Distortions of imprecise probabilities ⋆

David Nieto-Barba[0009−0001−2551−9605], Ignacio Montes[0000−0001−6534−1613], and Enrique Miranda[0000−0001−7763−3779]

University of Oviedo, Dep. of Statistics and Operations Research
davidphbarba@outlook.com, {imontes,mirandaenrique}@uniovi.es

Abstract. We generalise the idea of distortion models to the case where the starting model is an imprecise probability model instead of a precise probability. Specifically, we discuss the transformation of a lower probability or a credal set into a more imprecise model, and analyse a number of desirable properties any such transformation should satisfy. Then, we investigate in detail the extension of the total variation distortion model from this perspective.

Keywords: Distortion models · coherent lower probabilities · credal sets · 2-monotonicity · total variation distance.

1 Introduction

In situations of imprecise or ambiguous information, when we have missing data or when we should aggregate the different opinions of several experts, it may be sensible to consider an alternative to probability measures as a model of the uncertainty associated with an experiment. In the past decades, this idea has given rise to a number of *imprecise probability* models, such as coherent lower probabilities and previsions [21], belief functions [18] or possibility measures [7], just to name a few.

One context where imprecise probabilities arise naturally is when we consider distortion models. By this, we mean a model where a precise probability measure is transformed into an imprecise one by means of some distorting function d and considering some distortion factor δ that may represent the degree of robustness we are aiming for. Depending on the procedure, several different models have been proposed in the literature, such as the ϵ-*contamination* [10], *pari-mutuel* [21] or *total variation* models [9]. On the other hand, the term *distortion* also appeared in [3] as a direct transformation of a probability measure expressing a human misperception of a probability measure; a unified study of the main approaches can be found in [15, 16].

There are however scenarios in which it may be sensible to have an imprecise probability model as a starting point: for instance, it may be that the available information does not allow to come up with a precise model, and yet we may

⋆ This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "FEDER/UE".

want to robustify it; it could also be that the model suffers from some inconsistencies that we want to correct by making it more imprecise; or we could also want to transform our model into another that is more operational, while at the same time not removing any precise probability measure that has been deemed possible in our original formulation. In any of those cases, we would like to have a procedure that transforms our model into a more imprecise one, and where we are also able to quantify the degree of distortion we are introducing.

We should mention here that the idea of distorting an imprecise model is not new. Indeed, it appeared under the terminology of *discounting* in the field of Evidence Theory [18], where a belief function is combined with a vacuous model. Following this terminology, Serafín Moral [17] discussed the problem of distorting sets of probabilities from an axiomatic point of view, and analysed some instances, among which is the total variation model. Our focus here will be instead on lower probabilities and we shall also consider other rationality axioms. A somewhat closer approach to ours would be that of Sébastien Destercke [6], where a discounting rule for lower probabilities is defined, inspired by a rule established in [13] in the context of the transferable belief model.

The paper is organised as follows. After recalling some elementary notions about imprecise probabilities in Sect. 2, we analyse two avenues for this problem: the direct distortion of the lower probability or the aggregation of the distortions of the elements of its associated credal set. These possibilities and a number of desirable properties we may impose are discussed in Sect. 3. Our analysis is exemplified on the extension of the total variation distortion model in Sect. 4. Finally, we give some concluding remarks and additional insights in Sect. 5.

2 Preliminary concepts

Let us give the basics of the theory of imprecise probabilities we shall use in the paper; we refer to [1] for a more detailed introduction.

Let $\mathcal{X}$ be a finite possibility space. Any function $\underline{P} : \mathcal{P}(\mathcal{X}) \to [0, 1]$ that is monotone ($A \subseteq B \Rightarrow \underline{P}(A) \leq \underline{P}(B)$) and normalised ($\underline{P}(\emptyset) = 0$ and $\underline{P}(\mathcal{X}) = 1$) is called a *capacity* (see for example [8]). In this paper, we shall also call it *lower probability*, because the value $\underline{P}(A)$ for an event A may be interpreted as a lower bound for the probability of A under a 'true', but unknown, probability measure P_0. Under this interpretation, we may consider the set of probability measures compatible with $\underline{P}$, given by

$$\mathcal{M}(\underline{P}) := \{P \in \mathbb{P}(\mathcal{X}) : P(A) \geq \underline{P}(A) \ \forall A \subseteq \mathcal{X}\}, \tag{1}$$

where $\mathbb{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$. Then we say that a lower probability $\underline{P}$ *avoids sure loss* when the set $\mathcal{M}(\underline{P})$ it determines by means of Eq. (1) is non-empty, and that it is *coherent* when

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

The set $\mathcal{M}(\underline{P})$ is a closed and convex set of probability measures, and as a consequence it is a *credal set* in the sense of Levi [11].

There are a number of particular cases of coherent lower probabilities that are of interest. We have for instance the 2-*monotone* ones, that satisfy

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

the k-monotone ones satisfy, for any $p \leq k$ and $A_1, \ldots, A_p \subseteq \mathcal{X}$, the inequality

$$\underline{P}\left(\bigcup_{i=1}^{p} A_i\right) \geq \sum_{I \subseteq \{1, \ldots p\}} (-1)^{|I|+1} \underline{P}\left(\bigcap_{i \in I} A_i\right)$$

and the *minitive* ones, also known as *necessity measures* within possibility theory [7], that satisfy $\underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}$ for any $A, B \subseteq \mathcal{X}$. It turns out that minitive lower probabilities are k-monotone for any $k \in \mathbb{N}$, and these in fact satisfy 2-monotonicity. On the other hand, 2-monotone lower probabilities also satisfy the rationality condition of coherence.

