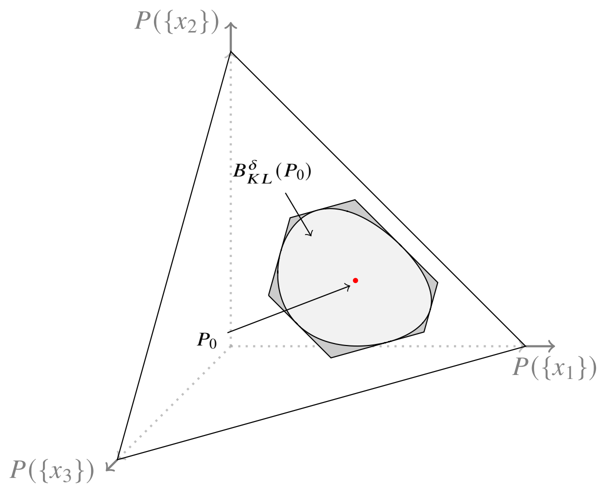

Figure 2: Graphical representation of the credal set $B^{\delta}_{KL}(P_0)$ (filled in light gray) in Example 3 and the credal set determined by the restriction to events (filled in dark gray), together with the probability measure P_0 (in red).

Taking $B = \{x_1, x_2, x_3\}$ and applying the regular extension to any $A \subseteq B$, we obtain the updated model:

$\lvert A\rvert$	1	2
$\underline{P}_{KL}(A\lvert B)$	0.097	0.3802

Let us assume that $\underline{P}_{KL}(\cdot|B)$ is a Kullback-Leibler model induced by P^*_0 and δ^*. From Lemma 10 we know that $P^*_0 = (1/3, 1/3, 1/3)$. Also, Theorem 7 implies that

$$\underline{P}_{KL}(A|B) = \frac{1}{3}\alpha_1, \text{ for } |A| = 1,$$
$$\underline{P}_{KL}(A|B) = \frac{2}{3}\alpha_2, \text{ for } |A| = 2,$$

where $\alpha_1 = 0.2899$ and $\alpha_2 = 0.5704$ follow by solving Equation (14). Also, from Equation (13) we obtain $\beta_1 = 1.355$ and $\beta_2 = 1.859$.

Then, the probability measures Q_1 and Q_2 given by

$$Q_1(\{x_i\}) = \begin{cases} \alpha_1 P^*_0(\{x_i\}) = 0.0967 & \text{if } x_i \in A_1. \\ \beta_1 P^*_0(\{x_i\}) = 0.4517 & \text{if } x_i \notin A_1. \end{cases}$$
$$Q_2(\{x_i\}) = \begin{cases} \alpha_2 P^*_0(\{x_i\}) = 0.1901 & \text{if } x_i \in A_2. \\ \beta_2 P^*_0(\{x_i\}) = 0.6197 & \text{if } x_i \notin A_2. \end{cases}$$

for $|A_1| = 1$ and $|A_2| = 2$ satisfy $Q_1(A_1) = \underline{P}_{KL}(A_1 \mid B)$ and $Q_2(A_2) = \underline{P}_{KL}(A_2 \mid B)$. These probability measures must satisfy

$$D_{KL}(Q_1, P^*_0) = D_{KL}(Q_2, P^*_0) = \delta^*,$$

but:

$$D_{KL}(Q_1, P^*_0) = \frac{\alpha_1}{3}\log(\alpha_1) + \frac{2}{3}\beta_1 \log(\beta_1) = 0.1548.$$
$$D_{KL}(Q_2, P^*_0) = \alpha_2 \frac{2}{3}\log(\alpha_1) + \frac{\beta_2}{3}\log(\beta_2) = 0.1708.$$

We conclude that the updated model $\underline{P}_{KL}(\cdot \mid B)$ is not a Kullback-Leibler model. ♦

5. Comparison of the Distortion Models

In this section we compare the Euclidean and Kullback-Leibler models investigated in the previous sections with other well-known distortion models: the pari mutuel, linear vacuous, total variation and constant odds ratio models. The comparison is done with respect to the amount of imprecision of the models for a fixed δ and the properties of the associated coherent lower prevision.

