A COMPLETE AXIOMATIZATION OF SOME DEPENDENCIES IN PROBABILISTIC MODELS

Dan Geiger

September 1987
CSD-870056

TECHNICAL REPORT
R-97
Aug 28, 1987

A COMPLETE AXIOMATIZATION OF SOME DEPENDENCIES IN PROBABILISTIC MODELS

Dan Geiger

Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024

ABSTRACT

This paper deals with the task of establishing a complete axiomatic basis for the probabilistic independency relation "x is independent of y". This relation is shown to be an Armstrong relation. A set of axioms is presented and shown to be complete for this relation. Some other dependencies in probabilistic models are presented for which a tractable membership algorithm is given. We also justify the use of Undirected Graphs as a represantation scheme for probabilistic dependencies.

* This work was supported in part by the National Science Foundation Grant, IRI-8610155.

2. COMPLETENESS AND ARMSTRONG RELATIONS.

As a metaphor for the concept of completeness, consider a multivariate probabilistic phenomena for which the probability model (distribution) is unknown. For example, the search of oil involves many variables such as: earth surface, soil, earthquake history etc. Each combination of these events (i.e instantiation of the variables) has some probability to occur, however the exact distribution is unknown. An expert may supply us with a list of dependencies and independencies (i.e constrains) between the various parameters. Such assessments, free of any numerical reference, describe constrains of the phenomena regardless of the exact distribution. Our task is to find inference rules (axioms) that are powerful enough to infer all the constrains that are *logically implied* by the original list given by the expert. Clearly the desired axioms depend on two factors: the type of constrains given by the expert and the family of distributions from which the model is drawn. The latter, if unknown, is taken to be the family of all distributions.

we consider three types of constrains:

1) Conditional independency constrains, syntactically denoted $I(x,z,y)$.Where x , y and z are non-intersecting sets of variables and $I(x,z,y)$ is assigned a truth value if

$$P(x,y \mid z) = P(x \mid z) \cdot P(y \mid z). \tag{1}$$

These constrains will be referred as *conditional statements*. For technical convenience we shall adopt the convention that every variable is independent of the null set, i.e., $I(x, z, \emptyset)$.

2) Fixed-context Conditional independency constrains. These constrains are denoted by $I(x, z, y)$ and their truth values is also determined by equation (1). The difference from the first type lies in the restriction that $x \cup y \cup z$ must always sum to a fixed set of variables U. These constrains are named *U-statements*. Clearly every U-statement is a conditional statement and the converse is not true.

3) Marginal independency constrains. These constrains, named *marginal statements*, are denoted by (x, y) and their truth value is determined via

$$P(x,y) = P(x) \cdot P(y).$$

Note that every marginal statement can be regarded also as a conditional statement by denoting

$$I(x, \emptyset, y) \text{ iff } P(x,y) = P(x) \cdot P(y).$$

The following families of distributions are considered in this paper:

1) The family of all models denoted PD.
2) The family of all non-extreme models, i.e models for which every instantiation of variables is assigned a non-zero probability, denoted PD^-.
3) The family of all binary models, i.e models for which every variable is binary, denoted PB.
4) The family of all Normal (Gausian) models, denoted PN.

In all definitions throughout this section, we assume a fixed family of models **P** and a set of constrains **S** from a fixed type. An element s_i of **S** is called a *statement.*

Definition: An axiom

$$s_1 \,\&\, s_2 \,\&\, \cdots \,\&\, s_n \;\rightarrow\; \hat{s}_1 \; or \; \hat{s}_2 \; or \; \cdots \; or \; \hat{s}_m$$

is *sound* in **P** if every model in **P** that obeys the antecedents of the axiom also obeys at least one of the statements of the disjunction on the right hand side of the implication. When m=1 the axiom is said to be a *Horn axiom.*

Notation: Let $\Sigma \subseteq$ **S** be a set of statements and let $\sigma \in$ **S** be a single statement. Let **A** be a set of axioms. When σ is *logically implied* by Σ, i.e, every model in **P** that obeys Σ also obeys σ we use the notation: $\Sigma \models \sigma$. When σ can be derived from Σ by using the axioms in **A** we write: $\Sigma \models_A \sigma$.

Definition: A set of axioms is *weakly complete* if for every set of statements $\Sigma \subseteq$ **S** we have

$$\Sigma \models \sigma \text{ iff } \Sigma \models_A \sigma.$$

In other words every logical consequence of Σ can actually be derived by using the axioms in **A** and visa versa every statement that can be derived is a logical consequence. Clearly, a necessary condition for **A** to be weakly complete is that every axiom of **A** is sound in **P**.

Intuitively, one would desire that the completeness of a set of axioms **A** would imply that every sound axiom is derivable from **A**. However the definition of weak completeness does not imply such a property. Weak completeness, as will be shortly shown, only implies that sound Horn axioms are derivable from **A** and not necessarily an arbitrary non-Horn axiom. For this reason we define the following:

Definition: A set of axioms is *complete* if for every set of statements Σ and for every disjunction $\sigma_1 \; or \; \sigma_2 \; or \; \cdots \; \sigma_m$ we have

$$\Sigma \models \sigma_1 \; or \; \sigma_2 \; or \; \cdots \; \sigma_m \text{ iff } \Sigma \models_A \sigma_1 \; or \; \sigma_2 \; or \; \cdots \; \sigma_m .$$

The symbols $\models$ and $\models_A$ are defined in the same way as for single statements.