More generally, we can consider lower *previsions*; these give lower expectations to real valued functions $f : \mathcal{X} \to \mathbb{R}$, called *gambles*. An instance of gambles are the indicator functions I_A of events $A \subseteq \mathcal{X}$, which take the value 1 on the elements of A and 0 elsewhere. We denote by $\mathcal{L}(\mathcal{X})$ the set of all gambles on $\mathcal{X}$.

A *lower prevision* on $\mathcal{L}(\mathcal{X})$ is a function $\underline{P} : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$. We say that it is coherent if and only if

$$\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} E_P(f) \quad \forall f \in \mathcal{L}(\mathcal{X}), \qquad \text{where}$$
$$\mathcal{M}(\underline{P}) := \left\{P \in \mathbb{P}(\mathcal{X}) : E_P(f) \geq \underline{P}(f) \ \forall f \in \mathcal{L}(\mathcal{X})\right\}$$

and E_P denotes the expectation operator with respect to P. While a probability measure has a unique extension as an expectation operator, this is not the case for coherent lower probabilities: two different coherent lower previsions may have the same restriction to indicators of events. One property that guarantees uniqueness is 2-monotonicity: if the coherent lower prevision satisfies

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

for any $f, g \in \mathcal{L}(\mathcal{X})$, then it can be computed as the Choquet integral with respect to its restriction to events [4].

3 Distortions of imprecise models

When robustifying a probability measure into an imprecise probability model, two approaches can be followed. As discussed in [15, 16], a *distortion model* transforms a probability measure P_0 into a lower probability $\underline{P}$ by means of two elements: a distortion factor $\delta > 0$ and a distorting measure d that compares probability measures, such as a distance or a divergence. With these tools, the distorted lower probability may be obtained as the lower envelope of the neighbourhood

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

i.e., as $\underline{Q}(A) = \inf\{P(A) : P \in B_d^\delta(P_0)\}$. Alternatively, the distorted lower probability may be obtained by directly applying a function on P_0. Given a distortion parameter $\delta > 0$ and some function $f_\delta : [0,1] \to [0,1]$ that is monotone and satisfies $f_\delta(0) = 0, f_\delta(1) = 1$, the lower probability would be defined as $\underline{P}(A) = f_\delta(P_0(A))$ for any $A \subseteq \mathcal{X}$. This procedure was introduced in [3] for representing the imprecise observation of a probability measure and it has been investigated in [2, 5].

For any distortion procedure for probability measures, we may consider its generalisation towards lower probabilities $\underline{P}$. We may follow here two avenues:

(a) The first of them would be to distort each of the probability measures in $\mathcal{M}(\underline{P})$, and to consider the lower envelope of the set of lower probabilities thus obtained; note however that this method is only applicable when $\mathcal{M}(\underline{P}) \neq \emptyset$, i.e., when $\underline{P}$ avoids sure loss. Moreover, when $\underline{P}$ avoids sure loss and it is not coherent, the procedure will not distinguish between the distortion of a lower probability $\underline{P}$ and its *natural extension* $\underline{E}$, which is the coherent lower probability that is the lower envelope of $\mathcal{M}(\underline{P})$.
(b) For this reason, it may be interesting also to consider a distortion procedure that applies directly on a lower probability $\underline{P}$, irrespective on the properties it satisfies. In this sense, if the distortion of a probability measure P_0 produces the lower probability $f_\delta(P_0)$, the idea would be to consider $f_\delta(\underline{P})$ as the lower probability that on the event A takes the value $f_\delta(\underline{P}(A))$.

Whichever the approach, there are a number of desirable properties that our distortion procedure may satisfy. For this, let $\underline{Q}_\delta(\underline{P})$ denote the distorted model that our procedure determines if we start from a distortion factor $\delta > 0$ and a lower probability $\underline{P}$. We consider the following:

P1.Expansion: Given $\delta_1 > \delta_2 > 0$, $\underline{Q}_{\delta_1}(\underline{P}) \leq \underline{Q}_{\delta_2}(\underline{P})$.
P2.Aggregation: For any $\delta_1, \delta_2 > 0$, $\underline{Q}_{\delta_1+\delta_2}(\underline{P}) = \underline{Q}_{\delta_1}\big(\underline{Q}_{\delta_2}(\underline{P})\big)$.
P3.Structure preservation: If $\underline{P}$ is coherent (resp., 2-monotone, k-monotone, minitive) so is $\underline{Q}_\delta(\underline{P})$ for every $\delta > 0$.
P4.Commutativity: If $\underline{P}$ is coherent, then for any $\delta > 0$

$$\mathcal{M}\big(\underline{Q}_\delta(\underline{P})\big) = \bigcup_{P \in \mathcal{M}(\underline{P})} \mathcal{M}\big(\underline{Q}_\delta(P)\big),$$

where $\underline{Q}_\delta(P)$ refers to the particular case in which the distortion procedure is applied to a probability measure P.

