Independence for Full Conditional Probabilities: Structure, Factorization, Non-uniqueness, and Bayesian Networks

Fabio G. Cozman[a]

[a]*Universidade de São Paulo – Av. Prof. Mello Moraes, 2231, São Paulo, SP – Brazil*

Abstract

This paper examines concepts of independence for full conditional probabilities; that is, for set-functions that encode conditional probabilities as primary objects, and that allow conditioning on events of probability zero. Full conditional probabilities have been used in economics, in philosopy, in statistics, in artificial intelligence. This paper characterizes the structure of full conditional probabilities under various concepts of independence; limitations of existing concepts are examined with respect to the theory of Bayesian networks. The concept of layer independence (factorization across layers) is introduced; this seems to be the first concept of independence for full conditional probabilities that satisfies the graphoid properties of Symmetry, Redundancy, Decomposition, Weak Union, and Contraction. A theory of Bayesian networks is proposed where full conditional probabilities are encoded using infinitesimals, with a brief discussion of hyperreal full conditional probabilities.

Keywords: Full conditional probabilities, Coherent probabilities, Independence concepts, Graphoid properties, Bayesian networks

1. Introduction

A standard probability measure is a real-valued, non-negative, countably additive set-function, such that the possibility space gets probability 1. In fact, if the space is finite, as we assume in this paper, there is no need to be concerned with countable additivity, and one deals only with finite additivity. In standard probability theory, the primitive concept is the "unconditional" probability $\mathbb{P}(A)$ of an event A; from this concept one defines conditional probability $\mathbb{P}(A|B)$ of event A given event B, as the ratio $\mathbb{P}(A \cap B) / \mathbb{P}(B)$. This definition however is only enforced if $\mathbb{P}(B) > 0$; otherwise, the conditional probability $\mathbb{P}(A|B)$ is left undefined.

A full conditional probability is a real-valued, non-negative set-function, but now the primitive concept is the conditional probability $\mathbb{P}(A|B)$ for event A given event B. This quantity is only restricted by the relationship $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$. Note that $\mathbb{P}(A|B)$ is a well-defined quantity even if $\mathbb{P}(B) = 0$.

Full conditional probabilities offer an alternative to standard probabilities that has found applications in economics [6, 7, 8, 35], decision theory [26, 45] and statistics [9, 40], in philosophy [24, 33], and in artificial intelligence, particularly in dealing with default reasoning [1, 11, 13, 15, 23, 30]. Applications in statistics and artificial intelligence are usually connected with the theory of *coherent probabilities*; indeed, a set of probability assessments is said to be *coherent* if and only if the assessments can be extended to a full conditional probability on some suitable space [19, 28, 39, 45]. Full conditional probabilities are related to other uncertainty representations such as lexicographic probabilities [7, 30], and hyperreal probabilities [25, 27].

In this paper we study concepts of independence applied to full conditional probabilities. We characterize the structure of joint full conditional probabilities when various judgments of independence are enforced. We examine difficulties caused by failure of some graphoid properties and by non-uniqueness of joint probabilities under judgments of independence. We discuss such difficulties within the usual theory of Bayesian networks [38].

We then propose the concept of *layer independence* as it satisfies the graphoid properties of Symmetry, Redundancy, Decomposition, Weak Union, and Contraction. We also propose a theory of Bayesian networks that accommodates full conditional probabilities by resorting to infinitesimals, and comment on a theory of hyperreal full conditional probabilities.

This paper should be relevant to researchers concerned with full conditional probabilities and their applications for instance in game theory and default reasoning, and also relevant to anyone interested in uncertainty modeling where conditional probabilities are the primary object of interest. The paper is organized as follows. Section 2 reviews the necessary background on full conditional probabilities. Section 3 characterizes the structure of full conditional probabilities under various judgments of independence. Section 4 introduces layer factorization, defines layer independence, and analyzes its graphoid properties. Section 5 examines the challenges posed by failure of graphoid properties and non-uniqueness, paying special attention to the theory of Bayesian networks. We suggest a strategy to specify joint full conditional probabilities through Bayesian networks, by resorting to infinitesimals. Section 6 offers brief remarks on a theory of hyperreal full conditional probabilities.

2. Background on full conditional probabilities

In this paper we focus on finite possibility spaces, and take that every subset of the possibility space Ω is an event. Any nonempty event is a *possible* event.

2.1. Axioms

A full conditional probability [20] is a two-place set-function $\mathbb{P} : \mathcal{B} \times (\mathcal{B} \backslash \emptyset) \to \Re$, where $\mathcal{B}$ is a Boolean algebra over a set Ω, such that for every event $C \neq \emptyset$:
(1) $\mathbb{P}(C|C) = 1$;
(2) $\mathbb{P}(A|C) \geq 0$ for every A;

(3) $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C)$ for disjoint A and B;
(4) $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|B \cap C)\,\mathbb{P}(B|C)$ for A and B such that $B \cap C \neq \emptyset$.

Whenever the conditioning event C is equal to Ω, we suppress it and write the "unconditional" probability $\mathbb{P}(A)$. Note that $\mathbb{P}(\Omega|C) = 1$ for any $C \neq \emptyset$, as: $1 = P(C|C) = P(C \cap \Omega|C) = P(C|\Omega \cap C)P(\Omega|C) = P(C|C)P(\Omega|C) = P(\Omega|C)$. If instead $\mathbb{P}(\Omega|C) = 1$ is assumed as an axiom, then the fourth axiom can be derived, in the presence of the others, from the following condition: $\mathbb{P}(A|C) = \mathbb{P}(A|B)\,\mathbb{P}(B|C)$ when $A \subseteq B \subseteq C$ and $B \neq \emptyset$ [17, Section 2].

There are other names for full conditional probabilities in the literature, such as *conditional probabilities* [31], *complete conditional probability systems* [35]. We sometimes use *joint full probability* or *marginal full probability* to emphasize that a particular full conditional probability is defined respectively for a set of variables or for a variable within a set of variables.

Full conditional probabilities allow conditioning on events of zero probability. Indeed, the axioms impose no restriction on the probability of a conditioning event. Consider Axiom (4). If $\mathbb{P}(B|C) = 0$, the constraint $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|B \cap C)\,\mathbb{P}(B \cap C)$ is trivially true, and $\mathbb{P}(A|B \cap C)$ must be elicited by different means.

Example 1. Consider $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and probability assessments

$$\mathbb{P}(\omega_1) = 1, \qquad \mathbb{P}(\omega_2|A) = \alpha, \qquad \mathbb{P}(\omega_3|A) = 1 - \alpha,$$

where $A = \{\omega_2, \omega_3, \omega_4\}$ and $\alpha \in (0, 1)$. Note that $\mathbb{P}(A) = 0$, hence we cannot condition on A in the standard Kolmogorovian setup. Note also that $\mathbb{P}(\omega_4|A) = 0$ given the assessments. $\square$

For a nonempty event C, the set-function $\mathbb{P}(\cdot| \cdot \cap C)$, defined whenever the conditioning event is nonempty, is a full conditional probability. That is, the restriction of a full conditional probability to conditional events $A|(B \cap C)$, for fixed C, remains a full conditional probability. We refer to it as the full conditional probability given C, and denote it by $\mathbb{P}_C$.

A sequence of positive probabilities $\{\mathbb{P}_n\}$ *approximates* a full conditional probability if $\mathbb{P}(A|B) = \lim_{n\to\infty} \mathbb{P}_n(A \cap B)/\mathbb{P}_n(B)$ for any event A and nonempty event B. Any full conditional probability can be associated with such an approximating sequence [36, Theorem 1].

2.2. Variables and their full distributions

Throughout we use letters W, X, Y, Z to denote random variables. For variable X, denote the set of values of X by Ω_X. As the possibility space is finite, there are no issues of measurability.

Whenever possible we use x to denote the event $\{X = x\}$ and $\{y, z\}$ to denote the event $\{Y = y\} \cap \{Z = z\}$, and likewise for similar events. We use y^c to denote $\{Y \neq y\}$.

Letters A, B, C, D denote events. We often write $A(X)$ to indicate that event A belongs to the algebra generated by X (that is, A is equal to $\{\omega : X(\omega) \in A'\}$ for some set A' of real-numbers).

Denote by $\mathbb{P}_X$ the set-function that evaluates to $\mathbb{P}(A(X)|B(X))$ for any events $A(X)$ and $B(X)$ in the algebra generated by X, where $B(X)$ is nonempty. Clearly $\mathbb{P}_X$ is a full conditional probability: the *full distribution* of X.

Example 2. Consider Example 1. Define binary variables as follows:

$$X(\omega_1) = X(\omega_3) = x_0, \qquad X(\omega_2) = X(\omega_4) = x_1,$$
$$Y(\omega_1) = Y(\omega_2) = y_0, \qquad Y(\omega_3) = Y(\omega_4) = y_1,$$

Here $\Omega_X = \{x_0, x_1\}$ and $\Omega_Y = \{y_0, y_1\}$. Then $\mathbb{P}(x_0, y_0) = 1$ (all other atomic events have unconditional probability equal to zero). Also, $\mathbb{P}(x_1, y_0|A) = \alpha$ and $\mathbb{P}(x_0, y_1|A) = 1 - \alpha$ for $A = \{x_1, y_0\} \cup \{x_0, y_1\} \cup \{x_1, y_1\}$. □

2.3. Layers

We can partition the possible elements of Ω into events $L_0, \dots, L_K$ as follows. First, take L_0 to be the set of elements of Ω that have positive unconditional probability. Then take L_1 to be the set of elements of Ω that have positive probability conditional on $\Omega\backslash L_0$. And then take L_i, for $i \in \{2, \dots, K\}$, to be the set of elements of Ω that have positive probability conditional on $\Omega\backslash\cup_{j=0}^{i-1} L_j$. The events L_i are called the *layers* of the full conditional probability. (Note that some authors use a different terminology, using instead the sequence $\{\cup_{j=i}^{K} L_j\}_{i=0}^{K}$ rather than $\{L_i\}_{i=0}^{K}$ [15, 31].)

Example 3. In Example 1, we have $L_0 = \{\omega_1\}$, $L_1 = \{\omega_2, \omega_3\}$, and $L_2 = \{\omega_4\}$. Note that $\mathbb{P}(\omega_4|L_2) = \mathbb{P}(\omega_4|\omega_4) = 1$. □

For nonempty B, denote by L_B the first layer such that $\mathbb{P}(B|L_B) > 0$, and refer to it as the *layer of* B. We then have $\mathbb{P}(A|B) = \mathbb{P}(A|B \cap L_B)$ [6, Lemma 2.1a]. Clearly, if we have $\mathbb{P}_C$, then $\mathbb{P}(A|B \cap C) = \mathbb{P}(A|B \cap C \cap L_{B\cap C})$.

Given an event A and a layer L_i, if $A \cap L_i \neq \emptyset$, then $\mathbb{P}(A|L_i) > 0$. This is true because A must contain some ω that belongs to L_i, and $\mathbb{P}(w|L_i) = \mathbb{P}\big(w|(\cup_{j=i}^{K} L_j) \cap L_i\big) = \mathbb{P}\big(w| \cup_{j=i}^{K} L_i\big) > 0$.

