2-monotonicity and independence: nice to have, hard to keep

Sébastien Destercke

Independence and 2-monotonicity: nice to have, hard to keep

Sebastien Destercke[1]

INRA/CIRAD, UMR1208, 2 place P. Viala, F-34060 Montpellier cedex 1, France
sdestercke@gmail.com

Abstract. When using lower probabilities to model uncertainty about the value assumed by a variable, 2-monotonicity is an interesting property to satisfy, as it greatly facilitates further treatments (such as the computation of lower/upper expectation bounds). In this paper, we show that multivariate joint models induced from marginal ones by strong independence, epistemic independence or epistemic irrelevance do not usually preserve such a property, even if it is satisfied by all marginal models. We then propose a joint model outer-approximating those induced by strong and epistemic independence and study some of its properties.
keywords: factorisation properties, credal sets, propagation, lower previsions.

1 Introduction

In imprecise probability theories where uncertainty is represented by so-called *credal sets* (i.e., convex sets of probabilities), or equivalently by lower expectation bounds (called coherent lower previsions by Walley), independence modeling and tractability are two important issues.

Indeed, the notion of independence plays an essential role in uncertainty theories when dealing with multivariate spaces, its associated factorization properties allowing to decompose a complex problem into simpler ones, or to easily build joint models from marginal ones. When probabilities or expectations are made imprecise, the notion of stochastic independence used in probability theory can be extended in several ways, and such extensions have been proposed and compared by many authors (see, for example, Walley [1] and Couso *et al.* [2]).

On the other hand, tractability is essential in many applications, and although using general uncertainty models is certainly attractive from a theoretical point of view, their complexity often makes them difficult to handle computationally. In practice, tractability can be improved by restricting oneself to classes of uncertainty models that presents a good trade-off between generality and computational convenience. 2-monotone lower probabilities, that encompass many useful uncertainty models (e.g., p-boxes [3], possibility distributions [4], belief functions [5], probability intervals [6]), correspond to such a class, as satisfying the property of 2-monotonicity greatly facilitates the handling of uncertainty in information treatments (e.g., to compute lower and upper expectation bounds). This is why researchers have devoted a lot of attention to such models [7,8].

In this paper, we consider the problem of whether the property of 2-monotonicity is preserved when building a joint representation induced from marginal representations and a strong independence, an epistemic irrelevance or an epistemic independence assumption. After introducing notations and required preliminaries in Section 2, we show in Section 3 that 2-monotonicity is not preserved under none of the assumptions of strong independence, epistemic independence or epistemic irrelevance. In order to solve this issue, we propose in Section 4 an outer approximation by extending the notion of random set independence to 2-monotone lower probabilities. We also study some properties of this approximation, concluding that, while this approximation may be useful in some cases, its usefulness within Walley's theory of imprecise probabilities may be limited. In order to simplify our exposure, we will limit ourselves to the case of two variables, however most presented results readily extend to any number of dimensions.

2 Preliminaries

This section recalls basic notions and introduces main notations used in the rest of the paper. Although we deal with marginal uncertainty models defined by 2-monotone lower probabilities, we will start from lower expectations, as they are needed to express the joint models resulting from different independence assumptions.

2.1 Lower expectations and credal sets

Consider a variable X whose value lies in a finite space $\mathscr{X}$. We assume here that the uncertainty on X is described by a lower expectation (or coherent lower prevision in Walley's terms) $\underline{P}:\mathscr{L}(\mathscr{X})\to\mathbb{R}$ defined over the set $\mathscr{L}(\mathscr{X})$ of all real-valued functions over $\mathscr{X}$. The lower probability of an event $A\subseteq\mathscr{X}$ corresponds to the value $\underline{P}(\mathbf{1}_A)$, where $\mathbf{1}_A$ is the indicator function of A. Here, it will be denoted by $\underline{P}(A)$ when no confusion is possible. From a lower expectation, one can consider the dual notion of upper expectation $\overline{P}$, linked to lower expectation by the relation $\underline{P}(f)=-\overline{P}(-f)$. In the specific case of lower probabilities, the dual notion of upper probability is such that $\underline{P}(A)=1-\overline{P}(\overline{A})$, with $\overline{A}$ the complement of A. A classical expectation operator will be denoted $P:\mathscr{L}(\mathscr{X})\to\mathbb{R}$, the corresponding mass function p being defined as $p(x):=P(\mathbf{1}_x), x\in\mathscr{X}$ with $P(f)=\sum_{x\in\mathscr{X}}p(x)f(x)$.

A lower expectation $\underline{P}$ induces a corresponding closed convex set $\mathscr{M}(\underline{P})$ of dominating probability distributions, here called *credal set*, such that

$$\mathscr{M}(\underline{P})=\{p\in\mathbb{P}_{\mathscr{X}}|P(f)\geq\underline{P}(f)\ \ \forall f\in\mathscr{L}(\mathscr{X})\},$$

where $\mathbb{P}_{\mathscr{X}}$ is the set of all probability masses over $\mathscr{X}$. One can show that there is a one-to-one correspondence between lower expectations and credal sets (that is, each credal set correspond to one and only one lower expectation, and vice-versa).

