



Logical and Algorithmic Properties of Conditional Independence and Graphical Models

Author(s): Dan Geiger and Judea Pearl

Reviewed work(s):

Source: *The Annals of Statistics*, Vol. 21, No. 4 (Dec., 1993), pp. 2001-2021

Published by: [Institute of Mathematical Statistics](#)

Stable URL: <http://www.jstor.org/stable/2242326>

Accessed: 09/11/2011 03:49

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Statistics*.

<http://www.jstor.org>

LOGICAL AND ALGORITHMIC PROPERTIES OF CONDITIONAL INDEPENDENCE AND GRAPHICAL MODELS¹

BY DAN GEIGER AND JUDEA PEARL

*Technion–Israel Institute of Technology and University of California,
Los Angeles*

This article develops an axiomatic basis for the relationship between conditional independence and graphical models in statistical analysis. In particular, the following relationships are established: (1) every axiom for conditional independence is an axiom for graph separation, (2) every graph represents a consistent set of independence and dependence constraints, (3) all binary factorizations of strictly positive probability models can be encoded and determined in polynomial time using their correspondence to graph separation, (4) binary factorizations of non-strictly positive probability models can also be derived in polynomial time albeit less efficiently and (5) unconditional independence relative to normal models can be axiomatized with a finite set of axioms.

1. Introduction. A useful approach to multivariate statistical modeling is to first define the conditional independence constraints that are likely to hold in the domain, and then to restrict the analysis to probability functions that satisfy those constraints. An increasingly popular way of specifying independence constraints are graphical models, such as Markov networks and Bayesian networks, where the constraints are encoded through the topological properties of the corresponding graphs [Lauritzen (1982), Lauritzen and Spiegelhalter (1988), Pearl (1988) and Whittaker (1990)].

The key idea behind these specification schemes is to utilize the correspondence between *separation* in graphs and *conditional independence* in probability; each node represents a variable and each missing edge encodes some conditional independence constraint. More specifically, if a set of nodes Z blocks all the paths between two nodes, then the corresponding two variables are asserted to be conditionally independent given the variables corresponding to Z .

The notions of graph separation and conditional independence, which at first glance seem to have little in common, share key properties which render graphs an effective language of specifying independence constraints. This

Received July 1989; revised November 1992.

¹Supported in part by NSF Grant IRI-8821444 while the first author was at UCLA. The revised version was prepared while the first author was at Northrop Research and Technology Center and completed at Technion.

AMS 1991 subject classifications. Primary 60A05, 60J99, 60G60; secondary 62A15, 62H25.

Key words and phrases. Conditional independence, Markov fields, Markov networks, graphical models.

article develops an axiomatic characterization of these properties, thus providing a theoretical basis for understanding the role of graphical models in statistical analysis.

The article is organized as follows. Section 2 provides preliminary definitions. Section 3 proves the existence of perfect probability models, that is, probability models that, given an arbitrary list of conditional independence statements, satisfy every statement on that list, every statement that logically follows from that list and none other. Using this result, Section 4 then shows that every axiom for conditional independence is an axiom for graph separation and that every graph represents a consistent set of independence and dependence constraints. In other words, graphs provide a “safe” language for encoding statistical associations; the set of conditional independencies and dependencies encoded by any graph is guaranteed to be realizable in some probability model.

Section 5 deals with special kinds of conditional independence relationships, those that permit the factorization of a probability model into a product of two functions. It is shown that graphs provide a parsimonious code (requiring polynomial space) for representing the entire set of binary factorizations that are realizable in strictly positive probability models. Graphs also facilitate a polynomial time algorithm for determining whether an arbitrary binary factorization logically follows from a given set of such factorizations.

The rest of the article provides a complete axiomatic characterization for special families of independence relationships. We first develop complete axiomatizations for saturated independence (Section 6) and marginal independence (Section 7) and then address the axiomatization of conditional independence in general (Section 8). Section 9 generalizes several results to qualitative independence, and Section 10 provides a tabulated summary of our results.

2. Preliminaries. Throughout this article, let U be a finite set of distinct symbols $\{u_1, \dots, u_n\}$, called *attributes* (or *variable names*). A *domain mapping* is a mapping that associates a set, $d(u_i)$, with each attribute u_i . This set is called the *domain* of u_i and each of its elements is a *value* for u_i . An attribute combined with a domain is a *variable*. For example, the variable describing the age of a person will be characterized by the attribute *age* and may be assigned a domain such as $\{i \mid 0 \leq i \leq 120\}$ or $\{\text{infant}, \text{child}, \text{young adult}, \text{other adult}\}$. The distinction between attributes and variables allows us to associate several domains with the same variable name, as done in some of the following.

DEFINITION. A *probability model* over a finite set of attributes $U = \{u_1, \dots, u_n\}$ is a pair (d, P) , where d is a domain mapping that maps each u_i to a finite domain $d(u_i)$, and $P: d(u_1) \times \dots \times d(u_n) \rightarrow [0, 1]$ is a probability distribution having the Cartesian product of these domains as its sample space. The class of probability models over U is denoted by \mathcal{P} .

Unless stated otherwise, U and its domain are assumed to be finite.

DEFINITION. The expression $I(X, Y|Z)$ where X , Y and Z are disjoint subsets of U is called an *independence statement*. Its negation $\neg I(X, Y|Z)$ is called a *dependence statement*. An independence or dependence statement is defined over $V \subseteq U$ if it mentions only attributes in V .

DEFINITION. Let (d, P) be a probability model over U . An independence statement $I(X, Y|Z)$ is said to *hold for* (d, P) if for every value \mathbf{X} , \mathbf{Y} and \mathbf{Z} of X , Y and Z , respectively,

$$(1) \quad P(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \cdot P(\mathbf{Z}) = P(\mathbf{X}, \mathbf{Z}) \cdot P(\mathbf{Y}, \mathbf{Z}).$$

Equivalently, (d, P) is said to *satisfy* $I(X, Y|Z)$. Otherwise, (d, P) is said to *satisfy* $\neg I(X, Y|Z)$.

DEFINITION. When $I(X, Y|Z)$ holds for (d, P) , then X and Y are *conditionally independent* relative to (d, P) , and if $Z = \emptyset$, then X and Y are *marginally independent* relative to (d, P) .

DEFINITION. A probability model over U is *strictly positive* if every combination of U 's values has a probability greater than 0. The class of strictly positive probability models is denoted by \mathcal{P}^+ .

DEFINITION. A probability model over U is *binary* if it assigns every attribute in U a domain with only two values, say 0 and 1. The class of binary probability models is denoted by \mathcal{B} .

Equations (2) through (6) list some properties of conditional independence. Variants of them were first introduced by Dawid (1979) and further studied by Spohn (1980), Pearl and Paz (1985), Pearl (1988) and Geiger (1990).