The idea of expansion has already been discussed by Moral in [17] in the context of discounting of credal sets, while the notion of structure preservation is also present in the work of Destercke [6]. On the other hand, the idea of commutativity corresponds to what Destercke calls *lower probability preservation*. In what follows, we shall investigate these properties in more detail in an example of a distortion procedure: the generalization of the total variation model.

In the following, and when no confusion is possible, we shall simplify the notation using $\underline{Q}$ to denote the lower probability $\underline{Q}_\delta(\underline{P})$.

4 The imprecise total variation

Next, we apply the ideas put forward in the previous section to analyse the extension of the total variation distortion model to the imprecise case. Recall that, given $P, Q \in \mathbb{P}(\mathcal{X})$, their *total variation distance* [12] is given by

$$d_{\mathrm{TV}}(P,Q) = \max_{A \subseteq \mathcal{X}} |P(A) - Q(A)|.$$

Using this distance, a probability measure P and a distortion factor $\delta > 0$, the lower envelope of the ball $B^{\delta}_{d_{\mathrm{TV}}}(P)$ coincides with the credal set $\mathcal{M}\big(\underline{Q}_P\big)$, where $\underline{Q}_P$ is given by [9]

$$\underline{Q}_P(A) = \max\{P(A) - \delta, 0\} \quad \forall A \subset \mathcal{X}, \quad \underline{Q}_P(\mathcal{X}) = 1. \tag{2}$$

We refer to [9] and [16, Sect.2] for a thorough study of this model.

4.1 First approach: distortion of the lower probability

The total variation model can be straightforwardly extended towards arbitrary lower probabilities.

Definition 1. *Let $\underline{P}$ be a coherent lower probability and $\delta > 0$. The* total variation model *induced by $(\underline{P}, \delta)$ is the lower probability $\underline{Q}$ given by*

$$\underline{Q}(A) = \max\big\{\underline{P}(A) - \delta, 0\big\} \quad \forall A \subset \mathcal{X}, \quad \underline{Q}(\mathcal{X}) = 1. \tag{3}$$

Clearly, this definition is a generalisation of that in Eq. (2). Interestingly, it has appeared, in the context of coalitional game theory, as the *strong δ-core* [19]. We refer to [8, 14, 20] for more information about the connections between game theory and imprecise probabilities.

Let us study this distortion model from the point of view of the properties discussed in Sect. 3. In this respect, it is clear that Eq. (3) complies with expansion (P1). Concerning aggregation, note that for any $A \subset \mathcal{X}$ it holds that

$$\max\{\underline{P}(A) - \delta_1 - \delta_2, 0\} = \max\{\max\{\underline{P}(A) - \delta_2, 0\} - \delta_1, 0\},$$

whence (P2) holds.

With respect to structure preservation (P3), we have the following result.

Proposition 1. *Let $\underline{P}$ be a coherent lower probability, $\delta > 0$ and let $\underline{Q}$ be the lower probability they induce by Eq. (3). If $\underline{P}$ has any of the following properties:*

(i) avoiding sure loss,
(ii) coherence,
(iii) 2-monotonicity,
(iv) minitivity,

so does $\underline{Q}$.

Proof. (i) Since by construction $\mathcal{M}(\underline{P}) \subseteq \mathcal{M}(\underline{Q})$, if $\underline{P}$ avoids sure loss then $\mathcal{M}(\underline{P}) \neq \emptyset$, whence $\mathcal{M}(\underline{Q}) \neq \emptyset$ and therefore $\underline{Q}$ avoids sure loss as well.

(ii) Assume that $\underline{P}$ is coherent. For any $P \in \mathcal{M}(\underline{P})$, let $\underline{Q}_P$ denote the coherent lower probability it induces by means of Eq. (2). It holds that $\mathcal{M}(\underline{Q}_P) \subseteq \mathcal{M}(\underline{Q})$, given that $\underline{Q}_P \geq \underline{Q}$ by construction.
Given $A^* \subset \mathcal{X}$, we aim to prove that there exists $Q \in \mathcal{M}(\underline{Q})$ such that $Q(A^*) = \underline{Q}(A^*)$. By coherence of $\underline{P}$, there exists $P^* \in \mathcal{M}(\underline{P})$ such that $P^*(A^*) = \underline{P}(A^*)$. Since $\underline{Q}_{P^*}$ is coherent, there exists $Q \in \mathcal{M}\big(\underline{Q}_{P^*}\big) \subseteq \mathcal{M}\big(\underline{Q}\big)$ such that $Q(A^*) = \underline{Q}_{P^*}(A^*)$. Hence,

$$\underline{Q}(A^*) = \max\{0, \underline{P}(A^*) - \delta\} = \max\{0, P^*(A^*) - \delta\} = \underline{Q}_{P^*}(A^*) = Q(A^*).$$

We conclude that $\underline{Q}$ is coherent.