In practice, the information contained in $\underline{P}$ can often be restricted to, or is given for, a finite subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X})$, and the induced credal set is then

$$\mathscr{M}(\underline{P})=\{p\in\mathbb{P}_{\mathscr{X}}|P(f)\geq\underline{P}(f)\ \ \forall f\in\mathscr{K}\}.$$

In such a case, the lower expectation, or *natural extension*[1] induced by $\underline{P}$ on any function $g \in \mathscr{L}(\mathscr{X})$ is given by $\underline{P}(g) = \min\{P(g) | p \in \mathscr{M}(\underline{P})\}$. This natural extension represents the most conservative inference one can make when all the information we have about X is represented by the initial lower prevision.

The evaluation of this natural extension, which plays an essential role in further inferences, may represent a heavy computational burden, especially when the space $\mathscr{X}$ is large (as happens in the multivariate case). An important case where this computational burden can be reduced is when $\underline{P}$ can be restricted to events (i.e., is a lower probability) and satisfy the property of 2-monotonicity. This property is satisfied if, for any pair $A, B \subseteq \mathscr{X}$ of events, the following inequality holds:

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) + \underline{P}(A \cap B).$$

Such a property ensures, for instance, that extreme points of $\mathscr{M}(\underline{P})$ can be easily determined [9], or that natural extension over any function can be computed thanks to a Choquet integral. Also, 2-monotonicity is a sufficient condition for $\underline{P}$ to be coherent.

2.2 2-monotone lower probability and Möbius inverse

Let $\underline{P}$ be a lower probability on $\mathscr{X}$. Its Möbius inverse $m : \wp(\mathscr{X}) \to \mathbb{R}$ is defined as a mapping from the power set of $\mathscr{X}$ to the real space such that, for every subset $E \subseteq \mathscr{X}$,

$$m(E) = \sum_{A \subseteq E} (-1)^{|E \setminus A|} \underline{P}(A), \tag{1}$$

with $|E \setminus A|$ the cardinality of $E \setminus A$. Note that for any lower probability, $\sum_{E \subseteq \mathscr{X}} m(E) = 1$, $m(\emptyset) = 0$ and $m(\{x\}) \geq 0$ for any $x \in \mathscr{X}$. From the Möbius inverse m, the lower probability $\underline{P}(A)$ of an event A can be found back through the formula

$$\underline{P}(A) = \sum_{E \subseteq A} m(E). \tag{2}$$

Chateauneuf and Jaffray [9] (among other things) have proved the following relation between 2-monotone lower probabilities and their Möbius inverse:

Proposition 1. *$\underline{P}$ is a 2-monotone lower probability if and only if its Möbius inverse m is such that, for any $A \subseteq \mathscr{X}$ and all $\{x_1, x_2\} \in A$, $x_1 \neq x_2$,*

$$\sum_{\{x_1, x_2\} \subseteq B \subseteq A} m(B) \geq 0$$

This proposition have the following corollary

Corollary 1. *If $\underline{P}$ is a 2-monotone lower probability, then $m(E) \geq 0$ for all E such that $|E| \leq 2$.*

[1] Note that here, we use the same notation for $\underline{P}$ and its natural extension, as we only deal with so-called coherent lower previsions.

However, the inverse is not true, i.e., any mapping m with $\sum_{E\subseteq\mathscr{X}} m(E) = 1$ and $m(E) \geq 0$ for all E such that $|E| \leq 2$ will not induce a 2-monotone lower probability, as shows the next example:

Example 1. Consider a 3 element space $\mathscr{X} = \{x_1,x_2,x_3\}$ with the mass function m such that

$$m(\{x_1\}) = 0.1, \quad m(\{x_2\}) = 0.2, \quad m(\{x_3\}) = 0.5, \quad m(\{x_1,x_2\}) = 0,$$
$$m(\{x_1,x_3\}) = 0.2, \quad m(\{x_2,x_3\}) = 0.3, \quad m(\mathscr{X}) = -0.3.$$

Using Eq (2), we get $\underline{P}(\{x_1\}) = 0.1$ and $\underline{P}(\{x_2,x_3\}) = 1$, a non-coherent lower probability which therefore cannot be 2-monotone (another means to see it is to consider the pair of events $A = \{x_1,x_3\}$ and $B = \{x_2,x_3\}$).

Chateauneuf and Jaffray have also shown that, in the case of 2-monotone lower probabilities, natural extension can be computed using the Möbius inverse.