The next two propositions give equivalent definitions for weak completeness and completeness. It demonstrate that completeness of a set of axioms **A** means that every sound axiom is derivable from **A**. The concepts of completeness and weak completeness are taken from database theory. In [10], Fagin proved a proposition similar to proposition 1, for which **A** consists of Horn axioms. This constrain is now relaxed. We point out, however, that Fagin's remarkable papers ([10], [11]) are an important basis for the work reported herein.

Proposition 1: The following conditions are equivalent.

a) **A** is a weakly complete set of axioms.

b) Every sound Horn axiom is derivable from A.

c) For every set of statements Σ closed under a set of sound axioms A and for every $\sigma \notin \Sigma$ there exist a probability model P_σ that obeys all statements in Σ but does not obey σ.

Proposition 2: The following conditions are equivalent.

a) A is a complete set of axioms.

b) Every sound axiom is derivable from A.

c) For every set of statements Σ closed under a set of sound axioms A there exist a probability model P that obeys exactly the statements in Σ.

Proof: We prove only the first proposition. The second proof is very similar and therefore omitted.

a→b: Let $A_1 : \sigma_1 \,\&\, \sigma_2 \,\&\, \cdots \,\&\, \sigma_n \rightarrow \sigma$ be a sound Horn axiom. Let $\Sigma = \{\sigma_1, \cdots, \sigma_n\}$. Due the soundness of A_1 we have $\Sigma \models \sigma$. A is weakly complete. Therefore given Σ we can derive σ by the axioms in A. Thus A_1 is derivable from A.

b→c: Assume, by contradiction, that $\sigma \notin \Sigma$ and that every model P that satisfies Σ also satisfies σ. Let $\Sigma = \{\sigma_1, \cdots, \sigma_n\}$. Consider the axiom $A_1 : \sigma_1 \,\&\, \sigma_2 \,\&\, \cdots \,\&\, \sigma_n \rightarrow \sigma$. This axiom is sound in P. Σ is closed under A. A_1 cannot be derived from A, for otherwise σ would belong to Σ, contradicting our selection of σ.

c→a: Assume A is a set of sound axioms that is not weakly complete in P. Thus, there exists a set Σ and a statement σ such that $\Sigma \models \sigma$ and $\Sigma \not\models_A \sigma$. Let $\hat{\Sigma}$ be a supper set of Σ that is closed under A and does not contain σ. Such a set always exists because $\Sigma \not\models_A \sigma$. Construct P_σ that satisfies $\hat{\Sigma}$ and $\neg\sigma$. P_σ also satisfies Σ and therefore σ is not logically implied by Σ, contradicting our selection of σ.
□

From the above definitions it is clear that completeness implies weak completeness. The converse, however does not always hold. {An example can be found in [10], but it is long and i still hope to bring an example from marginal statements in PB}.

We now turn on to an interesting property of dependencies in PD, namely, that every complete set of axioms is also complete. It is surprising that after differentiating these two concepts, we now state that they are equivalent. Indeed, this result does not follow from the definitions of completeness and weak completeness rather it is due to the following property of PD stated in the next theorem.

Theorem 3: Let S be a set of statements. Then there exist an operation $\oplus$ that maps finite sequences of probabilistic models into probabilistic models, such that if σ is a statement in S and if $P_i \;\; i=1..n$ are models in PD, then σ holds for $\oplus\{P_i \;\; i=1..n\}$ iff σ holds for each P_i.

Proof: We shall construct the operation $\oplus$ for conditional statements. The same construction holds for U-statements and marginal statements. First we construct a binary operation $\oplus'$ such that if $P = P_1 \oplus P_2$ then for every $\sigma \in$ S we get

$$\text{P obeys } \sigma \quad \text{iff} \quad P_1 \text{ obeys } \sigma \text{ and } P_2 \text{ obeys } \sigma.$$

The operation $\oplus$ is defined in terms of $\oplus'$ as follows:

$$\oplus\{P_i \mid i=1..n\} = ((P_1 \oplus' P_2) \oplus' P_3) \oplus' \cdots P_n).$$

Let $P_1(x_1, \cdots, x_n)$ and $P_2(x_1, \cdots, x_n)$ be probabilistic models (without loss of generality we assume both models use the same variables, because if σ contains a variable that does not appear in P_1 (or similarly in P_2) then σ trivialy does not hold in P_1). Let $A_1, \cdots, A_n$ be the domains of $x_1, \cdots, x_n$ in P_1 and let an instantiation of this variables be $\alpha_1, \cdots, \alpha_n$. Similarly, let $B_1, \cdots, B_n$ be the domains of $x_1, \cdots, x_n$ in P_2 and $\beta_1, \cdots, \beta_n$ an instantiation of these variables in P_2. The domain of $P = P_1 \oplus' P_2$ is the product domain $A_1 B_1, \cdots, A_n B_n$ and an instantiation of the variables of P is $\alpha_1\beta_1, \cdots, \alpha_n\beta_n$. Define

$$P(\alpha_1\beta_1, \alpha_2\beta_2, \cdots, \alpha_n\beta_n) = P_1(\alpha_1, \alpha_2, \cdots, \alpha_n) \cdot P_2(\beta_1, \beta_2, \cdots, \beta_n). \tag{2}$$

To prove that $\oplus'$ satisfies the required conditions, we need the following lemma:

