LEARNING HIDDEN CAUSES FROM EMPIRICAL DATA

Judea Pearl

December 1984
CSD-840065

R - 38
November 1984
CSD-840065
Revision I
August 1985

Learning Hidden Causes from Raw Data

Judea Pearl

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

from the Proceedings of the
1985 International Joint Conference on Artificial Intelligence
August 1985, Volume 1, pp. 567-572.

This work was supported in part by the National Science Foundation, Grant #DSR 83-13875, and by a UCLA-Hughes MICRO Grant, #HB840120

LEARNING HIDDEN CAUSES FROM EMPIRICAL DATA*

Judea Pearl

Computer Science Department, University of California at Los Angeles

ABSTRACT

Models of complex phenomena often consist of hypothetical entities called "hidden causes", which cannot be observed directly and yet play a major role in understanding, communicating, and predicting the dynamics of those phenomena. This paper examines the cognitive and computational roles of these constructs, and addresses the question of whether they can be discovered from empirical observations.

Causal models are treated as trees of binary random variables where the leaves are accessible to direct observation, and the internal nodes--representing hidden causes--account for inter-leaf dependencies. In probabilistic terms, every two leaves are conditionally independent given the value of some internal node between them.

We show that if the mechanism which drives the visible variables is indeed tree-structured, then it is possible to uncover the topology of the tree uniquely by observing pair-wise dependencies among the leaves. The entire tree structure, including the strengths of all internal relationships, can be reconstructed in time proportional to $n\log n$, where n is the number of leaves.

I. *INTRODUCTION: CAUSALITY, CONDITIONAL INDEPENDENCE AND TREES*

This study is motivated by the observation that human beings, facing complex phenomena, exhibit an almost obsessive urge to conceptually mold these phenomena into structures of cause-and-effect relationships. This tendency is, in fact, so compulsive that it sometimes comes at the expense of precision and often requires the invention of hypothetical, unobservable entities such as "ego", "elementary particles", and "supreme beings" to make theories fit the mold of causal schema. When we try to explain the actions of another person, for example, we invariably invoke abstract notions of mental states, social attitudes, beliefs, goals, plans and intentions. Medical knowledge, likewise, is organized into causal hierarchies of invading organisms, physical disorders, complications, pathological states, and only finally, the visible symptoms.

This paper takes the position that human obsession with causation is computationally motivated. Causal models are only attractive because they provide effective data-structures for representing empirical knowledge, and their effectiveness is a result of the high degree of decomposition they induce. More specifically, causes are viewed as names given to auxiliary variables which encode a summary of the interaction between the visible variables and, once calculated, would permit us to treat visible variables as if they were mutually independent.

The dual summarizing-decomposing role of a causal variable is analogous to that of an orchestra conductor; it achieves coordinated behavior through central communication and thereby relieves the players from having to communicate directly with each other. Such coordination is characteristic of tree structures and draws its effectiveness from the local nature of the data flow topology. In a management hierarchy, for example, where employees can only communicate with each other through their immediate superiors, the passage of information is swift, economical, conflict-free, and highly parallel. These computational attributes, we postulate, give rise to the satisfying sensation called "in-depth understanding", which people experience when they discover causal models consistent with observations.

Cast in probabilistic terms, central decomposition is embodied by the relation of *conditional independence*, which we claim constitutes the most universal and distinctive characteristic featured by the notion of causality. (See also [6] and [7].) In medical diagnosis, for example, a group of co-occurring symptoms often become independent of each other once we know the disease that caused them. When some of the symptoms directly influence each other, the medical profession invents a name for that interaction (e.g., complication, pathological state, etc.) and treats it as a new auxiliary variable which again assumes the decompositional role characteristic of causal agents. Knowing the exact state of the auxiliary variable renders the interacting symptoms independent of each other. Causes invoked to explain human behavior, such as motives and intentions, also induce conditional in-

*This work was supported in part by the National Science Foundation, Grant #DSR 83-13875

dependence. For example, once a murder suspect confesses to having wished the death of the victim, testimonies proving that he expressed such wishes in public or that he stood to gain from the victim's death are perceived to be irrelevant; they shed no further light on whether he actually performed the murder.