Proposition 2. *Let $\underline{P}$ be a 2-monotone lower probability and m its Möbius inverse. Then, its natural extension to any function $f \in \mathscr{L}(\mathscr{X})$ is given by*

$$\underline{P}(f) = \sum_{E\subseteq\mathscr{X}} m(E) \inf_{x\in E} f(x). \tag{3}$$

These results will be instrumental in the rest of the paper.

3 2-monotonicity preservation under independence assumptions

We now assume that the uncertainty about two variables X and Y taking their values on finite spaces $\mathscr{X}$ and $\mathscr{Y}$, respectively, are modeled by the 2-monotone lower probabilities $\underline{P}_X$ and $\underline{P}_Y$, respectively. In order to make inferences on the whole space $\mathscr{X} \times \mathscr{Y}$, one needs to build a joint uncertainty model $\underline{P} : \mathscr{L}(\mathscr{X} \times \mathscr{Y}) \to \mathbb{R}$ over it that respects the marginal information given by $\underline{P}_X$ and $\underline{P}_Y$.

As recalled in the introduction, independence assumptions allow one to easily build such a joint uncertainty model from marginal ones. In probability theory, this is done by using the notion of stochastic independence. When considering lower expectations as a model of uncertainty, there exist many ways in which stochastic independence can be extended [2]. Also, one may require, when building the joint uncertainty model, that this joint model remains 2-monotone, if only for computational convenience.

We will show in this section, by the means of simple counter-examples, that the joint models obtained from the marginals $\underline{P}_X$, $\underline{P}_Y$ and the various assumptions of strong independence, epistemic irrelevance or epistemic independence (each of them briefly recalled in the corresponding subsection) are not, in general, 2-monotone lower probabilities.

3.1 Strong independence

The concept of strong independence directly extends the concept of stochastic independence to sets of probabilities, in the sense that it corresponds to take the stochastic product of every probability mass function inside $\mathscr{M}(\underline{P}_X)$ and $\mathscr{M}(\underline{P}_Y)$. The joint lower expectation obtained by such an assumption, denoted by $\underline{P}_{SI}$, is then such that for any $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

$$\underline{P}_{SI}(f) = \inf\{P_{12}(f) | P_{12} = P_1 \otimes P_2, P_1 \in \mathscr{M}(\underline{P}_X), P_2 \in \mathscr{M}(\underline{P}_Y)\},$$

where $\otimes$ is the classical stochastic product. Let us now show that 2-monotonicity is, in general, not preserved by an assumption of strong independence.

Example 2. Consider two binary spaces $\mathscr{X} = \{x_1, x_2\}$ and $\mathscr{Y} = \{y_1, y_2\}$. Recall that any lower expectation on such spaces can be restricted to their values on singletons. Hence they are lower probabilities, which happens to always be 2-monotone. Consider then the following marginal lower probabilities:

$$\underline{P}_X(\{x_1\}) = 0.3, \underline{P}_X(\{x_2\}) = 0.5 \quad \text{and} \quad \underline{P}_Y(\{y_1\}) = 0.4, \underline{P}_Y(\{y_2\}) = 0.4$$

Now, consider the two events $A = \{\mathscr{X} \times y_1\}$ and $B = \{(x_1 \times y_2) \cup (x_2 \times y_1)\}$ on $\mathscr{X} \times \mathscr{Y}$. Under an assumption of strong independence, we have

$$\underline{P}_{SI}(A) = \underline{P}_Y(\{y_1\}) = 0.4,$$
$$\underline{P}_{SI}(B) > 0.4,$$

where the second inequality follows from the fact that all probability masses p which dominate $\underline{P}_{SI}$ must satisfy $p(y_1|x_2) \geq \underline{P}_Y(\{y_1\}) = 0.4$ and $p(y_2|x_1) \geq \underline{P}_Y(\{y_2\}) = 0.4$, whence

$$P(B) = p(y_1|x_2)p(x_2) + p(y_2|x_1)p(x_1) \geq 0.4(p(x_2) + p(x_1)) = 0.4$$

for all probabilities P which dominate $\underline{P}_{SI}$. The actual value is 0.46, obtained by choosing probability masses $p(x_1) = 0.3$ and $p(y_1) = 0.4$. Then, using the factorization properties of $\underline{P}_{SI}$ over events, we have

$$\underline{P}_{SI}(A \cap B) = \underline{P}(x_2 \times y_1) = \underline{P}(x_2)\underline{P}(y_2) = 0.2,$$
$$\underline{P}_{SI}(A \cup B) = \underline{P}(\overline{x_2 \times y_2}) = 1 - \overline{P}(x_2)\overline{P}(y_2) = 0.58,$$

hence, $\underline{P}_{SI}$ violates 2-monotonicity,as

$$\underline{P}_{SI}(A) + \underline{P}_{SI}(B) \geq 0.8 \geq \underline{P}_{SI}(A \cup B) + \underline{P}_{SI}(A \cap B) = 0.78.$$