Any full conditional probability can be represented as a sequence of strictly positive probability measures $\mathbb{P}_0, \dots, \mathbb{P}_K$, where the support of $\mathbb{P}_i$ is restricted to L_i. This useful result has been derived by several authors [7, 15, 26, 31].

Example 4. In Example 1, we have $\mathbb{P}_0, \mathbb{P}_1, \mathbb{P}_2$ as follows: $\mathbb{P}_0(\omega_1) = \mathbb{P}(\omega_1|L_0) = 1$; $\mathbb{P}_1(\omega_2) = \mathbb{P}(\omega_2|L_1) = \alpha$ and $\mathbb{P}_1(\omega_3) = \mathbb{P}(\omega_3|L_1) = 1 - \alpha$; and $\mathbb{P}_2(\omega_4) = \mathbb{P}(\omega_4|L_2) = 1$. If we take event $A = \{\omega_2, \omega_3, \omega_4\}$, then $\mathbb{P}(\omega_2|A) = \mathbb{P}(\omega_2|A \cap L_1) = \alpha$. □

Given a full conditional probability and its $K + 1$ layers, we can create an approximating sequence as follows:

$$\mathbb{P}_n = \gamma_n^{-1} \left(\mathbb{P} + \epsilon_n \mathbb{P}_1 + \epsilon_n^2 \mathbb{P}_2 + \dots + \epsilon_n^K \mathbb{P}_K \right), \tag{1}$$

where $\gamma_n = \sum_{i=1}^{K} \epsilon_n^i$ is a sequence of normalization constants, and where $\epsilon_n > 0$ goes to zero as $n \to \infty$ [29]. Such approximating sequences are used later.

Given a variable X, we can consider layers of $\mathbb{P}_X$ (subsets of Ω_X), denoted by L_i^X. The layers L_i^X form a partition of the possible elements of Ω_X.

$\mathbb{P}_0$	y_0	y_1
x_0	1	
x_1		

$\mathbb{P}_1$	y_0	y_1
x_0		$(1-\alpha)$
x_1	α	

$\mathbb{P}_2$	y_0	y_1
x_0		
x_1		1

Table 1: Joint full distribution of binary variables X and Y, with $\alpha \in (0,1)$, specified over three layers.

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1-\alpha \rfloor_1$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor 1 \rfloor_2$

Table 2: Compact representation of joint full distribution for binary variables X and Y.

Example 5. Consider Example 2. Table 1 shows a joint full distribution of (X,Y), as a series of positive distributions over three layers. The marginal full distribution of X is given by $\mathbb{P}_X(x_0) = 1$ and $\mathbb{P}_X(x_1|x_1) = 1$. Hence $L_0^X = \{x_0\}$ and $L_1^X = \{x_1\}$. Similarly, the marginal full distribution of Y is given by $\mathbb{P}_Y(y_0) = 1$ and $\mathbb{P}_Y(y_1|y_1) = 1$. Hence $L_0^Y = \{y_0\}$ and $L_1^Y = \{y_1\}$. □

2.4. Some notation

We often write $\lfloor \alpha \rfloor_i$ to denote a probability value α that belongs to the ith layer L_i.

Example 6. Table 2 shows the full distribution in Example 5, using a compact notation where probability values and layers are conveyed together. □

2.5. Layer numbers

For a nonempty event A, the index i of the first layer L_i such that $\mathbb{P}(A|L_i) > 0$ is the *layer number* of A, denoted by $\circ(A)$. Layer numbers have been studied by Coletti and Scozzafava [15], who refer to them as *zero-layers*. Given a nonempty event B, define the layer number of A given B to be $\circ(A|B) = \circ(A \cap B) - \circ(B)$. Inspired by Coletti and Scozzafava [15], we adopt $\circ(A) = \infty$ if $A = \emptyset$. We have, for $C \neq \emptyset$,

$$\circ(A \cup B|C) = \min(\circ(A|C), \circ(B|C)).$$

Also, if $\circ(A \cap B) > \circ(B)$, then $\mathbb{P}(A|B) = \mathbb{P}(A|B \cap B)\,\mathbb{P}(B|B) = \mathbb{P}(A \cap B|B) = \mathbb{P}(A \cap B|B \cap L_B) = 0$ because $A \cap B$ and L_B must be disjoint. Note that when we write $\circ(x)$ for the event $\{X = x\}$, we must compute the layer number with respect to the underlying full conditional probability, not with respect to the full distribution of X. For instance, in Table 3 we have $\circ(x_1) = 2$, but if we were to focus on the marginal full distribution of X, we would see that the event $\{X = x_1\}$ lies in layer L_1^X. Note also that the conditional layer number $\circ(x|y)$ is computed with respect to the underlying full conditional probability as $\circ(x,y) - \circ(y)$, and it may not be identical to the index of the

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_1$
x_1	$\lfloor 1 \rfloor_3$	$\lfloor 1 \rfloor_2$

Table 3: Joint full distribution of binary variables X and Y.

layer for x in the conditional full distribution $\mathbb{P}_{\{Y=y\}}$. For instance, for the full distribution in Table 3 we have $\circ(x_1, y_0) - \circ(y_0) = 3$, but x_1 lies in the layer of the full conditional probability $\mathbb{P}_{\{Y=y_0\}}$ associated with index 1. To some extent, layer numbers "carry" with them information about the underlying joint full probability.

2.6. Relative probability

A concept of independence discussed later employs *relative probabilities* [3, 29, 34]. A relative probability ρ is a two-place set-function that takes values in $[0, \infty]$, such that for every event A and every nonempty events B and C, we have [29, Definition 2.1]:
(1) $\rho(A; A) = 1$;
(2) $\rho(A \cup B; C) = \rho(A; C) + \rho(B; C)$ if $A \cap B = \emptyset$;
(3) $\rho(A; C) = \rho(A; B)\,\rho(B; C)$ if the latter product is not $0 \times \infty$.

If a relative probability is such that all values of $\rho(\cdot;\cdot)$ are positive and finite, then this relative probability can be represented by a positive probability measure $\mathbb{P}$ by making $\mathbb{P}(A)$ equal to $\rho(A; \Omega)$. The last axiom then implies $\rho(A; B) = \mathbb{P}(A)/\mathbb{P}(B)$. Note however that for any $\alpha > 0$, the measure $\alpha\mathbb{P}$ also offers a representation for the same relative probabilities.

Now if some values of $\rho(\cdot;\cdot)$ are equal to zero, the relative probability can be represented by a full conditional probability with more than one layer. The layers are formed by collecting pairs of atoms whose relative probability is finite; these layers are ordered so that if $\rho(\omega; \omega') = 0$, then $\omega \in L_i$ and $\omega' \in L_j$ with $i > j$. Now for each layer L, define $\mathbb{P}(A|L) = \rho(A; L)$ for any event A in the layer. We then obtain $\rho(A; B) = \mathbb{P}(A|L)/\mathbb{P}(B|L)$ whenever A and B belong to the same layer; otherwise, $\rho(A; B)$ is 0 if $\circ(A) > \circ(B)$ and ∞ if $\circ(A) < \circ(B)$. We thus have:

$$\begin{aligned}\rho(A; B) &= \frac{\mathbb{P}(A|L_{A\cup B})}{\mathbb{P}(B|L_{A\cup B})} = \frac{\mathbb{P}(A|(A\cup B)\cap L_{A\cup B})}{\mathbb{P}(B|(A\cup B)\cap L_{A\cup B})} \times \frac{\mathbb{P}(A\cup B|L_{A\cup B})}{\mathbb{P}(A\cup B|L_{A\cup B})} \\ &= \frac{\mathbb{P}(A|A\cup B)}{\mathbb{P}(B|A\cup B)},\end{aligned}$$

with the understanding that the ratio yields ∞ if its denominator is zero.

A sequence of positive probability measures $\{\mathbb{P}_n\}$ *approximates* a relative probability if $\rho(A; B) = \lim_{n\to\infty} \mathbb{P}_n(A)/\mathbb{P}_n(B)$. It is always possible to find such a sequence of probability measures for any given relative probability [29, footnote 4]; for instance, write down a full conditional probability that represents the relative probability, and generate an approximating sequence for this full conditional probability (Expression (1)).

2.7. Concepts of independence

The standard concept of stochastic independence for variables X and Y given variable Z requires

$$\mathbb{P}(x|y,z) = \mathbb{P}(x|z) \qquad \text{whenever } \mathbb{P}(y,z) > 0. \tag{2}$$

Throughout the paper we ignore Z if it is some constant variable, and discard the expression "given Z" in those cases; then we simply say that X and Y are stochastically independent.

The definition of stochastic independence is too weak for full conditional probabilities: consider Table 3, where X and Y are stochastically independent but

$$\mathbb{P}(y_0) = 1 \neq 0 = \mathbb{P}(y_0|x_1)\,.$$

To avoid this embarrassment, more stringent notions of independence have been proposed for full probabilities [7, 42, 15, 26].

First, say that X is *epistemically irrelevant* to Y given Z if $\mathbb{P}(y|x,z) = \mathbb{P}(y|z)$ whenever $\{x,z\} \neq \emptyset$, and then say that X and Y are *epistemically independent* given Z if X is epistemically irrelevant to Y given Z and vice-versa. Note that epistemic irrelevance is quite weak, and in particular it is not symmetric: in Table 3 we see that Y is epistemically irrelevant to X, but X is not epistemically irrelevant to Y.

Second, say that X is *h-irrelevant* to Y given Z when

$$\mathbb{P}(B(Y)|z \cap A(X) \cap D(Y)) = \mathbb{P}(B(Y)|z \cap D(Y))\,,$$

for all values z, all events $B(Y)$, $D(Y)$ in the algebra generated by Y, and all events $A(X)$ in the algebra generated by X, such that $z \cap A(X) \cap D(Y) \neq \emptyset$. And say that X and Y are *h-independent* given Z when X is h-irrelevant to Y given Z and vice-versa; in that case, we have:

$$\begin{aligned} \mathbb{P}(A(X) \cap B(Y)|z \cap C(X) \cap D(Y)) \quad = & \\ & \mathbb{P}(A(X)|z \cap C(X))\,\mathbb{P}(B(Y)|z \cap D(Y))\,, \end{aligned} \tag{3}$$

for $C(X)$ in the algebra generated by X such that $z \cap C(X) \cap D(Y) \neq \emptyset$. Hammond [26] refers to h-independence simply as "conditional independence," while Battigalli [5] refers to it as the "independence condition"; Swinkels [41] uses the term "quasi-independence" for Battigalli's independence condition, while Kohlberg and Reny [29] employ the term "weak independence" for a condition that is weaker than Battigalli's for several variables but equivalent to Battigalli's for two variables. When dealing with unconditional independence, Hammond and Battigalli assume the possibility space to be a product of possibility spaces ($\Omega = \Omega_X \times \Omega_Y$), while Kohlberg and Reny do not assume it but require, for independence of X and Y, the same condition on Ω. Hence Kohlberg and Reny's version of h-independence is stronger in that it imposes a condition on the possibility space.