Based on these observations we chose to represent causal models as trees of binary random variables, where the leaves are directly accessible to empirical observations and the internal nodes represent hidden causes; any two leaves become conditionally independent once we know the value of some internal variable on the path connecting them. The propagation of updated probabilities in such trees was analyzed by Pearl [6] and Kim and Pearl [3]. It was shown that the propagation can be accomplished by a network of parallel processors working autonomously, and that the impact of new information can be imparted to all variables in time proportional to the longest path in the tree.

Given that tree-dependence captures the main feature of causation and that it provides a convenient computational medium for performing updating and predictions, we now ask whether the internal structure of the tree can be determined from observations made solely on the leaves. If it can, then the structure found would constitute an operational definition for the hidden causes. Additionally, if we take the view that "learning" entails the acquisition of computationally effective representations for nature's regularities, then the procedure of configuring the tree may reflect an important component of human learning.

A related structuring task was treated by Chow and Liu [1], who also used tree-dependent random variables to approximate an arbitrary joint distribution. However, whereas in Chow's trees all nodes denote observed variables, the internal nodes in our trees denote dummy variables, artificially concocted to make the representation tree-like. The problem of configuring probabilistic models using auxiliary variables is mentioned by Hinton *et al.* [2] as one of the tasks that a Boltzmann machine should be able to solve. However, no performance results have been reported and it is not clear whether the relaxation techniques employed by the Boltzmann machine can easily handle the restriction that the resulting structure be a tree.

This paper is organized as follows: Section 2 presents nomenclature and precise definitions for the notions of star-decomposability and tree-decomposability. In section 3 we treat triplets of random variables and ask under what conditions one is justified in attributing the observed dependencies to one central cause. We show that these conditions are readily testable and, when the conditions are satisfied, that the parameters specifying the relations between the visible variables and the central cause can be determined uniquely. In section 4 we extend these results to the case of a tree with n leaves. We show that if a joint distribution of n variables has a tree-dependent representation, then the uniqueness of the triplets' decomposition enables us to configure that tree from pair-wise dependencies among the variables. Moreover, the configuration procedure takes only $0(n\log n)$ steps. In Section 5 we evaluate the merits of this method and address the difficult issues of estimation and approximations.

II. PROBLEM DEFINITION AND NOMENCLATURE

Consider a set of n binary-valued random variables $x_1, \cdots, x_n$ with a given probability mass function $P(x_1, \cdots, x_n)$. We first address the problem of representing P as a marginal of an $(n+1)$-variable distribution $P_S(x_1, \ldots, x_n, w)$, that renders $x_1, \cdots, x_n$ conditionally independent given w, i.e.

$$P_s(x_1, \cdots, x_n, w) = P_s(w) \prod_{i=1}^{n} P_s(x_i|w) \tag{1}$$

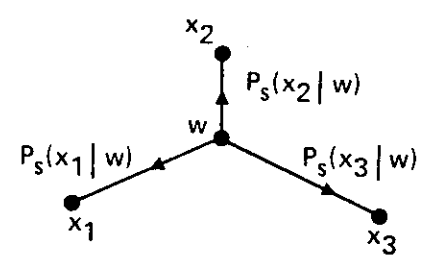
$$P(x_1, \ldots, x_n) = \alpha \prod_{i=1}^{n} P_s(x_i|w=1) + (1-\alpha) \prod_{i=1}^{n} P_s(x_i|w=0) \tag{2}$$

The functions $P_S(x_i \mid w)$, $w=0, 1$, $i=1, \ldots, n$, can be viewed as 2×2 stochastic matrices relating each x_i to the central hidden variable w (see Fig. 1a), hence we name P_S a *star-distribution* and call P *star-decomposable*. Each matrix contains two independent parameters, f_i and g_i, where

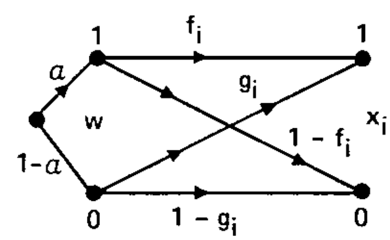
$$f_i = P_s(x_i = 1 \mid w=1)$$

$$g_i = P_s(x_i = 1 \mid w=0) \tag{3}$$

and the central variable w is characterized by its prior probability $P_s(w=1) = \alpha$ (see Figure 1b).



