Conditional Products: An Alternative Approach to Conditional Independence

A. P. Dawid
Department of Statistical Science
University College London
Gower Street, London WC1E 6BT, UK

M. Studený
Academy of Sciences of Czech Rep. and
University of Economics Prague
Pod vodárenskou věží 4, Prague 18208, CZ

Abstract

We introduce a new abstract approach to the study of conditional independence, founded on a concept analogous to the factorization properties of probabilistic independence, rather than the separation properties of a graph. The basic ingredient is the "conditional product", which provides a way of combining the objects under consideration while preserving as much independence as possible. We introduce an appropriate axiom system for conditional product, and show how, when these axioms are obeyed, it induces a derived concept of conditional independence which obeys the usual semi-graphoid axioms. The general structure is used to throw light on three specific areas: the familiar probabilistic framework (both the discrete and the general case); a set-theoretic framework related to "variation independence"; and a variety of graphical frameworks.

Key words: Directed graph independence, Probabilistic independence, Projection, Semi-graphoid, Undirected graph independence, Variation independence.

1 MOTIVATION

In several distinct mathematical areas, especially those describing uncertainty in Probability and Statistics (Dawid 1979) and Artificial Intelligence (Studený 1993; Shenoy 1994), some concept of *conditional independence* plays a fundamental role, permitting the decomposition of a complex object into simpler pieces. This concept was introduced in miscellaneous frameworks, but certain reasonable formal properties are shared. These formal properties are described by the abstract theory of conditional independence, as captured by the "semi-graphoid" axiom system (Dawid 1979; Spohn 1980; Pearl 1988). The archetypal model for this system is the concept of *separation* in an undirected graph, and the axioms can most readily be understood as being about a generalized form of separation.

An entirely different approach to conditional independence is to try and abstract the *factorization* properties which form the traditional basis for the probabilistic definition. This is the task we attempt here. We first introduce the abstract concept of *conditional product*, and propose a suitable axiom system for it. We use this to introduce conditional independence as a derived concept, and show that it does then satisfy the semi-graphoid axioms. Such a point of view provides a unifying framework for conditional independence, and suggests new forms and applications. However, not every semi-graphoid can arise in this way. In this paper we describe the appropriate conditional product construction in probabilistic and set-theoretic frameworks for conditional independence, as well as showing that, in the frameworks of undirected or directed graphical models, no suitable conditional product exists.

2 CONDITIONAL PRODUCTS

In this Section we give an axiomatic definition of projection and conditional product. Then we show that an induced concept of conditional independence satisfies the semi-graphoid axioms.

2.1 AXIOMS FOR PROJECTION AND CONDITIONAL PRODUCT

We have an index set N; a class $\mathcal{N}$ of subsets of N, containing $\emptyset$ and closed under union and intersection; and a set Υ of "objects". Associated with any $\phi \in \Upsilon$ is its *domain* $d(\phi) \in \mathcal{N}$.

2.1.1 Projection

For any $\phi \in \Upsilon$ and $E \in \mathcal{N}$ there exists an object $\phi^E \in \Upsilon$, the *projection* of ϕ onto E. Projection has the following properties:

P1 $d(\phi^E) = d(\phi) \cap E$ for $\phi \in \Upsilon$, $E \in \mathcal{N}$.

P2 $(\phi^E)^F = \phi^{E\cap F}$ for $\phi \in \Upsilon$, $E, F \in \mathcal{N}$.

P3 $\phi^{d(\phi)} = \phi$ for $\phi \in \Upsilon$.

The properties above imply that $(\phi^E)^F = (\phi^F)^E$ and $\phi^E = \phi^{d(\phi)\cap E}$ for $\phi \in \Upsilon$, $E, F \in \mathcal{N}$. In particular, $d(\phi) \subseteq E$ implies $\phi^E = \phi$ by P3.

We say that $\phi, \psi \in \Upsilon$ are *(weakly) compatible* if $\phi^{d(\psi)} = \psi^{d(\phi)}$ and *strongly compatible* if there exists $\tau \in \Upsilon$ such that $\tau^{d(\phi)} = \phi$ and $\tau^{d(\psi)} = \psi$. Strong compatibility implies weak compatibility, but not necessarily conversely. The collection of pairs of weakly compatible objects from Υ will be denoted by $\mathcal{C}$, and the collection of pairs of strongly compatible objects from Υ by $\mathcal{C}_*$.

2.1.2 Conditional Product

We consider a mapping $\otimes : \mathcal{D} \to \Upsilon$ having domain $\mathcal{D} \subseteq \mathcal{C}$ (often we shall have $\mathcal{D} = \mathcal{C}$). The conditional product operation $\otimes$ is required to have the following properties:

T1 $(\phi, \psi) \in \mathcal{D} \Rightarrow d(\phi \otimes \psi) = d(\phi) \cup d(\psi)$.

T2 $(\phi, \psi) \in \mathcal{D} \Rightarrow (\psi, \phi) \in \mathcal{D}$ and $\phi \otimes \psi = \psi \otimes \phi$.

T3 $\phi \in \Upsilon, E \in \mathcal{N} \Rightarrow (\phi, \phi^E) \in \mathcal{D}$ and $\phi \otimes \phi^E = \phi$.

T4 $(\phi, \psi) \in \mathcal{D}, d(\phi) \subseteq E \in \mathcal{N} \Rightarrow (\phi, \psi^E) \in \mathcal{D}$ and $(\phi \otimes \psi)^E = \phi \otimes \psi^E$.

Axiom T4 can be formulated in apparently stronger form:

T4′ $(\phi, \psi) \in \mathcal{D}, d(\phi) \cap d(\psi) \subseteq E \in \mathcal{N} \Rightarrow (\phi, \psi^E) \in \mathcal{D}$ and $(\phi \otimes \psi)^{d(\phi)\cup E} = \phi \otimes \psi^E$.

Indeed, if $d(\phi) \cap d(\psi) \subseteq E$ one has by P3 and P2 $\psi^{d(\phi)\cup E} = \psi^{d(\psi)\cap(d(\phi)\cup E)} = \psi^{d(\psi)\cap E} = \psi^E$.

T5 $(\phi, \psi) \in \mathcal{D}, d(\phi) \cap d(\psi) \subseteq E \in \mathcal{N} \Rightarrow (\phi, \psi^E) \in \mathcal{D}, ((\phi \otimes \psi^E), \psi) \in \mathcal{D}$, and $\phi \otimes \psi = (\phi \otimes \psi^E) \otimes \psi$.

T6 $\phi, \psi \in \Upsilon, d(\phi) \cap d(\psi) \subseteq E \in \mathcal{N}, (\phi, \psi^E) \in \mathcal{D}, ((\phi \otimes \psi^E), \psi) \in \mathcal{D} \Rightarrow (\phi, \psi) \in \mathcal{D}$.

Using T4′, we can also restate T5 as:

T5′ $(\phi, \psi) \in \mathcal{D}, d(\phi) \cap d(\psi) \subseteq E \in \mathcal{N} \Rightarrow ((\phi \otimes \psi)^{d(\phi)\cup E)}, \psi) \in \mathcal{D}$ and $\phi \otimes \psi = ((\phi \otimes \psi)^{d(\phi)\cup E)} \otimes \psi)$.

We can further re-express T4–T6 in terms of pairwise disjoint $A, B, C, D \in \mathcal{N}$ (using $\psi^E = \psi^{d(\phi)\cap E}$, by P2):

T4″ $(\phi, \psi) \in \mathcal{D}, d(\phi) = A \cup C, d(\psi) = B \cup C \cup D \Rightarrow (\phi, \psi^{C\cup D}) \in \mathcal{D}$ and $(\phi \otimes \psi)^{A\cup C\cup D} = \phi \otimes \psi^{C\cup D}$.