Another concept of independence has been proposed by Kohlberg and Reny [29] using relative probabilities. Their proposal is to consider X and Y independent when

$$\rho\left(A(X) \cap B(Y); A'(X) \cap B'(Y)\right) = \lim_{n\to\infty} \frac{\mathbb{P}_n(A(X))\mathbb{P}_n(B(Y))}{\mathbb{P}_n(A'(X))\mathbb{P}_n(B'(Y))} \qquad (4)$$

for all $A(X)$, $A'(X)$ in the algebra generated by X, and all $B(Y)$, $B'(Y)$ in the algebra generated by Y, for some sequence $\{\mathbb{P}_n\}$ of positive product probability measures [29, Definition 2.7]. They show that this concept is equivalent to the following condition [29, Lemma 2.8]: the set of possible values for (X, Y) is the product $\Omega_X \times \Omega_Y$, and moreover there is a sequence $\{\mathbb{P}_n\}$ of positive probability measures such that

$$\frac{\mathbb{P}(x, y|\{x, y\} \cup \{x', y'\})}{\mathbb{P}(x', y'|\{x, y\} \cup \{x', y'\})} = \lim_{n\to\infty} \frac{\mathbb{P}_n(x)\mathbb{P}_n(y)}{\mathbb{P}_n(x')\mathbb{P}_n(y')}. \qquad (5)$$

Kohlberg and Reny [29] then say that X and Y are *strongly independent*; because the term "strong" has been used in the literature before to refer to various concepts of independence, we simply say that X and Y are *kr-independent*. And we say that X and Y are *kr-independent* given Z when they are kr-independent with respect to $\mathbb{P}_{\{Z=z\}}$ for every possible z. Expression (4) can be adapted to such a concept of conditional independence, in the language of full conditional probabilities, as follows: there is a sequence $\{\mathbb{P}_n\}$ of positive probability measures such that, for all events in appropriate algebras,

$$\frac{\mathbb{P}(A(X) \cap B(Y)|((A(X) \cap B(Y)) \cup (A'(X) \cap B'(Y))) \cap \{z\})}{\mathbb{P}(A'(X) \cap B'(Y)|((A(X) \cap B(Y)) \cup (A'(X) \cap B'(Y))) \cap \{z\})} = \lim_{n\to\infty} \frac{\mathbb{P}_n(A(X)|z)\mathbb{P}_n(B(Y)|z)}{\mathbb{P}_n(A'(X)|z)\mathbb{P}_n(B'(Y)|z)}.$$

Expression (5) can be similarly adapted to conditional independence.

Coletti and Scozzafava [12, 13, 14, 15] have proposed conditions on zero-layers to capture aspects of independence. Here *cs-independence* of event B to event A, where $B \neq \emptyset \neq B^c$, requires:

$$\mathbb{P}(A|B) = \mathbb{P}(A|B^c), \quad \circ(A|B) = \circ(A|B^c), \quad \text{and} \quad \circ(A^c|B) = \circ(A^c|B^c). \qquad (6)$$

To understand the motivation for conditions on zero-layers, suppose that $A \cap B$, $A \cap B^c$, $A^c \cap B$ are nonempty, but $A^c \cap B^c = \emptyset$. Hence observation of B^c does provide information about A. However, the indicator functions of A and B may be epistemically/h-independent! Coletti and Scozzafava's conditions fail in this situation: B cannot be independent to A. Hence Coletti and Scozzafava's condition automatically handles logical dependence/independence of events. As noted before, other authors [6, 26, 29] instead require the possibility space to be the product of possibility spaces for the variables.

Vantaggi [42] has presented detailed analysis of Condition (6), and also has presented conditions aimed at independence of variables. First, consider an extension of cs-independence to conditional cs-independence, as follows [13, 42].

Say that B is cs-irrelevant to A given C if $B \cap C \neq \emptyset \neq B^c \cap C$, and $\mathbb{P}(A|B \cap C) = \mathbb{P}(A|B^c \cap C)$, $\circ(A|B \cap C) = \circ(A|B^c \cap C)$, and $\circ(A^c|B \cap C) = \circ(A^c|B^c \cap C)$. Say that Y is *strongly* cs-irrelevant to X given Z if any nonempty event $\{Y = y\}$ is cs-irrelevant to any event $\{X = x\}$ given any nonempty event $\{Z = z\}$ [42, Definition 7.1]. This is a very strong condition as in particular it demands logical independence of Y and Z. A weaker concept of independence has also been proposed by Vantaggi: Y is *weakly* cs-irrelevant to X given Z if $\{Y = y\}$ is cs-irrelevant to $\{X = x\}$ given $\{Z = z\}$ whenever $\{y, z\} \neq \emptyset \neq \{y^c, z\}$ [42, Definition 7.3]. Note that Vantaggi initially refers to both concepts as *stochastic independence* [42], remarking that the first concept leads to a "strong" form of independence when applied to indicator functions; later she uses *conditional cs-independence* for the second concept [44, Definition 3.4]. Cozman and Seidenfeld use *strong coherent irrelevance* for the first concept and *weak coherent irrelevance* for the second [16], but it is perhaps better to keep Vantaggi's name so has to indicate clearly the origin of the concept.

Focusing only on layer numbers, conditional cs-irrelevance of Y to X given Z implies $\circ(x|y, z) = \circ(x|z)$ whenever $\{y, z\} \neq \emptyset$. This is true because, assuming $\{y, z\} \neq \emptyset$, we have that either $\{y^c, z\} = \emptyset$, and in this case $\{y, z\} = \{z\}$ and then $\circ(x|y, z) = \circ(x|z)$ trivially, or else

$$
\begin{aligned}
\circ(x|z) &= \circ(\{x, y, z\} \cup \{x, y^c, z\}) - \circ(z) \\
&= \min\left(\circ(x, y, z), \circ(x, y^c, z)\right) - \circ(z) \\
&= \min\left(\circ(x|y, z) + \circ(y, z), \circ(x|y^c, z) + \circ(y^c, z)\right) - \circ(z) \\
&= \circ(x|y, z) + \min\left(\circ(y, z), \circ(y^c, z)\right) - \circ(z) \\
&= \circ(x|y, z) + \circ(\{y, z\} \cup \{y^c, z\}) - \circ(z) \\
&= \circ(x|y, z).
\end{aligned}
$$

Consequently, conditional cs-irrelevance of Y to X given Z implies

$$
\circ(x, y|z) = \circ(x|z) + \circ(y|z) \qquad \text{whenever } \{Z = z\} \neq \emptyset. \tag{7}
$$

Condition (7) is called the *conditional layer condition* by Cozman and Seidenfeld [16, Corollary 4.11]. Note that this condition is symmetric. One can obtain additional concepts of independence by combining the conditional layer condition with other conditions. For instance, say that X is *fully irrelevant* to Y given Z if X is h-irrelevant to Y given Z and they satisfy the conditional layer condition; say that X and Y are *fully independent* given Z if they are h-independent given Z and they satisfy the conditional layer condition [16]. Full independence is stronger than Kolhlberg and Reny's version of h-independence, because full independence not only implies conditions on the possibility space, but also imposes conditions on layer numbers.

2.8. Graphoid properties

Concepts of independence can be compared with respect to the *graphoid properties* they satisfy. Graphoid properties purport to encode the essence of conditional independence of X and Y given Z, as a ternary relation $(X \perp\!\!\!\perp Y \,|\, Z)$ [18, 38]. In this paper we are interested in the following five properties:

Symmetry: $(X \perp\!\!\!\perp Y \,|\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,|\, Z)$

Redundancy: $(X \perp\!\!\!\perp Y \,|\, X)$

Decomposition: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, Z)$

Weak Union: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp W \,|\, (Y, Z))$

Contraction: $(X \perp\!\!\!\perp Y \,|\, Z)$ & $(X \perp\!\!\!\perp W \,|\, (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$

Often the following property is considered:

Intersection: $(X \perp\!\!\!\perp W \,|\, (Y, Z))$ & $(X \perp\!\!\!\perp Y \,|\, (W, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$

We do not deal with Intersection in this paper, as this property holds for strictly positive probability measures, but fails already for standard probability measures when some events have probability zero [38]. The other five properties, namely Symmetry, Redundancy, Decomposition, Weak Union, and Contraction, are often used to define structures that are called *semi-graphoids* [38]. Whenever we refer to the "semi-graphoid" properties, we mean these five properties.

Epistemic independence satisfies Symmetry, Redundancy, Decomposition and Contraction, but it fails Weak Union, while h-/full independence satisfy Symmetry, Redundancy, Decomposition and Weak Union, but fail Contraction [16]. The full distribution in Table 4 displays both the failure of Weak Union for epistemic independence and the failure of Contraction for h-/full independence.

Concerning kr-independence, it does not seem that its graphoid properties have been analyzed in the literature. We have:

Theorem 1 *Symmetry, Redundancy, Decomposition and Weak Union are satisfied by kr-independence.*

Proof. Symmetry is immediate. To obtain Redundancy, consider a fixed value x of X. Now the range of (X, Y) given $\{X = x\}$ is exactly $\{x, y\}$ for all y in the range of Y with X fixed at x. Given any full distribution for (X, Y), we can construct a sequence $\mathbb{P}_n$ for the full distribution given $\{X = x\}$ by first taking $\mathbb{P}_n(X = x|x) = 1$, and then by multiplying it by each positive probability distribution $\mathbb{P}_n(Y|x)$ in an approximating sequence of the full distribution of Y given $\{X = x\}$. The resulting product distribution (positive over the range of (X, Y) given $\{X = x\}$) approximates the original full distribution given $\{X = x\}$, as

$$\frac{\mathbb{P}(x, y|\{x, y\} \cap \{x, y'\})}{\mathbb{P}(x, y'|\{x, y\} \cap \{x, y'\})} = \frac{\mathbb{P}(y|x \cap \{y \cup y'\})}{\mathbb{P}(y'|x \cap \{y \cup y'\})} = \lim_{n \to \infty} \frac{\mathbb{P}_n(x|x)\mathbb{P}_n(y|x)}{\mathbb{P}_n(x|x)\mathbb{P}_n(y'|x)};$$

hence X and Y are kr-independent given $\{X = x\}$. Now consider Decomposition and Weak Union. Take a sequence that satisfies kr-independence of X and

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\lfloor\alpha\rfloor_0$	$\lfloor\beta\rfloor_2$	$\lfloor 1-\alpha\rfloor_0$	$\lfloor 1-\beta\rfloor_2$
x_1	$\lfloor\alpha\rfloor_1$	$\lfloor\gamma\rfloor_3$	$\lfloor 1-\alpha\rfloor_1$	$\lfloor 1-\gamma\rfloor_3$

Table 4: Full distribution of W, X, Y, with distinct $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$.

