A Causal Calculus

(Draft Copy)

Judea Pearl

Cognitive Systems Laboratory

Computer Science Department

University of California, Los Angeles, CA 90024

judea@cs.ucla.edu

Abstract

Given an arbitrary causal graph, some of whose nodes are observable and some unobservable, the problem is to determine whether the causal effect of one variable on another can be computed from the joint distribution over the observables and, if the answer is positive, to derive a formula for the causal effect. We introduce a calculus which, using a step by step reduction of probabilistic expressions, derives the desired formulas.

1 Introduction

Networks employing *directed acyclic graphs* (DAGs) can be used to provide either

1. an economical scheme for representing conditional independence assumptions and joint distribution functions, or

2. a graphical language for representing causal influences.

Although the professed motivation for investigating such models lies primarily in the second category, [Wright, 1921, Blalock, 1971, Simon, 1954, Pearl 1988], causal inferences have been treated very cautiously in the statistical literature [Lauritzen & Spiegelhalter 1988, Cox 1992, Cox & Wermuth 1993, Spiegelhalter etal. 1993], as well as in the literature on influence diagrams [Howard, 1987, Shachter, 1986] and expert systems applications [Heckerman, 1990, Neapolitan, 1990]. The causal interpretation of the directed arcs has been de-emphasized in favor of the safer interpretation in terms of "relevance" and "dependence". This limited interpretation is deficient in several respects. First, causal associations are the primary source of judgments about dependence and relevance; they should therefore guide the process of knowledge elicitation [Pearl, 1988, page 123]. Second, rejecting the causal interpretation of directed arcs prevents us from using graphical models for making predictions about the effects of actions, unless actions have been considered in advance as part of the network. In other words, a joint distribution tells us how probable events are and how probabilities would change with subsequent observations, but a causal model also tells us how these probabilities would change as a result of external interventions in the system, including interventions that were not contemplated during the network construction. Such predictions are indispensable in most decision making applications, including policy analysis and treatment management.

In recent publications, the causal reading of belief networks has regained emphasis [Pearl and Verma 1991, Druzdzel & Simon 1993, Lemmer 1993, Goldszmidt 1992, Pearl 1993a Spirtes et al 1993 and Pearl 1993b]. In Spirtes et al [1993] and Pearl [1993b], for example, it was shown how complex information about external interventions can be organized and represented graphically and, conversely, how the graphical representation can be used to facilitate quantitative predictions of the effects of interventions. Section 2 reviews this aspect of causal networks, following the formulation in [Pearl 1993b].

The problem addressed in this paper is to quantify the effect of interventions when the causal network cannot be fully parameterized. In other words, we are given the topology of the network but not the conditional probabilities on all variables. The only probabilities available are those defined on a subset of the variables which are observed. To manage this problem, this paper introduces a calculus which integrates both statistical and causal information in a uniform symbolic machinery. The calculus admits two types of conditioning operators: ordinary Bayes conditioning, $P(y|X=x)$, and causal conditioning, $P(y|set(X=x))$, that is, the probability of $Y=y$ conditioned on holding X constant (at x) by external intervention. Given a causal graph and a joint distribution on a subset of its variables, the calculus derives conditional probabilities of both the Bayesian and the causal types, and, whenever possible, generates probabilistic formulas for the effect of interventions in terms of the observed distribution function.

2 The Manipulative Reading of Causal Networks: A Review

The connection between the probabilistic and the manipulative readings of directed acyclic graphs is formed through Simon's [1977] mechanism-based model of causal ordering[1]. In this model, each child-parent family in a DAG Γ represents a deterministic function

$$X_i = f_i(\mathbf{pa}_i, \epsilon_i), \tag{1}$$

where $\mathbf{pa}_i$ are the parents of variable X_i in Γ, and ϵ_i, $0 < i < n$, are mutually independent, arbitrarily distributed random disturbances. Characterizing each child-parent relationship as a deterministic function, instead of the usual conditional probability $P(x_i \mid \mathbf{pa}_i)$, imposes equivalent independence constraints on the resulting distributions and leads to the same recursive decomposition

$$P(x_1, ..., x_n) = \prod_i P(x_i \mid \mathbf{pa}_i) \tag{2}$$

that characterizes Bayesian Networks [Pearl, 1988]. This is so because each ϵ_i is independent on all non-descendants of X_i. However, the functional characterization $X_i = f_i(\mathbf{pa}_i, \epsilon_i)$ also specifies how the resulting distribution would change in response to external interventions, since, by convention, each function is presumed to remain constant unless specifically altered. Moreover, the non-linear character of f_i permits us to treat changes in the function f_i itself as a variable, F_i, by writing

$$X_i = f'_i(\mathbf{pa}_i, F_i, \epsilon_i) \tag{3}$$

[1]This mechanism-based model was adopted in [Pearl & Verma 1991] for defining probabilistic causal theories. It has been elaborated in Druzdzel & Simon [1993] and is also the basis for the "invariance" principle of [Spirtes et al, 1993].

