TECHNICAL REPORT
R-119
March 1991

Reprinted from INFORMATION AND COMPUTATION
All Rights Reserved by Academic Press, New York and London

Vol. 91, No. 1, March 1991
Printed in Belgium

Axioms and Algorithms for Inferences Involving Probabilistic Independence*

DAN GEIGER, AZARIA PAZ†, AND JUDEA PEARL

*Cognitive Systems Laboratory, Computer Science Department,
Los Angeles, California 90024-1596*

This paper offers an axiomatic characterization of the probabilistic relation "X is independent of Y (written (X, Y))," where X and Y are two disjoint sets of variables. Four axioms for (X, Y) are presented and shown to be *complete.* Based on these axioms, a polynomial *membership* algorithm is developed to decide whether any given *independence statement* (X, Y) *logically follows* from a set Σ of such statements, i.e., whether (X, Y) holds in every probability distribution that satisfies Σ. The complexity of the algorithm is $O(|\Sigma| \cdot k^2 + |\Sigma| \cdot n)$, where $|\Sigma|$ is the number of given statements, n is the number of variables in $\Sigma \cup \{(X, Y)\}$, and k is the number of variables in (X, Y). © 1991 Academic Press, Inc.

1. INTRODUCTION

Consider a collection of information sources, each reflecting a different aspect of some underlying probabilistic phenomenon. These sources can be regarded as a set of random variables, governed by an unknown probability distribution, some of which are dependent and some independent. In this paper, we are concerned with the following problem: Assume we know that some groups of variables are mutually independent, either by statistical analysis or by conceptual understanding of the underlying phenomenon; we need to infer new independencies without resorting to additional measurements or expensive numerical analysis.

We formalize this question as follows: Let an (*independence*) *statement* be a sentence of the form (X, Y), where $X = \{x_1, ..., x_n\}$ and $Y = \{y_1, ..., y_n\}$ are disjoint finite sets of variables. The meaning of (X, Y) is that X and Y are probabilistically independent, namely,

$$P(X, Y) = P(X) \cdot P(Y)$$

* This work was supported in part by the National Science Foundation Grant IRI-8821444.
† Permanent affiliation: Computer Science Department, Technion, Haifa, Israel 32000.

0890-5401/91 $3.00

which is a short-hand notation for the equality

$$\Pr\{x_1 = a_1, ..., x_n = a_n, y_1 = b_1, ..., y_m = b_m\}$$
$$= \Pr\{x_1 = a_1, ..., x_n = a_n\} \cdot \Pr\{y_1 = b_1, ..., y_n = b_m\}$$

for every choice of $a_1, ..., a_n, b_1, ..., b_m$ from the domains of $x_1, ..., x_n$, $y_1, ..., y_m$, respectively. We often refer to independence statements as *independencies* and to their negation as *dependencies*. We say that a statement σ is logically implied by a set of statements Σ iff every distribution that satisfies Σ satisfies σ as well. We ask: Is the statement $\sigma = (X, Y)$ logically implied by a given set Σ of such statements, each characterized by a different pair of subsets X' and Y'?

The answer is given in two steps. First, (in Section 2), we characterize the relation of independence by the following inference rules, considered as axioms:

Trivial independence	$(X, \varnothing)$	(1.a)
Symmetry	$(X, Y) \rightarrow (Y, X)$	(1.b)
Decomposition	$(X, Y \cup Z) \rightarrow (X, Y)$	(1.c)
Mixing	$(X, Y) \,\&\, (X \cup Y, Z) \rightarrow (X, Y \cup Z)$	(1.d)

These axioms clearly hold for all distributions and therefore are *sound*. For example, to prove (1.d), we observe that $P(X, Y) = P(X) \cdot P(Y)$ and $P(X, Y, Z) = P(X, Y) \cdot P(Z)$ imply that $P(X, Y, Z) = P(Y) \cdot P(X) \cdot P(Z)$. Moreover, summing over X yields $P(Y, Z) = P(Y) \cdot P(Z)$, hence $P(X, Y, Z) = P(X) \cdot P(Y, Z)$ which establishes the right-hand side of (1.d). We show that these axioms are also *complete*, i.e., capable of deriving by repeated applications all independencies that are logically implied by the input set of independencies.

The second step of our solution is a membership algorithm that efficiently answers whether a statement σ is a member of the closure $\mathbf{cl}(\Sigma)$ of Σ under axioms (1). In light of the soundness and completeness results, $\sigma \in \mathbf{cl}(\Sigma)$ iff σ holds in every distribution that satisfies Σ. This step is covered in Section 3. Section 4 extends the results to the problem of deciding consistency: Given a set of independence statements mixed with dependence statements, decide if the set is consistent, i.e., if there exists a probability distribution that satisfies all the statements simultaneously. We show that the consistency problem can be translated into a sequence of membership problems.