(a)



(b)

Figure 1

The advantages of having star-decomposable distributions are several. First, the product form of P, in (1) makes it extremely easy to compute the probability of any combination of variables. More importantly, it is also convenient for calculating the conditional probabilities $P(x_i|x_j)$, describing the impact of an observation x_j on the probabilities of unobserved variables. The computation requires only two vector multiplications.

Unfortunately, when the number of variables exceeds 3 the conditions for star-decomposability become very stringent, and are not likely to be met in practice. Indeed, a star-decomposable distribution for n variables has $2n+1$ independent parameters, while the specification of a general distribution requires 2^n-1 parameters. Lazarfeld [4] considered star-decomposable distributions where the hidden variable w is permitted to range over λ values, $\lambda>2$. Such an extension requires the solution of $\lambda n+\lambda-1$ non-linear equations to find the values of the $\lambda n+\lambda-1$ independent parameters. In this paper, we pursue a different approach, allowing a larger number of binary hidden variables, but insisting that they form a tree-like structure (see Figure 2), i.e., each triplet forms a star but the central variables may differ from triplet to triplet. Trees often portray meaningful conceptual hierarchies and, computationally, are almost as convenient as stars.

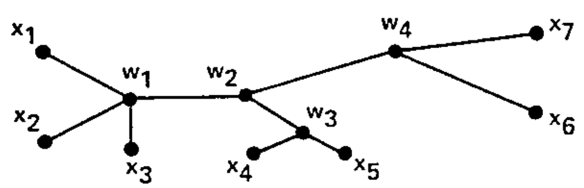


Figure 2

We shall say that a distribution $P(x_1,x_2, \cdots, x_n)$ is *tree-decomposable* if it is a marginal of a tree distribution

$$P_T(x_1,x_2, \cdots, x_n,w_1,w_2, \cdots, w_m) \qquad m \le n-2$$

where $w_1,w_2, \cdots, w_m$ correspond to the internal nodes of an unrooted tree T and $x_1, x_2, \cdots, x_n$ to its leaves. Given a tree structure and an assignment of variables to its nodes, the form of the corresponding distribution can be written by inspection. We first choose an arbitrary node as a root. This, in turn, defines a unique father $F(y_i)$ for each node $y_i \in \{x_1, \cdots, x_n, w_1, \cdots, w_n\}$ in the tree, except the chosen root, y_1. The joint distribution is simply given by the product form:

$$P_T(x_1 \cdots x_n, w_1 \cdots w_m) = P(y_1)\prod_{i=2}^{m+n} P[y_i|F(y_i)] \qquad (4)$$

For example, if in Figure 2 we choose w_2 as the root we obtain:

$$P_T(x_1 \cdots x_7, w_1 \cdots w_4) =$$

$$P(x_7|w_4)\,P(x_6|w_4)\,P(x_5|w_3)\,P(x_4|w_3)\,P(x_3|w_1)\,P(x_2|w_1)\,P(x_1|w_1)$$
$$P(w_1|w_2)\,P(w_3|w_2)\,P(w_4|w_2)\,P(w_2)$$

Throughout this discussion we shall assume that each w has at least three neighbors; otherwise it is superfluous. In other words, an internal node with two neighbors can simply be replaced by an equivalent direct link between the two. Note that any two leaves are conditionally independent given the value of any internal node on the path connecting them.

If we are given $P_T(x_1, \cdots x_n, w_1, \cdots w_m)$ then, clearly, we can obtain $P(x_1, \cdots x_n)$ by summing over the w's. We now ask whether the inverse transformation is possible, i.e., given a tree-decomposable distribution $P(x_1, \cdots x_n)$, can we recover its underlying extension $P_T(x_1 \cdots x_n, w_1 \cdots w_m)$? We shall show that: (1) the tree distribution P_T is unique, (2) it can be recovered from P using $n\log n$ computations, and (3) the structure of T is uniquely determined by the second order probabilities of P. The construction method depends on the analysis of star-decomposability for triplets which is presented next.