(iii) Assume that $\underline{P}$ is 2-monotone, meaning that for any $A, B \subset \mathcal{X}$ it satisfies $(\underline{P}(A \cap B) - \delta) + (\underline{P}(A \cup B) - \delta) \geq (\underline{P}(A) - \delta) + (\underline{P}(B) - \delta)$. Let us show that $\underline{Q}$ is 2-monotone as well. We have the following cases:

Case 1 Assume that $\underline{P}(A \cup B) - \delta \leq 0$. By monotonicity, $\underline{P}(A \cap B) - \delta \leq 0$, $\underline{P}(A) - \delta \leq 0$ and $\underline{P}(B) - \delta \leq 0$. Hence, $\underline{Q}(A \cap B) + \underline{Q}(A \cup B) = 0 = \underline{Q}(A) + \underline{Q}(B)$.

Case 2 If $\underline{P}(A \cap B) - \delta > 0$, it follows that $\underline{P}(A \cup B) - \delta > 0$, $\underline{P}(A) - \delta > 0$ and $\underline{P}(B) - \delta > 0$, whence

$$\begin{aligned}\underline{Q}(A \cap B) + \underline{Q}(A \cup B) &= \big(\underline{P}(A \cap B) - \delta\big) + \big(\underline{P}(A \cup B) - \delta\big) \\ &\geq \big(\underline{P}(A) - \delta\big) + \big(\underline{P}(B) - \delta\big) = \underline{Q}(A) + \underline{Q}(B).\end{aligned}$$

Case 3 Finally, assume that $\underline{P}(A \cap B) - \delta \leq 0$ and $\underline{P}(A \cup B) - \delta > 0$. Then

$$\underline{Q}(A \cap B) + \underline{Q}(A \cup B) = \underline{P}(A \cup B) - \delta.$$

If $\underline{P}(B) - \delta \leq 0$ (respectively, $\underline{P}(A) - \delta \leq 0$), then $\underline{P}(A \cup B) - \delta \geq \underline{P}(A) - \delta$ (respectively, $\underline{P}(B) - \delta$) by monotonicity, whence

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

When, instead, both $\underline{P}(B) - \delta$ and $\underline{P}(A) - \delta$ are greater than or equal to zero, the 2-monotonicity of $\underline{P}$ yields:

$$\begin{aligned}\underline{Q}(A \cap B) + \underline{Q}(A \cup B) &= \underline{P}(A \cup B) - \delta \\ &\geq \big(\underline{P}(A) - \delta\big) + \big(\underline{P}(B) - \delta\big) - \big(\underline{P}(A \cap B) - \delta\big) \\ &\geq \big(\underline{P}(A) - \delta\big) + \big(\underline{P}(B) - \delta\big) = \underline{Q}(A) + \underline{Q}(B).\end{aligned}$$

(iv) Let $\underline{P}$ be minitive, so that $\underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}$ for every $A, B \subset \mathcal{X}$. Given $A, B \subset \mathcal{X}$, if $\delta > 0$ is such that $\underline{P}(A \cap B) - \delta \geq 0$, then

$$\underline{Q}(A \cap B) = \underline{P}(A \cap B) - \delta = \min\{\underline{P}(A) - \delta, \underline{P}(B) - \delta\} = \min\{\underline{Q}(A), \underline{Q}(B)\}.$$

When, instead, $\underline{P}(A \cap B) - \delta \leq 0$, then either $\underline{P}(A) - \delta \leq 0$ or $\underline{P}(B) - \delta \leq 0$, whence $\underline{Q}(A) = 0$ or $\underline{Q}(B) = 0$ and as a consequence $\underline{Q}(A \cap B) = 0 = \min\{\underline{Q}(A), \underline{Q}(B)\}$. From this we conclude that $\underline{Q}$ is minitive as well. □

In spite of this positive result, not all properties of $\underline{P}$ hold onto $\underline{Q}$. Observe for instance that, when $\underline{P}$ is a precise probability measure (that is in particular k-monotone for all k), the lower probability it determines need not be k-monotone for $k > 2$, as shown in [16, Ex. 2.1]. See also [2, Prop. 5] for other interesting comments in this respect.

To see whether the procedure in Definition 1 complies with commutativity (P4), we next investigate the approach based on distorting each element in the credal set.

4.2 Second approach: distortion of the elements in the credal set

The extension of the total variation model towards lower probabilities may also be approached in terms of the distortion of credal sets, in the following manner: for any $P \in \mathcal{M}(\underline{P})$ we may consider the credal set $\mathcal{M}(\underline{Q}_P)$, where $\underline{Q}_P$ is given by Eq. (2). The dominance $P \geq \underline{P}$ leads immediately to the inclusion $\mathcal{M}(\underline{Q}_P) \subseteq \mathcal{M}(\underline{Q})$, whence $\cup_{P\in\mathcal{M}(\underline{P})}\mathcal{M}(\underline{Q}_P) \subseteq \mathcal{M}(\underline{Q})$. If we denote by $\underline{Q}'$ the coherent lower probability defined as[1]

$$\underline{Q}'(A) = \min\left\{Q(A) : Q \in \bigcup_{P\in\mathcal{M}(\underline{P})} \mathcal{M}(\underline{Q}_P)\right\} \quad \forall A \subseteq \mathcal{X}, \tag{4}$$

it follows that $\underline{Q}' \geq \underline{Q}$. It is not difficult to show that these two lower probabilities coincide on events.