Trivial independence:

$$(2) \quad I(X, \emptyset|Y).$$

Symmetry:

$$(3) \quad I(X, Y|Z) \Rightarrow I(Y, X|Z).$$

Decomposition:

$$(4) \quad I(X, Y \cup W|Z) \Rightarrow I(X, Y|Z).$$

Weak contraction [the axiomatic theory of Pearl and Paz (1985) invoked a stronger version of this axiom which is not needed in the discussion of this article]:

$$(5) \quad I(X \cup W, Y|Z), I(X, W|Z \cup Y) \Rightarrow I(X, Y \cup W|Z).$$

Weak union:

$$(6) \quad I(X, Y \cup W|Z) \Rightarrow I(X, Y|Z \cup W).$$

DEFINITION. An *independence Horn clause* is an implication of the form

$$I(X_1, Y_1|Z_1), I(X_2, Y_2|Z_2), \dots I(X_k, Y_k|Z_k) \Rightarrow I(X_{k+1}, Y_{k+1}|Z_{k+1}).$$

Each independence statement on the left of the implication is called an *antecedent* and the one on the right is called the *consequence*. Independence Horn clauses may also have no consequence [as in (2)].

DEFINITION. An independence Horn clause is *instantiated* if each of the X_i 's, Y_i 's and Z_i 's is substituted with a specific subset of U [e.g., $I(\{u_1, u_2\}, \emptyset|\{u_3, u_4\})$ is an instance of trivial independence].

We use σ , possibly subscripted, to denote an independence statement, $\neg \sigma$ to denote the negation of σ , Σ to denote a set of independence statements and \mathcal{F} to denote a subset of \mathcal{P} (i.e., a class of probability models over U such as \mathcal{B} or \mathcal{P}^+).

DEFINITION. An independence Horn clause is *sound* relative to \mathcal{F} iff for every instantiation of the clause, every probability model in \mathcal{F} that satisfies the clause's antecedents also satisfies its consequence.

DEFINITION. When an independence Horn clause is sound relative to \mathcal{F} , it is called an *axiom* relative to \mathcal{F} . An axiom relative to \mathcal{P} is simply called an *axiom*.

For example, (7) is an axiom relative to \mathcal{P}^+ but not relative to \mathcal{P} .

Intersection:

$$(7) \quad I(X, Y|Z \cup W), I(X, W|Z \cup Y) \Rightarrow I(X, Y \cup W|Z).$$

DEFINITION. Given a set of axioms \mathcal{A} , an independence statement σ is *derivable* from a set of statements Σ , denoted $\Sigma \vdash \sigma$, if there exists a *derivation sequence* $\sigma_1, \dots, \sigma_n$ such that $\sigma_n = \sigma$ and for each σ_j , either (1) $\sigma_j \in \Sigma$ or (2) σ_j is the consequence of some instantiated axiom in \mathcal{A} for which every antecedent is in $\{\sigma_1, \dots, \sigma_{j-1}\}$. The *closure* of Σ is the set of derivable statements, $\{\sigma|\Sigma \vdash \sigma\}$, and is denoted by Σ^+ .

For example, $I(u_1, u_3|\emptyset)$ is derivable from the set $\{I(\{u_1u_3\}, u_2|\emptyset), I(u_1, u_3|u_2)\}$ using axioms (2) through (6) via the derivation sequence $I(\{u_1u_3\}, u_2|\emptyset), I(u_1, u_3|u_2), I(u_1, \{u_2, u_3\}|\emptyset), I(u_1, u_3|\emptyset)$. The third and fourth statements in this sequence are derived from the previous ones by weak contraction and decomposition, respectively. [For simplicity, throughout, $I(u_i, u_j|u_k)$ stands for $I(\{u_i\}, \{u_j\}|\{u_k\})$].

DEFINITION. An independence statement σ is *entailed* by a set of statements Σ relative to a set of probability models \mathcal{F} , denoted $\Sigma \models \sigma$, if every probability model in \mathcal{F} that satisfies Σ satisfies σ as well. The set of entailed statements, $\{\sigma|\Sigma \models \sigma\}$, is denoted by Σ^* , keeping \mathcal{F} implicit.

PROPOSITION 1. *Let \mathcal{A} be a set of axioms relative to \mathcal{F} . For every set Σ of independence statements, we have $\Sigma^+ \subseteq \Sigma^*$, where Σ^+ is derived from Σ using the axioms in \mathcal{A} , and Σ^* is entailed relative to \mathcal{F} .*

PROOF. The proof follows by induction on the length of a derivation sequence of each σ in Σ^+ , using the fact that the axioms in \mathcal{A} are sound relative to \mathcal{F} . \square

Equality of Σ^+ and Σ^* holds only if no axioms are “missing.”

DEFINITION. A set of axioms \mathcal{A} is *complete* (relative to \mathcal{F}) if for every set Σ of independence statements, $\Sigma^* = \Sigma^+$.

PROPOSITION 2. *A set of axioms \mathcal{A} is complete (relative to \mathcal{F}) if and only if for every set of statements Σ and every statement $\sigma \notin \Sigma^+$ there exists a probability model (d_σ, P_σ) in \mathcal{F} that satisfies Σ and does not satisfy σ .*

PROOF. The proof follows immediately from the definition of completeness and Proposition 1. \square

Next, we seek conditions under which, for every set Σ of independence statements, there exists a probability model in a given class \mathcal{F} that satisfies precisely the statements in Σ^* and none other. Fagin (1982) spelled out such conditions and showed, in the context of database theory, that they imply the existence of an operator \otimes that maps a set of probability models to a probability model, such that an independence statement holds in the latter if and only if it holds in every constituent of the former. The next section constructs such an operator.

3. Perfect probability models. The main result of this section is that, for any given set Σ of independence statements, there exists a probability model in \mathcal{P} that satisfies precisely Σ^* and no other statements. (Fagin called models with this property “Armstrong models.”) An immediate application of it, as we shall see, lies in determining whether a given set of independence and dependence statements is consistent.

DEFINITION. Let Σ be a set of independence statements. A probability model is *perfect* for Σ (relative to \mathcal{F}) if it satisfies precisely the set of statements Σ^* entailed by Σ (relative to \mathcal{F}) and none other.

The key idea in showing the existence of perfect probability models rests with the notion of *direct product* defined below, which extends Fagin’s definition (1982) from database relations to probability models.

DEFINITION. The (*binary*) *direct product* for \mathcal{F} is a mapping, $\otimes : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$, where \mathcal{F} is a class of probability models over a finite set of attributes

$\{u_1, \dots, u_n\}$, and $(d, P) = (d_1, P_1) \otimes (d_2, P_2)$ is defined as follows: Let $d_1(u_i)$ and $d_2(u_i)$ be the domains associated with u_i in (d_1, P_1) and in (d_2, P_2) , respectively. Let a_i and b_i be values drawn respectively from these domains. Set the domain of u_i in (d, P) to be the Cartesian product $d_1(u_i) \times d_2(u_i)$, and let

$$(8) \quad P(a_1 b_1, a_2 b_2, \dots, a_n b_n) = P_1(a_1, a_2, \dots, a_n) \cdot P_2(b_1, b_2, \dots, b_n),$$

where $a_i b_i$ denotes a value of u_i in (d, P) .

A notable property of \otimes is the assignment of a new domain, $d_1(u_i) \times d_2(u_i)$, to each u_i . Thus u_i is treated as an attribute rather than a variable with a fixed domain. We will show at the end of this section that if the domain of each attribute is fixed, then the existence of perfect models is not guaranteed.

The next lemma shows that the product form of (8) remains valid after marginalization.