T5″ $(\phi, \psi) \in \mathcal{D}, d(\phi) = A \cup C, d(\psi) = B \cup C \cup D \Rightarrow ((\phi \otimes \psi)^{A\cup C\cup D}, \psi) \in \mathcal{D}$ and $\phi \otimes \psi = (\phi \otimes \psi)^{A\cup C\cup D} \otimes \psi$.

T6″ $d(\phi) = A \cup C, d(\psi) = B \cup C \cup D, (\phi, \psi^{C\cup D}) \in \mathcal{D}$ and $((\phi \otimes \psi^{C\cup D}), \psi) \in \mathcal{D} \Rightarrow (\phi, \psi) \in \mathcal{D}$.

Observation 2.1 $(\phi \otimes \psi)^{d(\phi)} = \phi$ and $(\phi \otimes \psi)^{d(\psi)} = \psi$ whenever $(\phi, \psi) \in \mathcal{D}$.

Proof: Use T4, the fact that $(\phi, \psi) \in \mathcal{C}$ and T3 to write $(\phi \otimes \psi)^{d(\phi)} = \phi \otimes \psi^{d(\phi)} = \phi \otimes \phi^{d(\psi)} = \phi$. Use T2 for the other equality. □

Thus, when all the above axioms hold and $\mathcal{D} = \mathcal{C}$, we shall have $\mathcal{C} = \mathcal{C}_*$, since then we can take $\tau = \phi \otimes \psi$.

2.2 SEMI-GRAPHOIDS AND CONDITIONAL INDEPENDENCE

2.2.1 Semi-Graphoid

A ternary operation $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ on $\mathcal{N}$ is called a (full) *semi-graphoid* if it satisfies:

C0 $A \perp\!\!\!\perp B \mid C$ if $B \subseteq C$.

C1 $A \perp\!\!\!\perp B \mid C \Rightarrow B \perp\!\!\!\perp A \mid C$.

C2 $A \perp\!\!\!\perp (B \cup D) \mid C \Rightarrow A \perp\!\!\!\perp D \mid C$.

C3 $A \perp\!\!\!\perp (B \cup D) \mid C \Rightarrow A \perp\!\!\!\perp B \mid (C \cup D)$.

C4 $A \perp\!\!\!\perp B \mid (C \cup D)$ and $A \perp\!\!\!\perp D \mid C \Rightarrow A \perp\!\!\!\perp (B \cup D) \mid C$.

A *partial semi-graphoid* on a class of triplets $\mathcal{K} \subseteq \mathcal{N} \times \mathcal{N} \times \mathcal{N}$ is a predicate $\perp\!\!\!\perp$ having $\mathcal{K}$ as domain such that C0–C4 hold under the additional constraint that all triplets involved belong to $\mathcal{K}$. We shall here limit attention to special partial semi-graphoids, where $\mathcal{K}$ is the class of triplets of pairwise disjoint finite subsets of N; equivalently, C0–C4 are only required when A, B, C, D are pairwise disjoint. These semi-graphoids will be called *disjoint semi-graphoids over N*.

2.2.2 Conditional Independence

For every $\varphi \in \Upsilon$ and pairwise disjoint finite sets $A, B, C \in \mathcal{N}$ we write $A \perp\!\!\!\perp B \mid C\ [\varphi]$ if $\varphi^{A\cup B\cup C} = \varphi^{A\cup C} \otimes \varphi^{B\cup C}$ (this includes the requirement $(\varphi^{A\cup C}, \varphi^{B\cup C}) \in \mathcal{D}$).

Observation 2.2 $A \perp\!\!\!\perp B \mid C\ [\phi]$ iff $A \perp\!\!\!\perp B \mid C\ [\psi]$ for $\phi, \psi \in \Upsilon$ and $\phi^{A\cup B\cup C} = \psi^{A\cup B\cup C}$.

Proof: A consequence of P2 and the definition. □

Observation 2.3 $A \perp\!\!\!\perp B \mid C\ [\varphi^{A\cup C} \otimes \varphi^{B\cup C}]$ whenever $\varphi \in \Upsilon$ and $(\varphi^{A\cup C}, \varphi^{B\cup C}) \in \mathcal{D}$.

Proof: Put $\phi = \varphi^{A\cup C}$, $\psi = \varphi^{B\cup C}$, and $\tau = \phi \otimes \psi$. Since $d(\tau) \subseteq A \cup B \cup C$ (use P1 and T1) $\tau^{A\cup B\cup C} = \tau$. By P2 $\psi^{A\cup C} = \varphi^C = \phi^C$ and one can write using T4 and T3: $\tau^{A\cup C} = \phi \otimes \psi^{A\cup C} = \phi \otimes \phi^C = \phi$. Similarly, using T2, $\tau^{B\cup C} = \psi$. □

Proposition 2.1 *For $\varphi \in \Upsilon$, the collection of triplets $\langle A, B|C \rangle$ of pairwise disjoint finite subsets of N for which $A \perp\!\!\!\perp B \mid C\ [\varphi]$ forms a disjoint semi-graphoid over N.*

Proof: Without loss of generality suppose below that the sets A, B, C, D in C0–C4 are subsets of $d(\varphi)$ (otherwise replace every set by its intersection with $d(\varphi)$). With B and C disjoint, C0 becomes $A \perp\!\!\!\perp \emptyset \,|\, C\,[\varphi]$, which follows from T3 with $\phi = \varphi^{A\cup C}$, $E = C$ (note that $\phi^C = \varphi^C$ by P2). C1 follows from T2 with $\phi = \varphi^{A\cup C}$, $\psi = \varphi^{B\cup C}$. From this point on, define $\phi = \varphi^{A\cup C}$, $\psi = \varphi^{B\cup C\cup D}$. If $A \perp\!\!\!\perp (B \cup D) \,|\, C\,[\varphi]$, then $\varphi^{A\cup B\cup C\cup D} = \phi \otimes \psi$. So, from P2, $\varphi^{A\cup C\cup D} = (\phi \otimes \psi)^{A\cup C\cup D}$, $= \phi \otimes \psi^{C\cup D}$ by T4″, $= \varphi^{A\cup C} \otimes \varphi^{C\cup D}$ (again using P2), so that C2 holds. Also, C3 then follows easily from T5″ and P2. Finally, $A \perp\!\!\!\perp B \,|\, (C \cup D)\,[\varphi]$ and $A \perp\!\!\!\perp D \,|\, C\,[\varphi]$ imply $\varphi^{A\cup B\cup C\cup D} = (\phi \otimes \psi^{C\cup D}) \otimes \psi$, which $= \phi \otimes \psi$ by T6″ and T5″ with $E = C \cup D$, so that C4 holds. □

Our axioms for projection and conditional product resemble those for marginalization and combination of valuations (Shenoy and Shafer 1990; Shenoy 1994), which were motivated by the desire to establish an axiomatic framework for belief propagation in join trees, rather than as a framework for conditional independence. The main difference is that their combination is defined also for non-compatible valuations. However, Shenoy (1994) also derived graphoid properties from his axioms for valuations.

3 PROBABILISTIC FRAMEWORK

The classic example of objects satisfying our axioms is given by probability measures on Cartesian products of arbitrary measurable spaces. We start with the important special case of discrete probability measures, then treat the general case. Throughout, we take $\mathcal{N}$ to be the class of finite subsets of N.

