1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June – 2 July 1999

Epistemic Independence for Imprecise Probabilities

Paolo Vicig
University of Trieste, Italy
paolov@econ.univ.trieste.it

Abstract

The aim of this paper is that of studying a notion of independence for imprecise probabilities which is essentially based on the intuitive meaning of this concept. This is expressed, in the case of two events, by the reciprocal irrelevance of the knowledge of the value of each event for evaluating the other one, and has been termed epistemic independence. In order to consider more general situations in the framework of coherent imprecise probabilities, a definition of (epistemic) independence is introduced referring to arbitrary sets of logically independent partitions. Logical independence is viewed as a natural prerequisite for epistemic independence. It is then proved that the definition is always consistent, its relationship with the factorization rule is analysed, and some of its more relevant implications are discussed.

Keywords. Coherent imprecise probabilities, (epistemic) independence, logical independence, factorization.

1 Introduction

Let A, B be two non-impossible events and m a measure of uncertainty on $\{A, B\}$. Then it is quite intuitive to call B *irrelevant* to A (with respect to the given m) if the knowledge of the logical value of B (either that B is true or that B is false) does not alter our evaluation on the uncertainty of A, which remains still equal to $m(A)$. A natural way of introducing independence between A and B (under m) is then that of identifying it with the mutual irrelevance of A to B and of B to A. When m is a coherent imprecise probability, this is the notion of *epistemic independence* for two events given by Walley in [11], where the concept is also generalised to two partitions or two independent experiments.

The main purpose of this paper is to extend the concept of epistemic independence, using the definition 2.3.2 of coherence due to Williams, to an arbitrary family of (finite or infinite) logically independent partitions of the certain event Ω (def. 3.1), to prove its consistency, i.e. to show that there *always* exist (coherent) imprecise probabilities that satisfy the conditions required by this definition (theor. 3.4), and to discuss some of its implications in sects. 3.6, 4, 5.

Operationally, the usual approach to (unconditional or conditional) independence under a given uncertainty measure aims at establishing some algebraic condition, like the factorization property, which greatly simplifies the computations and the underlying models in most applications, from sampling to bayesian networks.

However this approach and that based on irrelevance do not perfectly overlap. This does not even happen in the simplest case when m is a precise probability measure P: for instance, identifying independence of A, B with the property $P(A \wedge B) = P(A)P(B)$ would imply that A is independent from itself if $P(A) = 1$, A and B are at any rate independent if $P(A) = 0$ or $P(B) = 0$ (other interesting examples may be found in [1], [2], [4] p. 274, [11]). Things may become more complicated with other uncertainty measures (see, for instance, [3], [5]). In particular, several concepts of independence have been proposed for imprecise probabilities, among them the sensitivity analysis approach (we postpone a comment on it to sect. 3.5) and those in [3].

In our framework, (epistemic) independence necessarily implies, but is not equivalent to, a 'weak' factorization property (prop. 3.6.1); anyway 'strong' factorization can always be imposed, has a special probabilistic meaning and may be a reasonable choice on many occasions, as shown in sect. 3.6.3.

Another question is what does the proposed independence condition imply when extending the imprecise probability to larger sets of events, in particular in terms of independence propagation. Although a full understanding of the matter requires further investigation, some special interesting cases are

discussed in sects. 4.1.2, 4.1.3 and 4.2.

2 Preliminaries

We recall in this section the basic definitions and well-known results which will be used in the sequel.

2.1 Notation

If A is an event, A^c is its negation, $I(A)$ is the indicator function of A (i.e. $I(A) = 1$ if A is true, $I(A) = 0$ if A is false).

A' indicates an event which may be either A or A^c; any expression including A' (for instance (i) and (ii) at the beginning of sect. 3) is meant to hold when replacing A' with A or A^c.

Ω is the certain event, $\varnothing$ the impossible one. Following the logical interpretation of events, we write $\wedge$, $\vee$, $\Rightarrow$ to denote the logical product, logical sum and implication of events (corresponding, respectively, to $\cap$, $\cup$, and $\in$ or $\subset$ in the set-theoretic approach).

2.2 Partitions and logical independence

A *partition* $\mathcal{P}$ of the certain event Ω is an exhaustive set of pairwise disjoint events, called *constituents*, which might in general include also $\varnothing$. The non-impossible constituents of $\mathcal{P}$ are called *atoms*. Defining $\mathcal{P}^{\varnothing} = \mathcal{P} - \{\varnothing\}$, partition $\mathcal{P}^{\varnothing}$ is made up of atoms only. None of the partitions to be considered in the sequel will be the null partition $\mathcal{P}_0 = \{\Omega\}$.

Denote with $\mathcal{A}(\mathcal{P})$ the *algebra of the events logically dependent on* $\mathcal{P}$, i.e. the set formed by the events which are logical sums of constituents of $\mathcal{P}$ ($\mathcal{A}(\mathcal{P})$ is the power set of $\mathcal{P}$ in set-theory language); define also $\mathcal{A}^{\varnothing}(\mathcal{P}) = \mathcal{A}(\mathcal{P}) - \{\varnothing\}$.

Given an arbitrary set of partitions S, the *product partition* $\wedge S$ of the partitions of S is the partition whose constituents are the logical products obtained choosing as operands a constituent for every partition of S in all possible ways. Putting $S^{\varnothing} = \{\mathcal{P}^{\varnothing}: \mathcal{P} \in S\}$, $\wedge S^{\varnothing}$ is the product partition of the partitions of S, all leaving out $\varnothing$. For instance, if $S = \{\mathcal{P}_1,\ldots, \mathcal{P}_n\}$, $\mathcal{P}_i = \{A_i, A_i^c\}$, $A_i \neq \varnothing$, $A_i \neq \Omega$, i = 1,…,n, any constituent of $\wedge S^{\varnothing} = \wedge \mathcal{P}_i^{\varnothing}$ (= $\wedge \mathcal{P}_i^{\varnothing}$, in this case) may be written in the form A_1' $\wedge\ldots\wedge$ A_n'; some constituents of $\wedge S^{\varnothing}$ may in general be impossible.