where

$$f_i'(a,b,c) = f_i(a,c) \text{ whenever } b = f_i.$$

Thus, any external intervention F_i that alters f_i can be represented graphically as an added parent node of X_i, and the effect of such an intervention can be analyzed by Bayesian conditionalization, that is, by simply setting this added parent variable to the appropriate value f_i.

The simplest type of external intervention is one in which a single variable, say X_i, is forced to take on some fixed value, say, x_i'. Such intervention, which we call *atomic*, amounts to replacing the old functional mechanism $X_i = f_i(\mathbf{pa}_i, \epsilon_i)$ with a new mechanism $X_i = x_i'$ governed by some external force F_i that sets the value x_i'. If we imagine that each variable X_i could potentially be subject to the influence of such an external force F_i, then we can view the causal network Γ as an efficient code for predicting the effects of atomic interventions and of various combinations of such interventions.

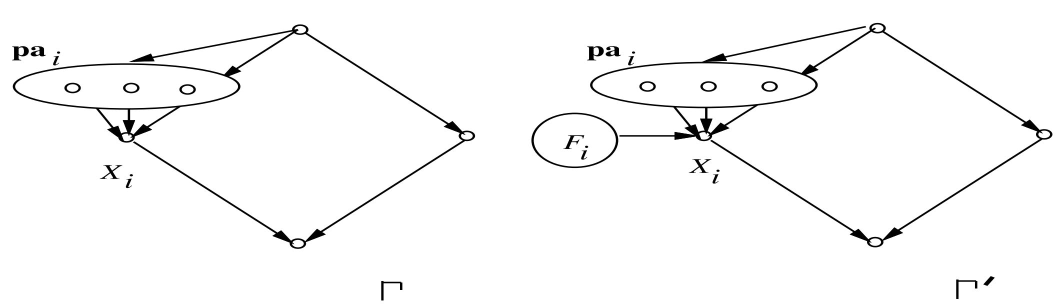

Figure 1: Representing external intervention F_i by an augmented network

$$\Gamma' = \Gamma \cup \{F_i \rightarrow X_i\}.$$

The effect of an atomic intervention $set(X_i = x_i')$ is encoded by adding to Γ a link $F_i \longrightarrow X_i$ (see Figure 1), where F_i is a new variable taking values in $\{set(x_i'),\ idle\}$, x_i' ranges over the domain of X_i, and $idle$ represents no intervention. Thus, the new parent set of X_i in the augmented network is $\mathbf{pa}'_i = \mathbf{pa}_i \cup \{F_i\}$, and it is related to X_i by the conditional probability

$$P(x_i \mid \mathbf{pa}'_i) = \begin{cases} P(x_i \mid \mathbf{pa}_i) & \text{if} \quad F_i = idle \\ 0 & \text{if} \quad F_i = set(x'_i) \quad \text{and} \quad x_i \neq x'_i \\ 1 & \text{if} \quad F_i = set(x'_i) \quad \text{and} \quad x_i = x'_i \end{cases} \tag{4}$$

The effect of the intervention $set(x'_i)$ is to transform the original probability function $P(x_1, ..., x_n)$ into a new function $P_{x'_i}(x_1, ..., x_n)$, given by

$$P_{x'_i}(x_1, ..., x_n) = P'(x_1, ..., x_n \mid F_i = set(x'_i)) \tag{5}$$

where P' is the distribution specified by the augmented network $\Gamma' = \Gamma \cup \{F_i \rightarrow X_i\}$ and Eq. (4), with an arbitrary prior distribution on F_i. In general, by adding a hypothetical intervention link $F_i \rightarrow X_i$ to each node in Γ, we can construct an augmented probability function $P'(x_1, ..., x_n; F_1, ..., F_n)$ that contains information about richer types of interventions. Multiple interventions would be represented by conditioning P' on a subset of the F_i's (taking values in their respective $set(x'_i)$), while the pre-intervention probability function P would be viewed as the posterior distribution induced by conditioning each F_i in P' on the value $idle$.