3.1 DISCRETE CASE

For each $i \in N$ we are given a non-empty finite set $\mathbf{X}_i$. For non-empty $D \in \mathcal{N}$ we define $\mathbf{X}_D = \prod_{i\in D} \mathbf{X}_i$, while $\mathbf{X}_\emptyset$ is taken to be some fixed singleton set $\{\bullet\}$. An object with domain D is a distribution over D, *i.e.* a non-negative function p on $\mathbf{X}_D$ such that $\sum\{\, p(\mathbf{x}) : \mathbf{x} \in \mathbf{X}_D\} = 1$.

The projection of p over D onto $E \in \mathcal{N}$ is the marginal distribution p^E on $\mathbf{X}_{E\cap D}$, defined by the formula:

$$p^E(\mathbf{y}) = \sum\{\, p(\mathbf{x}, \mathbf{y}) \;:\; \mathbf{x} \in \prod_{i\in D\setminus E} \mathbf{X}_i \,\}$$

for every $\mathbf{y} \in \mathbf{X}_{E\cap D}$ (with obvious modification if $E \cap D = \emptyset$). Of course, $p^E = p$ if $D \setminus E = \emptyset$.

The conditional product of a distribution p over $E \in \mathcal{N}$ and a distribution q over $F \in \mathcal{N}$ will be defined for every pair of weakly compatible distributions, so that $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{D}}$ in this case. Supposing p is defined on $\mathbf{X}_E$ and q on $\mathbf{X}_F$, $p \otimes q$ is defined on $\mathbf{X}_{E\cup F}$ as follows:

$$(p \otimes q)\,(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \begin{cases} \frac{p(\mathbf{x},\mathbf{y})\cdot q(\mathbf{y},\mathbf{z})}{p^{E\cap F}(\mathbf{y})} & \text{if } p^{E\cap F}(\mathbf{y}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

for every $\mathbf{x} \in \prod_{i\in E\setminus F} \mathbf{X}_i$, $\mathbf{y} \in \mathbf{X}_{E\cap F}$, $\mathbf{z} \in \prod_{i\in F\setminus E} \mathbf{X}_i$. Of course, if $E \setminus F = \emptyset$, then $\mathbf{x}$ is omitted; if $F \setminus E = \emptyset$, then $\mathbf{z}$ is omitted. Of course, $p^{E\cap F}$ in (1) can be replaced by $q^{E\cap F}$.

The reader can verify directly:

Proposition 3.1 *The properties P1–P3 and T1–T6 hold for the above defined projection and conditional product of discrete probability distributions.*

Thus, by Proposition 2.1 every probability distribution p over $D \in \mathcal{N}$ induces a disjoint semi-graphoid over N as follows: for every triplet $\langle A, B|C\rangle$ of pairwise disjoint finite subsets of N one has $A \perp\!\!\!\perp B \,|\, C\;[p]$ iff

$$p^{A\cup B\cup C}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \cdot p^C(\mathbf{y}) \;=\; p^{A\cup C}(\mathbf{x}, \mathbf{y}) \cdot p^{B\cup C}(\mathbf{y}, \mathbf{z})$$

for every $\mathbf{x} \in \prod_{i\in A\cap D} \mathbf{X}_i$, $\mathbf{y} \in \mathbf{X}_{C\cap D}$, $\mathbf{z} \in \prod_{B\cap D} \mathbf{X}_i$ (thus corresponding to the standard probabilistic definition of conditional independence for this case). Let us call these semi-graphoids (as triplets of pairwise disjoint subsets of N) *discrete probabilistic models*.

3.2 GENERAL CASE

Now for every $i \in N$ a measurable space $(\mathbf{X}_i, \mathcal{X}_i)$ is given and an object, with domain $D \in \mathcal{N}$, is a probability measure P on the product space $(\mathbf{X}_D, \mathcal{X}_D) \equiv \prod_{i\in D}(\mathbf{X}_i, \mathcal{X}_i)$. The projection of P onto $E \in \mathcal{N}$ is the marginal measure of P on $(\mathbf{X}_{E\cap D}, \mathcal{X}_{E\cap D})$:

$$P^E(\mathsf{T}) = P(\mathsf{T} \times \mathbf{X}_{D\setminus E}) \qquad \text{for every } \mathsf{T} \in \mathcal{X}_{E\cap D}\,.$$

Of course, $P^E = P$ if $D \setminus E = \emptyset$. An example showing that weak compatibility may not imply strong compatibility in this general framework (that is $\boldsymbol{\mathcal{C}} \neq \boldsymbol{\mathcal{C}}_*$) is given in Appendix B (Example B.1).

The definition of conditional product is now more technical (for relevant background see, for example, Neveu (1964)). Let P having domain $E \in \mathcal{N}$ and Q having domain $F \in \mathcal{N}$ be weakly compatible. Given $\mathsf{T} \in \mathcal{X}_{E\setminus F}$ (where $E\setminus F \neq \emptyset$), by a representative of the conditional probability of T given $\mathcal{X}_{E\cap F}$ induced by P is understood any $\mathcal{X}_{E\cap F}$-measurable non-negative function $P(\mathsf{T}|\cdot)$ on $\mathbf{X}_{E\cap F}$ such that

$$P(\mathsf{T} \times \mathsf{U}) = \int_{u\in\mathsf{U}} P(\mathsf{T}|u)\; dP^{E\cap F}(u) \tag{2}$$

for every $\mathsf{U} \in \mathcal{X}_{E\cap F}$. Observe that in the discrete case one has:

$$P(\{t\}|u) \;=\; \frac{p(t, u)}{p^{E\cap F}(u)}$$

for $t \in \mathbf{X}_{E\setminus F}$ and $u \in \mathbf{X}_{E\cap F}$ with $p^{E\cap F}(u) > 0$. For each T such a representative exists, and any two representatives can differ only on a set in $\mathcal{X}_{E\cap F}$ of P-probability 0. However, it is not always possible to choose representatives for all $\mathsf{T} \in \mathcal{X}_{E\setminus F}$ to ensure σ-additivity over T.

Conditional probability $Q(\mathsf{V}|\cdot)$ of $\mathsf{V} \in \mathcal{X}_{F\setminus E}$ given $\mathcal{X}_{E\cap F}$ induced by Q is defined analogously. Then one can introduce the following set function:

$$R(\mathsf{T} \times \mathsf{U} \times \mathsf{V}) = \int_{u \in \mathsf{U}} P(\mathsf{T}|u) \cdot Q(\mathsf{V}|u) \, dP^{E\cap F}(u) \quad (3)$$

for $\mathsf{T} \in \mathcal{X}_{E\setminus F}$, $\mathsf{U} \in \mathcal{X}_{E\cap F}$, $\mathsf{V} \in \mathcal{X}_{F\setminus E}$. Of course, Cartesian products over the empty set are omitted and $P^{E\cap F} = Q^{E\cap F}$ is the common projection of P and Q onto $E \cap F$.

Equation (3) defines R only on the subclass $\mathcal{S}$ of sets in $\mathcal{X}_{E\cup F}$ of the form $\mathsf{T} \times \mathsf{U} \times \mathsf{V}$, as described above. Then R is finitely additive, but need not be σ-additive, on $\mathcal{S}$. If it is σ-additive, it can be uniquely extended to a σ-additive probability measure on $\mathcal{X}_{E\cup F}$, which we also denote by R. In this case, $(P,Q) \in \boldsymbol{\mathcal{D}}$, and we define the conditional product $P \otimes Q = R$. Otherwise, $(P,Q) \notin \boldsymbol{\mathcal{D}}$, and $P \otimes Q$ is not defined.