3.2 Epistemic irrelevance

The concept of epistemic irrelevance [10] corresponds to an asymmetric concept, expressing the idea that learning the value of a variable does not modify the uncertainty (or the knowledge) about the value of another variable (not excluding the possibility that learning the value of the latter may modify our uncertainty about the former). Here, we consider the statement that X is epistemically irrelevant to Y and denote it by $X \not\to Y$. The corresponding joint lower expectation, denoted by $\underline{P}_{X \not\to Y}$, is such that, for any $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

$$\underline{P}_{X \not\to Y}(f) = \underline{P}_X(\underline{P}_Y(f(\mathscr{X}, \cdot))), \tag{4}$$

where $\underline{P}_Y(f(\mathscr{X}, \cdot))$ is a function on $\mathscr{X}$ assuming the value $\underline{P}_Y(f(x, \cdot))$ for every $x \in \mathscr{X}$. Note that, when X is epistemically irrelevant to Y, we have that the sets

$$\{P(\cdot|x) | P \in \mathscr{M}(\underline{P}_{X \not\to Y})\} = \mathscr{M}(\underline{P}_Y)$$

coincide for every $x \in \mathscr{X}$, with $P(\cdot|x)$ the conditional expectation of P. Recall that, given a joint probability mass p over $\mathscr{X} \times \mathscr{Y}$, the conditional expectation $P(f|x)$ of a function $f : \mathscr{Y} \to \mathbb{R}$ is the expectation of f w.r.t. the conditional probability mass $p(\cdot|x)$. This links epistemic irrelevance with credal sets.

Example 3. Consider the same model as in Example 2. The same arguments than for strong independence (factorization and bounds on conditional dominated probabilities) still hold, hence epistemic independence still violates 2-monotonicity. Note that, in this case, the value $\underline{P}_{X \not\to Y}(B) = 0.4$ is exact and can be computed by linear programming.

3.3 Epistemic independence

The concept of epistemic independence [11] is the symmetric counterpart of epistemic irrelevance. It corresponds to the statements that X and Y are epistemically irrelevant of each others, denoted by $X \not\leftrightarrow Y$. The corresponding joint lower expectation, denoted by $\underline{P}_{X \not\leftrightarrow Y}$, is such that, for any $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

$$\underline{P}_{X \not\leftrightarrow Y}(f) = \inf\{P(f) | P \in \left(\mathscr{M}(\underline{P}_{X \not\to Y}) \cap \mathscr{M}(\underline{P}_{Y \not\to X})\right)\}.$$

Similarly to epistemic irrelevance, we have that the sets

$$\{P(\cdot|x) | P \in \mathscr{M}(\underline{P}_{X \not\leftrightarrow Y})\} = \mathscr{M}(\underline{P}_Y) \text{ and } \{P(\cdot|y) | P \in \mathscr{M}(\underline{P}_{X \not\leftrightarrow Y})\} = \mathscr{M}(\underline{P}_X)$$

coincide for every $x \in \mathscr{X}$ and $y \in \mathscr{Y}$.

Example 4. Consider the same model as in Example 2. The same arguments than for strong independence (factorization and bounds on conditional dominated probabilities) still hold, hence epistemic independence still violates 2-monotonicity. Note that, in this case, the value $\underline{P}_{X \not\leftrightarrow Y}(B) = 0.4$ is again exact and can be computed by linear programming.

4 A 2-monotone outer-approximation

In this section, we propose and study a notion that allows one to easily build, from marginals, a joint lower probability that is still 2-monotone and outer-approximates the joint uncertainty models obtained by independence assumptions of Section 3.

4.1 Definition and basic properties

We start by defining how the uncertainty joint model is built, and call the associated notion *Möbius inverse independence* (MI).

Definition 1 (Möbius inverse independence). *Consider two lower probabilities $\underline{P}_X$, $\underline{P}_Y$ defined on finite spaces $\mathscr{X},\mathscr{Y}$ and their respective Möbius inverse m_X, m_Y. The Möbius inverse m_{MI} obtained under an assumption of Möbius inverse independence is defined as the mapping $m_{MI} : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ such that, for every $A \times B \subseteq \mathscr{X} \times \mathscr{Y}$,*

$$m_{MI}(A \times B) = m_X(A)m_Y(B) \tag{5}$$

This notion of independence is symmetrical. The joint lower probability $\underline{P}_{MI}$ induced by m_{MI} over $\mathscr{X} \times \mathscr{Y}$ is then defined for every event $E \subseteq \mathscr{X} \times \mathscr{Y}$ as

$$\underline{P}_{MI}(E) = \sum_{(A \times B) \subseteq E} m_{MI}(A \times B).$$

The MI notion can simply be seen as an extension of the notion of random set independence [2]. Random set independence notion applies to specific kinds of 2-monotone lower probabilities, i.e., belief functions. Recall that a belief function $\underline{P}_{bel} : \mathscr{X} \to [0,1]$ is a lower probability such that, for any collection of events $\{A_1, \ldots, A_n\} \subseteq \mathscr{X}$, the following inequality

