Independence concepts in evidence theory: some results about epistemic irrelevance and imprecise belief functions

Sebastien Destercke

Centre de cooperation internationale en
recherche agronomique pour le developpement (CIRAD)
UMR IATE, Campus Supagro, Montpellier, France
Email: sebastien.destercke@cirad.fr

Abstract—Many extensions of classical stochastic independence have been proposed when working with probability sets to represent uncertainty. As belief functions can be seen as particular instances of such probability sets, some authors have investigated how these extensions can be reinterpreted and retrieved in the particular framework of belief functions. They have mainly focused on the so-called notions of random set independence, fuzzy non-interaction, strong independence and unknown interaction. In this paper, we pursue this effort in two ways: first by showing that the notion of epistemic irrelevance, central in Walley theory of lower previsions, can be likewise reinterpreted in terms of belief functions; second by considering the more general case where mass assignments inducing belief functions are themselves imprecise.

Keywords: independence, belief functions, imprecise probability, random sets.

I. INTRODUCTION

The notion of stochastic independence is essential in probability theory, its associated factorization properties allowing to decompose a complex problem into simpler ones, or to easily build joint probabilities from marginal ones. When working with imprecise probabilities rather than precise ones, the concept of independence can be extended in many different ways, depending on the interpretation it is given. Such extensions have been proposed and compared by many authors (see, for example, Walley [16] and Couso *et al.* [4]).

Evidence theory [14] is formally embedded in the theory of imprecise probabilities, and belief functions can be seen as particular instances of generic lower probabilities. Therefore, independence concepts issued from imprecise probability theory can be reinterpreted in the particular framework of evidence theory. For instance, Fetz and Oberguggenberger [9], [10] consider different formal ways to combine both weights given to focal sets and probability distributions inside each focal sets. Depending on the conditions they impose to these combinations, they retrieve the different independence concepts that are unknown interaction, strong independence, random set independence, and fuzzy non-interaction. More specifically, they establish three types of condition, the first concerning focal set weights, and the two others concerning the combination and choice of probabilities inside each focal set. Couso [3] complete and pursue this study by providing interpretations for the formal links identified by Fetz and Oberguggenberger. She associates Fetz and Oberguggenberger's conditions to a 2-step random process (i.e., ball drawing from urns) with particular independence features.

In this paper, we go further in the study of independence concepts settled in the framework of evidence theory. First, we interest ourselves to the notion of epistemic irrelevance. Although this asymmetrical independence notion, central in Walley's interpretation of imprecise probabilities, has received important attention in the past [6], [15], it has, up to now, not been considered in the particular framework of evidence theory. We then consider a more general case

of imprecise mass assignments [1], and study to which extent some of the results of Fetz and Oberguggenberger still hold in this case.

We first recall the framework of evidence theory, as considered by Fetz and Couso (Section II). In Section III, we recall the results they obtain and show that, formally, they can be easily extended to the case of epistemic irrelevance or to the case where mass assignments are made imprecise. We also discuss brifely the possible interpretation of these extensions. Note that this paper is mainly concerned with belief functions interpreted as lower probabilities. Other interpretation of belief functions, such as the Transferable Belief Model, have led to other notions of independence [2], also generalising the classical notion of stochastic independence.

II. PRELIMINARIES

We recall here basic notions and introduce notations used in the rest of the paper

A. Imprecise probabilities and evidence theory

Let $\mathcal{X}$ be the finite domain on which a variable X assumes its values. In imprecise probability theory, the uncertainty about the value of X is described by a convex probability set $\mathcal{P}$. This set induces, on any event A, lower and upper probability bounds

$$\overline{P}(A) = \sup_{p \in \mathcal{P}} P(A); \quad \underline{P}(A) = \inf_{p \in \mathcal{P}} P(A)$$

with p a probability mass function on $\mathcal{X}$, and P the corresponding probability measure. These two bounding measures are dual, in the sense that $\overline{P}(A) = 1 - \underline{P}(A^c)$ for any event $A \subseteq \mathcal{X}$. A lower probability $\underline{P}$ induce a set of dominating probability measures, denoted by $\mathcal{P}(\underline{P})$ and such that $\mathcal{P}(\underline{P}) = \{p \in \mathbb{P}_{\mathcal{X}} | P(A) \geq \underline{P}(A)\}$, with $\mathbb{P}_{\mathcal{X}}$ the set of all probability mass functions over $\mathcal{X}$. In general, $\mathcal{P}(\underline{P})$ is a proper subset of $\mathcal{P}$.

In fact, to describe any probability set $\mathcal{P}$ by means of lower and upper bounds, one needs the richer language of bounded expected values [16]. Let f be any real-valued bounded function on $\mathcal{X}$. Then, its lower and upper expected values, denoted by $\underline{E}(f)$ and $\overline{E}(f)$, are

$$\underline{E}(f) = \inf_{p \in \mathcal{P}} \sum_{x \in \mathcal{X}} p(x)f(x); \quad \overline{E}(f) = \sup_{p \in \mathcal{P}} \sum_{x \in \mathcal{X}} p(x)f(x).$$

Lower and upper probabilities of an event A are retrieved when f is the indicator function I_A of A ($I_A(x) = 1$ when $x \in A$, zero otherwise). In his theory of lower previsions, Walley argues[1] that lower expectations modelling beliefs can be interpreted as behavioural dispositions of an agent toward a so-called gamble f, namely that $\underline{E}(f)$ is the maximum buying price an agent is ready to pay for f, given its beliefs about the value of variable X. Lower and upper

---

[1]In the same way as De Finetti does for classical expectations.

probabilities of an event A are retrieved when f is the indicator function I_A of A ($I_A(x) = 1$ when $x \in A$, zero otherwise).