(W,Y) given Z; clearly for each element of this sequence Decomposition and Weak Union hold, so X and Y are kr-independent given Z, and also X and W are kr-independent given (Y,Z). Hence Decomposition and Weak Union hold for kr-independence. While this reasoning is immediate for Decomposition, the argument for Weak Union deserves a more detailed description. Consider events $A(X)$, $A'(X)$, $B(W)$ and $B'(W)$ respectively in the algebras generated by X and by W, and denote by C the event $(A(X)\cap B(W))\cup(A'(X)\cap B'(W))$. Then, using the fact that a sequence $\{\mathbb{P}_n\}$ exists by hypothesis,

$$
\begin{aligned}
\frac{\mathbb{P}(A(X)\cap B(W)|C\cap y)}{\mathbb{P}(A'(X)\cap B'(W)|C\cap y)} &= \frac{\mathbb{P}(A(X)\cap B(W)\cap y|C\cap y)}{\mathbb{P}(A'(X)\cap B'(W)\cap y|C\cap y)} \\
&= \rho(A(X)\cap B(W)\cap y; A'(X)\cap B'(W)\cap y) \\
&= \lim_{n\to\infty} \frac{\mathbb{P}_n(A(X)\cap B(W)\cap y)}{\mathbb{P}_n(A'(X)\cap B'(W)\cap y)} \\
&= \lim_{n\to\infty} \frac{\mathbb{P}_n(A(X)\cap B(W)|y)}{\mathbb{P}_n(A'(X)\cap B'(W)|y)} \\
&= \lim_{n\to\infty} \frac{\mathbb{P}_n(A(X)|y)\mathbb{P}_n(B(W)|y)}{\mathbb{P}_n(A'(X)|y)\mathbb{P}_n(B'(W)|y)},
\end{aligned}
$$

as desired. □

Additionally, kr-independence fails Contraction: the full distribution in Table 4 satisfies kr-independence of X and Y, and also kr-independence of X and W given Y, and yet X and (W,Y) are not kr-independent.

3. The structure of epistemic, h-, and full independence

In this section we study the structure of joint full conditional probabilities subject to judgments of independence. To simplify notation we assume in this section that we only have two variables and a full conditional probability $\mathbb{P}$; all results hold if everything were stated given a variable Z, as in that case we would use a full conditional probability $\mathbb{P}_z$ for each possible values of Z.

The basic idea is to order the values of X by their layer numbers, then order the values of Y by their layer numbers, so as to write down the joint full conditional probability as a matrix of product measures. Table 5 depicts this idea, where

$$C_{i,j} = L_i^X \times L_j^Y,$$

	L_0^Y	$\dots$	L_n^Y
L_0^X	$C_{0,0}$	$\dots$	$C_{0,n}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
L_m^X	$C_{m,0}$	$\dots$	$C_{m,n}$

Table 5: Structure of the joint full conditional probability.

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1/3 \rfloor_2$
x_1	$\lfloor 1/6 \rfloor_1$	$\lfloor 1/3 \rfloor_1$
x_2	$\lfloor 1/3 \rfloor_1$	$\lfloor 1/6 \rfloor_1$

Table 6: Joint full distribution of stochastically independent variables X and Y.

and as before L_i^X and L_j^Y denote the layers of the full distributions of X and Y respectively. Throughout this section we denote by m the number of layers of the full distribution for X and by n the number of layers of the full distribution for Y.

Conditional stochastic independence, given by Condition (2), forces the elements of $C_{0,0}$ to have positive probability given by

$$\mathbb{P}(x, y) = \mathbb{P}(x)\,\mathbb{P}(y)\,,$$

but other cells in Table 5 need not resemble product measures in any way. For instance, take the full distribution in Table 6: probabilities conditional on $C_{1,0}$ are not products of marginal probabilities.

Epistemic independence extends the factorization into the first row and first column of Table 5 in the following sense.

Theorem 2 *X and Y are epistemically independent if and only if*

- *for $i \in \{0, \dots, m\}$, for all $x \in L_i^X$ and for all y:*

$$\mathbb{P}\big(x, y | L_i^X\big) = \mathbb{P}\big(x | L_i^X\big)\,\mathbb{P}(y)\,,$$

and

- *for $j \in \{0, \dots, n\}$, for all $y \in L_j^Y$ and for all x:*

$$\mathbb{P}\big(x, y | L_j^Y\big) = \mathbb{P}(x)\,\mathbb{P}\big(y | L_j^Y\big)\,.$$

Moreover, if X and Y are epistemically independent, then

- *for $i \in \{0, \dots, m\}$, for all possible pairs $(x, y) \in C_{i,0}$:*

$$\mathbb{P}(x, y | C_{i,0}) = \mathbb{P}\big(x | L_i^X\big)\,\mathbb{P}(y)\,;$$

- *for $j \in \{0, \ldots, n\}$, for all possible pairs $(x, y) \in C_{0,j}$:*

$$\mathbb{P}(x, y|C_{0,j}) = \mathbb{P}(x)\,\mathbb{P}\big(y|L_j^Y\big)\,.$$

Proof. Suppose X and Y are epistemically independent. For any pair (x, y) such that $x \in L_i^X$, we have:

$$\mathbb{P}\big(x, y|L_i^X\big) = \mathbb{P}\big(y|x \cap L_i^X\big)\,\mathbb{P}\big(x|L_i^X\big) = \mathbb{P}(y|x)\,\mathbb{P}\big(x|L_i^X\big) = \mathbb{P}(y)\,\mathbb{P}\big(x|L_i^X\big)\,.$$

By interchanging X and Y, we obtain $\mathbb{P}\big(x, y|L_j^Y\big) = \mathbb{P}(x)\,\mathbb{P}\big(y|L_j^Y\big)$ for any pair (x, y) such that $y \in L_j^Y$. Now suppose $\mathbb{P}$ satisfies $\mathbb{P}\big(x, y|L_i^X\big) = \mathbb{P}\big(x|L_i^X\big)\,\mathbb{P}(y)$ for $x \in L_i^X$ and $\mathbb{P}\big(x, y|L_j^Y\big) = \mathbb{P}(x)\,\mathbb{P}\big(y|L_j^Y\big)$ for $y \in L_j^Y$. Take $y \in L_j^Y$; then $\mathbb{P}\big(x, y|L_j^Y\big) = \mathbb{P}\big(x|y \cap L_j^Y\big)\,\mathbb{P}\big(y|L_j^Y\big) = \mathbb{P}(x|y)\,\mathbb{P}\big(y|L_j^Y\big)$; hence $\mathbb{P}(x|y)\,\mathbb{P}\big(y|L_j^Y\big) = \mathbb{P}(x)\,\mathbb{P}\big(y|L_j^Y\big)$ and we can cancel out $\mathbb{P}\big(y|L_j^Y\big)$ because it is larger than zero by definition. Consequently, $\mathbb{P}(x|y) = \mathbb{P}(x)$. By interchanging X and Y, we obtain $\mathbb{P}(y|x) = \mathbb{P}(y)$. Thus X and Y are epistemically independent.

Now consider the second part of the theorem. For any pair (x, y) such that $x \in L_i^X$ and $y \in L_0^Y$, use the fact that $C_{i,0} = (L_i^X \times \Omega_Y) \cap (L_0^Y \times \Omega_X)$ and that $1 = \mathbb{P}\big(L_0^Y\big) = \mathbb{P}\big(L_0^Y|L_i^X\big)$ (using Lemma 2.1 by Cozman and Seidenfeld [16]) to obtain:

$$\begin{aligned}
\mathbb{P}(x, y|C_{i,0}) &= \mathbb{P}\big(x, y|(L_i^X \times \Omega_Y) \cap (L_0^Y \times \Omega_X)\big) \\
&= \mathbb{P}\big(x, y|(L_i^X \times \Omega_Y) \cap (L_0^Y \times \Omega_X)\big)\,\mathbb{P}\big(L_0^Y|L_i^X\big) \\
&= \mathbb{P}\big(\{x, y\} \cap L_0^Y|L_i^X\big) \\
&= \mathbb{P}\big(x, y|L_i^X\big) \\
&= \mathbb{P}(y)\,\mathbb{P}\big(x|L_i^X\big)\,.
\end{aligned}$$

By interchanging X and Y, we obtain $\mathbb{P}(x, y|C_{0,j}) = \mathbb{P}(x)\,\mathbb{P}\big(y|L_j^Y\big)$. □

These results can be explained as follows. Define

$$p_i(x) = \mathbb{P}\big(x|L_i^X\big) \qquad \text{and} \qquad q_j(y) = \mathbb{P}\big(y|L_j^Y\big)\,.$$

Then for each cell (i, j) in the first column and in the first row of Table 5, we have a distribution that factorizes as $\mathbb{P}(x, y|C_{i,j}) = p_i(x)q_j(y)$; the other cells need not factorize.

We now move into h-independence. The following result is similar to Theorem 2.1 by Battigalli and Veronesi [6]:

Proposition 1 *If X and Y are h-independent, then for every possible pair $(x, y) \in C_{i,j}$,*

$$\mathbb{P}(x, y|C_{i,j}) = \mathbb{P}\big(x|L_i^X\big)\,\mathbb{P}\big(y|L_j^Y\big)\,. \tag{8}$$

Hence all possible (x, y) in a set $C_{i,j}$ share the same layer number.

Proof. In Expression (3), ignore z and take $A(X) = x$, $B(Y) = y$, $C(X) = L_i^X$, $D(Y) = L_j^Y$. □

$$
\begin{array}{c|c|c|c|}
 & L_0^Y & \cdots & L_n^Y \\
\hline
L_0^X & p_0 q_0 & \cdots & p_0 q_n \\
\hline
\vdots & \vdots & \ddots & \vdots \\
\hline
L_m^X & p_m q_0 & \cdots & p_m q_n \\
\hline
\end{array}
$$

Table 7: Structure of the joint conditional probability: factorization.

Note that it is important to restrict the result to possible pairs (x, y), because the event $\{\omega : (X(\omega), Y(\omega)) \in C_{i,j}\}$ may be empty even when $C_{i,j}$ contains pairs (x, y). In fact, each $C_{i,j}$ is either entirely possible or entirely impossible, given the constraints on layer numbers.

Returning to the matrix in Table 5, we see that now inside cell (i, j) we have factorization $p_i(x)p_j(y)$ given $C_{i,j}$. Table 7 depicts the structure of the joint full conditional probability.

The cells in Table 7 must satisfy some constraints concerning their "depth"; basically, layer numbers grow to the right and to the bottom. Recall that m and n are respectively the maximum layer number for the full distribution of X and of Y. Then:

Proposition 2 *If X and Y are h-independent, then for $i \geq 0$, $j \geq 0$, $k > 0$ such that conditioning events are well-defined and nonempty:*

$$\mathbb{P}(C_{i+k,j}|C_{i,j} \cup C_{i+k,j}) = 0 \quad \textit{and} \quad \mathbb{P}(C_{i,j+k}|C_{i,j} \cup C_{i,j+k}) = 0,$$

and

$$\circ(C_{i+k,j}) > \circ(C_{i,j}), \qquad \circ(C_{i,j+k}) > \circ(C_{i,j})$$

whenever $\circ(C_{i,j})$ is finite. Additionally, $\circ(C_{i,j}) \geq i + j$ for $i \in [0, m]$, $j \in [0, n]$.