Similar problems of membership and axiomatic characterization are treated in the literature on database dependencies, for example (Beeri *et al.*, 1977; Fagin, 1977; Beeri, 1980; Fagin, 1982; Sagiv and Walecka, 1982). Our notations and definitions were particularly influenced by (Beeri *et al.*, 1977; Fagin, 1977). A survey on database dependency theory can be found in (Rissanen, 1978; Fagin and Vardi, 1986; Vardi, 1988). An extension of this work to conditional independence has been developed in (Geiger and Pearl, 1987; Pearl, 1988; Pearl *et al.*, 1988; Geiger, 1990; Geiger *et al.*, 1990).

2. Axiomatic Characterization

The following symbols are used: σ for a statement, Σ for a set of statement, $\mathbf{A}$ for a set of axioms, and $\mathbf{P}$ for a class of distributions. For example, the class of all probability distributions will be denoted by *PD*, the class of normal distributions by *PN*, and the class of distributions over bi-valued variables by *PB*. A distribution in *PN* is characterized by the multivariate density function of the form

$$f(X)=\frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}}\exp\left\{-\frac{1}{2|\Lambda|}(x-m)\,\Lambda^{-1}(X-m)^t\right\},$$

where X is a vector of variables, $m=E[X]$ is a vector of averages, and $\Lambda=E[(X-m)(X-m)^t]$ is the covariance matrix. A distribution $P(X)$ in *PB* is any joint distribution of the set of variables X in which each $x_i \in X$ has a domain of size two (e.g., $\{0, 1\}$). We assume variable symbols are drawn from a finite set $\mathbf{U}=\{u_1, u_2, ...\}$. We use the letters x, y, z, u, v, w, possibly subscripted, for variables, and X, Y, Z, U, V, W for sets of variables. The set union symbol is often dropped and XY is written instead of $X \cup Y$.

Definition. σ is *logically implied* by Σ, denoted $\Sigma \models_{\mathbf{P}} \sigma$, iff every distribution in $\mathbf{P}$ that satisfies Σ also satisfies σ. σ is said to be *derivable* from Σ, denoted $\Sigma \vdash_{\mathbf{A}} \sigma$ 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 in* $\mathbf{P}$ iff for every statement σ and every set of statements Σ

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

The set $\mathbf{A}$ is *complete* for $\mathbf{P}$ iff

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

PROPOSITION 1. *Axioms* (1) *are sound for PD* (*i.e., hold for all distributions*).

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

PROPOSITION 2 (After Fagin, 1977). *A set of axioms* $\mathbf{A}$ *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* $\mathbf{P}$ *that satisfies* Σ *and does not satisfy* σ.

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

THEOREM 3 (Completeness). *Let* Σ *be a set of statements, and let* $\mathbf{cl}(\Sigma)$ *be the closure of* Σ *under the following axioms*:

Trivial independence	$(X, \varnothing)$	(1.a)
Symmetry	$(X, Y) \rightarrow (Y, X)$	(1.b)
Decomposition	$(X, YW) \rightarrow (X, Y)$	(1.c)
Mixing	$(X, Y)\ \&\ (XY, W) \rightarrow (X, YW)$	(1.d)

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

Remark. The sets X, Y, Z, and W may be empty (e.g., the derivation of $(\varnothing, YW)$ from $(\varnothing, Y)$ and (Y, W) by the mixing axiom is permitted although $(\varnothing, YW)$ can be derived directly by axiom (1.a)).

Proof. Let $\sigma = (X, Y)$ be an arbitrary statement not in $\mathbf{cl}(\Sigma)$. Without loss of generality we assume that for all non-empty sets X' and Y' obeying $X' \subseteq X$, $Y' \subseteq Y$, and $X'Y' \neq XY$ we have $(X', Y') \in \mathbf{cl}(\Sigma)$. A statement obeying this property is called a *minimal statement*. If $\sigma = (X, Y)$ is not a minimal statement then we can always find a minimal statement $\sigma' = (X', Y')$ not in $\mathbf{cl}(\Sigma)$, where $X' \subseteq X$ and $Y' \subseteq Y$, by deleting elements of X and Y until we obtain the desired property or until both X' and Y' become singletons, in which case, due to the trivial independence axiom (1.a), σ' is a minimal statement. For each such σ', we construct $P_{\sigma'}$ that satisfies $\mathbf{cl}(\Sigma)$ and violates σ'. Due the decomposition axiom (1.c), which holds for all distributions, we know that any distribution that violates σ', violates σ as well. In particular, $P_{\sigma'}$ violates σ (while satisfying $\mathbf{cl}(\Sigma)$) and therefore satisfies the conditions of the theorem.