5.1. Amount of Imprecision

We say that a distortion model $B^{\delta}_{d_1}(P_0)$ is more imprecise than other model $B^{\delta}_{d_2}(P_0)$ if d_1 induces a greater ball: $B^{\delta}_{d_1}(P_0) \supseteq B^{\delta}_{d_2}(P_0)$. Among the TV, PMM, LV and COR models, it is known that the TV is more imprecise than the other three models and the COR model is the less imprecise, while the PMM and the LV are not comparable [23, Sec.5.1].

First of all, we compare the amount of imprecision of the Euclidean and TV models. In general, they are not related. The reason is that for $n > 4$, there is not a general connection between the values $(n - |A|)|A|$ and n. Indeed:

$$(n - |A|)|A| > n \text{ if } |A| = 3.$$
$$(n - |A|)|A| < n \text{ if } |A| = 1.$$

For $n > 4$, the events A with $|A| = 1$ satisfy:

$$\underline{P}_{TV}(A) = P_0(A) - \delta < P_0(A) - \delta\sqrt{\frac{n-1}{n}} = \underline{P}_E(A),$$

while events A with $|A| = 3$ satisfy:

$$\underline{P}_{TV}(A) = P_0(A) - \delta > P_0(A) - \delta\sqrt{\frac{3(n-3)}{n}} = \underline{P}_E(A).$$

Therefore, there is not a connection between $B^{\delta}_{E}(P_0)$ and $B^{\delta}_{TV}(P_0)$.

The next example shows that there is not a connection between the Euclidean and the PMM and LV models either.

Example 5 *Consider a three element possibility space $\mathcal{X}$ and the probability measure $P_0 \in \mathbb{P}^*(\mathcal{X})$ given by:*

$$P_0(\{x_1\}) = P_0(\{x_2\}) = 0.1, \quad P_0(\{x_3\}) = 0.8.$$

Also, take δ small enough such that $B^\delta_E(P_0)$, $B^\delta_{LV}(P_0)$ and $B^\delta_{PMM}(P_0)$ are included in $\mathbb{P}^*(\mathcal{X})$. It holds that:

$$\underline{P}_{LV}(\{x_3\}) = P_0(\{x_3\}) - \delta P_0(\{x_3\}) = P_0(\{x_3\}) - \delta 0.8$$
$$> P_0(\{x_3\}) - \delta\sqrt{\frac{2}{3}} = \underline{P}_E(\{x_3\}).$$
$$\underline{P}_{LV}(\{x_1\}) = P_0(\{x_1\}) - \delta P_0(\{x_1\}) = P_0(\{x_1\}) - \delta 0.1$$
$$< P_0(\{x_1\}) - \delta\sqrt{\frac{2}{3}} = \underline{P}_E(\{x_1\}).$$

In a similar manner:

$$\underline{P}_{PMM}(\{x_3\}) = (1+\delta)P_0(\{x_3\}) - \delta = P_0(\{x_3\}) - 0.2\delta$$
$$> P_0(\{x_3\}) - \delta\sqrt{\frac{2}{3}} = \underline{P}_E(\{x_3\}).$$
$$\underline{P}_{PMM}(\{x_1\}) = (1+\delta)P_0(\{x_1\}) - \delta = P_0(\{x_1\}) - 0.9\delta$$
$$< P_0(\{x_1\}) - \delta\sqrt{\frac{2}{3}} = \underline{P}_E(\{x_1\}).$$

We conclude that there is not a connection between the Euclidean model and the PMM or LV models. ♦

In contrast, the Euclidean model is always more imprecise than the COR model.

Proposition 11 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_E(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ and $B^\delta_{COR}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Then, $B^\delta_{COR}(P_0) \subseteq B^\delta_E(P_0)$.*

Concerning the Kullback-Leibler model, the next result shows that this model is always more imprecise than the PMM and LV models.

Proposition 12 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, $B^\delta_{PMM}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ and $B^\delta_{LV}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Then, $B^\delta_{PMM}(P_0) \subseteq B^\delta_{KL}(P_0)$ and $B^\delta_{LV}(P_0) \subseteq B^\delta_{KL}(P_0)$.*

Next, we compare the two models we are analysing in this paper: the Kullback-Leibler and Euclidean models.