Lemma: For every subset $\{x_{i_1}, \cdots, x_{i_l}\}$ of the variables of P, the following equality holds:

$$P(\alpha_{i_1}\beta_{i_1}, \alpha_{i_2}\beta_{i_2}, \cdots, \alpha_{i_l}\beta_{i_l}) = P_1(\alpha_{i_1}, \alpha_{i_2}, \cdots, \alpha_{i_l}) \cdot P_2(\beta_{i_1}, \beta_{i_2}, \cdots, \beta_{i_l}). \tag{3}$$

Proof: We first prove the lemma for $i_1 = 1, i_2 = 2, \cdots i_l = l$. For $l=n$ equation (3) is identical to equation (2). We proceed by descending induction. Assume the lemma holds for $l = k < n$. Then,

$$P(\alpha_1\beta_1, \cdots, \alpha_{k-1}\beta_{k-1}) = \sum_{x_k} P(\alpha_1\beta_1, \cdots, \alpha_{k-1}\beta_{k-1}, x_k)$$

$$= \sum_{(\alpha_k, \beta_k) \in A_k B_k} P_1(\alpha_1, \cdots, \alpha_{k-1}, \alpha_k) \cdot P_2(\beta_1, \cdots, \beta_{k-1}, \beta_k)$$

$$= \left[\sum_{\alpha_k \in A_k} P_1(\alpha_1, \cdots, \alpha_{k-1}, \alpha_k) \right] \cdot \left[\sum_{\beta_k \in B_k} P_2(\beta_1, \cdots, \beta_{k-1}, \beta_k) \right]$$

$$= P_1(\alpha_1, \cdots, \alpha_{k-1}) \cdot P_2(\beta_1, \cdots, \beta_{k-1})$$

The proof is completed by repeating the induction step for all the $n!$ orderings of $\{x_1, \cdots, x_n\}$. □

Let $I(x, z, y)$ be an arbitrary statement. Let $\alpha_x, \alpha_y, \alpha_z$ be an instantiation of x, y, z in P_1 and let $\beta_x, \beta_y, \beta_z$ be an instantiation of x, y, z in P_2. It is left to prove that

$$P(x, z, y) = P(x, z) \cdot P(y \mid z) \ \textit{iff} \ \ P_1(x, z, y) = P_1(x, z) \cdot P_1(y \mid z) \text{ and }$$
$$P_2(x, z, y) = P_2(x, z) \cdot P_2(y \mid z)$$

<--:

$$\begin{aligned}
P(\alpha_x\beta_x, \alpha_y\beta_y, \alpha_z\beta_z) &= P_1(\alpha_x, \alpha_y, \alpha_z) \cdot P_2(\beta_x, \beta_y, \beta_z) \\
&= P_1(\alpha_x, \alpha_z) \cdot P_1(\alpha_y \mid \alpha_z) \cdot P_2(\beta_x, \beta_z) \cdot P_2(\beta_y \mid \beta_z) \\
&= P(\alpha_x\beta_x, \alpha_z\beta_z) \cdot \left[\frac{P_1(\alpha_y, \alpha_z) \cdot P_2(\beta_y, \beta_z)}{P_1(\alpha_z) \cdot P_2(\beta_z)} \right] \\
&= P(\alpha_x\beta_x, \alpha_z\beta_z) \cdot \left[\frac{P(\alpha_y\beta_y, \alpha_z\beta_z)}{P(\alpha_z\beta_z)} \right] \\
&= P(\alpha_x\beta_x, \alpha_z\beta_z) \cdot P(\alpha_y\beta_y \mid \alpha_z\beta_z)
\end{aligned}$$

-->:

$$\begin{aligned}
P_1(\alpha_x, \alpha_y, \alpha_z) &= \frac{P(\alpha_x\beta_x, \alpha_y\beta_y, \alpha_z\beta_z)}{P_2(\beta_x, \beta_y, \beta_z)} \\
&= \frac{P(\alpha_x\beta_x, \alpha_z\beta_z) \cdot P(\alpha_y\beta_y, \alpha_z\beta_z)}{P_2(\beta_x, \beta_y, \beta_z) \cdot P(\alpha_z\beta_z)} \\
&= \left[\frac{P_1(\alpha_x, \alpha_z) \cdot P_1(\alpha_y, \alpha_z)}{P_1(\alpha_z)} \right] \cdot \left[\frac{P_2(\beta_x, \beta_z) \cdot P_2(\beta_y, \beta_z)}{P_2(\beta_x, \beta_y, \beta_z) \cdot P_2(\beta_z)} \right] \\
&= P_1(\alpha_x, \alpha_z) \cdot P_1(\alpha_y \mid \alpha_z) \cdot \left[\frac{P_2(\beta_x, \beta_z) \cdot P_2(\beta_y \mid \beta_z)}{P_2(\beta_x, \beta_y, \beta_z)} \right] \\
&\stackrel{\Delta}{=} P_1(\alpha_x, \alpha_z) \cdot P_1(\alpha_y \mid \alpha_z) \cdot Const
\end{aligned}$$

By summing over all values α_x we get that $Const = 1$. Hence $I(x, z, y)$ holds both in P_1 and P_2. □

Definition: A set of constrains **S** is said to be an *armstrong relation* in **P** (a family of distributions), if exists an operation ⊗ that satisfies the requirement of theorem 3 in **P**. For example conditional independence is an armstrong relation in PD.

Corollary: If a set of constrains is an armstrong relation in **P** then every weakly complete set of axioms is also complete.