A probability set that will be of interest here is the vacuous probability set of an event $A \subseteq \mathcal{X}$, representing the fact that all we know is that $X \in A$. This set, that we denote by $\mathfrak{I}_A$ is such that $\mathfrak{I}_A = \{p \in \mathbb{P}_{\mathcal{X}} | P(A) = 1\}$.

In the mathematical theory of evidence [14], uncertainty is described by means of a mass assignment $m : 2^{|\mathcal{X}|} \to [0, 1]$ equivalent to a probability mass function defined over the power set of $\mathcal{X}$ (i.e. $\sum_{A \in \mathcal{X}} m(A) = 1$ and $m(\emptyset) = 0$). Subsets receiving positive mass are called focal sets, and are denoted here $\mathcal{F}$. $m(A)$ can be interpreted as the probability that our knowledge is $X \in A$, and nothing else. This mass assignment induces two measures called plausibility and belief, resp. denoted Pl and Bel and defined, for any event $E \in \mathcal{X}$, as follows:

$$Bel(E) = \sum_{A \subseteq E} m(A); \quad Pl(E) = 1 - Bel(E^c) = \sum_{A \cap E \neq \emptyset} m(A).$$

These measures induce a particular probability set $\mathcal{P}(Bel)$. Now, if $\mathcal{F} = \{A_1, \ldots, A_m\}$ and if we associate to each focal set the set $\mathfrak{I}_{A_i}$, $\mathcal{P}(Bel)$ can be described as follows:

$$\mathcal{P}(Bel) = \{p \in \mathbb{P}_{\mathcal{X}} | p = \sum_{i=1}^{m} m(A_i)p_i, p_i \in \mathfrak{I}_{A_i}\}. \quad (1)$$

That is, probabilities in the set can be recovered by picking, in each focal set, a probability mass function having this set for support. This is the view taken by Fetz.

B. Independence concepts for imprecise probabilities

Now, let X and Y be two variables respectively assuming their values on the finite spaces $\mathcal{X}$ and $\mathcal{Y}$. Marginal uncertainty about their true values are given by probability sets $\mathcal{P}_X$ and $\mathcal{P}_Y$. To make assessments about the value these two variables can jointly assume, it is necessary to build a joint uncertainty representation over $\mathcal{X} \times \mathcal{Y}$. Independence assessments are instrumental tools to build such joint representations. First, we assume X and Y to be logically independent, i.e., any joint observation $x, y \in \mathcal{X} \times \mathcal{Y}$ is deemed possible, as it is a necessary condition to have other kinds of independence. The independence concepts we consider here are the one of strong independence, epistemic irrelevance and random set independence.

The concept of strong independence extends directly the concept of stochastic independence, in the sense that we retrieve it when taking the stochastic product of every probability mass function inside $\mathcal{P}_X$ and $\mathcal{P}_Y$. The probability set induced by an assumption of strong independence, denoted by $\mathcal{P}_{SI}$, is such that

$$\mathcal{P}_{SI} = \{p \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}} | \forall x, y \in \mathcal{X} \times \mathcal{Y}, p(x, y) = p_{\mathcal{X}}(x)p_{\mathcal{Y}}(y),$$
$$p_{\mathcal{X}} \in \mathcal{P}_X, p_{\mathcal{Y}} \in \mathcal{P}_Y\}.$$

The concept of epistemic irrelevance corresponds to an asymmetric concept, expressing the idea that learning the value of a variable do not modify our beliefs or uncertainty about the value of another variable (not excluding the possibility that learning the value of the latter may modify our uncertainty about the former). Consider the assumption that X is epistemically irrelevant to Y and denote it by $X \not\to Y$. Let us denote by $\mathcal{P}_{X \not\to Y}$ a set of joint probability distributions over $\mathcal{X} \times \mathcal{Y}$, and for a given distribution $p \in \mathcal{P}_{X \not\to Y}$, let $p_{\mathcal{Y}}(\cdot | x)$ denote its conditional probability distribution on $\mathcal{Y}$, obtained by $p_{\mathcal{Y}}(y | x) = p(y, x) / P(\{y \times \mathcal{X}\})$, with P the measure induced by p. In this case, X is epistemically irrelevant to Y when

$$\{p_{\mathcal{Y}}(\cdot | x) | p \in \mathcal{P}_{X \not\to Y}\} = \mathcal{P}_Y,$$

that is, our conditional knowledge given the X agrees with our marginal uncertainty about Y. This captures the idea that learning the value of X does not change our uncertainty about Y. The probability set induced by this assumption, denoted by $\mathcal{P}_{X \not\to Y}$, is such that

$$\mathcal{P}_{X \not\to Y} = \{p \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}} | \forall x, y \in \mathcal{X} \times \mathcal{Y}, p(x, y) = p_{\mathcal{Y}}(y | x)p_{\mathcal{X}}(x),$$
$$p_{\mathcal{X}} \in \mathcal{P}_X, p_{\mathcal{Y}}(\cdot | x) \in \mathcal{P}_Y\},$$

with $p_{\mathcal{Y}}(\cdot | x)$ the conditional probability mass function on $\mathcal{Y}$ given $X = x$, whose particular values may depend on the value x. The notion can also be expressed in terms of lower expectation of a function f defined on $\mathcal{X} \times \mathcal{Y}$ as follows: for any function f whose values only depend on Y (i.e., for a given $y \in \mathcal{Y}$, $f(x, y)$ has the same value for every $x \in \mathcal{X}$), epistemic irrelevance of X towards Y implies that

$$\underline{E}(f | x) = \inf_{p \in \mathcal{P}_{X \not\to Y}} \sum p_{\mathcal{Y}}(y | x)f(y, x) = \underline{E}_{\mathcal{Y}}(f)$$

where $\underline{E}_{\mathcal{Y}}(f)$ is the lower expectation of f restricted to $\mathcal{Y}$ given $\mathcal{P}_Y$. Equality $\underline{E}_{\mathcal{Y}}(f | x) = \underline{E}(f)$ well express the fact that learning x do not change our uncertainty about Y.