Proposition 13 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ and $B^\delta_E(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Then $B^\delta_E(P_0) \subseteq B^\delta_{KL}(P_0)$.*

Finally, it only remains to see whether there is a connection between the Kullback-Leibler and the TV models. The next example shows that, in general, such connection does not exist.

Example 6 *Consider a three element possibility space $\mathcal{X} = \{x_1, x_2, x_3\}$, the probability measure P_0 given by the probability mass function $P_0 = (0.1, 0.3, 0.6)$ and the distortion parameter $\delta = 0.099$. It follows that:*

$$\underline{P}_{TV}(\{x_1\}) = 0.1 - \delta = 0.001 < 0.00114 = \underline{P}_{KL}(\{x_1\}),$$

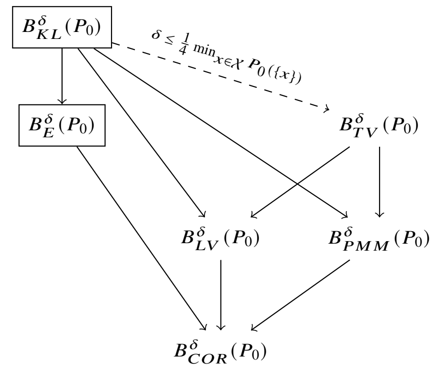

Figure 3: Connection between the different models in terms of imprecision, where the models investigated in this paper have been drawn with a box around.

$$\underline{P}_{TV}(\{x_2\}) = 0.3 - \delta = 0.201 > 0.1136 = \underline{P}_{KL}(\{x_2\}),$$

hence there is not a dominance relation between $\underline{P}_{TV}$ and $\underline{P}_{KL}$, so none of the Kullback-Leibler and TV models is more imprecise than the other. ♦

Even if there is not a general connection between these two models, some connections between the Kullback-Leibler divergence and the TV distance are known:

$$2d^2_{TV}(P, P_0) \leq D_{KL}(P, P_0) \leq \frac{4d^2_{TV}(P, P_0)}{\min_{x\in\mathcal{X}} P_0(x)}$$

where the former inequality follows from the Pinsker's inequality and the latter follows from [15, Lemma 4.1]. Using the second inequality we can prove that, when the distortion parameter is small enough, the Kullback-Leibler model is more imprecise than the TV model.

Proposition 14 *Consider $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $\delta > 0$ such that $B^\delta_{KL}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ and $B^\delta_{TV}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. If $\delta \leq \frac{1}{4}\min_{x\in\mathcal{X}} P_0(\{x\})$, then $B^\delta_{TV}(P_0) \subseteq B^\delta_{KL}(P_0)$.*

Figure 3 summarises the connections between all the models. In this figure, a continuous arrow between two nodes means that the parent is more imprecise than the children (i.e., an inclusion between the balls) and a missed arrow means that there is not a general connection between both models. The connection between the Kullback-Leibler and the TV models proven in Proposition 14 for small distortion parameters haven been indicated with a dashed line.

We conclude this subsection showing in Figure 4 the graphical representation of all the models for the probability measure $P_0 = (0.5, 0.3, 0.2)$ and the distortion parameter $\delta = 0.1$ from Examples 1 and 3. For this particular example, the Kullback-Leibler is the most imprecise model, followed

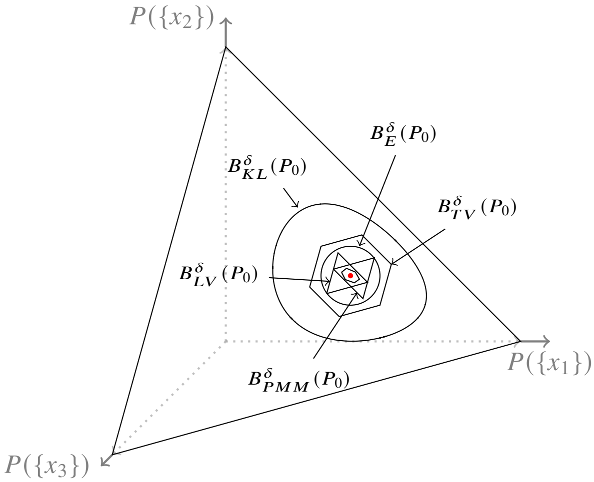

Figure 4: Graphical representation of the PMM, LV, COR (the smallest credal set), TV and the Euclidean and Kullback-Leibler models for the data from Examples 1 and 3.

by the TV and Euclidean models. Indeed, even if δ does not satisfy the sufficient condition in Proposition 14, the Kullback-Leibler model is more imprecise than the TV model, meaning that the condition in Proposition 14 is sufficient but not necessary. Also, although there is not a general connection between the Euclidean model and the TV, LV and PMM models, in this example the TV model is more imprecise than the Euclidean model, which is more imprecise than both the PMM and LV models.