This representation yields a simple and direct transformation between the pre-intervention and the post-intervention distributions:

$$P_{x'_i}(x_1, ..., x_n) = \begin{cases} \frac{P(x_1, ..., x_n)}{P(x_i \mid \mathbf{pa}_i)} & \text{if} \quad x_i = x'_i \\ 0 & \text{if} \quad x_i \neq x'_i \end{cases} \tag{6}$$

This transformation reflects the removal of the term $P(x_i \mid \mathbf{pa}_i)$ from the product decomposition of Eq. (2), since $\mathbf{pa}_i$ no longer influence X_i. Graphically, the removal of this term is equivalent to removing the links between $\mathbf{pa}_i$ and X_i, while keeping the rest of the network intact. Transformations involving conjunctive and disjunctive actions can be obtained by straightforward applications of Eq. (5) [Spirtes etal. 1993, Goldszmidt & Pearl 1992, Goldszmidt 1992]

The transformation (6) exhibits the following properties:

1. An intervention $set(x_i)$ can affect only the descendants of X_i in Γ.

2. For any set $\mathbf{S}$ of variables, we have

$$P_{x_i}(\mathbf{S} \mid \mathbf{pa}_i) = P(\mathbf{S} \mid x_i, \mathbf{pa}_i). \tag{7}$$

   In other words, given $X_i = x_i$ and $\mathbf{pa}_i$, it is superfluous to find out whether $X_i = x_i$ was established by external intervention or not. This can be seen directly from the augmented network Γ' (see Figure 1), since $\{X_i\} \cup \mathbf{pa}_i$ d-separates F_i from the rest of the network, thus legitimizing the conditional independence $\mathbf{S} \perp\!\!\!\perp F_i \mid (X_i, \mathbf{pa}_i)$.

3. A sufficient condition for an external intervention $set(X_i = x_i)$ to have the same effect on X_j as the passive observation $X_i = x_i$ is that X_i d-separates $\mathbf{pa}_i$ from X_j, that is,

$$P'(x_j | set(x_i)) = P(x_j \mid x_i) \ \text{ iff } \ X_j \perp\!\!\!\perp \mathbf{pa}_i \mid X_i. \tag{8}$$

The immediate implication of Eq. (6) is that, given the structure of the causal network Γ, one can infer post-intervention distributions from pre-intervention distributions; hence, we can reliably estimate the effects of interventions from passive (i.e., non-experimental) observations. Of course, Eq. (6) does not imply that we can always substitute observational studies for experimental studies, as this would require an estimation of $P(x_i \mid \mathbf{pa}_i)$. The mere identification of $\mathbf{pa}_i$ (i.e., the direct causal factors of X_i) requires substantive causal knowledge of the domain which is often unavailable. Moreover, even when we have sufficient substantive knowledge to structure Γ, some members of $\mathbf{pa}_i$ may be unobservable, or *latent*. Fortunately, there are conditions for which an unbiased estimate of $P(x_j|set(x_i))$ can be obtained even when the $\mathbf{pa}_i$ variables are latent and, moreover, simple graphical criteria can tell us when these conditions are satisfied. The characterization of those conditions and the symbolic derivation of expressions of the form $P(x_j|set(x_i))$ are the main topics of this paper.

3 A Causal Calculus

3.1 Preliminary Notation

Let X, Y, Z, W be four arbitrary disjoint sets of nodes in the dag G. We say that X and Y are independent given Z in G, denoted $(X \perp\!\!\!\perp Y|Z)_G$, if the set Z d-separates all paths from X to Y in G. We denote by $G_{\overline{X}}$ ($G_{\underline{X}}$, respectively) the graph obtained by deleting from G all arrows pointing to (emerging from, respectively) nodes in X.

Finally, we replace the expression $P(y|set(x), z)$ by a simpler expression $P(y|\hat{x}, z)$, using the $\hat{}$ symbol to identify the variables that are kept constant externally. In words, the expression $P(y|\hat{x}, z)$ stands for the probability of $Y = y$ given that $Z = z$ is observed and X is held constant at x.

3.2 Inference Rules

Armed with this notation we are now able to formulate the three basic inference rules of the proposed causal calculus.