Note that $\mathcal{P}_{SI}$ is recovered if $p_Y(\cdot | x)$ is constrained to remain the same whatever the value of x. The concept of epistemic independence corresponds to assert that both variables are epistemically irrelevant to each other. It can be retrieved by considering the intersection of the two sets generated by separated epistemic irrelevance assessments [4].

The concept of random set independence has, up to now, no clear intereptation in the imprecise probability theory (we will come back to it later), however it is still widely used [12], mainly for its practical interests. Assume that $\mathcal{P}_X$ and $\mathcal{P}_Y$ are induced by mass assignments m_X and m_Y, with $\mathcal{F}_X = \{A_1, \ldots, A_m\}$ and $\mathcal{F}_Y = \{B_1, \ldots, B_n\}$. Then, the probability set induced by an assumption of random set independence, denote by $\mathcal{P}_{RI}$, is the probability set induced by the joint mass assignment m_{RI} defined on $2^{|\mathcal{X}|} \times 2^{|\mathcal{Y}|}$ and such that

$$\forall A_i \times B_j \subseteq \mathcal{X} \times \mathcal{Y}, m_{RI}(A_i \times B_j) = m_X(A_i) \cdot m_Y(B_j).$$

The following inclusions hold between the probability sets induced by these independence assumptions

$$\mathcal{P}_{SI} \subseteq \left\{ \begin{array}{c} \mathcal{P}_{X \not\to Y} \\ \mathcal{P}_{Y \not\to X} \end{array} \right\} \subseteq \mathcal{P}_{RI}$$

Also recall that, when the two probability sets $\mathcal{P}_X$ and $\mathcal{P}_Y$ are vacuous probability sets $\mathfrak{I}_A, \mathfrak{I}_B$ with $A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}$, then the joint probability set induced by any of the above independence concept is simply $\mathfrak{I}_{A \times B}$, i.e., the vacuous probability set corresponding to the Cartesian product $A \times B$.

III. INDEPENDENCE IN EVIDENCE THEORY

We now proceed to transfer and re-examine all these concepts in the setting of evidence theory.

A. Existing results

Let $\mathcal{P}_X$ and $\mathcal{P}_Y$ be two probability sets induced by mass assignments m_X, m_Y with focal sets $\mathcal{F}_X = \{A_1, \ldots, A_m\}$ and $\mathcal{F}_Y = \{B_1, \ldots, B_n\}$. Then, any joint probability P over $\mathcal{X} \times \mathcal{Y}$ having marginals in $\mathcal{P}_X, \mathcal{P}_Y$ can be formed by the following procedure:

- m is a mass assignment over $2^{|\mathcal{X}|} \times 2^{|\mathcal{Y}|}$ such that, for any $i = 1, \ldots, m$ and $j = 1, \ldots, n$

$$m(A_i) = \sum_{j=1}^{n} m(A_i \times B_j); \quad m(B_j) = \sum_{i=1}^{m} m(A_i \times B_j)$$

- For each joint focal set $A_i \times B_j$, define a probability mass function $p^{ij} \in \mathfrak{I}_{A_i \times B_j}$
- The probability mass function $p = \sum_{j=1}^{n} \sum_{i=1}^{m} m(A_i \times B_j) p^{ij}$ has marginals in $\mathcal{P}_X, \mathcal{P}_Y$.

Given a probability mass function p^{ij} defined over $\mathcal{X} \times \mathcal{Y}$, we denote by $p_{\mathcal{X}}^{ij}$ and $p_{\mathcal{Y}}^{ij}$ its marginals over $\mathcal{X}$ and $\mathcal{Y}$, that is, for any $x \in \mathcal{X}$ ($y \in \mathcal{Y}$), $p_{\mathcal{X}}^{ij}(x) = P^{ij}(\{x\} \times \mathcal{Y})$ ($p_{\mathcal{Y}}^{ij}(y) = P^{ij}(\mathcal{X} \times \{y\})$), with P^{ij} the probability measure induced by p^{ij}.

Since any probability mass function having marginals in $\mathcal{P}_X, \mathcal{P}_Y$ can be built by the above procedure, considering all of them is equivalent to the joint probability set induced by an assumption of unknown interaction (i.e. considering all possible dependency structures). To retrieve smaller sets induced by specific independence assumptions, one has to constraint how joint probabilities are built from marginals. There are basically three levels at which (in)dependence structures and constraints can be specified:

1) how marginal mass assignments m_X, m_Y are combined;
2) how, for each (i,j), $i = 1, \ldots, n; j = 1, \ldots, m$, marginal probability mass functions $p_{\mathcal{X}}^{ij}, p_{\mathcal{Y}}^{ij}$ are combined into p^{ij};
3) how for each (i,j), $i = 1, \ldots, n; j = 1, \ldots, m$, marginal probability mass functions $p_{\mathcal{X}}^{ij}, p_{\mathcal{Y}}^{ij}$ are selected.

