Computer Science Department Technical Report
Cognitive Systems Laboratory
University of California
Los Angeles, CA 90024-1596

LOGICAL AND ALGORITHMIC PROPERTIES OF INDEPENDENCE AND THEIR APPLICATION TO BAYESIAN NETWORKS

Dan Geiger
Judea Pearl

June 1989
CSD-890035

Submitted to "Annals of Mathematics and AI," Special Issue on Statistics and AI.

TECHNICAL REPORT
R-123
February 1989

LOGICAL AND ALGORITHMIC PROPERTIES OF INDEPENDENCE AND THEIR APPLICATION TO BAYESIAN NETWORKS *

Dan Geiger & Judea Pearl

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

ABSTRACT

This paper establishes a partial axiomatic characterization of the predicate $I(X, Z, Y)$, to read "X is conditionally independent of Y, given Z". The main aim of such characterization is to facilitate a solution of the *implication* problem namely, deciding whether an arbitrary independency statement $I(X, Z, Y)$ *logically follows* from a given set Σ of such statements. In this paper, we provide a *complete axiomatization* and efficient algorithms for deciding implications in the case where Σ is limited to one of four types of independencies: *marginal* independencies, *fixed context* independencies, a *recursive* set of independencies or a *functional* set of independencies. The recursive and functional sets of independencies are the basic building blocks used in the construction of *Bayesian Networks*. For these models, we show that the implication algorithm can be used to efficiently identify which propositions are relevant to a task at hand at any given state of knowledge. We also show that conditional independence is an *Armstrong relation* [10], i.e, checking *consistency* of a mixed set of independencies and dependencies can be reduced to a sequence of implication problems. This property also implies a strong correspondence between conditional independence and graphical representations: for every undirected graph G there exists a probability distribution P that exhibits all the dependencies and independencies embodied in G.

Key words: Bayesian Networks, Conditional independence, Graphoids, Graphical representation, Probabilistic reasoning.

*This work was partially supported by the National Science Foundation Grant #IRI-8610155. "Graphoids: A Computer Representation for Dependencies and Relevance in Automated Reasoning (Computer Information Science)".

1. Introduction

The role of conditional independence stems from several practical considerations. First, knowledge about independencies saves space when storing an explicit distribution function (e.g., by a table), and saves time when computing and updating the probability of an event. For example, in the extreme case of representing the distribution of n independent binary variables, ignoring the independencies would require an explicit table of 2^n entries, and to calculate $P(x \text{ is } true)$ would require a summation over the other $n-1$ variables in the table. Recognizing the independencies among the variables enables us to replace the table by n parameters and to reduce computations to a single operation. Second, if we choose to represent and process random variables by networks [28], then the topology of such networks must reflect the conditional independencies that govern the variables in the domain and, therefore, the set of transformations [33] we are permitted to apply to the networks must, likewise, be determined by the rules that govern conditional independence. Third, in eliciting probabilistic models from human experts, dependencies among variables can often be asserted qualitatively while numerical assessments are subject to a great deal of uncertainties. Obtaining a direct representation scheme for judgements about dependencies would guard the model builder from assigning numerical values that lead to conceptually unrealistic dependencies. An example of such representation schemes are graphical dependency models such as Undirected Graphs (UG, section 3) [18,21,26], Directed Acyclic Graphs (DAG, section 4) [16,19,22,28,33], or Recursive models [20,41].

In this paper we concentrate on probabilistic conditional independence (although some results extend to other definitions of independence) because of its role in probabilistic reasoning systems (such as [1,4,7]). Information about dependencies can be specified by a list of *independency statements* (or simply *statements)* of the form $I(X, Z, Y)_P$ where, X, Y and Z are three finite disjoint sets of variables and $I(X, Z, Y)_P$ stands for "X is conditionally independent of Y, given Z", or equivalently,

$$I(X, Z, Y)_P \iff P(x, y \mid z) = P(x \mid z) \cdot P(y \mid z)$$

in any instantiation x, y and z of the variables X, Y and Z, for which $P(z)>0$. (By convention, we assume that $I(X, Z, \varnothing)$ always holds). When an independency statement does not hold we say it is a *dependency*.