Proof. Because all elements of a cell share the same layer number, we need only to focus on two events, $\{x, y\} \in C_{i,j}$ and $\{x', y\} \in C_{i+k,j}$. Note that $\mathbb{P}(x'|x \cup x') = 0$. Using h-independence:

$$\mathbb{P}(x', y|\{x, y\} \cup \{x', y\}) = \mathbb{P}(x', y|\{x \cup x'\} \cap y) = \mathbb{P}(x'|x \cup x)\,\mathbb{P}(y|y) = 0$$

and $\mathbb{P}(C_{i+k,j}|C_{i,j} \cup C_{i+k,j}) = 0$ whenever the conditioning event is nonempty (that is, $\circ(C_{i+k,j}) > \circ(C_{i,j})$ whenever $\circ(C_{i,j})$ is finite). By interchanging X and Y, we obtain $\mathbb{P}(C_{i,j+k}|C_{i,j} \cup C_{i,j+k}) = 0$ whenever the conditioning event is nonempty (that is, $\circ(C_{i,j+k}) > \circ(C_{i,j})$ whenever $\circ(C_{i,j})$ is finite). Finally, to reach a cell $C_{i,j}$ from cell $C_{0,0}$, at least $i + j$ layers are crossed (moving horizontally or vertically in Table 7). $\square$

Example 7. Given two h-independent variables X and Y, the "shallowest" possible joint full conditional probability is the one where cell (i, j) lives in the $(i + j)$th layer; that is, where both $C_{0,1}$ and $C_{1,0}$ are in the same layer, and so on. An example is the full distribution in Table 2. In such a configuration we

	y_0	y_1	y_2
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1/2 \rfloor_1$	$\lfloor 1/2 \rfloor_2$
x_1	$\lfloor 1/2 \rfloor_1$	$\lfloor 1 \rfloor_3$	$\lfloor 1/2 \rfloor_4$
x_2	$\lfloor 1/2 \rfloor_2$	$\lfloor 1/2 \rfloor_4$	$\lfloor 1 \rfloor_5$

Table 8: Joint full distribution of variables X and Y.

have $\circ(x,y) = \circ(x) + \circ(y)$. However this sort of equality may not hold: Table 8 presents a joint full distribution that satisfies h-independence of X and Y and where $\circ(x,y) > \circ(x) + \circ(y)$ for some (x,y).

The following theorem characterizes the structure of h-independence.

Theorem 3 *X and Y are h-independent if and only if*

- *for every nonempty $C_{i,j}$, for every pair $(x,y) \in C_{i,j}$:*

$$\mathbb{P}(x,y|C_{i,j}) = \mathbb{P}\big(x|L_i^X\big)\,\mathbb{P}\big(y|L_j^Y\big)\,,$$

and

- *for all $i \geq 0$, $j \geq 0$, $k > 0$ such that conditioning events are well-defined and nonempty:*

$$\mathbb{P}(C_{i+k,j}|C_{i,j} \cup C_{i+k,j}) = 0 \quad \textit{and} \quad \mathbb{P}(C_{i,j+k}|C_{i,j} \cup C_{i,j+k}) = 0.$$

In this theorem, the “only if ” direction is a combination of previous arguments, while the proof of the “if” direction requires an extended version of Lemma 2.2 by Battigalli and Veronesi [6]. The extension is needed because they assume $\Omega = \Omega_X \times \Omega_Y$ in their work. So, we start with:

Lemma 1 *X is h-irrelevant to Y if and only if*

$$\mathbb{P}(y|x \cap \{y \cup y'\}) = \mathbb{P}(y|x' \cap \{y \cup y'\})$$

whenever $x \cap \{y \cup y'\} \neq \emptyset \neq x' \cap \{y \cup y'\}$.

Proof. The “only if” direction is immediate. For the “if” direction, fix a value of X, say x', and values of Y, say y and y', such that $x' \cap \{y \cup y'\} \neq \emptyset$. Then:

$$\begin{aligned}
\mathbb{P}(y|y \cup y') &= \sum_{x \in \Omega_X} \mathbb{P}(x,y|y \cup y') \\
&= \sum_{x \in \Omega_X : x \cap \{y \cup y'\} \neq \emptyset} \mathbb{P}(x,y|y \cup y') \\
&= \sum_{x \in \Omega_X : x \cap \{y \cup y'\} \neq \emptyset} \mathbb{P}(y|x \cap \{y \cup y'\})\,\mathbb{P}(x|y \cup y')
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x\in\Omega_X : x\cap\{y\cup y'\}\neq\emptyset} \mathbb{P}(y|x'\cap\{y\cup y'\})\,\mathbb{P}(x|y\cup y') \\
&= \mathbb{P}(y|x'\cap\{y\cup y'\}) \sum_{x\in\Omega_X : x\cap\{y\cup y'\}\neq\emptyset} \mathbb{P}(x|y\cup y') \\
&= \mathbb{P}(y|x'\cap\{y\cup y'\})\,,
\end{aligned}
$$

where the condition in the lemma was used in the fourth equality (the other equalities are simply properties of full conditional probabilities). Because a full conditional probability is completely determined by its values on conditioning events given by the union of two atomic events [6, Lemma 2.1c], and because $\mathbb{P}(\cdot|x'\cap\cdot)$ is a full conditional probability, we obtain that the full distribution of Y and the full distribution of Y given x' must be identical except on a set $D'(Y)$ such that $x'\cap D'(Y)=\emptyset$ while $D'(Y)\neq\emptyset$. But $D'(Y)$ must belong to layers of the full distribution of Y that have higher layer numbers than events in $(D'(Y))^c$ (to see that, take any $y\in D'(Y)$ and any possible $y'\neq D'(Y)$; then $\mathbb{P}(y|y\cup y')=\mathbb{P}(y|x'\cap\{y\cup y'\})=0$). We obtain, for any $D(Y)$ such that $x'\cap D(Y)\neq\emptyset$:

$$
\begin{aligned}
\mathbb{P}(y|x'\cap D(Y)) &= \mathbb{P}(y|x'\cap D(Y)\cap(D'(Y))^c) \\
&= \mathbb{P}(y|D(Y)\cap(D'(Y))^c) \\
&= \mathbb{P}(y|D(Y))\,.
\end{aligned}
$$

And using Lemma 2.1 by Cozman and Seidenfeld [16], we obtain the equality $\mathbb{P}(B(Y)|A(X)\cap D(Y))=\mathbb{P}(B(Y)|D(Y))$ whenever $A(X)\cap D(Y)\neq\emptyset$; thus X is h-irrelevant to Y. □

We can now present the proof of Theorem 3.

Proof. As noted, the "only if" direction is basically a combination of Propositions 1 and 2. To prove the "if" direction, note that Lemma 1 implies: X and Y are h-independent if for any two distinct values x and x' and any two distinct values y and y',

$$
\mathbb{P}(x,y|\{x,y\}\cup\{x,y'\}) = \mathbb{P}(x',y|\{x',y\}\cup\{x',y'\}) \quad \text{whenever } \{x,y\}\cup\{x,y'\}\neq\emptyset\neq\{x',y\}\cup\{x',y'\}, \tag{9}
$$

$$
\mathbb{P}(x,y|\{x,y\}\cup\{x',y\}) = \mathbb{P}(x,y'|\{x,y'\}\cup\{x',y'\}) \quad \text{whenever } \{x,y\}\cup\{x',y\}\neq\emptyset\neq\{x,y'\}\cup\{x',y'\}. \tag{10}
$$

Note also that if we have two points (x,y) and (x',y') that belong to the same nonempty cell $C_{i,j}$,

$$
\begin{aligned}
\mathbb{P}(x,y|C_{i,j}) &= \mathbb{P}(x,y|\{\{x,y\}\cup\{x',y'\}\}\cap C_{i,j})\,\mathbb{P}(\{x,y\}\cup\{x',y'\}|C_{i,j}) \\
&= \mathbb{P}(x,y|\{x,y\}\cup\{x',y'\})\,(\mathbb{P}(x,y|C_{i,j})+\mathbb{P}(x',y'|C_{i,j}))\,,
\end{aligned}
$$

and because $\mathbb{P}(x, y|C_{i,j}) = p_i(x)q_j(y) > 0$ and $\mathbb{P}(x', y'|C_{i,j}) = p_i(x')q_j(y') > 0$,

$$\mathbb{P}(x, y|\{x, y\} \cup \{x', y'\}) = \frac{p_i(x)q_j(y)}{p_i(x)q_j(y) + p_i(x')q_j(y')}.$$

Now consider four points (x, y), (x, y'), (x', y), (x', y'). Given the second condition in the theorem, there are only four possible situations.

Case 1: The four points belong to the same cell $C_{i,j}$, and this cell is nonempty (if the cell is empty, there is nothing to verify). Expression (9) yields

$$\frac{p_i(x)q_j(y)}{p_i(x)q_j(y) + p_i(x)q_j(y')} = \frac{p_i(x')q_j(y)}{p_i(x')q_j(y) + p_i(x')q_j(y')};$$

that is (by cancelling terms),

$$\frac{q_j(y)}{q_j(y) + q_j(y')} = \frac{q_j(y)}{q_j(y) + q_j(y')},$$

a tautology. Expression (10) is likewise satisfied.

Case 2: Points (x, y) and (x', y) belong to the same cell $C_{i,j}$, while points (x, y') and (x', y') belong to cell $C_{i,j+k}$ for some $k > 0$. If both cells are empty, there is nothing to verify. Suppose instead that $C_{i,j} \neq \emptyset$. Constraints on layers yield

$$\mathbb{P}(x, y|\{x, y\} \cup \{x, y'\}) = \mathbb{P}(x', y|\{x', y\} \cup \{x', y'\}) = 1.$$

If $C_{i,j+k} = \emptyset$, there is nothing to verify concerning Expression (10); otherwise, Expression (10) yields

$$\frac{p_i(x)q_j(y)}{p_i(x)q_j(y) + p_i(x')q_j(y)} = \frac{p_i(x)q_j(y')}{p_i(x)q_j(y') + p_i(x')q_j(y')},$$

a tautology.

Case 3: Points (x, y) and (x, y') belong to the same cell $C_{i,j}$, while points (x', y) and (x', y') belong to cell $C_{i+k,j}$ for some $k > 0$. If both cells are empty, there is nothing to verify. Suppose instead that $C_{i,j} \neq \emptyset$. If $C_{i+k,j} = \emptyset$, there is nothing to verify concerning Expression (9); otherwise, Expression (9) yields

$$\frac{p_i(x)q_j(y)}{p_i(x)q_j(y) + p_i(x)q_j(y')} = \frac{p_i(x')q_j(y)}{p_i(x')q_j(y) + p_i(x')q_j(y')},$$

a tautology. Constraints on layers yield

$$\mathbb{P}(x, y|\{x, y\} \cup \{x', y\}) = \mathbb{P}(x, y'|\{x, y'\} \cup \{x', y'\}) = 1.$$

Case 4: All points belong to different cells. Suppose first that the four cells are nonempty, with (x, y) of lowest layer number, (x', y') of highest layer number, and (x, y') and (x', y') of intermediate layer numbers. Then constraints on layers yield

$$\begin{aligned} \mathbb{P}(x, y|\{x, y\} \cup \{x, y'\}) &= \mathbb{P}(x', y|\{x', y\} \cup \{x', y'\}) = 1, \\ \mathbb{P}(x, y|\{x, y\} \cup \{x', y\}) &= \mathbb{P}(x, y'|\{x, y'\} \cup \{x', y'\}) = 1. \end{aligned}$$

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1/2 \rfloor_1$
x_1	$\lfloor 1 \rfloor_2$	$\lfloor 1/2 \rfloor_1$

Table 9: Joint full distribution that satisfies layer factorization.

