PMLR: Proceedings of Machine Learning Research, vol. 62, 121-132, 2017 ISIPTA '17

Independent Natural Extension for Infinite Spaces: Williams-Coherence to the Rescue

Jasper De Bock JASPER.DEBOCK@UGENT.BE
Ghent University - imec, IDLab, ELIS
Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde (Belgium)

Abstract

We define the independent natural extension of two local models for the general case of infinite spaces, using both sets of desirable gambles and conditional lower previsions. In contrast to Miranda and Zaffalon (2015), we adopt Williams-coherence instead of Walley-coherence. We show that our notion of independent natural extension always exists—whereas theirs does not—and that it satisfies various convenient properties, including factorisation and external additivity.

Keywords: independent natural extension; epistemic independence; Williams-coherence; infinite spaces; external additivity; factorisation; sets of desirable gambles; conditional lower previsions.

1. Introduction

When probabilities are imprecise, in the sense that they are only partially specified, it is no longer clear what it means for two variables to be independent (Couso et al., 1999). One approach is to apply the standard notion of independence to every element of some set of probability measures. The alternative, called epistemic independence, is to define independence as mutual irrelevance, in the sense that receiving information about one of the variables will not effect our uncertainty model for the other. The advantage of this intuitive alternative is that it has a much wider scope: since epistemic independence is expressed in terms of uncertainty models instead of probabilities, it can easily be applied to a variety of such models, including non-probabilistic ones; we here consider sets of desirable gambles and conditional lower previsions.

When an assessment of epistemic independence is combined with local uncertainty models, it leads to a unique corresponding joint uncertainty model that is called the independent natural extension. If the variables involved can take only a finite number of values, this independent natural extension always exists, and it then satisfies various convenient properties that allow for the design of efficient algorithms (de Cooman et al., 2011; de Cooman and Miranda, 2012). If the variables involved take values in an infinite set, the situation becomes more complicated. On the one hand, for the specific case of lower probabilities, Vicig (2000) managed to obtain results that resemble the finite case. On the other hand, for the more general case of lower previsions, Miranda and Zaffalon (2015) recently found that the independent natural extension may not even exist.

Our present contribution generalises the results of Vicig (2000) to the case of conditional lower previsions, using sets of desirable gambles as an intermediate step. The key technical difference with Miranda and Zaffalon (2015) is that we use Williams-coherence instead of Walley-coherence. This difference turns out to be crucial because our notion of independent natural extension always exists. Furthermore, as we will see, it satisfies the same convenient properties that are known to hold in the finite case, including factorisation and external additivity.

Proofs are provided in the appendix of the arXiv version of this paper (De Bock, 2017), which had to be omitted from the published version because of the page limit constraint.

2. Preliminaries and Notation

We use $\mathbb{N}$ to denote the natural numbers without zero and let $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$. $\mathbb{R}$ is the set of real numbers and $\mathbb{Q}$ is the set of rational numbers. Sign restrictions are imposed with subscripts. For example, we let $\mathbb{R}_{>0}$ be the set of positive real numbers and let $\mathbb{Q}_{\geq 0}$ be the set of non-negative rational numbers. The extended real numbers are denoted by $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{-\infty, +\infty\}$.

For any non-empty set $\mathscr{X}$, the power set of $\mathscr{X}$—the set of all subsets of $\mathscr{X}$—is denoted by $\mathscr{P}(\mathscr{X})$, and we let $\mathscr{P}_\emptyset(\mathscr{X}) \coloneqq \mathscr{P}(\mathscr{X}) \setminus \{\emptyset\}$ be the set of all non-empty subsets of $\mathscr{X}$. Elements of $\mathscr{P}(\mathscr{X})$ are called events. A set of events $\mathscr{B} \subseteq \mathscr{P}(\mathscr{X})$ is called a field if it is non-empty and closed with respect to complements and finite intersections and unions. If it is also closed with respect to countable intersections and unions, it is called a sigma field. A partition of $\mathscr{X}$ is a set $\mathscr{B} \subseteq \mathscr{P}_\emptyset(\mathscr{X})$ of pairwise disjoint non-empty subsets of $\mathscr{X}$ whose union is equal to $\mathscr{X}$. We also adopt the notational trick of identifying $\mathscr{X}$ with the set of atoms $\{\{x\} \colon x \in \mathscr{X}\}$, which allows us to regard $\mathscr{X}$ as a partition of $\mathscr{X}$.

A bounded real-valued function on $\mathscr{X}$ will be called a gamble on $\mathscr{X}$. The set of all gambles on $\mathscr{X}$ is denoted by $\mathscr{G}(\mathscr{X})$, the set of all non-negative gambles on $\mathscr{X}$ is denoted by $\mathscr{G}_{\geq 0}(\mathscr{X})$, and we let $\mathscr{G}_{>0}(\mathscr{X}) \coloneqq \mathscr{G}_{\geq 0}(\mathscr{X}) \setminus \{0\}$ be the set of all non-negative non-zero gambles. For any set of gambles $\mathscr{A} \subseteq \mathscr{G}(\mathscr{X})$, we let

$$\operatorname{posi}(\mathscr{A}) \coloneqq \left\{ \sum_{i=1}^{n} \lambda_i f_i \colon n \in \mathbb{N}, \lambda_i \in \mathbb{R}_{>0}, f_i \in \mathscr{A} \right\} \tag{1}$$

and

$$\mathscr{E}(\mathscr{A}) \coloneqq \operatorname{posi}\left(\mathscr{A} \cup \mathscr{G}_{>0}(\mathscr{X})\right). \tag{2}$$

Indicators are a particular type of gamble. For any $A \in \mathscr{P}(\mathscr{X})$, the corresponding indicator $\mathbb{I}_A$ of A is a gamble in $\mathscr{G}(\mathscr{X})$, defined for all $x \in \mathscr{X}$ by $\mathbb{I}_A(x) \coloneqq 1$ if $x \in A$ and $\mathbb{I}_A(x) \coloneqq 0$ otherwise.