LEMMA 3. *Let (d_1, P_1) , (d_2, P_2) and (d, P) be probability models over U as in (8). Then, for every subset u_{i_1}, \dots, u_{i_l} of U ,*

$$(9) \quad P(a_{i_1} b_{i_1}, a_{i_2} b_{i_2}, \dots, a_{i_l} b_{i_l}) = P_1(a_{i_1}, a_{i_2}, \dots, a_{i_l}) \cdot P_2(b_{i_1}, b_{i_2}, \dots, b_{i_l}).$$

PROOF. Assume without loss of generality that in (9), $i_1 = 1, i_2 = 2, \dots, i_l = l$. (otherwise reorder u_1, \dots, u_n to meet this assumption.) When $l = n$ this equation is identical to (8). We proceed by descending induction. Assume (9) holds for $l = k \leq n$; then

$$\begin{aligned} & P(a_1 b_1, \dots, a_{k-1} b_{k-1}) \\ &= \sum_{x_k} P(a_1 b_1, \dots, a_{k-1} b_{k-1}, x_k) \\ &= \sum_{a_k \in d_1(u_k), b_k \in d_2(u_k)} P_1(a_1, \dots, a_{k-1}, a_k) \cdot P_2(b_1, \dots, b_{k-1}, b_k) \\ &= \left\{ \sum_{a_k \in d_1(u_k)} P_1(a_1, \dots, a_{k-1}, a_k) \right\} \cdot \left\{ \sum_{b_k \in d_2(u_k)} P_2(b_1, \dots, b_{k-1}, b_k) \right\} \\ &= P_1(a_1, \dots, a_{k-1}) \cdot P_2(b_1, \dots, b_{k-1}). \quad \square \end{aligned}$$

The key property of \otimes is given in the following lemma.

LEMMA 4. *Let (d_1, P_1) , (d_2, P_2) and (d, P) be probability models over U as in (8). Then, for any three disjoint subsets X, Y and Z of U ,*

$$(10) \quad \begin{aligned} & I(X, Y|Z) \text{ holds for } (d, P) \text{ iff } I(X, Y|Z) \\ & \text{holds for } (d_1, P_1) \text{ and for } (d_2, P_2). \end{aligned}$$

PROOF. Let a_x, a_y, a_z be respective values of X, Y, Z in (d_1, P_1) and b_x, b_y, b_z be respective values of X, Y, Z in (d_2, P_2) .

The *if* part of (10) follows from

$$\begin{aligned}
 & P(a_z b_z) P(a_x b_x, a_y b_y, a_z b_z) \\
 &= P_1(a_x, a_y, a_z) \cdot P_1(a_z) \cdot P_2(b_x, b_y, b_z) \cdot P_2(b_z) \\
 &= P_1(a_x, a_z) \cdot P_1(a_y, a_z) \cdot P_2(b_x, b_z) \cdot P_2(b_y, b_z) \\
 &= P(a_x b_x, a_z b_z) \cdot P(a_y b_y, a_z b_z).
 \end{aligned}$$

(Note the implicit use of Lemma 3.)

The *only if* part of (10) follows from

$$\begin{aligned}
 & P_1(a_x, a_y, a_z) \cdot P_1(a_z) \cdot P_2(b_x, b_y, b_z) \cdot P_2(b_z) \\
 &= P(a_x b_x, a_y b_y, a_z b_z) \cdot P(a_z b_z) \\
 &= P(a_x b_x, a_z b_z) \cdot P(a_y b_y, a_z b_z) \\
 &= P_1(a_x, a_z) \cdot P_1(a_y, a_z) \cdot P_2(b_x, b_z) \cdot P_2(b_y, b_z).
 \end{aligned}$$

By summing once over a_x and once over b_x , it is evident that $I(X, Y|Z)$ holds for (d_1, P_1) and for (d_2, P_2) . \square

Next, we extend the direct product to be a mapping from families of probability models (rather than pairs) into probability models.

THEOREM 5. *There exists an operator \otimes that any nonempty finite family $\{(d_i, P_i) | i = 1, \dots, n\}$ of probability models over a set of attributes U into a probability model over U , such that if σ is an independence statement, then σ holds for $\otimes \{(d_i, P_i) | i = 1, \dots, n\}$ if and only if σ holds for each (d_i, P_i) .*

PROOF. Since the binary direct product is commutative and associative, it can be extended to sets as follows:

$$\otimes \{(d_i, P_i) | i = 1, \dots, n\} = (((((d_1, P_1) \otimes (d_2, P_2)) \otimes (d_3, P_3)) \otimes \dots (d_n, P_n)).$$

Due to Lemma 4,

$$\sigma \text{ holds for } \otimes \{(d_i, P_i) | i = 1, \dots, n\} \text{ iff } \sigma \text{ hold for every } (d_i, P_i),$$

as stated by the theorem. \square

Consequently, the existence of perfect probability models can be established [similar to (Fagin 1982)].

COROLLARY 6. *For every set of independence statements Σ over the attributes of U , there exists a probability model (d, P) in \mathcal{P} such that (d, P) satisfies every statement in Σ^* and none other, that is, (d, P) is a perfect model relative to \mathcal{P} .*

PROOF. Let (d, P) be $\otimes\{(d_\sigma, P_\sigma) \mid \sigma \notin \Sigma^*\}$, where (d_σ, P_σ) is a probability model that satisfies Σ^* but does not satisfy σ . By the definition of Σ^* , a probability model (d_σ, P_σ) always exists except for the degenerated case where Σ^* renders all variables mutually independent, in which case Corollary 6 holds trivially. (Also note that the set $\{\sigma \mid \sigma \notin \Sigma^*\}$ is finite because U is finite.) Due to Theorem 5, (d, P) satisfies the statements in Σ^* and none other because these are the only statements that hold for every (d_σ, P_σ) . \square

The probability model $\otimes\{(d_i, P_i) \mid i = 1, \dots, n\}$ is strictly positive whenever each (d_i, P_i) is strictly positive. Consequently, we obtain the following result.

COROLLARY 7. *For every set of independence statements Σ , there exists a strictly positive probability model (d, P) such that (d, P) satisfies every statement in Σ^* (relative to \mathcal{P}^+) and none other, that is, (d, P) is a perfect model relative to \mathcal{P}^+ .*

The existence of a perfect model implies that any algorithm that determines whether a given statement is entailed by Σ can also determine whether a disjunction of statements is entailed by Σ . For example, to show that

$$(11) \quad \{I(u_1, u_2 \mid \emptyset), I(u_1, u_2 \mid u_3)\} \not\models I(u_1, u_3 \mid \emptyset) \vee I(u_2, u_3 \mid \emptyset),$$

we will see that one must merely check that each disjunct is not entailed by itself.

To refute the first disjunct, construct a probability model (d_1, P_1) in which u_1 and u_2 are two independent binary variables and u_1 equals u_3 . This probability model satisfies the antecedents but does not satisfy the first disjunct. To refute the second disjunct, construct a probability model (d_2, P_2) in which u_1 and u_2 are two independent binary variables and u_2 equals u_3 ; it satisfies the antecedents but not the second disjunct. The probability model $(d_1, P_1) \otimes (d_2, P_2)$ satisfies the antecedents but does not satisfy the disjunction. Hence, the disjunction is not entailed by the antecedents.

Notably, if we fix the domain of u_3 to be binary, the antecedents of (11) *do* entail the disjunctive consequence [Pearl (1988), pp. 129 and 137]; the construction of $(d_1, P_1) \otimes (d_2, P_2)$ fails because \otimes assigns a domain of size 4 to u_3 . Consequently, we obtain the following result.