An example showing that strong compatibility does not imply the existence of the conditional product (*i.e.* $\boldsymbol{\mathcal{D}} \neq \boldsymbol{\mathcal{C}}_*$) is given in Appendix B (Example B.2).

With the above definition, the corresponding concept of conditional independence is given by: for pairwise disjoint $A, B, C \in \mathcal{N}$, $A \perp\!\!\!\perp B \mid C \, [P]$ if:

$$\text{for } \mathsf{T} \in \mathcal{X}_A, \mathsf{V} \in \mathcal{X}_B, \mathsf{U} \in \mathcal{X}_C,$$
$$P(\mathsf{T} \times \mathsf{U} \times \mathsf{V}) = \textstyle\int_{u\in\mathsf{U}} P(\mathsf{T}|u) \cdot P(\mathsf{V}|u) \, dP^C(u). \quad (4)$$

We note, by Observation 2.3, that $A \perp\!\!\!\perp B \mid C \, [P \otimes Q]$ for $(P,Q) \in \boldsymbol{\mathcal{D}}$, $d(P) = A \cup C$, $d(Q) = B \cup C$, since $(P \otimes Q)^{A\cup C} = P$, $(P \otimes Q)^{B\cup C} = Q$.

We assert, without proof, that equivalent statements to (4) are:

$$\begin{aligned} P(\mathsf{T} \times \mathsf{V} \mid u) &= P(\mathsf{T} \mid u) P(\mathsf{V} \mid u) && \text{a.s. } [P^C] && (5)\\ P(\mathsf{T} \mid u, v) &= P(\mathsf{T} \mid u) && \text{a.s. } [P^{B\cup C}] && (6)\\ P(\mathsf{V} \mid t, u) &= P(\mathsf{V} \mid u) && \text{a.s. } [P^{A\cup C}]. && (7)\end{aligned}$$

Further, (6) and (7) are equivalent to the existence of a $\mathcal{X}_C$-measurable representative of $P(\mathsf{T} \mid u, v)$ or of $P(\mathsf{V} \mid t, u)$, respectively.

The following proposition is proved in Appendix A:

Proposition 3.2 *The axioms of §2.1 hold for the above defined conditional product of probability measures.*

It now follows from Propositions 3.2 and 2.1 that general probabilistic conditional independence, as defined by (4) above *via* conditional products, induces a disjoint semi-graphoid, which we may term a *probabilistic model*.

4 SET-THEORETIC FRAMEWORK

In this Section we consider a framework which is in some ways analogous to the probabilistic one, but much simpler. The corresponding concept of conditional independence will turn out to be *variation independence* (Dawid 1998), which arises naturally in the context of relational databases, and has applications in the statistical analysis of graphical models (Dawid and Lauritzen 1993).

We again have a space $\mathbf{X}_i$ for every $i \in N$ but the σ-algebra $\mathcal{X}_i$ is no longer required. An object S with domain $D \in \mathcal{N}$ will now be an arbitrary subset of $\mathbf{X}_D$. We start by defining projection and conditional product for points. Thus let $\boldsymbol{x} = (x_i : i \in D)$. Then for $E \in \mathcal{N}$ we define its projection onto E to be $\boldsymbol{x}^E = (x_i : i \in D \cap E)$. We say that two points $\boldsymbol{x} = (x_i : i \in E)$ and $\boldsymbol{y} = (y_i : i \in F)$ are compatible if $x_i = y_i$ for every $i \in E \cap F$, *i.e.* they have the same projections onto $E \cap F$; and in this case define their conditional product $\boldsymbol{x} \otimes \boldsymbol{y}$ as $\boldsymbol{z} = (z_i : i \in E \cup F)$ where $z_i = x_i$ if $i \in E$, $z_i = y_i$ if $i \in F$. Then projection of an object S is defined pointwise: $S^E = \{\boldsymbol{x}^E : \boldsymbol{x} \in S\}$. If S and T are two objects, with respective domains E and F, they will be compatible if their projections onto $E \cap F$ coincide; and in this case we take $(S,T) \in \boldsymbol{\mathcal{D}}$, defining $S \otimes T = \{\boldsymbol{x} \otimes \boldsymbol{y} : \boldsymbol{x} \in S, \boldsymbol{y} \in T, \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are compatible}\}$. It is then not difficult to verify:

Proposition 4.1 *The axioms of §2.1 hold for the above defined conditional product of sets.*

The corresponding definition of conditional independence is given by: for disjoint $A, B, C \subseteq N$, $A \perp\!\!\!\perp B \mid C \, [S]$ if, for each $\boldsymbol{z} \in S^C$, $\{(\boldsymbol{x}^A, \boldsymbol{x}^B) : \boldsymbol{x} \in S, \boldsymbol{x}^C = \boldsymbol{z}\}$ is a Cartesian product. That is to say, as a point varies in S subject to having a given projection onto C, there are no constraints relating its projections onto A and onto B. By Proposition 2.1, this concept of "variation independence" defines a disjoint semi-graphoid on N.

5 GRAPHICAL FRAMEWORK

In this Section we explore two specific graphical frameworks which are widely used in Artificial Intelligence: undirected graphs and directed acyclic graphs.

5.1 UNDIRECTED GRAPHS

An *undirected graph* G over a finite set of nodes $D \subseteq N$ is specified by a collection $\mathcal{L}$ of two-element subsets of D which are called edges. We call D the nodeset, and $\mathcal{L}$ the edgeset, of G, and write $G = (D, \mathcal{L})$. A path in G is a sequence of distinct nodes $w_1, \ldots, w_n$, $n \geq 1 \in D$ such that $\{w_i, w_{i+1}\} \in \mathcal{L}$ for $i = 1, \ldots, n-1$. We say that a triplet $\langle A, B | C \rangle$ of pairwise disjoint finite subsets of N is represented on G and write $A \perp\!\!\!\perp B \mid C \, [G]$ if for every path $w_1, \ldots, w_n$ in G with $w_1 \in A$ and $w_n \in B$ there exists $1 < i < n$ with $w_i \in C$. It is no problem to verify that $\cdot \perp\!\!\!\perp \cdot \mid \cdot \, [G]$ forms a disjoint semi-graphoid over N (see for example Pearl (1988)). Let us call such semi-graphoids *UG-models* over N.

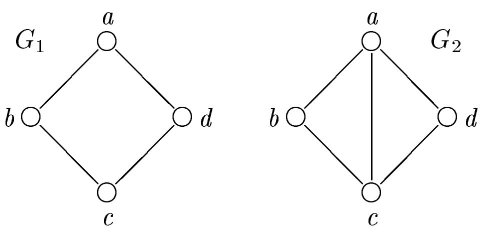

Figure 1: Two Undirected Graphs

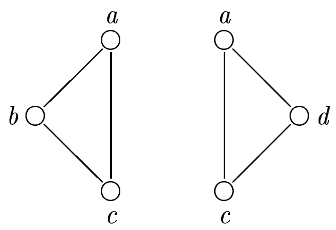

Figure 2: Their Common Projections

A natural question arises: Is it possible to define the above mentioned conditional independence predicate $\perp\!\!\!\perp$ by means of a suitable conditional product operation on undirected graphs? The answer is negative.