Finally, for any $\mathscr{B} \subseteq \mathscr{P}_\emptyset(\mathscr{X})$, we will also require the notion of a non-negative $\mathscr{B}$-measurable gamble, which we define as a uniform limit of simple $\mathscr{B}$-measurable gambles.

Definition 1 *Let $\mathscr{B} \subseteq \mathscr{P}_\emptyset(\mathscr{X})$. We call $g \in \mathscr{G}_{\geq 0}(\mathscr{X})$ a simple $\mathscr{B}$-measurable gamble if there are $c_0 \in \mathbb{R}_{\geq 0}$, $n \in \mathbb{N}_0$ and, for all $i \in \{1, \ldots, n\}$, $c_i \in \mathbb{R}_{\geq 0}$ and $B_i \in \mathscr{B}$, such that $g = c_0 + \sum_{i=1}^{n} c_i \mathbb{I}_{B_i}$.*

Definition 2 *Let $\mathscr{B} \subseteq \mathscr{P}_\emptyset(\mathscr{X})$. A gamble $g \in \mathscr{G}_{\geq 0}(\mathscr{X})$ is $\mathscr{B}$-measurable if it is a uniform limit of non-negative simple $\mathscr{B}$-measurable gambles, in the sense that there is a sequence $\{g_n\}_{n \in \mathbb{N}}$ of simple $\mathscr{B}$-measurable gambles in $\mathscr{G}_{\geq 0}(\mathscr{X})$ such that* $\lim_{n \to +\infty} \sup |g - g_n| = 0$.

Readers that are familiar with the concepts of simple and measurable functions that are common in measure theory will observe some similarities. However, there are also some important differences. On the one hand, our definitions are more restrictive: we only consider bounded non-negative functions, Definition 1 requires that the coefficients c_i are non-negative, and Definition 2 considers uniform limits instead of pointwise limits. On the other hand, our definitions are more general because we allow for $\mathscr{B}$ to be any subset of $\mathscr{P}_\emptyset(\mathscr{X})$. Nevertheless, if $\mathscr{B} \cup \{\emptyset\}$ is a sigma field, we have the following equivalence.

Proposition 3 *Consider any $\mathscr{B} \subseteq \mathscr{P}_{\emptyset}(\mathscr{X})$ such that $\mathscr{B}^* := \mathscr{B} \cup \{\emptyset\}$ is a sigma field. Then for any $g \in \mathscr{G}_{\geq 0}(\mathscr{X})$, g is $\mathscr{B}^*$-measurable in the measure-theoretic sense (Nielsen, 1997, Definition 10.1) if and only if it is $\mathscr{B}$-measurable in the sense of Definition 2.*

The proof of this result is based on the following sufficient condition for $\mathscr{B}$-measurability, which provides a convenient tool for establishing the $\mathscr{B}$-measurability of a given function. In particular, it implies that every non-negative gamble is $\mathscr{P}_{\emptyset}(\mathscr{X})$-measurable.

Proposition 4 *Let $\mathscr{B} \subseteq \mathscr{P}_{\emptyset}(\mathscr{X})$ and $g \in \mathscr{G}_{\geq 0}(\mathscr{X})$. If, for all $r \in \mathbb{Q}_{\geq 0}$, the set $\{x \in \mathscr{X} : g(x) \geq r\}$ is a finite union of pairwise disjoint events in $\mathscr{B} \cup \{\mathscr{X}, \emptyset\}$, then g is $\mathscr{B}$-measurable.*

Corollary 5 *Every $g \in \mathscr{G}_{\geq 0}(\mathscr{X})$ is $\mathscr{P}_{\emptyset}(\mathscr{X})$-measurable.*

3. Modelling Uncertainty

A subject's uncertainty about a variable X that takes values x in some non-empty set $\mathscr{X}$ can be mathematically represented in various ways. The most popular such method is perhaps probability theory, but it is by no means the only one, nor is it the most general one. In order for our results to have a broader scope, we here adopt the frameworks of sets of desirable gambles and conditional lower previsions.

The main aim of this section is to provide an overview of the basic technical aspects of these frameworks, as these will be essential to the rest of the paper. Notably, we do not impose any constraints on the cardinality of $\mathscr{X}$: it may be finite, countably infinite or uncountably infinite. Connections with other—perhaps better known—models for uncertainty, including probability theory, will be discussed briefly at the end.

The basic idea behind *sets of desirable gambles* is to model a subject's uncertainty about X by considering his attitude towards gambles—bets—on $\mathscr{X}$. In particular, we consider the gambles $f \in \mathscr{G}(\mathscr{X})$ that he finds *desirable*, in the sense that he is willing to engage in a transaction where, once the actual value $x \in \mathscr{X}$ of X is known, he will receive a—possibly negative—reward $f(x)$ in some linear utility scale. Even more so, he prefers these desirable gambles over the status quo, that is, over not conducting any transaction at all. A set of desirable gambles is called coherent if it satisfies the following rationality requirements.

Definition 6 *A coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}$ is a subset of $\mathscr{G}(\mathscr{X})$ such that, for any two gambles $f, g \in \mathscr{G}(\mathscr{X})$ and any non-negative real number $\lambda \in \mathbb{R}_{>0}$:*

- D1: *if $f \geq 0$ and $f \neq 0$, then $f \in \mathscr{D}$;*
- D2: *if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$;*
- D3: *if $f, g \in \mathscr{D}$, then $f + g \in \mathscr{D}$;*
- D4: *if $f \leq 0$, then $f \notin \mathscr{D}$.*

Despite their simplicity, sets of desirable gambles offer a surprisingly powerful framework for modelling uncertainty; see for example (Walley, 2000) and (Quaeghebeur, 2014). For our present purposes though, all we need for now is Definition 6.

