Proceedings of Machine Learning Research 103:304–313, 2019 ISIPTA 2019

A Unifying Frame for Neighbourhood and Distortion Models

Enrique Miranda MIRANDAENRIQUE@UNIOVI.ES
Ignacio Montes IMONTES@UNIOVI.ES
Statistics and Operations Research, University of Oviedo, Spain

Sébastien Destercke SEBASTIEN.DESTERCKE@HDS.UTC.FR
Sorbonne Universités, Université Technologique de Compiègne, CNRS, UMR 7253 - Heudiasyc, 57 Avenue de Landshut, Compiègne, France

Abstract

Neighbourhoods of precise probabilities are instrumental to perform robustness analysis, as they rely on very few parameters. Many such models, sometimes referred to as *distortion* models, have been proposed in the literature, such as the pari-mutuel model, linear vacuous mixtures or the constant odds ratio model. In this paper, we show that all of them can be represented as probability sets that are neighbourhoods defined over different (pre)-metrics, providing a unified view of such models. We also compare them in terms of a number of properties: precision, number of extreme points, n-monotonicity, . . . thus providing possible guidelines to pick a neighbourhood rather than another.

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

1. Introduction

Modelling of uncertainty about some quantity or about the outcome of an experiment by means of precise probability measures may not always be accurate nor reliable, due to a number of factors such as low-quality data, conflicts between sources of information, etc. As a consequence, being able to perform robustness analysis is essential in critical applications. One way to perform such an analysis is to explore in a principled way neighbourhoods around the precise estimates, by defining what is usually referred to as a *distorted* model. Such approaches have been applied for example in the analysis of graphical models [8, 15], in reinforcement learning [13] or in regression problems [31].

There are two basic procedures to determine this distorted model: the first one is to take the set of probabilities that are close to the original precise one under some criteria. We may for instance consider convex combinations with a set of probability measures where the weights model the amount of conflicting data, assume that there is some bounded error when reporting the values of the probability measure, or simply work with those measures at a given distance from the initial model. In all those cases, we end up with a set of probability measures, usually called *neighbourhood model* [14, 17, 25, 26, 30]. Under some conditions, this set is equivalent to the coherent lower probability that we obtain by taking lower envelopes. The second procedure consists in directly transforming the initial probability measure by means of a function on its set of values [6, 34]. This procedure also produces a lower probability, but the interpretation of this function may be more involved.

By means of these two procedures, several distortion models have been proposed in the literature [2, 16, 33], such as linear vacuous mixtures or the pari mutuel model. However, the problem of choosing an appropriate distortion model for a given situation is still somewhat unresolved. Our goal in this paper is to contribute to its solution by analyzing and comparing a number of distortion models from the literature, by means of: (a) the amount of imprecision they introduce in the original model for a fixed distortion factor; (b) the complexity of the sets of probabilities they determine, in terms of their number of extreme points; and (c) the properties of their associated coherent lower probabilities. We shall focus on neighbourhood models that induce polytopes in the space of probability distributions, meaning for instance that distortion models induced by the Euclidean norm or the Kullback-Leibler distance [13] are out of the scope of the present paper.

After introducing some preliminaries in Section 2, in Section 3 we define the general framework of neighbourhood models. Then, in Section 4, we study a number of prominent examples of distortion models: the pari mutuel, linear vacuous, constant odds ratio, and those associated with the total variation and Kolmogorov distances. A comparative study of these models is carried out in Section 5. Some conclusions are given in Section 6.

2. Preliminary Concepts

From now on, $\mathscr{X} = \{x_1, \ldots, x_n\}$ denotes a finite possibility space with cardinality n, and $\mathbb{P}(\mathscr{X})$ denotes the set of probability measures defined on $\mathscr{P}(\mathscr{X})$. We shall assume that $\mathscr{X}$ is totally ordered: $x_1 < x_2 < \cdots < x_n$.

2.1. Credal Sets, Lower Probabilities, Lower Previsions

A set of probability measures defined on $\mathscr{P}(\mathscr{X})$ is called a *credal set* [18]. By taking the lower envelope over events, a credal set $\mathscr{M}$ determines a *coherent lower probability* $\underline{P}$:

$$\underline{P}(A) = \inf\{P(A) : P \in \mathscr{M}\} \quad \forall A \subseteq \mathscr{X}. \tag{1}$$

If instead of lower envelopes we take upper envelopes we obtain a *coherent upper probability*, that we shall denote $\overline{P}$. Both of them are related by the conjugacy relation $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \mathscr{X}$.

Different credal sets $\mathscr{M}_1, \mathscr{M}_2$ may determine the same coherent lower probability $\underline{P}$ by using Eq. (1). The largest of them is denoted by $\mathscr{M}(\underline{P})$ and it is given by:

$$\mathscr{M}(\underline{P}) = \{P \in \mathbb{P}(\mathscr{X}) : P(A) \geq \underline{P}(A)\ \forall A \subseteq \mathscr{X}\}.$$

This set is usually referred to as the *credal set associated with* $\underline{P}$, and it is closed in the weak-* topology, that coincides with the Euclidean one in the finite-dimensional case we are considering in this paper. As a consequence, we obtain $\underline{P}(A) = \min\{P(A) : P \in \mathscr{M}(\underline{P})\}$ for every $A \subseteq \mathscr{X}$.

Also, since every probability measure P on $\mathscr{P}(\mathscr{X})$ is equivalent to its expectation operator, that for simplicity we denote also with the symbol P, we can use credal sets to obtain lower and upper expectation operators: given a credal set $\mathscr{M}$, we get $\underline{P}(f) := \inf\{P(f) : P \in \mathscr{M}\}$ and $\overline{P}(f) := \sup\{P(f) : P \in \mathscr{M}\}$. These two functionals are *coherent lower and upper previsions* in the sense of Walley, and are related by $\underline{P}(f) = -\overline{P}(-f)$ for any $f : \mathscr{X} \to \mathbb{R}$.