First, suppose for contradiction that separation in undirected graphs can be equivalently defined by means of some projection and conditional product satisfying the axioms from §2.1. Let G^E be the projection of an undirected graph G over $D \in \mathcal{N}$ onto $E \in \mathcal{N}$. Then Observation 2.2 and P2 imply $A \perp\!\!\!\perp B \,|\, C\ [G^E]$ iff $A \perp\!\!\!\perp B \,|\, C\ [G]$ for every triplet $\langle A, B|C\rangle$ of pairwise disjoint finite sets of $A, B, C \subseteq E \cap D$. Observe that $\{u, v\}$ is an edge in an undirected graph H over $E \cap D$ iff $\neg(\{u\} \perp\!\!\!\perp \{v\} \,|\, E \cap D \setminus \{u, v\}\ [H])$. Hence, $\{u, v\} \subseteq E \cap D$ is an edge in G^E iff there exists a path $u = w_1, \ldots, w_n = v$, $n \geq 2$ in G with $w_i \in D \setminus E$ for $1 < i < n$. Thus, the projection G^E is uniquely determined. Note that projection operation defined in this way satisfies P1–P3.

To show that no reasonable conditional product operation can be introduced within this framework, consider the two graphs over $\{a, b, c, d\}$ from Figure 1. Evidently $\{b\} \perp\!\!\!\perp \{d\} \,|\, \{a, c\}\ [G_i]$ for $i = 1, 2$. In case $\perp\!\!\!\perp$ can be defined by means of a conditional product operation $\otimes$ deduce from the definition in §2.2 $G_i = G_i^{\{a,b,c\}} \otimes G_i^{\{a,c,d\}}$ for $i = 1, 2$. However, the corresponding projections of G_i onto $\{a, b, c\}$ and onto $\{a, c, d\}$ coincide — they are as in Figure 2. Hence, a contradictory conclusion $G_1 = G_2$ is derived. Thus, we have:

Consequence 5.1
Conditional independence for undirected graphs cannot be defined by means of projection and conditional product operations on undirected graphs.

The reader may incline to conclude that conditional product operation cannot induce UG-models at all. However, this is not true. The reason is that the framework of undirected graphs can be considered as a subframework of the discrete probabilistic framework. Geiger and Pearl (1993) showed that for every UG-model over $D \subseteq N$ there exists a discrete probability distribution over D inducing it (see also Studený and Bouckaert (1998/9)). Proposition 3.1 says that every such discrete probabilistic model can be defined by means of a conditional product (on discrete probability distributions).

5.1.1 An Alternative Construction

Although we have seen that there is no way of defining operations of projection and conditional product on undirected graphs so as to reconstruct the standard definition of graphical conditional independence (separation), other constructions are possible, leading to new concepts of graphical separation. Thus let $G = (D, \mathcal{L})$ be an undirected graph on $D \subseteq N$. For $A \subseteq D$ the projection of G onto A is just the induced subgraph $G^A = (A, \mathcal{L} \cap (A \times A))$, while for more general A we define $G^A = G^{A \cap D}$. Then two graphs G and H are compatible if they have exactly the same edges between all nodes common to both their domains. In this case we take $(G, H) \in \boldsymbol{\mathcal{D}}$, and define their conditional product $G \otimes H$ as the graph whose nodeset and edgeset are obtained as the unions of the corresponding nodesets and edgesets of G and H. It is readily seen that all the axioms of Section 2.1 are satisfied for these definitions. Correspondingly, we have the following definition of conditional independence with respect to a graph $G = (D, \mathcal{L})$: for disjoint $A, B, C \subseteq N$, $A \perp\!\!\!\perp B \,|\, C\ [G]$ if there is no edge in $\mathcal{L}$ containing one element in A and the other in B. By Proposition 2.1, this definition yields a disjoint semi-graphoid over N. However, it is distinct from any of the usual forms motivated by probabilistic independence. In particular, we see that, for given A and B, $A \perp\!\!\!\perp B \,|\, C$ holds or fails simultaneously for every "conditioning set" C (so long only as the disjointness requirement is maintained).

5.2 DIRECTED ACYCLIC GRAPHS

A directed graph H over a finite set of nodes $D \in \mathcal{N}$ is specified by a collection $\mathcal{A}$ of ordered pairs (u, v) of distinct nodes, called arrows. A descending path in H is a sequence of distinct nodes $w_1, \ldots, w_n$, $n \geq 1$ such that $(w_i, w_{i+1}) \in \mathcal{A}$ for $i = 1, \ldots, n-1$. It is called a directed cycle if moreover $(w_n, w_1) \in \mathcal{A}$. A *directed acyclic graph* is a directed graph without directed cycles.

Testing whether a triplet $\langle A, B|C\rangle$ of pairwise disjoint finite subsets of N is represented in such a graph is more complicated than in the undirected case. Let us describe the moralization criterion of Lauritzen *et al.*

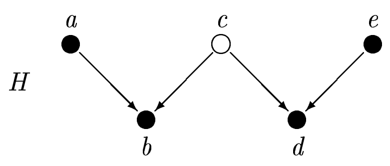

Figure 3: A Directed Acyclic Graph

(1990) which is equivalent to the d-separation criterion of Pearl (1988). First, consider the set E of ancestors of $A \cup B \cup C$, that is the set of nodes $u \in D$ such that there exists a descending path $u = w_1, \ldots, w_n \in A \cup B \cup C$, $n \geq 1$ in H. An undirected graph G over E, called the moral graph of H, is constructed as follows: $\{u, v\} \in \mathcal{L}$ iff $(u, v) \in \mathcal{A}$ or $(v, u) \in \mathcal{A}$ or there exists $w \in E$ with $(u, w) \in \mathcal{A}$ and $(v, w) \in \mathcal{A}$. The triplet $\langle A, B|C \rangle$ is represented in H, denoted by $A \perp\!\!\!\perp B \mid C\ [H]$, if it is represented in the moral graph G in the sense described in §5.1. Again, the reader can verify that $\cdot \perp\!\!\!\perp \cdot \mid \cdot\ [H]$ determines a disjoint semi-graphoid over N. Let us call such semi-graphoids *DAG-models*.

One can raise a question analogous to that in the previous case: Is there any conditional product operation on directed acyclic graphs inducing the conditional independence predicate above? The answer is again negative. The reason is even more basic than in undirected case: no reasonable projection operation exists. To see this, consider the graph H of Figure 3 and ask what could be the projection of H onto $F = \{a, b, d, e\}$? By Observation 2.2 the DAG-model over F induced by this projection is uniquely determined. However, we observe the following.

Lemma 5.1 There is no directed acyclic graph K over $\{a, b, d, e\}$ such that for every triplet $A, B, C \subseteq \{a, b, d, e\}$ of pairwise disjoint sets $A \perp\!\!\!\perp B \mid C\ [K]$ iff $A \perp\!\!\!\perp B \mid C\ [H]$.

Proof: The argument is based on two observations concerning an arbitrary directed acyclic graph K over F. The first observation is that $\{u, v\}$ is an edge in K (that is either (u, v) or (v, u) is an arrow in K) iff $\neg(\{u\} \perp\!\!\!\perp \{v\} \mid L\ [K])$ for every $L \subseteq F \setminus \{u, v\}$. Indeed, to show sufficiency of this condition take as L the set of ancestors of $\{u, v\}$ in K, excluding u and v. The second observation is that whenever $\{u, w\}$ and $\{v, w\}$ are edges in K but $\{u, v\}$ is not an edge in K then (u, w) and (v, w) are simultaneously arrows in K iff $\neg(\{u\} \perp\!\!\!\perp \{v\} \mid L\ [K])$ for every $\{w\} \subseteq L \subseteq F \setminus \{u, v\}$. Indeed, to show sufficiency of the condition take as L the set of ancestors of $\{u, v, w\}$ in K excluding u and v.