III. STAR-DECOMPOSABLE TRIPLETS

In order to test whether a given 3-variable distribution $P(x_1,x_2,x_3)$ is star-decomposable, we first solve Equation (2) and express the parameters α, f_i, g_i as a function of the parameters specifying P (see Figure 1). This task was carried out by Lazarfeld [4] in terms of the seven joint-occurrence probabilities

$$p_i = P\ (x_i=1)$$

$$p_{ij} = P\ (x_i=1, x_j=1) \qquad (5)$$

$$p_{ijk} = P\ (x_i=1, x_j=1, x_k=1)$$

and led to the following solution:

Define the quantities

$$[ij] = p_{ij} - p_i p_j \qquad (6)$$

$$S_i = \left(\frac{[ij][ik]}{[jk]}\right)^{\frac{1}{2}} \qquad (7)$$

$$\mu_i = \frac{p_i p_{ijk} - p_{ij} p_{ik}}{[jk]} \qquad (8)$$

$$K = \frac{S_i}{p_i} - \frac{p_i}{s_i} + \frac{\mu_i}{S_i p_i} \qquad (9)$$

and let t be the solution of

$$t^2 + Kt - 1 = 0 \qquad (10)$$

The parameters α, f_i, g_i are given by:

$$\alpha = \frac{t^2}{1+t^2} \qquad (11)$$

$$f_i = p_i + S_i \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{2}} \qquad (12)$$

$$g_i = p_i - S_i \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{2}} \qquad (13)$$

Moreover, the differences $f_i - g_i$ are independent of p_{ijk},

$$f_i - g_i = S_i \overset{\Delta}{=} \left(\frac{[ij][ik]}{[jk]}\right)^{\frac{1}{2}} \qquad (14)$$

The conditions for star-decomposability are obtained by requiring that the preceding solutions satisfy:
(a) S_i be real
(b) $0 \le f_i \le 1$
(c) $0 \le g_i \le 1$
Using the variances

$$\sigma_i = [p_i\,(1-p_i)]^{\frac{1}{2}} \qquad (15)$$

and the correlation coefficients

$$\rho_{ij} = \frac{p_{ij} - p_i p_j}{\sigma_i \sigma_j} \qquad (16)$$

requirement (a) is equivalent to the condition that all three correlation coefficients are non-negative. (If two of them are negative, we can rename two variables by their complements; the newly defined triplet will have all its pairs positively correlated.) We shall call triplets with this property *positively correlated*.

This, together with requirements (b) and (c), gives (after some manipulations):

Theorem 1: A necessary and sufficient condition for three dichotomous random variables to be star-decomposable is that they are positively correlated, and that the inequality:

$$\frac{p_{ik}p_{ij}}{p_i} \le p_{ijk} \le \frac{p_{ik}p_{ij}}{p_i} + \sigma_j\sigma_k\,(\rho_{jk} - \rho_{ij}\rho_{ik}) \qquad (17)$$

is satisfied for all $i \in \{1, 2, 3\}$. When this condition is satisfied, the parameters of the star-decomposed distribution can be determined uniquely, up to a complementation $w \to (1-w)$ of the central variable w (i.e., $w \to 1-w$, $f_i \leftrightarrow g_i$), $\alpha \to 1-\alpha$).

Obviously, in order to satisfy (17), the term $(\rho_{jk} - \rho_{ij}\rho_{ik})$ must be non-negative. This introduces a simple necessary condition for star-decomposability that may be used to quickly rule out many likely candidates.

Corollary -- A necessary condition for a distribution $P(x_1, x_2, x_3)$ to be star-decomposable is that all correlation coefficients obey the triangle inequality:

$$\rho_{jk} \ge \rho_{ji}\rho_{ik} \qquad (18)$$

(18) is satisfied with equality if w coincides with x_i; i.e., when x_j and x_k are independent given x_i. Thus, an intuitive interpretation of this corollary is that the correlation between any two variables must be stronger than that induced by their dependencies on the third variable; a mechanism accounting for direct dependencies must be present.