We see then that a credal set $\mathscr{M}$ can be used to determine both a coherent lower probability $\underline{P}$ and a coherent lower prevision $\underline{P}'$. However, the sets $\{P : P(A) \geq \underline{P}(A)\ \forall A \subseteq \mathscr{X}\}$ and $\{P : P(f) \geq \underline{P}'(f)\ \forall f : \mathscr{X} \to \mathbb{R}\}$ do not coincide in general, the second being smaller, or more informative.

2.2. K-Monotonicity

Some coherent lower probabilities satisfy mathematical properties that make them interesting from a practical standpoint, such as k-monotonicity [7]. Given a natural number $k \geq 2$, a coherent lower probability is *k-monotone* [7] iff

$$\underline{P}(\cup_{i=1}^{p} A_i) \geq \sum_{i=1}^{p} \underline{P}(A_i) - \sum_{i \neq j} \underline{P}(A_i \cap A_j) + \cdots + (-1)^p \underline{P}(\cap_{i=1}^{p} A_i) \tag{2}$$

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

We focus here on two extreme cases: if $k = 2$ we say that $\underline{P}$ is a *2-monotone* lower probability; and if $\underline{P}$ is k-monotone for every natural number k, we say that it is *completely monotone*, or a *belief function* [27]. One interesting property of 2-monotone lower probabilities is that the extreme points of their associated credal set are in correspondence with the permutations of the possibility space [28].

Also, if $\underline{P}'$ denotes a coherent lower prevision, it is called *2-monotone* [32] when:

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

for every $f, g : \mathscr{X} \to \mathbb{R}$, where $\wedge$ and $\vee$ denote the pointwise minimum and maximum. The restriction to events of a 2-monotone lower prevision is a 2-monotone lower probability. Conversely, if $\underline{P}$ is a 2-monotone lower probability, then it has a unique extension to gambles preserving the 2-monotonicity: its natural extension, that is given by the Choquet integral [10].

2.3. Probability Intervals

A particular instance of 2-monotone lower probabilities are the *probability intervals* [9]. A coherent lower probability is a probability interval if $\mathscr{M}(\underline{P})$ can be expressed as:

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

If $\underline{P}$ is a probability interval and $\overline{P}$ denotes its conjugate upper probability, it holds that [9]:

$$\underline{P}(A) = \max\left\{\sum_{x \in A} \underline{P}(\{x\}), 1 - \sum_{x \notin A} \overline{P}(\{x\})\right\} \quad \forall A \subseteq \mathscr{X}.$$

In general, any coherent lower probability $\underline{P}$ and its conjugate upper probability $\overline{P}$ define a probability interval l, u by considering their restriction to singletons: $\underline{P}(\{x\})$ and $\overline{P}(\{x\})$ for every $x \in \mathscr{X}$. In that case, the credal set induced by these bounds is more imprecise than $\mathscr{M}(\underline{P})$, in the sense that is included in it, being equal if and only if $\underline{P}$ is itself a probability interval.

2.4. P-Boxes

Another particular case of 2-monotone lower probability (indeed, completely monotone), is the lower envelope of the credal set associated with *p-boxes* [12]. If we consider two distribution functions $\underline{F}, \overline{F} : \mathscr{X} \to [0,1]$ on the totally ordered space $\mathscr{X}$, with $\underline{F} \leq \overline{F}$, and the credal set

$$\mathscr{M}(\underline{F}, \overline{F}) = \{P \in \mathbb{P}(\mathscr{X}) : \underline{F}(x) \leq F_P(x) \leq \overline{F}(x)\ \forall x \in \mathscr{X}\},$$

then the lower envelope of this set produces a coherent lower probability that is completely monotone [29].

3. Distortion Models

Given a function $d : \mathbb{P}(\mathscr{X}) \times \mathbb{P}(\mathscr{X}) \to [0,1]$, a probability measure $P_0 \in \mathbb{P}(\mathscr{X})$ and a parameter $\delta > 0$, we can consider the set

$$\mathscr{M}(P_0, d, \delta) = \{P \in \mathbb{P}(X) \mid d(P, P_0) \leq \delta\}$$

of those probability measures that differ at most in δ from P_0. We call it the *distorted* model on P_0 associated with the distorting function d and the factor $\delta > 0$. It can be used to determine the coherent lower probability

$$\underline{P}(A) = \inf\{P(A) \mid d(P,P_0) \le \delta\} \quad \forall A \subseteq \mathscr{X}.$$

For the sake of simplicity, we will assume throughout the paper that $P_0(A) > 0$ for every non-empty $A \subseteq \mathscr{X}$. In the remainder of this paper, we investigate in detail the properties of several distortion models.

Before we tackle this problem, we should mention that sometimes in the literature [6] the term *distorted* model is referred to a non-additive measure of the type $\underline{P} = f(P_0)$, where $f : [0,1] \to [0,1]$ is a suitable distortion function that is increasing and satisfies $f(0) = 0, f(1) = 1$. In this sense, if f is convex [7, 11], then $\underline{P} = f(P_0)$ is 2-monotone. Some other results in this direction can be found in [4, 5], in connection with the theory of aggregation operators; see also [34]. For example, in [4, Thm. 7] it is shown that the condition $f(t) \le t$ guarantees that $\underline{P}$ is monotone. Next we prove that such models can be incorporated into our formalism:

Proposition 1 *Let $P \in \mathbb{P}(\mathscr{X})$ be a probability measure. Consider an increasing function $f : [0,1] \to [0,1]$ such that $f(0) = 0, f(1) = 1$ and $f(t) \le t$ for every $t \in [0,1]$, and let $\underline{P} = f(P)$. Then there exists a premetric d (i.e., a non-negative function d satisfying $d(P,P) = 0$ for every $P \in \mathbb{P}(\mathscr{X})$) and some $\delta > 0$ such that:*

$$\mathscr{M}(\underline{P}) = \{Q \in \mathbb{P}(\mathscr{X}) \mid d(P,Q) \le \delta\}.$$

4. Examples of Distortion Models

Throughout this section we consider a number of known neighbourhood models known in the literature, such as the pari mutuel, the linear vacuous and the constant odds ratio. Furthermore, we also consider the neighbourhood models induced by the total variation and Kolmogorov distances.