Let $\sigma = (X, Y)$ be a minimal statement, where $X = \{x_1, x_2, ..., x_l\}$, $Y = \{y_1, y_2, ..., y_m\}$, and let $Z = \{z_1, z_2, ..., z_k\}$ stand for the rest of the variables, namely, $\mathbf{U} - XY$. Construct P_σ as follows: Let all variables, except x_1, be independent binary variables with probability $\frac{1}{2}$ for each of their two values (e.g., fair coins), and let

$$x_1 = \sum_{i=2}^{l} x_i + \sum_{j=1}^{m} y_j \pmod 2.$$

Clearly, P_σ has the product form:

$$P_\sigma(XYZ) = P_\sigma(XY) \cdot \prod_{z_i \in Z} P_\sigma(z_i). \tag{2}$$

We first show that $\sigma = (X, Y)$ does not hold in P_σ. Instantiate x_1 to one and all other variables in XY to zero. For this assignment of values we have

$$P_\sigma(x_1 \cdots x_l, y_1 \cdots y_m) \neq P_\sigma(x_1 \cdots x_l) \cdot P_\sigma(y_1 \cdots y_m) \tag{3}$$

because the LHS of Eq. (3) is equal to 0, whereas the RHS consists of a product of two non-zero quantities.

It is left to show that every statement in **cl**(Σ) holds in P_σ, or equivalently, that for an arbitrary statement (V, W) we have

$$(V, W) \in \mathbf{cl}(\Sigma) \Rightarrow P_\sigma(V, W) = P_\sigma(V) \cdot P_\sigma(W).$$

This is done by examining the statement (V, W) for every possible assignment of variables to the sets V and W and showing that either $P_\sigma(V, W) = P_\sigma(V) \cdot P_\sigma(W)$ or that $(V, W) \notin \mathbf{cl}(\Sigma)$.

Case 1. Either V or W contains only elements of Z. By Eq. (2), we obtain $P_\sigma(V, W) = P_\sigma(V) \cdot P_\sigma(W)$.

Case 2. Both V and W include an element of $X \cup Y$.

Case 2.1. $V \cup W$ does not include all the variables of $X \cup Y$. To verify whether (V, W) holds in P_σ amounts to checking this statement in the projection of P_σ on the set $V \cup W$. Since the probability of every value assignment to a proper subset $S \subsetneqq X \cup Y$ is $(\frac{1}{2})^{|S|}$, this projection assumes the product form $\prod_{w_i \in V \cup W} P_\sigma(w_i)$. Hence, again, $P_\sigma(V, W) = P_\sigma(V) \cdot P_\sigma(W)$.

Case 2.2. $V \cup W$ includes all elements of $X \cup Y$. This is the only case for which (V, W) is definitely not in **cl**(Σ). Let $V = X'Y'Z'$, $W = X''Y''Z''$, where $X = X'X''$, $Y = Y'Y''$, and $Z'Z'' \subseteq Z$. We continue by contradiction. Assume $(V, W) = (X'Y'Z', X''Y''Z'')$ is in **cl**(Σ). **cl**(Σ) is closed under decomposition. Therefore, $(X'Y', X''Y'') \in \mathbf{cl}(\Sigma)$. To reach a contradiction we show that this statement implies that σ must have been in **cl**(Σ), contradicting our selection of σ. The proof uses the mixing and symmetry axioms to infer $(X'X'', Y'Y'')$ (i.e., σ) from $(X'Y', X''Y'')$ by "pushing" all the X's to one side and all Y's to the other side. The following is a derivation of σ.

First, (X', Y') belongs to $\mathbf{cl}(\Sigma)$ because (X, Y) is a minimal statement. Due to the mixing axiom

$$(X', Y') \,\&\, (X'Y', X''Y'') \rightarrow (X', Y'X''Y'').$$

We conclude that $(X', X''Y) \in \mathbf{cl}(\Sigma)$. Due to symmetry $(X''Y, X') \in \mathbf{cl}(\Sigma)$ as well. $(X'', Y) \in \mathbf{cl}(\Sigma)$ because σ is a minimal statement and therefore (by symmetry) also (Y, X'') is a member of $\mathbf{cl}(\Sigma)$. Using the mixing axiom again, we obtain

$$(Y, X'') \,\&\, (YX'', X') \rightarrow (Y, X'X'')$$

which leads to the conclusion that $(Y, X) \in \mathbf{cl}(\Sigma)$, and by symmetry that (X, Y) is in $\mathbf{cl}(\Sigma)$, a contradiction (note that the derivation of σ remains valid when some of X', X'', Y', and Y'' are empty, as long as $X = X'X''$ and $Y = Y'Y''$). ∎

Theorem 3 implies that the problem of verifying $\Sigma \models_P \sigma$ is *decidable* because, for a finite set of variables $\mathbf{U}$, the process of generating $\mathbf{cl}(\Sigma)$ by successive application of axioms (1.a) through (1.d) will always terminate. We note that the construction of P_σ uses bi-valued variables, therefore, axioms (1.a) through (1.d) are complete also in PB, namely a statement is derivable iff every distribution in PB that satisfies Σ also satisfies σ.