Suppose for contradiction the existence of a graph K over $\{a, b, d, e\}$ satisfying the required conditions. Then the fact $\neg(\{d\} \perp\!\!\!\perp \{e\} \mid L\ [H])$ for every $L \subseteq \{a, b\}$ implies that $\{d, e\}$ is an edge in K. Similarly, the fact $\neg(\{b\} \perp\!\!\!\perp \{d\} \mid L\ [H])$ for every $L \subseteq \{a, e\}$ implies that $\{b, d\}$ is an edge in K. The fact $\{b\} \perp\!\!\!\perp \{e\} \mid \emptyset\ [H]$ implies that $\{b, e\}$ is not an edge in K. Finally, the fact $\neg(\{b\} \perp\!\!\!\perp \{e\} \mid L\ [H])$ for $\{d\} \subseteq L \subseteq \{a, d\}$ implies that (b, d) and (e, d) are arrows in K. However, by the same consideration, interchanging a with e and b with d, derive that (d, b) and (a, b) are arrows in K. Then the conclusion that both (b, d) and (d, b) are arrows in K contradicts the assumption that K is acyclic. □

It follows from Lemma 5.1 and Observation 2.2 that a suitable projection of H onto F cannot be defined. Thus, we have:

Consequence 5.2 *Conditional independence for directed acyclic graphs cannot be defined by means of projection and conditional product operations on directed acyclic graphs.*

Again, as in the undirected case, the framework of directed acyclic graphs can be considered as a subframework of the discrete probabilistic framework. The result that every DAG-model is a discrete probabilistic model is given by Geiger and Pearl (1990); see also Studený and Bouckaert (1998/9).

6 CONCLUSION

We have introduced an abstract concept of conditional product, $\otimes$, and shown how it can be used to derive an induced concept of conditional independence, $\perp\!\!\!\perp$. Conversely, if we start with a definition of $\perp\!\!\!\perp$ obeying the semi-graphoid axioms, we can search for an associated conditional product $\otimes$: in particular, for $(\phi, \psi) \in \boldsymbol{\mathcal{D}}$ with $d(\phi) = A$, $d(\psi) = B$, $\tau = \phi \otimes \psi$ would have to satisfy $\tau^A = \phi$, $\tau^B = \psi$, and $A \perp\!\!\!\perp B \mid C\,[\tau]$ (see Observation 2.3). As we have seen in the framework of graphical models, this is not always possible. In other contexts, the required construction may be possible but non-unique.

We have developed the theory of semi-graphoids and conditional products for the special case of domains which are subsets of some given index set N. Correspondingly, our definitions in the probabilistic and set-theoretic frameworks have been based on sample spaces which are Cartesian products. In fact it is possible to develop the theory in a still more general framework, in which we need only require that the class of possible domains be an abstract join semi-lattice (Birkhoff 1949).

The general abstract algebraic structures underlying conditional independence would seem to have considerable independent mathematical interest, and promise to richly repay further study. Also, they bring a unifying point of view to the study of many application areas, and it will be fruitful to identify new applications, both within and beyond the motivating areas of probability and other uncertainty formalisms, and graph theory. In particular, in further work we plan to investigate the existence, nature and properties of conditional products for a number of special problem areas of natural interest, including:

1. *Possibility theory* and Spohn's theory of κ-*functions*, as mentioned in Shenoy (1994).
2. *Meta Markov models* and *hyper Markov models* (Dawid and Lauritzen 1993). These are formed by suitably combining the concepts of probabilistic (§3) and set-theoretic (§4) conditional product. The combination process involved can be abstracted to apply to conditional products more generally.
3. *Dempster-Shafer belief functions* (Shafer 1976). Again, suitable definitions of conditional product and conditional independence here will involve combining probabilistic and set-theoretic concepts.
4. Semi-graphoids induced by *imsets* (supermodular integer-valued set functions on subsets of N) (Studený 1994/5).
5. Completely general *semi-graphoids* over N. Is there any reasonable definition of conditional product for semi-graphoids? Or perhaps for interesting subclasses, such as those arising from probabilistic models?

A PROOF OF PROPOSITION 3.2

T1 and T2 are immediate from the definition. For T3, take P to have domain E, and $Q = P^F$ where, without loss of generality, $F \subseteq E$. In (3), $\mathsf{V} \in \mathcal{X}_\emptyset = \{\emptyset, \{\bullet\}\}$, and we have $Q(\emptyset \,|\, u) = 0$ a.s.$[Q]$, $Q(\{\bullet\} \,|\, u) = 1$ a.s. $[Q]$. We can regard R as defined on a subclass of $\mathcal{X}_E$ (on identifying $\mathsf{T} \times \mathsf{U} \times \{\bullet\}$ with $\mathsf{T} \times \mathsf{U}$), and then $R(\mathsf{T} \times \mathsf{U}) = \int_{u \in \mathsf{U}} P(\mathsf{T}|u) dP^F(u)$, with $\mathsf{U} \in \mathcal{X}_F$, $= P(\mathsf{T} \times \mathsf{U})$. So clearly $(P, P^F) \in \boldsymbol{\mathcal{D}}$ and and $P \otimes P^F = P$, *i.e.* T3 holds.

Now suppose $(P, Q) \in \boldsymbol{\mathcal{D}}$, with $d(P) = A \cup C$, $d(Q) = B \cup C \cup D$, where A, B, C, D are pairwise disjoint. Let $S = P \otimes Q$, so that $S^{A \cup C} = P$, $S^{B \cup C \cup D} = Q$. Then $A \perp\!\!\!\perp (B \cup D) \,|\, C\,[S]$. Using (4) $\Rightarrow$ (6), we thus have $S(\mathsf{T} \,|\, u, w, z) = S(\mathsf{T} \,|\, u)$ a.s. $[S^{B \cup C \cup D}]$, with $u \in \mathbf{X}_C$, $(w, z) \in \mathbf{X}_{B \cup D} = \mathbf{X}_B \times \mathbf{X}_D$. Equivalently, $S(\mathsf{T} \,|\, u, w, z)$ is (more precisely, has a representative that is) $\mathcal{X}_C$-measurable. It readily follows that $S(\mathsf{T} \,|\, u, z)$ is $\mathcal{X}_C$-measurable, so that $A \perp\!\!\!\perp D \,|\, C\,[S^{A \cup C \cup D}]$, *i.e.* $(P \otimes Q)^{A \cup C \cup D} = P \otimes Q^{C \cup D}$, *viz.* T4$''$ holds; and that $S(\mathsf{T} \,|\, u, w, z)$ is $\mathcal{X}_{C \cup D}$-measurable, so $A \perp\!\!\!\perp B \,|\, (C \cup D)\,[S]$, *i.e.* $P \otimes Q = (P \otimes Q)^{A \cup C \cup D} \otimes Q$, *viz.* T5$''$ holds.