Our objection is to answer the following two questions:

1. Given a set of independence statements, is the set *redundant ?* i.e., are some of the statements implied by the others ?

2. Given a mixed set of independence and dependence statements, is the set *consistent ?* i.e, could they be realized simultaneously by some probability distribution ?

Both problems are expressible in terms of a sequence of *implication problems*. The implication problem is to decide whether an independency statement σ *logically follows* from a set of such statements Σ, namely, to decide whether σ holds in every distribution that obeys all statements in Σ. To answer the first question we simply select each statement σ in Σ and check whether σ logically follows from $\Sigma - \{\sigma\}$. The answer to the second problem is covered in section 2.

The most common strategy for solving the implication problem involves a two step process [2,3,8]; First finding a *complete* set of inference rules called *axioms,* and second, finding an efficient algorithm to repeatedly apply these axioms to determine whether σ is derivable from Σ. Completeness guarantees that, by repeated application of the axioms, each statement σ that logically follows from Σ will eventually be derived. Examples of axioms for conditional independence are [28]:

Symmetry $\quad I(X, Z, Y) \Rightarrow I(Y, Z, X)$ (1.a)

Decomposition $\quad I(X, Z, Y \cup W) \Rightarrow I(X, Z, Y)$ (1.b)

Weak union $\quad I(X, Z, W \cup Y) \Rightarrow I(X, Z \cup W, Y)$ (1.c)

Contraction $\quad I(X, Z, Y) \;\&\; I(X, Z \cup Y, W) \Rightarrow I(X, Z, Y \cup W)$ (1.d)

These axioms are used in [30,31] to generalize the concept of independence; Any ternary predicate obeying axioms (1.a) through (1.d) is called a *semi-graphoid* [31]. Other examples of semi-graphoids include Partial Correlation [5,30], Ordinal Conditional Functions [17,37], Embedded Multivalued Dependencies [9], Qualitative Independence [34], Vertex Separation [30], d-separation [26] and D-separation [15] (hence the name semi-graphoid). Similar axioms for probabilistic independence were used by Dawid [6], Spohn [36], and Smith [35].

In this paper we carry out these two steps for four classes of independency statements: marginal independence (Section 2), fixed context independence (Section 3) and both a *recursive* and a *functional* set of independence (Section 4)(for which axioms (1) are shown to be complete). The general implication problem for unrestricted sets of conditional independence statements Σ, remains unsolved. The results reported in [11,23,38] suggest that a finite complete set of axioms for conditional independence does not exists. This however, does not exclude a possible solution to the implication problem because a non-axiomatizable set of statements can still admit an efficient implication algorithm (e.g., [32]), but the existence of such solution is less likely [39].

2. Some Completeness Results

The following notations are employed; σ, possibly scripted, denotes an independency statement, Σ denotes a set of independency statements and $\mathbf{P}$ denotes a class of probability distributions, for example strictly positive distributions (PD^+), normal distributions (PN), distributions over binary variables (PB) and the class of all probability distributions (PD).

Definition: An *axiom*

$$\sigma_1 \;\&\; \sigma_2 \;\& \;\cdots\; \& \;\sigma_n \Rightarrow \sigma$$

is *sound* for $\mathbf{P}$ if every distribution $P \in \mathbf{P}$ that obeys the antecedents of the axiom also obeys σ. Axioms (1.a) through (1.d) are examples of sound axioms in PD.

Definition: σ is *logically implied (logically follows)* by Σ, denoted $\Sigma \models_{\mathbf{P}} \sigma$, iff every distribution in $\mathbf{P}$ that obeys Σ also obeys σ. $\Sigma \vdash_{\mathbf{A}} \sigma$ iff $\sigma \in \mathbf{cl}_{\mathbf{A}}(\Sigma)$, i.e., there exists a *derivation chain* $\sigma_1, ..., \sigma_n = \sigma$ such that for each σ_j, either $\sigma_j \in \Sigma$, or σ_j is derived by an axiom in $\mathbf{A}$ from the previous statements.

Definition: A set of axioms $\mathbf{A}$ is *sound* for $\mathbf{P}$ iff for every statement σ and every set of statements Σ

$\Sigma \vdash_{\mathbf{A}} \sigma$ *only* if $\Sigma \models_{\mathbf{P}} \sigma$

The set **A** is *complete* for **P** iff

$$\Sigma \vdash_{\mathbf{A}} \sigma \text{ if } \Sigma \models_{\mathbf{P}} \sigma .$$

Proposition 1: A set of axioms is sound for **P** iff each axiom in the set is sound for **P**.