THEOREM 4 (Completeness in PN) (Geiger and Pearl, 1987). *Let Σ be a set of statements, and let $\mathbf{cl}(\Sigma)$ be the closure of Σ under the following axioms*:

Trivial independence	$(X, \varnothing)$	(4.a)
Symmetry	$(X, Y) \rightarrow (Y, X)$	(4.b)
Decomposition	$(X, YW) \rightarrow (X, Y)$	(4.c)
Composition	$(X, Y) \,\&\, (X, W) \rightarrow (X, YW)$	(4.d)

Then there exists a normal distribution $P \in PN$ that satisfies all statements in $\mathbf{cl}(\Sigma)$ and none other.

Axioms (4) are stronger than axioms (1) since the mixing axiom can be derived from axioms (4), but composition cannot be derived from axioms (1). Composition, clearly does not hold in all distributions. For example, letting x and y be the outcomes of two fair coins and $z = x + y \pmod 2$ yields a distribution where z is independent of each coin separately, yet it is completely determined by the joint outcome of x and y. However, composition holds for every normal distribution and Theorem 4 shows that adding this property is sufficient to render the axioms complete for *PN*. For normal distributions the membership algorithm is trivial; (V, W) is logically implied by Σ iff for each $v_i \in V$ and $w_j \in W$ there exists a statement (X, Y) in Σ such that v_i appears in X and w_j appears in Y (or vice versa, $w_j \in X$ and $v_i \in Y$); its complexity is $O(n^2 \cdot |\Sigma|)$. The next section provides a membership algorithm for axioms (1.a) through (1.d), having similar complexity.

3. The Membership Algorithm

The following notation is employed. σ and γ denote single statements, Σ and Γ sets of statements, and s a set of elements (variables). $\gamma = (X, Y)$ is *trivial* if either X or Y is empty. The notation span(γ) stands for the set of elements represented in a statement γ, and similarly, span(Γ) denotes the set of elements represented in all the statements of Γ; for example, span$(\{(x_1, x_2), (x_1, x_3)\})$ is $\{x_1, x_2, x_3\}$. The *projection* of γ on s, denoted $\gamma(s)$, is the statement derived from γ by removing all elements not in s from γ, e.g., if $\gamma = (x_1 x_2 x_3, x_4 x_5)$ then $\gamma(x_1 x_2 x_3) = (x_1 x_2 x_3, \emptyset)$ and $\gamma(x_1 x_3 x_4 x_6) = (x_1 x_3, x_4)$. Similarly, the projection of Γ on s, denoted $\Gamma(s)$, stands for $\{\gamma(s) \mid \gamma \in \Gamma\}$. The number of elements appearing in γ is denoted by $|\gamma|$ and is called the *size* of γ. The membership algorithm, presented below, uses the procedure *Find* to answer whether a statement σ is derivable from Σ by axioms (1.a) through (1.d).

Algorithm Membership.

Procedure Find (Σ, σ):

1. $\Sigma' := \Sigma(\text{span}(\sigma))$ $\{\Sigma'$ is the projection of Σ on the variables of the target statement $\sigma\}$
2. If σ is trivial, or σ (or its symmetric image) belongs to Σ' then set Find$(\Sigma, \sigma) :=$ True and return.
3. Else if for all nontrivial $\sigma' \in \Sigma'$, span$(\sigma') \neq$ span(σ) then set Find$(\Sigma, \sigma) :=$ False.
4. Else there exists a statement $\sigma' \in \Sigma'$ such that span$(\sigma') =$ span(σ), and up to symmetry, $\sigma' = (AP, BQ)$ and $\sigma = (AQ, BP)$, where one of the sets A, B, P, Q may be empty (if several such σ' exist, then choose one arbitrarily).

Set $\sigma_1 := (A, P)$, $\sigma_2 = (B, Q)$,
Find(Σ, σ) := Find$(\Sigma', \sigma_1) \wedge$ Find(Σ', σ_2).
return.
Begin {Membership}
    Input(Σ, σ)
    Print Find(Σ, σ)
End.

We will first show that the algorithm is correct and then prove its complexity.