5.2. Practical Properties of the Models

One interesting property a coherent lower prevision may satisfy is that of 2-monotonicity, as was deeply discussed for example in [11, Sec.2.3]. We have seen that the coherent lower prevision of none the Euclidean ($\underline{P}_E$) or Kullback-Leibler models ($\underline{P}_{KL}$) is 2-monotone. Nevertheless, we have seen in Propositions 5 and 9 that they do satisfy 2-monotonicity on events.

In any case, both the Euclidean and Kullback-Leibler models are not 2-monotone, in contrast with the PMM, LV and TV models which do satisfy 2-monotonicity. In particular, the lack of 2-monotonicity means that the Euclidean and Kullback-Leibler models are neither determined by the restriction to singletons (i.e., they are not probability intervals [10, 35]) nor completely monotone, meaning that its Möbius inverse given by:

$$m_{\underline{P}}(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \underline{P}(B)$$

is non-negative for any $A \subseteq \mathcal{X}$. One may wonder whether the restriction to events of $\underline{P}_E$ and $\underline{P}_{KL}$ satisfy these two properties or not. In the next example, we answer this question.

Example 7 *Consider a four element possibility space $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and a uniform probability measure P_0 ($P_0(\{x_i\}) = \frac{1}{4}$ for $i = 1, 2, 3, 4$). Taking $\delta = 0.1$, the (approximate values of the) lower and upper probabilities of the Euclidean model are given by:*

$\|A\|$	1	2	3	4
$\underline{P}_E(A)$	0.1634	0.4	0.6634	1
$\overline{P}_E(A)$	0.3366	0.6	0.8366	1

If it was a probability interval, according to [10, Prop.6] it should happen that:

$$\underline{P}_E(\{x_1, x_2\}) = \{\underline{P}_E(\{x_1\}) + \underline{P}_E(\{x_1\}), \\ 1 - \overline{P}_E(\{x_3\}) + \overline{P}_E(\{x_4\})\} = 0.3267949,$$

which does not coincide with the value $\underline{P}_E(\{x_1, x_2\}) = 0.4$. Moreover, it is neither completely monotone because its Möbius inverse is negative in some events:

$$\begin{aligned} m_{\underline{P}_E}(\{x_1, x_2, x_3\}) &= \underline{P}_E(\{x_1, x_2, x_3\}) - \underline{P}_E(\{x_1, x_2\}) \\ &- \underline{P}_E(\{x_1, x_3\}) - \underline{P}_E(\{x_2, x_3\}) \\ &+ \underline{P}_E(\{x_1\}) + \underline{P}_E(\{x_2\}) + \underline{P}_E(\{x_3\}) \\ &= 0.6634 - 3 \cdot 0.4 + 3 \cdot 0.1634 = -0.0464 < 0. \end{aligned}$$

Not surprisingly, the KL model neither satisfies the properties of complete monotonicity or being a probability interval. Consider the same example, where we obtain the following values:

$\|A\|$	1	2	3	4
$\underline{P}_{KL}(A)$	0.07765	0.28021	0.54333	1
$\overline{P}_{KL}(A)$	0.45667	0.71979	0.92235	1