The proof is achieved by induction on the length of a derivation.

Proposition 2 (After Fagin [8]): A set of axioms A is complete iff for every set of statements Σ and every statement $\sigma \notin \mathbf{cl}_{\mathbf{A}}(\Sigma)$ there exists a distribution P_σ in **P** that satisfies Σ and does not satisfy σ.

Proof: This is the counter-positive form of the completeness definition, if $\sigma \notin \mathbf{cl}_{\mathbf{A}}(\Sigma)$ (i.e., $\Sigma \nvdash_{\mathbf{A}} \sigma$) then $\Sigma \not\models_{\mathbf{P}} \sigma$. □

Next we present a complete set of axioms for the class of marginal independency statements.

Definition: A *marginal statement* (or *marginal independency)* is an independency statement $I(X, Z, Y)$ where Z is $\varnothing$ i.e.,

$$I(X, \varnothing, Y)_P \quad \textit{iff} \quad P(X\ Y) = P(X) \cdot P(Y)$$

for each instantiation of X and Y. The inequality $P(X\ Y) \neq P(X) \cdot P(Y)$ is called a *marginal dependency.*

Theorem 3 (Completeness for marginal independence) [12]: Let Σ be a set of marginal statements, and let $\mathbf{cl}(\Sigma)$ be the closure of Σ under the following axioms:

Symmetry	$I(X, \varnothing, Y)_P \rightarrow I(Y, \varnothing, X)_P$	(2.a)
Decomposition	$I(X, \varnothing, Y \cup W)_P \rightarrow I(X, \varnothing, Y)_P$	(2.b)
Mixing	$I(X, \varnothing, Y)_P \ \& \ I(X \cup Y, \varnothing, W)_P \rightarrow I(X, \varnothing, Y \cup W)_P$	(2.c)

Then for every marginal statement $\sigma = I(X, \varnothing, Y)_P \notin \mathbf{cl}(\Sigma)$ there exists a probability distribution P_σ that obeys all statements in $\mathbf{cl}(\Sigma)$ but does not obey σ.

Proposition 4: Axioms (2) are sound for PD (i.e., holds for all distributions).

Theorem 3 and Proposition 4 guarantee that by repeatedly applying axioms (2.a) through (2.c) on a set of marginal statements Σ, any marginal statement σ that logically follows from Σ will eventually be derived and, conversely, any marginal statement that is derivable, logically follows from Σ. Paz [12,24] provides an efficient algorithm to check whether a marginal statement σ is derivable from Σ. The complexity of his algorithm is $O(|\Sigma| \cdot n^2)$ where n is the number of distinct variables in $\Sigma \cup \{\sigma\}$.

A complete set of axioms does not provide sufficient means for deriving all the information that is implied by a given set of statements. For example, assume that the set $\Sigma = \{ I(X, Z \cup \{\alpha\}, Y), I(X, Z, Y)\}$ is given, where α is a single variable and all variables are bi-valued i.e., drawn from PB. It can be shown [28] that the disjunction $I(X, Z, \alpha)$ *or* $I(Y, Z, \alpha)$ logically follows from Σ and, yet, it cannot be derived by a complete set of axioms. Such a set only guarantees to reveal (correctly) that neither of the disjuncts is logically implied by Σ but would not show that one of the two statements must hold. To obtain all disjunctions, a strongly complete set of axioms is needed.

Definition (after [3,8]): A set of axioms **A** is *strongly complete* in a class of distributions **P**, if for every set of statements Σ and for every set of single statements $\{\sigma_i \mid i = 1, \cdots, n\}$ the following relation holds:

$$\Sigma \models_{\mathbf{P}} \sigma_1 \; or \; \cdots \; or \; \sigma_n \quad iff \quad \Sigma \vdash_{\mathbf{A}} \sigma_1 \; or \; \cdots \; or \; \sigma_n$$

Similar to Prposition 2, the following holds:

Proposition 5 (After [3,8]): A set of axioms **A** is *strongly complete* iff for every set of statements Σ closed under axioms **A**, there exists a distribution P in **P** that satisfies all statements in Σ and none other.

