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

Sébastien Destercke

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

S. Destercke

Université de Technologie de Compiègne - CNRS, Heudiasyc UMR 7253 BP 20529, 60205 Compiègne, France

Abstract

In imprecise probability theories, independence modeling and computational tractability are two important issues. The former is essential to work with multiple variables and multivariate spaces, while the latter is essential in practical applications. When using lower probabilities to model uncertainty about the value assumed by a variable, satisfying the property of 2-monotonicity decreases the computational burden of inference, hence answering the latter issue. In a first part, this paper investigates whether the joint uncertainty obtained by main existing notions of independence preserve the 2-monotonicity of marginal models. It is shown that it is usually not the case, except for the formal extension of random set independence to 2-monotone lower probabilities. The second part of the paper explores the properties and interests of this extension within the setting of lower probabilities.

Keywords: factorisation properties, credal sets, propagation, lower previsions

1. Introduction

Independence modelling and computational tractability have always been two important issues in uncertainty theories. On one hand, independence notions allow one to easily deal with multivariate spaces, their associated factorization properties allowing one to decompose a complex problem into simpler ones (using, e.g., graphical models), or to easily build joint models from marginal ones. They are also essential to derive statistical results such as laws of large numbers. On the other hand, ensuring computational tractability is essential in many applications, especially those involving complex systems.

Lower expectation bounds, better known as *lower previsions* [31], are very general models of uncertainty that have been shown to be formally equivalent to convex sets of probabilities, or *credal sets*. They include most known uncertainty models as special cases [32], but this generality can result in a prohibitive computational cost of information processing.

☆This paper is an extended version of [10]

Email address: sebastien.destercke@hds.utc.fr (S. Destercke)

This computational issue is even more critical in multivariate problems. A first mean to reduce the computational burden of dealing with multivariate models is to assume independence between the variables: this facilitates the joint model construction and avoids the need to identify dependence structures. When expectations or probabilities are made imprecise, the classical notion of stochastic independence between precise probabilities can be extended in several ways. Such extensions have been proposed and compared by many authors (see, for example, Walley [31] and Couso *et al.* [4]). These independence notions ensure that the joint uncertainty model can be built from the sole knowledge of the marginal models, facilitating their handling. However, these joint models may still be hard to handle from a computational viewpoint. A possible way to facilitate even more this computational handling is for the joint model to share some convenient properties with the marginals (2-monotonicity in this paper).

In practice, tractability can be further improved by restricting oneself to classes of uncertainty models that present a good trade-off between generality and computational convenience. 2-monotone lower probabilities constitute such a trade-off: they still encompass many useful uncertainty models (e.g., belief functions [26, 9], probability intervals [5], possibility distributions [13], p-boxes [14, 11], Pari-mutuel models [22]) while the property of 2-monotonicity greatly facilitates several computational aspects [21, 2, 1].

This paper is divided in two parts. The first part investigates whether joint lower expectations built through various independence assumptions preserve the 2-monotonicity of marginals. Only the most common type of independence assumptions are considered: the case of strong independence is studied in Section 3, the cases of epistemic irrelevance and independence are studied in Section 4, while Section 5 studies an extension of random set independence to 2-monotone lower probabilities, that we call *Möbius independence*. As this extension turns out to satisfy 2-monotonicity preservation, the second part of the paper explores it in more details. Section 6, whose results also apply to random set independence (of which Möbius independence is a formal extension), investigates its properties as well as its links with the lower prevision approach. Finally, Section 7 illustrates Möbius independence on an illustrative practical example in the domain of multi-criteria decision making. Preliminaries about lower previsions and 2-monotone lower probabilities needed in this paper are recalled in Section 2.

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

Lower expectations have been introduced by Williams [33] and formalized by Walley [31] as a unifying reasoning framework encompassing most known uncertainty models, which generalizes de Finetti's [16] subjective account of probabilities (and is therefore coherent with this latter). This section recalls the basics needed in this paper, and we refer to Refs. [20] and [31] for more details. Note that since this study mainly focuses on formal properties, we will prefer the term lower expectation to lower prevision, as the latter is related to a particular interpretation of imprecise probabilities.

Consider a variable X whose value lies in a finite space $\mathscr{X}$. We assume that the uncertainty on X is described by a lower expectation $\underline{P} : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ defined over the set $\mathscr{L}(\mathscr{X})$ of all real-valued functions over $\mathscr{X}$. The quantity $\underline{P}(f)$ then denotes the lower expectation of a function f. A lower expectation $\underline{P}$ is associated with its dual upper expectation $\overline{P}$, defined as $\overline{P}(f) = -\underline{P}(-f)$ for any $f \in \mathscr{L}(\mathscr{X})$.

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 ($\mathbf{1}_A(x) = 1$ if $x \in A$, zero otherwise). This lower probability will be denoted by $\underline{P}(A)$ when no confusion is possible. In the specific case of lower probabilities, the dual notion of upper probability is such that $\overline{P}(A) = 1 - \underline{P}(\overline{A})$ where $\overline{A}$ is the complement of A.

A *coherent lower expectation* $\underline{P}$ on $\mathscr{L}(\mathscr{X})$ is defined as a lower expectation satisfying the following conditions:

1. $\underline{P}(f) \geq \inf_{x\in\mathscr{X}} f(x)$ for all $f \in \mathscr{L}(\mathscr{X})$ (accepting sure gain);
2. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for each $f \in \mathscr{L}(\mathscr{X})$ and $\lambda \geq 0$ (positive homogeneity);
3. $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathscr{L}(\mathscr{X})$ (superadditivity).

Note that, in this paper, we will exclusively deal with coherent lower expectations.

A coherent lower expectation is said to be a *linear expectation* P if it is self-dual, that is if $\underline{P}(f) = \overline{P}(f) = P(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ (in particular, $\underline{P}(A) = \overline{P}(A) = P(A)$ for any event A). A linear expectation P is additive, in the sense that $P(f+g) = P(f) + P(g)$ for all $f, g \in \mathscr{L}(\mathscr{X})$. A linear expectation P corresponds to the expectation of a probability measure, and we will denote by p the corresponding probability mass function p defined as $p(x) := P(\mathbf{1}_x), x \in \mathscr{X}$ and such that $P(f) = \sum_{x\in\mathscr{X}} p(x) f(x)$ for any $f \in \mathscr{L}(\mathscr{X})$.

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 $\mathscr{X}$ and $P(f)$ is the linear expectation of f induced by p. One can show that there is a one-to-one correspondence between lower expectations and credal sets (that is, each credal set corresponds to one and only one lower expectation).