$$\underline{P}_{bel}(\bigcup_{i=1^n} A_i) \geq \sum_{\mathscr{I} \subseteq \{1,\ldots,n\}} (-1)^{|\mathscr{I}+1|} \underline{P}(\bigcap_{i \in \mathscr{I}} A_i)$$

holds. Belief functions are also characterised by the fact that their Möbius inverse are non-negative. Given this similarity, we can expect the resulting joint uncertainty model $\underline{P}_{MI}$ to share some properties fo the joint model obtained under an assumption of random set independence (i.e. preservation of n-monotonicity and outer-approximating other joint models studied in Section 3). It should be noted that the Möbius inverse and the corresponding independence notion are here used as a mathematically and computationally convenient tool, and that no semantic is associated to it. Indeed, how to interpret non-positive weights on subsets is still an open problem.

Proposition 3. *Let $\underline{P}_X$, $\underline{P}_Y$ be 2-monotone lower probabilities, then $\underline{P}_{MI}$ is a 2-monotone lower probability.*

Proof. In order to show that $\underline{P}_{MI}$ is 2-monotone, we have to show that m_{MI} has the following properties:

1. $m_{MI}(\emptyset) = 0$
2. $\sum_{A \times B \subseteq \mathscr{X} \times \mathscr{Y}} m_{MI}(A \times B) = 1$
3. For any $A \times B \subseteq \mathscr{X}$ and all $\{x_1 \times y_1, x_2 \times y_2\} \in A \times B$, $\sum_{\{x_1 \times y_1, x_2 \times y_2\} \subseteq C \subseteq A \times B} m(C) \geq 0$ holds (using Prop. 1).

The first property is easily shown, as $m_X(\emptyset) = m_Y(\emptyset) = 0$. The second property follows from

$$\sum_{A \times B \subseteq \mathscr{X} \times \mathscr{Y}} m_{MI}(A \times B) = \sum_{A \subseteq \mathscr{X}} \sum_{B \subseteq \mathscr{Y}} m_X(A)m_Y(B) = \sum_{A \subseteq \mathscr{X}} m_X(A) \sum_{B \subseteq \mathscr{Y}} m_Y(B) = 1.$$

Now, let us show the third property. We have

$$\begin{aligned}\sum_{\{x_1\times y_1, x_2\times y_2\}\subseteq C\subseteq A\times B} m(C) &= \sum_{\{x_1\times y_1, x_2\times y_2\}\subseteq A'\times B'\subseteq A\times B} m(A')m(B') \\ &= \sum_{\{x_1,x_2\}\subseteq A'\subseteq A} m(A') \sum_{\{y_1,y_2\}\subseteq B'\subseteq A} m(B') \geq 0,\end{aligned}$$

where the last inequality comes from the fact that the two sums are positive (according to Prop. 1). □

Let us now show that the joint lower probability $\underline{P}_{MI}$ outer-approximates the joint uncertainty models obtained by other independence notions.

Proposition 4. *Let $\underline{P}_X$, $\underline{P}_Y$ be 2-monotone lower probabilities, then the joint uncertainty model $\underline{P}_{MI}$ outer-approximates the joint uncertainty models $\underline{P}_{X\not\to Y}, \underline{P}_{Y\not\to X}, \underline{P}_{X\not\leftrightarrow Y}, \underline{P}_{SI}$, in the sense that for any $f \in \mathscr{L}(\mathscr{X}\times\mathscr{Y})$,*

$$\underline{P}_{MI}(f) \leq \min\{\underline{P}_{X\not\to Y}(f), \underline{P}_{Y\not\to X}(f), \underline{P}_{X\not\leftrightarrow Y}(f), \underline{P}_{SI}(f)\}.$$

Proof. First, recall that joint models obtained by independence assumptions are related in the following way:

$$\max\{\underline{P}_{X\not\to Y}, \underline{P}_{Y\not\to X}\} \leq \underline{P}_{X\not\leftrightarrow Y} \leq \underline{P}_{SI}$$

where the joint uncertainty models are obtained from the same marginals $\underline{P}_X, \underline{P}_Y$. Hence, it is sufficient to show that $\underline{P}_{MI} \leq \underline{P}_{X\not\to Y}$ to prove that $\underline{P}_{MI}$ outer-approximates the other joint uncertainty models.