Let us now summarise results previously obtained by Fetz [9] (proofs are provided in his paper)

Proposition 1. *Consider the two marginal probability sets $\mathcal{P}_X, \mathcal{P}_Y$ induced by mass assignments m_X, m_Y and the joint set $\mathcal{P}_{RI}$. The set of joint probabilities built with the following constraints coincide with the set $\mathcal{P}_{RI}$:*

1) $\forall i = 1, \ldots, n; j = 1, \ldots, m, m(A_i \times B_j) = m_X(A_i)m_Y(B_j)$
2) $\forall i = 1, \ldots, n; j = 1, \ldots, m, p^{ij}$ *has marginal probability mass functions $p_{\mathcal{X}}^{ij}, p_{\mathcal{Y}}^{ij}$;*
3) $\forall i = 1, \ldots, n; j = 1, \ldots, m, p_{\mathcal{X}}^{ij} \in \mathfrak{I}_{A_i}, p_{\mathcal{Y}}^{ij} \in \mathfrak{I}_{B_j}$

Proposition 2. *Consider the two marginal probability sets $\mathcal{P}_X, \mathcal{P}_Y$ induced by mass assignments m_X, m_Y and the joint set $\mathcal{P}_{SI}$. The set of joint probabilities built with the following constraints coincide with the set $\mathcal{P}_{SI}$:*

1) $\forall A_i \times B_j, i = 1, \ldots, n; j = 1, \ldots, m, m(A_i \times B_j) = m_X(A_i)m_Y(B_j)$
2) $\forall i = 1, \ldots, n; j = 1, \ldots, m, p^{ij} = p_{\mathcal{X}}^{ij} \cdot p_{\mathcal{Y}}^{ij}$, *i.e. p^{ij} is the stochastic product of $p_{\mathcal{X}}^{ij}, p_{\mathcal{Y}}^{ij}$;*
3) $\forall i = 1, \ldots, n; j = 1, \ldots, m, p_{\mathcal{X}}^{i} \in \mathfrak{I}_{A_i}, p_{\mathcal{Y}}^{j} \in \mathfrak{I}_{B_j}$ *with*

$$p_{\mathcal{X}}^{i} := p_{\mathcal{X}}^{i1} = \ldots = p_{\mathcal{X}}^{im} \tag{2}$$
$$p_{\mathcal{Y}}^{j} := p_{\mathcal{Y}}^{1j} = \ldots = p_{\mathcal{X}}^{nj} \tag{3}$$

Note that logical constraints (2) and (3) on the probability mass functions selected in focal sets are essential to retrieve $\mathcal{P}_{SI}$. Without them, the obtained probability set is $\mathcal{P}_{RI}$.

B. Epistemic irrelevance

Let us now focus on the concept of epistemic irrelevance. Without loss of generality, we assume that variable Y is epistemically irrelevant to variable X. We now show that results recalled in the previous section can easily be extended to the case of epistemic irrelevance.

Proposition 3. *Consider the two marginal probability sets $\mathcal{P}_X, \mathcal{P}_Y$ induced by mass assignments m_X, m_Y and the joint set $\mathcal{P}_{Y \not\to X}$. The set of joint probabilities built with the following constraints coincide with the set $\mathcal{P}_{Y \not\to X}$:*

1) $\forall A_i \times B_j, i = 1, \ldots, n; j = 1, \ldots, m, m_X(A_i \times B_j) = m_X(A_i)m_Y(B_j)$

2) $\forall i = 1, \ldots, n; j = 1, \ldots, m, p^{ij} \in \mathfrak{I}_{A_i \times B_j}$;
3) $\forall i = 1, \ldots, n; j = 1, \ldots, m$ *and* $y \in \mathcal{Y}$, $p_{\mathcal{X}}^{i}(\cdot|y) \in \mathfrak{I}_{A_i}$, $p_{\mathcal{Y}}^{j} \in \mathfrak{I}_{B_j}$ *with*

$$p_{\mathcal{X}}^{i}(\cdot|y) := p_{\mathcal{X}}^{i1}(\cdot|y) = \ldots = p_{\mathcal{X}}^{im}(\cdot|y) \tag{4}$$
$$p_{\mathcal{Y}}^{j} := p_{\mathcal{Y}}^{1j} = \ldots = p_{\mathcal{X}}^{nj} \tag{5}$$

But we can have $p_{\mathcal{X}}^{i}(\cdot|y) \neq p_{\mathcal{X}}^{i}(\cdot|y')$ for $y \neq y'$

Proof: Let us first show that the set described by the constraints includes $\mathcal{P}_{Y \not\to X}$. Consider a probability mass function p that is in $\mathcal{P}_{Y \not\to X}$. For any couple $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $p(x, y) = p_{\mathcal{X}}(x|y)p_{\mathcal{Y}}(y)$, with $p_{\mathcal{X}}(\cdot|y) \in \mathcal{P}_X$ and $p_{\mathcal{Y}} \in \mathcal{P}_Y$. Using the relation (1), $p(x, y)$ can be rewritten as

$$p(x, y) = \sum_{i=1}^{n} m(A_i) p_{\mathcal{X}}^{i}(x|y) \sum_{j=1}^{m} m(B_j) p_{\mathcal{Y}}^{j}(y) \tag{6}$$

with $p_{\mathcal{X}}^{i}(x|y) \in \mathfrak{I}_{A_i}$ and $p_{\mathcal{Y}}^{j}(y) \in \mathfrak{I}_{B_j}$. This probability satisfy constraints of Proposition 3, hence the set described by constraints of Proposition 3 includes $\mathcal{P}_{Y \not\to X}$.