Theorem 3.1 *Given a causal theory $< P, G >$, for any sets of variables X, Y, Z, W we have:*

Rule 1 *Insertion/deletion of Observations (Bayes conditioning)*

$$P(y|\hat{x}, z, w) = P(y|\hat{x}, w) \quad \textit{if} \quad (Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{X}}} \tag{9}$$

Rule 2 *Action/observation Exchange*

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, z, w) \quad \textit{if} \quad (Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{X}\underline{Z}}} \tag{10}$$

Rule 3 *Insertion/deletion of actions*

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, w) \quad \textit{if} \quad (Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{X}\,\overline{Z}}} \tag{11}$$

Each of the inference rules above can be proven from the basic interpretation of the "set(x)" operation as a replacement of the causal mechanism which connects X to its parent prior to the action by a new mechanism $X = x$ introduced by the intervening force (as in Eqs. (4) - (5)).

Rule 1 reaffirms d-separation as a legitimate test for Bayesian conditional independence in the distribution determined by the intervention $set(X = x)$, hence the graph $G_{\overline{X}}$.

Rule 2 provides conditions for an external intervention set $(Z = z)$ to have the same effect on Y as the passive observation $Z = z$. It is equivalent to Eq. (8), also named the "back-door" criterion [Pearl, 1993b].

Rule 3 provides conditions for introducing (or deleting) an external intervention set ($Z = z$) without affecting the probability of $Y = y$. The validity of this rule stems, again, from simulating the intervention $set(Z = z)$ by severing all relations between Z and its parent (hence the graph $G_{\overline{XZ}}$).

3.3 Example

We will now demonstrate how these inference rules can be used to quantify the effect of actions, given partially specified causal theories. Consider the graph G given in Figure 2 below. It represents the theory $< P(x, y, z), G >$ where X, Y, Z are the observed variables.

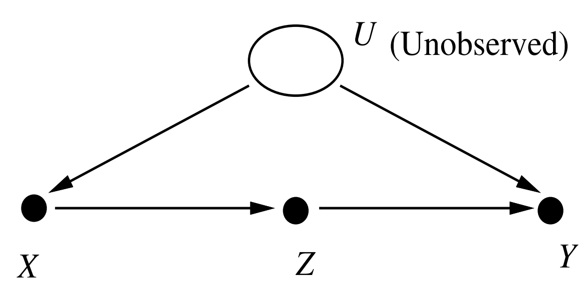

Figure 2

Since U is unobserved, the theory is only partially specified; it will be impossible to infer all required parameters such as $P(u)$, or $P(y|z, u)$. We will see however that this structure still permits us to quantify, using causal calculus, the effect of every action on every observed variable.

The applicability of the three inference rules depends on d-separation conditions holding in various graphs, the structure of which varies with the expressions to be manipulated. Figure 3 displays the graphs that will be needed for the derivations that follow.