Conditional lower previsions also model a subject's uncertainty about X by considering his attitude towards gambles on $\mathscr{X}$. However, in this case, instead of considering sets of gambles, we consider the prices at which a subject is willing to buy these gambles. Let

$$\mathscr{C}(\mathscr{X}) \coloneqq \mathscr{G}(\mathscr{X}) \times \mathscr{P}_{\emptyset}(\mathscr{X})$$

be the set of all pairs (f,B), where f is a gamble on $\mathscr{X}$ and B is a non-empty subset of $\mathscr{X}$—an event. A conditional lower prevision is then defined as follows.

Definition 7 *A conditional lower prevision $\underline{P}$ on $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X})$ is a map*

$$\underline{P}\colon \mathscr{C} \to \overline{\mathbb{R}}\colon (f,B) \to \underline{P}(f|B).$$

For any (f,B) in the domain $\mathscr{C}$, the lower prevision $\underline{P}(f|B)$ of f conditional on B is interpreted as a subject's supremum price μ for buying f, under the condition that the transaction is called off when B does not happen—if $x \notin B$. In other words, $\underline{P}(f|B)$ is the supremum value of μ for which he is willing to engage in a transaction where he receives $f(x) - \mu$ if $x \in B$ and zero otherwise, and furthermore prefers this transaction to the status quo.

It is also possible to consider conditional upper previsions $\overline{P}(f|B)$, which are interpreted as infimum selling prices. However, since selling f for μ is equivalent to buying $-f$ for $-\mu$, we have that $\overline{P}(f|B) = -\underline{P}(-f|B)$. For that reason, we will mainly focus on conditional lower previsions. Unconditional lower previsions correspond to the special case where $B = \mathscr{X}$ for all $(f,B) \in \mathscr{C}$; we then use the shorthand notation $\underline{P}(f) \coloneqq \underline{P}(f|\mathscr{X})$ and call $\underline{P}(f)$ the lower prevision of f. Similarly, we refer to $\overline{P}(f) \coloneqq \overline{P}(f|\mathscr{X})$ as the upper prevision of f.

Because of their interpretation in terms of buying prices for gambles, a particularly intuitive way to obtain a conditional lower prevision $\underline{P}$ is to derive it from a set of gambles $\mathscr{D}$. In particular, for every $\mathscr{D} \subseteq \mathscr{G}(\mathscr{X})$, we let

$$\underline{P}_{\mathscr{D}}(f|B) \coloneqq \sup\{\mu \in \mathbb{R}\colon [f-\mu]\mathbb{I}_B \in \mathscr{D}\} \text{ for all } (f,B) \in \mathscr{C}(\mathscr{X}). \tag{3}$$

A conditional lower prevision is then called coherent if can be derived from a coherent set of desirable gambles in this way.

Definition 8 *A conditional lower prevision $\underline{P}$ on a domain $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X})$ is coherent if there is a coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}$ such that $\underline{P}$ coincides with $\underline{P}_{\mathscr{D}}$ on $\mathscr{C}$.*

This definition of coherence is heavily inspired by the work of Williams (1975, 2007). The only two minor differences are that our rationality axioms on $\mathscr{D}$ are slightly different from his, and that we do not impose any structure on the domain $\mathscr{C}$. Nevertheless, when the domain $\mathscr{C}$ satisfies the structural constraints in (Williams, 2007), Definition 8 is equivalent to that of Williams. More generally, as the following result establishes, it is equivalent to the structure-free notion of Williams-coherence that was developed by Pelessoni and Vicig (2009).

Proposition 9 *A conditional lower prevision $\underline{P}$ on $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X})$ is coherent if and only if it is real-valued and, for all $n \in \mathbb{N}_0$ and all choices of $\lambda_0, \ldots, \lambda_n \in \mathbb{R}_{\geq 0}$ and $(f_0,B_0), \ldots, (f_n,B_n) \in \mathscr{C}$:*

$$\sup_{x \in B}\Big(\sum_{i=1}^{n} \lambda_i \mathbb{I}_{B_i}(x)[f_i(x) - \underline{P}(f_i|B_i)] - \lambda_0 \mathbb{I}_{B_0}(x)[f_0(x) - \underline{P}(f_0|B_0)]\Big) \geq 0, \tag{4}$$

where we let $B \coloneqq \cup_{i=0}^{n} B_i$.

The advantage of this alternative characterisation is that it is expressed directly in terms of lower previsions. Nevertheless, we consider Equation (4) to be less intuitive than Definition 8, which is why we prefer the latter.

From a mathematical point of view, Definition 8 also has the advantage that it allows for simple and elegant proofs of some well-known results. For example, it follows trivially from our definition of coherence that the domain of a coherent conditional lower prevision can be arbitrarily extended while preserving coherence, whereas deriving this result directly from Equation 4 is substantially more involved; see for example the proof of (Pelessoni and Vicig, 2009, Proposition 1). Furthermore, our definition also allows for a very natural derivation of the so-called *natural extension* of $\underline{P}$, that is, the most conservative extension of $\underline{P}$ to $\mathscr{C}(\mathscr{X})$. In particular, instead of having to derive this natural extension directly, Definition 8 allows us to rephrase this problem into a closely related yet simpler question: what is the smallest coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}$ such that $\underline{P}_{\mathscr{D}}$ coincides with $\underline{P}$ on $\mathscr{C}$? The answer turns out to be surprisingly simple.

Proposition 10 *Consider a coherent conditional lower prevision $\underline{P}$ on $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X})$ and let*

$$\mathscr{A}_{\underline{P}} \coloneqq \left\{[f-\mu]\mathbb{I}_B \colon (f,B) \in \mathscr{C}, \mu < \underline{P}(f|B)\right\} \text{ and } \mathscr{E}(\underline{P}) \coloneqq \mathscr{E}(\mathscr{A}_{\underline{P}}). \tag{5}$$

Then $\mathscr{E}(\underline{P})$ is a coherent set of desirable gambles on $\mathscr{X}$ and $\underline{P}_{\mathscr{E}(\underline{P})}$ coincides with $\underline{P}$ on $\mathscr{C}$. Furthermore, for any other coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}$ such that $\underline{P}_{\mathscr{D}}$ coincides with $\underline{P}$ on $\mathscr{C}$, we have that $\mathscr{E}(\underline{P}) \subseteq \mathscr{D}$.