COROLLARY 8. *There exists a set of independence statements Σ for which no binary probability model is perfect.*

PROOF. Let $\Sigma = \{I(u_1, u_2 \mid \emptyset), I(u_1, u_2 \mid u_3)\}$. Every binary probability model that satisfies Σ satisfies either $I(u_1, u_3 \mid \emptyset)$ or $I(u_2, u_3 \mid \emptyset)$. However, neither statement in itself is entailed by Σ (relative to \mathcal{B}) and therefore none is in Σ^* . \square

Another application of Theorem 5 is facilitating tests for *consistency*.

DEFINITION. A set of independence statements Σ_p and a set of negated independence statements (i.e., *dependence statements*) Σ_n is *consistent* if there

exists a probability model that satisfies $\Sigma_p \cup \Sigma_n$. The task of deciding whether a set of independence and dependence statements is consistent is called the *consistency problem*. The task of determining whether a set of independence statements entails an independence statement is called the *implication problem*.

The following algorithm determines whether or not $\Sigma_p \cup \Sigma_n$ is consistent: For every member $\neg\sigma$ of Σ_n , determine whether $\Sigma_p \models \sigma$. If the answer is negative for all members of Σ_n , then $\Sigma_p \cup \Sigma_n$ is consistent; otherwise it is not consistent [(Geiger, Paz and Pearl (1991)].

This algorithm works when the following two conditions are met: (1) we can efficiently check whether or not $\Sigma \models \sigma$ and (2) entailment is taken with respect to a class of probability models that has perfect models (i.e., \mathcal{P}^+ but not \mathcal{B}). In Section 5 we examine a class of independence statements, called *saturated*, for which these conditions are met.

The correctness of the algorithm stems from the fact that if the negation of each member $\neg\sigma$ of Σ_n is not entailed by Σ_p , that is, each member of Σ_n is individually consistent with Σ_p , then there exists a probability model (d_σ, P_σ) that satisfies Σ_p and does not satisfy σ . The probability model $(d, P) = \otimes \{(d_\sigma, P_\sigma) \mid \neg\sigma \in \Sigma_n\}$ satisfies every statement in $\Sigma_p \cup \Sigma_n$, and therefore the algorithm's decision that the two sets are consistent is correct. In the other direction, namely, when the algorithm detects an inconsistent member of Σ_n , then the decision is obviously correct.

4. Graphs and independence. The use of graphs for representing probability distributions is well documented in the statistical literature [Whittaker (1990) and reference therein]. The basis of these representation schemes is the similarity between *separation* in graphs and *conditional independence* in probability. We will show that these two concepts are related in a stronger sense than was previously known; we will show that every axiom for conditional independence must also be an axiom for graph separation, and that the set of separation–connection conditions embodied in any graph always corresponds to a consistent set of independence–dependence statements in probability.

DEFINITION. An *undirected graph* is a pair (U, E) , where U is a finite set of attributes, called *nodes*, and E is a set of unordered pairs of distinct nodes, called *edges*. When (u_1, u_2) is an edge, u_1 and u_2 are *directly connected*. A *path* between two nodes is a sequence of nodes for which every pair of adjacent nodes is directly connected and no node appears twice.

DEFINITION. Let X, Y and Z be disjoint subsets of nodes in a graph $G = (U, E)$. A *separation* statement $J(X, Y|Z)$ is said to *hold for G* if every path between a node in X and a node in Y includes a node in Z . Equivalently, we say that G *satisfies* $J(X, Y|Z)$ or X and Y are *separated* by Z in G .

Connection (negated separation) statements, separation Horn clauses and separation Horn axioms for a set of graphs are defined analogously to the corresponding concepts of independence defined in Section 2.

It is easy to see that axioms (2) through (7) remain sound when I is replaced with J ; that is, whenever the antecedent of one of these axioms holds in some graph, its consequence holds as well. For example, if X and $Y \cup W$ are separated by Z in some graph G , then X and Y are also separated by $Z \cup W$ as dictated by the weak-union axiom (6). This correspondence between independence and graph separation is not a coincidence; we show next that every axiom of conditional independence is an axiom for separation. The converse does not hold [Pearl (1988)]. A preliminary definition and a lemma are needed.

DEFINITION. Let (d, P) be a probability model over a finite set of attributes U , and let G be a graph whose nodes are the elements of U (i.e., each node is associated with an attribute). Then G is said to be a *Markov network* of (d, P) if for every three disjoint subsets X, Y and Z of U ,

$$J(X, Y|Z) \text{ holds for } G \text{ implies that } I(X, Y|Z) \text{ holds for } (d, P).$$

For example, a language in which the probability of the i th letter is determined solely by the $(i - 1)$ th letter via $P(l_i|l_{i-1})$ can be represented by the Markov network of Figure 1. This graph shows, for example, that l_1 and l_3 are conditionally independent given l_2 , since l_2 separates l_1 and l_3 . Note that this independence statement holds regardless of the domain associated with each l_i (i.e., the alphabet of the language need not be specified). Markov networks are discussed in Darroch, Lauritzen and Speed (1980) and Lauritzen (1982).

A variant of the next lemma was independently derived by Frydenberg (1990).

LEMMA 9. *Let G be an undirected graph with U as its set of nodes. Let X, Y and Z be disjoint subsets of U such that X and Y are not separated by Z . Then there exists a strictly positive probability model (d, P) over a set of attributes U , such that G is a Markov network of (d, P) and $I(X, Y|Z)$ does not hold for (d, P) .*

PROOF. Since X and Y are not separated by Z , there exists a path r_1, r_2, \dots, r_l which contain no nodes of Z and which connects a node r_1 in X to a node r_l in Y . Let every node r_i be associated with a binary variable v_i and

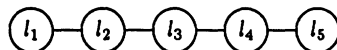


FIG. 1. A five-node chain.

every node not on the path be associated with a binary variable s_i . A probability model (d, P) where

$$P(v_1, \dots, v_l, s_1, \dots) = (1/2) \cdot \prod_{i=1}^{l-1} f(v_i, v_{i+1}) \cdot \prod_i g(s_i),$$

$g(s_i) = 1/2$, and

$$f(v_i, v_{i+1}) = \begin{cases} 1/2, & \text{if } v_i = 0, v_{i+1} = 0, \\ 1/2, & \text{if } v_i = 0, v_{i+1} = 1, \\ 1/4, & \text{if } v_i = 1, v_{i+1} = 0, \\ 3/4, & \text{if } v_i = 1, v_{i+1} = 1, \end{cases}$$

satisfies the requirements; $I(v_1, v_l|Z)$ does not hold and if $J(X', Y'|Z')$ holds, $I(X', Y'|Z')$ holds as well. \square

THEOREM 10. *Every independence Horn clause $\sigma_1, \sigma_2, \dots, \sigma_n \Rightarrow \sigma$ that is an axiom for independence relative to \mathcal{P}^+ is also an axiom for separation, where each σ_i is interpreted as a separation statement.*

PROOF. Suppose by contradiction that there exists a graph that satisfies $\Sigma = \{\sigma_1, \dots, \sigma_n\}$ and does not satisfy σ . Then by Lemma 9 there exists a strictly positive probability model that satisfies Σ and does not satisfy σ . Thus $\sigma_1, \sigma_2, \dots, \sigma_n \Rightarrow \sigma$ is not sound relative to \mathcal{P}^+ . \square

Consequently, in particular, axioms (2) through (7) as well as those discussed by Studeny (1992) are axioms for separation. A complete list of axioms for separation was found by Pearl and Paz (1985).