Proposition 2. *Let $\underline{P}$ be a coherent lower probability, $\delta > 0$ and $\underline{Q}, \underline{Q}'$ the lower probabilities induced by Eqs.* (3) *and* (4)*. Then $\underline{Q}' = \underline{Q}$.*

Proof. The equality is trivial for $A = \mathcal{X}$, so let us fix an event $A \subset \mathcal{X}$. Since $\underline{P}$ is coherent, there exists some $P^* \in \mathcal{M}(\underline{P})$ such that $P^*(A) = \underline{P}(A)$. By construction $\underline{Q}' \leq \underline{Q}_{P^*}$, hence

$$\underline{Q}'(A) \leq \underline{Q}_{P^*}(A) = \max\{P^*(A) - \delta, 0\} = \max\{\underline{P}(A) - \delta, 0\} = \underline{Q}(A).$$

Since on the other hand $\underline{Q}' \geq \underline{Q}$, we deduce that they are equal. □

The set $\cup_{P\in\mathcal{M}(\underline{P})}\mathcal{M}(\underline{Q}_P)$ corresponds to the discounting credal set from the total variation distance proposed in [17]; interestingly, it is established in [17, Thm. 4.1] a connection with the distortion of *sets of almost-desirable gambles*, which are another imprecise probability model we are not considering in this paper.

We can more succinctly express $\cup_{P\in\mathcal{M}(\underline{P})}\mathcal{M}(\underline{Q}_P)$ by defining a suitable (pre)metric extending the TV distance. Indeed, for any two lower probabilities $\underline{P}, \underline{Q}$ avoiding sure loss, we may define $d_{\text{TV}}^{\min}$ as the minimum of the total

[1] That this is indeed a minimum and not an infimum holds because $\bigcup_{P\in\mathcal{M}(\underline{P})} \mathcal{M}(\underline{Q}_P)$ is closed, as we shall establish in Proposition 3 later on.

variation distances between the probabilities in their respective credal sets:

$$d_{\mathrm{TV}}^{\min}(\underline{P},\underline{Q}) = \min_{\substack{P\in\mathcal{M}(\underline{P})\\ Q\in\mathcal{M}(\underline{Q})}} d_{\mathrm{TV}}(P,Q) = \min_{\substack{P\in\mathcal{M}(\underline{P})\\ Q\in\mathcal{M}(\underline{Q})}} \max_{A\subseteq\mathcal{X}} |P(A)-Q(A)|.$$

Note that $d_{\mathrm{TV}}^{\min}$ is *not* a distance: for instance, it will be $d_{\mathrm{TV}}^{\min}(\underline{P},\underline{Q})=0$ as soon as $\mathcal{M}(\underline{P})\cap\mathcal{M}(\underline{Q})\neq\emptyset$, and it does not satisfy the triangle inequality in general. Indeed, it can be checked that it is only a premetric.

Given the lower probability $\underline{P}$ and $\delta>0$, $d_{\mathrm{TV}}^{\min}$ determines the neighbourhood

$$\begin{aligned} B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P}) &= \{Q\in\mathbb{P}(\mathcal{X}) : d_{\mathrm{TV}}^{\min}(Q,\underline{P})\leq\delta\} \\ &= \{Q\in\mathbb{P}(\mathcal{X}) : \exists P\in\mathcal{M}(\underline{P}) \text{ such that } d_{\mathrm{TV}}(Q,P)\leq\delta\}. \end{aligned} \tag{5}$$

It is not difficult to prove that $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ coincides with $\cup_{P\in\mathcal{M}(\underline{P})}\mathcal{M}(\underline{Q}_P)$. Let us establish that this set is closed and convex:

Proposition 3. *Let $\underline{P}$ be a coherent lower probability and $\delta>0$ a distortion factor. The ball $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ given by Eq.* (5) *is closed and convex.*

Proof. We begin by showing that $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ is closed. Consider a sequence $(Q_n)_n\subset B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ such that $(Q_n)_n\to Q$ for some $Q\in\mathbb{P}(\mathcal{X})$, and let us show that $Q\in B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$. That $(Q_n)_n\subset B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ implies that for every $n\in\mathbb{N}$ there exists $P_n\in\mathcal{M}(\underline{P})$ such that $d_{\mathrm{TV}}(Q_n,P_n)\leq\delta$. The sequence $(P_n)_n$ is included in the compact set $\mathcal{M}(\underline{P})$, which is also sequentially compact; as a consequence, there exists a subsequence $(P_{n_k})_k$ such that $(P_{n_k})_k\to P$ for certain $P\in\mathcal{M}(\underline{P})$. Then:

$$d_{\mathrm{TV}}(P,Q) = d_{\mathrm{TV}}\Big(\lim_{k\to\infty}P_{n_k},\lim_{k\to\infty}Q_{n_k}\Big) = \lim_{k\to\infty} d_{\mathrm{TV}}(P_{n_k},Q_{n_k})\leq\delta,$$

using for the second equality that d_{TV} is continuous. Thus, $Q\in B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$.