Abstracting away some technical details, the reason why this result holds should be intuitively clear. First, since conditional lower previsions are interpreted as called-off supremum buying prices, we see that the gambles in $\mathscr{A}_{\underline{P}}$ should be desirable. Combined with D1–D3, the desirability of the gambles in $\mathscr{E}(\underline{P})$ then follows.

Since smaller sets of desirable gambles lead to more conservative—pointwise smaller—lower previsions, we conclude that the natural extension of $\underline{P}$ is given by

$$\underline{E}(f|B) \coloneqq \underline{P}_{\mathscr{E}(\underline{P})}(f|B) \text{ for all } (f,B) \in \mathscr{C}(\mathscr{X}). \tag{6}$$

The following proposition provides a formal statement of this result.

Proposition 11 *Let $\underline{P}$ be a coherent conditional lower prevision on $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X})$. Then $\underline{E}$, as defined by Equation (6), is the pointwise smallest coherent conditional lower prevision on $\mathscr{C}(\mathscr{X})$ that coincides with $\underline{P}$ on $\mathscr{C}$.*

All in all, we conclude that Definition 8 provides an intuitive as well as mathematically convenient characterisation of Williams-coherence that is furthermore equivalent to the structure-free version of Pelessoni and Vicig (2009). From a technical point of view, this equivalence will not be important further on, since all of our arguments will be based on the connection with sets of desirable gambles. From a practical point of view though, this equivalence is highly important, because the Williams-coherent conditional lower previsions that are considered in (Pelessoni and Vicig, 2009) are well-known to include as special cases a variety of other uncertainty models, including expectations, lower expectations, probabilities, lower probabilities and belief functions; lower probabilities, for example, can be obtained by restricting the domain of $\underline{P}$ to indicators. For that reason, all of our results can be immediately applied to these special cases as well. A detailed treatment of these special cases, however, does not fit within the page constraints of this contribution, and therefore falls beyond the scope our present work.

4. Epistemic Independence

Having introduced our main tools for modelling uncertainty, the next step towards developing a notion of independent natural extension is to agree on what we mean by independence. Within the context of lower previsions, there are basically two main options.

The first approach, which we will not consider here, is to interpret lower previsions as lower expectations, that is, as tight lower bounds on the expectations that correspond to some set of probability measures, and to then impose the usual notion of independence on each of the probability measures in that set. This approach has the advantage of being familiar, but is restricted in scope because it can only be applied to uncertainty models that are expressed in terms of probabilities.

The second approach, which is the one that we will adopt here, is to regard independence as an assessment of mutual irrelevance. In particular, we say that X_1 and X_2 are independent if our uncertainty model for X_1 is not affected by conditioning on information about X_2, and vice versa. This definition can easily be applied to a probability measure, and then yields the usual notion of independence. However, and that is what makes this approach powerful and intuitive, it can just as easily be applied to lower previsions, sets of desirable gambles, or any other type of uncertainty model. This type of independence is usually referred to as epistemic independence. The aim of this section is to formalize this concept for the case of two variables, in terms of sets of desirable gambles and conditional lower previsions.

Consider two variables X_1 and X_2 where, for every $i \in \{1,2\}$, X_i takes values x_i in a non-empty set $\mathscr{X}_i$ that may be uncountably infinite, and let $X \coloneqq (X_1,X_2)$ be the corresponding joint variable that takes values $x \coloneqq (x_1,x_2)$ in $\mathscr{X}_1 \times \mathscr{X}_2$. In this context, whenever convenient, we will identify $B \in \mathscr{P}_\emptyset(\mathscr{X}_1)$ with $B \times \mathscr{X}_2$ and $B \in \mathscr{P}_\emptyset(\mathscr{X}_2)$ with $\mathscr{X}_1 \times B$. Similarly, for any $i \in \{1,2\}$, we will identify $f \in \mathscr{G}(\mathscr{X}_i)$ with its cylindrical extension to $\mathscr{G}(\mathscr{X}_1 \times \mathscr{X}_2)$, defined by

$$f(x_1,x_2) \coloneqq f(x_i) \text{ for all } x = (x_1,x_2) \in \mathscr{X}_1 \times \mathscr{X}_2.$$

In order to make this explicit, we will then often denote this cylindrical extension by $f(X_i)$. In this way, for example, for any $f \in \mathscr{G}(\mathscr{X}_2)$ and $B \in \mathscr{P}(\mathscr{X}_1)$, we can write $f(X_2)\mathbb{I}_B(X_1)$ to denote a gamble in $\mathscr{G}(\mathscr{X}_1 \times \mathscr{X}_2)$ whose value in (x_1,x_2) is equal to $f(x_2)$ if $x_1 \in B$ and equal to zero otherwise. Using these conventions, for any set of gambles $\mathscr{D}$ on $\mathscr{X}_1 \times \mathscr{X}_2$, we define the marginal models

$$\operatorname{marg}_1(\mathscr{D}) \coloneqq \{f \in \mathscr{G}(\mathscr{X}_1)\colon f(X_1) \in \mathscr{D}\} \text{ and } \operatorname{marg}_2(\mathscr{D}) \coloneqq \{f \in \mathscr{G}(\mathscr{X}_2)\colon f(X_2) \in \mathscr{D}\}$$

and, for any events $B_1 \in \mathscr{P}_\emptyset(\mathscr{X}_1)$ and $B_2 \in \mathscr{P}_\emptyset(\mathscr{X}_2)$, the conditional models

$$\operatorname{marg}_1(\mathscr{D}\rfloor B_2) \coloneqq \{f \in \mathscr{G}(\mathscr{X}_1)\colon f(X_1)\mathbb{I}_{B_2}(X_2) \in \mathscr{D}\}$$

and

$$\operatorname{marg}_2(\mathscr{D}\rfloor B_1) \coloneqq \{f \in \mathscr{G}(\mathscr{X}_2)\colon f(X_2)\mathbb{I}_{B_1}(X_1) \in \mathscr{D}\}.$$

Conditioning and marginalisation both preserve coherence: if $\mathscr{D}$ is a coherent set of desirable gambles on $\mathscr{X}_1 \times \mathscr{X}_2$, then $\operatorname{marg}_1(\mathscr{D})$ and $\operatorname{marg}_1(\mathscr{D}\rfloor B_2)$ are coherent sets of desirable gambles on $\mathscr{X}_1$, and $\operatorname{marg}_2(\mathscr{D})$ and $\operatorname{marg}_2(\mathscr{D}\rfloor B_1)$ are coherent sets of desirable gambles on $\mathscr{X}_2$.