The partitions of S are *logically independent* (in short, lg. i.) if the constituents of $\wedge S^{\varnothing}$ are all possible. This implies in particular that choosing arbitrarily partitions $\mathcal{P}_1,\ldots, \mathcal{P}_n$ from S, it is always $A_1 \wedge\ldots\wedge A_n \neq \varnothing$, $\forall A_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i)$, i = 1,…,n.

Logical independence is a concept which does not require any uncertainty evaluation, but should be a prerequisite for notions of independence involving uncertainty measures (see sect. 3.1.1).

2.3 Precise and imprecise coherent probabilities

We shall denote with $\mathcal{E}$ an arbitrary (finite or infinite) non-empty set of conditional events. Then, following [7]

2.3.1 Definition

$P(\cdot|\cdot)$ is a *coherent conditional probability* on $\mathcal{E}$ iff, $\forall$ m, $\forall A_i|B_i \in \mathcal{E}$, $\forall s_i \in \Re$, i = 1,…,m, defining

$$G = \sum_{i=1}^{m} s_i\, I(B_i)[I(A_i) - P(A_i|B_i)] \text{ and } B = \vee_{i=1}^{m} B_i,$$

it is true that $\max G|B \geq 0$.

The coherence concepts for precise and imprecise probabilities we recall here require a conditional random number ($G|B$ in def. 2.3.1, $\underline{G}|B$ in the next def. 2.3.2) not to be strictly negative. G can be interpreted as the gain obtained from an arbitrary finite number of bets, each on a conditional event in $\mathcal{E}$, and conditioning (G or $\underline{G}$) on B takes account of the fact that the bet on $A_i|B_i$ is called off iff $I(B_i) = 0$, so that it is meaningful to evaluate the gains only for $I(B) = 1$; in fact $I(B) = 0$ identifies the situation where no bet actually takes place.

It is possible in def. 2.3.1 to bet both in favour ($s_i > 0$) and against ($s_i < 0$) any $A_i|B_i$, whilst coherence for imprecise lower probabilities (def. 2.3.2) allows betting against at most one event, for each choice of the distinct events $A_i|B_i \in \mathcal{E}$ appearing in the expression of the gain $\underline{G}$.

2.3.2 Definition

$\underline{P}(\cdot|\cdot)$ is a *coherent lower probability* on $\mathcal{E}$ iff, $\forall$ m, $\forall A_i|B_i \in \mathcal{E}$, $\forall s_i \geq 0$, i = 0,…,m, defining

$$\underline{G} = \sum_{i=1}^{m} s_i\, I(B_i)[I(A_i) - \underline{P}(A_i|B_i)] - s_0 I(B_0)[I(A_0) - \underline{P}(A_0|B_0)]$$

and $B = \vee_{i=0}^{m} B_i$, it is true that $\max \underline{G}|B \geq 0$.

This is the coherence notion by Williams [13], which coincides with Walley's definition of coherence in [11], sect. 7.1.4 (b), if $\mathcal{E}$ is finite, is weaker if $\mathcal{E}$ is infinite (hence not all of the results in the sequel still apply with Walley's definition). See also [6], [10], [11] for other concepts of coherence with imprecise probabilities.

Upper probabilities $\overline{P}(\cdot|\cdot)$ are customarily related to lower probabilities by the conjugacy relation

$$(1) \qquad \overline{P}(A|B) = 1 - \underline{P}(A^c|B)$$

which we assume to hold throughout; hence it is sufficient to refer to, for instance, lower probabilities only.

A *precise probability* $P(\cdot|\cdot)$ on $\mathcal{E}$ is the special case of imprecise probability where $P(A|B) = \underline{P}(A|B) = \overline{P}(A|B)$ for all events in $\mathcal{E}$ (while in general $\underline{P}(A|B) \leq \overline{P}(A|B)$).

Further, coherent lower probabilities can be characterized as lower envelopes of coherent precise probabilities [13]:

2.3.3 Lower envelope theorem

$\underline{P}(\cdot|\cdot)$ is a coherent lower probability on $\mathcal{E}$ iff there exists a non-empty set M of coherent conditional probabilities on $\mathcal{E}$ such that

(2) $\underline{P}(A|B) = \min_{P \in M} \{P(A|B)\}, \forall A|B \in \mathcal{E}$.

The lower envelope theorem gives in particular an indirect but useful way of assessing coherent lower probabilities as lower envelopes of precise probabilities. $\underline{P}$ is called the *lower envelope* of M, and $\forall P \in M$, it is $P \geq \underline{P}$, or in words P *dominates* $\underline{P}$.

2.3.4 Extensions of probabilities

Coherent lower or precise probabilities on $\mathcal{E}$ can be extended to any $\mathcal{E}' \supset \mathcal{E}$, i.e. if $\underline{P}$ (P) is a coherent lower (precise) probability on $\mathcal{E}$, then $\forall \mathcal{E}' \supset \mathcal{E}$ there exists $\underline{P}'$ (P') which is coherent on $\mathcal{E}'$ and coincides with $\underline{P}$ (with P) on $\mathcal{E}$ (*extension theorem* [7], [13]).