Proof: If **A** is weakly complete, then by part (c) of proposition 1, for every $\Sigma \subseteq$ S closed under **A** and $\sigma \notin \Sigma$ there exist a model $P_\sigma \in$ **P** that realizes Σ and $\neg\sigma$. Let $P = \oplus\{P_\sigma \mid \sigma \notin \Sigma\}$. P satisfies part (c) of proposition 2, hence **A** is complete.

Remark 1: The operation $\oplus$ is well defined in **P**, if for every P_i $i = 1..n$ we get $\oplus P_i \in$ **P**. For example, conditional independence is an armstrong relation in PD^-, because the model $\oplus P_i$ has a product form of non-zero quantities and therefore $\oplus P_i$ is also a non-extreme model that belongs to PD^-. However, conditional independence is not an armstrong relation in PB. We first note that if P_1 and P_2 are both binary, $P_1 \oplus P_2$ is not binary. We now need to show that there exist no operation $\otimes$ that satisfies theorem 3. This proof is postponed until after equivalent definitions of armstrong relation are given.

Remark 2: If conditional independence is an armstrong relation in **P** then also marginal independence is an armstrong relation in **P**. The converse, however, does not hold. For example consider the Normal distributions PN. In normal models $P(\alpha, \beta) = P(\alpha) \cdot P(\beta)$ iff $\rho_{\alpha\beta} = 0$ where $\rho_{\alpha\beta}$ is the correlation factor of the variables α and β. Given a set of normal models we construct the normal standard distribution $\oplus P_i$ by assigning $\rho_{\alpha\beta} = 0$ in $\oplus P_i$ iff $\rho_{\alpha\beta} = 0$ in every P_i. All other correlation factors are assigned a non-zero quantity ρ (ρ satisfies $n \cdot \rho^2 < 1$ to assure that the covariance matrix of $\oplus P_i$ is positive defined). Therefore, marginal independence is an armstrong relation in PN. The proof that conditional independence is not an armstrong relation in PN is postponed until after equivalent definitions of armstrong relation are given.

Theorem 4 (Fagin [10]): Let **S** be a set of statements and **P** be a family of distributions. The following conditions are equivalent.

(a) **S** is an armstrong relation in **P**.

(b) For every set $\Sigma \subseteq$ **S** and $\sigma \in$ **S** there exist a model P $\in$ **P** such that

$$\text{P obeys } \sigma \text{ iff } \Sigma \models \sigma$$

(c) For every set $\Sigma \subseteq$ S and $\sigma_i \in$ **S** $i = 1..n$ the following holds.

$$\Sigma \models \sigma_i \text{ or } \sigma_i \cdots \text{ or } \sigma_n \text{ iff There exist an i such that } \Sigma \models \sigma_i$$

Remark 3: This theorem holds regardless of the existence of a finite complete set of axioms for **P**. For example define **P** by specifying infinite sequence of independent Horn axioms that every P in **P** obeys. **S** obeys part (c) and therefor would be an armstrong relation in **P**. By definition **S** does not have a complete set of axioms in **P**.

Remark 4 : We can now justify our previous claim that conditional independence is not an armstrong relation neither in PB nor in PN. We present the following axiom that does not satisfy condition (c) of theorem 4 neither in PB ([5]) nor in PN (appendix):

$$I(\alpha, \varnothing, \beta) \ \& \ I(\alpha, \gamma, \beta) \ \rightarrow I \alpha, \varnothing, \gamma) \ or \ I(\gamma, \varnothing, \beta).$$

3. A COMPLETE AXIOMATIZATION OF MARGINAL DEPENDENCIES

Completeness theorem: Let Σ be a set of marginal statements closed under the following axioms:

Symmetry $\quad (x\ ,y) \rightarrow (y\ ,x)$ (5.a)

Decomposition $\quad (x\ ,yw) \rightarrow (x\ ,y)$ (5.b)

Mixing $\quad (x\ ,y)\ \&\ (xy\ ,w) \rightarrow (x\ ,yw)$ (5.c)

Then for every marginal statement $\sigma = (x\ ,y)$ there exist a probability model P_σ that obeys all statements in Σ but does not obey σ.

Proof: Let σ be an arbitrary marginal statement not in Σ. Without loss of generality assume that for every non-empty sets x' and y' obeying $x' \subseteq x$, $y' \subseteq y$ and $x' \cup y' \neq x \cup y$ we have $(x'\ ,y') \in \Sigma$. A marginal statement obeying this property, is called a *minimal dependency*. If $\sigma = (x\ ,y)$ is not a minimal dependency then find a minimal dependency $\sigma' = (x', y')$ where $x' \subseteq x$ and $y' \subseteq y$. Clearly, σ' always exists because $\subseteq$ is a partial order. Construct $P_{\sigma'}$ that satisfies Σ and violates $(x'\ ,y')$. Due the decomposition axiom (5.b) we have

$$\neg\,(x'\ ,y') \rightarrow \neg\,(x\ ,y)$$

thus $P_{\sigma'}$ violates σ. Therefor, pick $P_\sigma = P_{\sigma'}$.