4.1. Pari Mutuel Model

Our analysis begins with the *pari mutuel model*.

Definition 2 *Given a probability distribution P_0 and a distortion parameter $\delta > 0$, the associated* pari mutuel model *(PMM) is given by the conjugate coherent lower and upper probabilities:*

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

This model has its origins as a betting scheme in horse racing. There, δ is interpreted as a taxation from the house, so $\underline{P}_{PMM}, \overline{P}_{PMM}$ can be given a behavioural interpretation as betting rates for and against the event A, as discussed in depth in [33, Sec. 2.9.3]. The credal set $\mathscr{M}(\underline{P}_{PMM})$ they determine is given by ([22, Corollary 1]):

$$\{P \in \mathbb{P}(\mathscr{X}) \mid P(\{x\}) \le (1+\delta)P_0(\{x\}) \; \forall x \in \mathscr{X}\}.$$

For a detailed study of the PMM from the point of view of imprecise probabilities, we refer to [22, 24, 33]. In particular, it was established in [22, Thm. 1] that the PMM is a particular case of probability interval, meaning that it is determined by its restriction to singletons. This implies that the coherent lower probability $\underline{P}_{PMM}$ is 2-monotone.

Let us show that $\underline{P}_{PMM}, \overline{P}_{PMM}$ can be seen as distortion models, in the manner we introduced in Section 3.

Theorem 3 *Consider the PMM associated with a probability measure P_0 and a distortion factor $\delta > 0$. Then $\mathscr{M}(\underline{P}_{PMM}) = \mathscr{M}(P_0, d_{PMM}, \delta)$, where d_{PMM} is given by*

$$d_{PMM}(P,Q) = \max_{A: Q(A)<1} \frac{Q(A) - P(A)}{1 - Q(A)} \quad \forall P,Q \in \mathbb{P}(\mathscr{X}).$$

In the case of the pari mutuel model, the distortion factor δ is not always attained in the credal set $\mathscr{M}(\underline{P}_{PMM})$, meaning that $d(P,P_0) \le \delta$ for every $P \in \mathscr{M}(\underline{P}_{PMM})$. The next result characterizes when the upper bound δ is attained.

Proposition 4 *Let $\mathscr{M}(P_0, d_{PMM}, \delta)$ be the neighbourhood PMM determined by a probability measure P_0 and a distortion factor $\delta > 0$. Then*

$$\sup_{P \in \mathscr{M}(P_0, d_{PMM}, \delta)} d_{PMM}(P,P_0) = \delta \Leftrightarrow \delta \le \max_{A \subsetneq \mathscr{X}} \frac{P_0(A)}{1 - P_0(A)}.$$

In practice, this means that choosing any value δ higher than $\max_{A \subsetneq \mathscr{X}} \frac{P_0(A)}{1-P_0(A)}$ would be useless. Note that due to the monotonicity of P_0, the maximal value can be obtained by testing the n events $\mathscr{X} \setminus \{x\}$ for all $x \in \mathscr{X}$.

4.2. Linear Vacuous Model

Another relevant class of neighbourhood models is that of *linear-vacuous* mixtures, also referred to as ε-*contamination* models in the literature [16, 33].

Definition 5 *Given a probability distribution P_0 and a distortion parameter $\delta \in (0,1)$, the associated* linear vacuous mixture *(LV) is given by the conjugate coherent lower and upper probabilities:*

$$\underline{P}_{LV}(\mathscr{X}) = 1, \quad \underline{P}_{LV}(A) = (1-\delta)P_0(A) \quad \forall A \neq \mathscr{X}.$$
$$\overline{P}_{LV}(\emptyset) = 0, \quad \overline{P}_{LV}(A) = (1-\delta)P_0(A) + \delta \quad \forall A \neq \emptyset.$$

The associated credal set is given by:

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

The lower probability $\underline{P}_{LV}$ associated with a linear-vacuous model is completely monotone, because it is a convex combination of two completely monotone lower probabilities:

P_0 and the vacuous lower probability given by $\underline{P}(A) = 0\ \forall A \neq \mathscr{X}$, $\underline{P}(\mathscr{X}) = 1$. In addition, it is a probability interval: if $P(\{x\}) \geq \underline{P}_{LV}(\{x\}) = (1-\delta)P_0(\{x\})$, then by additivity we conclude that $P(A) \geq (1-\delta)P_0(A) = \underline{P}_{LV}(A)$.

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

Theorem 6 *Consider the linear vacuous mixture associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then $\mathscr{M}(\underline{P}_{LV}) = \mathscr{M}(P_0, d_{LV}, \delta)$, where d_{LV} is given by*

$$d_{LV}(P,Q) = \max_{A:Q(A)>0} \frac{Q(A)-P(A)}{Q(A)} \quad \forall P,Q \in \mathbb{P}(\mathscr{X}).$$

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

Proposition 7 *Let $\mathscr{M}(P_0, d_{LV}, \delta)$ be the neighbourhood linear-vacuous model determined by a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then*

$$\sup_{P \in \mathscr{M}(P_0, d_{LV}, \delta)} d_{LV}(P,P_0) = \delta.$$

4.3. Constant Odds Ratio

Next we consider the constant odds ratio model. It was given a behavioural interpretation by Peter Walley in [33, Sec. 2.9.4], and studied in [1, 2, 25, 30].