Each graph can be thought of as a specification language for independence and dependence statements; whenever a separation condition holds in the graph, the corresponding independence statement is asserted, and whenever a connection condition holds in the graph, the corresponding dependence statement is asserted. We will show next that, in any graph, the two sets of statements are always consistent. This result justifies the use of undirected graphs as a general language for encoding intricate patterns of statistical associations. Similar results hold for *directed acyclic graphs* [Geiger and Pearl (1988)].

THEOREM 11. *For every graph G with U as its nodes, there exists a strictly positive probability model (d, P) over U , such that for every three disjoint sets X, Y and Z of U ,*

$$J(X, Y|Z) \text{ holds for } G \text{ if and only if } I(X, Y|Z) \text{ holds for } (d, P).$$

PROOF. Let Σ be the set of separation statements that hold in G . For every statement $\sigma \notin \Sigma$, there exists a probability model (d_σ, P_σ) that satisfies Σ and does not satisfy σ where Σ and σ are interpreted as independence statements

(Lemma 9). Let (d, P) be $\otimes\{(d_\sigma, P_\sigma) \mid \sigma \notin \Sigma\}$. (The set $\{\sigma \mid \sigma \neg \in \Sigma\}$ is finite because U is finite.) Due to Theorem 5, (d, P) satisfies precisely the statements in Σ and none other. \square

Note, however, that \otimes assigns to each attribute in U an arbitrary domain size. We conjecture that this arbitrariness is not needed.

CONJECTURE 1. *For every graph G with u_1, \dots, u_n as its nodes and for every n integers k_1, \dots, k_n all greater than 2, there exists a strictly positive probability model (d, P) over U , such that (1) $|d(u_i)| = k_i$ and (2) for every three disjoint sets X, Y and Z of U ,*

$$J(X, Y|Z) \text{ holds for } G \text{ if and only if } I(X, Y|Z) \text{ holds for } (d, P).$$

5. Graphs and binary factorizations. The relationship between graph separation and conditional independence is even stronger than that shown so far if we restrict ourselves to strictly positive probability models and to *saturated statements*.

DEFINITION. An independence statement $I(X, Y|Z)$ or a separation statement $J(X, Y|Z)$ is *saturated* if $X \cup Y \cup Z = U$, where U is the finite set of attributes of interest.

In the following discussion we show that saturated independence statements (relative to \mathcal{P}^+) and saturated separation statements satisfy *precisely* the same axioms. This correspondence provides us with an efficient algorithm to determine all saturated independence statements entailed (relative to \mathcal{P}^+) by a given set of such statements.

Moreover, each statement $I(X, Y|Z)$ holds for a probability model (d, P) if and only if (d, P) has a *binary factorization*, namely,

$$P(X, Y, Z) = f(X, Y) \cdot g(Y, Z),$$

where g and f are any functions [Lauritzen (1982)]. Consequently, the proposed algorithm provides an efficient way to determine all binary factorizations entailed (relative to \mathcal{P}^+) by a given set of binary factorizations. [The terms saturated independence and binary factorizations are borrowed, respectively, from Lee and Buehler (1986) and Malvestuto (1992)].

We use the following theorem of Pearl and Paz (1985) which generalizes a result by Lauritzen (1982).

THEOREM 12. *Let Σ be a set of independence statements over a finite set of attributes U , and let Σ^+ be the closure of Σ with respect to trivial independence, symmetry, decomposition, intersection and weak union. Let G_0 be the graph having U as its nodes and an edge between x and y if and only if*

$I(\{x\}, \{y\} | U \setminus \{x, y\}) \in \Sigma^+$. Then (1) for every three disjoint subsets X, Y and Z of U ,

$J(X, Y | Z)$ holds for G_0 implies that $I(X, Y | Z) \in \Sigma^+$,

and (2) if any edge is removed from G_0 property 1 ceases to hold.

Next, we strengthen Theorem 12 when Σ consists of saturated independence statements.

THEOREM 13. *Let Σ be a set of saturated independence statements over a finite set of attributes U , and let Σ^+ be the closure of Σ with respect to saturated trivial independence [i.e., all statements of the form $I(X, \emptyset | Z)$ where $X \cup Z = U$], symmetry, intersection and weak union. Let G_0 be the graph defined in Theorem 12. Then for every three disjoint subsets X, Y and Z of U , such that $X \cup Y \cup Z = U$,*

$J(X, Y | Z)$ holds for G_0 iff $I(X, Y | Z) \in \Sigma^+$.

PROOF. The key point to notice is that $I(X, Y | Z) \in \Sigma^+$ if and only if $I(\{x\}, \{y\} | Z \cup (X \setminus \{x\}) \cup (Y \setminus \{y\}))$ is in Σ^+ for every $x \in X$ and $y \in Y$. Each of these independence statements is derivable from $I(X, Y | Z)$ by an application of weak union followed by symmetry, weak union and finally followed by symmetry. The statement $I(X, Y | Z)$ is derivable by repeated applications of intersection and symmetry. The same equivalence holds when I is replaced by J because separation satisfies the three axioms we have used in the preceding argument. Consequently, $J(X, Y | Z)$ holds for G_0 iff $J(\{x\}, \{y\} | Z \cup (X \setminus \{x\}) \cup (Y \setminus \{y\}))$ holds for G_0 for every $x \in X$ and $y \in Y$. By the definition of G_0 , the latter set of statements holds if and only if $I(\{x\}, \{y\} | Z \cup (X \setminus \{x\}) \cup (Y \setminus \{y\}))$ is in Σ^+ for every $x \in X$ and $y \in Y$. In addition, these statements are in Σ^+ iff $I(X, Y | Z) \in \Sigma^+$. An additional minor observation is that each trivial independence statement holds in every graph (U, E) in particular in G_0 . \square

Similarly, we obtain the following result.

THEOREM 14 (Completeness relative to \mathcal{P}^+). *Let Σ be a set of saturated independence statements, and let Σ^+ be the closure of Σ with respect to saturated trivial independence, symmetry, intersection and weak union. Then, for every $\sigma \notin \Sigma^+$, there exists a strictly positive probability model (d_σ, P_σ) over U , where U is the set of attributes that appears in Σ , that satisfies Σ^+ and does not satisfy σ .*

PROOF. By Theorem 13 there exists a graph G_0 that satisfies Σ^+ and no other independence statement. By Lemma 9 there exists a strictly positive probability model (d_σ, P_σ) that satisfies the statements that hold in G_0 and does not satisfy σ . Thus (d_σ, P_σ) satisfies the requirement of the theorem. \square

Theorems 13 and 14 together show that saturated independence statements and saturated separation statements share precisely the same axioms (relative to \mathcal{P}^+). This equivalence permits us to compute the set of all saturated independence statements entailed relative to \mathcal{P}^+ by a given set of saturated statements, using a purely graph-theoretic approach.

The algorithm is simple: Given a set of saturated independence statements Σ over U , construct the graph $G_0 = (U, E)$ as follows.

Step 1. Replace each given statement $I(X, Y|Z)$ with a set of independence statements $\{I(\{x\}, \{y\}|Z \cup (X \setminus \{x\}) \cup (Y \setminus \{y\})) \mid x \in X, y \in Y\}$.