That being said, let us now recall our informal definition of epistemic independence, which was that the uncertainty model for X_1 is not affected by conditioning on information about X_2, and vice versa. In the context of sets of desirable gambles, this can now be formalized as follows:

$$\operatorname{marg}_1(\mathscr{D}|B_2) = \operatorname{marg}_1(\mathscr{D}) \text{ and } \operatorname{marg}_2(\mathscr{D}|B_1) = \operatorname{marg}_2(\mathscr{D}).$$

The only thing that is left to specify are the conditioning events B_1 and B_2 for which we want this condition to hold. We think that the most intuitive approach is to impose this for every $B_1 \in \mathscr{P}_\emptyset(\mathscr{X}_1)$ and $B_2 \in \mathscr{P}_\emptyset(\mathscr{X}_2)$, and will call this epistemic subset-independence. However, this is not what is usually done. The conventional approach, which we will refer to as epistemic value-independence, is to focus on singleton events of the type $B_1 = \{x_1\}$ and $B_2 = \{x_2\}$; see for example (Walley, 1991) and (de Cooman and Miranda, 2012). We believe this conventional approach to be flawed and will argue against it further on. Until then, we postpone this debate by adopting a very general approach that subsumes the former two as special cases. In particular, for every $i \in \{1,2\}$, we simply fix a generic set of conditioning events $\mathscr{B}_i \subseteq \mathscr{P}_\emptyset(\mathscr{X}_i)$. Epistemic value-independence corresponds to choosing $\mathscr{B}_i = \mathscr{X}_i$, whereas epistemic subset-independence corresponds to choosing $\mathscr{B}_i = \mathscr{P}_\emptyset(\mathscr{X}_i)$.

For sets of desirable gambles, this leads us to the following definition.

Definition 12 *Let $\mathscr{D}$ be a coherent set of desirable gambles on $\mathscr{X}_1 \times \mathscr{X}_2$. Then $\mathscr{D}$ is epistemically independent if, for any i and j such that $\{i,j\} = \{1,2\}$:*

$$\mathrm{marg}_i(\mathscr{D}\rfloor B_j) = \mathrm{marg}_i(\mathscr{D}) \text{ for all } B_j \in \mathscr{B}_j.$$

For coherent lower previsions, as a prerequisite for defining epistemic independence, we require that the domain $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X}_1 \times \mathscr{X}_2)$ is independent, by which we mean that for any i and j such that $\{i,j\} = \{1,2\}$, any pair $(f_i, B_i) \in \mathscr{C}(\mathscr{X}_i)$ and any event $B_j \in \mathscr{B}_j$:

$$(f_i, B_i) \in \mathscr{C} \Leftrightarrow (f_i, B_i \cap B_j) \in \mathscr{C}. \tag{7}$$

Other than that, we impose no restrictions on $\mathscr{C}$; its elements $(f,B) \in \mathscr{C}$ are for example not restricted to the types that appear in Equation (7). As a result, the following definition of epistemic independence is applicable beyond the context of lower previsions. For example, by restricting the domain to indicators, we obtain a notion of epistemic independence that applies to conditional lower probabilities. A detailed discussion of these special cases, however, is left as future work.

Definition 13 *Let $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X}_1 \times \mathscr{X}_2)$ be an independent domain. A coherent conditional lower prevision $\underline{P}$ on $\mathscr{C}$ is then epistemically independent if, for any i and j such that $\{i,j\} = \{1,2\}$:*

$$\underline{P}(f_i|B_i) = \underline{P}(f_i|B_i \cap B_j) \text{ for all } (f_i, B_i) \in \mathscr{C} \text{ and } B_j \in \mathscr{B}_j.$$

Another important feature of this definition is that B_j is not only irrelevant to unconditional local lower previsions of the form $\underline{P}(f_i)$—in the sense that $\underline{P}(f_i) = \underline{P}(f_i|B_j)$—but also to conditional local lower previsions such as $\underline{P}(f_i|B_i)$—in the sense that $\underline{P}(f_i|B_i) = \underline{P}(f_i|B_i \cap B_j)$. This type of irrelevance is called h-irrelevance; see Cozman (2013) and De Bock (2015). Note, however, that this feature is optional within our framework; it only appears when $\mathscr{C}$ is sufficiently large. If $B_i = \mathscr{X}_i$ for all $(f_i, B_i) \in \mathscr{C}$, our definition reduces to the simple requirement that $\underline{P}(f_i) = \underline{P}(f_i|B_j)$.

5. The Independent Natural Extension

All of that being said, we are now finally ready to introduce our central object of interest, which is the *independent natural extension*. Basically, the question to which this concept provides an answer is always the same: given two local uncertainty models and an assessment of epistemic independence, what then should be the corresponding joint model? The answer, however, depends on the specific framework that is being considered.