Definition 8 *Given a precise probability P_0 and a distortion parameter $\delta \in (0,1)$, the associated* constant odds ratio *model is given by the coherent lower prevision that is the unique solution to the equation:*

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

where $g^+ = \max\{g,0\}$ and $g^- = \max\{-g,0\}$. Its conjugate coherent upper prevision $\overline{P}_{COR}(f)$ is the unique solution of the equation:

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

As was explained in [33], the restriction to events of $\underline{P}_{COR}, \overline{P}_{COR}$, denoted by $\underline{Q}_{COR}, \overline{Q}_{COR}$, has a simpler expression. For every $A \subseteq \mathscr{X}$:

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

The reason why in this case we are first considering the coherent lower prevision instead of the coherent lower probability that is its restriction to events is that the constant odds ratio model is not 2-monotone in general, as we show in Example 1 later on. As a consequence, its value on gambles is not uniquely determined by the restriction to events. We will therefore make a separate study of $\underline{P}_{COR}$ and its restriction to events $\underline{Q}_{COR}$.

Constant odds ratio on gambles We start with $\underline{P}_{COR}$ in this section. The following proposition summarizes some first properties of this model:

Proposition 9 ([33]) *Let $\underline{P}_{COR}$ be the constant odds ratio determined by a probability measure P_0 and a distortion factor $\delta \in (0,1)$.*

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

$$\Big\{P \in \mathbb{P}(\mathscr{X}) \Big| P(A)P_0(B) \geq (1-\delta)P_0(A)P(B) \ \ \forall A,B\Big\}. \tag{3}$$

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

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

We deduce that $\underline{P}_{COR}$ is not 2-monotone in general.

Example 1 *Let us give an example where $\mathscr{M}(\underline{P}_{COR})$ and $\mathscr{M}(\underline{Q}_{COR})$ do not coincide. Consider a three element space $\mathscr{X}$, the probability measure $P_0 = (0.5, 0.3, 0.2)$ and the distortion factor $\delta = 0.2$. Then $\underline{Q}_{COR}$ is given by:*

A	P_0	$\underline{Q}_{COR}$
$\{x_1\}$	0.5	$4/9$
$\{x_2\}$	0.3	$12/47$
$\{x_3\}$	0.2	$1/6$
$\{x_1,x_2\}$	0.8	$16/21$
$\{x_1,x_3\}$	0.7	$28/43$
$\{x_2,x_3\}$	0.5	$4/9$

Hence, $P = (\frac{4}{9}, \frac{20}{63}, \frac{5}{21}) \in \mathscr{M}(\underline{Q}_{COR})$, but $P \notin \mathscr{M}(\underline{P}_{COR})$, because it violates Eq. (3) with $A = \{x_1\}, B = \{x_3\}$. ♦

From Proposition 9, it follows that:

Proposition 10 *Consider the constant odds ratio model associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then $\mathscr{M}(\underline{P}_{COR})$ always has $2^n - 2$ different extreme points.*

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

Theorem 11 *Consider the constant odds ratio model associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then $\mathscr{M}(\underline{P}_{COR}) = \mathscr{M}(P_0, d_{COR}, \delta)$, where for every $P,Q \in \mathbb{P}(\mathscr{X})$, d_{COR} is given by:*

$$d_{COR}(P,Q) = \max_{\substack{A,B \subseteq \mathscr{X} \\ P(B),Q(A)>0}} \left\{1 - \frac{P(A)\cdot Q(B)}{P(B)\cdot Q(A)}\right\}.$$

Finally, we show that all distortion factors $\delta \in (0,1)$ can be attained.

Proposition 12 *Let $\mathscr{M}(P_0, d_{COR}, \delta)$ be the neighbourhood constant odds ratio determined by a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then*

$$\sup_{P \in \mathscr{M}(P_0, d_{COR}, \delta)} d_{COR}(P,P_0) = \delta.$$

Constant odds ratio on events We turn now to the restriction to events of the constant odds ratio model, $\underline{Q}_{COR}$. It is easy to see that $\underline{Q}_{COR}$ can be expressed by $\underline{Q}_{COR} = f(P_0)$, where f is the convex function given by:

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

As a consequence, $\underline{Q}_{COR}$ is a 2-monotone lower probability. Its main properties are given in the following proposition:

Proposition 13 *Let $\underline{Q}_{COR}$ be the 2-monotone lower probability that is the restriction to events of the constant odds ratio model.*

(a) $\underline{Q}_{COR}$ is completely monotone.

(b) The credal set determined by $\underline{Q}_{COR}$ is given by

$$\mathscr{M}(\underline{Q}_{COR}) = \left\{ P : \frac{P(A)}{P(A^c)} \geq (1-\delta)\frac{P_0(A)}{P_0(A^c)} \quad \forall A \subsetneq \mathscr{X} \right\}.$$

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

The maximal number extreme points, $n!$, can be attained for instance with P_0 the uniform distribution and δ small enough ($\delta < \frac{1}{n}$). In that case, the extreme points of $\mathscr{M}(\underline{Q}_{COR})$ are given by:

$$P_\sigma(\{x_{\sigma(k)}\}) = \frac{n(1-\delta)}{(n-k\delta)(n-(k-1)\delta)} \quad \forall k = 1,\ldots,n$$

and for all the permutations σ of $\{1,\ldots,n\}$. Moreover, it is easy to verify that any two permutations induces two different extreme points, whence we have $n!$ extreme points overall.

From Proposition 13 we can prove that the coherent lower probability $\underline{Q}_{COR}$ is a distortion model.