Step 2. Introduce an edge between x and y if $I(\{x\}, \{y\}|Z \cup (X \setminus \{x\}) \cup (Y \setminus \{y\}))$ is not among the statements generated in Step 1.

Step 3. Output $I(X', Y'|Z') \in \Sigma^+$ if $J(X', Y'|Z')$ holds in the graph produced in Step 2. Otherwise output $I(X', Y'|Z') \notin \Sigma^+$.

The algorithm requires $O(|\Sigma| \cdot n^2)$ steps to construct G_0 where n is the number of attributes because it scans each statement of the input once and each statement may require checking n^2 pairs of attributes. Once G_0 is constructed, it permits us to check whether a specific saturated statement $\sigma = I(X, Y|Z)$ is entailed (relative to \mathcal{P}^+) by Σ in only $O(n)$ steps—the time needed to check whether Z separates X and Y in G_0 .

This method allows us to represent in polynomial space (in the number of attributes) the entire set of binary factorizations entailed (relative to \mathcal{P}^+) by a given set of binary factorizations and to determine, in polynomial time, whether or not a specific binary factorization is in this set. We will see next that a similar implication algorithm, albeit less efficient, can be developed without the assumption of strict positiveness.

6. Saturated independence. The next completeness theorem is the analog of Theorem 14 with weak contraction replacing intersection. This change is needed because intersection is sound relative to \mathcal{P}^+ but not relative to \mathcal{P} .

THEOREM 15 (Completeness relative to \mathcal{P}). *Let Σ be a set of saturated independence statements over a finite set of attributes U , and let Σ^+ be the closure of Σ with respect to saturated trivial independence, symmetry, weak contraction and weak union. Then, for every $\sigma \notin \Sigma^+$, there exists a probability model (d_σ, P_σ) that satisfies Σ^+ and does not satisfy σ .*

PROOF. Let $\sigma = I(X, Y|Z)$ be a saturated statement not in Σ^+ where $X \cup Y \cup Z = U$. At first we assume that σ is *maximal*, that is, for all sets $X'X''$ and $Y'Y''$ partitioning X and Y , respectively, the statement $I(X', Y'|ZX''Y'')$ is in Σ^+ . (In this proof AB stands for $A \cup B$.) At the end of the proof we relax this assumption.

Let each attribute in U be associated with a binary domain $\{0, 1\}$. Denote all attributes in X by x_1, x_2, \dots, x_l , those in Y by y_1, y_2, \dots, y_m and those in Z

by z_1, z_2, \dots, z_k . The probability model (d_σ, P_σ) is defined as follows:

$$P_\sigma(X, Y, Z) = \prod_{z_i \in Z} f(z_i) \cdot \begin{cases} 1/2, & \text{if all attributes in } X \cup Y \text{ are assigned 0,} \\ 1/2, & \text{if all attributes in } X \cup Y \text{ are assigned 1,} \\ 0, & \text{otherwise,} \end{cases}$$

where $f(z_i) = 1/2$.

This probability model does not satisfy σ because $P_\sigma(X = 0, Y = 1, Z = 0)$ is 0, while $P_\sigma(X = 0, Z = 0)$ and $P_\sigma(Y = 1, Z = 0)$ are not.

It remains to show that every saturated statement in Σ^+ holds for (d_σ, P_σ) , or equivalently that every saturated statement either holds for (d_σ, P_σ) or does not belong to Σ^+ . Any saturated statement γ can be written as $I(X_1Y_1Z_1, X_3Y_3Z_3|X_2Y_2Z_2)$, where $X = X_1X_2X_3$, $Y = Y_1Y_2Y_3$ and $Z = Z_1Z_2Z_3$ and the X_i 's, Y_i 's and Z_i 's are all disjoint. If $X_2Y_2 \neq \emptyset$, then γ holds for (d_σ, P_σ) because every instance of $X_1Y_1Z_1$ and of $X_3Y_3Z_3$ that is consistent with the values of X_2Y_2 has the same probability of occurring, namely, $1/2^{|Z_1|} \cdot 1/2^{|Z_3|}$. If $X_1Y_1 = \emptyset$, then, again, γ holds for (d_σ, P_σ) because Z_1 is marginally and conditionally independent of any other set of attributes of P_σ . (Symmetrically when $X_3Y_3 = \emptyset$.) Otherwise, γ is of the form $I(X_1Y_1Z_1, X_3Y_3Z_3|Z_2)$, where $X_1Y_1 \neq \emptyset$ and $X_3Y_3 \neq \emptyset$. We continue by contradiction and show that in this case γ does not belong to Σ^+ .

Assume, by contradiction, that the statement $I(X_1Y_1Z_1, X_3Y_3Z_3|Z_2)$ is in Σ^+ . Then $I(X_1Y_1, X_3Y_3|Z)$ is in Σ^+ as well because it can be derived by weak union and symmetry. To reach a contradiction, we show that the latter statement implies that σ must be in Σ^+ , contradicting our selection of σ . The proof uses weak contraction and symmetry to derive $I(X_1X_3, Y_1Y_3|Z)$ (i.e., σ) from $I(X_1Y_1, X_3Y_3|Z)$ by "joining" the X 's and the Y 's. The following is a derivation of σ .

First, $I(X_1, Y_1|ZX_3Y_3)$ is in Σ^+ because $I(X, Y|Z)$ is maximal. Due to weak contraction,

$$I(X_1Y_1, X_3Y_3|Z), I(X_1, Y_1|ZX_3Y_3) \Rightarrow I(X_1, Y_1X_3Y_3|Z),$$

we conclude that $I(X_1, YX_3|Z) \in \Sigma^+$. Due to symmetry, we conclude $I(YX_3, X_1|Z) \in \Sigma^+$ as well. $I(X_3, Y|ZX_1) \in \Sigma^+$ because σ is maximal. Therefore, by symmetry, $I(Y, X_3|ZX_1)$ is also in Σ^+ . Using weak contraction, we obtain

$$I(YX_3, X_1|Z), I(Y, X_3|ZX_1) \Rightarrow I(Y, X_1X_3|Z).$$

Thus $I(Y, X|Z) \in \Sigma^+$, and, by symmetry, $I(X, Y|Z) \in \Sigma^+$, a contradiction. (Note that if some sets out of X_1, X_3, Y_1 and Y_3 are empty, the derivation just described remains valid.)

If $\sigma = I(X, Y|Z)$ is not maximal, then either $I(X \setminus \{x\}, Y|Z \cup \{x\}) \notin \Sigma^+$ for some $x \in X$ or $I(X, Y \setminus \{y\}|Z \cup \{y\}) \notin \Sigma^+$ for some $y \in Y$. Without loss of generality assume the first statement is not in Σ^+ . If this statement is maximal, denote it σ' . Otherwise, repeat the process of augmenting Z with additional elements from X and Y . When this process can no longer continue, we denote the resulting statement $\sigma' = I(R, S|T)$. Clearly, σ' is maximal; it is

not in Σ^+ and for all sets $R'R''$ and $S'S''$ partitioning S and T , respectively, the statement $I(R', S'|TR''S'')$ is in Σ^+ .