For verification of T6$''$ take P, Q with $d(P) = A \cup C$, $d(Q) = B \cup C \cup D$, where A, B, C, D are pairwise disjoint. Let $R = P \otimes Q^{C \cup D}$, $S = R \otimes Q$, so that $S^{A \cup C \cup D} = R$. Then $A \perp\!\!\!\perp B \,|\, (C \cup D)\,[S]$ and using (4) $\Rightarrow$ (6) we have that $S(\mathsf{T} \,|\, u, w, z) = S(\mathsf{T} \,|\, u, z)$ a.s. $[S]$, with $u \in \mathbf{X}_C$, $w \in \mathbf{X}_B$, $z \in \mathbf{X}_D$. However $A \perp\!\!\!\perp D \,|\, C\,[R]$ implies that there exists a $\mathcal{X}_C$-measurable representative of $R(\mathsf{T} \,|\, u, z)$ and thus a $\mathcal{X}_C$-measurable representative of $S(\mathsf{T} \,|\, u, z)$. Hence, there exists a $\mathcal{X}_C$-measurable representative of $S(\mathsf{T} \,|\, u, w, z)$, which is equivalent to $A \perp\!\!\!\perp (B \cup D) \,|\, C\,[S]$. Since $S^{A \cup C} = P$, $S^{B \cup C \cup D} = Q$, this implies $S = P \otimes Q$.

B TWO EXAMPLES (Studený 1987)

To give examples that $\boldsymbol{\mathcal{C}} \neq \boldsymbol{\mathcal{C}}_*$ and $\boldsymbol{\mathcal{C}}_* \neq \boldsymbol{\mathcal{D}}$ in the framework described in §3.2 we utilize the following lemma.

Lemma B.1 Let $(\mathbf{T}, \mathcal{A})$ be a measurable space, and $\mathcal{B} \subseteq \mathcal{A}$ a sub-σ-algebra such that the diagonal $\{(\mathbf{t}, \mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$ is $\mathcal{A} \times \mathcal{B}$-measurable. Let us put $(\mathbf{X}_1, \mathcal{X}_1) = (\mathbf{X}_3, \mathcal{X}_3) = (\mathbf{T}, \mathcal{A})$, $(\mathbf{X}_2, \mathcal{X}_2) = (\mathbf{T}, \mathcal{B})$.

(i) Every probability measure P on $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ whose marginals on $\mathbf{X}_1 \times \mathbf{X}_2$ and $\mathbf{X}_2 \times \mathbf{X}_3$ are concentrated on their diagonals is concentrated on the diagonal $\mathsf{D} = \{(\mathbf{t}, \mathbf{t}, \mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$.

(ii) Every probability measure π on $(\mathbf{T}, \mathcal{A})$ induces, by the formula

$$P(\mathsf{A} \times \mathsf{B} \times \mathsf{C}) = \pi(\mathsf{A} \cap \mathsf{B} \cap \mathsf{C})$$

for $\mathsf{A} \in \mathcal{X}_1$, $\mathsf{B} \in \mathcal{X}_2$, $\mathsf{A} \in \mathcal{X}_3$, a probability measure P on $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ concentrated on the diagonal.

(iii) Conversely, every such measure can be introduced in this way.

(iv) The marginal of P on $\mathbf{X}_1$ and on $\mathbf{X}_3$ is π, the marginal of P on $\mathbf{X}_2$ is the restriction of π to $\mathcal{B}$.

Proof: For (i), realize that both $\mathsf{D}_1 = \{(\mathbf{t}, \mathbf{t}, \mathbf{v}) : \mathbf{t}, \mathbf{v} \in \mathbf{T}\}$ and $\mathsf{D}_2 = \{(\mathbf{v}, \mathbf{t}, \mathbf{t}) : \mathbf{t}, \mathbf{v} \in \mathbf{T}\}$ belong to $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$. Of course, $\mathsf{D} = \mathsf{D}_1 \cap \mathsf{D}_2$ and the assumption $P(\mathsf{D}_1) = P(\mathsf{D}_2) = 1$ imply $P(\mathsf{D}) = 1$. Consider the mapping $\mathbf{t} \mapsto (\mathbf{t}, \mathbf{t}, \mathbf{t})$ from $\mathbf{T}$ into $\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3$. The reader can easily verify that it is a measurable one-to-one transformation of $(\mathbf{T}, \mathcal{A})$ into $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ (realize that the inverse image of $\mathsf{A} \times \mathsf{B} \times \mathsf{C}$ is $\mathsf{A} \cap \mathsf{B} \cap \mathsf{C}$) whose inverse transformation is measurable as well (the image of A can be written as $(\mathsf{A} \times \mathbf{T} \times \mathbf{T}) \cap \mathsf{D}$). Since the image of $\mathbf{T}$ is D, probability measures on $(\mathbf{T}, \mathcal{A})$ are transformed to probability measures on $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ concentrated on D. Thus, both (ii) and (iii) are evident; (iv) follows from the formula in (ii). $\square$

Let us put $\mathbf{T} = \langle 0, 1 \rangle$ and consider the σ-algebra $\mathcal{B}$ of Borel subsets of $\mathbf{T}$. Let us denote by λ the Lebesgue measure on $(\mathbf{T}, \mathcal{B})$. It was shown in Halmos (1974) (Theorem E in §16) that there exists a set $\mathsf{M} \subset \mathbf{T}$ such that $\lambda_*(\mathsf{M}) = \lambda_*(\mathbf{T} \setminus \mathsf{M}) = 0$. Here λ_* denotes Lebesgue inner measure defined by the formula:

$$\lambda_*(\mathsf{M}) = \sup\{\, \lambda(\mathsf{K}) : \mathsf{K} \subseteq \mathsf{M},\ \mathsf{K} \in \mathcal{B} \,\}.$$

Put $\mathcal{A} = \{ (\mathsf{E} \cap \mathsf{M}) \cup (\mathsf{F} \setminus \mathsf{M}) \,:\, \mathsf{E}, \mathsf{F} \in \mathcal{B} \}$: it is the σ-algebra generated by $\mathcal{B} \cup \{\mathsf{M}\}$. Observe that whenever $(\mathsf{E}\cap\mathsf{M})\cup(\mathsf{F}\setminus\mathsf{M}) = (\tilde{\mathsf{E}}\cap\mathsf{M})\cup(\tilde{\mathsf{F}}\setminus\mathsf{M})$ for $\mathsf{E}, \tilde{\mathsf{E}}, \mathsf{F}, \tilde{\mathsf{F}} \in \mathcal{B}$ then $\mathsf{E}\Delta\tilde{\mathsf{E}} \subseteq \mathbf{T} \setminus \mathsf{M}$ and $\mathsf{F}\Delta\tilde{\mathsf{F}} \subseteq \mathsf{M}$ and therefore $\lambda(\mathsf{E}\Delta\tilde{\mathsf{E}}) = \lambda(\mathsf{F}\Delta\tilde{\mathsf{F}}) = 0$. In particular, for every $\alpha \in \langle 0, 1\rangle$ the set function on $\mathcal{A}$ defined by:

$$\pi_\alpha(\mathsf{G}) \,=\, \alpha \cdot \lambda(\mathsf{E}) + (1 - \alpha) \cdot \lambda(\mathsf{F}), \qquad (8)$$

where $\mathsf{G} = (\mathsf{E} \cap \mathsf{M}) \cup (\mathsf{F} \setminus \mathsf{M})$, $\mathsf{E}, \mathsf{F} \in \mathcal{B}$, is well-defined. The reader can show by the procedure in the proof of Theorem A, §17 of Halmos (1974) that π_α is a probability measure on $(\mathbf{T}, \mathcal{A})$. Evidently, the restriction of π_α to $\mathcal{B}$ is λ for every $\alpha \in \langle 0, 1\rangle$, π_1 is concentrated on M and π_0 is concentrated on $\mathbf{T} \setminus \mathsf{M}$. Put $(\mathbf{X}_1, \mathcal{X}_1) = (\mathbf{T}, \mathcal{A})$, $(\mathbf{X}_2, \mathcal{X}_2) = (\mathbf{T}, \mathcal{B})$, $(\mathbf{X}_3, \mathcal{X}_3) = (\mathbf{T}, \mathcal{A})$ and introduce the measure R_α on $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ for $\alpha \in \langle 0, 1\rangle$ by the formula:

$$R_\alpha(\mathsf{A} \times \mathsf{B} \times \mathsf{C}) = \pi_\alpha(\mathsf{A} \cap \mathsf{B} \cap \mathsf{C}) \qquad (9)$$

for $\mathsf{A} \in \mathcal{X}_1$, $\mathsf{B} \in \mathcal{X}_2$, $\mathsf{A} \in \mathcal{X}_3$. By Lemma B.1(ii) R_α is a probability measure concentrated on the diagonal.