Generally, there is no unique way of extending $\underline{P}$ from $\mathcal{E}$ to $\mathcal{E}'$, but there always exists a *least-committal* coherent extension $\underline{P}^*$ which is such that every other coherent extension $\underline{P}'$ on $\mathcal{E}'$ dominates it, i.e. $\underline{P}' \geq \underline{P}^*$.

$\underline{P}^*$ is the *natural extension* in the terminology of [11] (actually the concept of natural extension is wider than the one recalled here, which is at any rate sufficient for the purposes of this paper).

A different notion – which will be also employed in the sequel – is that of *common extension* of imprecise (in particular, also precise) probabilities. It is used to denote a mapping $\underline{P}$ defined on a set $S^* \supseteq \cup S_i$ and such that the restriction $\underline{P}|_{S_i}$ of $\underline{P}$ on each S_i is a coherent lower probability. In general, separate coherence of the lower probabilities $\underline{P}|_{S_i}$ (each on its domain S_i) does not imply coherence of their common extension $\underline{P}$ on S^* and not even on $\cup S_i$, apart from some special cases. One such case is considered in lemma 3.3.

Coherent imprecise probabilities are quite a general tool in uncertainty reasoning, and include several common approaches to uncertainty as special cases (see [11], [12] for a wide discussion). Note further that the set $\mathcal{E}$ is completely arbitrary, which ensures a great flexibility too.

It is also important to recall that if $\{A|B, B\} \subset \mathcal{E}$, the assignment $\underline{P}(B) = 0$ is in general not ruled out a priori, i.e. it may be possible to condition on events whose lower probability is zero. This fact should not be particularly unusual in practice: for instance, recall that under a sensitivity analysis interpretation the assessment $\underline{P}(B) = 0$, $\overline{P}(B) = \alpha > 0$ means that the assessor's feeling about the 'correct' probability of B is just that it cannot be more than α, but his information is lower vacuous (entirely vacuous if also $\alpha = 1$).

3 Independent partitions

Two events A, B (A, $B \neq \varnothing$, A, $B \neq \Omega$) are called in [11] *(epistemically) independent*

(i) if $\underline{P}(A|B') = \underline{P}(A)$, $\overline{P}(A|B') = \overline{P}(A)$ (*irrelevance* of B to A) and $\underline{P}(B|A') = \underline{P}(B)$, $\overline{P}(B|A') = \overline{P}(B)$ (*irrelevance* of A to B); or also

(ii) if $\underline{P}(A'|B') = \underline{P}(A')$, $\underline{P}(B'|A') = \underline{P}(B')$.

Conditions (i) and (ii) are equivalent, assuming that (1) holds.

We can generalise this definition to an arbitrary number of partitions, each with an arbitrary cardinality, in the following way:

3.1 Definition

Let S be a set of logically independent (lg. i.) partitions, define the set of events

(3) $\mathcal{C} = \{E_1|E_2 \wedge \ldots \wedge E_n: E_1 \in \mathcal{A}(\mathcal{P}_1), E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i), \mathcal{P}_i \neq \mathcal{P}_1, i = 2,\ldots,n, \{\mathcal{P}_1,\ldots,\mathcal{P}_n\} \subset S\}$,

and let $\underline{P}$ be a lower probability on $\mathcal{C}$.

The partitions of S are *epistemically independent*, or simply independent, iff

(4) $\underline{P}(E_1|E_2 \wedge \ldots \wedge E_n) = \underline{P}(E_1), \forall E_1|E_2 \wedge \ldots \wedge E_n \in \mathcal{C}$.

In words, the definition requires that the lower probability of any event E_1 logically dependent on an arbitrarily chosen partition in S is independent of the knowledge of the truth values of any finite logical product of events logically dependent on partitions in S other than that of E_1.

3.1.1 The logical independence assumption

We obtain in particular (ii) from def. 3.1 putting $S = \{\mathcal{P}_a, \mathcal{P}_b\}$, $\mathcal{P}_a = \{A, A^c\}$, $\mathcal{P}_b = \{B, B^c\}$. Note that logical independence of $\mathcal{P}_a$ and $\mathcal{P}_b$ is essential to apply (ii) being capable of evaluating *freely* A' and B'. In fact, if for instance one constituent of $\mathcal{P}_a \wedge \mathcal{P}_b$ is impossible, let it be $A \wedge B$, it is necessary (to ensure that both the independence conditions in (ii) and the coherence requirements $\underline{P}(A|B) = \underline{P}(B|A) = 0$, $\underline{P}(A^c|B) = \underline{P}(B^c|A) = 1$ may hold) to evaluate $\underline{P}(A) = \underline{P}(B) = 0$, $\underline{P}(A^c) = \underline{P}(B^c) = 1$, which means by (1) that both A and B must be given *precise* probability zero.

More generally, the logical independence assumption in def. 3.1 is motivated by the fact that if it is not assumed either it may be impossible to meet conditions (4) or very strong constraints on the values of $\underline{P}$ must be assumed.

See [1], [2], where def. 3.1 has been introduced referring to precise probabilities, for a wide discussion on this point (many of the arguments there can be easily transposed to imprecise probabilities).

Logical independence is regarded as an essential prerequisite for independence also in [4], [9].

In particular, logical independence for 'frames of discernment' (i.e. finite partitions) is required by Shafer in [9], pp. 127-129, for both evidential and cognitive independence of support functions.