Let $\sigma = (x\ ,y)$ be a minimal dependency where $x = \{x_1, x_2 \cdots x_l\}$, $y = \{y_1, y_2 \cdots y_m\}$ and all other variables appearing in Σ are $z = \{z_1, z_2 \cdots z_k\}$. Construct P_σ as follows: Let all variables except x_1, be independent fair coins and let

$$x_1 = \sum_{i=2}^{l} x_i + \sum_{i=1}^{m} y_j \quad (mod 2).$$

Clearly, P_σ has the product form:

$$P_\sigma(x_1, \ldots, x_l, y_1, \ldots, y_m, z_1, \ldots, z_k) = P_\sigma(x_1, \ldots x_l, y_1, \ldots y_m) \cdot \prod_{i=1}^{k} P_\sigma(z_i)$$

We first show that $\sigma = (x\ ,y)$ does not hold in P_σ. Instantiate x_1 to one and all other variables in $x \cup y$ to zero. Clearly for this assignment of values

$$P_\sigma(x_1 \cdots x_l, y_1 \cdots y_m) \neq P_\sigma(x_1 \cdots x_l) \cdot P_\sigma(y_1 \cdots y_m) \qquad (6)$$

because the LHS of equation (6) is equal to 0 where the RHS consists of a product of two non-zero quantities.

It is left to show that every statement in Σ holds in P_σ, or equivalently, that for an arbitrary marginal statement (u, v) we have:

$$(u, v) \in \Sigma \rightarrow P_\sigma(u, v) = P_\sigma(u) \cdot P_\sigma(v).$$

This is done by examining (u, v) for every possible assignment of variables to the sets u and v and showing that either $P_\sigma(u, v) = P_\sigma(u) \cdot P_\sigma(v)$ or that $(u, v) \notin \Sigma$.

Case 1: Either $u \cap (x \cup y) = \emptyset$ or $v \cap (x \cup y) = \emptyset$.
Assume $u \cap (x \cup y) = \emptyset$ (the other case is similar), then $u \subseteq z$. Therefor, $P_\sigma(u, v) = P_\sigma(u) \cdot P_\sigma(v)$.

Case 2: Both u and v include an element of $x \cup y$.

Case 2.1: $x \cup y \neq u \cup v - z$.

Projecting P_σ on the set $u \cup v$ results $P_\sigma(u, v) = \prod_{w_i \in u \cup v} P_\sigma(w_i)$

hence clearly $P_\sigma(u, v) = P_\sigma(u) \cdot P_\sigma(v)$.

Case 2.2: $x \cup y = u \cup v - z$

This is the only case for which (u, v) is definitely not in Σ.

Let $u = x'y'z'$, $v = x''y''z''$ were $x = x'x''$, $y = y'y''$ and $z' \cup z'' \subseteq z$. We continue by contradiction. Assume $(u, v) = (x'y'z', x''y''z'')$ belongs to Σ. Σ is close under decomposition. Therefore, $(x'y', x''y'') \in \Sigma$. To reach a contradiction we shall show that $\sigma = (x'x'', y'y'')$ must have been in Σ, contradicting our selection of σ. The proof uses the mixing axiom to infer σ from $(x'y', x''y'')$ by "pushing" all the x's to one side and all y's to the other side. We further assume that x', x'', y', y'' are non-empty sets. If some of these sets is empty, not all the derivations that follow need to be performed to reach the contradicting conclusion that $(x, y) \in \Sigma$.

(x', y') belongs to Σ because (x, y) is a minimal dependency. Due to the mixing axiom

$$(x', y') \;\&\; (x'y', x''y'') \rightarrow (x', y'x''y'')$$

we conclude that $(x', x''y) \in \Sigma$. Using the mixing axiom again,

$$(y, x'') \;\&\; (yx'', x') \rightarrow (y, x'x'')$$

leads to the conclusion that $(x, y) \in \Sigma$, contradiction. □

Remark 1: Axioms 1.a-1.c are also complete for PD^- because marginal independence is an armstrong relation in PD^- and because we can modify the construction of P_σ such that it will be an non-extreme model.

Remark 2: It is straight forward to show that symmetry, decomposition and

$$\textit{composition} \quad I(x,z,y) \ \& \ I(x,z,w) \ \rightarrow \ I(x,z,yw)$$

are complete for marginal independence in PN.

Remark 3: This theorem shows that axioms 1.a-1.c are weakly complete for PB. However I believe that they are not complete for PB. For the proof I need to find a non Horn axiom that holds in PB (verma [?]). **Regardless of the completeness in PB,** axioms 1.a-1.c are powerful enough to solve the membership problem in PB. Thus an algorithm that solves the membership problem in PD would work also for PB. This is why weakly complete is considered "complete" in database and our completeness is called "strongly complete".

4. THE SOUNDNESS OF GRAPHICAL REPRESENTATIONS

Undirected graphs have been used ([1]) as a graphical representation of probabilistic dependencies. This representation scheme uses the device of vertex separation as a machinery for capturing the dependencies embedded in a probabilistic model.

In undirected graphs we define

$I(x, z, y)_G$ iff Removing the vertexes z from G renders x and y on two disconnected components.

An undirected graph G is said to be a *perfect-map* of a probabilistic model P if

$$P(x, y \mid z) = P(x \mid z) \cdot P(y \mid z) \quad \textit{iff} \quad I(x, z, y)_G$$

Clearly not every probabilistic model can be perfectly represented by an undirected graph. The next theorem, however, states that the converse is true, i.e every undirected graph can be perfectly represented by a probabilistic model.