To see that $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ is convex, let $Q_1,Q_2\in B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ and $\alpha\in[0,1]$. Then, there exists $P_i\in\mathcal{M}(\underline{P})$ such that $d_{\mathrm{TV}}(P_i,Q_i)\leq\delta$ for $i=1,2$. Since $\mathcal{M}(\underline{P})$ is convex, $\alpha P_1+(1-\alpha)P_2\in\mathcal{M}(\underline{P})$. By the convexity of d_{TV},

$$\begin{aligned} &d_{\mathrm{TV}}(\alpha Q_1+(1-\alpha)Q_2,\alpha P_1+(1-\alpha)P_2) \\ &\qquad\leq \alpha d_{\mathrm{TV}}(P_1,Q_1)+(1-\alpha)d_{\mathrm{TV}}(P_2,Q_2)\leq\delta, \end{aligned}$$

and as a consequence $\alpha Q_1+(1-\alpha)Q_2\in B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$. ◻

Next we show a somewhat surprising result: the two distortion procedures so far introduced do not coincide in general, and $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$ may be a proper subset of $\mathcal{M}(\underline{Q})$. This means that while $\underline{Q}'$ and $\underline{Q}$ agree on events, they will not agree on gambles in general. In other words, the first procedure does not satisfy commutativity. From the practical viewpoint, this means that while both approaches agree on events, the approach based on $d_{\mathrm{TV}}^{\min}$ produces a more informative lower prevision; notwithstanding, $\mathcal{M}(\underline{Q})$ is the credal set of a lower probability, making it easier to handle than $B^{\delta}_{d_{\mathrm{TV}}^{\min}}(\underline{P})$.

Example 1. Let $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, and define the coherent lower probability $\underline{P}$ as the lower envelope of $\{P_1, \dots, P_{17}\}$, where the mass functions of these probability measures are given by

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$	$\{x_6\}$
$P_1(A)$	0.2	0.061	0.061	0.19	0.199	0.289
$P_2(A)$	0.161	0.1	0.1	0.238	0.141	0.26
$P_3(A)$	0.239	0.099	0.12	0.161	0.17	0.211
$P_4(A)$	0.161	0.177	0.13	0.161	0.102	0.269
$P_5(A)$	0.22	0.041	0.109	0.21	0.199	0.221
$P_6(A)$	0.22	0.041	0.041	0.23	0.199	0.269
$P_7(A)$	0.178	0.16	0.111	0.161	0.1	0.29
$P_8(A)$	0.161	0.157	0.14	0.181	0.073	0.288
$P_9(A)$	0.19	0.073	0.078	0.21	0.15	0.299
$P_{10}(A)$	0.161	0.158	0.159	0.161	0.112	0.249
$P_{11}(A)$	0.2	0.071	0.14	0.199	0.199	0.191
$P_{12}(A)$	0.161	0.177	0.138	0.161	0.073	0.29
$P_{13}(A)$	0.239	0.071	0.1	0.189	0.199	0.202
$P_{14}(A)$	0.22	0.041	0.081	0.238	0.199	0.221
$P_{15}(A)$	0.219	0.119	0.061	0.161	0.17	0.27
$P_{16}(A)$	0.218	0.043	0.062	0.238	0.168	0.271
$P_{17}(A)$	0.2	0.0705	0.1585	0.1805	0.199	0.1915

Take now the distortion factor $\delta = 0.011 > 0$ and let $\underline{Q}$ be the distorted model determined by $\underline{P}, \delta$ by means of Eq. (2). Consider also the probability measure Q with probability mass function

$$Q := (0.2, 0.05, 0.05, 0.201, 0.199, 0.3) \in \mathcal{M}(\underline{Q}).$$

Since $Q(\{x_1, x_2\}) = 0.25 < \underline{P}(\{x_1, x_2\}) = 0.261$, we deduce that $Q \in \mathcal{M}(\underline{Q}) \setminus \mathcal{M}(\underline{P})$. To see that there is no $P \in \mathcal{M}(\underline{P})$ such that $d_{\mathrm{TV}}(Q, P) \leq \delta$, observe that $Q(A) = \underline{P}(A) - \delta$ for the events $A = \{x_1, x_2\}$, $\{x_1, x_3\}$, $\{x_2, x_3, x_4\}$, $\{x_1, x_2, x_3, x_4\}$ and $\{x_2, x_3, x_4, x_5\}$. Then, if $P \in \mathcal{M}(\underline{P})$ satisfied $d_{\mathrm{TV}}(Q, P) \leq \delta$ then it should be $P(A) = \underline{P}(A)$ on those events. Thus, it should be

$$P(\{x_1, x_2\}) = P(\{x_1, x_3\}) = 0.261, \; P(\{x_2, x_3, x_4\}) = 0.312$$
$$P(\{x_1, x_2, x_3, x_4\}) = 0.512, \; P(\{x_2, x_3, x_4, x_5\}) = 0.511.$$

But all these conditions imply that P should be given by the mass function $P_1 = (0.2, 0.061, 0.061, 0.19, 0.199, 0.289) \in \mathcal{M}(\underline{P})$, and this probability measure satisfies $d_{\mathrm{TV}}(Q, P_1) = 0.22 > 0.11 = \delta$. Therefore, $Q \notin B^{\delta}_{d^{\min}_{\mathrm{TV}}}(\underline{P})$ and as a consequence $\mathcal{M}(\underline{Q}) \supsetneq B^{\delta}_{d^{\min}_{\mathrm{TV}}}(\underline{P})$. ♦

The two credal sets coincide when the original model satisfies 2-monotonicity.