The first question now is to make sure that def. 3.1 is not vacuous, i.e. that there exists a coherent lower probability on $\mathscr{C}$ which realizes conditions (4). It will be useful for this the following result, proved in [1] or [2]:

3.2 Theorem

Let S be a set of lg. i. partitions, define $\mathscr{C}$ as in (3), and assign, for every $\mathscr{P} \in S$, a coherent (precise) probability P on $\mathscr{A}(\mathscr{P})$. Then the common extension on $\mathscr{C}$ of the probabilities on each $\mathscr{A}(\mathscr{P})$, defined for all events in $\mathscr{C}$ by $P(E_1 \mid E_2 \wedge \ldots \wedge E_n) = P(E_1)$, is a coherent precise probability. Further, P has a unique extension on $\mathscr{C} \cup \Pi$, where Π is defined by

(5) $\Pi = \{E_1 \wedge \ldots \wedge E_n : E_i \in \mathscr{A}(\mathscr{P}_i),\ i = 1,\ldots,n,\ \{\mathscr{P}_1,\ldots,\mathscr{P}_n\} \subset S\}$;

the extension is obtained by factorization, i.e. putting, for every $E_1 \wedge \ldots \wedge E_n \in \Pi$,

(6) $P(E_1 \wedge \ldots \wedge E_n) = P(E_1)\ldots P(E_n)$.

We shall also need the following lemma 3.3, whose proof can be easily achieved exploiting defs. 2.3.1 or 2.3.2 and the logical independence assumption (the part concerning precise probabilities is also implied by thm. 3.2).

Note that lemma 3.3 is of some interest in itself, because it concerns a case where logical independence makes it possible to obtain an aggregate uncertainty evaluation simply by grouping together separate evaluations.

3.3 Lemma

Let S be a set of lg. i. partitions and assign, for every $\mathscr{P} \in S$, a lower (precise) probability coherent on $\mathscr{A}(\mathscr{P})$. Then the common extension $\underline{P}$ (P) on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$ of the lower (precise) probabilities given on each $\mathscr{A}(\mathscr{P})$ is a lower (precise) probability coherent on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$.

We prove now that:

3.4 Consistency theorem

Let S be a set of lg. i. partitions, define $\mathscr{C}$ as in (3), and assign, for every $\mathscr{P} \in S$, a lower probability $\underline{P}$ coherent on $\mathscr{A}(\mathscr{P})$. Then:

(a) the mapping defined on $\mathscr{C}$ by putting, for each $E_1 \mid E_2 \wedge \ldots \wedge E_n \in \mathscr{C}$, $\underline{P}(E_1 \mid E_2 \wedge \ldots \wedge E_n) = \underline{P}(E_1)$, is a coherent lower probability;

(b) given $\underline{P}$ on $\mathscr{C}$ as in (a) and defining Π as in (5), there exists a coherent extension of $\underline{P}$ on $\mathscr{C} \cup \Pi$ (also named $\underline{P}$) such that, for all events in Π,

(7) $\underline{P}(E_1 \wedge \ldots \wedge E_n) = \underline{P}(E_1)\ldots\underline{P}(E_n)$.

Proof. (a) By lemma 3.3, the common extension $\underline{P}$ on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$ of the given lower probabilities is coherent on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$.

Define the set $M_0 = \{P: P$ is a coherent precise probability on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$, $P \geq \underline{P}$ on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})\}$, which is non-empty as a consequence of thm. 2.3.3. More precisely, a generic $P \in M_0$ is obtained defining for each $\mathscr{P} \in S$ a probability which dominates $\underline{P}$ on $\mathscr{A}(\mathscr{P})$ and naming P the common extension on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$ of these probabilities. By varying in all admissible ways the choice of the dominating probability on each $\mathscr{A}(\mathscr{P})$, we get all the probabilities in M_0 (each one is coherent by lemma 3.3).

Extend now every $P \in M_0$ on $\mathscr{C}$ by putting $P(E_1 \mid E_2 \wedge \ldots \wedge E_n) = P(E_1)$ for every $E_1 \mid E_2 \wedge \ldots \wedge E_n \in \mathscr{C}$. By thm. 3.2, every such extension is a coherent probability on $\mathscr{C}$.

Call M_1 the set of these extensions. By thm. 2.3.3, the mapping $\underline{P}$ defined on every event in $\mathscr{C}$ by

$$\underline{P}(\cdot|\cdot) = \min_{P \in M_1} \{\underline{P}(\cdot|\cdot)\}$$

is a coherent lower probability on $\mathscr{C}$.

But clearly $\underline{P}$ is an extension on $\mathscr{C}$ of the coherent lower probability previously defined on $\cup_{\mathscr{P} \in S} \mathscr{A}(\mathscr{P})$ ($\subset \mathscr{C}$), and further, by thm. 2.3.3 and by construction,

$$\underline{P}(E_1 \mid E_2 \wedge \ldots \wedge E_n) = \min_{P \in M_1} \{P(E_1 \mid E_2 \wedge \ldots \wedge E_n)\} =$$
$$\min_{P \in M_1} \{P(E_1)\} = \min_{P \in M_0} \{P(E_1)\} = \underline{P}(E_1).$$

(b) By construction, the hypotheses of thm. 3.2 hold for every $P \in M_1$, which then has a unique coherent extension on $\mathscr{C} \cup \Pi$, given for every $E_1 \wedge \ldots \wedge E_n \in \Pi$ by:

(8) $P(E_1 \wedge \ldots \wedge E_n) = P(E_1)\ldots P(E_n) \geq \underline{P}(E_1)\ldots\underline{P}(E_n)$.

Call M_2 the set formed by the extensions on $\mathscr{C} \cup \Pi$ of all $P \in M_1$. Now consider an arbitrary $E_1 \wedge \ldots \wedge E_n \in \Pi$ and note that there exists $P^* \in M_2$ such that $P^*(E_i) = \underline{P}(E_i)$, $i = 1,\ldots,n$ (for every $\mathscr{A}(\mathscr{P})$, choose among the

probabilities dominating $\underline{P}$ on $\mathcal{A}(\mathcal{P})$ one such that $P^*(E_i) = \underline{P}(E_i)$ (2.3.3), note that the common extension of these P^* on $\cup_{\mathcal{P} \in S} \mathcal{A}(\mathcal{P})$ belongs to M_0, then follow the procedure in (a) and above to extend it to M_1 and M_2). Then it is:

$P^*(E_1 \wedge \ldots \wedge E_n) = \underline{P}(E_1) \ldots \underline{P}(E_n)$.