As $\mathscr{M}(\underline{P})$ is a convex set of probabilities, another way to represent is by its set $\mathscr{E}(\mathscr{M}(\underline{P}))$ of extreme points. Although this representation is in one-to-one correspondence with coherent lower expectations, it will often be less convenient than using bounds over expectations, simply because the number of extreme points can be very large. However, the computation of such extreme points may be necessary in some algorithms, for example in some extensions of graphical models [23].

In practice, the information contained in $\underline{P}$ will often be given or restricted to a finite subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X})$. A lower expectation defined on such a subset $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ is

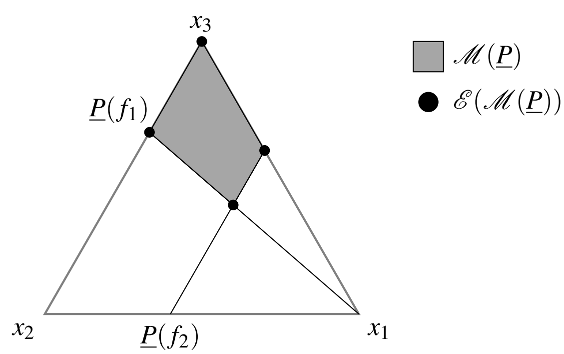

Figure 1: Lower expectations, credal set and extreme points of Example 1

called coherent if it is the restriction on $\mathscr{K}$ of some coherent lower expectation on $\mathscr{L}(\mathscr{X})$. The induced credal set is then

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

and 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 on g when all the information we have about X is represented by the initial lower expectation $\underline{P}$ defined on $\mathscr{K}$.

Example 1. Consider a 3 elements space $\mathscr{X} = \{x_1, x_2, x_3\}$. The information (e.g., given by some expert) we have about the ill-known variable X is that

- x_3 is at least two times more probable than x_2: $2P(\{x_2\}) \leq P(\{x_3\})$, and
- the probability of x_1 is not higher than 0.4: $P(\{x_1\}) \leq 0.4$.

The first statement can be transformed into $0 \leq P(\{x_3\}) - 2P(\{x_2\})$, meaning that the lower expectation of the function $f_1(x_1) = 0, f_1(x_2) = -2, f_1(x_3) = 1$ is $\underline{P}(f_1) = 0$. Similarly, the second statement says that $\underline{P}(f_2 = -\mathbf{1}_{\{x_1\}}) = -0.4$ (obtained from $\overline{P}(\{x_1\}) = 0.4$ and duality of lower/upper expectations). The coherent lower expectation $\underline{P}$ defined on $\mathscr{K} = \{f_1, f_2\}$ then induces the credal set depicted[2] in Figure 1.

Manipulating coherent lower expectations to make various inferences (e.g., compute the natural extension, conditioning) may represent a heavy computational burden, especially when the space $\mathscr{X}$ is large (as it happens in the multivariate case). An important case where it is reduced is when $\underline{P}$ is restricted to events (i.e., $\mathscr{K}$ is the

[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 expectations.

[2]The figure displays the unit simplex in barycentric coordinates where each point represent a probability$p = (p_1, p_2, p_3)$ with $p_i = P(\{x_i\})$. Recall that given the extreme points x_1, x_2, x_3 of the simplex, if $p = p_1x_1 + p_2x_2 + p_3x_3$, the probability masses (p_1, p_2, p_3) are the barycentric coordinates of p. See Ref. [24] for details about this representation.

set of indicator functions and $\underline{P}$ is a lower probability) and satisfies the property of 2-monotonicity. Before summarizing the main advantages of 2-monotonicity in Section 2.3, we recall the property of n-monotonicity and its links with the Möbius inverse, as these notions will be used in the following.

2.2. *n-monotone lower probabilities and Möbius inverse*

Many of the results presented in this section and the next one are contained in the paper of Chateauneuf and Jaffray [2].

A lower probability $\underline{P}$ is n-monotone, where $n > 1$ and $n \in \mathbb{N}$, if and only if for any set $\mathscr{A} = \{A_i \subseteq \mathscr{X} | i \in \mathbb{N}, 0 < i \leq n\}$ of events A_i, it holds that

$$\mu\left(\bigcup_{A_i \in \mathscr{A}} A_i\right) \geq \sum_{I \subseteq \mathscr{A}} (-1)^{|I|+1} \mu\left(\bigcap_{A_i \in I} A_i\right) \tag{1}$$

where $|I|$ is the cardinality of I. If a lower probability is n-monotone, then it is $(n-1)$-monotone. Any n-monotone lower probability $\underline{P}$ for $n \geq 2$ is coherent (i.e., is the lower envelope of $\mathscr{M}(\underline{P})$ on events).

In particular, a lower probability $\underline{P}$ satisfies the property of 2-monotonicity when, for any pair $A, B \subseteq \mathscr{X}$ of events, the following inequality holds:

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) + \underline{P}(A \cap B). \tag{2}$$

Chateauneuf and Jaffray [2] have shown that there are strong links between the n-monotonicity of a lower probability $\underline{P}$ and the *mass assignment* induced by its Möbius inverse. The Möbius inverse $m : \wp(\mathscr{X}) \to \mathbb{R}$ of $\underline{P}$ is defined as a mapping from the power set of $\mathscr{X}$ to the real space such that, for every subset $E \subseteq \mathscr{X}$,

$$m(E) = \sum_{A \subseteq E} (-1)^{|E \setminus A|} \underline{P}(A), \tag{3}$$

with $|E \setminus A|$ the cardinality of $E \setminus A = \{x | x \in E \text{ and } x \notin A\}$. We will call *Möbius mass* the mass $m(E)$ attributed to a set E. For any lower probability, $\sum_{E \subseteq \mathscr{X}} m(E) = 1$, $m(\emptyset) = 0$ and $m(\{x\}) \geq 0$ for any $x \in \mathscr{X}$. The Möbius inverse is a bijective transformation, meaning that the lower probability $\underline{P}$ can be retrieved from m through the formula

$$\underline{P}(A) = \sum_{E \subseteq A} m(E) \tag{4}$$

for any $A \subseteq \mathscr{X}$. Using the language of belief functions [26], we will call *focal* a set E whose Möbius mass is non-null (i.e., $m(E) \neq 0$).

2-monotonicity of $\underline{P}$ can be checked through its Möbius inverse, thanks to the following proposition [2]:

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 $\{x, y\} \in A$, $x \neq y$,*

$$\sum_{\{x,y\} \subseteq B \subseteq A} m(B) \geq 0.$$

This proposition has the following corollary

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

However, the converse 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 the next example shows:

Example 2. Consider a 3 elements 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 (4), 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 $B = \{x_2,x_3\}$).

2.3. Some advantages of 2-monotonicity

According to Walley [30, P. 51], there is not "... any rationality argument for two-monotonicity, beyond its computational convenience." However, as said by other authors [1] and as recalled in the Introduction, computational convenience is a critical issue in imprecise probabilistic approaches, and having easy-to-find 2-monotone approximations at our disposal can help to solve practical problems.