For a maximal statement σ' , we have shown how to construct a probability model $(d_{\sigma'}, P_{\sigma'})$ that satisfies Σ and does not satisfy σ' . Due to symmetry and weak union, which hold for all probability models, any probability model that does not satisfy σ' , does not satisfy σ as well. In particular, $(d_{\sigma'}, P_{\sigma'})$ does not satisfy σ while satisfying Σ^+ , as required by the theorem. \square

The probability model (d_{σ}, P_{σ}) constructed previously has an additional property; each combination of values for $X \cup Y \cup Z$ has either zero probability or a constant probability of $1/2^{|Z|+1}$. Thus the probability model (d_{σ}, P_{σ}) can be viewed as a *database*, categorically distinguishing between possible and impossible value combinations. Consequently, the proof of Theorem 15 shows that the previously mentioned axioms are also complete for MVD statements of relational databases [Fagin (1978)]. Indeed, the only difference between our axioms and the ones governing MVD's is that the latter allow overlapping sets X, Y and Z in $I(X, Y|Z)$ whereas we do not [Beeri, Fagin and Howard (1977)]. This equivalence permits the employment of a polynomial implication algorithm devised for MVDs [Beeri (1980)] to determine whether a saturated statement is entailed by a set of saturated statements, just as the equivalence between graph separation and conditional independence (relative to \mathcal{P}^+) provided us with an implication algorithm in the previous section.

Malvestuto (1992) has independently observed this equivalence and used it to produce an indirect proof of Theorem 15 by showing that MVD and saturated independence statements must satisfy the same set of axioms.

The complexity of the implication algorithm for saturated statements relative to \mathcal{P} [Beeri (1980)] differs from that needed relative to \mathcal{P}^+ ; the former requires $O(|\Sigma| \cdot n^2)$ operations to decide $\Sigma \models \sigma$ for each σ , while the latter requires only $O(n)$ operations, regardless of $|\Sigma|$. These savings are achieved at the cost of investing $O(|\Sigma| \cdot n^2)$ steps in constructing a graphical representation of the closure of Σ (relative to \mathcal{P}^+), but this cost is encountered only once. This difference in complexity can be significant since, in principle, $|\Sigma|$ can be exponential in n .

7. Marginal independence. This section summarizes two completeness results for statements of the form $I(X, Y|\emptyset)$ (*marginal statements*).

THEOREM 16 (Completeness). *Let Σ be a set of marginal statements, and let Σ^+ be the closure of Σ with respect to axioms (12) through (15). Then for every marginal statement $\sigma = I(X, Y|\emptyset)$ not in Σ^+ , there exists a binary probability model (d_{σ}, P_{σ}) that satisfies Σ^+ and does not satisfy σ .*

Marginal trivial independence:

$$(12) \quad I(X, \emptyset|\emptyset).$$

Marginal symmetry:

$$(13) \quad I(X, Y|\emptyset) \Rightarrow I(Y, X|\emptyset).$$

Marginal decomposition:

$$(14) \quad I(X, Y \cup W|\emptyset) \Rightarrow I(X, Y|\emptyset).$$

Marginal mixing:

$$(15) \quad I(X, Y|\emptyset), I(X \cup Y, W|\emptyset) \Rightarrow I(X, Y \cup W|\emptyset).$$

The proof of Theorem 16 uses the same technique as that of Theorem 15. It can be found in Geiger, Paz and Pearl (1991), together with an $O(|\Sigma| \cdot n^2)$ implication algorithm that is based on these axioms. The implication algorithm and the axiomatization hold relative to \mathcal{B} and \mathcal{P} .

DEFINITION. A *Gaussian model* over a finite set of attributes $U = \{u_1, \dots, u_n\}$ is a pair (d, P) , where d is a domain mapping that maps each u_i to $(-\infty, +\infty)$, and $P: d(u_1) \times \dots \times d(u_n) \rightarrow [0, 1]$ is a multivariate Gaussian probability distribution. (For the sake of brevity, we will not define multivariate Gaussian probability distributions.) The class of Gaussian models is denoted by \mathcal{N} .

Gaussian models share stronger properties for marginal independence than the ones listed previously; in particular, it is well known that Gaussian models satisfy the following additional property:

Marginal composition:

$$(16) \quad I(X, Y|\emptyset), I(X, W|\emptyset) \Rightarrow I(X, Y \cup W|\emptyset).$$

Theorem 17 shows that marginal composition is the only axiom that was “missing” relative to Gaussian models.

THEOREM 17 (Completeness). *Let Σ be a set of marginal statements, and let Σ^+ be the closure with respect to marginal trivial independence, marginal symmetry, marginal decomposition and marginal composition. Then there exists a Gaussian model that satisfies all statements in Σ^+ and none other.*

PROOF. Let $U = u_1, \dots, u_n$ be the attributes of interest. Let P be a zero-mean multivariate normal distribution, with the following covariance matrix:

$$\Gamma = (\rho_{i,j}) \quad \text{where } \rho_{i,j} = \begin{cases} 0, & \text{if } \exists I(X, Y|\emptyset) \in \Sigma \text{ s.t. } u_i \in X, u_j \in Y, \\ \rho, & \text{otherwise,} \end{cases}$$

where $\rho^2 \ll 1$. Simple algebra shows that this matrix is positive definite.

We need to show that P satisfies Σ^+ and no other marginal statement or, equivalently, that $I(X, Y|\emptyset) \in \Sigma$ if and only if $I(X, Y|\emptyset)$ holds for P . This is

proven by the following chain of relationships:

$$I(X, Y|\emptyset) \in \Sigma^+ \text{ iff } \forall u_i \in X, u_j \in Y, I(u_i, u_j|\emptyset) \in \Sigma^+ \text{ iff} \\ \forall u_i \in X, u_j \in Y, \rho_{i,j} = 0 \text{ iff } \forall u_i \in X, u_j \in Y, I(u_i, u_j|\emptyset) \\ \text{holds for } P \text{ iff } I(X, Y|\emptyset) \text{ holds for } P.$$

The first and last equivalences hold due to marginal decomposition and composition, making any statement $I(X, Y|\emptyset)$ completely determined by statements on singletons. The second equivalence holds by the construction of Γ and the third equivalence is a property of normal distributions. \square

The construction of the matrix Γ requires $O(|\Sigma| \cdot n^2)$ steps, where n is the number of attributes appearing in statements of Σ . Testing whether a marginal statement $I(X, Y|\emptyset)$ is entailed (relative to Gaussian models) by a set of marginal statements amounts to checking that $\rho_{i,j} = 0$ for every $u_i \in X$ and $u_j \in Y$, which requires on the order of n^2 steps.

8. Nonaxiomatizability of conditional independence. The previous two sections provide finite sets of Horn axioms for marginal and saturated independence statements. These axiom sets remain fixed when the size of U increases because our proofs depend only on the finiteness of $|U|$ but not on its actual size. Unfortunately, analogous results for independence statements (without restrictions) cannot be obtained.

THEOREM 18 [Studeny (1992)]. *There exists no finite set of Horn axioms for independence statements (relative to \mathcal{P}) that is complete for every finite $|U|$.*

Studeny proved the preceding theorem by presenting an infinite set of Horn axioms for conditional independence that is not implied by any finite set of such axioms.

The nonexistence of a complete set of axioms does not exclude the possibility of an efficient implication algorithm for conditional independence; Sagiv and Walecka (1982) provide an example of a class of sentences, called Z -EMVD, which admits an efficient polynomial implication algorithm but for which there exists no finite set of axioms. Nevertheless, we make the following conjecture.

CONJECTURE 2. *The task of determining whether an independence statement is entailed (in \mathcal{P}) by a set of independence statements requires at least exponential time.*

Moreover, the preceding task might even be undecidable; that is, there might exist no algorithm for deciding entailment of conditional independence. For related problems, consult Fagin and Vardi (1986).