From the arbitrariness of $E_1 \wedge \ldots \wedge E_n$ and recalling also (8) it appears that the lower probability which satisfies (7) is the lower envelope of M_2. ■

3.5 Independent envelopes and independence

The sets M_1, M_2 defined in the proof of thm. 3.4 are instances of sets whose elements (precise probabilities) all realize independence for the partitions of S in the sense of def. 3.1, applied to precise probabilities. Such sets may be called, following [11], *independent envelopes*. Their use is instrumental in the proof of 3.4 as well as, often, in building up independent lower probabilities.

A common way of introducing independence for imprecise probabilities, the *sensitivity analysis* approach, defines an independent lower probability as the lower envelope of a set of independent precise probabilities. Anyway we recall that the set of *all* precise probabilities dominating an independent lower probability $\underline{P}$ is usually not an independent envelope, not even in the simplest cases ([10], lemma 9.1). In other words, independence for a lower probability does not imply independence for all its dominating precise probabilities.

It is not even true that lack of independence for $\underline{P}$ implies that none of its dominating precise probabilities realizes independence. In fact, under quite general conditions, to achieve the *dilation* condition $\underline{P}(A \mid B') \leq \underline{P}(A) \leq \overline{P}(A) \leq \overline{P}(A \mid B')$ (which, if the first or last inequality is strict, is incompatible with irrelevance of B to A and hence with independence of A and B) it is necessary that an unconditional probability P exists, which dominates $\underline{P}$ (and is dominated by $\overline{P}$) and is such that $P(A \wedge B) = P(A)P(B)$ (see [8]). But this implies, if $0 < P(A)P(B) < 1$, that A and B are independent under P.

3.6 The factorization property

The factorization condition (6) on the events of the set Π is a necessary, but not sufficient, condition for independence under a precise probability [2]. The following result holds for imprecise probabilities:

3.6.1 Proposition

Let S be a set of lg. i. partitions, and $\underline{P}$ a coherent lower probability on $\mathcal{C} \cup \Pi$. A necessary condition for $\underline{P}$ to realize independence of the partitions of S by def. 3.1 is that for every $E_1 \wedge \ldots \wedge E_n \in \Pi$ the following inequality holds:

(9) $\underline{P}(E_1 \wedge \ldots \wedge E_n) \geq \underline{P}(E_1) \ldots \underline{P}(E_n)$

(*weak factorization*).

Proof. The following condition is necessary for coherence of lower probabilities (supposing they are defined on the events of interest):

(10) $\underline{P}(E_1 \wedge \ldots \wedge E_n) \geq \underline{P}(E_1)\underline{P}(E_2 \mid E_1) \ldots \underline{P}(E_n \mid E_1 \wedge \ldots \wedge E_{n-1})$;

it is easy to prove (10) for n = 2 (applying def. 2.3.2 to the particular gain $\underline{G}$ where m = 2, $A_1 \mid B_1 = E_1 \mid \Omega = E_1$, $A_2 \mid B_2 = E_2 \mid E_1$, $A_0 \mid B_0 = E_1 \wedge E_2$, $s_1 = \underline{P}(E_2 \mid E_1)$, $s_2 = 1$, $s_0 = -1$) and by induction for n > 2. Then (9) follows from (10) and (4). ■

3.6.2 Note

Under hypotheses analogous to those of 3.6.1, the condition corresponding to (9) for upper probabilities is:

(9') $\overline{P}(E_1 \wedge \ldots \wedge E_n) \leq \overline{P}(E_1)\overline{P}(E_2) \ldots \overline{P}(E_n)$.

3.6.3 Factorization and independence

Inequalities (9) in prop. 3.6.1 impose only a *weak factorization* condition as necessary for epistemic independence of a set of partitions. Anyway, by thm. 3.4 (b), it is always possible to apply *strict factorization* – replacing ≥ with = in (9) – to extend on $\mathcal{C} \cup \Pi$ a lower probability which satisfies the independence conditions (4) in def. 3.1, therefore reaching the lower bound in (9) for lower probability assignments on every event in Π. This implies that:

(a) strict factorization has an important probabilistic meaning: the extension of $\underline{P}$ from $\mathcal{C}$ to $\mathcal{C} \cup \Pi$ by strict factorization is the *natural extension* of $\underline{P}$ on $\mathcal{C} \cup \Pi$, that is, applying (9) we extend a lower probability which satisfies (4) in the vaguest, or least-committal, possible way (while preserving coherence, cfr. 2.3.4).

(b) Strict factorization is computationally simple to apply, and we know *a priori* that what we obtain is still a coherent probability; alternative choices for the probability evaluations on the events in Π would generally require checking coherence of the extension so obtained, which might be heavy to perform.

(c) Although points (a) and (b) are strong arguments in favour of strict factorization, other independence-preserving extensions of $\underline{P}$ to $\mathcal{C} \cup \Pi$ are possible. Probably the simplest way to build them is to apply thm. 3.2 for extending to $\mathcal{C} \cup \Pi$ some, but not all, probabilities in the independent envelope of $\underline{P}$ on $\cup_{\mathcal{P} \in S} \mathcal{A}(\mathcal{P})$, and to compute then their lower envelope on $\mathcal{C} \cup \Pi$.