Consider a function $f \in \mathscr{L}(\mathscr{X}\times\mathscr{Y})$. Using the fact that $\underline{P}_X$, $\underline{P}_Y$ are 2-monotone lower probabilities and combining Eq. (3) with Eq. (4), we obtain that $\underline{P}_{X\not\to Y}(f)$ can be reformulated as follows:

$$\underline{P}_{X\not\to Y}(f) = \sum_{A\subseteq\mathscr{X}} m_X(A) \inf_{x\in A}\left(\sum_{B\subseteq\mathscr{Y}} m_Y(B) \inf_{y\in B} f(x,y)\right).$$

Similarly, since we have shown that $\underline{P}_{MI}$ is 2-monotone, we can use Eq. (3) and obtain

$$\begin{aligned}\underline{P}_{MI}(f) &= \sum_{A\times B\subseteq\mathscr{X}\times\mathscr{Y}} m_{MI}(A\times B) \inf_{x,y\in A\times B} f(x,y) \\ &= \sum_{A\subseteq\mathscr{X}} \sum_{B\subseteq\mathscr{Y}} m_X(A)m_Y(B) \inf_{x\in A}\inf_{y\in B} f(x,y) \\ &= \sum_{A\subseteq\mathscr{X}} m_X(A) \sum_{B\subseteq\mathscr{Y}} m_Y(B) \inf_{x\in A}\inf_{y\in B} f(x,y).\end{aligned}$$

This shows that $\underline{P}_{MI}(f) \leq \underline{P}_{X\not\to Y}(f)$, since

$$\sum_{B\subseteq\mathscr{Y}} m_Y(B) \inf_{x\in A}\inf_{y\in B} f(x,y) \leq \inf_{x\in A}\left(\sum_{B\subseteq\mathscr{Y}} m_Y(B) \inf_{y\in B} f(x,y)\right).$$

□

Next section discusses the interest of the proposed approximation for various applications.

4.2 Discussion about practical interest

To simplify notations, we identify in this section a function g defined on space $\mathscr{X}$ with its cylindrical extension to the cartesian product $\mathscr{X} \times \mathscr{Y}$ (defined, for every $x \in \mathscr{X}$ and all $y \in \mathscr{Y}$, as $g(x,y) = g(x)$), and we identify similarly functions defined on space $\mathscr{Y}$. Within the theory of lower prevision, recent works [12,13] have focused at characterising interesting factorisation properties of joint models. One of the weakest properties developed in these works is the one of productivity, defined as follows in the case of two variables:

Definition 2 (Productivity). *Consider a joint lower expectation $\underline{P}$ on $\mathscr{L}(\mathscr{X} \times \mathscr{Y})$. This lower expectation is called* productive *if for all $g \in \mathscr{L}(\mathscr{X})$ (resp. all $g \in \mathscr{L}(\mathscr{Y})$) and all non-negative $f \in \mathscr{L}(\mathscr{Y})$ (resp. all $f \in \mathscr{L}(\mathscr{X})$), $\underline{P}(f[g - \underline{P}(g)]) \geq 0$.*

Unfortunately, the next example shows that the joint uncertainty model $\underline{P}_{MI}$ obtained under an MI assumption does not satisfy this property.

Example 5. Let $\mathscr{X} = \{x_1, x_2\}$ and $\mathscr{Y} = \{y_1, y_2\}$ be two binary spaces. Consider two 2-monotone lower probabilities $\underline{P}_Y$ and $\underline{P}_X$ defined on this space and their Möbius inverses m_X and m_Y (note that they are positive), such that

$$m_X(\{x_1\}) = \alpha_1, m_X(\{x_2\}) = \alpha_2 \text{ and } m_X(\mathscr{X}) = 1 - \alpha_1 - \alpha_2;$$

$$m_Y(\{y_1\}) = \beta_1, m_Y(\{y_2\}) = \beta_2 \text{ and } m_Y(\mathscr{Y}) = 1 - \beta_1 - \beta_2;$$

Now consider two functions $g \in \mathscr{L}(\mathscr{X})$ and $f \in \mathscr{L}(\mathscr{Y})$ such that $g(x_1) = a < g(x_2) = b$ and $0 < f(y_1) = c < f(y_2) = d$. Consider now $\underline{P}_{MI}$ as a joint uncertainty model, and let us calculate $\underline{P}_{MI}(f[g - \underline{P}_{MI}(g)])$. Let us first consider $\underline{P}_{MI}(g)$. As $g \in \mathscr{L}(\mathscr{X})$, we have that

$$\underline{P}_{MI}(g) = \alpha_2 b + (1 - \alpha_2)a,$$

and the function $h = f[g - \underline{P}(g)]$ on $\mathscr{X} \times \mathscr{Y}$ is summarised in Table 1 below. The

$h = f[g - \underline{P}_{MI}(g)]$	x_1		x_2
y_1	$c\alpha_2(a-b)$	$<$	$c(1-\alpha_2)(b-a)$
	$\vee$		$\wedge$
y_2	$d\alpha_2(a-b)$	$<$	$d(1-\alpha_2)(b-a)$