Let us now show that $\mathcal{P}_{Y \not\to X}$ includes all probabilities that could be built with the specified constraints. Consider a generic probability mass function p built from marginal mass functions in $\mathcal{P}_X, \mathcal{P}_Y$, such that, for any $x, y \in \mathcal{X} \times \mathcal{Y}$

$$p(x, y) = \sum_{j=1}^{n} \sum_{i=1}^{m} m(A_i \times B_j) p^{ij}(x, y)$$

with $p^{ij} \in \mathfrak{I}_{A_i \times B_j}$. Now, the two first constraints and conditions (4) and (5) of Proposition 3 means that we can rewrite $p(x, y)$ as

$$p(x, y) = \sum_{j=1}^{n} \sum_{i=1}^{m} m(A_i)m(B_j) p_{\mathcal{X}}^{i}(x|y) \cdot p_{\mathcal{Y}}^{j}(y) \tag{7}$$

$$= \sum_{j=1}^{n} m(B_j) \cdot p_{\mathcal{Y}}^{j}(y) \sum_{i=1}^{m} m(A_i) p_{\mathcal{X}}^{i}(x|y), \tag{8}$$

Which, being equivalent to Eq. (6), shows that the probability masses built from the specified constraints are indeed in $\mathcal{P}_{Y \not\to X}$. Therefore, the two sets coincide. ∎

Note that, if we consider $p_{\mathcal{X}}^{i}(\cdot|y)$ to be the same for all $y \in \mathcal{Y}$, we retrieve the constraints (2) and (3). Also, simply dropping the third condition gives the joint probability set induced by random set independence. Note that, as for strong independence [3], we can expect some specific joint probability masses inside $\mathcal{P}_{Y \not\to X}$ to be reachable by other combination of focal set weights and marginal probabilities inside each focal sets. However, our aim here is to describe constraints needed to retrieve exactly all probability masses in $\mathcal{P}_{Y \not\to X}$.

Extreme points of $\mathcal{P}_{Y \not\to X}$ are reached by taking Dirac measures for $p_{\mathcal{X}}^{i}(\cdot|y), p_{\mathcal{Y}}^{j}$ on each focal sets. Joint probabilities are then convex mixtures of Dirac measures where the weights given to the Dirac measures are respectively $m(A_i)$ and $m(B_j)$. Lower expectation for a given real-valued bounded function f can therefore be written as the solution of the following optimisation problem:

$$\underline{E}(f) = min \sum_{j=1}^{n} \sum_{i=1}^{m} f(x_i, y_j)m(A_i)m(B_j)\delta_{x_{i,y_j}} \cdot \delta_{y_j}$$

with the constraints

$$\text{for } j = 1, \ldots, m, \quad y_j \in B_j$$

$$\text{for } i = 1, \ldots, n; y \in \mathcal{Y} \quad x_{i,y} \in A_j$$

wih δ_{y_j} and $\delta_{x_{i,y}}$ dirac measures respectively on y_j and $x_{i,y}$. Upper expectation can be obtained by replacing the minimum by a maximum in the objective function.

The next example, inspired from Couso [3] and using the classical urn example, shows that constraints (4) and (5) are necessary to retrieve the probabilistic set corresponding to epistemic irrelevance.

Example 1. *Consider two urns X and Y, each of them with 10 balls either painted red, white, or unpainted (to be painted either in white or red). The first urn has four red, two white and four unpainted balls. The second one has two red, three white and five unpainted balls. We have $\mathcal{X} = \{r, w\} = \mathcal{Y} = \{r, w\}$, focal elements $\mathcal{F}_X = \{A_1 = \{r\}, A_2 = \{w\}, A_3 = \{r, w\}\}$ and $\mathcal{F}_Y = \{B_1 = \{r\}, B_2 = \{w\}, B_3 = \{r, w\}\}$. Mass assignments for the random selection of balls in the urn are the following:*

$$m_X(A_1) = 0.4 \quad m_X(A_2) = 0.2 \quad m_X(A_3) = 0.4$$

$$m_Y(B_1) = 0.2 \quad m_Y(B_2) = 0.3 \quad m_Y(B_3) = 0.5$$

From Proposition 3, epistemic irrelevance between $\mathcal{P}_X$ and $\mathcal{P}_Y$ is retrieved by the following set-up. One ball is selected stochastically from each urns. If the ball from the second urn is uncoloured, it is painted by a fixed random process. If the ball coming from the first urn is uncoloured, then it is coloured by a random process whose characteristics may depend on the colour of the ball drawn from the second urn, independently of how this second colour has been obtained (i.e. by drawing directly a painted ball or by painting an uncoloured one).

As we have recalled, extreme points of $\mathcal{P}_{Y \nrightarrow X}$ can be reached by considering Dirac measures on each focal element and then taking the convex mixture of these measures. For example, a first extreme point can be reached by choosing $p_{\mathcal{Y}}^3(w) = 1$, $p_{\mathcal{X}}^3(w|r) = 1$ and $p_{\mathcal{X}}^3(r|w) = 1$. Considering the element $\{r, w\} \in \mathcal{X} \times \mathcal{Y}$ and the combination of focal elements for which Dirac measures concentrate on $\{r, w\}$, this gives

$$
\begin{aligned}
p(\{r, r\}) = {} & m_X(A_1)m_Y(B_2) + m_X(A_1)m_Y(B_3)p_{\mathcal{Y}}^3(w) \\
& + m_X(A_3)p_{\mathcal{X}}^3(r|w)m_Y(B_3)p_{\mathcal{Y}}^3(w) \\
& + m_X(A_3)p_{\mathcal{X}}^3(r|w)m_Y(B_2) = 0.64.
\end{aligned}
$$

The same calculation for each element of $\mathcal{X} \times \mathcal{Y}$ provides an extreme points of $\mathcal{P}_{Y \nrightarrow X}$. Considering every combination of Dirac measures gives all extreme points of $\mathcal{P}_{Y \nrightarrow X}$ (and possibly other points inside $\mathcal{P}_{Y \nrightarrow X}$). The following tabular specify all extreme points of the probability set $\mathcal{P}_{Y \nrightarrow X}$. Example detailed above for $\{r, w\}$ is the fifth probability mass function

n^o	$p(\{r, r\})$	$p(\{r, w\})$	$p(\{w, r\})$	$p(\{w, w\})$
1	0.08	0.32	0.12	0.48
2	0.28	0.12	0.42	0.18
3	0.16	0.64	0.04	0.16
4	0.56	0.24	0.14	0.06
5	0.08	0.64	0.12	0.16
6	0.16	0.24	0.04	0.56
7	0.28	0.24	0.42	0.06
8	0.56	0.12	0.14	0.18