Having established the criterion for star-decomposability we may address a related problem: Suppose P is not star-decomposable, can it be approximated by a star-decomposable distribution $\hat{P}$ that has the same second-order probabilities?

The preceding analysis contains the answer to this question. Note that the 3rd order dependencies are represented only by the term p_{ijk}, and this term is confined by Eq. (17) to a region whose boundaries are determined by 2nd- order parameters. Thus, if we insist on keeping all 2nd-order dependencies of P in tact and are willing to choose p_{ijk} so as to yield a star-decomposable distribution, we can only do so if the region circumscribed by (17) is non-empty. This leads to the statement:

Theorem 2: A necessary and sufficient condition for the 2nd order dependencies among the triplet x_1, x_2, x_3 to support a star-decomposable extension is that the six inequalities:

$$\frac{p_{ij}p_{ik}}{p_i} \le x \le \frac{p_{ij}p_{ik}}{p_i} + \sigma_j\sigma_k\,(\rho_{jk} - \rho_{ij}\rho_{ik}) \quad i=1,2,3 \qquad (19)$$

possess a solution for x.

IV. A TREE-RECONSTRUCTION PROCEDURE

We are now ready to confront the central problem of this paper: Given a tree-decomposable distribution $P(x_1, \cdots, x_n)$, can we recover its underlying topology and the underlying tree-distribution $P_T(x_1, \cdots, x_n, w_1, \cdots, w_m)$?

The construction method is based on the observation that any three leaves in a tree have one and only one internal node that can be considered their *center*, i.e., it lies on all the paths connecting the leaves to each other. If one removes the center, the three leaves become disconnected from each other. This means that if P is tree-decomposable then the joint distribution of any triplet of variables x_i, x_j, x_k is star-decomposable, i.e., $P(x_i, x_j, x_k)$ uniquely determines the parameters α, f_i, g_i as in Equations (11), (12), and (13), where α is the marginal probability of the central variable. Moreover, if we compute the star decompositions of two triplets of leaves, both having the same central node w, the two distributions should have the same value for $\alpha = P_T(w=1)$. This provides us with a basic test for verifying whether two arbitrary triplets of leaves share a common center and a successive application of this test is sufficient for determining the structure of the entire tree.

Consider a 4-tuple x_1, x_2, x_3, x_4 of leaves in T. These leaves are interconnected through one of the four possi-

ble topologies shown in Figure 3. The topologies differ in the identity of the triplets which share a common center. For example, in the topology of Figure 3(a), the pair [(1,2,3), (1,2,4)] share a common center and so does the pair [(1,3,4), (2,3,4)]. In Figure 3(b), on the other hand, the sharing pairs are [(1,2,4), (2,4,3)] and [(1,3,4), (2,1,3)], and in Figure 3(d) all triplets share the same center. Thus, the basic test for center-sharing triplets enables us to decide the topology of any 4-tuple and, eventually, to configure the entire tree.

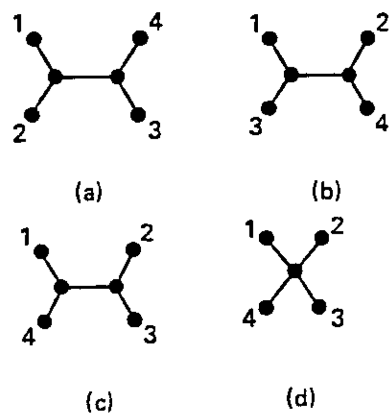


Figure 3

We start with any three variables x_1, x_2, and x_3, form their star decomposition, choose a fourth variable x_4, and ask to which leg of the star should x_4 be joined. We can answer this question easily by testing which pairs of triplets share centers, decide on the appropriate topology, and connect x_4 accordingly. Similarly, if we already have a tree structure T_i, with i leaves, and wish to know where to join the $(i+1)^{th}$ leaf, we can choose any triplet of leaves from T_i with central variable w, and test to which leg of w should x_{i+1} be joined. This, in turn, identifies a subtree T_i' of T_i that should receive x_{i+1} and permits us to remove from further considerations the subtrees emanating from the unselected legs of w. Repeating this operation on the selected subtree T_i' will eventually reduce it to a single branch, to which x_{i+1} is joined.