If it was a probability interval, it should satisfy the following:

$$\underline{P}_{KL}(\{x_1, x_2\}) = \max\{\underline{P}_{KL}(\{x_1\}) + \underline{P}_{KL}(\{x_1\}), \\ 1 - \overline{P}_{KL}(\{x_3\}) + \overline{P}_{KL}(\{x_4\})\} = 0.155303,$$

which does not coincide with the value of $\underline{P}_{KL}(\{x_1, x_2\})$ reported in the table. Moreover, its Möbius inverse may be negative:

$$\begin{aligned} m_{\underline{P}_{KL}}(\{x_1, x_2, x_3\}) &= \underline{P}_{KL}(\{x_1, x_2, x_3\}) - \underline{P}_{KL}(\{x_1, x_2\}) \\ &- \underline{P}_{KL}(\{x_1, x_3\}) - \underline{P}_{KL}(\{x_2, x_3\}) \\ &+ \underline{P}_{KL}(\{x_1\}) + \underline{P}_{KL}(\{x_2\}) + \underline{P}_{KL}(\{x_3\}) \\ &= 0.54333 - 3 \cdot 0.28021 + 3 \cdot 0.07765 \\ &= -0.064337 < 0. \blacklozenge \end{aligned}$$

With this example we conclude that the restriction to events of $\underline{P}_E$ and $\underline{P}_{KL}$, in spite of being 2-monotone, are neither probability intervals nor completely monotone.

Another important feature is the complexity of the credal set: those credal sets with finitely many extreme points are easier to handle than those with infinitely many extreme points. Both the Euclidean and KL-models have infinite extreme points, in contrast with the other four models which have a finitely many number of extreme points.

Finally, other important property that some distortion models satisfy is that of being closed under conditioning. This is the case of the PMM, LV, TV and COR models. However, neither the Euclidean nor the Kullback-Leibler models satisfy this property, as Examples 2 and 4 show, respectively.

On the whole, it seems that from the practical point of view, the Euclidean and Kullback-Leibler models do not satisfy as good properties as the other models. We refer to [23, Tables 3, 4 and 5] for a summary of the properties of the other models.

6. Conclusions

This paper aims at analysing the neighbourhood or distortion models induced by the Euclidean distance and the Kullback-Leibler divergence. The choice of these models is based on the following facts (i) the Euclidean distance is a reasonable choice as a distance between n-dimensional vectors, as those determined by probability mass functions in a n-element possibility space; and (ii) the Kullback-Leibler divergence has been argued to be a more adequate function to compare probability measures than distances between probability mass functions, and it has already been used in the imprecise framework [14, 28].

Since both the Euclidean distance and the Kullback-Leibler divergences are continuous and convex, it follows from [22, Prop.3.1] that the sets of probability measures they induce are closed and convex, meaning that they are equivalent to their associated coherent lower prevision.

In this paper we have explored the main properties of these two models. On the one hand, for the Euclidean model we have given an explicit formula for the associated coherent lower prevision, we have seen that the extreme points of the ball coincide with its boundary and we have proven that its coherent lower prevision is 2-monotone on events but not on gambles. On the other hand, for the Kullback-Leibler model we have also seen that the extreme points of the ball are the probability measures on the boundary of the ball and that its associated coherent lower prevision is 2-monotone on events but not in gambles. In contrast with the Euclidean model, it does not seem easy to obtain an explicit expression of the lower prevision associated with the Kullback-Leibler model; instead, we can express it as the solution of a convex optimisation problem. Nevertheless, we were able to give an expression of its restriction to events.

Finally, we have performed a comparative analysis of the Euclidean and Kullback-Leibler models with the LV, PMM, COR and TV models in terms of the amount of imprecision and the properties of the associated coherent lower previsions. In light of the obtained results, it seems that the lack of 2-monotonicity (on gambles) of the Euclidean and Kullback-Leibler models make them more difficult to handle than the other models, all of them 2-monotone except the COR. Also, the Euclidean and Kullback-Leibler models are not closed under conditioning by means of the regular extension, in contrast to the LV, PMM, COR and TV models. Moreover, after the comparative analysis in terms of imprecision performed in Section 5.1, the Kullback-Leibler model adds, probably, too imprecision.

Consequently, it seems reasonable to conclude that, from the practical viewpoint, the Euclidean and Kullback-Leibler models are less adequate than other models such as the LV, PMM or TV models, which do satisfy the 2-monotonicity property (hence they can make use of all the practical advantages of this property [11]), and that are preserved under conditioning by means of the regular extension (property also satisfied by the COR model). Notwithstanding, we do *not* aim at concluding that the Euclidean or Kullback-Leibler models are useless or that they should not be used in the imprecise setting. The utility of the Euclidean distance and the Kullback-Leibler divergence in a probability setting is unquestionable, and even the Kullback-Leibler divergence has already been used with sets of probability measures [14, 28]. Instead, we claim that the practical properties of these two models make them more difficult to use in practice than other neighbourhood or distortion models.

