A Probabilistic Calculus of Actions

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Abstract

We present a symbolic machinery that admits both probabilistic and causal information about a given domain, and produces probabilistic statements about the effect of actions and the impact of observations. The calculus admits two types of conditioning operators: ordinary Bayes conditioning, $P(y|X=x)$, which represents the observation $X=x$, and causal conditioning, $P(y|do(X=x))$, read: the probability of $Y=y$ conditioned on holding X constant (at x) by deliberate action. Given a mixture of such observational and causal sentences, together with the topology of the causal graph, the calculus derives new conditional probabilities of both types, thus enabling one to quantify the effects of actions and observations.

1 Introduction

Probabilistic methods, especially those based on graphical models have proven useful in tasks of predictions, abduction and belief revision [Pearl 1988, Heckerman 1990, Goldszmidt 1992, Darwiche 1993]. Their use in planning, however, remains less popular[1] partly due to the unsettled, strange relationships between probability and actions. In principle, actions are not part of standard probability theory, and understandably so –– probabilities captures normal relationships in the world, while actions represent interventions that perturb those relationships. It is no wonder, therefore, that actions are treated as foreign entities in traditional probability calculus; they do not serve as arguments of probability expressions nor as events for conditioning such expressions. Whenever an action is given a formal symbol, that symbol serves merely as an index for distinguishing one probability function from another, but not to convey any information about the purpose or effect of the action.

The peculiar status of actions in probability theory can be seen most clearly in comparison to the status of observations. By specifying a probability function $P(s)$ on the possible states of the world, we automatically specify how probabilities would change with every conceivable observation e, since $P(s)$ permits us to compute (using Bayes rule) the posterior probabilities $P(E|e)$ for every pair of events E and e. However, specifying $P(s)$ tells us nothing about how our probabilities should be revised as a response to an external action A. In general, if

[1] works by Dean & Kanazawa [1989] and Kushmerick et al, [1993] not withstanding

an action A is to be described as a function that takes $P(s)$ and transforms it to $P_A(s)$, then Bayesian conditioning is clearly inadequate for encoding this transformation. For example, consider the statements: "I have observed the barometer reading to be x" and "I intervened and set the barometer reading to x" If processed by Bayes conditioning on the event "the barometer reading is x", these two reports would have the same impact on our current probability function, yet we certainly do not consider the two reports equally informative about an incoming storm.

Philosophers [Lewis, 1974] studied another probability transformation called "imaging" (to be distinguished from "conditioning") which was deemed useful in the analysis of subjunctive conditionals and which more adequately represent transformations associated with actions. Whereas Bayes conditioning $P(s|e)$ transfers the entire probability mass from states excluded by e to the remaining states (in proportion to their current $P(s)$), imaging works differently; each excluded state s transfers its mass individually to a select set of states $S^*(s)$, which are considered "closest" to s. While providing a more adequate and general framework for actions, imaging leaves the precise specification of the selection function $S^*(s)$ totally unconstrained. The task of formalizing and representing these specifications can be viewed as the probabilistic version of the infamous *frame problem.* and its two satellites, the *ramification* and *concurrent actions* problems.

An assumption commonly found in the literature is that the effect of an elementary action $do(q)$ is merely to change $\neg q$ to q in case the current state satisfies $\neg q$, but, otherwise, to leave things unaltered[2]. We can call this assumption the "delta" rule, variants of which are embedded in STRIPS as well as in probabilistic planning systems. In BURIDAN [Kushmerick et al, 1993], for example, every action is specified as a probabilistic mixture of several elementary actions, each operating under the delta rule.

The problem with the delta rule and its variants is that they do not take into account the indirect ramifications of an action such as, for example, those triggered by chains of causally related events. To handle such ramifications we must construct a causal theory of the domain, specifying which event chains are likely to be triggered by a given action (the ramification problem) and how these chains interact when triggered by several actions (the concurrent action problem). Elaborating on the works of Dean and Wellman [Dean & Wellman 1991], this paper shows how the frame, ramification and concurrency problem can be handled effectively using the language of causal graphs.

The key idea is that causal knowledge can efficiently be organized in terms of just a few basic mechanisms, each involving a relatively small number of variables, and each encoded as a set of functional constraints perturbed by random disturbances. Each external elementary action overrules just one mechanism while leaving the others unaltered. The specification of an action then requires only the identification of the mechanisms which are overruled by that action. Once this is identified, the effect of the action (or combinations thereof) can be computed from the constraints imposed by the remaining mechanisms.