Soundness theorem: For every undirected graph G, there exists a probability model P such that

$$P(x, y \mid z) = P(x \mid z) \cdot P(y \mid z) \quad \textit{iff} \quad I(x, z, y)_G$$

Proof: Let Σ be the set of all statements that hold in G. For every statement $\sigma \notin \Sigma$ we construct a probabilistic model P_σ that satisfies Σ and does not satisfy σ. Note that the statements in Σ are the only statements obeyed by all the P_σ's. Let P be $\oplus\{P_\sigma \mid \sigma \notin \Sigma\}$. P obeys exactly the statements in Σ, therefore P satisfies the requirements of the theorem.

Let $\sigma = I(x, z, y)_G$ be an arbitrary statement not in Σ. From the properties of $I(x, z, y)_G$, there exist an element $\alpha \in x$ and $\beta \in y$ such that $\sigma' = I(\alpha, z, \beta)_G \notin \Sigma$. Construct P_σ such that Σ and $\neg\sigma'$ are realized. Due the decomposition axiom for probabilistic models, P_σ does not satisfy σ. The construction of P_σ follows:

$I(\alpha, z, \beta)_G$ does not hold in G, therefore, there is a path bypassing z between α and β. Assume this path is $(r_1 = \alpha, r_2, \ldots, r_n = \beta)$. Let R be the set of all nodes on the path (α, β) and $S = \{s_1, s_2, \ldots, s_k\}$ be all the nodes in G not appearing in R. In P_σ all the variables are binary, and obey the distribution

$$P_\sigma(r_1, \ldots, r_n, s_1, \ldots, s_k) = K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{j=1}^{k} f(s_j)\right] \tag{7}$$

where

$$f(s, t) = \begin{cases} a & \text{if } s=0\ t=0 \\ \tfrac{1}{2}-a & \text{if } s=0\ t=1 \\ \tfrac{1}{2}-a & \text{if } s=1\ t=0 \\ a & \text{if } s=1\ t=1 \end{cases} \qquad f(s) = \begin{cases} \tfrac{1}{2} & \text{if } s=0 \\ \tfrac{1}{2} & \text{if } s=1 \end{cases}$$

and K is a normalizing factor

$$K^{-1} = \sum_{(r_1, \ldots, r_n, s_1, \ldots, s_k) \in \{0,1\}^{n+k}} \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{j=1}^{k} f(s_j)\right] = (\tfrac{1}{2})^{n-2}$$

The value of a is determined such that $P_\sigma(\alpha, \beta \mid z) \neq P_\sigma(\alpha \mid z) \cdot P_\sigma(\beta \mid z)$. The following argument shows that such a value always exists. Form equation (7) it is clear that for every subset z of S we have $P_\sigma(\alpha, \beta \mid z) = P_\sigma(\alpha, \beta)$. Therefore, it is enough to satisfy $P_\sigma(\alpha, \beta) \neq P_\sigma(\alpha) \cdot P_\sigma(\beta)$. In fact it is enough to show that this inequality holds for one specific assignment of values, say, $\alpha = 0, \beta = 0$. A straight forward calculation reveals that

$$P_\sigma(\alpha = 0) = P_\sigma(\beta = 0) = K \cdot \sum_{(r_2 \cdots r_n) \in \{0,1\}^{n-1}} f(0, r_2) \cdot \left[\prod_{i=2}^{n-1} f(r_i, r_{i+1})\right] = K \cdot \tfrac{1}{2}^{n-1} = \tfrac{1}{2}$$

Hence it is left to show that for some value of a a we get

$$P_\sigma(\alpha = 0, \beta = 0) \neq \tfrac{1}{4}$$

However, $P_\sigma(\alpha = 0, \beta = 0)$ is a polynomial in a of degree $n-1$. Therefore there must be a value of a in the interval $(0, \tfrac{1}{2})$ that satisfies this inequality.

It is left to verify that P_σ satisfies all the statements in Σ, or equivalently that for every non-intersecting sets u, v, w we get:

$$I(u, w, v)_G \in \Sigma \rightarrow P(u, v \mid w) = P(u \mid w) \cdot P(v \mid w)$$

Case 1: $u \cap R = \emptyset$ (or $v \cap R = \emptyset$)

Assume $u \cap R = \emptyset$ (the other case is similar). An equivalent condition is that u is a subset of S. We show that

$$P_\sigma(u, v \mid w) = P_\sigma(u \mid w) \cdot P_\sigma(v \mid w).$$

In fact, we prove a stronger version of this equality, namely,

$$P_\sigma(u, v, w) = P_\sigma(u) \cdot P_\sigma(v, w). \tag{7.a}$$

From equation (7), by summing over all variables in R, it is evident that

$$P_\sigma(s_1, \ldots, s_k) = \prod_{j=1}^{k} f(s_j).$$

Therefore, for every $u \subseteq S$ we get,

$$P_\sigma(u) = \prod_{s_j \in u} f(s_j).$$

The proof of equation (7.a) follows. From equation (7),

$$P_\sigma(u, v, w) = \sum_{r_i, s_j \notin uvw} K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{j=1}^{k} f(s_j)\right]$$

By first summing over the s_j's, we may rewrite the last equation.

$$P_\sigma(u, v, w) = \sum_{r_i \notin uvw} K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{s_j \in uvw} f(s_j)\right]$$

Since u contains only s_j's and, in addition, u, v and w are disjoint, we may further modify the last equation to obtain equation (7.a).

$$P_\sigma(u, v, w) = \left[\prod_{s_j \in u} f(s_j)\right] \cdot \left[\sum_{r_i \notin uvw} K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{s_j \in vw} f(s_j)\right]\right]$$

$$= \left[\prod_{s_j \in u} f(s_j)\right] \cdot \left[\sum_{r_i, s_j \notin vw} K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{s_j \in vw} f(s_j)\right]\right]$$

$$= P_\sigma(u) \cdot P_\sigma(v, w).$$

Case 2: Both u and v contain an element of R

Case 2.1: There exists an element r_{i_1} in u and r_{i_2} in v for which no element r_{i^*} of w satisfies $i_1 < i^* < i_2$ (or $i_2 < i^* < i_1$).