Besides our comparative analysis given in Section 5, it would be interesting to complement our study in [12] investigating the behaviour of the Euclidean and Kullback-Leibler models under different processing operations: aggregation (intersection, union or convex mixtures), marginalisation or construction of a joint model from marginal information under some (in)dependence assumption. In this respect, some preliminary thoughts suggest that their behaviour under these operations will not be satisfactory either. Moreover, other distortion models deserve some attention, such as the one induced by the Wasserstein distance [18, 32].

Acknowledgments

I want to thank Enrique Miranda for the productive discussions on the topic of distortion models, his suggestions and his carefull reading of this manuscript. I also appreciate the suggestions put forward by the reviewers after carefully reviewing the paper. Undoubtedly, their comments helped me to improve the manuscript.

References

[1] P.R. Alonso-Martín, I. Montes, and E. Miranda. Distortion models for estimating human error probabilities. *Safety Science*, 105915, 2023.

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

[3] A. Benavoli and M. Zaffalon. Density-ratio robustness in dynamic state estimation. *Mechanical Systems and Signal Processing*, 37(1–2):54–75, 2013.

[4] B. Bigi. Using kullback-leibler distance for text categorization. In F. Sebastiani, editor, *Advances in Information Retrieval*, pages 305–319. Springer Berlin Heidelberg, 2003.

[5] A. Bronevich. On the closure of families of fuzzy measures under eventwise aggregations. *Fuzzy Sets and Systems*, 153:45–70, 2005.

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

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

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

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

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

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

[12] S. Destercke, I. Montes, and E.. Miranda. Processing distortion models: A comparative study. *International Journal of Approximate Reasoning*, 145C:91–120, 2022.

[13] S. Eguchi and J. Copas. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, 97(9):2034–2040, 2006.

[14] 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 (Allerton)*, pages 115–122, 2010.

[15] F. Götze, H. Sambale, and A. Sinulis. Higher order concentration for functions of weakly dependent random variables. *arXiv:1801.06348*, 2018.

[16] T. Herron, T. Seidenfeld, and L. Wasserman. Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444, 1997.

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

[18] L.V. Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6 (4):366–422, 1939.

[19] E. Miranda. Updating coherent lower previsions on finite spaces. *Fuzzy Sets and Systems*, 160(9):1286–1307, 2009.

[20] E. Miranda and I. Montes. Coherent updating of non-additive measures. *International Journal of Approximate Reasoning*, 56:159–177, 2015.

[21] I. Montes, E. Miranda, and S. Destercke. Pari-mutuel probabilities as an uncertainty model. *Information Sciences*, 481:550 – 573, 2019.

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

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

[24] C. Morais, R. Moura, M. Beer, and E. Patelli. Analysis and estimation of human errors from major accident investigation reports. *ASME J. Risk Uncertainty Part B*, 6(1):011014, 2019.

[25] C. Morais, S. Tolo, R. Moura, M. Beer, and E. Patelli. Tackling the lack of data for human error probability with credal network. In *Proceedings of the ESREL*, 2019.

[26] C. Morais, H. Estrada-Lugo, S. Tolo, T. Jacques, R. Moura, M. Beer, and E. Patelli. Robust data-driven human reliability analysis using credal networks. *Reliability Engineering and System Satefy*, 218(107990), 2022.

[27] C. Morais, K.L. Yung, K. Johnson, R. Moura, M. Beer, and E. Patelli. Identification of human errors and influencing factors: A machine learning approach. *Safety Science*, 146:105528, 2022.

[28] 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.

[29] 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.

[30] L.S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

[31] L. Utkin. A framework for imprecise robust one-class classification models. *Journal of Machine Learning Research and Cybernetics*, 5(3):379–393, 2014.

[32] L.N. Vaserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72, 1969.

[33] P. Walley. Coherent lower (and upper) probabilities. Statistics research report, University of Warwick, 1981.

[34] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[35] K. Weichselberger and S. Pöhlmann. *A Methodology for Uncertainty in Knowledge-Based Systems*, volume 419 of *LNAI*. Springer, 1990.