(d) We present a simple example on point (c) (similar examples may be found in [11]): suppose $S = \{\mathcal{P}_1, \mathcal{P}_2\}$, $\mathcal{P}_1 = \{A, A^c\}$, $\mathcal{P}_2 = \{B, B^c\}$, $A' \wedge B' \neq \varnothing$. The assessments $P_1(A) = 0.7$, $P_1(B) = 0.2$, $P_2(A) = 0.4$, $P_2(B) = 0.5$ uniquely identify two coherent precise probabilities P_1, P_2 on $\mathcal{A}(\mathcal{P}_1) \cup \mathcal{A}(\mathcal{P}_2) = \{\varnothing, \Omega, A', B'\}$. Extend them on $\mathcal{C} = \{\varnothing, \Omega, A', B', A' \mid B', B' \mid A'\}$ putting $P_i(A' \mid B') = P_i(A')$, $P_i(B' \mid A') = P_i(B')$, and then on $\mathcal{C} \cup \Pi$ ($\Pi = \{A' \wedge B'\}$) by putting $P_i(A' \wedge B') = P_i(A')P_i(B')$, i = 1, 2. By 3.2, the extensions are coherent. Defining now $\underline{P}$ as $\underline{P}(\cdot|\cdot) = \min\{P_1(\cdot|\cdot), P_2(\cdot|\cdot)\}$ for all events in $\mathcal{C} \cup \Pi$, $\underline{P}$ is coherent by 2.3.3 and it is easily seen that it realizes conditions (4) in def. 3.1. Therefore $\mathcal{P}_1$ and $\mathcal{P}_2$ are independent partitions (under $\underline{P}$), but $\underline{P}(A \wedge B) = 0.14 > \underline{P}(A)\underline{P}(B) = 0.08$ (note that, on the set $\mathcal{A}(\mathcal{P}_1) \cup \mathcal{A}(\mathcal{P}_2)$, $\{P_1, P_2\}$ is strictly included in the independent envelope of $\underline{P}$).

4 Some implications of epistemic independence

Suppose, throughout this section, that $\underline{P}$ is coherent on $\mathcal{C} \cup \Pi$, that the partitions in S are (lg. i. and) independent with respect to $\underline{P}$, and that whenever upper probabilities are defined they must also realize independence for the partitions, so that (9') must hold for them.

A general question is to investigate the implications of these hypotheses on the coherent extensions of $\underline{P}$. Some special interesting cases are discussed below.

4.1.1 External n-monotonicity

Define $U = \{E_1 \vee \ldots \vee E_n: E_i \in \mathcal{A}(\mathcal{P}_i), \mathcal{P}_1 \neq \ldots \neq \mathcal{P}_n, i = 1,\ldots,n, \{\mathcal{P}_1,\ldots,\mathcal{P}_n\} \subset S\}$. Say that an extension of $\underline{P}$ to $\mathcal{C} \cup \Pi \cup U$ is *externally n-monotone* if n-monotonicity holds for each event in U, i.e. if $\forall\, E_1 \vee \ldots \vee E_n \in U$,

$$(11)\quad \underline{P}(E_1 \vee \ldots \vee E_n) \geq \sum_{i=1}^{n} \underline{P}(E_i) - \sum_{i>j} \underline{P}(E_i \wedge E_j) + \ldots + (-1)^{n+1} \underline{P}(E_1 \wedge \ldots \wedge E_n).$$

4.1.2 Proposition

(a) If $\underline{P}$ strictly factors on Π, so that equality always holds in (9), then $\underline{P}$ is externally n-monotone;

(b) in any case, also when strict factorization does not apply, $\underline{P}$ is externally 2-monotone.

Proof. Apply (1) and (9') to write

$$\underline{P}(E_1 \vee \ldots \vee E_n) = 1 - \overline{P}(E_1^c \wedge \ldots \wedge E_n^c) \geq 1 - \overline{P}(E_1^c)\overline{P}(E_2^c)\ldots\overline{P}(E_n^c) = 1 - (1 - \underline{P}(E_1))\ldots(1 - \underline{P}(E_n)).$$

Developing the last expression we obtain

$$(12)\quad \underline{P}(E_1 \vee \ldots \vee E_n) \geq \sum_{i=1}^{n} \underline{P}(E_i) - \sum_{i>j} \underline{P}(E_i)\underline{P}(E_j) + \ldots + (-1)^{n+1} \underline{P}(E_1) \ldots \underline{P}(E_n).$$

It is immediate from (11) and (12) to see that (a) holds. To prove (b), write (12) for n = 2 and apply (9). ■

4.1.3 Comment

External n-monotonicity differs from standard n-monotonicity just because it is applied to a set of events which is not the usual algebra of events logically dependent on a partition. It is external to each of the partitions in S in the sense that the restriction of $\underline{P}$ on $\mathcal{A}(\mathcal{P})$, $\forall\, \mathcal{P} \in S$, may not be n-monotone, and in general not even 2-monotone.

So (b) seems to suggest that (external) 2-monotonicity arises in strict connection with independence, which following def. 3.1 operates among and not necessarily within partitions (property (b) is mentioned in [11], note 7 to sect. 9.1, in the case of independence for two events; see also sect. 5.13.4 for a significant example where lack of independence is incompatible with (standard) 2-monotonicity, but not with coherence).

As for (a), my feeling at present is that this property might concern a much wider class of independent lower probabilities than those which strictly factor on Π: to be sure, it is not confined to them.