LEMMA 5. *$\Sigma \vdash \sigma$ iff $\Sigma' \vdash \sigma$, where $\Sigma' = \Sigma(s)$ and $s = \text{span}(\sigma)$.*

Proof. Assume that $\Sigma \vdash \sigma$. Then there is a derivation chain for σ, $\gamma_1, \gamma_2, ..., \gamma_k$ with each γ_i either in Σ, or derived from previous statements in the chain by one of the axioms. Consider an arbitrary statement γ_j in the chain. If γ_j is derived from γ_k, $k < j$ by symmetry or decomposition, then $\gamma_j(s)$ is derived from $\gamma_k(s)$ by symmetry or decomposition, respectively. Similarly, if γ_j is derived from γ_k and γ_l by the mixing axiom, then $\gamma_j(s)$ is derived from $\gamma_k(s)$ and $\gamma_l(s)$ by the mixing axiom. It follows that $\gamma_1(s) \cdots \gamma_k(s)$ is a derivation chain for $\sigma = \sigma(s)$ in $\Sigma' = \Sigma(s)$.

Assume now that $\Sigma' \vdash \sigma$, then clearly $\Sigma \vdash \sigma$, because all the statements in Σ' can be derived from statements in Σ by decomposition. ∎

LEMMA 6. *For any nontrivial statement σ, $\Sigma' \vdash \sigma$ only if there is a statement $\sigma' \in \Sigma'$ such that* $\text{span}(\sigma) = \text{span}(\sigma')$.

Proof. The span of any statement in Σ' is included in, or equal to, the span of σ. If no statement in Σ' has the same span as the span of σ, then the derivation of σ is impossible; an inspection of the axioms in (1) shows that no axiom can add variables to the span of a derived statement. ∎

Lemma 5 shows that to derive a statement σ from Σ one may start, without loss of generality, by projecting all statements in Σ on the span of σ. Thus justifies Step 1 and 2 of the procedure used by the algorithm. Step 3 stems from the fact that if there exists no statement σ' with the same span as σ, then by Lemma 6, σ cannot be derived. Hence, Find(Σ, σ) is correctly set to False. Step 4 is justified by Lemma 7 and Theorem 8.

LEMMA 7. *Let $\sigma = (AQ, BP)$, $\sigma' = (AP, BQ)$, $\sigma_1 = (A, P)$, $\sigma_2 = (B, Q)$ be statements. If $\sigma' \in \mathbf{cl}(\Sigma)$ then $\sigma \in \mathbf{cl}(\Sigma)$ iff $\sigma_1 \in \mathbf{cl}(\Sigma_1)$ and $\sigma_2 \in \mathbf{cl}(\Sigma_2)$, where $\Sigma_i = \Sigma(\text{span}(\sigma_i))$ (notice that σ, σ', σ_1, and σ_2 are defined as in Step 4 of Procedure Find).*

Proof. If $\sigma' \in \mathbf{cl}(\Sigma)$, $\sigma_i \in \mathbf{cl}(\Sigma_i)$ $i = 1, 2$, then σ can be derived as follows:

(i) $(AP, BQ) = \sigma'$; $(A, P) = \sigma_1$; $(B, Q) = \sigma_2$

(ii) (A, PBQ): Apply the mixing axiom (1.d) on σ_1 and σ'

(iii) (APB, Q): Apply the mixing and decomposition axioms (1.b), (1.d) on σ_2 and σ'

(iv) (PB, Q): Apply the decomposition axiom (1.c) on (iii)

(v) $(AQ, BP) = \sigma$: Apply the symmetry, decomposition, and mixing axioms (1.b), (1.c), (1.d) on (ii) and (iv).

If $\sigma \in \mathbf{cl}(\Sigma)$ then let $\gamma_1, \gamma_2, ..., \gamma_k = \sigma = (AQ, BP)$ be a derivation chain for σ in Σ. Let $s = \text{span}((\sigma_1))$. Then $\gamma_1(s), \gamma_2(s), ..., \gamma_k(s) = \sigma_1 = (A, P)$ is a derivation chain for σ_1 in Σ_1. Thus, $\sigma_1 \in \mathbf{cl}(\Sigma_1)$. Similarly, a derivation chain for σ_2 can be constructed. ∎

Lemma 7 shows that the selection of σ' in Step 4 can be made arbitrarily because any selection provides a necessary and sufficient means to check whether σ belongs to $\mathbf{cl}(\Sigma)$.