Clearly, a complete set of axioms is strongly complete but the converse is not always true [8]. The notion of *Armstrong relation* [10] provides a condition under which strong completeness is equivalent to completeness.

Definition: Conditional independence is an *Armstrong relation* in a class of distributions **P** if there exists an operation $\otimes$ that maps finite sequences of distributions of **P** into a distribution of **P**, such that if σ is a conditional independence statement and if P_i i=1..n are distributions in **P**, then σ holds for

$\otimes\{P_i \mid i=1..n\}$ iff σ holds for each P_i.

We concentrate on two families of distributions **P**: All distributions, denoted PD and strictly positive distributions, denoted PD^+.

Theorem 6 [14]: Conditional independence is an Armstrong relation in PD and in PD^+.

The operation $\otimes$ is realized by the direct product.

Corollary (Strong completeness for marginal independence) : For every set of marginal statements Σ there exists a distribution P that satisfies all statements in the closure of Σ under axioms (2) and none other.

Proof: Let P be $\otimes\{P_\sigma \mid \sigma \notin \Sigma\}$ where P_σ are the distributions guaranteed by Theorem 3. P satisfies all marginal statements in $\mathbf{cl}_A(\Sigma)$ because each P_σ satisfies these statements. P does not satisfy any $\sigma_0 \notin \mathbf{cl}_A(\Sigma)$ because P_{σ_0} does not satisfy σ_0. □

An immediate application of strong completeness is the reduction of the *consistency problem* to a set of implication problems.

Definition: A set of marginal dependencies Σ^- and a set of marginal independencies Σ^+ are consistent iff there exists a distribution that satisfies $\Sigma^+ \cup \Sigma^-$. The task of deciding whether a set is consistent is called the *consistency problem.*

The following algorithm answers whether $\Sigma^+ \cup \Sigma^-$ is consistent: For each member of Σ^- determine, using the implication algorithm, whether its negation logically follows from Σ^+. If the answer is negative for all members of Σ^-, then the two sets are consistent, otherwise they are inconsistent.

The correctness of the algorithm stems from the fact that if the negation of each member σ of Σ^- does not follow from Σ^+ i.e, each member of Σ^- is individually consistent with Σ^+, then there is a distribution P_σ that realizes Σ^+ and $\neg\sigma$. The distribution $P = \otimes\{P_\sigma \mid \neg\sigma \in \Sigma^-\}$ then realizes both Σ^+ and Σ^-, therefore the algorithm correctly identifies that the sets are consistent. In the other direction, namely when the algorithm detects an inconsistent member of Σ^-, the decision is obviously correct.

3. A Graph-Based Closure Algorithm

In this section we concentrate on another class of statements called U-statements.

Definition: Let U be a set of variables (the universe). A statement $I(X, Z, Y)$ is called a U-statement if $X \cup Y \cup Z = U$. A set of U-statements is called *fixed-context.*

The interest in this set of statements stems from two reasons. First, for the class of strictly positive distributions, U-statements obey the same axioms as vertex separation in undirected graphs. Second, there exists a compact representation of all independence statements that logically follow (in PD^+) from a given set of U-statements, requiring only $O(|U|^2)$ bits of storage. These two properties rely on an additional axiom of independence that holds in every strictly positive distribution:

$$\textit{Intersection} \qquad I(X, Z \cup Y, W) \ \& \ I(X, Z \cup W, Y) \Rightarrow I(X, Z, Y \cup W) \qquad (3)$$

(That intersection does not hold in all distributions can be seen by examining the case, $X = Y = W$ and $Z = \emptyset$).

To exemplify the representation of probabilistic knowledge by undirected graphs, consider a language governed by a Markov process, namely, the probability of the i-th letter is determined solely by the (i-1)-th letter via $P(l_i \mid l_{i-1}) > 0$. The dependencies embedded in the distribution function can be represented by the Markov chain of Figure 1.