Theorem 14 *Consider the constant odds ratio model on events associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then $\mathscr{M}(\underline{Q}_{COR}) = \mathscr{M}(P_0, d'_{COR}, \delta)$, where for every $P,Q \in \mathbb{P}(\mathscr{X})$ d'_{COR} is given by:*

$$d'_{COR}(P,Q) = \max_{\substack{A \subseteq \mathscr{X} \\ P(A^c), Q(A) > 0}} \left\{ 1 - \frac{P(A)}{P(A^c)} \frac{Q(A^c)}{Q(A)} \right\}.$$

On the other hand, $\underline{Q}_{COR}$ is not a probability interval:

Example 2 *Consider $\mathscr{X} = \{x_1,x_2,x_3,x_4\}$, let P_0 be the uniform probability distribution and consider $\delta = 0.1$. Denoting by l,u the probability interval determined by $\underline{Q}_{COR}, \overline{Q}_{COR}$ we obtain the following values:*

$\lvert A\rvert$	P_0	$\underline{P}_{COR}$	$\overline{P}_{COR}$	l	u
1	1/4	9/39	10/37	9/39	10/37
2	1/2	9/19	10/19	18/39	21/39
3	3/4	27/37	30/39	27/37	30/39
4	1	1	1	1	1

Since $\underline{Q}_{COR}$ and l do not coincide for events of cardinality 2, we deduce that $\underline{Q}_{COR}$ is not a probability interval. ♦

We conclude this section showing that the distortion factor δ is always attained for some $P \in \mathscr{M}(P_0, d'_{COR}, \delta)$.

Proposition 15 *Let $\mathscr{M}(P_0, d'_{COR}, \delta)$ be the neighbourhood constant odds ratio on events determined by a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then*

$$\sup_{P \in \mathscr{M}(P_0, d'_{COR}, \delta)} d'_{COR}(P, P_0) = \delta.$$

4.4. Total Variation Distance

The previous models were proposed without an explicit link to a distance, and part of the current work is to establish such a link. Considering other distorting functions than the ones studied in the previous sections will lead to new neighbourhood models. This is what we will now do. We begin with the total variation. Given two probabilities $P,Q \in \mathbb{P}(\mathscr{X})$, their *total variation* [19] is defined by:

$$d_{TV}(P,Q) = \sup_{A \subset \mathscr{X}} |P(A) - Q(A)|.$$

This motivates the following definition:

Definition 16 *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 $\mathscr{M}(P_0, d_{TV}, \delta)$.*

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

Theorem 17 ([14]) *Let $\underline{P}_{TV}$ be the lower envelope of the credal set $\mathscr{M}(P_0, d_{TV}, \delta)$ induced by the probability measure P_0 and the distortion factor $\delta \in (0,1)$. It holds that:*

$$\begin{aligned} \underline{P}_{TV}(A) &= \max\{0, P_0(A) - \delta\}, \\ \overline{P}_{TV}(A) &= \min\{1, P_0(A) + \delta\} \quad \forall A \subseteq \mathscr{X}. \end{aligned} \tag{4}$$

Next we establish under which conditions the distortion factor δ is informative, that is, when it is attained by some element in the credal set:

Proposition 18 *Let $\mathscr{M}(P_0, d_{TV}, \delta)$ be the neighbourhood model associated with the total variation distance determined by a probability measure P_0 and a distortion factor $\delta \in (0,1)$. Then:*

$$\sup_{P \in \mathscr{M}(P_0, d_{TV}, \delta)} d_{TV}(P, P_0) = \delta \Leftrightarrow \exists A \subsetneq \mathscr{X} \mid \delta \leq P_0(A).$$

Let us study the properties of $\underline{P}_{TV}$ as a non-additive measure. It is not difficult to prove the following:

Proposition 19 *Let $\underline{P}_{TV}$ be the lower envelope of the credal set $\mathscr{M}(P_0, d_{TV}, \delta)$ induced by the probability measure P_0 and the distortion factor $\delta \in (0,1)$. Then, $\underline{P}_{TV}$ is 2-monotone.*

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

Example 3 *Let us continue with Example 2. From Eq. (4) we obtain $\underline{P}_{TV}, \overline{P}_{TV}$ as well as its induced probability interval l, u:*

$\|A\|$	$\underline{P}_{TV}$	$\overline{P}_{TV}$	l	u
1	0.15	0.35	0.15	0.35
2	0.40	0.60	0.30	0.70
3	0.65	0.85	0.65	0.85
4	1	1	1	1

Since $l(A) < \underline{P}_{TV}(A)$ for events A of cardinality 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:

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

while $\underline{P}_{TV}(A_1 \cup A_2 \cup A_3) = 0.60$. Hence, Eq. (2) is not satisfied, so $\underline{P}_{TV}$ is not completely monotone. ♦

To discuss the number of extreme points induced by the total variation model, we will denote by

$$\mathscr{L} := \{A \subseteq \mathscr{X} \mid \underline{P}_{TV}(A) = 0\}$$

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

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

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

Proposition 20 *Let $\mathscr{M}(P_0, d_{TV}, \delta)$ be the neighbourhood model associated with a probability measure P_0 and a distortion factor $\delta \in (0,1)$ by means of the total variation distance. Then the number of extreme points of $\mathscr{M}(P_0, d_{TV}, \delta)$ is $\sum_{A \in \mathscr{L}} s_A$. As a consequence, if $|\mathscr{X}| = n$, the maximal number of extreme points of $\mathscr{M}(P_0, d_{TV}, \delta)$ 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 $\frac{n}{2}$.