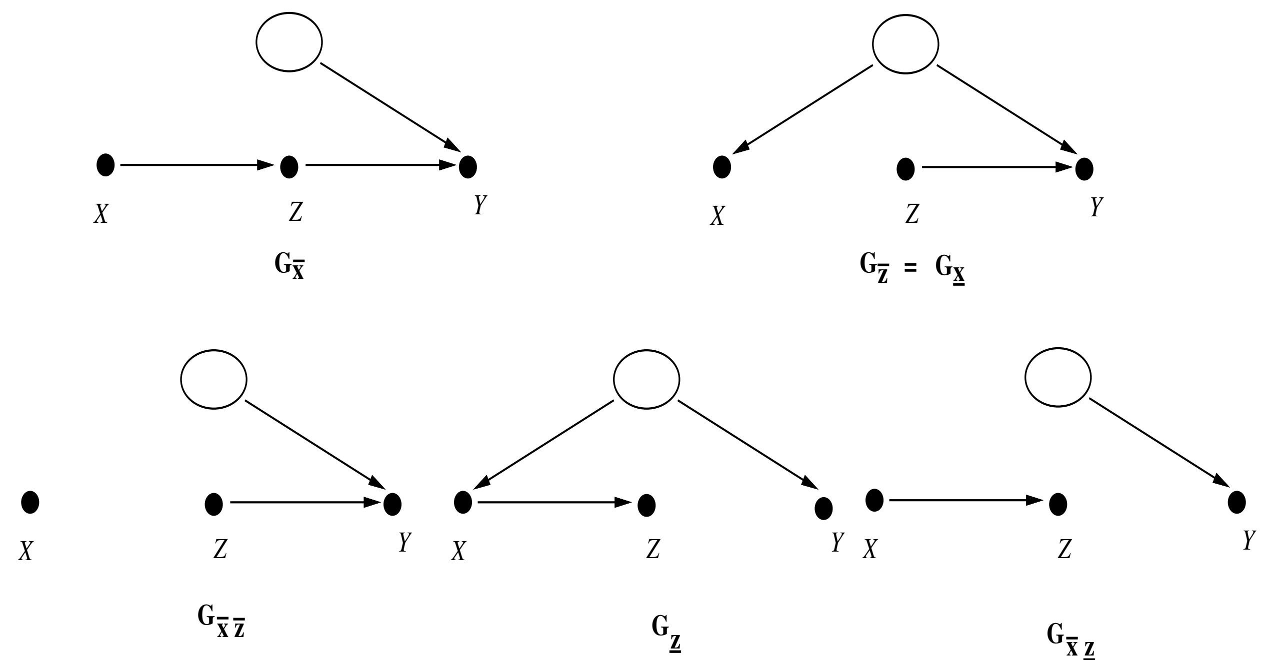

Figure 3

Task-1, compute $P(z|\hat{x})$

This task can be accomplished in one step, since G satisfies the applicability condition for Rule 2, namely $X \perp\!\!\!\perp Z$ in $G_{\underline{X}}$ (because the path $X \leftarrow U \rightarrow Y \leftarrow Z$ is blocked by the collider at Y) and we can write

$$P(z|\hat{x}) = P(z|x) \tag{12}$$

Task-2, compute $P(y|\hat{z})$

Here we cannot apply Rule 2 to exchange $\hat{z}$ by z, because $G_{\underline{Z}}$ contains a path from Z to Y (so called a "back-door" path [Pearl, 1993b]). Naturally, we would like to "block" this path by "adjusting for" covariates (such as X) that reside on that path. Symbolically, the "adjustment" operation involves conditioning and summing over all values of X,

$$P(y|\hat{z}) = \sum_x P(y|x, \hat{z}) P(x|\hat{z}) \tag{13}$$

We now have to deal with two expressions involving $\hat{z}$, $P(y|x, \hat{z})$ and $P(x|\hat{z})$. The latter can be readily computed by applying Rule 3 for action deletion.

$$P(x|\hat{z}) = P(x) \quad \text{if} \quad (Z \perp\!\!\!\perp X)_{G_{\overline{Z}}} \tag{14}$$

noting that, indeed, X and Z are d-separated in $G_{\overline{Z}}$. (This can be seen immediately from Figure 2; manipulating Z will have no effect on X.) To reduce the former quantity, $P(y|x, \hat{z})$, we consult Rule 2

$$P(y|x, \hat{z}) = P(y|x, z) \quad \text{if} \quad (Z \perp\!\!\!\perp Y|X)_{G_{\underline{Z}}} \tag{15}$$

and note that X d-separates Z from Y in $G_{\underline{Z}}$. This allows us to write Eq. (13) as

$$P(y|\hat{z}) = \sum_x P(y|x, z)P(x) = E_x P(y|x, z) \tag{16}$$

which is a special case of the "back-door" formula [Pearl, 1993b, Eq. (14)] with $S = X$. This formula appears in a number of treatments on causal effects (see for example [Rosenbaum & Rubin, 1983; Rosenbaum, 1989; Pratt & Schlaifer, 1988]) where the legitimizing condition, $(Z \perp\!\!\!\perp Y|X)_{G_{\underline{Z}}}$ was given a variety of names, all based on conditional-independence judgments of one sort or another. The causal calculus replaces such judgments by formal tests (d-separation) on a single graph (G) which represents the domain knowledge.

We are now ready to tackle a harder task, the evaluation of $P(y|\hat{x})$, which cannot be reduced to an observational expression by direct application of any of the inference rules.

Task-3, compute $P(y|\hat{x})$

Writing

$$P(y|\hat{x}) = \sum_z P(y|z,\hat{x})P(z|\hat{x}) \tag{17}$$

we see that the term $P(z|\hat{x})$ was reduced in Eq. (12) while no rule can be applied to eliminate the manipulation symbol $\hat{}$ from the term $P(y|z,\hat{x})$. However, we can add a $\hat{}$ symbol to this term via Rule 2

$$P(y|z,\hat{x}) = P(y|\hat{z},\hat{x}) \tag{18}$$

since Figure 3 shows:

$$(Y \perp\!\!\!\perp Z|X)_{G_{\overline{X}\underline{Z}}}$$

We can now delete the action $\hat{x}$ from $P(y|\hat{z},\hat{x})$ using Rule 3, since $Y \perp\!\!\!\perp X|Z$ holds in $G_{\overline{XZ}}$. Thus, we have

$$P(y|z,\hat{x}) = P(y|\hat{z}) \tag{19}$$

which was calculated in Eq. (16). Substituting, (16), (19, and (12) back in (17), finally yields

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_{x'} P(y|x',z)P(x') \tag{20}$$

Eq. (20) was named the Mediating Variable formula in [Pearl, 1993c], where it was derived by algebraic manipulation of the joint distribution and taking the expectation over U.

Task-4, compute $P(y,z|\hat{x})$

$$P(y,z|\hat{x}) = P(y|z,\hat{x})P(z|\hat{x}) \tag{21}$$

The two terms on the r.h.s. were derived before in Eqs. (12) and (19), from which we obtain

$$\begin{aligned} P(y, z|\hat{x}) &= P(y|\hat{z})P(z|x) \\ &= P(z|x) \textstyle\sum_{x'} P(y|x', z)P(x') \end{aligned}$$

3.4 Discussion

In this example we were able to compute answers to all possible queries of the form $P(y|z, \hat{x})$ where Y, Z, and X are subsets of observed variables. In general, this will not be the case. For example, there is no general way of computing $P(y|\hat{x})$ from the observed joint distribution whenever the causal model contains the subgraph shown in Figure 4, where X and Y are adjacent, and the dashed line represents a path traversing unobserved variable[2]. Similarly,

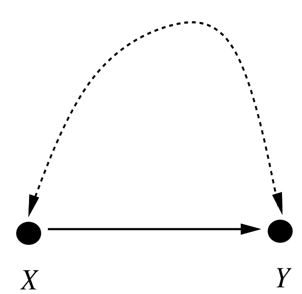

Figure 4

our ability to compute $P(y|\hat{x})$ for every pair of singleton variables does not ensure our ability to compute joint distributions, e.g. $P(y_1, y_2|\hat{x})$. Figure 5, for example, shows a causal graph where both $P(y_1|\hat{x})$ and $P(y_2|\hat{x})$ are computable, but $P(y_1, y_2|\hat{x})$ is not. Consequently, we cannot compute $P(z|\hat{x})$. Interestingly, the graph of Figure 5 is the smallest graph which does not contain the pattern of Figure 4 and still presents an uncomputable causal effect.

[2]One can calculate upper and lower bounds on $P(y|\hat{x})$ and these bounds may coincide for special distributions $P(x, y, z)$ [Balke & Pearl, 1993] but there is no way of computing $P(y|\hat{x})$ for *every* distribution $P(x, y, z)$.

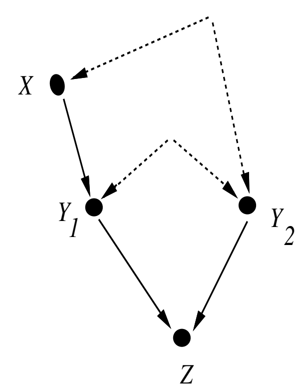

Figure 5

Another interesting feature demonstrated by the network in Figure 5 is that it is often easier to compute the effect of a joint action than the effects of its constituent singleton actions[3]. In this example, it is possible to compute $P(z|\hat{x}, \hat{y_1})$, yet there is no way of computing $P(z|\hat{x})$. For example, the former can be evaluated by invoking Rule 2, writing

$$
\begin{aligned}
P(z|\hat{x}, \hat{y_2}) &= \sum_{y_1} P(z|y_1, \hat{x}, \hat{y_2}) P(y_1|\hat{x}, \hat{y_2}) \\
&= \sum_{y_1} P(z|y_1, x_1, y_2) P(y_1|x)
\end{aligned}
$$

On the other hand, Rule 2 cannot be applied to the computation of $P(y_1|\hat{x}, y_2)$ because, conditioned on Y_2, X and Y_1 are d-connected in $G_{\underline{X}}$ (through the dashed lines). We conjecture, however, that whenever $P(y|\hat{x}_i)$ is computable for every singleton x_i, then $P(y|\hat{x}_1, \hat{x}_2, ...\hat{x}_l)$ is computable as well, for any subset of variables $\{X_1, ...X_l\}$.

BIBLIOGRAPHY

Balke, A. and Pearl, J. (1993) Nonparametric bounds on treatment effects in partial compliance studies. em Technical Report R-199, UCLA Computer Science Department.

[3]The fact that the two tasks are not equivalent was brought to my attention by James Robins who has worked out many of these computations in the context of sequential treatment management [Robins 1989].

(Submitted.)

Blalock, H.M. (1971) *Causal Models in the Social Sciences*. London: Macmillan.

Cox, D.R. (1992) Causality: Some statistical aspects. *Journal of the Royal Statistical Society*, Series A, **155**, 291-301.

Cox, D.R. and N. Wermuth (1993) Linear dependencies represented by chain graphs. *Statistical Science*, **8** (3), 204-218.

Druzdzel, M.J., and Simon, H.A. (1993) Causality in Bayesian Belief, in *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (D. Heckerman and A. Mamdani, eds.), Morgan Kaufmann, San Mateo, CA, 3-11.

Goldszmidt, M. (1992) Qualitative Probabilities: A Normative Framework for Commonsense Reasoning. *Technical Report R-190*, UCLA Cognitive Systems Laboratory, Ph.D. Thesis.

Goldszmidt, M. and J. Pearl (1992) Default ranking: A practical framework for evidential reasoning, belief revision and update. in *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, pp. 661-672.

Heckerman, D. (1990) Probabilistic Similarity Networks. *Networks*, **20**, (5), 607-636.

Howard (1987) From Influence to Relevance to Knowledge, in Oliver & Smith (eds.), 3-23.

Lauritzen, S.L. and D.J. Spiegelhalter (1988) Local computations with probabilities on graphical structures and their applications to expert systems. *Proceedings of the Royal Statistical Society*, Series B, **50**, 154-227.

Lemmer, J.F. (1993) Causal Modeling, in *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (D. Heckerman and A. Mamdani, eds.), Morgan Kaufmann, San Mateo, CA, 143-151.

Neapolitan, R. (1990) *Probabilistic Reasoning in Expert Systems*. New York: Wiley.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

Pearl, J., (1993a) From Conditional Oughts to Qualitative Decision Theory. in *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (D. Heckerman and A. Mamdani (eds.)), Morgan Kaufmann, San Mateo, CA,

Pearl, J. (1993b) Graphical Models, Causality, and Intervention. *Statistical Science*, **8**, (3), 266-273.

Pearl, J. (1993c) Mediating Instrumental Variables. *Technical Report R-210*, Cognitive Systems Laboratory, UCLA Computer Science Department.

Pearl, J. and T. Verma (1991) A Theory of Inferred Causation, in *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference*, (eds J.A. Allen, R. Fikes and E. Sandewall), Morgan Kaufmann, San Mateo, CA, pp. 441-452.

Pratt, J., and Schlaifer, R. (1988) On the Interpretation and Observation of Laws. *Journal of Economics*, **39**, 23-52.

Robins, J.M. (1989) The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies, in *Health Service Re-*

search Methodology: A Focus on AIDS, (eds L. Sechrest, H. Freeman, and A. Mulley), NCHSR, U.S. Public Health Service, 113-159, 1989.

Rosenbaum, P.R. (1989) The role of known effects in observational studies. *Biometrics*, **45**, 557-569.

Rosenbaum, P. and D. Rubin (1983) The central role of propensity score in observational studies for causal effects. *Biometrica*, **70**, 41-55.

Schacter, R.D. (1986) Evaluating Influence Diagrams. *Oper. Res.*, **34**, (6), 871-882.

Simon, H.A. (1954) Spurious Correlations: A Causal Interpretation. *J. Amer. Statist. Assoc.*, **49**, 469-492.

Simon, H.A. (1977) *Models of Discovery: and Other Topics in the Methods of Science*, D. Reidel, Dordrecht, Holland.

Spiegelhalter, D.J., S.L. Lauritzen, P.A. Dawid and R.G. Cowell (1993), Bayesian analysis in expert systems. *Statistical Science*, **8** (3), 219-247.

Spirtes, P., C. Glymour and R. Schienes (1993) *Causation, Prediction, and Search*, Springer-Verlag, New York.

Wright, S. (1921) Correlation and Causation. *Journal of Agricultural Research*, **20**, 557-585.