The semantics behind causal graphs and their relations to actions and belief networks have been discussed in prior publications [Pearl and Verma 1991, Druzdzel & Simon 1993, Goldszmidt & Pearl 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

[2]This assumption corresponds to Dalal's [1988] database update, which uses the Hamming distance to define the "closest world" in Lewis' imaging.

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, including interventions that were not contemplated during the network construction[3]. 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 graph is not fully parameterized, that is, we are given the topology of the graph but not the conditional probabilities on all variables. Numerical probabilities will be given to only a subset of variables, in the form of unstructured conditional probability sentences. This is a more realistic setting in AI applications, where the user/designer might not have either the patience or the knowledge necessary for the specification of a complete distribution function. Some combinations of variables may be too esoteric to be assigned probabilities, and some variables may be too hypothetical (e.g., "life style" or "attitude") to even be parametrized numerically.

To manage this problem, this paper introduces a calculus which operates on whatever probabilistic and causal information is available, and, using symbolic transformations on the input sentences, produces probabilistic assessments of the effect of actions. The calculus admits two types of conditioning operators: ordinary Bayes conditioning, $P(y|X = x)$, and causal conditioning, $P(y|do(X = x))$, that is, the probability of $Y = y$ conditioned on holding X constant (at x) by deliberate external action. Given a causal graph and an input set of conditional probabilities, the calculus derives new 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 input information.

2 The Manipulative Reading of Causal Networks: 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[4]. In this model, each child-parent family in a DAG G 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 G, and ϵ_i, $0 < i < n$, are mutually independent, arbitrarily distributed random disturbances. A *causal theory* is a pair $< P, G >$, where G is a dag and P is the probability distribution that results from the functions f_i in (1).

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

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

[3]Influence diagrams, in contrast, require that actions be considered in advance as part of the network

[4]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].

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

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

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 G as an efficient code for predicting the effects of atomic interventions and of various combinations of such interventions.

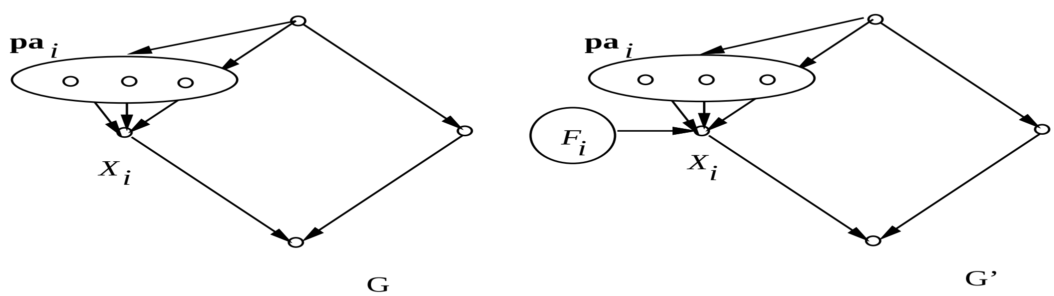

Figure 1: Representing external intervention F_i by an augmented network $G' = G \cup \{F_i \rightarrow X_i\}$.

The effect of an atomic intervention $do(X_i = x_i')$ is encoded by adding to G a link $F_i \longrightarrow X_i$ (see Figure 1), where F_i is a new variable taking values in $\{do(x_i'),\ idle\}$, 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

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

The effect of the intervention $do(x_i')$ 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

$$P_{x_i'}(x_1, ..., x_n) = P'(x_1, ..., x_n \mid F_i = do(x_i')) \tag{5}$$

where P' is the distribution specified by the augmented network $G' = G \cup \{F_i \rightarrow X_i\}$ 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 G, 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 $do(x_i'))$, 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:

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

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 $do(x_i)$ can affect only the descendants of X_i in G.

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 G' (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 $do(X_i = x_i)$ 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,

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

The immediate implication of Eq. (6) is that, given the structure of the causal network G, 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. However, the use of Eq. (6) is limited for several reasons. First, the formula was derived under the assumption that the pre-intervention probability P is given by the product of Eq. (2), which represents general domain knowledge prior to making any specific observation. Second, the formula in Eq. (6) is not very convenient in practical computations, since the joint distribution $P(x_1, ..., x_n)$ is not represented explicitly, but implicitly, in the form of probabilistic sentences from which it can be computed. Finally, the formula in Eq. (6) presumes that we have sufficient information at hand to define a complete joint distribution function. In practice, a complete specification of P is rarely available, and we must predict the effect of actions from a knowledge base containing unstructured collection of probabilistic statements, some are observational and some causal.