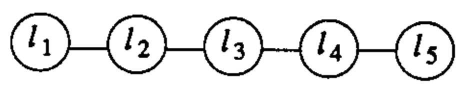

Figure 1

This graph asserts, for example, that the variables l_1 and l_3 are conditionally independent given l_2, since the node l_2 blocks all paths from l_1 to l_3. More generally, for every two disjoint sets X, Y and Z of nodes in a graph G, let us define the predicate $I(X, Z, Y)_G$ by

$$I(X, Z, Y)_G <=> Z \text{ separates } X \text{ from } Y \text{ in } G.$$

We then say that G perfectly represents the dependencies of P if there exists a 1-1 correspondence between the variables in P and the vertices of G such that,

$$I(X, Z, Y)_G <=> I(X, Z, Y)_P.$$

Such a graph is called a *perfect-map* of P [28].

Suppose that by sampling 5-letter words from some unknown language (see previous example), the following two independencies (U-statements) were identified:

$$\Sigma = \{ \; I(\{l_1, l_2\}, l_3, \{l_4, l_5\}), \; I(l_3, \{ l_2 \; l_4 \}, \{ l_1 \; l_5 \}) \; \}$$

The question arises: Are these statement sufficient to guarantee the Markov nature of the language and, moreover, is the chain structure a complete representation of all independencies that logically follow from Σ ? The main result of this section is a polynomial-time algorithm that generates all independency statements that logically follow (in PD^+) from a set of U-statements Σ i.e., it generates all statements that hold in every strictly-positive distribution which obeys Σ. In particular it reveals that Σ implies the Markov nature of the language via

$$P(l_i \mid l_{i-1}, l_{i-2}, \cdots, l_1) = P(l_i \mid l_{i-1}) \qquad i = 2, \ldots, 5$$

or, in our notation, that $I(l_i, l_{i-1}, \{l_{i-2} \cdots l_1\})$ must hold for $i = 2,\ldots,5$ and, moreover, that no other statement is logically implied by Σ. This *closure algorithm*, generates from Σ the graph of Figure 1. It starts with a complete graph (over all the variables) and simply deletes every edge (α,β) for which a statement of the form $I(X \cup \{ \alpha \}, Z, Y \cup \{ \beta \})$ is found in Σ, for some X, Y and Z. We show that the resulting graph represents all the independencies that are shared by all strictly-positive distributions that obey Σ, and must therefore, be obeyed by the language. Thus the two statements of Σ constitute a sufficient code for the chain structure of the distribution P, and the algorithm uncovers this structure without resorting to numerical calculations. The next two theorems justify this algorithm.

Theorem 7 [14]: For every undirected graph G, there exists a non-extreme distribution P, and a 1-1 correspondence between the variables in P and the nodes of G such that for every three disjoint sets of nodes X, Y and Z the following holds:

$$I(X, Z, Y)_G \quad \textit{iff} \quad I(X, Z, Y)_P .$$

Theorem 8 (strong completeness) [14,25]: Let Σ be a set of U-statements and let $\mathbf{cl}(\Sigma)$ be the closure of Σ under symmetry (1.a), decomposition (1.b), weak union (1.c) and intersection (3). Then, 1) there exists an undirected graph G for which $I(X, Z, Y)_G$ satisfies exactly the statements in $\mathbf{cl}(\Sigma)$ and 2) there exists a strictly positive distribution P that satisfies exactly the statements in $\mathbf{cl}(\Sigma)$ (i.e., G is a perfect map of P). G is constructed by removing from the complete graph over U every edge (α,β), such that $\alpha \in X$, $\beta \in Y$ for some statement $\sigma = I(X, Z, Y) \in \Sigma$, and only these edges.

The closure algorithm is now clear: given Σ, it constructs G by the procedure of Theorem 8 and uses $I(X, Z, Y)$ to identify the elements of $\mathbf{cl}(\Sigma)$. The construction of G requires $O(k \cdot n^2)$ steps, where k is the size of Σ and n is the number of variables, while to verify if a specific statement belongs to $\mathbf{cl}(\Sigma)$ requires $O(n)$ steps. The simplicity of this algorithm stems from the fact that axioms (1.a) through (1.c) and (2) are complete for fixed context statements both when interpreted as vertex separation (Theorem 8, part 1) as well as when interpreted as conditional independence in PD^+ (Theorem 8, part 2).