4.5. Kolmogorov Distance

We conclude this section by considering the neighbourhood model based on the Kolmogorov distance, that makes a comparison between the distribution functions associated with the probability measures. Given two probabilities $P, Q \in \mathbb{P}(\mathscr{X})$, their *Kolmogorov distance* is defined by:

$$d_K(P, Q) = \sup_{x \in \mathscr{X}} |F_P(x) - F_Q(x)|,$$

where F_P and F_Q denote the cumulative distribution functions associated with P and Q, respectively. This leads to the following definition:

Definition 21 *Let P_0 be a probability measure and consider a distortion factor $\delta \in (0,1)$. Its associated* Kolmogorov model *$\underline{P}_K$ induced by the Kolmogorov distance is the lower envelope of the credal set $\mathscr{M}(P_0, d_K, \delta)$.*

There exists an obvious connection between the credal sets $\mathscr{M}(P_0, d_{TV}, \delta)$ and $\mathscr{M}(P_0, d_K, \delta)$: any $P \in \mathscr{M}(P_0, d_{TV}, \delta)$ satisfies

$$d_{TV}(P, P_0) \leq \delta \Leftrightarrow \sup_{A \subseteq \mathscr{X}} |P(A) - P_0(A)| \leq \delta, \text{ whence}$$
$$|P(\{x_1, \ldots, x_k\}) - P_0(\{x_1, \ldots, x_k\})| \leq \delta,$$

hence $P \in \mathscr{M}(P_0, d_K, \delta)$. This implies that $\mathscr{M}(P_0, d_K, \delta) \supseteq \mathscr{M}(P_0, d_{TV}, \delta)$. However, both sets do not coincide in general, as our next example shows.

Example 4 *Let us continue with Example 2. Consider the probability P given by the mass function $(0.35, 0.05, 0.35, 0.25)$. Then $P \in \mathscr{M}(P_0, d_K, \delta)$, 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
$\|F_P - F_{P_0}\|$	0.1	0.1	0	0

However, P does not belong to $\mathscr{M}(P_0, d_{TV}, \delta)$, because:

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

Thus, $\mathscr{M}(P_0, d_K, \delta) \supsetneq \mathscr{M}(P_0, d_{TV}, \delta)$. ♦

We next establish for which values of δ we can find $P \in \mathscr{M}(P_0, d_K, \delta)$ such that $d_K(P, P_0) = \delta$.

Proposition 22 *Let $\mathscr{M}(P_0, d_K, \delta)$ the neighbourhood model associated with P_0 by means of Kolmogorov distance and a distortion factor $\delta \in (0,1)$. Then, $\sup_{P \in \mathscr{M}(P_0, d_K, \delta)} d_K(P, P_0) = \delta \Leftrightarrow F_{P_0}(x) + \delta \leq 1$ or $F_{P_0}(x) - \delta \geq 0$ for some $x \in \mathscr{X}$.*

Since we can rewrite $\mathscr{M}(P_0, d_K, \delta)$ as:

$$\{P \in \mathbb{P}(\mathscr{X}) \mid \delta \geq |F_P(x) - F_{P_0}(x)|, \ \forall x \in \mathscr{X}\} =$$
$$\{P \in \mathbb{P}(\mathscr{X}) \mid F_{P_0}(x) - \delta \leq F_P(x) \leq F_{P_0}(x) + \delta, \ \forall x \in \mathscr{X}\},$$

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

$$\underline{F}(x) = \max\{0, F_{P_0}(x) - \delta\}, \ \ \overline{F}(x) = \min\{1, F_{P_0}(x) + \delta\}$$

for every $x \in \mathscr{X}$, and it holds that $\mathscr{M}(P_0, d_K, \delta) = \mathscr{M}(\underline{F}, \overline{F})$. This means that $\mathscr{M}(P_0, d_K, \delta) = \mathscr{M}(\underline{P}_K)$ is the credal set associated with a p-box. As a consequence, $\underline{P}_K$ is completely monotone [29], and $\underline{P}_K$ can be expressed as:

$$\underline{P}_K(A) = \inf\{P(A) \mid F_{P_0}(x) - \delta \leq F_P(x) \leq F_{P_0}(x) + \delta, \ \forall x\}$$

for every $A \subseteq \mathscr{X}$. This means that, although $\underline{P}_K$ and $\underline{P}_{TV}$ induce the same p-box, they are not the same coherent lower probability: $\underline{P}_K \lneq \underline{P}_{TV}$. If we follow the notation in [3, 21], we deduce [23, Prop. 15] that $\underline{P}_K$ is the unique undominated outer approximation of $\underline{P}_{TV}$ in terms of p-boxes.

Example 5 *If we consider our running Example 2, we obtain* $\underline{P}_K(\{x_1\}) = \underline{P}_K(\{x_4\}) = 0.15, \underline{P}_K(\{x_2\}) = \underline{P}_K(\{x_3\}) = 0.05, \overline{P}_K(\{x_1\}) = \overline{P}_K(\{x_4\}) = 0.35$ *and* $\overline{P}_K(\{x_2\}) = \overline{P}_K(\{x_3\}) = 0.45$. *Then, if we consider the probability interval* $[l,u]$ *that is the restriction of* $\underline{P}_K, \overline{P}_K$ *to events, taking* $A = \{x_1, x_2\}$, *it holds that:*

$$l(A) = \max\left\{\sum_{x\in A} l(\{x\}), 1 - \sum_{x\notin A} u(\{x\})\right\} = 0.2, \text{ while}$$
$$\underline{P}_K(\{x_1,x_2\}) = \underline{F}(x_2) = \max\{0, F_{P_0}(x_2) - \delta\} = 0.4.$$

Since $\underline{P}_K$ *and* l *do not coincide for the event* $\{x_1,x_2\}$, *we conclude that* $\underline{P}_K$ *is not a probability interval.* ♦

Finally, we investigate the maximal number of extreme points in $\mathscr{M}_K(P_0,\delta)$. It is known [20, 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 $\mathscr{X}$. Pell numbers are recursively defined by:

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

The next result shows that a p-box induced by the Kolmogorov distance can also attain this maximal value.