Within the framework of sets of desirable gambles, the local uncertainty models are coherent sets of desirable gambles. In particular, for each $i \in \{1,2\}$, we are given a coherent set of desirable gambles $\mathscr{D}_i$ on $\mathscr{X}_i$. The aim is to combine these local models with an assessment of epistemic independence to obtain a coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}_1 \times \mathscr{X}_2$. The first requirement on $\mathscr{D}$, therefore, is that it should have $\mathscr{D}_1$ and $\mathscr{D}_2$ as its marginals, in the sense that $\mathrm{marg}_i(\mathscr{D}) = \mathscr{D}_i$ for all $i \in \{1,2\}$. The second is that $\mathscr{D}$ should be epistemically independent. If both requirements are met, $\mathscr{D}$ is called an independent product of $\mathscr{D}_1$ and $\mathscr{D}_2$. The most conservative among these independent products is called the independent natural extension.

Definition 14 *An independent product of $\mathscr{D}_1$ and $\mathscr{D}_2$ is an epistemically independent coherent set of desirable gambles $\mathscr{D}$ on $\mathscr{X}_1 \times \mathscr{X}_2$ that has $\mathscr{D}_1$ and $\mathscr{D}_2$ as its marginals.*

Definition 15 *The independent natural extension of $\mathscr{D}_1$ and $\mathscr{D}_2$ is the smallest independent product of $\mathscr{D}_1$ and $\mathscr{D}_2$.*

If all we know is that $\mathscr{D}$ is epistemically independent and has $\mathscr{D}_1$ and $\mathscr{D}_2$ as its marginal models, then the safest choice for $\mathscr{D}$—the only choice that does not require any additional assessments—is their independent natural extension, provided of course that it exists. In order to show that it always does, we let

$$\mathscr{D}_1 \otimes \mathscr{D}_2 \coloneqq \mathscr{E}\left(\mathscr{A}_{1\to 2} \cup \mathscr{A}_{2\to 1}\right), \tag{8}$$

with

$$\mathscr{A}_{1\to 2} \coloneqq \{f_2(X_2)\mathbb{I}_{B_1}(X_1) \colon f_2 \in \mathscr{D}_2, B_1 \in \mathscr{B}_1 \cup \{\mathscr{X}_1\}\} \tag{9}$$

and

$$\mathscr{A}_{2\to 1} \coloneqq \{f_1(X_1)\mathbb{I}_{B_2}(X_2) \colon f_1 \in \mathscr{D}_1, B_2 \in \mathscr{B}_2 \cup \{\mathscr{X}_2\}\}. \tag{10}$$

The following result establishes that $\mathscr{D}_1 \otimes \mathscr{D}_2$ is the independent natural extension of $\mathscr{D}_1$ and $\mathscr{D}_2$.

Theorem 16 *$\mathscr{D}_1 \otimes \mathscr{D}_2$ is the independent natural extension of $\mathscr{D}_1$ and $\mathscr{D}_2$.*

Similar concepts can be defined for conditional lower previsions as well. In that case, the local uncertainty models are coherent conditional lower previsions. In particular, for every $i \in \{1,2\}$, we are given a coherent conditional lower prevision $\underline{P}_i$ on some freely chosen local domain $\mathscr{C}_i \subseteq \mathscr{C}(\mathscr{X}_i)$. The aim is now to construct an epistemically independent coherent conditional lower prevision $\underline{P}$ on $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X}_1 \times \mathscr{X}_2)$ that has $\underline{P}_1$ and $\underline{P}_2$ as its marginals, in the sense that $\underline{P}$ coincides with $\underline{P}_1$ and $\underline{P}_2$ on their local domain: $\underline{P}(f_i|B_i) = \underline{P}_i(f_i|B_i)$ for all $i \in \{1,2\}$ and $(f_i, B_i) \in \mathscr{C}_i$. As before, a model that meets these criteria is then called an independent product, and the most conservative among them is called the independent natural extension. Clearly, in order for these notions to make sense, the global domain $\mathscr{C}$ must at least include the local domains $\mathscr{C}_1$ and $\mathscr{C}_2$ and must furthermore be independent in the sense of Equation (7). The definitions and results below take this for granted.

Definition 17 *An independent product of $\underline{P}_1$ and $\underline{P}_2$ is an epistemically independent coherent conditional lower prevision on $\mathscr{C}$ that has $\underline{P}_1$ and $\underline{P}_2$ as its marginals.*

Definition 18 *The independent natural extension of $\underline{P}_1$ and $\underline{P}_2$ is the point-wise smallest independent product of $\underline{P}_1$ and $\underline{P}_2$.*

Here too, if all we know is that $\underline{P}$ is epistemically independent and has $\underline{P}_1$ and $\underline{P}_2$ as its marginal models, then the safest choice for $\underline{P}$—the only choice that does not require any additional assessments—is the independent natural extension, provided that it exists. The following result establishes that it does, by showing that it is a restriction of the operator $\underline{P}_1 \otimes \underline{P}_2$, defined by

$$(\underline{P}_1 \otimes \underline{P}_2)(f|B) \coloneqq \underline{P}_{\mathscr{D}}(f|B) \text{ for all } (f,B) \in \mathscr{C}(\mathscr{X}_1 \times \mathscr{X}_2), \text{ with } \mathscr{D} = \mathscr{E}(\underline{P}_1) \otimes \mathscr{E}(\underline{P}_2). \tag{11}$$

Theorem 19 *The independent natural extension of $\underline{P}_1$ and $\underline{P}_2$ is the restriction of $\underline{P}_1 \otimes \underline{P}_2$ to $\mathscr{C}$.*

Interestingly, as can be seen from this result, the choice of the joint domain $\mathscr{C}$ does not affect the resulting independent natural extension, in the sense that any $\mathscr{C}$ that includes (f,B) will lead to the same value of $(\underline{P}_1 \otimes \underline{P}_2)(f|B)$. For that reason, we will henceforth assume without loss of generality that $\mathscr{C} = \mathscr{C}(\mathscr{X}_1 \times \mathscr{X}_2)$.

6. On the Choice of Conditioning Events

The fact that the existence results in the previous section are valid regardless of the choice of $\mathscr{B}_1$ and $\mathscr{B}_2$ should not be taken to mean that this choice does not affect the model. In some cases, it most definitely does. In the remainder of this contribution, we will study the extend to which it does, and how it affects the properties of the resulting notion of independent natural extension.