Example B.1 There are measurable spaces $(\mathbf{X}_i, \mathcal{X}_i)$ $i = 1, 2, 3$ and probability distributions P on $(\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{X}_1 \times \mathcal{X}_2)$ and Q on $(\mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_2 \times \mathcal{X}_3)$ which are weakly compatible but not strongly compatible.

Repeat the construction above and define P to be the marginal of R_0 on $\mathbf{X}_1 \times \mathbf{X}_2$ and Q to be the marginal of R_1 on $\mathbf{X}_2 \times \mathbf{X}_3$. By Lemma B.1(iv) the marginal of P on $\mathbf{X}_2$ is λ; the same conclusion holds for Q. Thus, P and Q are weakly compatible. Suppose that T is a distribution on $(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ having P and Q as marginals. Then T has on $\mathbf{X}_1$ the same marginal as P, and therefore as R_0, that is π_0 (use Lemma B.1(iv)). Similarly, T has on $\mathbf{X}_3$ the same marginal as Q, that is π_1. However, by Lemma B.1(i) T is concentrated on the diagonal and by Lemma B.1(iv) the marginals of T on $\mathbf{X}_1 = \mathbf{X}_3$ coincide. This leads to the contradictory conclusion $\pi_0 = \pi_1$. Therefore P and Q are not strongly compatible.

Example B.2 There are measurable spaces $(\mathbf{X}_i, \mathcal{X}_i)$ $i = 1, 2, 3$ and probability distributions P on $(\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{X}_1 \times \mathcal{X}_2)$ and Q on $(\mathbf{X}_2 \times \mathbf{X}_3, \mathcal{X}_2 \times \mathcal{X}_3)$ which are strongly compatible but their conditional product from §3.2 is not defined.

Repeat the construction before Example B.1 and put $R = R_\alpha$ for $\alpha = \frac{1}{2}$. Define P and Q to be the respective marginals of R. They are evidently strongly compatible. Suppose for contradiction that T is the conditional product of P and Q as defined in §3.2. By Lemma B.1(i) T is concentrated on the diagonal. Moreover, by Lemma B.1(iv) the marginal of T on $\mathbf{X}_1$ is $\pi_{\frac{1}{2}}$. Thus, using Lemma B.1(iii), we derive $T = R$. The reader can observe that for $\mathsf{G} \in \mathcal{X}_1 = \mathcal{A}$ of the form $\mathsf{G} = (\mathsf{E}\cap\mathsf{M}) \cup (\mathsf{F} \setminus \mathsf{M})$ where $\mathsf{E}, \mathsf{F} \in \mathcal{B}$ the formula

$$P(\mathsf{G}|u) \,=\, \frac{1}{2} \cdot \chi_{\mathsf{E}}(u) + \frac{1}{2} \cdot \chi_{\mathsf{F}}(u)$$

gives a representative of the conditional probability of G given $\mathcal{X}_2 = \mathcal{B}$ induced by P (verify (2) using (9) and (8) with $\alpha = \frac{1}{2}$). A similar conclusion holds for Q. Put $\mathsf{T} = (\langle 0, \frac{1}{2}) \cap \mathsf{M}) \cup (\langle \frac{1}{2}, 1\rangle \setminus \mathsf{M}) \in \mathcal{X}_1$, $\mathsf{U} = \langle 0, 1\rangle \in \mathcal{X}_2$, $\mathsf{V} = (\langle \frac{1}{2}, 1\rangle \cap \mathsf{M}) \cup (\langle 0, \frac{1}{2}) \setminus \mathsf{M}) \in \mathcal{X}_3$. Then $p(\mathsf{T}|\cdot) = \frac{1}{2}$ and $q(\mathsf{V}|\cdot) = \frac{1}{2}$ by the formula above and (3) implies $R(\mathsf{T} \times \mathsf{U} \times \mathsf{V}) = T(\mathsf{T} \times \mathsf{U} \times \mathsf{V}) = \frac{1}{4}$. However, since $\mathsf{T} \cap \mathsf{V} = \emptyset$, (9) gives a contradictory conclusion $R(\mathsf{T} \times \mathsf{U} \times \mathsf{V}) = 0$. Therefore P and Q have no conditional product.

Acknowledgments

Milan Studený's work was supported by grants GAČR n. 201/98/0478, MŠMT n. 96008, and GAAVČR n. A1075801.

References

G. Birkhoff (1949). *Lattice Theory*. AMS Colloquium Publications **25**.

A. P. Dawid (1979). Conditional independence in statistical theory (with Discussion). *J. Roy. Statist. Soc. B* **41**, 1–31.

A. P. Dawid (1998). Conditional independence. To appear in *Encyclopedia of Statistical Sciences (Update Volume 3)*, Wiley.

A. P. Dawid and S. L. Lauritzen (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21**, 1272–1317.

D. Geiger and J. Pearl (1990). On the logic of causal models. In *Uncertainty in Artificial Intelligence* **4** (R. D. Shachter, T. S. Lewitt, L. N. Kanal, J. F. Lemmer, eds.), North-Holland, 3–14.

D. Geiger and J. Pearl (1993). Logical and algorithmic properties of conditional independence and graphical models. *Ann. Statist.* **21**, 2001–2021.

P. R. Halmos (1974). *Measure Theory*. Graduate Texts in Mathematics **18**, Springer.

S. L. Lauritzen, A. P. Dawid, B. N. Larsen and H.-G. Leimer (1990). Independence properties of directed Markov fields. *Networks* **20**, 491-505.

J. Neveu (1964). *Bases Mathematiques du Calcul des Probabilités*. Masson et Cie.

J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

G. Shafer (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

P. P. Shenoy and G. Shafer (1990). Axioms for probability and belief propagation. In *Uncertainty in Artificial Intelligence* **4** (R. D. Shachter, T. S. Lewitt, L. N. Kanal, J. F. Lemmer, eds.), North-Holland, 169–198.

P. P. Shenoy (1994). Conditional independence in valuation-based systems. *Internat. J. Approx. Reasoning* **10**, 203–234.

W. Spohn (1980). Stochastic independence, causal independence and shieldability. *J. Philos. Logic* **9**, 73–99.

M. Studený (1987). The notion of multiinformation in probabilistic decision making. Thesis (in Czech), Institute of Information Theory and Automation, Prague.

M. Studený (1993). Formal properties of conditional independence in different calculi of AI. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty* (M. Clarke, R. Kruse, S. Moral eds.), Springer, 241–348.

M. Studený (1994/5). Description of structures of stochastic conditional independence by means of faces and imsets (a series of 3 papers). *Internat. J. General Systems* **23**, 123–137, 201–219, 323–341.

M. Studený and R. R. Bouckaert (1998/9). On chain graph models for description of conditional independence structures. To appear in *Ann. Statist.*