Now suppose $\{x, y'\}$ is empty; hence $\{x', y'\}$ is empty as well. The first equality holds while the second is irrelevant. Likewise, if $\{x', y\}$ is empty, then $\{x', y'\}$ is empty as well; the second equality holds while the first one is irrelevant. Finally, if both $\{x', y\}$ and $\{x, y'\}$ are empty, then $\{x', y'\}$ is empty as well, and there is nothing to verify. $\Box$

Corollary 1 *X and Y are fully independent if and only if for every $C_{i,j}$, for every pair $(x, y) \in C_{i,j}$,*

$$\mathbb{P}(x, y|C_{i,j}) = \mathbb{P}\big(x|L_i^X\big)\,\mathbb{P}\big(y|L_j^Y\big) \qquad \textit{and} \qquad \circ(x, y) = \circ(x) + \circ(y)\,.$$

As noted by Kohlberg and Reny [29], kr-independence implies h-independence when variables are logically independent. Hence the structure of joint full conditional probabilities under kr-independence must be given by product measures as in Table 7. However, kr-independence imposes considerable stronger conditions; Kohlberg and Reny [29] tie kr-independence to an exchangeability condition, while Swinkels [41] tie kr-independence to an "extendibility" condition. These conditions are rather complex; we have not been able to derive any further insight on kr-independence, and leave its structure to future work. We however come back, in Section 5, to insights behind approximating sequences in order to build a theory of Bayesian networks for full conditional probabilities.

4. Factorization by layer

H-independence and full independence are quite attractive, but still they do not satisfy the Contraction property. In this section we examine a different route to concepts of independence, one that will take us to a concept of independence satisfying the semi-graphoid properties.

Say that X and Y satisfy *layer factorization* given Z when, for each layer L_i of the underlying full conditional probability $\mathbb{P}$,

$$\mathbb{P}(x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\,\mathbb{P}(y|z \cap L_i) \qquad \text{whenever } z \cap L_i \neq \emptyset. \tag{11}$$

By itself, the layer factorization condition is quite weak. For instance, the joint full distribution in Table 9 satisfies layer factorization, even though $\mathbb{P}(x_0|y_0) = 1 \neq 1/2 = \mathbb{P}(x_0|y_1)$ (that is, even epistemic independence fails).

We might then combine layer factorization with other conditions, to obtain stronger concepts of independence that satisfy desirable properties. As an exercise, suppose for instance that two variables X and Y are h-independent

	y_0	y_1	y_2
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_2$
x_1	$\lfloor 1 \rfloor_3$	$\lfloor 1 \rfloor_5$	$\lfloor 1 \rfloor_6$
x_2	$\lfloor 1 \rfloor_4$	$\lfloor 1 \rfloor_7$	$\lfloor 1 \rfloor_8$

	y_0	y_1	y_2
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_3$	$\lfloor 1 \rfloor_6$
x_1	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_4$	$\lfloor 1 \rfloor_7$
x_2	$\lfloor 1 \rfloor_2$	$\lfloor 1 \rfloor_5$	$\lfloor 1 \rfloor_8$

Table 10: Joint full distributions of variables X and Y that satisfy layer factorization and are h-independent; in the right table, X and Y are even fully independent.

and satisfy layer factorization. Table 10 shows two full distributions that satisfy both conditions. Alas, the combination of h-/fully independence and layer factorization does not yield the Contraction property. Indeed, for the full distribution in Table 4 we have that X and Y are h-/fully independent and satisfy layer factorization, X and W are h-/fully independent given Y and satisfy layer factorization given Y, and yet X and (W, Y) are not h-independent.

Nonetheless, we can use layer factorization to produce an interesting new concept:

Definition 1. X and Y are *layer independent* given Z if, for each layer L_i of the underlying full conditional probability $\mathbb{P}$,

- $\mathbb{P}(x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\,\mathbb{P}(y|z \cap L_i)$ whenever $z \cap L_i \neq \emptyset$,

and

- $\circ(x, y|z) = \circ(x|z) + \circ(y|z)$ whenever $\{Z = z\} \neq \emptyset$.

For a fixed $z \cap L_i \neq \emptyset$, consider the sets

$$A_i(X) = \{x : x \cap z \cap L_i \neq \emptyset\} \quad \text{and} \quad B_i(X) = \{y : y \cap z \cap L_i \neq \emptyset\}.$$

Then $\mathbb{P}(x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\,\mathbb{P}(y|z \cap L_i) > 0$ for every $(x, y, z) \in A_i(X) \times B_i(Y) \times \{Z = z\}$, while for every other (x, y, z) we have $\mathbb{P}(x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\,\mathbb{P}(y|z \cap L_i) = 0$. Hence $z \cap L_i = A_i(X) \times B_i(Y) \times \{Z = z\}$; in other words, every set $z \cap L_i$ is a rectangle.

Moreover, we obtain the semi-graphoid properties:

Theorem 4 *Layer independence satisfies Symmetry, Redundancy, Decomposition, Weak Union and Contraction.*

Proof. Symmetry is immediate.

For Redundancy: Whenever $x \cap L_i \neq \emptyset$,

$$\mathbb{P}(x, y|x \cap L_i) = \mathbb{P}(y|x \cap x \cap L_i)\,\mathbb{P}(x|x \cap L_i) = \mathbb{P}(x|x \cap L_i)\,\mathbb{P}(y|x \cap L_i)\,.$$

Also, $\circ(x, y|x) = \circ(x, y, x) - \circ(x) = \circ(x, y) - \circ(x) = \circ(y|x) = \circ(x|x) + \circ(y|x)$.

For Decomposition: We have $\mathbb{P}(w, x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(w, y|z \cap L_i)$ for $z \cap L_i \neq \emptyset$; then

$$\begin{aligned}
\mathbb{P}(x, y|z \cap L_i) &= \sum_w \mathbb{P}(w, x, y|z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i) \sum_w \mathbb{P}(w, y|z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(y|z \cap L_i)\,.
\end{aligned}$$

Also, we start with $\circ(w, x, y, z) + \circ(z) = \circ(w, y, z) + \circ(x, z)$; then $\circ(x, y, z) + \circ(z) = \min_w \circ(w, x, y, z) + \circ(z) = \min_w \circ(w, y, z) + \circ(x, z) = \circ(y, z) + \circ(x, z)$ as desired.

For Weak Union: If $y \cap z \cap L_i \neq \emptyset$,

$$\begin{aligned}
\mathbb{P}(w, x|y \cap z \cap L_i)\, \mathbb{P}(y|z \cap L_i) &= \mathbb{P}(w, x, y|z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(w, y|z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i)\, \mathbb{P}(y|z \cap L_i) \\
&= \mathbb{P}(x, y|z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i) \\
&= \mathbb{P}(x|y \cap z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i)\, \mathbb{P}(y|z \cap L_i)\,,
\end{aligned}$$

while the second equality comes from the layer independence of X and (W, Y), the fourth equality comes from the layer independence of X and Y (using Decomposition), and the other equalities are properties of full conditional probabilities. By dividing both sides by $\mathbb{P}(y|z \cap L_i)$ (this is possible because, as $y \cap z \cap L_i \neq \emptyset$, we have $\mathbb{P}(y|z \cap L_i) > 0$, so we can divide both sides by this quantity to obtain $\mathbb{P}(w, x|y \cap z \cap L_i) = \mathbb{P}(x|y \cap z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i)$ as desired. Also, we have $\circ(w, x, y, z) + \circ(z) = \circ(w, y, z) + \circ(x, z)$ and by Decomposition we have $\circ(x, z) + \circ(y, z) = \circ(x, y, z) + \circ(z)$; by adding both sides, $\circ(w, x, y, z) + \circ(y, z) = \circ(w, y, z) + \circ(x, y, z)$ as desired.

For Contraction: We have $\mathbb{P}(x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(y|z \cap L_i)$ for $z \cap L_i \neq \emptyset$ and $\mathbb{P}(w, x|y \cap z \cap L_i) = \mathbb{P}(x|y \cap z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i)$ for $y \cap z \cap L_i \neq \emptyset$. Suppose $z \cap L_i \neq \emptyset$: if $y \cap z \cap L_i = \emptyset$, then $\mathbb{P}(w, x, y|z \cap L_i) = \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(w, y|z \cap L_i) = 0$; if instead $y \cap z \cap L_i \neq \emptyset$, then

$$\begin{aligned}
\mathbb{P}(w, x, y|z \cap L_i) &= \mathbb{P}(w, x|y \cap z \cap L_i)\, \mathbb{P}(y|z \cap L_i) \\
&= \mathbb{P}(x|y \cap z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i)\, \mathbb{P}(y|z \cap L_i) \\
&= \mathbb{P}(x, y|z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(y|z \cap L_i)\, \mathbb{P}(w|y \cap z \cap L_i) \\
&= \mathbb{P}(x|z \cap L_i)\, \mathbb{P}(w, y|z \cap L_i)\,,
\end{aligned}$$

as desired. Also, we have $\circ(w, x, y, z) + \circ(y, z) = \circ(w, y, z) + \circ(x, y, z)$ and $\circ(x, y, z) + \circ(z) = \circ(x, z) + \circ(y, z)$; by adding both sides, $\circ(w, x, y, z) + \circ(z) = \circ(w, y, z) + \circ(x, z)$ as desired. $\square$

Note that this result is obtained because we keep track of the layers of the underlying full conditional probability, not just layers of the marginal and conditional pieces that appear in graphoid properties. It is the cost of keeping track of these layers that pays for the semi-graphoid properties. Similarly, all layer numbers are computed with respect to the underlying full conditional probability; hence the whole idea requires considerable bookkepping when many variables are interacting.[1]

5. Building joint full conditional probabilities: non-uniqueness and Bayesian networks

Independence relations are often used to build joint probability distributions out of marginal and conditional distributions. One ubiquitous example is the construction of a sequence of independent identically distributed variables so as to prove concentration inequalities. Another example is the combination of marginal and conditional probabilities in Bayesian networks and Markov random fields [38]. In this section we examine to what extend this modeling strategy can be used to establish a theory of Bayesian networks with full conditional probabilities.

5.1. Challenges in building a joint full probability with a Bayesian network

In the theory of Bayesian networks, directed acyclic graphs are employed to organize marginal and conditional distributions into a single standard joint distribution [38]. Alas, as the examples in Appendix A show, the standard theory of Bayesian networks does not apply when concepts of independence fail semi-graphoid properties. This suggests that a concept such as layer independence, that satisfies all semi-graphoid properties, should be important in specifying full conditional probabilities through Bayesian networks. Concepts such as *enhanced basis* and *d-separation* could then be defined without difficulty [22]. (Of course, a different path would be to build alternatives to Bayesian networks that do not require all semi-graphoid properties [4, 43, 44].)

However, failure of graphoid properties is not the only challenge when one tries to build a joint full conditional probability out of conditional and marginal pieces. Another challenge is the non-uniqueness of joint full conditional probabilities.

We start by examining epistemic independence, the weakest concept of independence that makes sense for full conditional probabilities. Consider the full distribution in Table 2. The marginal full distribution of X is given by $\mathbb{P}(x_0) = \mathbb{P}(x_1|x_0^c) = 1$; likewise, the marginal full distribution of Y is given by $\mathbb{P}(y_0) = \mathbb{P}(y_1|y_0^c) = 1$. The marginal distributions do not provide any information about α; indeed, any value of α produces identical marginals. In addition, both full distributions in Table 11 produce the same marginal distributions.