$\mathcal{P}_{SI}$ (which is included in $\mathcal{P}_{Y \nrightarrow X}$) is retrieved when taking the convex hull of the four first probability mass functions. $\mathcal{P}_{Y \nrightarrow X}$ is the convex hull of all eight probability mass functions. Now, assume that we drop constraints (4) and (5) in Proposition 3, and that we consider the following selection: $p_{\mathcal{X}}^{33}(r|w) = 1, p_{\mathcal{X}}^{33}(w|r) = 1, p_{\mathcal{X}}^{32}(r|w) = 1,$

$p_{\mathcal{X}}^{31}(w|r) = 1$, $p_{\mathcal{Y}}^{13}(r) = 1$, $p_{\mathcal{Y}}^{33}(w) = p_{\mathcal{Y}}^{23}(w) = 1$. The probability mass function resulting from such a choice is

$$p(\{r, r\}) = 0.28 \quad p(\{r, w\}) = 0.32$$

$$p(\{r, r\}) = 0.12 \quad p(\{r, w\}) = 0.28$$

and it is not in $\mathcal{P}_{Y \nrightarrow X}$ (i.e. it cannot be formulated as a convex combination of the eight extreme points of $\mathcal{P}_{Y \nrightarrow X}$)

C. Interpreting epistemic irrelevance

Setting aside the practical aspect of interpreting independence concepts in evidence theory (i.e. the possibility to use specific optimisation problems), let us now focus on the interpretation we can give to these results.

The conditions of Propositions 2 and 3 are easier to interpret if a mass assignment m_X is associated to a so-called hierarchical model [7], where the mass assignment is the second-order precise probabilistic model, modelling our uncertainty about the characteristics of the first order model, here reduced to vacuous (i.e. completely imprecise) models $\mathfrak{I}_{A_1}, \ldots, \mathfrak{I}_{A_m}$ corresponding to focal sets. This corresponds to cases where there is a "correct" first order model, but we are unsure about what is it. Couso [3] relates this model to random sets and interpret m as an observational process and focal sets as possible selections for the variables. In Shafer's view [14], the mass assignment describe our uncertainty about what is our actual knowledge about the variable, this knowledge being in the form $X \in A$. De Cooman [7] provides yet other interpretations, and shows that such second-order models can be transformed into equivalent first-order models (at least from a behavioural viewpoint). Actually, interpreting a mass assignment m as a probability set $\mathcal{P}(Bel)$ is such a transformation.

Independence concepts can be reinterpreted with this particular view. For instance, taking the mass assignments product can be interpreted as an assumption of independence between second-order models. It can be independence in the observational process or between the sources (depending on the particular interpretation given to m). Whether we interpret this independence as a stochastic or as an epistemic concept does not formally matter, since the model is precise.

Now, condition 2 and constraints (2),(3) of Proposition 2 indicate that strong independence corresponds to the assumptions (1) that the first order model for each variable is a precise model, whose probability mass function is unknown and depend only of the considered variable and (2) that the joint uncertainty of these first-order model is described by the product of marginal probabilities. This can translate the fact that the two variables X and Y follow random and stochastically independent processes, or that the agent is forced to select precise models at the first order. This is well in accordance with a step-wise application at the second- and first-order levels of the strong independence concept, which makes the assumption that imprecise probabilities model our incomplete knowledge about a precise probability.

Similarly, constraints (4) and (5) of Proposition 3 indicate that epistemic irrelevance corresponds to the assumptions that the first-order models for the conditional uncertainty about the value of X given Y and for Y are unique precise (yet unknown) models. Contrary to the case of strong independence, these constraints appear to be surprising. Indeed, the classical assumption of epistemic irrelevance [16, Ch. 9] makes no reference to an underlying precise probability that is only partially known. Therefore, it is hard to interpret condition 1 and constraints (4), (5) of Proposition 3 as a

step-wise application at the second- and first-order level of epistemic irrelevance. Constraints (4),(5) are therefore questionable, if classical epistemic irrelevance is to be applied between first order models $\mathfrak{I}_{A_1}, \ldots, \mathfrak{I}_{A_m}$ and $\mathfrak{I}_{B_1}, \ldots, \mathfrak{I}_{B_n}$.

An intereesting remark is that, if one drops constraints (4) and (5) of Proposition 3 (applying classical epistemic irrelevance for each combination of focal sets), the obtained probability set would coincide with the one obtained with an assumption of random set independence. This view could then provide some theoretical justification to the use of random set independence, even in the case of imprecise probability theory.

D. Taking one step further: imprecise belief functions

To settle this work in the more general context of generic second-order uncertainty models, the next step is to consider imprecise belief functions. In this setting, uncertainty about X and Y are no longer modelled by single mass assignments but by convex sets $\mathcal{M}_X$ and $\mathcal{M}_Y$ of mass assignments. $\mathcal{M}_X$ and $\mathcal{M}_Y$ focal sets $\mathcal{F}_X = \{A_1, \ldots, A_m\}$ and $\mathcal{F}_Y = \{B_1, \ldots, B_n\}$ are the sets to which at least one mass assignment in $\mathcal{M}_X$ and $\mathcal{M}_Y$ assign strictly positive mass. If imprecise belief functions as uncertainty models have already been discussed by different authors [1], [8], [11], they have not considered the problems of independence assessments and of joint models construction for such models.