It is possible to show [8] that by choosing, in each state, a central variable that splits the available tree into subtrees of roughly equal-size, the joining branch of x_{i+1} can be identified in at most $\log_{\frac{k}{k-1}}(i)$ tests, where k is the maximal degree of the tree T_i. This amounts to $O(n\log n)$ tests for constructing an entire tree of n leaves.

So far we have shown that the structure of the tree T can be uncovered uniquely. Next we show that the distribution P_T, likewise, is uniquely determined from P, i.e., that we can determine all the functions $P(x_i \mid w_j)$ and $P(w_j \mid w_k)$ in (4), for $i=1, \cdots n$ and $j, k=1, 2, \cdots m$. The functions $P(x_i \mid w_j)$ assigned to the peripheral branches of the tree are determined directly from the star decomposition of triplets involving adjacent leaves. In Figure 2, for example, the star decomposition of $P(x_1, x_2, x_5)$ yields $P(x_1 \mid w_1)$ and $P(x_2 \mid w_1)$. The conditional probabilities $P(w_i \mid w_k)$ assigned to interior branches are determined by solving matrix equations. For example, $P(x_1 \mid w_2)$ is obtained from the star decomposition of (x_1, x_5, x_7), and it is related to $P(x_1 \mid w_1)$ via

$$P(x_1 \mid w_2) = \sum_{w_1} P(x_1 \mid w_1) P(w_1 \mid w_2)$$

This matrix equation has a solution for $P(w_1 \mid w_2)$ because $P(x_1 \mid w_1)$ must be non-singular. It is only singular when $f_1 = g_1$, i.e., when x_1 is independent of w_1 and, therefore, independent of all other variables. Hence, we can determine the parameters of the branches next to the periphery, use those to determine more interior branches, and so on, until all the interior conditional probabilities $P(w_i \mid w_j)$ are determined.

Next, we shall show that the tree structure can be recovered without resorting to 3rd order probabilities; correlations among pairs of leaves suffice. This feature stems from the observation that when two triplets of a 4-tuple are star-decomposable with respect to the same central variable w (e.g. 1,2,3 and 1,2,4 in Fig. 3(a)), then not only the values of α are the same but also the f and g parameters associated with the two common variables (e.g. 1 and 2 in Fig. 3(a)) must be the same. Whereas the value of α depends on a 3rd order probability, the difference $f_i - g_i$ depends only on 2nd order terms via Equation (14). Thus, requiring that $f_1 - g_1$ in Fig. 3(a) obtains the same value in the star decomposition of (1,2,3) as in that of (1,2,4), leads to the equation:

$$\frac{[12][13]}{[23]} = \frac{[12][14]}{[24]} \tag{20}$$

and, using (6), this yields

$$\rho_{13}\rho_{42} = \rho_{14}\rho_{32} \,. \tag{21}$$

An identical equality will be obtained for each $f_i - g_i$, $i=1,2,3,4$, relative to the topology of Figure 3(a). Similarly, the topology of Figure 3(b) dictates

$$\rho_{12}\rho_{43} = \rho_{14}\rho_{23} \tag{22}$$

and that of Figure 3(c):

$$\rho_{12}\rho_{34} = \rho_{13}\rho_{24} \tag{23}$$

Thus, we see that each of these three topologies is characterized by its own distinct equality, while the topology of Figure 3(d) by having all three equalities hold simultaneously. This provides the necessary 2nd-order criterion for deciding the topology of any 4-tuple tested; if the equality $\rho_{ij}\rho_{kl} = \rho_{ik}\rho_{jl}$ holds for some permutation of the indices, we decide on the topology $\genfrac{}{}{0pt}{}{i}{l}\!>\!\!-\!\!<\!\genfrac{}{}{0pt}{}{j}{k}$, if it holds for two such permutations, the entire 4-tuple is star decomposable. Note that the equality $\rho_{ij}\rho_{kl} = \rho_{ik}\rho_{jl}$ must hold for at least one permutation of the variables, or else the 4-tuple would not be tree-decomposable.