Proposition 23 *Let* $\mathscr{M}(P_0, d_K, \delta)$ *be the neighbourhood model associated with a probability measure* P_0 *and a distortion factor* $\delta \in (0,1)$ *by means of the Kolmogorov distance. If* $|\mathscr{X}| = n$, *the maximal number of extreme points of* $\mathscr{M}(P_0, d_K, \delta)$ *is* $\mathscr{P}_n$.

This maximal number of extreme points is attained for instance when P_0 is the uniform distribution and $\delta \in \left(\frac{1}{2n}, \frac{1}{n}\right)$.

5. Comparison of the Distortion Models

Next we compare all the models according to: the amount of imprecision introduced by the distortion model, the properties of the associated coherent lower probability, and the number of extreme points of the neighbourhood model.

5.1. Amount of Imprecision

Let us compare the amount of imprecision that is introduced by the different examples of distortion model we have considered in the previous section once the initial probability measure P_0 and the distortion factor $\delta > 0$ are fixed. Given two credal sets $\mathscr{M}_1, \mathscr{M}_2$, $\mathscr{M}_1$ is *more informative* than $\mathscr{M}_2$ when $\mathscr{M}_1 \subseteq \mathscr{M}_2$; in terms of their lower envelopes, this means that $\underline{P}_1(f) \geq \underline{P}_2(f)$ for every $f : \mathscr{X} \to \mathbb{R}$.

Example 6 *Consider again Example 1. The associated distorted models are:*

A	$\underline{P}_{PMM}(A)$	$\underline{P}_{LV}(A)$	$\underline{P}_{COR}(A)$	$\underline{P}_{TV}(A)$	$\underline{P}_K(A)$
$\{x_1\}$	0.4	0.4	0.444	0.3	0.3
$\{x_2\}$	0.16	0.24	0.255	0.1	0
$\{x_3\}$	0.04	0.16	0.166	0	0
$\{x_1,x_2\}$	0.76	0.64	0.761	0.6	0.6
$\{x_1,x_3\}$	0.64	0.56	0.651	0.5	0.3
$\{x_2,x_3\}$	0.4	0.4	0.444	0.3	0.3

Considering the events $A = \{x_2\}$ *and* $B = \{x_1,x_2\}$, *the PMM and the linear vacuous are not comparable, in the sense that none of them is more imprecise than the other.* ♦

It was already stated in [33, 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,

$$\mathscr{M}(\underline{Q}_{COR}) \subseteq \mathscr{M}(\underline{P}_{LV}) \cap \mathscr{M}(\underline{P}_{PMM}).$$

Since $\underline{Q}_{COR}$ is the restriction to events of $\underline{P}_{COR}$, it follows that $\mathscr{M}(\underline{P}_{COR}) \subseteq \mathscr{M}(\underline{Q}_{COR})$. Next we compare these models with the one associated with the total variation.

Proposition 24 *For any probability measure* P_0 *and any distortion factor* $\delta > 0$, *it holds that*

$$\mathscr{M}(\underline{P}_{PMM}) \cup \mathscr{M}(\underline{P}_{LV}) \subseteq \mathscr{M}(\underline{P}_{TV}).$$

If $\underline{P}_{PMM}, \underline{P}_{LV}, \underline{P}_{TV}$ denote their lower envelopes, we can equivalently state that $\underline{P}_{TV} \leq \min\{\underline{P}_{PMM}, \underline{P}_{LV}\}$. Finally, taking into account the comments given in Section 4.5, the model based on the Kolmogorov distance is more imprecise than the total variation distance: $\mathscr{M}(\underline{P}_{TV}) \subseteq \mathscr{M}(\underline{P}_K)$. These relationships are summarized in Figure 1.

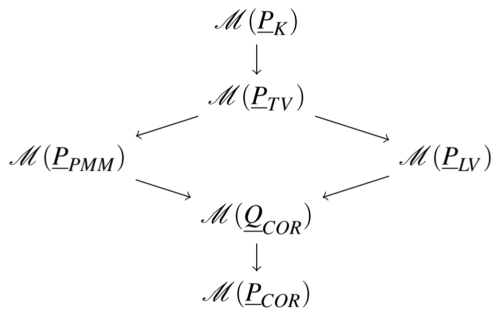

Figure 1: Relationships between the different models. An arrow between two nodes means that parent includes the child.

More generally, as suggested by a reviewer, we may also compare the models by means of imprecision indices or the generalized Hartley measure; this approach is left as a future line of research.

5.2. Properties of the Envelopes of the Neighbourhood Models

We may also compare the different distortion models in terms of the properties of the coherent lower probability they determine. Since all the distorting functions are convex and continuous, it can be shown that the credal set $\mathscr{M}(P_0, d, \delta)$ coincides with the one associated with the coherent lower probability it induces by taking lower envelopes.

As we mentioned in Section 2, 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: it guarantees that the distortion model has a unique extension to gambles [7] and it allows to use a simple formula to compute the extreme points of the credal set [28]. It turns out that the models we have considered in this paper are 2-monotone, except for the constant odds ratio, that only satisfies 2-monotonicity when it is restricted to events.

Two particular cases of 2-monotone lower probabilities are the k-monotone ones and probability intervals. The latter corresponds to those that are uniquely determined by their restrictions to singletons. With respect to the examples considered in this paper, only the pari mutuel model and the linear vacuous mixture satisfy this property. With respect to k-monotonicity for $k > 2$, or, more specifically, complete monotonicity, both the linear vacuous mixture and Kolmogorov's model are completely monotone: the former, because it is a convex combination of two completely monotone models, and the latter because all coherent lower probabilities associated with a p-box are. On the other hand, neither the pari mutuel ([22, Prop.5]) nor the total variation (see Example 3) are 3-monotone, and therefore none of them is completely monotone. Table 1 summarizes these results.