$I(u, w, v)_G \notin \Sigma$ because no element of w blocks the path between r_{i_1} and r_{i_2}.

Case 2.2: For every element r_{i_1} in u and r_{i_2} in v there exists an element r_{i^*} in w that satisfies $i_1 < i^* < i_2$ (or $i_2 < i^* < i_1$).

In this case, we show that there exists two functions g, h such that

$$P_\sigma(u, v, w) = g(u, w) \cdot h(v, w). \qquad (9.a)$$

This equality is equivalent ([13]) to the standard definition of conditional independence, i.e. it is equivalent to asserting

$$P_\sigma(u, v \mid w) = P_\sigma(u \mid w) \cdot P_\sigma(v \mid w).$$

Let z be all the variables of R not appearing in uvw. Let z' and z'' be a partition of z. We prove a stronger version of equation (9.a), namely, that

$$P_\sigma(u, v, w, z) = g(u, w, z') \cdot h(v, w, z''). \qquad (9.b)$$

From equation (7) we have

$$P_\sigma(u, v, w, z) = K \cdot \left[\prod_{i=1}^{n-1} f(r_i, r_{i+1})\right] \cdot \left[\prod_{s_j \in uvw} f(s_j)\right]$$

The factoring of P_σ into g and h follows:

If $s_j \in u$ then $f(s_j)$ is a factor of g. If $\{r_i, r_{i+1}\} \cap u \neq \emptyset$ then $f(r_i, r_{i+1})$ is a factor of g. All other terms are factors of h. The restriction that an element of w immediates between every element of u and v assures that every factor $f(r_i, r_{i+1})$ of g, does not contain an element of v and every factor $f(r_i, r_{i+1})$ of h, does not contain an element of u. Hence this factorization of $P_\sigma(u, v, w, z)$ satisfies equation (9.b). □

5. FIXED CONTEXT STATEMENTS

Fixed context conditional statements (U-statements) are an interesting relation in non-extreme probability models (PD^-) mostly because a tractable algorithm is given for inferring all statements which are logically implied from an arbitrary set of U-statements. In this section we find a complete set of axioms for U-statements, both in PD and PD^-. For PD^- we translate these axioms to a tractable algorithm.

Theorem 5: For every set of U-statements Σ satisfying the axioms:

symmetry $\quad I(x, z, y) \rightarrow I(y, z, x) \qquad (11.a)$

weak union $I(x, z, yw) \rightarrow I(x, zy, w)$ (11.b)

Weak contraction $I(xy, z, w)$ & $I(x, zw, y) \rightarrow I(x, z, yw)$ (11.c)

there exists a probability model P obeying all statements in Σ and no other U-statement. Moreover, if Σ is closed also under

intersection $I(x, zw, y)$ & $I(x, zy, w) \rightarrow I(x, z, yw)$ (11.d)

then P is a non-extreme model.

Proof's main ideas: Let $\sigma = I(x, z, y)$ be an arbitrary U-statement not in Σ. Let x=$\{x_1, \cdots, x_k\}$, y=$\{y_1, \cdots, y_l\}$. and z=$\{z_1, \cdots, z_m\}$. Σ is closed under weak union. Therefore, without loss of generality, assume that σ is a *maximal dependency* i.e., for every non-empty sets x', x'', y', y'' such that $x' \cup x'' = x$ and $y' \cup y'' = y$ the statement $I(x', zx''y'', y')$ is in Σ. The construction of P_σ follows:

$$P_\sigma(x_1,...,x_k, y_1,...,y_k) = \begin{cases} \frac{1}{2} & \text{if } (\underline{x}, \underline{y}) = \underline{0} \\ \frac{1}{2} & \text{if } (\underline{x}, \underline{y}) = \underline{1} \\ 0 & \textit{otherwise} \end{cases}$$

where all other variable are independent fair coins. P_σ can be shown to realize Σ and $\neg\sigma$.

If Σ obeys also the intersection axiom then without loss of generality we may assume that $\sigma = I(\alpha, z, \beta)$ where α and β are single variables. Thus the following non-extreme model can serve as the required P_σ.

$$P_\sigma(\cdot) = f(\alpha,\beta) \cdot \prod_{z_i \in U-\alpha\beta} f(z_i) \ .$$

Remark: Weak contraction is implied by weak union and intersection therefore symmetry, weak union and intersection are complete for U-statements in PD^-.

Our task at hand is to find an efficient way to use axioms 11.a, 11.b and 11.d to infer all the U-statements that are logically implied from a given set Σ in the environment of non-extreme models. The constructive proof of the following lemma provides the desired algorithm. In fact, the algorithm uses an additional axiom

Decomposition $I(x, z, yw) \rightarrow I(x, z, y)$ (11.f)

That guarantees that every statement (and not only U-statements) are inferred from Σ.