4.2 A result on independence propagation

Suppose, for instance, that $S = \{\mathcal{P}_1,\ldots,\mathcal{P}_i,\ldots,\mathcal{P}_j,\ldots\}$, that i < j in the partition indexes implies that the experiment whose outcomes are described by (the atoms of) $\mathcal{P}_i$ takes place before that described by $\mathcal{P}_j$, and that we believe the partitions of S are epistemically independent (under a certain given $\underline{P}$). By def. 3.1, this implies in particular $\underline{P}(E_3 \mid E_1 \wedge E_2) = \underline{P}(E_3)$, $E_3 \in \mathcal{A}(\mathcal{P}_3)$, $E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i)$, i = 1, 2, which means that the evaluation on E_3 is independent from our knowing 'what happened in the past', if this is meant in the sense of knowing *jointly* the truth values of any couple of events (E_1, E_2), $E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i)$, i = 1, 2. Does this also imply independence from less punctual information concerning the past, like knowing the truth value of, say, $E_1 \vee E_2$?

More generally, given an arbitrary partition in S, let it be $\mathcal{P}_1$, consider a finite number of partitions in $S - \{\mathcal{P}_1\}$, for instance $\mathcal{P}_2,\ldots,\mathcal{P}_n$. If def. 3.1 holds, any event $E_1 \in \mathcal{A}^{\varnothing}(\mathcal{P}_1)$ is independent of our knowing jointly the truth values of any n–1 events $E_2,\ldots,E_n$ such that $E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i)$, i = 2,…,n, i.e. is independent of any event belonging to the set $\mathcal{A}_1 = \mathcal{A}^{\varnothing}(\mathcal{P}_2) \wedge \ldots \wedge \mathcal{A}^{\varnothing}(\mathcal{P}_n) = \{E_2 \wedge \ldots \wedge E_n: E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i), i = 2,\ldots,n\}$. On the other hand, if we know the truth values of the atoms of partitions $\mathcal{P}_2,\ldots,\mathcal{P}_n$ we also know the truth values of the events in the set $\mathcal{A}^* = \mathcal{A}^{\varnothing}(\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n)$ of the non-impossible events logically dependent on the product partition

$\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$. Clearly, $\mathcal{A}^*$ is a larger set than $\mathcal{A}_1$ (for instance, $\mathcal{A}_1$ is not closed under complementation whilst $\mathcal{A}^* \cup \{\varnothing\}$ is an algebra).

Now, the question is: does independence of E_1 from any event in $\mathcal{A}_1$ propagate to independence of E_1 from any event in $\mathcal{A}^*$?

When all partitions in S are finite, the following proposition assures in particular that independence from information arising from $\mathcal{A}^{\varnothing}(\mathcal{P}_2) \wedge \ldots \wedge \mathcal{A}^{\varnothing}(\mathcal{P}_n)$ implies independence from information depending on $\mathcal{A}^{\varnothing}(\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n)$.

4.2.1 Proposition

Let S be an arbitrary set of lg. i. finite partitions and let $\underline{P}$ be a lower probability on $\mathcal{C}$ which realizes independence for the partitions in S by def. 3.1; choose arbitrarily a partition $\mathcal{P} \in S$ and then a product partition, let us call it $\wedge\mathcal{P}$, obtained as the product of a finite number of partitions in $S - \{\mathcal{P}\}$. Then $\underline{P}$ has a unique coherent extension on every event $E \mid F$, $E \in \mathcal{A}(\mathcal{P})$, $F \in \mathcal{A}^{\varnothing}(\wedge\mathcal{P})$, given by

(13) $\underline{P}(E \mid F) = \underline{P}(E)$.

Proof. It is not restrictive, but simplifies notation, to prove that (13) holds for $\mathcal{P} = \mathcal{P}_1$, $\wedge\mathcal{P} = \mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$.

Call for this $e^{(i)}$ the generic atom of the product partition $\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$. From the finiteness hypothesis, $\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$ is made up of a finite number s of atoms (by the logical independence assumption, $s = s_2 \ldots s_n$, where s_i is the number of atoms of $\mathcal{P}_i$, i = 2,…,n). Hence any $F \in \mathcal{A}(\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n)$ may be written as a logical sum of a finite number of atoms of $\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$.

We shall prove the thesis in three steps.

(a) Proving (13) involves operating with extensions of $\underline{P}$. Now, by 2.3.3 every extension of $\underline{P}$ on an arbitrary $\mathcal{C}_1 \supset \mathcal{C}$ may be obtained as the lower envelope on $\mathcal{C}_1$ of a set M_1 of precise probabilities such that the lower envelope of the restriction of M_1 on $\mathcal{C}$, let us call it M, is the starting lower probability $\underline{P}$. Hence, possibly different coherent extensions of $\underline{P}$ on $\mathcal{C}_1$ could be obtained by varying the probabilities in the set M.

Anyway, defining $M^*(A \mid B) = \{P: P \text{ is a coherent precise probability on } \mathcal{C},\ P \geq \underline{P},\ P(A \mid B) = \underline{P}(A \mid B)\}$, it is $M^* \neq \varnothing$ (by 2.3.3) and *at least one probability* $P^* \in M^*$ *must be included in every set* M whose lower envelope on $\mathcal{C}$ is $\underline{P}$, and this for every $A \mid B \in \mathcal{C}$ (again by 2.3.3).

(b) Let us now consider an arbitrary event $E \mid F$, $E \in \mathcal{A}(\mathcal{P}_1)$, $F \in \mathcal{A}^{\varnothing}(\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n)$.