As a first observation, we note that larger sets of conditioning events correspond to stronger assessments of epistemic independence, and therefore lead to more informative joint models. For example, as can be seen from Equations (8)–(10), adding events to $\mathscr{B}_1$ and $\mathscr{B}_2$ leads to a larger—more informative—set of desirable gambles $\mathscr{D}_1 \otimes \mathscr{D}_2$. Similarly, as can be seen from Equation (11), it leads to a joint lower prevision that is higher—and therefore again more informative. There is one important exception to this observation though, which occurs when we add conditioning events that are a finite disjoint union of other conditioning events. In that case, the resulting notion of independent natural extension does not change.

Proposition 20 *For each $i \in \{1,2\}$, let $\mathscr{B}'_i$ be a superset of $\mathscr{B}_i$ that consists of finite disjoint unions of events in $\mathscr{B}_i$. Replacing $\mathscr{B}_1$ by $\mathscr{B}'_1$ and $\mathscr{B}_2$ by $\mathscr{B}'_2$ then has no effect on the resulting independent natural extension $\mathscr{D}_1 \otimes \mathscr{D}_2$ or $\underline{P}_1 \otimes \underline{P}_2$.*

As a particular case of this result, it follows that if $\mathscr{B}_i$ is a finite partition of $\mathscr{X}_i$, we can replace it by the generated algebra—minus the empty event. As an even more particular case, if $\mathscr{X}_1$ and $\mathscr{X}_2$ are finite, we find that epistemic value- and subset-independence lead to the same notion of independent natural extension. For that reason, in the finite case, it does not really matter which of these two types of epistemic independence is adopted.

In the infinite case though, the difference does matter, and the debate between epistemic value- and subset-independence remains open. For lower previsions, Miranda and Zaffalon (2015) recently adopted epistemic value-independence in combination with Walley-coherence. Unfortunately, they found that the corresponding notion of independent natural extension does not always exist. They also considered the combination of epistemic value-independence with Williams-coherence, and argued that the resulting model was too weak. For the case of lower probabilities, Vicig (2000) adopted epistemic subset-independence in combination with Williams-coherence, showed that the corresponding independent natural extension always exists, and proved that it satisfies factorisation properties. Our results so far can be regarded as a generalisation of the existence results of Vicig (2000). As we are about to show, his factorisation results can be generalised as well.

7. Factorisation and External Additivity

When $\mathscr{X}_1$ and $\mathscr{X}_2$ are finite, the independent natural extension of two lower previsions $\underline{P}_1$ and $\underline{P}_2$ is well-known to satisfy the properties of factorisation and external additivity (de Cooman et al., 2011). Factorisation, on the one hand, states that

$$(\underline{P}_1 \otimes \underline{P}_2)(gh) = \underline{P}_1(g\underline{P}_2(h)) = \begin{cases} \underline{P}_1(g)\underline{P}_2(h) & \text{if } \underline{P}_2(h) \geq 0 \\ \overline{P}_1(g)\underline{P}_2(h) & \text{if } \underline{P}_2(h) \leq 0, \end{cases} \tag{12}$$

where g is a non-negative gamble on $\mathscr{X}_1$, h is a gamble on $\mathscr{X}_2$ and $\overline{P}_1(g) \coloneqq -\underline{P}_1(-g)$. By symmetry, the role of 1 and 2 can of course be reversed. External additivity, on the other hand, states that

$$(\underline{P}_1 \otimes \underline{P}_2)(f+h) = \underline{P}_1(f) + \underline{P}_2(h) \tag{13}$$

where f and h are gambles on $\mathscr{X}_1$ and $\mathscr{X}_2$, respectively.

Compared to the properties that are satisfied by the joint expectation of a product measure of two precise probability measures, these notions of factorisation and external additivity are rather weak. For example, for a precise product measure, additivity is not 'external', in the sense that f and h do not have to be defined on separate variables, nor does factorisation require g to be non-negative. Nevertheless, even in this weaker form, these properties remain of crucial practical importance. For example, in the context of credal networks—Bayesian networks whose local models are imprecise—they turned out to be the key to the development of efficient inference algorithms; see for example de Cooman et al. (2010), De Bock and de Cooman (2014) and De Bock (2015). Any notion of independent natural extension that aims to extend such algorithms to infinite spaces, therefore, should preserve some suitable version of Equations (12) and (13).

The aim of this section is to study the extent to which these equations are satisfied by the notion of independent natural extension that was developed in this paper. As we will see, the answer ends up being surprisingly positive.

For all $i \in \{1,2\}$, let $\underline{P}_i$ be a coherent conditional lower prevision on $\mathscr{C}_i \subseteq \mathscr{C}(\mathscr{X}_i)$, let $\underline{E}_i$ be its natural extension to $\mathscr{C}(\mathscr{X}_i)$, and let $\mathscr{B}_i$ be a subset of $\mathscr{P}_\emptyset(\mathscr{X}_i)$. The independent natural extension of $\underline{P}_1$ and $\underline{P}_2$ then satisfies the following three properties, the first of which implies the other two as special cases.

Theorem 21 *Let $\{i,j\} = \{1,2\}$. For any $f \in \mathscr{G}(\mathscr{X}_i)$, $h \in \mathscr{G}(\mathscr{X}_j)$ and $\mathscr{B}_i$-measurable $g \in \mathscr{G}_{\geq 0}(\mathscr{X}_i)$, we then have that*

$$(\underline{P}_1 \otimes \underline{P}_2)(f+gh) = \underline{E}_i\big(f + g\underline{E}_j(h)\big).$$

Corollary 22 (Factorisation) *Let $\{i,j\} = \{1,2\}$. For any $h \in \mathscr{G}(\mathscr{X}_j)$ and any $g \in \mathscr{G}_{\geq 0}(\mathscr{X}_i)$ that is $\mathscr{B}_i$-measurable, we then have that*

$$(\underline{P}_1 \otimes \underline{P}_2)(gh) = \underline{E}_i\big(g\underline{E}_j(h)\big) = \begin{cases} \underline{E}_i(g)\underline{E}_j(h) & \text{if } \underline{E}_j(h) \geq 0; \\ \overline{E}_i(g)\underline{E}_j(h) & \text{if } \underline{E}_j(h) \leq 0. \end{cases}$$