Table 1. Function $f[g - \underline{P}(g)]$ of Example 5

inequalities in Table 1 are due to the two inequalities $a \leq b$ and $0 \leq c \leq d$ and to the fact that $(a-b) \leq 0$, $(1-\alpha_2) \geq 0$. Note that the four values are totally ordered. Using Eq. (3) and Definition 1, we have that

$$\begin{aligned}
\underline{P}_{MI}(h) &= (1-\alpha_2)(1-\beta_1)h(x_1,y_2) + \beta_1(1-\alpha_2)h(x_1,y_1) + \alpha_2(1-\beta_2)h(x_2,y_1) + \alpha_2\beta_2 h(x_2,y_2) \\
&= (1-\alpha_2)((1-\beta_1)h(x_1,y_2) + \beta_1 h(x_1,y_1)) + \alpha_2((1-\beta_2)h(x_2,y_1) + \beta_2 h(x_2,y_2)) \\
&= ((1-\alpha_2)\alpha_2(a-b)(d-\beta_1 d + \beta_1 c)) + (\alpha_2(1-\alpha_2)(b-a)(c-\beta_2 c + \beta_2 d)) \\
&= (1-\alpha_2)\alpha_2(b-a)(c-d)(1-\beta_2-\beta_1) = (1-\alpha_2)\alpha_2(b-a)(c-d)\beta_3
\end{aligned}$$

If we assume that $0 < \alpha_2 < 1$, then this value is negative (as $b-a > 0$ and $c-d < 0$), unless $\beta_3 = 0$, that is unless $\underline{P}_Y$ is a precise probability. If we extend these conclusions to all possible f and g satisfying Def. 2, this means that $\underline{P}_{MI}(f[g-\underline{P}_{MI}(g)]) \geq 0$ only in degenerated cases (that is, when $\underline{P}_X$ and $\underline{P}_Y$ are either both precise probabilities or vacuous models).

This example shows that we cannot expect the notion of Möbius inverse independence (and also of random set independence) to satisfy productivity as well as other stronger factorization properties that imply productivity. In the framework of lower previsions, such factorisation properties allows to easily derive laws of large numbers, or are instrumental in the construction of generalisation of Bayesian networks. However, it should be noted that random set independence (of which Möbius inverse independence is a direct extension) has been used in graphical models [14], hence not satisfying productivity does not mean that this independence notion cannot be useful in such models.

Also, the computational convenience of this approximation may be useful in some practical applications involving the computation of natural extension. One such application, illustrated by the following (simple) example, may be multi-criteria decision-making under uncertainty.

Example 6. Assume that some decision maker (DM) wants to build a new airport in a region, and has retained some sites to do so. After selecting sites whose building costs are roughly equivalent, the DM decides to base his/her decision on some additional criteria: the easiness of access to main roads (variable X defined on $\mathscr{X}$), the generated pollution impact on nearby lands (variable Y defined on $\mathscr{Y}$) and the public opinion (variable Z defined on $\mathscr{Z}$). Each criterion is evaluated on a utility scale ranging from 1 to 4, 1 being the worst case, 4 the best. Criteria values are then aggregated according to a weighted average $f = w_X X + w_Y Y + w_Z Z$ to obtain the global utility of a given alternative, where $w_X = 0.2, w_Y = 0.4, w_Z = 0.4$ are the importance weights given to each criterion.

Now, consider an alternative where the utility of each criterion is uncertainly known. The uncertainty concerning variable X is given by the following probability intervals (i.e. upper and lower probabilities over singletons):

$$\overline{P}(\{1\}) = 0.1,\ \overline{P}(\{2\}) = 0.2,\ \overline{P}(\{3\}) = 0.6,\ \overline{P}(\{4\}) = 0.7$$

$$\underline{P}(\{1\}) = 0,\ \underline{P}(\{2\}) = 0,\ \underline{P}(\{3\}) = 0.3,\ \underline{P}(\{4\}) = 0.3$$

This uncertainty can correspond to the fact that a major road is likely to be built in the future in the region, but that this fact is not fully certain. Uncertainty can come, for example, from an expert. These probability intervals are 2-monotone (we refer to [6] for details on probability intervals) and their Möbius inverse is such that

$$m_X(\{3\}) = m_X(\{4\}) = 0.3,\ m_X(\{3,4\}) = m_X(\{1,2,4\}) = m_X(\{1,3,4\}) = 0.1,$$

$$m_X(\{2,3,4\}) = 0.2,\ m_X(\mathscr{X}) = -0.1.$$

Concerning variable Y, risk analysis shows that pollution impact may be high, and the related uncertainty is modeled by the possibility distribution (recall that possibility distributions have Möbius inverses which are positive and are such that non-null masses are given to nested sets)

$$m_Y(\{1\}) = 0.3,\; m_Y(\{1,2\}) = 0.7.$$

Finally, public opinion has been gathered by a survey where answers could be imprecise (hence, frequencies are given to set of values). The results are such that

$$m_Z(\{2\}) = 0.3,\; m_Z(\{4\}) = 0.2,\; m_Z(\{1,2\}) = 0.2,\; m_Z(\mathscr{Z}) = 0.3.$$

The weighted average (or any other aggregation functions) is a mapping $f : \mathscr{X} \times \mathscr{Y} \times \mathscr{Z} \to \mathbb{R}$, and as it seems reasonable to assume that each criterion is independent of the other, we can use m_{MI} as a joint model over $\mathscr{X} \times \mathscr{Y} \times \mathscr{Z}$ to compute lower and upper expectations outer approximating results given by other (more complex) joint models. Using m_X, m_Y, m_Z as uncertainty models, the results are (for lower and upper expectations)

$$\underline{P}_{MI}(f) = 1.936 \quad ; \quad \overline{P}_{MI}(f) = -\underline{P}_{MI}(-f) = 2.62.$$

Note that, in the above example, f can be replaced by any mapping or by any indicator function on the resulting output of f, thus allowing one to perform uncertainty propagation through f.

Finally, let us make two remarks concerning complexity related issues:

- storing information in terms of the Möbius inverse means storing at most $2^{|\mathscr{X}|}$ values, as for lower probabilities on every events. This can be compared to the maximum number of extreme points of a credal set induced by a 2-monotone lower probability [9], which is $|\mathscr{X}|!$ (i.e., the number of permutations among elements of $\mathscr{X}$);
- when working in a multivariate space, computing the lower expectation $\underline{P}_{MI}(f)$ has a complexity that increases exponentially with the number of variables. This is comparable to the complexity associated to the computations under an assumption of forward irrelevance [10]. Also, if an important number of Möbius inverses are positive (i.e., if marginal probabilities often correspond to belief functions), then exact computations could be combined with efficient simulation techniques [15].

Acknowledgements

Examples of Section 3 are the results of discussion with M. Troffaes and E. Miranda.

5 Conclusions

Independence notions play a central role in many applications of uncertainty reasoning. We have shown that the joint models obtained by independence notions proposed in the theory of imprecise probabilities, in which uncertainty is modeled by the means of credal sets or lower previsions, do not preserve the 2-monotonicity property of marginal uncertainty models (when these latter models satisfy it).

This is a practical downside of these independence notions, as satisfying 2-monotonicity increases the computational tractability of imprecise probabilistic models. To solve this issue, we have proposed a 2-monotone outer-approximation by simply extending the notion of random set independence to 2-monotone lower probabilities.

This approximation does not satisfy the weak property of productivity, which is implied by many other factorization properties of joint models. This means that this approximation cannot benefit from results associated to such properties. Still, there remains applications where this approximation may be useful, such as the one involving uncertainty propagation or expectation bound computations. Especially, since this approximation is an extension of the random set independence, it may benefits from algorithms and methods originating from random set and evidence theory.

References

1. Walley, P.: Statistical reasoning with imprecise Probabilities. Chapman and Hall, New York (1991)
2. Couso, I., Moral, S., Walley, P.: A survey of concepts of independence for imprecise probabilities. Risk Decision and Policy **5** (2000) 165–181
3. Ferson, S., Ginzburg, L., Kreinovich, V., Myers, D., Sentz, K.: Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories (2003)
4. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
5. Shafer, G.: A mathematical Theory of Evidence. Princeton University Press, New Jersey (1976)
6. de Campos, L., Huete, J., Moral, S.: Probability intervals: a tool for uncertain reasoning. I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **2** (1994) 167–196
7. Bronevich, A., Augustin, T.: Approximation of coherent lower probabilities by 2-monotone measures. In: ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications, SIPTA (2009) 61–70
8. Miranda, E., Couso, I., Gil, P.: Extreme points of credal sets generated by 2-alternating capacities. I. J. of Approximate Reasoning **33** (2003) 95–115
9. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Mobius inversion. Mathematical Social Sciences **17**(3) (1989) 263–283
10. de Cooman, G., Miranda, E.: Forward irrelevance. Journal of Statistical Planning and Inference **139** (2009) 256–276
11. Vicig, P.: Epistemic independence for imprecise probabilities. Int. J. of approximate reasoning **24** (2000) 235–250
12. de Cooman, G., Miranda, E., Zaffalon, M.: Independent natural extension. In: IPMU. (2010) 737–746
13. de Cooman, G., Miranda, E., Zaffalon, M.: Factorisation properties of the strong product. In: Proceedings of the Fifth International Conference on Soft Methods in Probability and Statistics (SMPS'2010). (2010)
14. Xu, H., Smets, P.: Reasoning in evidential networks with conditional belief functions. Int. J. Approx. Reasoning **14**(2-3) (1996) 155–185
15. Wilson, N.: Algorithms for Dempster-Shafer Theory. In: Handbook of Defeasible Reasoning and Uncertainty Management. Vol. 5: Algorithms. Kluwer Academic (2000) 421–475