[1] Matthias Troffaes has suggested a different concept (unpublished) that uses a condition similar to $\mathbb{P}(x, y, z|L_{x,y,z})\,\mathbb{P}(z|L_z) = \mathbb{P}(x, z|L_{x,z})\,\mathbb{P}(y, z|L_{y,z})$. This is an interesting alternative path, where each probability value is associated with a particular layer.

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_2$
x_1	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_3$

	y_0	y_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_1$
x_1	$\lfloor 1 \rfloor_2$	$\lfloor 1 \rfloor_3$

Table 11: Joint full distributions of binary variables X and Y.

	y_0	y_1
x_0	$\lfloor \alpha \rfloor_0$	$\lfloor 1-\alpha \rfloor_0$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor 1-\alpha \rfloor_1$

Table 12: Joint full distribution of (X, Y) from Table 4.

The full distributions in Tables 2 and 11 have identical marginals, and they satisfy h-/full independence of X and Y; hence non-uniqueness of joint distributions can happen for h-/full independence (non-uniqueness is already discussed by Battigalli [5]).

We can further understand the difficulties with non-uniqueness for h-/full independence by considering how they fail the Contraction property. Consider again Table 4 and the marginalized full distribution of (X, Y) in Table 12. The problem here is that the full distribution for (X, Y) does not contain any information about β and γ, but these values become crucial once we condition on w_1. The marginal full distribution of (X, Y) "hides" β and γ because the probabilities in deeper layers disappear when we marginalize over W. In a sense, the deeper layers are "covered" by the shallower layers. That is, the joint full distribution contains more information than its marginal pieces.

Now note that both full distributions in Table 11 satisfy layer independence of X and Y, so non-uniqueness can happen for this concept of independence as well.

Uniqueness also fails with kr-independence (as already noted by Kohlberg and Reny [29]). Both full distributions in Tables 2 and 11 display kr-independence of X and Y with identical marginals.

We might wonder whether non-uniqueness crops up even in the absence of any judgment of independence. For instance, suppose we have variables X and Y, and we obtain $\mathbb{P}(x|y)$ and $\mathbb{P}(y)$ for all possible (x, y). Alas, we cannot necessarily build a single joint distribution of (X, Y) out of these assessments:

Example 8. Consider two variables X and Y respectively with three and two values, and suppose we have the following assessments:

$$\mathbb{P}(y_0) = \mathbb{P}(y_1|y_1) = 1,$$

$$\mathbb{P}(x_0|y_0) = \mathbb{P}(x_1|y_0)/2 = 1/3, \quad \mathbb{P}(x_0|y_1) = \mathbb{P}(x_1|y_1) = 1/2.$$

The joint full distributions in Table 13 satisfy these assessments, for any $\alpha \in (0, 1)$. $\square$

	y_0	y_1
x_0	$\lfloor 1/3 \rfloor_0$	$\lfloor (1-\alpha)/2 \rfloor_1$
x_1	$\lfloor 2/3 \rfloor_0$	$\lfloor (1-\alpha)/2 \rfloor_1$
x_2	$\lfloor \alpha \rfloor_1$	$\lfloor 1 \rfloor_2$

	y_0	y_1
x_0	$\lfloor 1/3 \rfloor_0$	$\lfloor 1/2 \rfloor_2$
x_1	$\lfloor 2/3 \rfloor_0$	$\lfloor 1/2 \rfloor_2$
x_2	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_3$

	y_0	y_1
x_0	$\lfloor 1/3 \rfloor_0$	$\lfloor 1/2 \rfloor_1$
x_1	$\lfloor 2/3 \rfloor_0$	$\lfloor 1/2 \rfloor_1$
x_2	$\lfloor 1 \rfloor_2$	$\lfloor 1 \rfloor_3$

Table 13: Joint full distributions discussed in Example 8, with $\alpha \in (0, 1)$.

Hence we cannot expect to generate unique full distributions out of a Bayesian network whose assessments are interpreted as a collection of full conditional probabilities, unless more information is input into the network concerning the relative layer numbers of various events. One possibility is to view a Bayesian network as a representation for a set of full conditional probabilities [10, 46]. But here we wish to consider the specification of a single full conditional probability over a set of variables, out of marginal and conditional pieces; we defer the direct treatment of sets of full conditional probabilities to the future. So, how can we specify a single full conditional probability within the framework of Bayesian networks?

We might, for instance, adopt layer independence, and ask the user to specify a standard Bayesian network per layer of the joint full conditional probability. Another, more direct, and much more attractive, idea is to introduce more information explicitly into Bayesian networks, as discussed in the next subsection.

5.2. Specifying an approximating sequence with a single extended Bayesian network

Our proposal is that, to specify a joint full conditional probability, one must specify an approximating sequence through a single suitably extended Bayesian network. To understand the proposal, suppose we have a set of variables and we start building a standard Bayesian network for them. We proceed as usual, by assigning variables to nodes and by placing edges between nodes, so as to build a directed acyclic graph. We must then specify probability values. In a standard Bayesian network, every probability value is given as a real number that may be zero. In our extended Bayesian network we do not allow a probability value to be zero; instead, all probability values must be given as strictly positive ratios of polynomials in $\epsilon > 0$. This ϵ-parametrized Bayesian network encodes an approximating sequence that is obtained by taking ϵ to zero. The resulting full distribution is the semantics of the extended Bayesian network.

The following example illustrates the idea.

Example 9. Consider a Bayesian network with two binary variables X and Y and no arrow between them (hence X and Y are independent). If all probability

	y_0	y_1
x_0	$\frac{1}{1+(\alpha/(1-\alpha))\epsilon}\frac{1}{1+\epsilon}$	$\frac{1}{1+(\alpha/(1-\alpha))\epsilon}\frac{\epsilon}{1+\epsilon}$
x_1	$\frac{(\alpha/(1-\alpha))\epsilon}{1+(\alpha/(1-\alpha))\epsilon}\frac{1}{1+\epsilon}$	$\frac{(\alpha/(1-\alpha))\epsilon}{1+(\alpha/(1-\alpha))\epsilon}\frac{\epsilon}{1+\epsilon}$

$\xrightarrow{\epsilon\to 0}$

	y_0	y_1
x_0	$\lfloor 1\rfloor_0$	$\lfloor 1-\alpha\rfloor_1$
x_1	$\lfloor \alpha\rfloor_1$	$\lfloor 1\rfloor_2$

Table 14: Extended distribution of binary variables X and Y, and resulting full conditional probability as ϵ goes to zero.

values were zero, we would have no difficulty specifying a single probability measure displaying this independence relation. However, suppose both $\mathbb{P}(x_1)$ and $\mathbb{P}(y_1)$ are equal to zero. As we have noted already, a naive specification of marginal probabilities is not sufficient to fix the complete joint full conditional probability. However, suppose we have the following assessments:

$$\mathbb{P}(x_0) \propto 1, \quad \mathbb{P}(x_1) \propto \frac{\alpha}{1-\alpha}\epsilon, \quad \mathbb{P}(y_0) \propto 1, \quad \mathbb{P}(y_1) \propto \epsilon.$$

Note that these assessments are only proportional to the probabilities, as the obvious normalizing constants are easy to compute. The joint distribution is given by Table 14; note that by taking the limit as ϵ goes to zero, we obtain the full distribution in Table 2. □

As this example shows, the layer L_i of the joint full distribution consists of those polynomial coefficients associated with ϵ^i. By being explicit about ϵ, one can specify precisely the relative probabilities of cells $C_{i,j}$ and $C_{i',j'}$.

The following simple elicitation method builds approximating sequences for joint full conditional probabilities using extended Bayesian networks. First, build a directed acyclic graph where nodes are variables and edges denote dependence, as in a standard Bayesian network. Now consider a variable X and a configuration y of its parents. Specify each layer of the full distribution of X given y; say that layer L_i is associated with the positive probability measure $p_i(X|y)$. Then write, for any x,

$$\mathbb{P}(x|y) = \frac{\sum_{i=0}^{K} \beta_i \epsilon^i p_i(x|y)}{\sum_{i=0}^{K} \beta_i \epsilon^i}.$$

The resulting network represents a single joint full conditional probability that is obtained by taking ϵ to zero. The specification of numbers β_i guarantees that the relative probabilities of cells are given. For instance, in Example 9, the full conditional probability of X was encoded using $\beta_1 = \alpha/(1-\alpha)$.

The stochastic independence relations in the approximating probability distributions are inherited as kr-independence relations in the resulting full conditional probability. Hence d-separation in the graph of the Bayesian network implies kr-independence in the resulting joint full distribution.

The simplest way to interpret ϵ, and to determine the rules to handle it, is to take it to be an infinitesimal, and to consider the specification of a Bayesian network to happen in the hyperreal line $\Re(\epsilon)$; that is, the real-numbers plus an infinitesimal ϵ. Hammond has forcefully argued for such a representation of full conditional probabilities [26, 27].

6. Hyperreal full conditional probabilities

Once we move to hyperreals as a specification device, one possibility is to adopt hyperreals as the basic machinery underlying probabilities. The use of hyperreals in probability theory has been intensely explored [2, 21, 37, 48], sometimes explicitly in connection with full conditional probabilities [25, 27]. Of course, one can take hyperreal "unconditional" probabilities as the primitive notion [32, 37, 47], assume then to be always positive, and then define conditional probability simply as a derived concept: $\mathbb{P}(A|B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$.

Instead, it seems a better idea to adopt conditional probability as the main primitive concept, even if hyperreal probabilities are always positive. That is, take that $\mathbb{P}$ is a two-place set-function from $\mathcal{B} \times (\mathcal{B} \backslash \emptyset)$ into the hyperreals, where $\mathcal{B}$ is a Boolean algebra over a set Ω, such that for any event $C \neq \emptyset$:
(1) $\mathbb{P}(C|C) = 1$;
(2) $\mathbb{P}(A|C) > 0$ whenever $A \cap C \neq \emptyset$;
(3) $\mathbb{P}\big(\cup_{i=1}^{N} A_i|C\big) = \sum_{i=1}^{N} \mathbb{P}(A_i|C)$ for disjoint A_i;
(4) $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|B \cup C)\,\mathbb{P}(B|C)$ when $B \cap C \neq \emptyset$.

We can then define independence of X and Y given Z as

$$\mathbb{P}(x|y, z) = \mathbb{P}(x|z) \qquad \text{for every nonempty } \{y, z\}. \tag{12}$$

Then, with the usual proof of semi-graphoid properties [38], we obtain:

Theorem 5 *For an hyperreal conditional probability that satisfies the axioms in this section, independence as defined by Expression (12) satisfies Symmetry, Redundancy, Decomposition, Weak Union, and Contraction.*

7. Conclusion

We have studied concepts of independence for full conditional probabilities, and the construction of joint full distributions from marginal and conditional ones using judgments of independence. We have derived the structure of joint full conditional probabilities under epistemic/h-/full independence, and examined the semi-graphoid properties of these (and other) concepts of independence. We have introduced the condition of layer factorization; the derived concept of layer independence is particularly interesting because it satisfies all semi-graphoid properties.