Given a set of mass assignment $\mathcal{M}_X$, Augustin [1] has shown that we can come back to a first order probability set $\mathcal{P}(\mathcal{M}_X)$ by considering the following set:

$$\mathcal{P}(\mathcal{M}_X) = \{ \sum_{A_i \in \mathcal{F}_X} m(A_i) p^i | m \in \mathcal{M}_X, p^i \in \mathfrak{I}_{A_i} \}, \quad (9)$$

This means that any probability mass function in $\mathcal{P}(\mathcal{M}_X)$ or $\mathcal{P}(\mathcal{M}_Y)$ can be built by picking a mass assignment resp. in $\mathcal{M}_X$ or $\mathcal{M}_Y$ and probability mass functions resp. inside $\mathfrak{I}_{A_1}, \ldots, \mathfrak{I}_{A_m}$ or $\mathfrak{I}_{B_1}, \ldots, \mathfrak{I}_{B_n}$. Therefore, any joint probability mass function on $\mathcal{X} \times \mathcal{Y}$ having marginals in $\mathcal{P}(\mathcal{M}_X)$ or $\mathcal{P}(\mathcal{M}_Y)$ can be built in the same way as for precise mass assignments. Augustin [1] have also shown that imprecise mass assignments could model any first-order probability set $\mathcal{P}$. This means that the scope of the present (short) study goes beyond the simple framework of evidence theory, and also concerns to some extent generic imprecise probabilities.

As in the previous sections, we can now consider the joint probability set resulting from an independence assumption between marginal sets $\mathcal{P}(\mathcal{M}_X)$ and $\mathcal{P}(\mathcal{M}_Y)$, and search under which conditions it is equivalent to first separately build joint sets over $\mathcal{X} \times \mathcal{Y}$ for second-order and first-order models, and then to consider the obtained equivalent first-order model. For example, the set $\mathcal{P}_{SI}$ resulting from a strong independence assumption between $\mathcal{P}(\mathcal{M}_X)$ and $\mathcal{P}(\mathcal{M}_Y)$ is

$$\mathcal{P}_{SI} = \{ p \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}} | \forall x, y \in \mathcal{X} \times \mathcal{Y}, p(x, y) = p_{\mathcal{X}}(x) p_{\mathcal{Y}}(y), \\ p_{\mathcal{X}} \in \mathcal{P}(\mathcal{M}_X), p_{\mathcal{Y}} \in \mathcal{P}(\mathcal{M}_Y) \}.$$

Again, if we want to relate this independence concept and this joint set with imprecise belief functions, there are three levels at which (in)dependence structures and constraints can be specified. Extending proposition 2, we can show that:

Proposition 4. *Consider the two marginal probability sets $\mathcal{P}(\mathcal{M}_X), \mathcal{P}(\mathcal{M}_Y)$ induced by mass assignment sets $\mathcal{M}_X, \mathcal{M}_Y$ and the joint set $\mathcal{P}_{SI}$. The set of joint probabilities built with the following constraints coincide with the set $\mathcal{P}_{SI}$:*

1) $\forall A_i \times B_j, i = 1, \ldots, n; j = 1, \ldots, m, \ m(A_i \times B_j) = m_X(A_i) m_Y(B_j)$ with $m_X \in \mathcal{M}_X$ and $m_Y \in \mathcal{M}_Y$

2) $\forall i = 1, \ldots, n; j = 1, \ldots, m, \ p^{ij} = p_{\mathcal{X}}^{ij} \cdot p_{\mathcal{Y}}^{ij}$, i.e. p^{ij} is the stochastic product of $p_{\mathcal{X}}^{ij}, p_{\mathcal{Y}}^{ij}$;

3) $\forall i = 1, \ldots, n; j = 1, \ldots, m, \ p_{\mathcal{X}}^i \in \mathfrak{I}_{A_i}, \ p_{\mathcal{Y}}^j \in \mathfrak{I}_{B_j}$ with

$$p_{\mathcal{X}}^i := p_{\mathcal{X}}^{i1} = \ldots = p_{\mathcal{X}}^{im} \quad (10)$$

$$p_{\mathcal{Y}}^j := p_{\mathcal{Y}}^{1j} = \ldots = p_{\mathcal{X}}^{nj} \quad (11)$$

Proof: Let us first show that the described set includes all probabilities inside $\mathcal{P}_S I$. Any probability mass function in $\mathcal{P}_S I$ can be decomposed into the products of two marginals, which by Eq. (9) can be rewritten, for any $(x, y) \in \mathcal{X} \times \mathcal{X}$, as

$$p(x, y) = \sum_{x \in A_i} m_X(A_i) p^i \sum_{y \in B_j} m_Y(B_j) p^j(y),$$

with m_X, m_Y in $\mathcal{M}_X, \mathcal{M}_Y$ and p^i, p^j in $\mathfrak{I}_{A_i}, \mathfrak{I}_{B_j}$. Such decompositions are indeed included in the constraints described in Prop. 4

Let us now show that any joint probability obtained through constraints of Proposition 4 is included in $\mathcal{P}_S I$. First, consider the fact that any joint probability having marginals in $\mathcal{P}(\mathcal{M}_X), \mathcal{P}(\mathcal{M}_Y)$ can be written as

$$p(x, y) = \sum_{j=1}^{m} \sum_{i=1}^{n} m(A_i \times B_j) p^{ij}(x, y),$$

with m a joint mass assignment having marginals in $\mathcal{M}_X$ and $\mathcal{M}_Y$, and $p^{ij} \in \mathfrak{I}_{A_i} \times \mathfrak{I}_{B_i}$. If constraints of Proposition 4 are satisfied, we have

$$p(x, y) = \sum_{i=1}^{m} \sum_{i=1}^{n} m_X(A_i) m_Y(B_j) p^i(x) p^j(y) \quad (12)$$

$$= \sum_{i=1}^{m} m_X(A_i) p^i(x) \sum_{j=1}^{n} m_Y(B_j) p^j(y), \quad (13)$$

by Eq. (9), $\sum_{i=1}^{m} m_X(A_i) p^i \in \mathcal{P}(\mathcal{M}_X)$ and $\sum_{j=1}^{n} m_Y(B_j) p^j \in \mathcal{P}(\mathcal{M}_Y)$, therefore any probability satisfying constraints of Proposition 4 is in $\mathcal{P}_S I$. This shows that the two sets coincide. ∎