9. Qualitative independence. Similar to conditional independence, we can define a *qualitative independence statement* $\hat{I}(X, Y|Z)$ by saying that $\hat{I}(X, Y|Z)$ holds for P if

$$P(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) > 0 \text{ and } P(\mathbf{Z}) > 0 \text{ iff } P(\mathbf{X}, \mathbf{Z}) > 0 \text{ and } P(\mathbf{Y}, \mathbf{Z}) > 0,$$

for every respective value of \mathbf{X} , \mathbf{Y} and \mathbf{Z} .

This definition is identical to that of EMVD in database theory and is also discussed by Shafer, Shenoy and Mellouli (1988). Theorems 5, 14 and 15 hold when I is replaced with \hat{I} . For details consult Geiger (1990).

10. Summary. Table 1 summarizes properties of classes of probability models versus classes of independence statements. A question mark means that the problem remains open as of the writing of this article. The symbol \mathcal{N} denotes the class of normal models and \mathcal{B} probability models over binary variables.

Some properties of Gaussian models are listed in Table 1 which have not been proven in this article. The axioms for saturated independence (relative to \mathcal{N}) consist of trivial independence, symmetry, weak union and intersection [Geiger (1990)]. The fact that perfect Gaussian models do not exist for some sets of statements can be proven in the same way as in Corollary 8 (with the same Σ selected). The nonexistence of a finite set of Horn axioms can be proven in the same way as in Theorem 18.

In addition, we have shown a strong relationship between graph separation and conditional independence. In particular, every undirected graph represents a consistent set of independence and dependence statements (Theorem 11), every axiom for conditional independence is also an axiom for graph separation (Theorem 10) and saturated separation and saturated independence (relative

TABLE 1
Properties of conditional independence

	Properties	Marginal statements	Saturated statements	Unrestricted statements
\mathcal{P}	Complete finite axiomatization	Yes	Yes	No
	Polynomial implication algorithm	Yes	Yes	?
	Perfect models	Yes	Yes	Yes
\mathcal{P}^+	Complete finite axiomatization	?	Yes	?
	Polynomial implication algorithm	?	Yes	?
	Perfect models	Yes	Yes	Yes
\mathcal{N}	Complete finite axiomatization	Yes	Yes	?
	Polynomial implication algorithm	Yes	Yes	?
	Perfect models	Yes	Yes	No
\mathcal{B}	Complete finite axiomatization	Yes	Yes	?
	Polynomial implication algorithm	Yes	Yes	?
	Perfect models	?	?	No

to \mathcal{P}^+) share the same axiomatic structure (Theorems 13 and 14). Analogous correspondence exists between separation in directed acyclic graphs (*d-separation*) and conditional independence. See Geiger and Pearl (1988) and Verma (1986) for details.

Acknowledgments. Our notation and definitions were influenced by Beeri, Fagin and Howard (1977) and Fagin (1976). We are indebted to Ron Fagin for pointing out the usefulness of the notion of Armstrong models. We thank Azaria Paz for his help in proving Theorem 13 by referring us to Paz (1987), and to Norman Dalkey and Thomas Verma for many useful discussions. We also thank Glenn Shafer and several reviewers for suggesting numerous improvements on earlier drafts.

REFERENCES

- BEERI, C. (1980). On the membership problem for functional and multivalued dependencies in relational databases. *ACM Trans. Database Systems* **5** 241–249.
- BEERI, C., FAGIN, R. and HOWARD, J. H. (1977). A complete axiomatization of functional dependencies and multi valued dependencies in database relations. *Proceedings of the 1977 ACM SIGMOD International Conference on Management of Data* 47–61, ACM, New York.
- DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and loglinear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.
- DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31.
- FAGIN, R. (1976). Functional dependencies in a relational database and propositional logic. *IBM J. Res. Develop.* **21** 262–278.
- FAGIN, R. (1978). Multivalued dependencies and a new normal form for relational databases. *ACM Trans. Database Systems* **2** (262–278).
- FAGIN, R. (1982). Horn clauses and database dependencies. *J. Assoc. Comput. Mach.* **29** 952–985.
- FAGIN, R. and VARDI, M. Y. (1986). The theory of data dependencies—a survey. In *Mathematics of Information Processing* (M. Anshel and W. Gewirtz, eds.) *Proc. Symp. Appl. Math.* **34** 19–72. Amer. Math. Soc., Providence, RI.
- FRYDENBERG, M. (1990). Marginalization and collapsibility in graphical association models. *Ann. Statist.* **18** 790–805.
- GEIGER, D. (1990). Graphoids: a qualitative framework for probabilistic inference. Ph.D. dissertation, Computer Science Dept., Univ. California, Los Angeles.
- GEIGER, D., PAZ, A. and PEARL, J. (1991). Axioms and algorithms for inferences involving probabilistic independence. *Inform. and Comput.* **91** 128–141.
- GEIGER, D. and PEARL, J. (1988). On the logic of causal models. *Proceedings of the Fourth Workshop on Uncertainty in AI* (R. D. Shacter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, eds.) 136–147. North-Holland, Amsterdam.
- LAURITZEN, S. L. (1982). *Lectures on Contingency Tables*, 2nd ed. Univ. Aalborg Press, Aalborg, Denmark.
- LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their applications to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 154–227.
- LEE, Y.G. and BUEHLER, R. J. (1986). Independence relationships for multivariate distributions. Technical Report 464, Univ. Minnesota.
- MALVESTUTO, F. M. (1992). A unique formal system for binary decompositions of database relations, probability distributions, and graphs. *Inform. Sci.* **59** 21–52.
- PAZ, A. (1987). A full characterization of pseudographoids in terms of families of undirected graphs. Technical Report R-95, Cognitive Systems Laboratory, Univ. California, Los Angeles.

- PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.
- PEARL, J. and PAZ, A. (1985). Graphoids: a graph-based logic for reasoning about relevance relations. In *Advances in Artificial Intelligence—II* (B. D. Boulay, D. Hogg and L. Steels, eds.) 357–363. North-Holland, Amsterdam.
- SAGIV, Y. and WALECKA, S. (1982). Subset dependencies and completeness results for a subset of EMVD. *J. Assoc. Comput. Mach.* **29** 103–117.
- SHAFER, G., SHENOY, P. and MELLOULI, K. (1988). Propagating belief functions in qualitative Markov trees. *Internat. J. Approx. Reason.* **1** 349–400.
- SPOHN, W. (1980). Stochastic independence, causal independence, and shieldability. *J. Philos. Logic* **9** 73–99.
- STUDENY, M. (1992). Conditional independence relations have no complete characterization. *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Foundation and Random Processes Vol. B* 377–396. Kluwer, Dordrecht.
- VERMA, T. S. (1986). Causal networks: semantics and expressiveness. Technical Report R-65, Cognitive Systems Laboratory, Univ. California, Los Angeles. [Also in VERMA, T. S. and PEARL, J. (1990). *Uncertainty in Artificial Intelligence* (R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, eds.) 69–76. North-Holland, Amsterdam.]
- WHITTAKER J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.

COMPUTER SCIENCE DEPARTMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA 32000,
ISRAEL

COMPUTER SCIENCE DEPARTMENT
6291 BOELTER HALL
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA 90024-1596