Proposition 4. *Let $\underline{P}$ be a 2-monotone lower probability, $\delta > 0$ and let $\underline{Q}$ be the lower probability defined by Eq. (3). Then, $\mathcal{M}(\underline{Q}) = B^{\delta}_{d^{\min}_{\mathrm{TV}}}(\underline{P})$.*

Proof. Let $\mathcal{H} \subseteq \mathcal{L}(\mathcal{X})$ be the set of gambles that take values in $[0,1]$. Given a probability measure Q, let us define the map

$$f_Q : \mathcal{M}(\underline{P}) \times \mathcal{H} \to \mathbb{R}$$
$$(P, g) \hookrightarrow P(g) - Q(g).$$

We shall prove that we can apply the minimax theorem in [21, App. E6], so that

$$\min_{\mathcal{M}(\underline{P})} \max_{\mathcal{H}} f_Q(P,g) = \max_{\mathcal{H}} \min_{\mathcal{M}(\underline{P})} f_Q(P,g). \tag{6}$$

For this, it suffices to verify the following conditions:

- $\mathcal{M}(\underline{P}), \mathcal{H}$ are compact convex sets in $\mathbb{R}^n$, where $n = |\mathcal{X}|$.
- For any $g \in \mathcal{H}$ and any $\mu \in \mathbb{R}$ fixed, $\{P \in \mathcal{M}(\underline{P}) : f_Q(P,g) \leq \mu\} = \{P \in \mathcal{M}(\underline{P}) : P(g) \leq Q(g) + \mu\}$ is closed and convex.
- For any $P \in \mathcal{M}(\underline{P})$ and any $\mu \in \mathbb{R}$ fixed, $\{g \in \mathcal{H} : f_Q(P,g) \geq \mu\} = \{g \in \mathcal{H} : P(g) - Q(g) \geq \mu\}$ is closed (since $P - Q$ is continuous on $\mathcal{H}$) and convex (since P, Q are linear).

Thus, the minimax theorem is applicable, and we have the equality in Eq. (6).

Now, given $\delta > 0$,

$$\max_{\mathcal{H}} \min_{\mathcal{M}(\underline{P})} f_Q(P,g) \leq \delta \Leftrightarrow \forall g \in \mathcal{H}\ \exists P \in \mathcal{M}(\underline{P}) \text{ such that } P(g) - Q(g) \leq \delta$$
$$\Leftrightarrow \forall g \in \mathcal{H}, Q(g) \geq \underline{P}(g) - \delta \Rightarrow \forall A \subseteq \mathcal{X}, Q(A) \geq \underline{P}(A) - \delta \Leftrightarrow Q \in \mathcal{M}(\underline{Q}).$$

But since $\underline{P}$ is 2-monotone the inequality $Q(A) \geq \underline{P}(A) - \delta\ \forall A \subseteq \mathcal{X}$ implies that $Q(g) = (C)\int g dQ \geq (C)\int g d\underline{P} - \delta = \underline{P}(g) - \delta$, using that g takes values in $[0,1]$ and the expression of the Choquet integral. Thus,

$$\max_{\mathcal{H}} \min_{\mathcal{M}(\underline{P})} f_Q(P,g) \leq \delta \Leftrightarrow Q \in \mathcal{M}\big(\underline{Q}\big). \tag{7}$$

On the other hand,

$$\min_{\mathcal{M}(\underline{P})} \max_{\mathcal{H}} f_Q(P,g) \leq \delta \Leftrightarrow \exists P \in \mathcal{M}(\underline{P}) \text{ such that } P(g) - Q(g) \leq \delta\ \forall g \in \mathcal{H}$$
$$\Rightarrow \exists P \in \mathcal{M}(\underline{P}) \text{ such that } P(A) - Q(A) \leq \delta\ \forall A \subseteq \mathcal{X}$$
$$\Leftrightarrow \exists P \in \mathcal{M}(\underline{P}) \text{ such that } Q \in \mathcal{M}\big(\underline{Q}_P\big).$$

But by linearity we also deduce that

$$P(A) - Q(A) \leq \delta\ \forall A \subseteq \mathcal{X} \Rightarrow P(g) - Q(g) \leq \delta\ \forall g \in \mathcal{H},$$

using that any $g \in \mathcal{H}$ can be expressed as $g = \sum_{i=1}^{k} x_i I_{A_i}$, for $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k, x_i \geq 0, \sum_{i=1}^{k} x_i \leq 1$ and $1 \leq k \leq n$, and that therefore

$$P(g) - Q(g) = \sum_{i=1}^{k} x_i (P(A_i) - Q(A_i)) \leq \delta \sum_{i=1}^{k} x_i \leq \delta.$$

We conclude then that

$$\min_{\mathcal{M}(\underline{P})} \max_{\mathcal{H}} f_Q(P,g) \leq \delta \Leftrightarrow Q \in \cup_{P \in \mathcal{M}(\underline{P})} \mathcal{M}(\underline{Q}_P) \Leftrightarrow Q \in B^{\delta}_{d^{\min}_{\mathrm{TV}}}(\underline{P}). \tag{8}$$

Putting together Eqs. (6), (7) and (8) we deduce the desired equality. □

5 Conclusions

In this paper we have explored how to distort a coherent lower probability. We have considered two approaches: either directly distorting the lower probability or distorting the elements in the credal set. When the distortion is done by means of the total variation distance, a summary of the results is shown in Figure 1. There we can see that the lower probabilities obtained by both approaches coincide (Prop. 2), but somewhat surprisingly their credal sets only coincide under 2-monotonicity (Prop. 4). This means the lower envelopes of $\mathcal{M}(\underline{Q})$ and $B^{\delta}_{d^{\min}_{TV}}(\underline{P})$ in gambles, i.e. their associated lower previsions, only coincide under 2-monotonicity, while in general we only have the inclusion $B^{\delta}_{d^{\min}_{TV}}(\underline{P}) \subseteq \mathcal{M}(\underline{Q})$.