Let us recall some of the computational advantages of a 2-monotone lower probability $\underline{P}$:

- the natural extension of $\underline{P}$ can be exactly computed using the Choquet integral. If f is a positive bounded function over $\mathscr{X}$, its natural extension $\underline{P}(f)$ is

$$\underline{P}(f) = \sum_{i=1}^{|\mathscr{X}|} f(x_{\sigma(i)})(\underline{P}(A_{\sigma(i)}) - \underline{P}(A_{\sigma(i+1)})) \tag{5}$$

  with σ the permutation of $\mathscr{X}$ elements such that $f(x_{\sigma(1)}) \leq \ldots \leq f(x_{\sigma(|\mathscr{X}|)})$ and with $A_{\sigma(i)} = \{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\}$ ($A_{\sigma(|\mathscr{X}|+1)} = \emptyset$). This Choquet integral can also be computed as follows

$$\underline{P}(f) = \sum_{E \subseteq \mathscr{X}} m(E) \inf_{x \in E} f(x) \tag{6}$$

  with m the Möbius inverse of $\underline{P}$;

- the vertices of the credal set $\mathscr{M}(\underline{P})$ can be generated in a straightforward way. If σ denotes a permutation of elements of $\mathscr{X}$, then the probability mass function p_σ such that, for any $i \in 1, \ldots, |\mathscr{X}|$ we have

$$p_\sigma(x_{\sigma(i)}) = \underline{P}(A_{\sigma(i)}) - \underline{P}(A_{\sigma(i+1)})$$

belongs to the set of extreme points $\mathscr{E}(\mathscr{M}(\underline{P}))$. Sampling extreme points of credal sets generated by 2-monotone lower probabilities then comes down to sample permutations of elements of $\mathscr{X}$, thus avoiding the need to enumerate them. Such a sampling can then be used to approximate the results of inference procedures whose exact solution is known to lie on some vertex (or on a combination of vertices), such as inferences in Credal networks [23];

- conditional lower and upper probabilities $\underline{P}(A|B) = \inf_{p \in \mathscr{M}(\underline{P})} P(A|B)$ and $\overline{P}(A|B) = \sup_{p \in \mathscr{M}(\underline{P})} P(A|B)$, where $P(A|B) = P(A\cap B)/P(B)$ is the conditional probability measure on A computed from p, can be evaluated through the closed-form formulas

$$\underline{P}(A|B) = \frac{\underline{P}(A \cap B)}{\underline{P}(A \cap B) + \overline{P}(\overline{A} \cap B)}; \quad \overline{P}(A|B) = \frac{\overline{P}(A \cap B)}{\overline{P}(A \cap B) + \underline{P}(\overline{A} \cap B)}. \tag{7}$$

  It can be shown that $\underline{P}(\cdot|B)$ remains 2-monotone if $\underline{P}$ is. Similar arguments apply to other conditioning rules such as Dempster's conditioning [9] that can be extended to 2-monotone lower probabilities if assimilated to a maximum likelihood principle [17];

- robust statistics based on minimax test and least-favourable pairs can be efficiently performed by using the Huber-Strassen theory [18];

- propagating the uncertainty represented by $\underline{P}$ over some input variable X through a function $h(X) = Y$ to estimate the uncertainty over some output variable Y (assuming its values on a space $\mathscr{Y}$) can be done easily through the Möbius inverse, as it simply comes down to transfer the Möbius mass $m(E)$ to the image $h(E)$ (obtaining $m(h(E))$).

From now on, we will assume that the uncertainty concerns two variables X and Y taking their values on two finite spaces $\mathscr{X}$ and $\mathscr{Y}$, respectively, and is modeled by the 2-monotone lower probabilities $\underline{P}_X$ and $\underline{P}_Y$, respectively. To make inferences on the product space $\mathscr{X} \times \mathscr{Y}$, one needs to build a joint uncertainty model $\underline{P} : \mathscr{L}(\mathscr{X} \times \mathscr{Y}) \to \mathbb{R}$ that complies with the marginal information given by $\underline{P}_X$ and $\underline{P}_Y$, i.e. for any $A \subseteq \mathscr{X}$ and any $B \subseteq \mathscr{Y}$, we should have

$$\underline{P}(A \times \mathscr{Y}) = \underline{P}_X(A) \text{ and } \underline{P}(\mathscr{X} \times B) = \underline{P}_Y(B). \tag{8}$$

A classical mean to build the joint model $\underline{P}$ is to assume that X and Y are independent. Given the practical interest of having a 2-monotone model, the next sections of this paper address two problems:

- whether usual independence assumptions [4] applied to marginal models that are 2-monotone lower probabilities will lead to 2-monotone joint models (Sections 3 and 4);

- if the answer to this questions is negative, whether or not it is possible to find an easy-to-build 2-monotone approximation (Section 5).

3. The case of strong independence

When uncertainty is modelled by precise probabilities, a first way to express independence between two variables X and Y is to require the joint model P_{XY} to satisfy

$$P_{XY}(A \times B) = P_X(A)P_Y(B) \tag{9}$$

for any $A \subseteq \mathscr{X}$ and any $B \subseteq \mathscr{Y}$. Note that Equation (9) is by nature symmetric. The joint probability mass p_{XY} is then obtained by the stochastic product $p_{XY} := p_X \otimes p_Y$ defined, for any $x \in \mathscr{X}$ and $y \in \mathscr{Y}$, as

$$p_{XY}(x,y) = p_X(x)p_Y(y). \tag{10}$$

The concept of strong independence directly extends Equation (10) to sets of probabilities, in the sense that it corresponds to taking 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 $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

$$\underline{P}_{SI}(f) = \inf\{P_{12}(f) | p_{12} = p_1 \otimes p_2, p_1 \in \mathscr{M}(\underline{P}_X), p_2 \in \mathscr{M}(\underline{P}_Y)\},$$

with P_{12} the linear expectation induced by p_{12}. The set of extreme points of $\mathscr{M}(\underline{P}_{SI})$ can be obtained by computing the stochastic products of every extreme points of $\mathscr{M}(\underline{P}_X)$ and $\mathscr{M}(\underline{P}_Y)$. Strong independence satisfies Equations (8), and it can be shown [4] that on Cartesian products of events $A \times B, A \subseteq \mathscr{X}, B \subseteq \mathscr{Y}$, the following factorization property holds:

$$\underline{P}_{SI}(A \times B) = \underline{P}_X(A)\underline{P}_Y(B); \quad \overline{P}_{SI}(A \times B) = \overline{P}_X(A)\overline{P}_Y(B). \tag{11}$$

The next example shows that 2-monotonicity is, in general, not preserved by the assumption of strong independence.

Example 3. Consider two binary spaces $\mathscr{X} = \{x_1, x_2\}$ and $\mathscr{Y} = \{y_1, y_2\}$. Recall that any lower expectation on binary spaces only depends on its values on singletons and is always a 2-monotone lower probability. Consider then the following marginal lower probabilities:

$$\underline{P}_X(\{x_1\}) = 0.2, \underline{P}_X(\{x_2\}) = 0.3 \quad \text{and} \quad \underline{P}_Y(\{y_1\}) = 0.3, \underline{P}_Y(\{y_2\}) = 0.4.$$

The extreme points of $\mathscr{E}(\mathscr{M}(\underline{P}_X))$ and $\mathscr{E}(\mathscr{M}(\underline{P}_Y))$ are respectively:

$$\mathscr{E}(\mathscr{M}(\underline{P}_X)) = \{(p(x_1) = 0.2, p(x_2) = 0.8), (p(x_1) = 0.7, p(x_2) = 0.3)\},$$

$$\mathscr{E}(\mathscr{M}(\underline{P}_Y)) = \{(p(y_1) = 0.3, p(y_2) = 0.7), (p(y_1) = 0.6, p(y_2) = 0.4)\}.$$

The four extreme points of $\mathscr{M}(\underline{P}_{SI})$ obtained from these two sets are summarized in Table 1. Now, consider the two events $A = \mathscr{X} \times \{y_2\}$ and $B = (\{x_1\} \times \{y_1\}) \cup (\{x_2\} \times \{y_2\})$ on $\mathscr{X} \times \mathscr{Y}$. Under an assumption of strong independence, we have

$$\underline{P}_{SI}(A) = \underline{P}_Y(\{y_2\}) = 0.4,$$
$$\underline{P}_{SI}(B) = 0.42.$$

$\mathscr{E}(\mathscr{M}(\underline{P}_{SI}))$	(x_1,y_1)	(x_1,y_2)	(x_2,y_1)	(x_2,y_2)
p^1	0.06	0.14	0.24	0.56
p^2	0.21	0.49	0.09	0.21
p^3	0.12	0.08	0.48	0.32
p^4	0.42	0.28	0.18	0.12

Table 1: Probability mass functions in $\mathscr{E}(\mathscr{M}(\underline{P}_{SI}))$ of Example 3.

$\underline{P}_{SI}(A)$ is obtained by applying Equation (8), while $\underline{P}_{SI}(B)$ is obtained by selecting p^2 in Table 1. Then, using the factorization properties (11) of $\underline{P}_{SI}$ over Cartesian products, we have

$$\underline{P}_{SI}(A\cap B) = \underline{P}(\{x_2\}\times\{y_2\}) = \underline{P}(\{x_2\})\underline{P}(\{y_2\}) = 0.12,$$
$$\underline{P}_{SI}(A\cup B) = \underline{P}(\overline{\{x_2\}\times\{y_1\}}) = 1-\overline{P}(\{x_2\})\overline{P}(\{y_1\}) = 0.52,$$

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

$$0.82 = \underline{P}_{SI}(A)+\underline{P}_{SI}(B) \geq \underline{P}_{SI}(A\cup B)+\underline{P}_{SI}(A\cap B) = 0.64.$$

Strong independence therefore does not preserve 2-monotonicity. The next section deals with another extension of stochastic independence to imprecise probabilities: epistemic irrelevance and independence.

4. The case of epistemic irrelevance and independence

A second way to express independence between precise probabilities is through conditional probabilities, by requiring that the conditional probability computed from the joint model satisfies

$$P_{XY}(B|A) = P_Y(B) \tag{12}$$

for any $A\subseteq\mathscr{X}$ and any $B\subseteq\mathscr{Y}$. Equation (12) is by nature asymmetric, and translates the fact that learning the value of X does not change our information about Y, or in other words that X is *epistemically irrelevant* to Y. It is only the axioms of probability theory that make this condition symmetric (i.e., Equation (12) implies Equation (9)). Using Equation (12), the joint probability is

$$p_{XY}(x,y) = p_Y(y|x)p_X(x) = p_Y(y)p_X(x). \tag{13}$$

The notion of epistemic irrelevance with imprecise probabilities aims at expressing the same idea, that is the fact 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). We consider the statement that X is epistemically irrelevant to Y and denote it by $X \not\rightarrow Y$. We refer to Ref. [7] for the extension to any number of variables.

$\mathscr{E}(\mathscr{M}(\underline{P}_{X \not\to Y}))$	(x_1, y_1)	(x_1, y_2)	(x_2, y_1)	(x_2, y_2)
p^5	0.06	0.14	0.48	0.32
p^6	0.21	0.49	0.18	0.12
p^7	0.12	0.08	0.24	0.56
p^8	0.42	0.28	0.09	0.21

Table 2: Probability mass functions in $\mathscr{E}(\mathscr{M}(\underline{P}_{X \not\to Y}))$ of Example 4.

The corresponding joint lower expectation, denoted by $\underline{P}_{X \not\to Y}$, is such that, for any $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

$$\underline{P}_{X \not\to Y}(f) = \inf\{P_{12}(f) | p_{12}(x,y) = p_2(y|x)p_1(x), p_1 \in \mathscr{M}(\underline{P}_X), p_2(\cdot|x) \in \mathscr{M}(\underline{P}_Y)\}, \tag{14}$$

where $p_2(y|x)$ may depend on the value of x, i.e, for two distinct values $x, x' \in \mathscr{X}$, we may have $p_2(y|x) \neq p_2(y|x')$. The constraint $p_2(\cdot|x) \in \mathscr{M}(\underline{P}_Y)$ corresponds to the epistemic irrelevance assumption that learning X value does not change the information about Y (given by $\mathscr{M}(\underline{P}_Y)$). The extreme points $\mathscr{E}(\mathscr{M}(\underline{P}_{X \not\to Y}))$ can be obtained by computing $p(x,y) = p_1(x)p_2(y|x)$ for each $x, y \in \mathscr{X} \times \mathscr{Y}$, with $p_1 \in \mathscr{E}(\mathscr{M}(\underline{P}_X))$ and $p_2(\cdot|x) \in \mathscr{E}(\mathscr{M}(\underline{P}_Y))$ (possibly choosing $p_2(\cdot|x) \neq p_2(\cdot|x')$ for $x \neq x'$). Note that when working with imprecise probabilities, an assumption of epistemic irrelevance is truly asymmetric, as the two models $\underline{P}_{X \not\to Y}$ and $\underline{P}_{Y \not\to X}$ will in general be different. The symmetrisation of the notion is called epistemic independence, and will be studied later in this section. In contrast with strong independence, the joint lower expectation $\underline{P}_{X \not\to Y}(f)$ can easily be expressed in terms of marginal lower expectations as we have [7]

$$\underline{P}_{X \not\to Y}(f) = \underline{P}_X(\underline{P}_Y(f(\mathscr{X}, \cdot))), \tag{15}$$

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 $x \in \mathscr{X}$.

The joint lower expectation $\underline{P}_{X \not\to Y}$ satisfies Equation (8) and also factorizes over Cartesian products of events [4] (see Equation (11)). The next example shows that, as for strong independence, 2-monotonicity is in general not preserved by epistemic irrelevance.

Example 4. Consider the same marginal models as in Example 3 with the assumption that $X \not\to Y$, and the same events A and B. As $\underline{P}_{X \not\to Y}$ still factorizes over products of events, the values for $\underline{P}_{X \not\to Y}(A \cup B) = \underline{P}_{SI}(A \cup B)$ and $\underline{P}_{X \not\to Y}(A \cap B) = \underline{P}_{SI}(A \cap B)$ remain unchanged. Table 2 summarizes the extreme points in $\mathscr{E}(\mathscr{M}(\underline{P}_{X \not\to Y}))$ not already given in Table 1 (the first point p^5 is obtained by taking $p_1(x_1) = 0.2$, $p_2(y_1|x_1) = 0.3 = 1 - p_2(y_2|x_1)$ and $p_2(y_1|x_2) = 0.7 = 1 - p_2(y_2|x_2)$ in (14)). Under the assumption $X \not\to Y$, we have

$$\underline{P}_{X \not\to Y}(A) = \underline{P}_Y(\{y_2\}) = 0.4,$$
$$\underline{P}_{X \not\to Y}(B) = 0.33,$$

where $\underline{P}_{X \not\to Y}(A)$ value is due to the marginal preservation (Equation (8)) and $\underline{P}_{X \not\to Y}(B)$ value is reached by considering p^6 in Table 2. We then get

$$0.73 = \underline{P}_{X \not\to Y}(A) + \underline{P}_{X \not\to Y}(B) \geq \underline{P}_{X \not\to Y}(A \cup B) + \underline{P}_{X \not\to Y}(A \cap B) = 0.64.$$

Note that the assumption $Y \not\to X$ leads to $\underline{P}_{Y \not\to X}(B) = 0.27$, hence also resulting in an inequality violating 2-monotonicity.

Let us now look at the symmetric counter-part of epistemic irrelevance: epistemic independence [29]. It corresponds to the statements that X and Y are epistemically irrelevant of each other, and is denoted by $X \not\leftrightarrow Y$. Its extension to any number n of variables has recently been investigated by de Cooman *et al.* [8]. The corresponding joint lower expectation, denoted by $\underline{P}_{X \not\leftrightarrow Y}$, is such that, for any $f \in \mathscr{L}(\mathscr{X} \times \mathscr{Y})$,

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

As for epistemic irrelevance, epistemic independence preserves marginal information and factorizes over Cartesian products of events.

Example 5. Consider the same marginals and events A, B as in Example 3. . As $\underline{P}_{X \not\leftrightarrow Y}$ still factorizes over Cartesian products, its values on $A, A \cap B$ and $A \cup B$ do not change. Also, as $\mathscr{M}(\underline{P}_{X \not\leftrightarrow Y}) = \mathscr{M}(\underline{P}_{X \not\to Y}) \cap \mathscr{M}(\underline{P}_{Y \not\to X})$, we have that $\underline{P}_{X \not\leftrightarrow Y}(B) \geq \max\{\underline{P}_{Y \not\to X}(B), \underline{P}_{X \not\to Y}(B)\}$, hence $\underline{P}_{X \not\leftrightarrow Y}$ is also not 2-monotone.

Examples 3, 4 and 5 indicate that neither strong independence, epistemic irrelevance nor epistemic independence preserve the 2-monotonicity of marginals. The next question is to know whether or not we can find an easy-to-build 2-monotone approximation. In the next section, we explore a formal extension of random set independence to 2-monotone lower probabilities, and show that it does preserve 2-monotonicity.

5. Extending random set independence

The notion of random set independence [4] only applies to lower probabilities that are n-monotone for any number n, also said to be completely- or ∞-monotone. A lower probability $\underline{P}$ on a space $\mathscr{X}$ is a completely-monotone lower probability if and only if its Möbius inverse is non-negative [26], that is $m : \wp(\mathscr{X}) \to [0,1]$. In this case, the Möbius mass $m(E)$ can be interpreted as a probability bearing over sets E rather than singletons (hence it can be assimilated to a random set [9] or a belief function [26]).

Let m_X and m_Y be the two Möbius inverses obtained from completely-monotone lower probabilities $\underline{P}_X$ and $\underline{P}_Y$. Then the joint model $\underline{P}_{RI}$ obtained by an assumption of random set independence is defined as a the lower probability having for Möbius inverse the joint distribution $m_{RI} : \wp(\mathscr{X} \times \mathscr{Y}) \to [0,1]$ such that, for any $A \subseteq \mathscr{X}$ and any $B \subseteq \mathscr{Y}$,

$$m_{RI}(A \times B) = m_X(A) m_Y(B). \tag{16}$$

$\underline{P}_{RI}$ can then be estimated via Equation (4) (and as Möbius inverse is a bijective transformation, m_{RI} can then be retrieved by applying Equation (3)). Note that m_{RI} has only Cartesian products of events for focal elements. As m_{RI} is non-negative by definition, then $\underline{P}_{RI}$ is completely-monotone, meaning that the joint model preserves monotonicity in the case of random set independence. As for strong independence, epistemic irrelevance and epistemic independence, random set independence does satisfy marginal preservation and factorizes over Cartesian product of events. Similarly to strong independence, it is a symmetric notion.

Formally extending the notion of random set independence and Equation (16) to 2-monotone lower probabilities can be done in a simple way:

Definition 1 (Möbius independence)**.** Let $\underline{P}_X$, $\underline{P}_Y$ be 2-monotone lower probabilities defined on finite spaces $\mathscr{X}, \mathscr{Y}$ with m_X, m_Y their respective Möbius inverses. The joint model $\underline{P}_{MI}$ corresponding to a *Möbius independence* notion is defined as the lower probability having for Möbius inverse the distribution $m_{MI} : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ such that, for every $A \times B \subseteq \mathscr{X} \times \mathscr{Y}$,

$$m_{MI}(A \times B) = m_X(A)m_Y(B). \tag{17}$$

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

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

Note that, as Möbius inversion remains a bijection in the 2-monotone case, applying Equation (3) to $\underline{P}_{MI}$ gives back m_{MI}.

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. As $\underline{P}_{MI}$ is entirely defined by its Möbius inverse m_{MI}, proving that $\underline{P}_{MI}$ is 2-monotone comes down to show that m_{MI} has the following properties:

1. $m_{MI}(\emptyset) = 0$;
2. $\sum_{A \times B \subseteq \mathscr{X} \times \mathscr{Y}} m_{MI}(A \times B) = 1$ (as m_{MI} is non-null only on Cartesian products);
3. For any $E \subseteq \mathscr{X} \times \mathscr{Y}$ and all $(\{x_1\} \times \{y_1\} \cup \{x_2\} \times \{y_2\}) \subseteq E$,

   $$\sum_{(\{x_1\} \times \{y_1\} \cup \{x_2\} \times \{y_2\}) \subseteq C \subseteq E} m_{MI}(C) \geq 0$$

   holds (using Prop. 1). To simplify notation, we will denote $\{x, y\}_i := \{x_i\} \times \{y_i\}$ in this proof.

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

$$\begin{aligned}
\sum_{(\{x,y\}_1\cup\{x,y\}_2)\subseteq C\subseteq E} m_{MI}(C) &= \sum_{(\{x,y\}_1\cup\{x,y\}_2)\subseteq A'\times B'\subseteq E} m_{MI}(A'\times B') \\
&= \sum_{(\{x,y\}_1\cup\{x,y\}_2)\subseteq A'\times B'\subseteq E} m_X(A')m_Y(B') \\
&= \sum_{A'\times B'\subseteq E}\left(\sum_{(\{x_1\}\cup\{x_2\})\subseteq A'\subseteq A} m_X(A')\right)\left(\sum_{(\{y_1\}\cup\{y_2\})\subseteq B'\subseteq B} m_Y(B')\right) \geq 0.
\end{aligned}$$

The first equality follows from the fact that m_{MI} is non-null only on Cartesian products, and the second from Definition 1 of Möbius independence ($m_{MI}(A'\times B') = m_X(A')m_Y(B')$). Finally, the last inequality comes from the fact that both sums are positive (according to Prop. 1). □

Proposition 3 shows that by extending the notion of random set independence we get an independence notion (here understood as a formal means to build a joint model from marginals) that preserves 2-monotonicity. Since Equation (17) is symmetric, extending the notion to any number N of variables is straightforward. As we now have an easy way to get a 2-monotone approximation (that benefits from the computational advantages recalled in Section 2.3), we can explore its properties and links with respect to the other joint models described in Sections 3 and 4.

6. Properties of Möbius independence

An immediate remark is that, as Möbius independence is a direct extension of random independence, the cases where it is likely to be the most useful are those where the notion of random set independence is commonly used. This includes, in particular, problems of uncertainty propagation [15] or of reliability analysis [25].

There are no natural semantics to Möbius independence within frameworks based on lower expectations and lower probabilities, as the Möbius inverse itself has no semantic within such frameworks. Actually, Möbius independence cannot even directly benefits from a random set-like interpretation, due to the presence of focal sets receiving negative masses. Providing Möbius independence with a semantic interpretation would therefore require to give a proper interpretation to focal sets with negative masses, an open question that is out of the scope of the present paper (some attempts at such interpretation have been made in other approaches such as Smets Transferable Belief Model [27], but they hardly apply to the present case).

This is why we will concentrate on the formal properties of Möbius independence that link it to strong independence, epistemic irrelevance and epistemic independence.

6.1. Factorization over Cartesian products

All joint models studied in the previous sections factorize over Cartesian products of events, i.e., if $\underline{P}_X$, $\underline{P}_Y$ are the marginal models and $\underline{P}_\omega$ the joint model with $\omega \in \{SI, Y \not\to X, X \not\to Y, X \not\leftrightarrow Y, RI\}$, then for any $A\times B \subseteq \mathscr{X}\times\mathscr{Y}$, we have

$$\underline{P}_\omega(A\times B) = \underline{P}_X(A)\underline{P}_Y(B), \quad \overline{P}_\omega(A\times B) = \overline{P}_X(A)\overline{P}_Y(B).$$

The same holds for Möbius independence: let $\underline{P}_X$, $\underline{P}_Y$ be 2-monotone lower probabilities, and $\underline{P}_{MI}$ the joint model. Then for any $A \times B \subseteq \mathscr{X} \times \mathscr{Y}$

$$\begin{aligned}\underline{P}_{MI}(A \times B) = & \sum_{C \subseteq A \times B} m_{MI}(C) = \sum_{(A' \times B') \subseteq (A \times B)} m_X(A') m_Y(B') \\ = & \sum_{A' \subseteq A} m_X(A') \sum_{B' \subseteq B} m_Y(B') = \underline{P}_X(A) \underline{P}_Y(B)\end{aligned}$$

since, by construction, m_{MI} is non-null only on Cartesian products. Similar arguments can be used to prove $\overline{P}_{MI}(A \times B) = \overline{P}_X(A) \overline{P}_Y(B)$.

That $\underline{P}_{MI}$ as an approximation preserves marginal information (Equation (8)) is immediate, since

$$\underline{P}_{MI}(A \times \mathscr{Y}) = \underline{P}_X(A) \underline{P}_Y(\mathscr{Y}) = \underline{P}_X(A) \tag{18}$$

as $\underline{P}_Y(\mathscr{Y}) = 1$. The same argument applies to events $\mathscr{X} \times B$.

Let us now show that Möbius independence can be used as a conservative approximation of other independence notions.

6.2. Möbius independence as a conservative approximation

In this paper, a model is said to be a conservative approximation of another one if inferences made using the former are more cautious than inferences made using the latter. In practice, this translates in the fact that a credal set $\mathscr{M}'$ defined on a space $\mathscr{X}$ is a conservative approximation of another set $\mathscr{M}$ defined on the same space $\mathscr{X}$ if and only if $\mathscr{M} \subseteq \mathscr{M}'$, or equivalently if $\underline{P}' \leq \underline{P}$ with $\underline{P}'$ and $\underline{P}$ the lower expectations induced by $\mathscr{M}'$ and $\mathscr{M}$, respectively. We will also say that $\underline{P}'$ outer-approximate $\underline{P}$ (since for any function $f \in \mathscr{L}(\mathscr{X})$ we will have $[\underline{P}(f), \overline{P}(f)] \subseteq [\underline{P}'(f), \overline{P}'(f)]$).

In uncertainty reasoning, if an approximation is going to be used as a replacement of an exact but more complex model, being conservative ensures that inferences made using the approximation will be at least as cautious as inferences made using the exact model. Although this may make conclusions more imprecise, it ensures that these same conclusions will not be misleading (e.g., by providing overly precise results or bad estimates). This is particularly critical in applications such as risk analysis.

The next proposition shows that Möbius independence is a conservative approximation of independence notions studied in Sections 3 and 4.

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 $\underline{P}_{X \not\to Y}, \underline{P}_{Y \not\to X}, \underline{P}_{X \not\leftrightarrow Y}$, and $\underline{P}_{SI}$, 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 via the independence assumptions of Sections 3 and 4 are related in the following way [4]:

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

when the joint uncertainty models are obtained from the same marginals $\underline{P}_X, \underline{P}_Y$. Hence, it is sufficient to show that $\underline{P}_{MI} \leq \underline{P}_{X \not\to Y}$ and $\underline{P}_{MI} \leq \underline{P}_{Y \not\to X}$ to prove that $\underline{P}_{MI}$ outer-approximates the other joint uncertainty models. We will only prove that $\underline{P}_{MI} \leq \underline{P}_{X \not\to Y}$, as the proof of $\underline{P}_{MI} \leq \underline{P}_{Y \not\to X}$ follows similar arguments.

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. (6) with Eq. (15), 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. (6) 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

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

□

Proposition 4 and Equation (18) indicate that Möbius independence can be used as a conservative approximation while not losing information about the marginals. This is in contrast with other (simpler) conservative approximations of independence notions using models such as possibility distributions [12] or p-boxes [28], which use implies a partial loss of the marginal information.

One question that remains, though, is whether $\underline{P}_{MI}$ is the most specific 2-monotone outer-approximation of $\max\{\underline{P}_{X \not\to Y}, \underline{P}_{Y \not\to X}\}$, in the sense that there is no other model $\underline{P}$ that is 2-monotone and such that $\underline{P}_{MI} \leq \underline{P} \leq \max\{\underline{P}_{X \not\to Y}, \underline{P}_{Y \not\to X}\}$. The answer is no, as shown by the next example.

Example 6. Consider again the marginal models of Example 3. We already know that none of $\underline{P}_{X \not\to Y}, \underline{P}_{Y \not\to X}$ does satisfy the 2-monotonicity property. As all studied independence concepts of this paper factorizes over events, the only events on which the joint models may differ are $\{x_1\} \times \{y_1\} \cup \{x_2\} \times \{y_2\}$ and $\{x_1\} \times \{y_2\} \cup \{x_2\} \times \{y_1\}$. We have, using Möbius inverse independence, that

$$\underline{P}_{MI}(\{x_1\} \times \{y_1\} \cup \{x_2\} \times \{y_2\}) = 0.17 \text{ and } \underline{P}_{MI}(\{x_1\} \times \{y_2\} \cup \{x_2\} \times \{y_1\}) = 0.18,$$

however it can be checked that the most specific joint model $\underline{P}$ being 2-monotone and outer-approximating $\underline{P}_{X \not\to Y}, \underline{P}_{Y \not\to X}$ is such that $\underline{P}(\{x_1\} \times \{y_2\} \cup \{x_2\} \times \{y_1\}) = 0.23$ and $\underline{P}(\{x_1\} \times \{y_1\} \cup \{x_2\} \times \{y_2\}) = 0.24$.

It must be noted that obtaining such a "best" approximation (that is not unique in general [1]) would require to estimate $\underline{P}_{X \not\to Y}$ on every event of $\mathscr{X} \times \mathscr{Y}$ and then run some linear program [1] to find an approximation. In this case, the approximation would not longer be easy to get, therefore losing the main advantages of an approximated model.

Also, the joint model m_{MI} requires storing at most $2^{|\mathscr{X}|+|\mathscr{Y}|}$ values, as m_{MI} is non-null only on Cartesian products. This can be compared to the maximal number of $2^{|\mathscr{X}|\cdot|\mathscr{Y}|}$ values needed to store a generic 2-monotone lower probability, or to the maximal factorial number $(|\mathscr{X}|\cdot|\mathscr{Y}|)!$ of extreme points of the induced credal set.

Remark 1. As a side result, Example 6 (and the examples before) also answers to the following interesting question: given marginals $\underline{P}_X, \underline{P}_Y$, is it possible to find a 2-monotone joint model $\underline{P}$ that inner approximates the model obtained by an assumption of epistemic irrelevance and outer approximates the model obtained by an assumption of epistemic independence? That is, is it always possible to find a 2-monotone lower probability $\underline{P}$ such that $\max\{\underline{P}_{X\not\to Y}, \underline{P}_{Y\not\to X}\} \leq \underline{P} \leq \underline{P}_{X\not\leftrightarrow Y}$? The answer is no, since if we consider marginal models and events of Example 3, lower probabilities on events $A, A\cap B$ and $A\cup B$ are the same for all independence assumptions and all approximations $\underline{P}$ would be such that $\underline{P}_{X\not\leftrightarrow Y}(B) \geq \underline{P}(B) \geq \max\{\underline{P}_{X\not\to Y}(B), \underline{P}_{Y\not\to X}(B)\}$. Therefore $\underline{P}$ would not be 2-monotone.

In summary, Möbius independence provides an easy means to build a (not too) conservative approximation of the usual independence assumptions for imprecise probabilities.

6.3. Möbius independence and productivity

Within the theory of lower prevision, recent works [8] have focused on characterizing other extensions of stochastic independence in the form of properties that a joint model could satisfy. One of the weakest property developed in these works is the one of productivity, from which very general laws of large numbers can be derived [3]. To simplify notations, we identify in this section a function g defined on the 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 similarly identify functions defined on the space $\mathscr{Y}$. The property of productivity for two variables is then defined as follows:

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 non-negative $f\in\mathscr{L}(\mathscr{X})$), $\underline{P}(f[g-\underline{P}(g)]) \geq 0$.

Intuitively, this property states that a random variable with positive expected value ($g-\underline{P}(g)$ has a lower expectation equal to zero) multiplied by a positive value (f) should have a positive expected value. Unfortunately, the next example shows that the joint uncertainty model $\underline{P}_{MI}$ obtained under an MI assumption does not satisfy this property.

Example 7. 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

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

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

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 $h = f[g - \underline{P}_{MI}(g)]$ on $\mathscr{X} \times \mathscr{Y}$ is summarised in Table 3 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 3: Function $f[g - \underline{P}(g)]$ of Example 7

inequalities in Table 3 are due to the two inequalities $a \le b$ and $0 \le c \le d$ and to the fact that $(a-b) \le 0$, $(1-\alpha_2) \ge 0$. Note that the four values are totally ordered. Using Eq. (6) and Definition 1, we have that

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

If we assume that $0 < \alpha_2 < 1$, then this value is negative (as $b-a > 0$ and $c-d < 0$), unless $((1-\beta_2-\beta_1) = 0$, that is unless $\underline{P}_Y$ is a precise probability. If we extend these conclusions to all possible f and g satisfying Definition 2, this means that $\underline{P}_{MI}(f[g - \underline{P}_{MI}(g)]) \ge 0$ 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 independence (and also of random set independence) to satisfy productivity as well as other stronger properties that imply productivity. This unfortunately means that results from Ref. [6] do not apply to Möbius independence nor with random set independence.

However, it should be noted that the notion random set independence (of which Möbius independence is a direct extension) has been used in graphical models [34] as well as basic assumptions to prove laws of large numbers [19]. This means that although Möbius independence cannot use recent results concerning lower expectations, it may still be useful if considered as an extension of random set independence. This direction is not explored here, as our main focus is the relation between Möbius independence and lower expectations.

7. An illustrative example

In this section, we provide a simple example of the use of Möbius independence in the computation of natural extension. This example can be seen as a particular instance of an uncertainty propagation problem, where one searches to estimate the output uncertainty from the uncertainty on input variables. The problem considered in this example is one of multi-criteria decision-making, where input variables are criteria and the function through which they are propagated is an aggregation function (i.e., a weighted mean).

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 [5] 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 can be imprecise (hence, frequencies can be given to sets 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)

$$\underline{P}_{MI}(f) = 1.74 \quad ; \quad \overline{P}_{MI}(f) = -\underline{P}_{MI}(-f) = 2.62.$$

We can compare the results to the bounds obtained from an assumption of strong independence between X, Y, Z, which are

$$\underline{P}_{SI}(f) = 1.74 \quad ; \quad \overline{P}_{SI}(f) = -\underline{P}_{SI}(-f) = 2.62.$$

In this case, we have equality between the two inferences. We may conjecture that this is due to the fact that f is non-decreasing in each variable X, Y and Z. So, in this case, the same conclusion can be computed more efficiently using Möbius independence, as evaluating $\underline{P}_{MI}(f)$ and $\overline{P}_{MI}(f)$ is simpler than evaluating $\underline{P}_{SI}(f)$ and $\overline{P}_{SI}(f)$.

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.

8. Conclusions

In this paper, we have first explored whether the most usual independence notions that can be found in the imprecise probabilistic literature preserve the 2-monotonicity of marginal models. The main interest of knowing the answer to such a question is practical, as 2-monotonicity is before all a convenient property that allows for more efficient computations.

As the answer to this question is negative, we have first shown that by extending the notion of random set independence to 2-monotone lower probabilities, we get a 2-monotone joint model that is easy to build. We call Möbius independence the corresponding notion (as it consists in taking the product of the Möbius inverses of the marginal models). We have also explored the properties of this model.

On the upside, we have shown that it can be used as a conservative approximation of the other independence notions, and therefore can be useful in applications where computational tractability and cautiousness of inferences are important issues. This is typically the case in risk and reliability studies, or in application requiring some robustness. Other areas where it could be useful is those where random set independence plays an important role, as Möbius independence is a direct extension of it.

On the downside, it appears that recent works concerning lower expectations and independence notions do not apply to this approximation (nor to classical random set independence), as Möbius independence fails to satisfy the property of productivity. Also, providing Möbius independence with a proper semantic appears difficult, as it would require to give an interpretation to focal sets with negative masses (an open issue).

Acknowledgements

I wish to thank Matthias Troffaes and Enrique Miranda, the counter-examples of Sections 3 and 3 being the results of discussions with them, as well as Cassio P. de

Campos whose questions are the reasons for Remark 1. I also wish to thank the two anonymous reviewers whose comments have allowed me to significantly improve the clarity of this paper.

References

[1] A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In *ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications*, pages 61–70. SIPTA, 2009.

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

[3] G. D. Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 138(8):2409 – 2432, 2008.

[4] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

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

[6] G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal nets: The case of imprecise markov trees. *Int. J. Approximate Reasoning*, 51(9):1029–1052, 2010.

[7] G. de Cooman and E. Miranda. Forward irrelevance. *Journal of Statistical Planning and Inference*, 139:256–276, 2009.

[8] G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. *Artif. Intell.*, 175(12-13):1911–1950, 2011.

[9] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[10] S. Destercke. Independence and 2-monotonicity: Nice to have, hard to keep. In *ECSQARU*, pages 263–274, 2011.

[11] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I generalized p-boxes. *Int. J. of Approximate Reasoning*, 49:649–663, 2008.

[12] S. Destercke, D. Dubois, and E. Chojnacki. A consonant approximation of the product of independent consonant random sets. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17(6):773–792, 2009.

[13] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[14] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.

[15] T. Fetz and M. Oberguggenberger. Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliability Engineering and System Safety*, 85:73–87, 2004.

[16] B. Finetti. *Theory of probability*, volume 1-2. Wiley, NY, 1974. Translation of 1970 book.

[17] I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory*, 59:33–49, 1993.

[18] P. J. Huber and V. Strassen. Minimax Tests and the Neyman-Pearson Lemma for Capacities. *The Annals of Statistics*, 1(2):251–263, 1973.

[19] S. Li, Y. Ogura, and V. Kreinovich. *Limit theorems and applications of set-valued and fuzzy set-valued random*. Springer, 2002.

[20] E. Miranda. A survey of the theory of coherent lower previsions. *Int. J. of Approximate Reasoning*, 48:628–658, 2008.

[21] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *Int. J. of Approximate Reasoning*, 33:95–115, 2003.

[22] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *Int. J. of Approximate Reasoning*, 51(9):1145–1158, 2010.

[23] A. Piatti, A. Antonucci, and M. Zaffalon. Building knowledge-based systems by credal networks: a tutorial. *Advances in Mathematics Research*, 11, 2010.

[24] E. Quaeghebeur and G. D. Cooman. Extreme lower probabilities. *Fuzzy Sets and Systems*, 159:2193–2175, 2008.

[25] M. Sallak, W. Schön, and F. Aguirre. Transferable belief model for reliability analysis of systems with data uncertainties and failure dependencies. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 224(4):266–278, 2010.

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

[27] P. Smets. The canonical decomposition of a weighted belief. In *Proc. Int. Joint. Conf. on Artificial Intelligence*, pages 1896–1901, Montreal, 1995.

[28] M. C. M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *Int. J. Approximate Reasoning*, 52(6):767–791, 2011.

[29] P. Vicig. Epistemic independence for imprecise probabilities. *Int. J. of Approximate Reasoning*, 24:235–250, 2000.

[30] P. Walley. Coherent lower (and upper) probabilities. Technical report, University of Warwick, department of Statistics, Warwick, 1981.

[31] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.

[32] P. Walley. Measures of uncertainty in expert systems. *Artifical Intelligence*, 83:1–58, 1996.

[33] P. M. Williams. Notes on conditional previsions. *Int. J. of Approximate Reasoning*, 44:366–383, 2007.

[34] H. Xu and P. Smets. Reasoning in evidential networks with conditional belief functions. *Int. J. of Approximate Reasoning*, 14(2-3):155–185, 1996.