Theorem 7 is important by itself because it justifies the use of undirected graphs as a representation scheme for probabilistic dependencies. It allows one to choose any UG for representing dependencies and be guaranteed that the model is supported by probability theory (similar results for Directed Acyclic Graphs are presented in [13]). The converse, however, does not hold; there are many distributions that do not have a perfect representation neither in DAGs nor in UGs and the challenge remains to devise graphical representations that minimize this deficiency.

The construction presented in the proof of Theorem 7 [14] leads to a rather complex distribution, where the domain of each variable is unrestricted. It still does not guarantee that a set of dependencies and independencies represented by UGs is realizable in a more limited class of distributions such as normal or those defined on binary variables. We conjecture that these two classes of distributions are sufficiently rich to permit the consistency of undirected graph representations.

4. Application to Bayesian Networks

A Bayesian network encodes properties of a probability distribution using a Directed Acyclic Graph (DAG). Each node i in a Bayesian network corresponds to a variable X_i, a set of nodes I correspond to to a set of variables X_I and x_i is a value from the domain of X_i. Each node in the network is regarded as a storage cell for the distribution $P(x_i \mid x_{\pi(i)})$ where $X_{\pi(i)}$ is a set of variables that correspond to the parent nodes $\pi(i)$ of i. The distribution represented by a Bayesian network is composed via

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{\pi(i)}) \tag{4}$$

(when i has no parents, then $X_{\pi(i)} = \emptyset$). The role of a Bayesian network is to record a state of knowledge P, to provide means for updating the knowledge as new information is accumulated and to facilitate query answering mechanisms for knowledge retrieval [22,27]. A standard query for a Bayesian network is to find the current belief distribution of a hypothesis variable X_I, given a composed evidence set $X_J = x_J$ i.e., to compute $P(x_I \mid x_J)$ for each value of X_I and for a given combination of values of X_J. We examine the following related problem: Given a variable X_k, a Bayesian network D and the task of computing $P(x_I \mid x_J)$ determine, without resorting to numeric calculations, whether the answer to the query is sensitive to the value of X_k.

This question is an instance of the implication problem because it amounts to verifying whether the statement $I(X_I, X_J, X_k)$ logically follows from the set of independencies that define the topology of the DAG. Let the nodes (and variables) be arranged in a total ordering d that agrees with the directionality of the DAG namely, i must precede j whenever there is a link from i to j. Let $U(i)$ stand for the set of nodes that precede i in the ordering d. The rule of composition (4) implicitly encodes a set of equalities $P(x_i \mid x_{\pi(i)}) = P(x_i \mid x_{U(i)})$, $i = 1..n$, or in our notations, a set of n independence statements $I(X_i, X_{\pi(i)}, X_{U(i)} - X_{\pi(i)})$. This set of statements, denoted L, is said to *define* a Bayesian network and is called a *recursive basis*. Thus, the problem at hand is to determine whether $I(X_I, X_J, X_k)$ logically follows from L. If it does, then $P(x_I \mid x_J) = P(x_I \mid x_J, x_k)$ for all instantiations x_I, x_k and x_J, hence the value of X_k will not affect the computation. If the statement does not follow, then the value of X_k may have an effect on $P(x_I \mid x_J)$.

Theorem 9 [13,40]: An independency statement σ logically follows from a recursive basis L iff σ can be derived from L using axioms (1).

Theorem 9 provides a characterization of all statements that logically follow from L, whereas [29] also establishes a compact representation of these statements, based on a graphical criteria called *d-separation.* [15] employs d-separation to facilitate a linear time (in the number of links) implication algorithm for inputs forming a recursive basis. This result is analogous to Theorem 8, since it provides a compact representation of the closure of L (via a DAG) and a linear time algorithm to determine whether a statement belongs to the closure.

Theorem 9 assumes that L contains only predecessor independencies of the form $I(X_i, X_{\pi(i)}, X_{U(i)}-X_{\pi(i)})$. Occasionly, however, we are in possession of stronger forms of independence relationships, in which case additional independencies should be read of the DAG. A common example is the case of a variable that is functionally dependent on its corresponding parents in the DAG (*deterministic variable,* [33]). The existence of each such variable X_i could be encoded in L by a statement of *global* independence $I(X_i, X_{\pi(i)}, X_U-X_{\pi(i)}-X_i)$ asserting that conditioned on $X_{\pi(i)}$, X_i is independent of **all** other variables, not merely of its predecessors ($X_{U(i)}$). A set combining predecessor independencies and global independencies is called a *functional basis.*

Theorem 10 [15]: An independency statement σ logically follows from a functional basis L iff σ can be derived from L using axioms (1).

[15] shows that the independencies implied by a functional basis can be identified in linear time using an enhanced version of d-separation, named *D-separation.*

ACKNOWLEDGEMENTS

We thank Ron Fagin for pointing out the usefulness of the notion of Armstrong relation and Norman Dalkey, Azaria Paz and Thomas Verma for many useful discussions on the properties of dependency models.

REFERENCES

[1] S. Andreassen, M. Woldbye, B. Falck and S.K. Andersen, MUNIN - A causal probabilistic network for interpretation of electromyographic findings, *Proc. 10th International Joint Conference on AI, (IJCAI-87)*, Milan, Italy, 1987, pp. 366-372.

[2] C. Beeri, On the membership problem for functional and multivalued dependencies in relational database, *ACM, transactions on Database systems 5,3*, Sept. 1980, pp. 241-249.

[3] C. Beeri, R. Fagin and J.H. Howard, A complete axiomatization of functional dependencies and multi valued dependencies in database relations, *Proc. ACM SIGMOD Int. Conf. on Mgmt of Data*, Toronto, Canada pp 47-61, 1977.

[4] M. Ben-Basat, R.W. Carlson, V.K. Puri, E. Lipnic, L.D. Portigal and M.H. Weil, Pattern-based interactive diagnosis of multiple disorders: The MEDAS system, *IEEE transactions on Pattern Analysis & Machine Intelligence, PAMI* 2:148-160, 1980.

[5] H.M. Cramér, *Mathematical methods of statistics*, Princeton University Press, Princeton, N.J, 1946.

[6] A.P. Dawid, Conditional independence in statistical theory (with discussion), *Journal of the Royal Statistical Society B*, 41,1979, 1-31.

[7] R.O. Duda, P.E. Hart, P. Barnett, J. Gaschnig, K. Konolige, R. Reboh and J. Slocum, Development of the PROSPECTOR consultant system for mineral exploration, Final Report for SRI Projects 5821 and 6915, Artificial Intelligence Center, SRI International, 1978.

[8] R. Fagin, Functional dependencies in a relational database and propositional logic, *IBM J.* Res Dev. 21,6, 1977, pp. 534-544.

[9] R. Fagin, Multivalued dependencies and a new form for relational Databases, *ACM, transactions on Database systems*, Vol. 2, No. 3, 1977, pp. 262-278.

[10] R. Fagin, Horn clauses and database dependencies, *JACM*, Vol. 29, No. 4, 1982, pp. 952-985.

[11] D. Geiger, Towards the formalization of informational dependencies, (M.Sc. thesis). UCLA, Cognitive Systems Laboratory, Technical Report R-102, Dec. 1987.

[12] D. Geiger, A. Paz and J. Pearl, Axioms and algorithms for inferences involving probabilistic independence, UCLA, Cognitive System Laboratory, Technical Report R-119, 1988, submitted to *information and computation.*

[13] D. Geiger and J. Pearl, On the logic of causal models, in: *Proc. 4th workshop on uncertainty in AI*, Minneapolis, Minnesota, Aug. 1988, pp. 136-147.

[14] D. Geiger and J. Pearl, Logical and algorithmic properties of conditional independence and qualitative independence, UCLA, Cognitive Systems Laboratory, Technical Report R-97, march 1988.

[15] D. Geiger and, T. Verma and J. Pearl, *Identifying independence in Bayesian networks*, UCLA, Cognitive Systems Laboratory, Technical Report R-116, 1989.

[16] R. Howard and J. E. Matheson, 1981, Influence diagrams, in *Principles and applications of Decision Analysis*, Menlo Park, CA: Strategic Decision Group.

[17] D. Hunter, Graphoids, semi-graphoids, and ordinal conditional independence, in preparation, 1988.

[18] V. Isham, An introduction to spatial point processes and Markov random fields, *International Statistical Review*, Vol. 49, 1981, pp. 21-43.