Lemma: Let Σ be a finite set of U-statements of size k and let Σ^* be the closure of Σ under symmetry, decomposition, intersection and weak union. Then, there exists a non-extreme probability model P that satisfies exactly the statements in Σ^*.

Proof: We first construct an undirected graph G that satisfies exactly the statements in Σ. The soundness of undirected graphs assures that a non-extreme probability model that is a perfect map of G, always exists. The construction of G is due to Paz ([6]) and is repeated for the integrity of this section.

Construct G, a graph over U by removing from the complete graph over U every edge (α,β), such that $\alpha \in x$, $\beta \in y$ for some statement $\sigma_i = I(x, z, y) \in \Sigma$, and only these edges. The constructed graph satisfies the requirements. This can be shown as follows:

Let $\sigma_1, \sigma_2 \cdots \sigma_k, \sigma_{k+1}, \cdots, \sigma_m$ be the list of all statements in Σ^* ordered in a way such that ant σ_i, $i > k$ is derived from previous statements in the list by one of the axioms. We can show now that every statement in the list is represented in the graph by finite induction: This proposition is clearly true for $\sigma_1, \cdots, \sigma_k$ and the truth for $j > k$ is implied by the fact that graphs satisfy the 4 axioms, and by the induction hypothesis. The other direction, namely that every statement satisfied by G belongs to Σ^* follows from [Pearl & Paz] (The minimal-edge I-mapness theorem). □

Remark: With slight variation of the proof of proposition 1 it can be shown that this lemma implies that G represents exactly all the statements logically implied by Σ in PD^-.

The algorithm is now clear. Given Σ, construct G. This requires $O(k \cdot n^2)$ time units, where k is the size of Σ and n is the number of variables. To verify if a specific statement belongs to Σ^* would require $O(n)$ time units.

COMMENT

In section 5 an algorithm was presented to find the closure of a set of U-statements in PD^-. The idea was to represent the closure in a compact way and then query this representation for any statement of interest. This comment shows that a similar approach for marginal statements is not feasible.

Using the methodology of [5], it is enough to show that there are $O(2^{2^n})$ different probabilistic models (with respect to marginal statements). Because then, a representation scheme must require, in average, exponential number of bits. Let U = $\{x_1, x_2, \cdots, x_n\}$ be all the variables. Consider the following set of marginal statements: $B = \{(x_1, C) \mid C \textit{ contains exactly } \lfloor n/2 \rfloor \textit{ elements}\}$. This set contains $O(2^n)$ marginal statements. For each truth assignment T of the statements of B we construct a probabilistic model P that realizes T. Let POS be the statements that are assigned a truth value under T and let NEG be the statements that are assigned a false value. Let $\sigma = (x_1, x_{i_1}, \cdots, x_{i_k})$ be an arbitrary marginal statement in NEG. Construct a probability model as follows: Let all the variables x_i $i = 2..n$ be independent fair coins and let

$$x_1 = \sum_{j=1}^{k} x_{i_j} \quad (mod 2)$$

This model P_σ clearly satisfies every marginal statement in B except σ. Hence the model $\oplus\{P_\sigma \mid \sigma \in NEG\}$ satisfies the truth assignment T. There are $O(2^{2^n})$ truth assignments for B hence at least that many different probabilistic models exist.

Hence solving the membership problem by specifying a *generic model* for the closure is not feasible. The question now is if there is any other polynomial solution for this seamingly easy problem. It might be still possible that given an arbitrary set of marginal statements Σ and a single marginal statement σ to answer the query is σ implied by Σ in polynomial time. The question is only HOW ?

REFERENCES

[1] J Pearl. Markov and Bayes networks: A comparison of two Graphical Representations of Probabilistic Knowledge. UCLA TR R-46, October 86.

[2] J Pearl, A Paz. Graphoids: A graph based logic for reasoning about relevance relations. TR R-53-L-1 ,UCLA, CSD 850038, April 1986.

[3] J Pearl, T Verma. The logic of representing dependencies by directed graphs. TR R-79-II, UCLA, CSD 870004, April 1987.

[4] T Verma. Causal Networks: Semantics and Expressiveness. TR R-65, Cognitive Systems Laboratory, UCLA, 1987.

[5] T Verma. Some Mathematical Properties of Dependency Models. TR R-xxx, Cognitive Systems Laboratory, UCLA, 1987.

[6] A Paz. On the intersection Property in Undirected Graphs. (In preparation).

[7] Y Sagiv, S Walecka. Subset Dependencies and Completeness result for a subset of EMVD. JACM 1982.

[8] Beeri, Fagin & Howard. A complete axiomatization of Functional Dependencies and Multi Valued Dependencies in Database relations. Proc 1977 ACM SIGMOD Int. Conf. on Mgmt of Data,Toronto, Canada pp 47-61.

[9] Fagin R. Multivalued dependencies and a new normal form for relational database. TODS 2,3 (Sept 1977),262-278.

[10] Fagin R. Horn clauses and database dependencies. Proc. ACM SIGACT Symp. on Theory of Computing, Los Angeles, April 1980.

[11] Fagin R. Functional dependencies in a relational database and propositional logic. IBM J. Res Dev. 21,6 (Nov 1977), 534-544.

[12] Beeri C. On the membership problem for functional and multivalued dependencies in relational database. ACM Trans. Database Syst. 5,3 (Sepy 1980), 241-249.

[13] S.L Lauritzen. Lectures on contingency Tables, 2nd Ed, University of Aalborg Press, Aalborg, Denemark, 1982.