THEOREM 8. *The procedure in the algorithm halts and when it halts*

$$\text{Find}(\Sigma, \sigma) = \text{true} \qquad \textit{iff} \quad \sigma \in \mathbf{cl}(\Sigma).$$

Proof. Every time the algorithm passes through Step 4 the size of the statements involved strictly decrease. If it did not halt before, it will halt when the size of the two statements have reached the value 2 (at Step 2 or 3). We show correctness by induction on the size of σ. If $|\sigma| = 1$ then σ is trivial, $\sigma \in \mathbf{cl}(\Sigma)$, and $\text{Find}(\Sigma, \sigma) = \text{true}$. If $|\sigma| = 2$ then $\sigma \in \mathbf{cl}(\Sigma)$ iff $\text{Find}(\Sigma, \sigma) = \text{true}$ as follows from Steps 2 and 3 of the algorithm.

Assume that the theorem holds for all $|\gamma| < k$ and let σ be a statement such that $|\sigma| = k$ and $\sigma = (AQ, BP)$. Then $\text{Find}(\Sigma, \sigma) = \text{true}$ iff (by the definition of Step 4) $\text{Find}(\Sigma', \sigma_1) = \text{true}$ and $\text{Find}(\Sigma', \sigma_2) = \text{true}$ iff (by the definition of Step 1) $\text{Find}(\Sigma_1, \sigma_1) = \text{true}$ and $\text{Find}(\Sigma_2, \sigma_2) = \text{true}$, where $\Sigma_i = \Sigma(\text{span}(\sigma_i))$, respectively, iff (by induction) $\sigma_i \in \mathbf{cl}(\Sigma_i)$ $i = 1, 2$ iff (by Lemma 7) $\sigma \in \mathbf{cl}(\Sigma)$. ∎

Next we analyze the time complexity. We measure the complexity in terms of basic operations of two types: comparison of two statements and a projection of a statement. Both operations are bounded by n, the number of distinct variables in $\Sigma \cup \{\sigma\}$. Let $\text{Cost}(k)$ be the number of basic operations needed to solve a size k problem, where $k = |\sigma|$ and assume (initially) that $\text{span}(\sigma) = \text{span}(\Sigma)$. By Step 4, $\text{Cost}(k)$ must satisfy the equation:

$$\text{Cost}(k) \leqslant \text{Cost}(k_1) + \text{Cost}(k_2) + |\Sigma|,$$

where $k_1 + k_2 = k$, $k_1 = |\sigma_1|$, and $k_2 = |\sigma_2|$. The solution to this equation is

$O(|\Sigma| \cdot k)$ measured in basic operations. Adding the cost of projecting Σ over the variables of σ which is $O(|\Sigma| \cdot n)$, yields the theorem below.

THEOREM 9. *The complexity of the membership algorithm is* $O(|\Sigma| \cdot k^2 + |\Sigma| \cdot n)$ *(which is* $O(|\Sigma| \cdot n^2)$*, since* $k \leqslant n$*).*

Remarks. 1. It is reasonable to assume that the bound is pessimistic at least in its $|\Sigma|$ part, since as the algorithm proceeds the number of statements in Σ' decreases.

2. The algorithm can be slightly modified so as to produce a derivation chain for σ if $\sigma \in \mathbf{cl}(\Sigma)$, whose length is $O(k)$.

3. The algorithm can be expanded into a polynomial algorithm (provided that $|\Sigma|$, and $|\Gamma|$ are polynomial) for the following problems:

a. Given Σ and Γ, is $\mathbf{cl}(\Sigma) = \mathbf{cl}(\Gamma)$, or is $\mathbf{cl}(\Sigma) \subseteq \mathbf{cl}(\Gamma)$?

b. Minimize the size of Σ while preserving $\mathbf{cl}(\Sigma)$: To solve this problem start with a maximal size statement and probe all statements derivable from it. In each step add an additional statement in a non-increasing order (by size) that has not been previously probed, and probe all statements derivable from the augmented set.

4. EXTENSIONS

In this section, we show that probabilistic independence enjoys an interesting property known in the literature as having *Armstrong models.* The concept of Armstrong models has evolved in the theory of relational databases and has been stated in rather general terminology that makes it applicable to probabilistic independence (Fagin, 1982). We will use this property to show that axioms (1) characterize the independence relation in a stronger sense than that defined in Section 2. We will employ this characterization to check whether a given mixture of independencies and dependencies is consistent.

Fagin's general setting consists of a class of *models* (which in our case is a class of probability distributions), a class of *sentences* $\mathscr{S}$ (for our purposes independence statements) and a relationship *Holds* that states whether a sentence holds in a given model. Holds(P, σ) means that σ *holds for* P or that P *satisfies* σ. We say that σ is a logical consequence of Σ, written $\Sigma \models_{\mathbf{P}} \sigma$, if every model that satisfies the set of sentences Σ satisfies the sentence σ as well. $\Sigma^* \triangleq \{\sigma \mid \Sigma \models_{\mathbf{P}} \sigma\}$. A set of sentences Σ is *consistent* if there exists a model that satisfies every sentence in Σ.