We shall prove in (b) that $\forall\ P^* \in M^*(E)$, (the extension of P^* on $\mathcal{C} \cup \{E \mid F\}$ is such that) $P^*(E \mid F) = \underline{P}(E)$. This implies $\underline{P}(E \mid F) \leq \underline{P}(E)$, since by (a) at least one P^* is included in any set M. Let us preliminarily note that

(14) $P^*(E) = \Sigma\ P^*(E \mid e^{(i)})\ P^*(e^{(i)}) \geq \underline{P}(E)\ \Sigma\ P^*(e^{(i)}) = \underline{P}(E)$,

where the summations are extended to all atoms $e^{(i)}$ of $\mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$ and the inequality holds since $P^*(E \mid e^{(i)}) \geq \underline{P}(E \mid e^{(i)}) = \underline{P}(E)$ (recall that $E \mid e^{(i)} \in \mathcal{C}$). But $P^*(E) = \underline{P}(E)$, so (14) implies

(15) $P^*(E \mid e^{(i)}) = \underline{P}(E),\ \forall\ e^{(i)} \in \mathcal{P}_2 \wedge \ldots \wedge \mathcal{P}_n$.

From what seen just before (a), we can write $F = \vee e^{(i)}$, the logical sum being extended to the *finite* set of those atoms $e^{(i)}$ which imply F. Therefore we can apply first the conglomerative property and then (15) to write (the summations are of course extended to the atoms $e^{(i)}$ which imply F and only to them):

(16) $P^*(E \mid F) = P^*(E \wedge \vee e^{(i)} \mid \vee e^{(i)}) =$

$\Sigma\ P^*(E \mid e^{(i)})\ P^*(e^{(i)} \mid \vee e^{(i)}) = \underline{P}(E)\ \Sigma\ P^*(e^{(i)} \mid \vee e^{(i)}) = \underline{P}(E)$.

(c) It is now easy to prove that every precise probability P dominating $\underline{P}$ on $\mathcal{C}$ is such that $P(E \mid F) \geq \underline{P}(E)$ (follow the steps in (16), substituting P^* with P and replacing the third equality by $\geq$ since $P(E \mid e^{(i)}) \geq \underline{P}(E \mid e^{(i)}) = \underline{P}(E)$). From this and the conclusions of (b) we obtain $\underline{P}(E \mid F) = \underline{P}(E)$. ■

4.2.2 Comment

The result in 4.2.1 shows that there are instances where independence is necessarily maintained when extending a lower probability. It should be possible to generalise 4.2.1 in some directions, as long as the hypothesis of finiteness of the partitions is maintained.

5 Conclusions

The approach followed in this note regards irrelevance as the essential concept for interpreting and defining independence. Nevertheless we saw that the property of strict factorization is always admissible and significant, not only computationally.

We focused on the essential aspects, starting from consistency, of the notion of independence proposed, which includes many concepts of independence as special cases: for instance, independence among n events, putting $S = \{\mathcal{P}_1, \ldots, \mathcal{P}_n\}$, $\mathcal{P}_i = \{A_i, A_i^c\}$, i = 1,…,n.

Although we preferred to discuss independence for the more immediate case of events, it is also possible to consider the atoms of each partition $\mathcal{P}_i$ in def. 3.1 as the distinct outcomes of a random number X_i, so that every event $E_i \in \mathcal{A}^{\varnothing}(\mathcal{P}_i)$ identifies a subset v_i of the set V_i of the admissible values for X_i. Then epistemic independence for a family F of random numbers can be defined by

requiring that their underlying partitions are logically independent and that the following equalities hold, $\forall X_1, X_2,\ldots, X_n \in F, \forall v_i \in V_i$, i = 1,…, n:

$\underline{P}(X_1 \in v_1 \,|\, X_2 \in v_2 \wedge \ldots \wedge X_n \in v_n) = \underline{P}(X_1 \in v_1)$.

A noteworthy feature is that, by the extension theorem (sect. 2.3.4), it is always possible to assign lower probabilities only on subsets of $\mathscr{A}(\mathscr{P})$, for some or all $\mathscr{P} \in S$, and to extend them on a subset of $\mathscr{C}$ realizing *partially* the independence conditions (4): this will be the rule in many practical situations.

Directions for further work include possible generalisations of independence propagation, the treatment of conditional independence, and applications.

Acknowledgements

I wish to thank the referees for their helpful comments and suggestions.

References

[1] L. Crisma. *Lezioni di calcolo delle probabilità.* Edizioni Goliardiche, Trieste, 1997.

[2] L. Crisma. *The notion of stochastic independence in the theory of coherent probabilities.* Tech. Rep. n. 1/99, Dipartimento di Matematica Applicata "B. de Finetti", University of Trieste, 1999.

[3] L.M. De Campos and S. Moral. *Independence concepts for convex sets of probabilities.* Tech. Rep., Departamento de Ciencias de la Computaciòn e I. A., Universidad de Granada, 1995.

[4] B. de Finetti. *Theory of Probability*, Vol. 1, English translation of *Teoria delle Probabilità,* (1970), Wiley, London, 1974.

[5] P. Ghirardato. On independence for non-additive measures, with a Fubini theorem. *J. Econ. Theory,* 73:261-291, 1997.

[6] A. Gilio and S. Ingrassia. Totally coherent set-valued probability assessments. *Kybernetica,* 34:3-15, 1998.

[7] S. Holzer. On coherence and conditional prevision. *Boll. Un. Mat. Ital.*, Serie VI (IV-C):441-460, 1985.

[8] T. Seidenfeld and L. Wasserman. Dilation for sets of probabilities. *Ann. Statist.*, 21:1139-1154, 1993.

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

[10] P. Walley. *Coherent lower (and upper) probabilities*. Research Report of the Dep. of Statistics, University of Warwick, 1981.

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

[12] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence,* 83:1-58, 1996.

[13] P. M. Williams. *Notes on conditional previsions.* Research Report of the School of Math. and Phys. Sci., University of Sussex, 1975.