Again, it may be the case that some joint probabilities in $\mathcal{P}_{SI}$ can be retrieved by combinations of mass assignments and probabilities on focal sets that do not form a decomposable probability. This result indicates that there should be no major problems in formally extending classical independence concepts to hierarchical models (at least limited to second-order) such as probability sets describing our uncertainty about which probability sets is the "correct" right first-order models. The results also indicates that, at least for strong independence, it is assumed that both second- and first-order probabilities are combined by taking the stochastic product, therefore assuming that there is a precise yet imprecisely known model at both levels. Such precise model could correspond to the hypothesis that variable values follow a random process or that we are forced to pick a precise model and then to consider the product combination.

Indeed, constraints in Proposition 4 shows that to extend strong independence to imprecise belief functions, sets of mass assignments should themselves be combined by assuming strong independence between them (i.e. taking the products of every mass functions in $\mathcal{M}_X$ and $\mathcal{M}_Y$). This comes down to assume that there are "ideal" but unknown precise mass assignments in $\mathcal{M}_X$ and $\mathcal{M}_Y$, and that they are (stochastically) independent. Such an assumption can be questioned, especially if $\mathcal{M}_X$ and $\mathcal{M}_Y$ describe the subjective uncertainty of some agent (expert,...). Indeed, one of the main motivation for lower previsions, possibility theory, the transferable belief models and other imprecise probability theories is that not all belief states can and should be modelled by precise probabilities.

This shows that one has to be careful when modelling joint models when marginal models are hierarchical models, and cannot just apply classical independence concepts to equivalent first-order models without paying close attention to interpretation issues.

IV. CONCLUSION

In this paper, we have extended some previous results obtained by Fetz and Couso concerning independence concepts and their interpretation in evidence theory. We have considered the concept of epistemic irrelevance and the case where belief functions are imprecise. In both cases, our results show that formal extensions can be easily obtained and that classical independence concepts of imprecise probability theory can be reinterpreted in the framework of evidence theory.

However, interpretations that can be associated to these results are at the very least questionable. Indeed, extensions studied in this paper implicitly assume the existence of a precise, yet ill-known, probability distribution describing our uncertainty about the value assumed by a variable or about the first-order model describing our knowledge about this variable. If assuming the existence of such precise models is justified when uncertainty describe a random process, such an assumption is less convincing when modelling the belief of an agent. It is interesting to note that, if one interprets a mass assignment as a hierarchical model where the second-order model is precise, and if one drops the assumption that there exists a precise but unknown probability describing our (first-order) uncertainty, then considering independence at every level corresponds to the notion of random set independence.

There are two directions in which this work could be extended to encompass other models considered in the litterature:

1) relax the assumption that mass assignments bear on crisp sets, and consider the case where "focal sets" are generic probability sets, possibly having specific properties. Such models encompass, for example, fuzzy random variables [5], [13], that have been studied by many in the literature. Extending the current work in this direction would then provide insight about how independence can be handled with such fuzzy random variables.
2) consider generic hierarchical models, where both second-order and first-order models are probability sets. Note that De Cooman [7] shows that for a broad class of such models and from a behavioural point of view, there is an imprecision-precision equivalence, i.e. decisions made by an agent remain the same whether the first-order level model is precise or imprecise. However, results recalled and obtained in this paper suggests that a similar equivalence is unlikely to hold for structural judgements concerning the same hierarchical models.

REFERENCES

[1] T. Augustin. Generalized basic probability assignments. *I. J. of General Systems*, 34(4):451–463, 2005.

[2] B. ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. the marginal case. *I. J. of Approximate Reasoning*, 29(1):47–70, 2002.

[3] I. Couso. Independence concepts in evidence theory. In *Proc. of the 5th Int. Symp. on Imprecise Probability: Theories and Applications*, 2007.

[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] I. Couso and L. Sanchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, 159:237–258, 2008.

[6] F. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artifical Intelligence*, 45:173–195, 2005.

[7] G. de Cooman. Precision-imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. *J. of Statistical Planning and Inference*, 105:175–198, 2002.

[8] T. Denoeux. Reasoning with imprecise belief structures. *Int. J. of Approximate Reasoning*, 20:79–111, 1999.

[9] T. Fetz. Sets of joint probability measures generated by weighted marginal focal sets. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. 2st International Symposium on Imprecise Probabilities and Their Applications*, 2001.

[10] 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.

[11] E. Miranda, G. de Cooman, and I. Couso. Lower previsions induced by multi-valued mappings. *Journal of Statistical Planning and Inference*, 133:173–197, 2005.

[12] M. Oberguggenberger, J. King, and B. Schmelzer. Imprecise probability methods for sensitivity analysis in engineering. In *proc. of the 5th Int. Symp. on Imprecise Probabilities: Theories and Applications*, pages 317–326, 2007.

[13] M. Puri and D. Ralescu. Fuzzy random variables. *J. Math. Anal. Appl.*, 114:409–422, 1986.

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

[15] P. Vicig. Epistemic independence for imprecise probabilities. *Int. J. of approximate reasoning*, 24:235–250, 2000.

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