The first issue is treated in [Pearl 1993a], where assumptions about persistence were added to distinguish properties that terminate as a result of an action from those that persist despite that action. This paper addresses the latter two issues, and offers a set of sound (and possibly complete) inference rules by which probabilistic sentences involving actions and observations can be transformed to other such sentences, thus providing a syntactic method of deriving (or verifying) claims about actions and observations. We will assume, however, that the knowledge base contains the topological structure of the causal network G, some of its links are annotated with conditional probabilities while others remain unspecified. Given such partially specified causal theory, our main problem will be to facilitate the syntactic derivation of expressions of the form $P(x_j|do(x_i))$.

3 A Calculus of Actions

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 $P(y|do(x), z)$ 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 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 "do(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 $do(X = x)$, hence the graph $G_{\overline{X}}$.

Rule 2 provides conditions for an external intervention $do(Z = z)$ 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 $do(Z = z)$ without affecting the probability of $Y = y$. The validity of this rule stems, again, from simulating the intervention $do(Z = z)$ 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 causal theory $< P(x, y, z), G >$ where G is the graph given in Figure 2 below, and $P(x, y, z)$ is the distribution over the

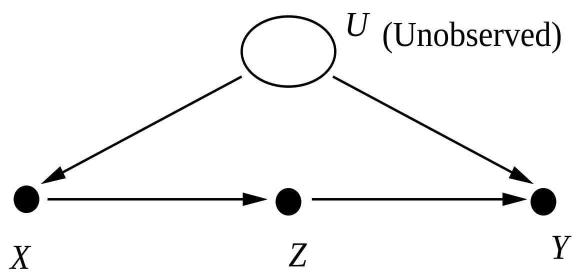

Figure 2

observed variables X, Y, Z. 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 our calculus, the effect of every action on every observed variable.

The applicability of each inference rule requires that certain d-separation conditions hold in some graph, the structure of which would vary with the expressions to be manipulated. Figure 3 displays the graphs that will be needed for the derivations that follow.

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

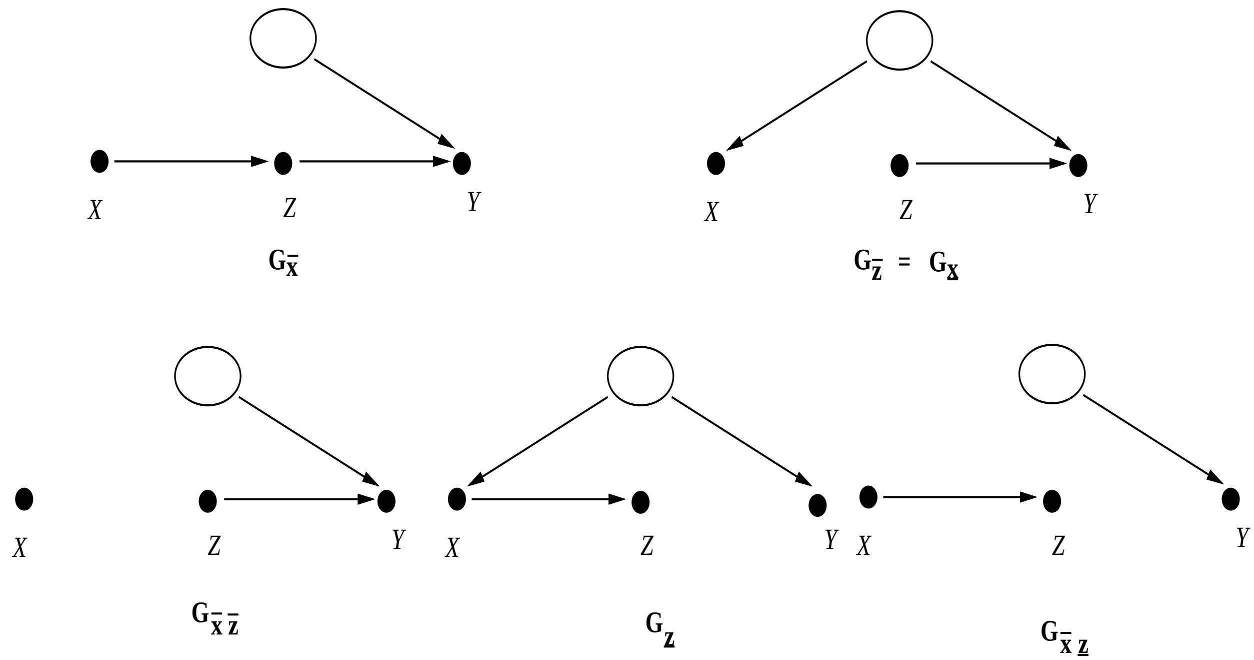

Figure 3

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 conditioning on variables (such as X) that reside on that path. Symbolically, this 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.