We have also examined non-uniqueness of full joint conditional probabilities under various concepts of independence. We suggested an specification strategy that adapts the theory of Bayesian networks to full conditional probabilities, by parameterizing probability values with an infinitesimal ϵ. We closed by commenting on a theory of hyperreal full conditional probabilities.

Our proposal concerning modeling tools, such as Bayesian networks, can be summarized as follows. Whenever a modeling tool, originally built for standard probability measures, is to be used to specify full conditional probabilities, the most effective way to do so is to extend the tool into the hyperreal line, so that specification of probability values only deals with positive values. Instead of trying to change completely the semantics of modeling tools so as to cope with failure of graphoid properties and of uniqueness, it is better to view these modeling tools as devices that specify approximating sequences. Full conditional probabilities are then obtained in the limit, and there are no concerns about non-uniqueness.

Acknowledgements

Thanks to Teddy Seidenfeld and Matthias Troffaes for help with full conditional probabilities, and to Anderson de Araújo for help with hyperreal numbers.

Thanks to the reviewers for excellent reviews; in particular for clarifying Coletti and Scozzafava's concept of independence, and for suggesting a simple version of axioms for full conditional probabilities.

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Project 2008/03995-5 (LOGPROB). The author was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grant PQ 305395/2010-6.

Appendix A. Failure of semi-graphoid properties and Bayesian networks

Failure of semi-graphoid properties does cause damage to the theory of Bayesian networks, assuming that a theory as developed by Geiger et al. [22] is desired. Note that Geiger et al. define *deterministic nodes* in terms of conditional independence; for full conditional probabilities one must instead define *deterministic nodes* directly, as nodes that are functions of their parents.

The next example shows the difficulties caused by failure of Weak Union for epistemic independence.

Example 10. Consider four binary variables ordered as Z, Y, X and W, and the following pair of judgements of epistemic independence: $(X \text{ EIN } Y \mid Z)$ and $(W \text{ EIN } (X, Y) \mid Z)$, where EIN stands for epistemic independence. These variables and judgements form an enhanced basis as defined by Geiger et al. [22]. The network induced by this enhanced basis has root Z with three children (the other variables), and no other edges. Clearly X and Y are d-separated by (W, Z). However X and Y may not be epistemically independent given (W, Z): suppose $\mathbb{P}(z_0) = \mathbb{P}(z_1)$, take $\mathbb{P}_{\{Z=z_1\}}$ to be uniform, and $\mathbb{P}_{\{Z=z_0\}}$ to be given by Table 4.

Suppose, conversely, that one receives a directed acyclic graph with three nodes, W, X, and Y, where Y is the sole parent of W, and where X is disconnected from W and Y. The Markov condition on this graph requires: X

independent of (W, Y), Y independent of X, and W independent of X conditional on Y. These epistemic independence relations are all satisfied by the full conditional probability in Table 4, but the d-separation of X and Y given W does not imply epistemic independence of X and Y conditional on W. $\Box$

The next example shows the difficulties caused by failure of Contraction for h-/full independence.

Example 11. Consider four binary variables ordered as Z, Y, X and W, and the following enhanced basis: $(X \,\textsf{FIN}\, Y \mid Z)$ and $(W \,\textsf{FIN}\, X \mid (Y, Z))$, where $\textsf{FIN}$ stands for full independence. The resulting network has a root Z with three children (the other variables); there is only one other edge from Y to W. Clearly X and (W, Y) are d-separated by Z; however X and (W, Y) may not be h-independent given Z: just take the same full conditional probability constructed in the first paragraph of Example 10.

Suppose conversely that one receives a directed acyclic graph with four nodes, where X is the only root, the only parent of Z is X, the only parent of Y is Z, and W has both Y and Z as parents. The Markov condition on this graph requires: W and X are independent conditional on (Y, Z); Y and X are independent conditional on Z. Again, the full conditional probability constructed in the first paragraph of Example 10 satisfies these judgements of full independence, but the d-separation of X and (W, Y) given Z does not imply full independence of X and (W, Y) conditional on Z. $\Box$

References

[1] Adams, E.W.: A Primer of Probability Logic. CSLI Publications, Stanford, CA (2002)

[2] Albeverio, S., Fenstad, J.E., Hoegh-Krohn, R., Lindstrom, T.: Nonstandard Methods in Stochastic Analysis and Mathematical Physics. Academic Press Inc. (1986)

[3] Armstrong, T. E.: Countably additive full conditional probabilities. Proceedings of the American Mathematical Society **107**(4), 977–987 (1989)

[4] Baioletti, M., Busanello, G., Vantaggi, B.: Conditional independence structure and its closure: inferential rules and algorithms. International Journal of Approximate Reasoning **50**, 1097–1114 (2009)

[5] Battigalli, P.: Strategic independence and perfect Bayesian equilibria. Journal of Economic Theory **70**, 201–234 (1996)

[6] Battigalli, P., Veronesi, P.: A note on stochastic independence without Savage-null events. Journal of Economic Theory **70**(1), 235–248 (1996)

[7] Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and choice under uncertainty. Econometrica **58**(1), 61–79 (1991)

[8] Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and equilibrium refinements. Econometrica **58**(1), 81–98 (1991)

[9] Brozzi, A., Capotorti, A., Vantaggi, B.: Incoherence correction strategies in statistical matching. International Journal of Approximate Reasoning **53**, 1124–1136 (2012)

[10] Capotorti, A., Coletti, G., Vantaggi, B.: Preferences representable by a lower expectation: Some characterizations. Theory and Decision **64**, 119–146 (2008)

[11] Capotorti, A., Regoli, G., Vattari, F.: Correction of incoherent conditional probability assessments. International Journal of Approximate Reasoning **51**(6), 718–727 (2010)

[12] Coletti, G., Scozzafava, R.: Zero probabilities in stochastic independence. Information, Uncertainty and Fusion, pp. 185–196 (2000)

[13] Coletti, G., Scozzafava, R., Vantaggi, B.: Probabilistic reasoning as a general unifying tool. In S. Berferhat, P. Besnard (eds.) European Conference on Symbolic and Qualitative Approaches to Reasoning with Uncertainty (ECSQARU), pp. 120–131 (2001)

[14] Coletti, G., Scozzafava, R.: Stochastic independence in a coherent setting. Annals of Mathematics and Artificial Intelligence **35**, 151–176 (2002)

[15] Coletti, G., Scozzafava, R.: Probabilistic Logic in a Coherent Setting. Trends in logic, 15. Kluwer, Dordrecht (2002)

[16] Cozman, F.G., Seidenfeld, T.: Independence for full conditional measures and their graphoid properties. In: B. Lowe, E. Pacuit, J.W. Romeijn (eds.) Reasoning about Probabilities and Probabilistic Reasoning, *Foundations of the Formal Sciences VI*, vol. 16, pp. 1–29. College Publications, London (2009)

[17] Császár, A.: Sur la structure des espaces de probabilité conditionnelle. *Acta Mathematica Academiae Scientiarum Hungarica*, **6**(3-4), 337–361 (1955)

[18] Dawid, A.P.: Conditional independence. In: S. Kotz, C.B. Read, D.L. Banks (eds.) Encyclopedia of Statistical Sciences, Update Volume 2, pp. 146–153. Wiley, New York (1999)

[19] de Finetti, B.: Theory of Probability, vol. 1-2. Wiley, New York, 1974.

[20] Dubins, L.E.: Finitely additive conditional probability, conglomerability and disintegrations. Annals of Statistics **3**(1), 89–99 (1975)

[21] Fajardo, S., Keisler, H.J.: Model Theory of Stochastic Processes. A.K. Peters (2002)

[22] Geiger, D., Verma, T., Pearl, J.: Identifying independence in Bayesian networks. Networks **20**, 507–534 (1990)

[23] Gilio, A.: Generalizing inference rules in a coherence-based probabilistic default reasoning. International Journal of Approximate Reasoning **53**(3), 413–434 (2012)

[24] Hajek, A.: What conditional probability could not be. Synthese **137**, 273–323 (2003)

[25] Halpern, J.Y.: Lexicographic probability, conditional probability, and non-standard probability. Games and Economic Behavior **68**, 155–179 (2010)

[26] Hammond, P.J.: Elementary non-Archimedean representations of probability for decision theory and games. In: P. Humphreys (ed.) Patrick Suppes: Scientific Philosopher; Volume 1, pp. 25–59. Kluwer, Dordrecht, The Netherlands (1994)

[27] Hammond, P.J.: Consequentialism, non-Archimedean probabilities, and lexicographic expected utility. In: C. Bicchieri, R. Jeffrey, B. Skyrms (eds.) The Logic of Strategy, chap. 2, pp. 39–66. Oxford University Press (1999)

[28] Holzer, S.: On coherence and conditional prevision. Boll. Un. Mat. Ital. Serie VI **IV-C**(1), 441–460 (1985)

[29] Kohlberg, E., Reny, P.J.: Independence on relative probability spaces and consistent assessments in game trees. Journal of Economic Theory **75**, 280–313 (1997)

[30] Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **14**(1), 167–207 (1990)

[31] Krauss, P.: Representation of conditional probability measures on Boolean algebras. Acta Mathematica Academiae Scientiarum Hungaricae **19**(3-4), 229–241 (1968)

[32] Maturo, A.: Conditional and non-standard probability. ESIT, 227–232 (2000)

[33] McGee, V.: Learning the impossible. In: E. Bells, B. Skyrms (eds.) Probability and Conditionals, pp. 179–199. Cambridge University Press (1994)

[34] McLennan, A.: Consistent conditional systems in noncooperative game theory. International Journal of Game Theory **18**, 141–174 (1989)

[35] Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, MA (1991)

[36] Myerson, R.B.: Multistage games with communication. Econometrica **54**(2), 323–358 (1986)

[37] Nelson, E.: Radically Elementary Probability Theory. Princeton University Press (1987)

[38] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, California (1988)

[39] Regazzini, E.: Finitely additive conditional probability. Rend. Sem. Mat. Fis. **55**, 69–89 (1985)

[40] Regazzini, E.: De Finetti's coherence and statistical inference. The Annals of Statistics **15**(2), 845–864 (1987)

[41] Swinkels, J.M.: Independence for conditional probability systems. Tech. Rep. 1076, Northwestern University, Center for Mathematical Studies in Economics and Management Science (1993)

[42] Vantaggi, B.: Conditional independence in a coherent finite setting. Annals of Mathematics and Artificial Intelligence **32**(1-4), 287–313 (2001)

[43] Vantaggi, B.: Graphical models for conditional independence structures. In: Second International Symposium on Imprecise Probabilities and Their Applications, pp. 332–341. Shaker (2001)

[44] Vantaggi, B.: The L-separation criterion for description of cs-independence models. International Journal of Approximate Reasoning **29**, 291–316 (2002)

[45] Vantaggi, B.: Incomplete preferences on conditional random quantities: Representability by conditional previsions. Mathematical Social Sciences **60**, 104–112 (2010)

[46] Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall (1991)

[47] Wenmackers, S., Horsten, L.: Fair infinite lotteries. Synthese **190**, 37–61 (2013)

[48] Wilson, N.: An order of magnitude calculus. In: Conference on Uncertainty in Artificial Intelligence, pp. 548–555 (1995)