V. CONCLUSIONS AND OPEN QUESTIONS

This paper provides an operational definition for entities called "hidden causes", which are not directly observable but facilitate the acquisition of effective causal models from empirical data. Hidden causes are viewed as dummy variables which, if held constant, induce probabilistic independence between sets of visible variables. It is shown that if all variables are bi-valued and if the activities of the visible variables are governed by a tree-decomposable probability distribution, then the topology of the tree can be uncovered uniquely from the observed correlations between pairs of variables. Moreover, the structuring algorithm requires only $n \log n$ steps.

The method introduced in this paper has two major shortcomings: It requires precise knowledge of the correlation coefficients and it only works when the underlying model is tree-structured. In practice, we often have only sample estimates of the correlation coefficients, and it is therefore unlikely that criteria based on equalities (as in Eq. (21)) will ever be satisfied exactly. It is possible, of course, to relax these criteria and make topological decisions on the basis of proximities rather than equalities. For example, instead of searching for an equality $\rho_{ij}\rho_{kl} = \rho_{ik}\rho_{jl}$, we can decide the 4-tuple topology on the basis of the permutation of indices that minimizes the difference $\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl}$. Experiments show, however, that the structure which evolves by such a method is very sensitive to inaccuracies in the estimates ρ_{ij}, because no mechanism is provided to retract erroneous decisions made in the early stages of the structuring process. Ideally, the topological membership of the $(i+1)^{th}$ leaf should be decided not merely by its relations to a single triplet of leaves chosen to represent an internal node w, but also by its relations to all previously structured triplets which share w as a center. This, of course, will substantially increase the complexity of the algorithm.

Similar difficulties plague the task of finding the best tree-structured *approximation* to a distribution which is not tree-decomposable. Even though we argued that natural data which lend themselves to causal modeling should be representable as tree-decomposable distributions, these distributions may contain internal nodes with more than two values. The task of determining the parameters associated with such nodes is much more complicated and, in addition, rarely yields unique solutions. Unique solutions, as shown in section 4, are essential for building large structures from smaller ones. We leave open the question of explaining how approximate causal modeling, an activity which humans seem to perform with relative ease, can be embodied in computational procedures that are both sound and efficient.

REFERENCES

[1] Chow, C.K. and C.N. Liu, "Approximating Discrete Probability Distributions with Dependence Trees". *IEEE Trans. Inf. Theory*. IT-14 (1968), 462-467.

[2] Hinton, G.E., T.J. Sejnowski, and D.H. Ackley, "Boltzmann Machines: Constraint Satisfaction Networks that Learn". Technical Report CMU-CS-84-119, Department of Computer Science, Carnegie-Mellon University, 1984.

[3] Kim, J. and J. Pearl, "A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems" In *Proceedings of IJCAI-83*. Washington, D.C., August, 1983, pp. 190-193.

[4] Lazarfeld, . (1966) "Latent Structure Analysis". in Stouffer, Guttman, Slachman, Lazarfeld, Star, and Claussen (eds.), *Measurement and Prediction:* Wiley: New York.

[5] Pearl, J., "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach" In *Proceedings of the AAAI National Conf. on Artificial Intelligence*. Pittsburgh, PA, August, 1982, pp. 133-136.

[6] Simon, H.A. "Spurious Correlations: A Causal Interpretation". *Journal of American Statistical Association* 49 (1959) 469-492.

[7] Suppes, P., *A Probabilistic Theory of Causality*. North-Holland: Amsterdam, 1970.

[8] Tarsi, M. and Pearl, J., "Algorithmic Reconstruction of Trees", Technical Report UCLA-CSD-840061, Cognitive Systems Laboratory, University of California, Los Angeles, December 1984.