Corollary 23 (External additivity) *For any $f \in \mathscr{G}(\mathscr{X}_1)$ and $h \in \mathscr{G}(\mathscr{X}_2)$, we have that*

$$(\underline{P}_1 \otimes \underline{P}_2)(f+h) = \underline{E}_1(f) + \underline{E}_2(h).$$

In each of these results, if the local domains $\mathscr{C}_1$ and $\mathscr{C}_2$ are sufficiently large—that is, if they include the gambles that appear in the statement of the results—it follows from Proposition 11 that $\underline{E}_i$ and $\underline{E}_j$ can be replaced by $\underline{P}_i$ and $\underline{P}_j$, respectively, and similarly for $\overline{E}_i$ and $\overline{P}_i$.

That being said, let us now go back to the question of whether or not Equations (12) and (13) can be generalised to the case of infinite spaces. For the case of external additivity, it clearly follows from Corollary 23 that the answer is fully positive. Furthermore, this conclusion holds regardless of our choice for $\mathscr{B}_1$ and $\mathscr{B}_2$; they can even be empty. For factorisation, the answer does depend on $\mathscr{B}_1$ and $\mathscr{B}_2$. If we adopt epistemic subset-independence—that is, if we choose $\mathscr{B}_1 = \mathscr{P}_\emptyset(\mathscr{X}_1)$ and $\mathscr{B}_2 = \mathscr{P}_\emptyset(\mathscr{X}_2)$—it follows from Corollaries 5 and 22 that the answer is again fully positive, because $\mathscr{P}_\emptyset(\mathscr{X}_i)$-measurability then holds trivially. If $\mathscr{B}_1 \cup \{\emptyset\}$ and $\mathscr{B}_2 \cup \{\emptyset\}$ are sigma fields, the answer remains fairly positive as well, because Proposition 3 then implies that it suffices for g to be measurable in the usual, measure-theoretic sense. If we adopt epistemic value-independence—that is, if we choose $\mathscr{B}_1 = \mathscr{X}_1$ and $\mathscr{B}_2 = \mathscr{X}_2$—it is necessary for g to be $\mathscr{X}_i$-measurable, which is a rather strong requirement that easily fails. For that reason, we think that for the case of infinite spaces, when it comes to choosing between epistemic value- and subset-independence, the latter should be preferred over the former.

8. Conclusions and Future Work

The main conclusion of this work is that by combining Williams-coherence with epistemic subset-independence, we obtain a notion of independent natural extension that always exists, and that furthermore satisfies factorisation and external additivity. For weaker types of epistemic independence, including epistemic value-independence, the existence result and the external additivity property remain valid, but factorisation then requires measurability conditions.

We foresee several lines of future research. The first, which we expect to be rather straightforward, is to extend our results from the case of two variables to that of any finite number of variables. Next, these extended versions of our results could then be used to develop efficient algorithms for credal networks whose variables take values in infinite spaces, by suitably adapting existing algorithms for the finite case. On the more technical side, it would be useful to see whether our results can be extended to the case of unbounded functions. Finally, for variables that take values in Euclidean space, $\mathscr{B}_1$ and $\mathscr{B}_2$ could be restricted to the Lebesgue measurable events. Combined with an assessment of continuity, we think that this could lead to the development of a notion of independent natural extension that includes sigma additive product measures as a special case.

Acknowledgments

I am a Postdoctoral Fellow of the Research Foundation - Flanders (FWO) and wish to acknowledge its financial support. The research that lead to this paper was conducted during a research visit—funded by an FWO travel grant—to the Imprecise Probability Group of IDSIA (Institute Dalle Molle for Artificial Intelligence), the members of which I would like to thank for their warm hospitality. Finally, I would also like to thank two anonymous reviewers, for their generous constructive comments, and Enrique Miranda, for commenting on a preliminary version of this paper and for suggesting the idea of adopting a general notion of epistemic independence where $\mathscr{B}_1$ and $\mathscr{B}_2$ are allowed to be arbitrary.

References

I. Couso, S. Moral, and P. Walley. Examples of Independence for Imprecise Probabilities. In *ISIPTA '99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications*, pages 121–130. 1999.

F. G. Cozman. Independence for sets of full conditional probabilities, sets of lexicographic probabilities, and sets of desirable gambles. In *ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*, pages 87–97. 2013.

J. De Bock. *Credal networks under epistemic irrelevance: theory and algorithms*. PhD thesis, Ghent University, 2015.

J. De Bock. *Independent Natural Extension for Infinite Spaces: Williams-Coherence to the Rescue*. ArXiv report 1701.07295, 2017.

J. De Bock and G. de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. *Journal of Artificial Intelligence Research*, 50:189–233, 2014.

G. de Cooman and E. Miranda. Irrelevant and independent natural extension for sets of desirable gambles. *Journal of Artificial Intelligence Research*, 45:601–640, 2012.

G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9): 1029–1052, 2010.

G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. *Artificial Intelligence*, 175(12):1911–1950, 2011.

E. Miranda and M. Zaffalon. Independent products in infinite spaces. *Journal of Mathematical Analysis and Applications*, 425(1):460 – 488, 2015.

O. A. Nielsen. *An introduction to integration and measure theory*. Wiley, 1997.

R. Pelessoni and P. Vicig. Williams coherence and beyond. *International Journal of Approximate Reasoning*, 50(4):612–626, 2009.

E. Quaeghebeur. Desirability. In T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 1–27. John Wiley & Sons, 2014.

P. Vicig. Epistemic independence for imprecise probabilities. *International Journal of Approximate Reasoning*, 24(2-3):235–250, 2000.

P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148, 2000.

P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, 1975.

P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.