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

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

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

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. Action 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

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

In contrast to the "back-door" formula of (16), Eq. (20) computes the causal effect of X on Y using an intermediate variable Z.

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 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[5]. Similarly,

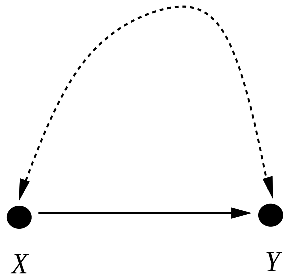

Figure 4

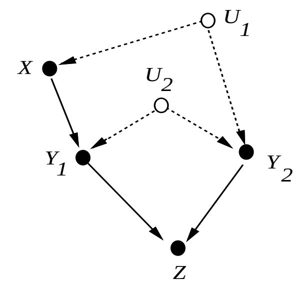

Figure 5

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.

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[6]. 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, giving

$$P(z|\hat{x}, \hat{y_2}) = \sum_{y_1} P(z|y_1, \hat{x}, \hat{y_2}) P(y_1|\hat{x}, \hat{y_2})$$

$$= \sum_{y_1} P(z|y_1, x_1, y_2) P(y_1|x)$$

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 $\{X_1, ...X_l\}$.

Computing the effect of actions from partial theories in which probabilities are specified on a select subset of (observed) variables is an extremely important task in statistics and socio-economic modeling, since it determines when a parameter of a causal theory are (so called) "identifiable" from non-experimental data, hence, when randomized experiments are not needed. The calculus proposed above, indeed uncovers possibilities that have remained unnoticed by economists and statisticians. For example, the structure of Figure 3 uncovers a class of observational studies in which the causal effect of an action (X) can be determined by measuring a variable (Z) that mediates the interaction between the action and its effect (Y). The relevance of such structures in practical situations can be seen, for instance, if we identify X with smoking, Y with lung cancer, Z with the amount of tar deposits in one's lung and U with an unobserved carcinogenic genotype which, according to the tobacco industry

[5] 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 $P(x, y, z)$.

[6] 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].

also induces an inborn crave for nicotine. Eq. (20) would provide us in this case with the means for quantifying, from non-experimental data, the causal effect of smoking on cancer. (Assuming, of course, that the data $P(x, y, z)$ is made available, and that we believe that smoking does not have a direct effect on lung cancer except that mediated by tar deposits).

However, our calculus is not limited to the derivation of causal probabilities from non-causal probabilities; we can reverse the role, and derive conditional and causal probabilities from causal expressions as well. For example, given the graph of figure 3 together with the quantities $P(z|\hat{x})$ and $P(y|\hat{z})$, we can derive an expression for $P(y|\hat{x})$,

$$P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}) \quad (22)$$

using the steps that led to Eq. (19). Note that the derivation is still valid when we add a common cause to X and Z, which is the most general condition under which the transitivity of causal relationships holds.

BIBLIOGRAPHY

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

Dalal, M. (1988) Investigations into a Theory of Knowledge Base Revision: Preliminary Report. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, 475-479.

Darwiche, A. (1993) Argument Networks: Explicating independence in propositional databases, in *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, (eds. D. Heckerman and A. Mamdani), 420-427.

Dean, T., and K. Kanazawa (1989) A Model for Reasoning about Persistence and Causation. *Computational Intelligence*, **5**, 142-150.

Dean, T. and M. Wellman (1991) *Planning and Control*, Morgan Kaufmann, San Mateo, CA.

Druzdzel, M.J., and H.A. Simon (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, 661-672.

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

Kushmerick, N., S. Hanks, and D. Weld (1993) An Algorithm for Probabilistic Planning. *Technical Report 93-06-03.* Department of Computer Science and Engineering, University of Washington.

Lewis, D.K. (1974) Probability of Conditionals and Conditional Probabilities. *The Philosophical Review*, **85**, 297-315.

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. 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, 441-452.

Pratt, J., and R. Schlaifer (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 Research 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.

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

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