[19] D. A. Kenny, *Correlation and causality*. New York: Wiley, 1979.

[20] H. Kiiveri, T.P. Speed and J.B. Carlin, Recursive causal models, *Journal of Australian Math. Society*, Vol. 36, 1984, pp. 30-52.

[21] S. L. Lauritzen, *Lectures on contingency tables*, 2nd Ed., University of Aalborg Press, Aalborg, Denemark, 1982.

[22] S. L. Lauritzen and D. Spiegelhalter, Local computation with probabilities on graphical structures, and their application to expert systems (with discussion), *Journal Royal Statistical Society*, B 50, pp. 157-244, 1988.

[23] D. S. Parker and K. Parsaye-Ghomi, Inferences involving embedded multivalued dependencies and transitive dependencies, in: *Proc. 1980 ACM SIGMOD Int. Conf. on Mgmt of Data*, pp. 52-57.

[24] A. Paz, Membership algorithm for marginal independence UCLA, Cognitive Systems Laboratory, Technical Report R-117, September 1988.

[25] A. Paz, A full characterization of pseudographoids in terms of families of undirected graphs, UCLA, Cognitive Systems Laboratory, Technical Report R-95, September 1987.

[26] J. Pearl, Markov and Bayes networks: A comparison of two graphical representations of probabilistic knowledge, UCLA, Cognitive System Laboratory, Technical Report R-46, October 1986. Also in [28] (Chapter 3).

[27] J. Pearl, Fusion, propagation and structuring in belief networks, *Artificial Intelligence*, 29, (3):241-288, September 1986.

[28] J. Pearl, *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, Morgan Kaufmann: San Mateo, 1988.

[29] J. Pearl, D. Geiger and T. Verma, The logic of Influence Diagrams, *Proc.*, Conference on Influence Diagrams for Decision Analysis, Inference, and Prediction. Berkeley, California, May 1988. Also a shorter version, in: *Kybernetica*, Vol. 25, 1989.

[30] J. Pearl and A. Paz. Graphoids: A graph-based logic for reasoning about relevance relations, *Advances in Artificial Intelligence-II*, Ed. B. Du Boulay et al. North-Holland Publishing Co., pp. 357-363, 1987.

[31] J. Pearl and T. Verma, The logic of representing dependencies by directed acyclic graphs, *Proc. AAA-I, Seattle Washington, pp. 374-379, July 1987.*

[32] Y. Sagiv, S. Walecka, Subset dependencies and completeness result for a subset of EMVD, *JACM Vol. 29, No. 1, Jan 1982, pp. 103-117.*

[33] R. Shachter, Probabilistic inference and influence diagrams, *Operations Research*, Vol. 36, pp. 589-604, 1988.

[34] P. Shenoy, G. Shafer, Propagating belief functions in qualitative Markov trees. *International Journal of Approximate Reasoning*, Vol. 1, No. 4, pp. 349-400.

[35] J. Q. Smith, Influence diagrams for statistical modeling, Technical Report 117, Department of Statistics, University of Warwick (Coventry, U.K.).

[36] W. Spohn, Stochastic independence, causal independence, and shieldability, *Journal of Philosophical Logic 9*, 1980, pp. 371-396.

[37] W. Spohn, Ordinal conditional functions: A dynamic theory of epistemic states. *Causation in Decision, Belief Change, and Statistics, II*, W. L. Harper and B. Skyrms (eds.), 1988, pp. 105-134.

[38] M. Studeny, Attempts at axiomatic description of conditional independence, *Workshop on uncertainty in expert systems*, Alsovice, Chechoslovakia, June 20-23, 1988.

[39] M. Vardi, Fundamentals of dependency theory, in *Trends in Theoretical Computer Science* ed. E. Borger, pp. 171-224, Computer Science Press, Rockvile, Md.,1988.

[40] T. Verma and J. Pearl, Causal Networks: Semantics and Expressiveness, in: *Proc. 4th workshop on uncertainty in AI*, Minneapolis, Minnesota, Aug. 1988, pp. 352-358.

[41] N. Wermouth and S.L. Lauritzen, Graphical and recursive models for contingency tables, *Biometrika*, Vol. 70, 1983, pp. 537-552.