International Journal of Approximate Reasoning 54 (2013) 1261–1278

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Independence for full conditional probabilities: Structure, factorization, non-uniqueness, and Bayesian networks

Fabio G. Cozman

Universidade de São Paulo – Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form 18 June 2013
Accepted 5 August 2013
Available online 29 August 2013

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

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 philosophy, 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.

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 [6–8,35], 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

E-mail address: fgcozman@usp.br.

0888-613X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2013.08.001

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) $\mathbb{P}(A|C) \geqslant 0$ 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.

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 $\{Y \neq y\}$.