THEOREM 10 (Fagin, 1982). *Let $\mathscr{S}$ be a set of sentences. The following properties of $\mathscr{S}$ are equivalent*:

(a) Existence of a faithful operator. *There exists an operator* $\otimes$ *that maps nonempty families of models into models, such that if* σ *is a sentence in* $\mathscr{S}$ *and* $\langle P_i : i \in I \rangle$ *is a nonempty family of models, then* σ *holds for* $\otimes \langle P_i : i \in I \rangle$ *if and only if* σ *holds for each* P_i.

(b) Existence of Armstrong models. *Whenever* Σ *is a consistent subset of* $\mathscr{S}$ *and* Σ^* *is the set of sentences in* $\mathscr{S}$ *that are logical consequences of* Σ, *then there exists a model (an "Armstrong model") that obeys* Σ^* *and no other sentences in* $\mathscr{S}$.

(c) Splitting of disjunctions. *Whenever* Σ *is a subset of* $\mathscr{S}$ *and* $\{\sigma_i : i \in I\}$ *is a nonempty subset of* $\mathscr{S}$, *then* $\Sigma \models \bigvee \{\sigma_i : i \in I\}$ *if and only if there exists some* i *in* I *such that* $\Sigma \models \sigma_i$.

Fagin provides several applications for his theorem and this paper provides an additional one. We first show the existence of a faithful operator for probabilistic independence. We note that while the theorem holds for any cardinality of the index set I, we use it only for finite nonempty I.

THEOREM 11. *Let* $\{P_i | i = 1 \cdots n\}$ *be a finite set of distributions. There exists an operation* $\otimes$ *that maps finite sets of distributions to distributions such that for each independency* σ,

$$\sigma \textit{ holds for } \otimes \{P_i | i = 1 \cdots n\} \textit{ iff } \sigma \textit{ holds for every } P_i,\ i = 1 \cdots n. \tag{5}$$

We shall construct the operation $\otimes$ using a binary operation $\otimes'$ such that if $P = P_1 \otimes' P_2$ then for every independence statement σ we obtain

$$\sigma \text{ holds for } \otimes' P_i \qquad \text{iff} \qquad \sigma \text{ holds for } P_1 \text{ and for } P_2. \tag{6}$$

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

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

Clearly, if $\otimes'$ satisfies Eq. (6), then $\otimes$ satisfies Eq. (5). Therefore, it suffices to show that $\otimes'$ satisfies (6).

Let P_1 and P_2 be two distributions sharing the variables $x_1, ..., x_n$. Let $A_1, ..., A_n$ be the domains of $x_1, ..., x_n$ in P_1 and let $\alpha_1, ..., \alpha_n$ be an instantiation of these variables. Similarly, let $B_1, ..., B_n$ be the domains of $x_1, ..., x_n$ in P_2 and $\beta_1, ..., \beta_n$ an instantiation of these variables. Let the domain of $P = P_1 \otimes' P_2$ be the product domain $A_1 B_1, ..., A_n B_n$ and denote an instantiation of the variables of P by $\alpha_1 \beta_1, ..., \alpha_n \beta_n$. Define $P_1 \otimes' P_2$ by the following equation:

$$P(\alpha_1 \beta_1, \alpha_2 \beta_2, ..., \alpha_n \beta_n) = P_1(\alpha_1, \alpha_2, ..., \alpha_n) \cdot P_2(\beta_1, \beta_2, ..., \beta_n).$$

The proof that P satisfies Eq. (5) can be found in (Geiger and Pearl, 1987). It follows from the basic definition of independence and can be extended to conditional independence as well.

Theorem 10 suggests two alternative formulations of Theorem 11. Part (b) states that given a set of independence statements Σ, there exists a distribution that satisfies exactly all statements that logically follow from Σ and none other (note that any set of independence statements is consistent, which is a requirement of part (b)). Part (c) states that to check whether a disjunction of statements logically follows from Σ, it suffices to check each disjunct separately. Interestingly, if the class of models is taken to be the class of normal distributions then the construction of $\otimes$ is not adequate because $\otimes \{P_i \mid i = 1 \cdots n\}$ is not a normal distribution. However, it is possible to define another faithful operator $P = \oplus \{P_i \mid i = 1 \cdots n\}$ for normal distributions as follows: For each pair of variables x, y, let the correlation factor ρ_{xy} in P be zero iff ρ_{xy} is zero in every P_i, $i = 1 \cdots n$. All other correlation factors are assigned a non-zero quantity ρ small enough to assure that the covariance matrix of $\oplus P_i$ is positive definite. Finding a similar operation for PB remains an open problem.