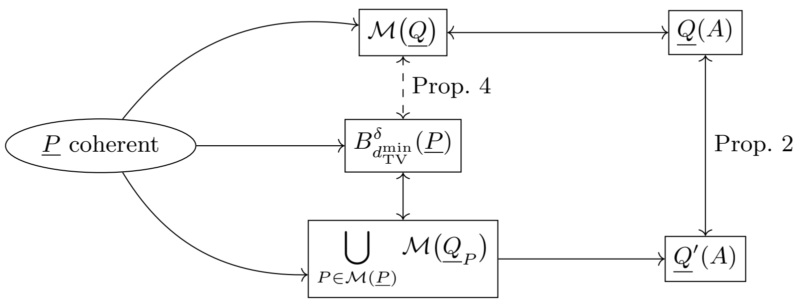

Fig. 1. Connection between the different approaches for distorting a lower probability using the TV-distance. The dashed line expresses the partial correspondence, in this case, under 2-monotonicity.

Even though this paper constitutes our first approximation to the problem, there are other results we have not reported here due to space limitations, as well as some problems still pending. First of all, regarding the properties that a distortion approach may satisfy, we should also consider some other rationality conditions from [6, 17] such as the invariance under permutations of the possibility space $\mathcal{X}$ and under marginalization.

Secondly, our extension $d^{\min}_{TV}$ of the total variation distance to credal sets is not the only possibility; we may instead consider other definitions that are applicable on lower probabilities that are not coherent or that are based on the maximum distance between the credal sets instead of the minimum.

Thirdly, the connection between our definition and the notion of strong δ-core in coalitional game theory leads us naturally towards a distortion model that connects with the *weak* δ*-core*; this would require us to consider a penalised version of the total variation distance.

Finally, our analysis should be completed by comparing the model we have introduced with similar extensions of other distortion models, such as the ϵ-contamination, pari-mutuel or Wasserstein models. This analysis, together with a thorough study of the connection with the results by Moral [17] and Destercke [6], is the main future line of research.

References

1. Augustin, T., Coolen, F., de Cooman, G., Troffaes, M. (eds.): Introduction to Imprecise Probabilities. Wiley Series in Probability and Statistics, Wiley (2014)
2. Bronevich, A.: On the clousure of families of fuzzy measures under eventwise aggregations. Fuzzy Sets and Systems **153**, 45–70 (2005)
3. Chateauneuf, A.: Decomposable capacities, distorted probabilities and concave capacities. Mathematical Social Sciences **31**(1), 19–37 (1996)
4. de Cooman, G., Troffaes, M.C.M., Miranda, E.: n-Monotone exact functionals. Journal of Mathematical Analysis and Applications **347**, 143–156 (2008)
5. Corsato, C., Pelessoni, R., Vicig, P.: Nearly-Linear uncertainty measures. International Journal of Approximate Reasoning **114**, 1–28 (2019)
6. Destercke, S.: A new contextual discounting rule for lower probabilities. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) Proceedings of IPMU'2010, Part II. pp. 198–207. Communications in Computer and Information Science, 81 (2010)
7. Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)
8. Grabisch, M.: Set functions, games and capacities in decision making. Springer (2016)
9. Herron, T., Seidenfeld, T., Wasserman, L.: Divisive conditioning: further results on dilation. Philosophy of Science **64**, 411–444 (1997)
10. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
11. Levi, I.: The enterprise of knowledge. MIT Press, Cambridge (1980)
12. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society (2009)
13. Mercier, D., Quost, B., Denoeux, T.: Refined modeling of sensor reliability in the belief function framework using contextual discounting. Information Fusion **9**, 246–258 (2008)
14. Miranda, E., Montes, I.: Centroids of the core of exact capacities: a comparative study. Annals of Operations Research **321**, 409–449 (2023)
15. Montes, I., Miranda, E., Destercke, S.: Unifying neighbourhood and distortion models: Part I- New results on old models. International Journal of General Systems **49**(6), 602–635 (2020)
16. Montes, I., Miranda, E., Destercke, S.: Unifying neighbourhood and distortion models: Part II- New models and synthesis. International Journal of General Systems **49**(6), 636–674 (2020)
17. Moral, S.: Discounting imprecise probabilities. In: Gil, E., Gil, J., Gil, M. (eds.) The Mathematics of the Uncertain, Studies in Systems, Decision and Control, vol. 142. Springer (2018)
18. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, New Jersey (1976)
19. Shapley, L., Shubik, M.: Quasi-cores in a monetary economy with nonconvex preferences. Econometrica **34**(4), 805–827 (1966)
20. Studený, M., Kroupa, T.: Core-based criterion for extreme supermodular functions. Discrete Applied Mathematics **206**, 122–151 (2016)
21. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)