Table 1: Properties of the coherent lower probabilities.

Model	2-monotone	∞-monotone	Prob. interval
$\underline{P}_{PMM}$	YES	NO	YES
$\underline{P}_{LV}$	YES	YES	YES
$\underline{P}_{TV}$	YES	NO	NO
$\underline{P}_{COR}$	NO	NO	NO
$\underline{Q}_{COR}$	YES	YES	NO
$\underline{P}_{K}$	YES	YES	NO

We therefore conclude that, from the point of view of these properties, the most adequate model is the linear vacuous, while the only model which does not satisfy 2-monotonicity is the constant odds ratio (on gambles).

5.3. Complexity

One important feature of a neighbourhood model is that it has a simple representation in terms of a finite number of extreme points. In this respect, it was established in [28] that when its lower envelope is 2-monotone, then there are at most $n!$ different extreme points, and these are related to the permutations of the possibility space. 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 [35]; 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 [22] and [33], respectively. Here, we have computed the maximum number of extreme points also for the neighbourhood models $\mathscr{M}(P_0, d_{TV}, \delta)$, $\mathscr{M}(P_0, d_K, \delta)$, $\mathscr{M}(P_0, d_{COR}, \delta)$ and $\mathscr{M}(P_0, d'_{COR}, \delta)$. Table 2 summarizes the results.

Table 2: Comparison of the number of extreme points.

Model	Maximal number of extreme points
$\underline{P}_{PMM}$	$\frac{n!}{\lfloor\frac{n}{2}\rfloor(\lfloor\frac{n}{2}\rfloor-1)!(n-\lfloor\frac{n}{2}\rfloor-1)!}$
$\underline{P}_{LV}$	n
$\underline{P}_{TV}$	$\frac{n!}{(\lfloor\frac{n}{2}\rfloor-1)!(n-\lfloor\frac{n}{2}\rfloor-1)!}$
$\underline{P}_{COR}$	2^n-2
$\underline{Q}_{COR}$	$n!$
$\underline{P}_{K}$	$\mathscr{P}_n$

We observe that the the simplest model is the linear vacuous, followed (for $n \geq 6$) by the constant odds ratio, the Kolmogorov model, the total variation model and the pari mutuel (that have the same maximal number of extreme points), 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.

6. Conclusions

This paper introduces a general framework of neighbourhoods and distorted models. We have seen that some of the well-known neighbourhoods models from the literature, such as the pari mutuel model, the linear vacuous mixture or the constant odds ratio can be embedded into our general framework, as well as some models defined using the total variation or the Kolmogorov distances.

From our results, we see that the smallest distortion from the original model is done when we consider the constant odds ratio, while the largest one is done by the Kolmogorov model. In terms of the properties of the coherent lower probability, the best one is the linear vacuous model, that is the only one that is simultaneously a belief function and a probability interval. It is also the one whose associated credal set has the smallest number of extreme points.

We have also characterized under which cases the distortion model is a probability interval. It would also be interesting to give necessary and sufficient conditions, in terms of the distance, for the associated distorted model to be 2-monotone or a belief function. 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, for instance in terms of their behaviour under conditioning and combination, as well as the study of other neighbourhood models associated with other distances.

Acknowledgments

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 projects PGC2018-098623-B-I00 and GRUPIN/IDI/2018/000176.

References

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

[2] J. Berger. Robust bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25:303–328, 1990.

[3] A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2009)*, pages 61–70, 2009.

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

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

[6] A. Chateauneuf. Decomposable capacities, distorted probabilities and concave capacities. *Mathematical Social Sciences*, 31:19–37, 1996.

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

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

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

[10] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n-Monotone exact functionals. *Journal of Mathematical Analysis and Applications*, 347:143–156, 2008.

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

[12] 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, 2003.

[13] S. Filippi, O. Cappé, and A. Garivier. Optimism in reinforcement learning and Kullback-Leibler divergence. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference*, pages 115–122. IEEE, 2010.

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

[15] M. Hourbracq, C. Baudrit, P.-H. Wuillemin, and S. Destercke. Dynamic credal networks: introduction and use in robustness analysis. In *Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2013)*, pages 159–169, 2013.

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

[17] P. J. Huber and V. Strassen. Minimax tests and the Neyman–Pearson lemma for capacities. *The Annals of Statistics*, 1:251–263, 1973.

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

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

[20] I. Montes and S. Destercke. On extreme points of p-boxes and belief functions. *Annals of Mathematics and Artificial Intelligence*, 81:405–428, 2017.

[21] I. Montes, E. Miranda, and P. Vicig. 2-monotone outer approximations of coherent lower probabilities. *International Journal of Approximate Reasoning*, 101: 181–205, 2018.

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

[23] 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.

[24] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51: 1145–1158, 2010.

[25] L. R. Pericchi and P. Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 59:1–23, 1991.

[26] H. Rieder. Least favourable pairs for special capacities. *Annals of Statistics*, 5(5):909–921, 1997.

[27] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.

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

[29] 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.

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

[31] L. Utkin and A. Wiencierz. An imprecise boosting-like approach to regression. In *Proceedings of the Eight International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'2013)*, pages 345–354, 2013.

[32] P. Walley. Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry, 1981.

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

[34] A. Wallner. Bi-elastic neighbourhood models. In *Proceedings of the Eight International Symposium on Imprecise Probability: Theory and Applications (ISIPTA 2003)*, pages 593–607, 2003.

[35] A. Wallner. Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning*, 44(3):339–357, 2007.