We now apply Theorem 11 to show that axioms (1) characterize probabilistic independence. The proof repeats the argument that derives part (b) from part (a) in Theorem 10.

THEOREM 12 (Characterization of independence). *For every set of statements Σ closed under axioms* (1.a) *through* (1.d) *there exists a distribution P such that for each independency σ,*

$$\sigma \textit{ holds for } P \qquad \textit{iff} \qquad \sigma \in \Sigma$$

Proof. By Theorem 3, for each $\sigma \notin \Sigma$ there exists a distribution P_σ that satifies Σ and does not satisfy σ. Let $P = \otimes \{P_\sigma \mid \sigma \notin \Sigma\}$. The distribution P is well defined because the set of all statements that use a finite set of variables $\mathbf{U}$ is finite. Due to Eq. (5), P satisfies all statements in Σ and none other; hence P satisfies the requirements of the theorem.

The immediate consequence of these theorems is that axioms (1.a) through (1.d) are powerful enough to derive all disjunctions of independence statements that are logically implied by a given set of such statements and not merely single statements as advertized in Section 3 (see part (c) of Theorem 10).

Another application of Theorem 11 is the reduction of the *consistency problem* to a set of membership problems.

DEFINITION. A set of dependencies Σ^- and a set of 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 membership 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 $P = \bigotimes \{P_\sigma \mid \neg\sigma \in \Sigma^-\}$ 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.

ACKNOWLEDGMENTS

We are grateful to Ron Fagin for pointing out the usefulness of the notion of the Armstrong relation and for a thorough reading of the manuscript. We also thank Thomas Verma and Norman Dalkey for many discussions on the properties of independence.

RECEIVED January 23, 1989; FINAL MANUSCRIPT RECEIVED October 26, 1989

REFERENCES

BEERI, C., FAGIN, R., AND HOWARD, J. H. (1977), A complete axiomatization of functional dependencies and multi valued dependencies in database relations, *in* "Proceedings, 1977 ACM SIGMOD Int. Conf. on Mgmt of Data, Toronto, Canada," pp. 47–61.

BEERI, C. (1980), On the membership problem for functional and multivalued dependencies in relational database, *ACM Trans. Database Systems* **5**, No. 3, 241–249.

FAGIN, R. (1977), Functional dependencies in a relational database and propositional logic, *IBM J. Res. Develop.* **21**, No. 6, 534–544.

FAGIN, R. (1982), Horn clauses and database dependencies, *J. Assoc. Comput. Mach.* **29**, No. 4, 952–985.

FAGIN, R., AND VARDI, M. (1986), The theory of data dependencies—A survey, *in* "Mathematics of Information Processing," Proceedings, Sympos. Appl. Math., Vol. 34, pp. 19–72, Amer. Math. Soc., Providence, RI.

GEIGER, D. (1990), Graphoids: a qualitative framework for probabilistic inference, Ph.D. dissertation UCLA, Computer Science Department.

GEIGER, D., AND PEARL, J. (1987), "Logical and Algorithmic Properties of Conditional Independence," UCLA Cognitive Systems Laboratory, Technical Report 870056 R-97, August; shorter version R-97-I, *in* "Proceedings, 2nd International Workshop on Artificial Intelligence and Statistics, January 1989," forthcoming, "The Annals of Statistics."

GEIGER, D., VERMA, T., AND PEARL, J. (1990), Identifying independence in Bayesian networks, *Networks* **20**, 1–48.

PEARL, J. (1988), "Probabilistic reasoning in intelligent systems: Networks of plausible inference," Morgan Kaufmann, San Mateo, CA.

PEARL, J., GEIGER, D., AND VERMA, T. (1988), The logic of influence diagrams, *in* "Influence Diagrams, Belief Nets and Decision Analysis" (R. M. Oliver and J. Q. Smith, Eds.), Wiley, London; shorter version, *Kybernetica* **25**, No. 2 (1989), 33–44.

RISSANEN, J. (1978), Theory of relations for databases—A tutorial survey, *in* "Proceedings, 7th Symp. on Math. Found. of Comput. Science," Lecture Notes in Computer Sciences, Vol. 64, pp. 537–551, Springer-Verlag, New York/Berlin.

SAGIV, Y., AND WELECKA, S. (1982), Subset dependencies and completeness results for a subset of EMVD, *J. Assoc. Comput. Mach.* **29**, No. 1, 103–117.

VARDI, M. (1988), Fundamentals of dependency theory, *in* "Trends in Theoretical Computer Science" (E. Borger, Ed.), pp. 171–224, Computer Sci., Rockville, MD.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium