On Identifying Causal Effects

Jin Tian
Department of Computer Science
Iowa State University
Ames, IA 50011
jtian@cs.iastate.edu

Ilya Shpitser
Department of Epidemiology
Harvard School of Public Health
ishpitse@hsph.harvard.edu

August 22, 2009

Abstract

A variety of questions in causal inference can be represented as probability distributions over hypothetical worlds where idealized randomized experiments known as interventions have taken place. Some such questions are really questions of *causal effect* of a particular intervention, while others are counterfactual and consider results of interventions which violate the state of affairs actually observed. Randomized experiments are expensive and often illegal. It is therefore imperative to find ways of evaluating, or identifying causal effect and counterfactual questions from available information, and causal assumptions.

In this paper, we review the state of the art in identification of causal effects and related counterfactual quantities in the framework of graphical causal models, a formalism where a causal domain of interest is represented by directed acyclic graphs with vertices representing variables of interest, and arrows representing direct causal influences.

1 Introduction

This paper deals with the problem of inferring cause-effect relationships from a combination of data and theoretical assumptions. This problem arises in diverse fields such as artificial intelligence, statistics, cognitive science, economics, and the health and social sciences. For example, investigators in the health and social sciences are often required to elucidate cause-effect relationships (e.g., the effects

of treatments on diseases) from observational studies of populations under natural conditions. Policymakers are concerned with the effects of policy decisions. One of the goals of artificial intelligence research is constructing agents able to create and execute plans in uncertain environments where trying actions to observe their effects directly is costly.

To estimate causal effects, scientists normally perform randomized experiments where a sample of units drawn from the population of interest is subjected to the specified manipulation directly. In many cases, however, such a direct approach is not possible due to expense or ethical considerations. Instead, investigators have to rely on observational studies to infer effects. A fundamental question in causal analysis is to determine when effects can be inferred from statistical information, encoded as a joint probability distribution, obtained under normal, intervention-free behavior. A key point here is that it is not possible to make causal conclusions from purely probabilistic premises – it is necessary to make causal assumptions. This is because without any assumptions it is possible to construct multiple "causal stories" which can disagree wildly on what effect a given intervention can have, but agree precisely on all observables. For instance, smoking may be highly correlated with lung cancer either because it causes lung cancer, or because people who are genetically predisposed to smoke may also have a gene responsible for a higher cancer incidence rate. In the latter case there will be no effect of smoking on cancer.

In this paper, we assume that the causal assumptions will be represented in by directed acyclic causal graphs [Pearl, 2000, Spirtes *et al.*, 2001] in which arrows represent the potential existence of direct causal relationships between the corresponding variables and some variables are presumed to be unobserved. Our task will be to decide whether the qualitative causal assumptions represented in any given graph are sufficient for assessing the strength of causal effects from nonexperimental data.

This problem of identifying causal effects has received considerable attention in the statistics, epidemiology, and causal inference communities [Robins, 1986, Robins, 1987, Pearl, 1993, Robins, 1997, Kuroki and Miyakawa, 1999, Glymour and Cooper, 1999, Pearl, 2000, Spirtes *et al.*, 2001]. In particular Judea Pearl and his colleagues have made major contributions in solving the problem. In his seminal paper Pearl (1995) established a *calculus of interventions* known as *do-calculus* - three inference rules by which probabilistic sentences involving interventions and observations can be transformed into other such sentences, thus providing a syntactic method of deriving claims about interventions. Later, *do*-calculus was shown to be complete for identifying causal effects, that is, every causal effects that can be identified can be derived using the three *do*-calculus rules[Shpitser and Pearl, 2006a, Huang and Valtorta, 2006b]. Pearl (1995) also established the popular "back-door" and "front-door" criteria - sufficient graphical conditions for ensuring identification of causal effects. Using *do*-calculus as a guide, Pearl and his collaborators developed a number of sufficient graphical criteria: a criterion for identifying causal

effects between singletons that combines and expands the "front-door" and "back-door" criteria [Galles and Pearl, 1995], a condition for evaluating the effects of plans in the presence of unmeasured variables, each plan consisting of several concurrent or sequential actions [Pearl and Robins, 1995]. More recently, an approach based on c-componenet factorization has been developed in [Tian and Pearl, 2002a, Tian and Pearl, 2003] and complete algorithms for identifying causal effects have been established [Tian and Pearl, 2003, Shpitser and Pearl, 2006b].

In this paper, we summarize the state of the art in identification of causal effects. The rest of the paper is organized as follows. Section 2 introduces causal models and gives formal definition for the identifiability problem. Section 3 presents Pearl's *do*-calculus and a number of easy to use graphical criteria. Section 4 presents the results on identifying (unconditional) causal effects. Section 5 shows how to identify conditional causal effects. Section 6 considers identification of counterfactual quantities which arise when we consider effects of additive interventions. Section 7 concludes the paper.

2 Notation, Definitions, and Problem Formulation

In this section we review the graphical causal models framework and introduce the problem of identifying causal effects.

2.1 Causal Bayesian Networks and Interventions

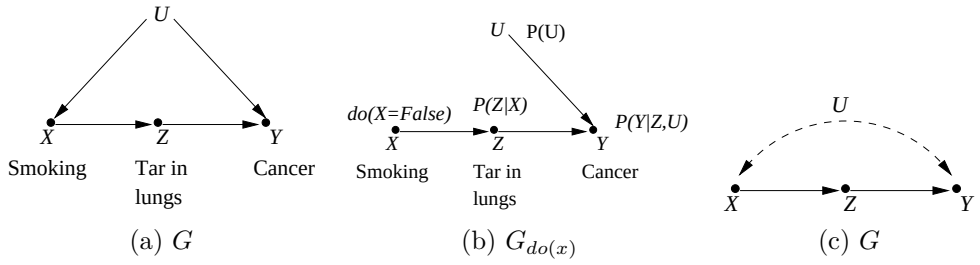
The use of graphical models for encoding distributional and causal assumptions is now fairly standard [Heckerman and Shachter, 1995, Lauritzen, 2000, Pearl, 2000, Spirtes *et al.*, 2001]. A *causal Bayesian network* consists of a DAG G over a set $V = \{V_1, \ldots, V_n\}$ of variables, called a *causal diagram*. The interpretation of such a graph has two components, probabilistic and causal. The probabilistic interpretation views G as representing conditional independence assertions: Each variable is independent of all its non-descendants given its direct parents in the graph.[1] These assertions imply that the joint probability function $P(v) = P(v_1, \ldots, v_n)$ factorizes according to the product [Pearl, 1988]

$$P(v) = \prod_i P(v_i|pa_i) \tag{1}$$

where pa_i are (values of) the parents of variable V_i in the graph. Here use uppercase letters to represent variables or sets of variables, and use corresponding lowercase letters to represent their values (instantiations).

The causal interpretation views the arrows in G as representing causal influences between the corresponding variables. In this interpretation, the factorization

[1] We use family relationships such as "parents," "children," and "ancestors" to describe the obvious graphical relationships.



(a) G (b) $G_{do(x)}$ (c) G

Figure 1: A causal diagram illustrating the effect of smoking on lung cancer

of (1) still holds, but the factors are further assumed to represent autonomous data-generation processes, that is, each parents-child relationship characterized by a conditional probability $P(v_i|pa_i)$ represents a stochastic process by which the values of V_i are assignedin response to the values pa_i (previously chosen for V_i's parents), and the stochastic variation of this assignment is assumed independent of the variations in all other assignments in the model. Moreover, each assignment process remains invariant to possible changes in the assignment processes that govern other variables in the system. This modularity assumption enables us to infer the effects of interventions, such as policy decisions and actions, whenever interventions are described as specific modifications of some factors in the product of (1). The simplest such intervention, called *atomic*, involves fixing a set T of variables to some constants $T = t$ denoted by $do(T = t)$ or $do(t)$, which yields the post-intervention distribution[2]

$$P_t(v) = \begin{cases} \prod_{\{i|V_i \notin T\}} P(v_i|pa_i) & v \text{ consistent with } t. \\ 0 & v \text{ inconsistent with } t. \end{cases} \quad (2)$$

Eq. (2) represents a truncated factorization of (1), with factors corresponding to the manipulated variables removed. This truncation follows immediately from (1) since, assuming modularity, the post-intervention probabilities $P(v_i|pa_i)$ corresponding to variables in T are either 1 or 0, while those corresponding to unmanipulated variables remain unaltered. If T stands for a set of treatment variables and Y for an outcome variable in $V \setminus T$, then Eq. (2) permits us to calculate the probability $P_t(y)$ that event $Y = y$ would occur if treatment condition $T = t$ were enforced uniformly over the population. This quantity, often called the "causal effect" of T on Y, is what we normally assess in a controlled experiment with T randomized, in which the distribution of Y is estimated for each level t of T.

As an example, consider the model shown in Figure 1(a) from [Pearl, 2000] that concerns the relations between smoking (X) and lung cancer (Y), mediated by the amount of tar (Z) deposited in a person's lungs. The model makes qualitative

[2] [Pearl, 1995, Pearl, 2000] used the notation $P(v|set(t))$, $P(v|do(t))$, or $P(v|\hat{t})$ for the post-intervention distribution, while [Lauritzen, 2000] used $P(v||t)$.

causal assumptions that the amount of tar deposited in the lungs depends on the level of smoking (and external factors) and that the production of lung cancer depends on the amount of tar in the lungs but smoking has no effect on lung cancer except as mediated through tar deposits. There might be (unobserved) factors (say some unknown carcinogenic genotype) that affect both smoking and lung cancer, but the genotype nevertheless has no effect on the amount of tar in the lungs except indirectly (through smoking). Quantitatively, the model induces the joint distribution factorized as

$$P(u, x, z, y) = P(u)P(x|u)P(z|x)P(y|z, u). \tag{3}$$

Assume that we could perform an ideal intervention on variable X by banning smoking[3], then the effect of this action is given by

$$P_{X=False}(u, z, y) = P(u)P(z|X = False)P(y|z, u), \tag{4}$$

which is represented by the model in Figure 1(b).

2.2 The Identifiability Problem

We see that, whenever all variables in V are observed, given the causal graph G, all causal effects can be computed from the observed distribution $P(v)$ as given by Eq. (2). However, if some variables are not measured, or two or more variables in V are affected by unobserved confounders, then the question of *identifiability* arises. The presence of such confounders would not permit the decomposition of the observed distribution $P(v)$ in (1). For example, in the model shown in Figure 1(a), assume that the variable U (unknown genotype) is unobserved and we have collected a large amount of data summarized in the form of (an estimated) join distribution P over the observed variables (X, Y, Z). We wish to assess the causal effect $P_x(y)$ of smoking on lung cancer.

Let V and U stand for the sets of observed and unobserved variables, respectively. If each U variable is a root node with exactly two observed children, then the corresponding model is called a *semi-Markovian* model. In this paper, we will focus on semi-Markovian models as they have simpler structures and it has been shown that causal effects in a model with arbitrary sets of unobserved variables can be identified by first projecting the model into a semi-Markovian model [Tian and Pearl, 2002b, Huang and Valtorta, 2006a].

In a semi-Markovian model, the observed probability distribution, $P(v)$, becomes a mixture of products:

$$P(v) = \sum_u \prod_i P(v_i|pa_i, u^i)P(u) \tag{5}$$

[3]Whether or not any actual action is an ideal manipulation of a variable (or is feasible at all) is not part of the theory - it is input to the theory.

where Pa_i and U^i stand for the sets of the observed and unobserved parents of V_i respectively, and the summation ranges over all the U variables. The post-intervention distribution, likewise, will be given as a mixture of truncated products

$$P_t(v) = \begin{cases} \sum_u \prod_{\{i|V_i \notin T\}} P(v_i|pa_i, u^i)P(u) & v \text{ consistent with } t. \\ 0 & v \text{ inconsistent with } t. \end{cases} \quad (6)$$

And, the question of identifiability arises, i.e., whether it is possible to express some causal effect $P_t(s)$ as a function of the observed distribution $P(v)$, independent of the unknown quantities, $P(u)$ and $P(v_i|pa_i, u^i)$.

It is convenient to represent a semi-Markovian model with a graph G that does not show the elements of U explicitly but, instead, represents the confounding effects of U variables using (dashed) bidirected edges. A bidirected edge between nodes V_i and V_j represents the presence of unobserved confounders that may influence both V_i and V_j. For example the model in Figure 1(a) will be represented by the graph in Figure 1(c).

In general we may be interested in identifying conditional causal effects $P_t(s|c)$, the causal effects of T on S conditioned on another set C of variables. This problem is important for evaluating *conditional plans* and stochastic plans [Pearl and Robins, 1995], where action T is taken to respond in a specified way to a set C of other variables – say, through a functional relationship $t = g(c)$. The effects of such actions may be evaluated through identifying conditional causal effects in the form of $P_t(s|c)$ [Pearl, 2000, chapter 4].

Definition 1 (Causal-Effect Identifiability) *The* causal effect *of a set of variables T on a disjoint set of variables S conditioned on another set C is said to be* identifiable *from a graph G if the quantity $P_t(s|c)$ can be computed uniquely from any positive probability of the observed variables—that is, if $P_t^{M_1}(s|c) = P_t^{M_2}(s|c)$ for every pair of models M_1 and M_2 with $P^{M_1}(v) = P^{M_2}(v) > 0$ and $G(M_1) = G(M_2) = G$.*

3 *Do*-calculus and Graphical Criteria

In general the identifiability of causal effects can be decided using Pearl's *do*-calculus – a set of inference rules by which probabilistic sentences involving interventions and observations can be transformed into other such sentences. A finite sequence of syntactic transformations, each applying one of the inference rules, may reduce expressions of the type $P_t(s)$ to subscript-free expressions involving observed quantities.

Let X, Y, and Z be arbitrary disjoint sets of nodes in G. We denote by $G_{\overline{X}}$ the graph obtained by deleting from G all arrows pointing to nodes in X. We denote by $G_{\underline{X}}$ the graph obtained by deleting from G all arrows emerging from nodes in X.

Theorem 1 (Rules of *do*-Calculus) *[Pearl, 1995] For any disjoint subsets of variables X, Y, Z, and W we have the following rules.*

Rule 1 (*Insertion/deletion of observations*) :

$$P_x(y|z,w) = P_x(y|w) \quad \textit{if} \quad (Y \perp\!\!\!\perp Z|X,W)_{G_{\overline{X}}}. \tag{7}$$

Rule 2 (*Action/observation exchange*) :

$$P_{x,z}(y|w) = P_x(y|z,w) \quad \textit{if} \quad (Y \perp\!\!\!\perp Z|X,W)_{G_{\overline{X}\underline{Z}}}. \tag{8}$$

Rule 3 (*Insertion/deletion of actions*) :

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

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

Theorem 2 *Do-calculus is complete for identifying causal effects of the form $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$.*

In principle we can apply *do*-calculus to identify any causal effects. The difficulty lies in that there is no general heuristics as to how to use those inference rules, that is, there is no general guidance on which *do*-calculus rule to apply at each step so as to finally decide whether a causal effect is identifiable or not.

In practice, there are a number of graphical criteria which can be used for quickly judging the identifiability by looking at the causal graph G.

Definition 2 (Back-Door) *A set of variables Z satisfies the* back-door *criterion relative to an ordered pair of variables (X_i, X_j) in a DAG G if:*

(i) *no node in Z is a descendant of X_i; and*

(ii) *Z blocks every path between X_i and X_j that contains an arrow into X_i.*

Similarly, if X and Y are two disjoint subsets of nodes in G, then Z is said to satisfy the back-door criterion relative to (X, Y) if it satisfies the criterion relative to any pair (X_i, X_j) such that $X_i \in X$ and $X_j \in Y$.

The name "back-door" echoes condition (ii), in which the paths with arrows pointing at X_i are called back door.

Theorem 3 (Back-Door Criteria) *[Pearl, 1995] If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula*

$$P_x(y) = \sum_z P(y|x,z)P(z). \tag{10}$$

For example, in Figure 1(c) X satisfies the back-door criterion relative to (Z, Y) and we have

$$P_z(y) = \sum_x P(y|x, z)P(x) \tag{11}$$

Definition 3 (Front-Door) *A set of variables Z is said to satisfy the* front-door *criterion relative to an ordered pair of variables (X, Y) if:*

(i) *Z intercepts all directed paths from X to Y;*

(ii) *there is no back-door path from X to Z; and*

(iii) *all back-door paths from Z to Y are blocked by X.*

Theorem 4 (Front-Door Criterion) *[Pearl, 1995] If Z satisfies the front-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula*

$$P_x(y) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x'). \tag{12}$$

For example, in Figure 1(c) Z satisfies the front-door criterion relative to (X, Y) and the causal effect $P_x(y)$ is given by Eq. (12).

There is a simple yet powerful graphical criterion for identifying the causal effects of a singleton. For any set S, let $An(S)$ denote the union of S and the set of ancestors of the variables in S. For any set C, let G_C denote the subgraph of G composed only of variables in C.

Theorem 5 *[Tian and Pearl, 2002a] $P_x(s)$ is identifiable if there is no bidirected path connecting X to any of its children in $G_{An(S)}$.*

In fact, for X and S being singletons, this criterion covers both "back-door" and "front-door" [Tian and Pearl, 2002a], and also the criteria in [Galles and Pearl, 1995].

These criteria are simple to use but are not necessary for identification. In the next sections we present complete systematic procedures for identification.

4 Identification of $P_t(s)$

In this section, we present a systematic procedure for identifying causal effects $P_t(s)$ using so-called c-component decomposition.

4.1 C-component Decomposition

Let a path composed entirely of bidirected edges be called a *bidirected path*. The set of variables V in G can be partitioned into disjoint groups by assigning two variables to the same group if and only if they are connected by a bidirected path.

Assuming that V is thus partitioned into k groups $S_1, \ldots, S_k$, each set S_j is called a *c-component* of V in G or a c-component of G. For example, the graph in Figure 1(c) consists of two c-components $\{X, Y\}$ and $\{Z\}$.

For any set $C \subseteq V$, define the quantity $Q[C](v)$ to denote the post-intervention distribution of C under an intervention to all other variables:[4]

$$Q[C](v) = P_{v \setminus c}(c) = \sum_{u} \prod_{\{i|V_i \in C\}} P(v_i|pa_i, u^i)P(u). \tag{13}$$

In particular, we have $Q[V](v) = P(v)$. For convenience, we will often write $Q[C](pa(C))$ as $Q[C]$. If there is no bidirected edges connected with a variable V_i, then $U^i = \emptyset$ and $Q[\{V_i\}] = P(v_i|pa_i)$.

The importance of the c-component steps from the following lemma.

Lemma 1 (C-component Decomposition) *[Tian and Pearl, 2002a] Assuming that V is partitioned into c-components $S_1, \ldots, S_k$, we have*

(i) $P(v) = \prod_i Q[S_i]$.

(ii) Each $Q[S_i]$ is computable from $P(v)$. Let a topological order over V be $V_1 < \ldots < V_n$, and let $V^{(i)} = \{V_1, \ldots, V_i\}$, $i = 1, \ldots, n$, and $V^{(0)} = \emptyset$. Then each $Q[S_j]$, $j = 1, \ldots, k$, is given by

$$Q[S_j] = \prod_{\{i|V_i \in S_j\}} P(v_i|v^{(i-1)}) \tag{14}$$

The lemma says that for each c-component S_i the causal effect $Q[S_i] = P_{v \setminus s_i}(s_i)$ is identifiable. For example, in Figure 1(c), we have $P_{x,y}(z) = Q[\{Z\}] = P(z|x)$ and $P_z(x, y) = Q[\{X, Y\}] = P(y|x, z)P(x)$.

Lemma 1 can be generalized to the subgraphs of G as given in the following lemma.

Lemma 2 (Generalized C-component Decomposition) *[Tian and Pearl, 2003] Let $H \subseteq V$, and assume that H is partitioned into c-components $H_1, \ldots, H_l$ in the subgraph G_H. Then we have*

(i) $Q[H]$ decomposes as

$$Q[H] = \prod_i Q[H_i]. \tag{15}$$

(ii) Each $Q[H_i]$ is computable from $Q[H]$. Let k be the number of variables in H, and let a topological order of the variables in H be $V_{h_1} < \cdots < V_{h_k}$ in G_H. Let $H^{(i)} = \{V_{h_1}, \ldots, V_{h_i}\}$ be the set of variables in H ordered before V_{h_i} (including V_{h_i}), $i = 1, \ldots, k$, and $H^{(0)} = \emptyset$. Then each $Q[H_j]$, $j = 1, \ldots, l$, is given by

$$Q[H_j] = \prod_{\{i|V_{h_i} \in H_j\}} \frac{Q[H^{(i)}]}{Q[H^{(i-1)}]}, \tag{16}$$

[4] Set $Q[\emptyset](v) = 1$ since $\sum_u P(u) = 1$.

where each $Q[H^{(i)}]$, $i = 0, 1, \ldots, k$, *is given by*

$$Q[H^{(i)}] = \sum_{h \setminus h^{(i)}} Q[H]. \tag{17}$$

Lemma 2 says that if the causal effect $Q[H] = P_{v \setminus h}(h)$ is identifiable then for each c-component H_i of the subgraph G_H the causal effect $Q[H_i] = P_{v \setminus h_i}(h_i)$ is identifiable.

Next, we show how to use Lemmas 1 and 2 to identify causal effects.

4.2 Computing $P_t(s)$

First we present a facility lemma. For $W \subseteq C \subseteq V$, the following lemma gives a condition under which $Q[W]$ can be computed from $Q[C]$ by summing over $C \setminus W$, like ordinary marginalization in probability theory.

Lemma 3 *[Tian and Pearl, 2003] Let* $W \subseteq C \subseteq V$, *and* $W' = C \setminus W$. *If* W *contains its own ancestors in the subgraph* G_C *(*$An(W)_{G_C} = W$*), then*

$$\sum_{w'} Q[C] = Q[W]. \tag{18}$$

Note that we always have $\sum_c Q[C] = 1$.

Next, we show how to use Lemmas 1–3 to identify the causal effect $P_t(s)$ where S and T are arbitrary (disjoint) subsets of V. We have

$$P_t(s) = \sum_{(v \setminus t) \setminus s} P_t(v \setminus t) = \sum_{(v \setminus t) \setminus s} Q[V \setminus T]. \tag{19}$$

Let $D = An(S)_{G_{V \setminus T}}$. Then by Lemma 3, variables in $(V \setminus T) \setminus D$ can be summed out:

$$P_t(s) = \sum_{d \setminus s} \sum_{(v \setminus t) \setminus d} Q[V \setminus T] = \sum_{d \setminus s} Q[D]. \tag{20}$$

Assume that the subgraph G_D is partitioned into c-components $D_1, \ldots, D_l$. Then by Lemma 2, $Q[D]$ can be decomposed into products of $Q[D_i]$'s, and Eq. (20) can be rewritten as

$$P_t(s) = \sum_{d \setminus s} \prod_i Q[D_i]. \tag{21}$$

We obtain that $P_t(s)$ is identifiable if all $Q[D_i]$'s are identifiable.

Let G be partitioned into c-components $S_1, \ldots, S_k$. Then any D_i is a subset of certain S_j since if the variables in D_i are connected by a bidirected path in a subgraph of G then they must be connected by a bidirected path in G. Assuming $D_i \subseteq S_j$, $Q[D_i]$ is identifiable if it is computable from $Q[S_j]$. In general, for $C \subseteq$

Algorithm Identify(C, T, Q)
INPUT: $C \subseteq T \subseteq V$, $Q = Q[T]$. G_T and G_C are both composed of one single c-component.
OUTPUT: Expression for $Q[C]$ in terms of Q or FAIL.

Let $A = An(C)_{G_T}$.

- IF $A = C$, output $Q[C] = \sum_{t \setminus c} Q$.
- IF $A = T$, output FAIL.
- IF $C \subset A \subset T$
  1. Assume that in G_A, C is contained in a c-component T'.
  2. Compute $Q[T']$ from $Q[A] = \sum_{t \setminus a} Q$ by Lemma 2.
  3. Output Identify$(C, T', Q[T'])$.

Figure 2: An algorithm for determining if $Q[C]$ is computable from $Q[T]$.

$T \subseteq V$, whether $Q[C]$ is computable from $Q[T]$ can be determined recursively by repeated applications of Lemma 3 and 2, as given in the recursive algorithm shown in Figure 2. At each step of the algorithm, we either find an expression for $Q[C]$, find $Q[C]$ unidentifiable, or reduce the problem to a simpler one.

In summary, an algorithm for computing $P_t(s)$ is given in Figure 3 and the algorithm has been shown to be complete.

Theorem 6 *[Shpitser and Pearl, 2006b, Huang and Valtorta, 2006a] The algorithm* ***ID*** *in Figure 3 is complete.*

5 Identification of Conditional Causal Effects

An important refinement to the problem of identifying causal effects $P(\mathbf{y}|do(\mathbf{x}))$ is concerned with identifying *conditional causal effects*, in other words causal effects in a particular subpopulation where variables $\mathbf{Z}$ are known to attain values $\mathbf{z}$. These conditional causal effects are written as $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$, and defined just as regular conditional distributions as

$$P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = \frac{P_{\mathbf{x}}(\mathbf{y}, \mathbf{z})}{P_{\mathbf{x}}(\mathbf{z})}$$

Despite the fact that do-calculus is complete for identifying such effects, it is desirable to obtain a closed form algorithm which can be applied in polynomial time, since this is preferable to searching for a valid do-calculus derivation, which, absent a general purpose heuristic, could take a long time.

Algorithm ID$(s, t, P(.), G)$
INPUT: two disjoint sets $S, T \subset V$.
OUTPUT: the expression for $P_t(s)$ or FAIL.
Phase-1:

1. Find the c-components of G: $S_1, \ldots, S_k$. Compute each $Q[S_i]$ by Lemma 1.
2. Let $D = An(S)_{G_{V \setminus T}}$ and the c-components of G_D be D_i, $i = 1, \ldots, l$.

Phase-2:
For each set D_i such that $D_i \subseteq S_j$:
Compute $Q[D_i]$ from $Q[S_j]$ by calling **Identify**$(D_i, S_j, Q[S_j])$ in Figure 2. If the function returns FAIL, then stop and output FAIL.

Phase-3: Output $P_t(s) = \sum_{D \setminus S} \prod_i Q[D_i]$.

Figure 3: A complete algorithm for computing $P_t(s)$.

One existing approach [Tian, 2004], generalizes the algorithm for identifying *unconditional* causal effects $P_{\mathbf{x}}(\mathbf{y})$ found in section 4. There is, however, an easier approach which works.

The idea is to reduce the expression $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$, which we don't know how to handle to something like $P_{\mathbf{x}'}(\mathbf{y}')$, which we do know how to handle via the algorithm already presented. This reduction would have to find a way to get rid of variables $\mathbf{z}$ in the conditional effect expression.

Ridding ourselves of some variables in $\mathbf{Z}$ can be accomplished via rule 2 of do-calculus. Recall that applying rule 2 to an expression allows us to replace conditioning on some variable set $\mathbf{W} \subseteq \mathbf{Z}$ by fixing $\mathbf{W}$ instead. Rule 2 states that this is possible in the expression $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ whenever $\mathbf{W}$ contains no back-door paths to $\mathbf{Y}$ conditioned on the remaining variables in $\mathbf{Z}$ and $\mathbf{X}$ (that is $\mathbf{X} \cup \mathbf{Z} \setminus \mathbf{W}$), in the graph where all incoming arrows to $\mathbf{X}$ have been cut.

It's not difficult to show the following uniqueness theorem.

Lemma 4 ([Shpitser and Pearl, 2006a]) *For every conditional effect $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ there exists a unique maximal $\boldsymbol{W} \subseteq \boldsymbol{Z}$ such that $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ is equal to $P_{\boldsymbol{x},\boldsymbol{w}}(\boldsymbol{y}|\boldsymbol{z} \setminus \boldsymbol{w})$ according to rule 2 of do-calculus.*

Lemma 4 states that we only need to apply rule 2 once to rid ourselves of as many conditioned variables as possible in the effect of interest. However, even after this is done, we may be left with some variables in $\mathbf{Z} \setminus \mathbf{W}$ past the conditioning bar in our effect expression. If we insist on using unconditional effect identification, we may try to identify the joint distribution $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}, \mathbf{z} \setminus \mathbf{w})$ to obtain an expression α, and obtain the conditional distribution $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z} \setminus \mathbf{w})$ by taking $\frac{\alpha}{\sum_{\mathbf{y}} \alpha}$. But what

function **IDC**(**y**, **x**, **z**, P, G)
INPUT: **x**,**y**,**z** value assignments, P a probability
distribution, G a causal diagram (an I-map of P).
OUTPUT: Expression for $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ in terms of P or **FAIL**.

1 if $(\exists W \in \mathbf{Z})(\mathbf{Y} \perp\!\!\!\perp W|\mathbf{X}, \mathbf{Z} \setminus \{Z\})_{G_{\overline{\mathbf{X}},\underline{Z}}}$,
return $\mathbf{IDC}(\mathbf{y}, \mathbf{x} \cup \{w\}, \mathbf{z} \setminus \{w\}, P, G)$.

2 else let $P' = \mathbf{ID}(\mathbf{y} \cup \mathbf{z}, \mathbf{x}, P, G)$.
return $P'/\sum_{\mathbf{y}} P'$.

Figure 4: A complete identification algorithm for conditional effects.

if $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}, \mathbf{z} \setminus \mathbf{w})$ is not identifiable? Are there cases where $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}, \mathbf{z} \setminus \mathbf{w})$ is not identifiable, but $P_{\mathbf{x},\mathbf{w}}(\mathbf{y}|\mathbf{z} \setminus \mathbf{w})$ is? Fortunately it turns out the answer is no.

Lemma 5 ([Shpitser and Pearl, 2006a]) *Let $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ be a conditional effect of interest in a causal model inducing G, and $\boldsymbol{W} \subseteq \boldsymbol{Z}$ the unique maximal set such that $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ is equal to $P_{\boldsymbol{x},\boldsymbol{w}}(\boldsymbol{y}|\boldsymbol{z} \setminus \boldsymbol{w})$. Then $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ is identifiable from G and $P(\boldsymbol{v})$ if and only if $P_{\boldsymbol{x},\boldsymbol{w}}(\boldsymbol{y}, \boldsymbol{z} \setminus \boldsymbol{w})$ is identifiable from G and $P(\boldsymbol{v})$.*

Lemma 5 gives us a simple algorithm for identifying arbitrary conditional effects by first reducing the problem into one of identifying an unconditional effect – and then invoking the complete algorithm. This simple algorithm is actually complete since the statement in Theorem 5 is if and only if. The algorithm itself is shown in Fig. 4. The algorithm as shown picks elements W of $\mathbf{W}$ one at a time, although the set of W it picks as it iterates will equal the maximal set $\mathbf{W}$ due to the following lemma.

Lemma 6 *Let $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ be a conditional effect of interest in a causal model inducing G, and $\boldsymbol{W} \subseteq \boldsymbol{Z}$ the unique maximal set such that $P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z})$ is equal to $P_{\boldsymbol{x},\boldsymbol{w}}(\boldsymbol{y}|\boldsymbol{z} \setminus \boldsymbol{w})$. Then $\boldsymbol{W} = \{W|P_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{z}) = P_{\boldsymbol{x},w}(\boldsymbol{y}|\boldsymbol{z} \setminus \{w\})\}$.*

Completeness of the algorithm easily follows from the results we presented.

Theorem 7 ([Shpitser and Pearl, 2006a]) *The algorithm **IDC** is complete.*

6 Relative Interventions and the Effect of Treatment on the Treated

Interventions considered in the previous sections are what we term "absolute," since the values $\mathbf{x}$ to which variables are set by $do(\mathbf{x})$ bear no relationship to whatever natural values were assumed by variables $\mathbf{X}$ prior to an intervention. Such absolute

interventions correspond to clamping a wire in a circuit to ground, or performing a randomized clinical trial for a drug which does not naturally occur in the body.

By contrast, many interventions are *relative*, in other words, the precise level x to which the variable X is set depends on the values X naturally attains. A typical relative intervention is the addition of insulin to the bloodstream. Since insulin is naturally synthesized by the human body, the effect of such an intervention depends on the initial, pre-intervention concentration of insulin in the blood, even if a constant amount is added for every patient. The insulin intervention can be denoted by $do(i + X)$, where i is the amount of insulin added, and X denotes the random variable representing pre-intervention insulin concentration in the blood. More generally, a relative intervention on a variable X takes the form of $do(f(X))$ for some function f.

How are we to make sense of a relative intervention $do(f(X))$ on X applied to a given population where the values of X are not known? Can relative interventions be reduced to absolute interventions? It appears that in general the answer is "no." Consider: if we knew that X attained the value x for a given unit, then the the effect of an intervention in question on the outcome variable Y is really $P(y|do(f(x)), x)$. This expression is almost like the (absolute) conditional causal effect of $do(f(x))$ on y, except the evidence that is being conditioned on is on the same variable that is being intervened. Since x and $f(x)$ are not in general the same, it appears that this expression contains a kind of value conflict. Are these kinds of probabilities always 0? Are they even well defined?

In fact, expressions of this sort are a special case of a more general notion of a *counterfactual distribution*, which can be derived from functional causal models [Pearl, 2000], Chapter 7.

Such models consist of two sets of variables, the observable set $\mathbf{V}$ representing the domain of interest, and the unobservable set $\mathbf{U}$ representing the background to the model that we are ignorant of. Associated with each observable variable V_i in $\mathbf{V}$ is a function f_i which determines the value of V_i in terms of values of other variables in $\mathbf{V} \cup \mathbf{U}$. Finally, there is a joint probability distribution $P(\mathbf{u})$ over the unobservable variables, signifying our ignorance of the background conditions of the model.

The causal relationships in functional causal models are represented, naturally, by the functions f_i; each function causally determines the corresponding V_i in terms of its inputs. Causal relationships entailed by a given model have an intuitive visual representation using a graph called a causal diagram. Causal diagrams contain two kinds of edges. Directed edges are drawn from a variable X to a variable V_i if X appears as an input of f_i. Directed edges from the same unobservable U_i to two observables V_j, V_k can be replaced by a bidirected edge between V_j to V_k. We will consider models which induce acyclic graphs where $P(\mathbf{u}) = \prod_i P(u_i)$, and each U_i has at most two observable children. A graph obtained in this way from a model is said to be induced by said model.

Unlike causal Bayesian networks introduced in Section 2, functional causal models represent fundamentally deterministic causal relationships which only appear stochastic due to our ignorance of background variables. This inherent determinism allows us to define counterfactual distributions which spam multiple worlds under different interventions regimes. Formally, a joint counterfactual distribution is a distribution over events of the form $Y_{\mathbf{x}}$ where Y is a post-intervention random variable in a causal model (the intervention in question being $do(\mathbf{x})$). A single joint distribution can contain multiple such events, with different, possibly conflicting interventions.

Such joint distributions are defined as follows:

$$P(Y^1_{\mathbf{x}^1} = y^1, ..., Y^k_{\mathbf{x}^k} = y^k) = \sum_{\{\mathbf{u}|Y^1_{\mathbf{x}^1}(\mathbf{u})=y^1 \wedge ... \wedge Y^k_{\mathbf{x}^k}(\mathbf{u})=y^k\}} P(\mathbf{u})$$

where $\mathbf{U}$ is the set of unobserved variables in the model. In other words, a joint counterfactual probability is obtained by adding up the probabilities of every setting of unobserved variables in the model that results in the observed values of each counterfactual event $Y_{\mathbf{x}}$ in the expression. The query with the conflict we considered above can then be expressed as a conditional distribution derived from such a joint, specifically $P(Y_{f(x)} = y|X = x) = \frac{P(Y_{f(x)}=y, X=x)}{P(X=x)}$. Queries of this form are well known in the epidemiology literature as the effect of treatment (ETT) on the treated [Heckman, 1992, Robins *et al.*, 2006].

In fact, relative interventions aren't quite the same as ETT since we don't actually know the original levels of X. To obtain effects of relative interventions, we simply average over possible values of X, weighted by the prior distribution $P(x)$ of X. In other words, the relative causal effect $P(y|do(f(X)))$ is equal to $\sum_x P(Y_{f(x)} = y|X = x)P(X = x)$.

Since relative interventions reduce to ETT, and because ETT questions are of independent interest, identification of ETT is an important problem. If interventions are performed over multiple variables, it turns out that identifying ETT questions is almost as intricate as general counterfactual identification [Shpitser and Pearl, 2009], [Shpitser and Pearl, 2007]. However, in the case of a singleton intervention, there is a formulation which bypasses most of the complexity of counterfactual identification. This formulation is the subject of this section.

We want to approach identification of ETT in the same way we approached identification of causal effects in the previous section, namely by providing a graphical representation of conditional independences in joint distributions of interest, and then expressing the identification algorithm in terms of this graphical representation. In the case of causal effects, we were given as input the causal diagram representing the original, pre-intervention world, and we were asking questions about the post-intervention world where arrows pointing to intervened variables were cut. In the case of counterfactuals we are interested in joint distributions that span multiple worlds each with its own intervention. We want to construct a graph for these

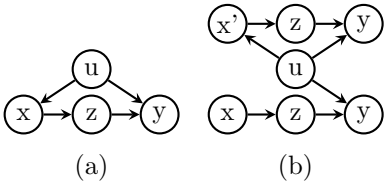


Figure 5: (a) A causal diagram G. (b) The counterfactual graph for $P(Y_x = y|x')$ in G.

distributions.

The intuition is that each interventional world is represented by a copy of the original causal diagram, with the appropriate incoming arrows cut to represent the changes in the causal structure due to the intervention. All worlds are assumed to share history up to the moment of divergence due to differing interventions. This is represented by all worlds sharing unobserved variables U. In the special case of two interventional worlds the resulting graph is known as the *twin network graph* [Balke and Pearl, 1994b, Balke and Pearl, 1994a].

In the general case, a refinement of the resulting graph (to account for the possibility of duplicate random variables) is known as the *counterfactual graph* [Shpitser and Pearl, 2007]. The counterfactual graph represents conditional independences in the corresponding counterfactual distribution via the d-separation criterion just as the causal diagram represents conditional independences in the observed distribution of the original world. A graph in Figure 5(b) is a counterfactual graph for the query $P(Y_x = y|X = x')$ obtained from the original causal diagram shown in Figure 5(a).

There exists a rather complicated general algorithm for identifying arbitrary counterfactual distributions from either interventional or observational data [Shpitser and Pearl, 2007], [Shpitser and Pearl, 2008], based on ideas from the causal effect algorithm from the previous section, only applied to the counterfactual graph, rather than the causal diagram. It turns out that while identifying ETT of a single variable X can be represented as an identification problem of ordinary causal effects, ETT of multiple variables is significantly more complex [Shpitser and Pearl, 2009]. In this paper, we will concentrate on single variable ETT with multiple outcome variables $\mathbf{Y}$.

What makes single variable ETT $P(\mathbf{Y}_x = \mathbf{y}|X = x')$ particularly simple is the form of its counterfactual graph. For the case of all ETTs, this graph will have variables from two worlds – the "natural" world where X is observed to have taken the value x' and the interventional world, where X is fixed to assume the value x. There two key points that simplify matters. The first is that no descendant of X (including variables in $\mathbf{Y}$) is of interest in the "natural" world, since we are only interested in the outcome $\mathbf{Y}$ in the interventional world. The second is that all non-descendants of X behave the same in both worlds (since interventions do

not affect non-descendants). Thus, when constructing the counterfactual graph we don't need to make copies of non-descendants of X, and we can ignore descendants of X in the "natural" world. But this means the only variable in the "natural" world we will construct is a copy of X itself.

What this implies is that a problem of identifying the ETT $P(\mathbf{Y}_x = \mathbf{y}|X = x')$ can be rephrased as a problem of identifying a certain conditional causal effect.

Theorem 8 ([Shpitser and Pearl, 2009]) *$P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is identifiable in G if and only if $P(\boldsymbol{y}|w, do(x))$ is identifiable in G', where G' is obtained from G by adding a new node W with the same set of parents (both observed and unobserved) as X, and no children. Moreover, the estimand for $P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is equal to that of $P(\boldsymbol{y}|w, do(x))$ with all occurrences of w replaced by x'.*

We illustrate the application of Theorem 8 by considering the graph G in Fig. 5 (a). The query $P(Y_x = y|X = x')$ is identifiable by considering $P(y|w, do(x))$ in the graph G' shown in Fig. 5 (b), while the counterfactual graph for $P(Y_x = y|x')$ is shown in Fig. 5 (c). Identifying $P(y|w, do(x))$ in G' using the algorithms in the previous section $\sum_z P(z|x) \sum_x P(y|z, w, x)P(w, x)/P(w)$. Replacing w by x' yields the expression $\sum_z P(z|x) \sum_{x''} P(y|z, x', x'')P(x', x'')/P(x')$.

Ordinarily, we know that $P(y|z, x', x'')$ is undefined if x' is not equal to x''. However, in our case, we know that observing $X = x'$ in the natural world implies $X = x'$ in any other interventional world which shares ancestors of X with the natural world. This implies the expression $\sum_{x''} P(y|z, x', x'')P(x', x'')/P(x')$ is equivalent to $P(y|z, x')$, thus our query $P(Y_x = y|X = x')$ is equal to $\sum_z P(y|z, x')P(z|x)$.

It is possible to use Theorem 8 to derive analogues of the Backdoor and Frontdoor criteria for ETT.

Corollary 1 (Backdoor Criterion for ETT) *If a set $\boldsymbol{Z}$ satisfies the Backdoor Criterion relative to $(X, \boldsymbol{Y})$, then $P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is identifiable and equal to $\sum_{\boldsymbol{z}} P(\boldsymbol{y}|\boldsymbol{z}, x)P(\boldsymbol{z}|x')$.*

The intuition for the Backdoor Criterion for ETT is that $\mathbf{Z}$, by assumption, screens X and Y from observed values of X in other counterfactual worlds. Thus, the first term in the Backdoor expression does not change. The second term changes in an obvious way since $\mathbf{Z}$ depends on observing $X = x'$.

Corollary 2 (Frontdoor Criterion for ETT) *If a set $\boldsymbol{Z}$ satisfies the Frontdoor Criterion relative to $(X, \boldsymbol{Y})$ in G, then $P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is identifiable and equal to $\sum_{\boldsymbol{z}} P(\boldsymbol{y}|\boldsymbol{z}, x')P(\boldsymbol{z}|x)$.*

Proof: We will be using a number of graphs in this proof. G is the original graph. G^w is the graph obtained from G by adding a copy of X called W with the same parents (including unobserved parents) as X and no children. G' is a graph representing independences in $P(X, \mathbf{Y}, \mathbf{Z})$. It is obtained from G by removing all nodes

other than $X, \mathbf{Y}, \mathbf{Z}$, by adding a directed arrow between any remaining A and B in $X, \mathbf{Y}, \mathbf{Z}$ if there is a d-connected path containing only nodes not in $X, \mathbf{Y}, \mathbf{Z}$ which starts with a directed arrow pointing away from A and ends with any arrow pointing to B. Similarly, a bidirected arrow is added between any A and B in $X, \mathbf{Y}, \mathbf{Z}$ if there is a d-connected path containing only nodes not in $X, \mathbf{Y}, \mathbf{Z}$ which starts with any arrow pointing to A and ends with any arrow pointing to B. (This graph is known as a latent projection [Pearl, 2000]). The graphs $G'^w, G'^w_{\overline{x}}$ are defined similarly as above.

We want to identify $P(\mathbf{y}, \mathbf{z}, w|do(x))$ in G'^w. First, we want to show that no node in $\mathbf{Z}$ shares a C-component with W or any node in $\mathbf{Y}$ in $G'^w_{\overline{x}}$. This can only happen if a node in $\mathbf{Z}$ and W or a node in $\mathbf{Y}$ share a bidirected arc in $G'^w_{\overline{x}}$. But this means that either there is a backdoor d-connected path from $\mathbf{Z}$ to $\mathbf{Y}$ in $G_{\overline{x}}$, or there is a backdoor d-connected path from X to $\mathbf{Z}$ in G. Both of these claims are contradicted by our assumption that $\mathbf{Z}$ satisfies the Frontdoor Criterion for $(X, \mathbf{Y})$.

This implies $P(\mathbf{y}, \mathbf{z}, w|do(x)) = P(\mathbf{y}, w|do(\mathbf{z}, x))P(\mathbf{z}|do(x, w))$ in G^w.

By construction of G^w and the Frontdoor Criterion, $P(\mathbf{z}|do(x, w)) = P(\mathbf{z}|do(x)) = P(\mathbf{z}|x)$. Furthermore, since no nodes in $\mathbf{Z}$ and $\mathbf{Y}$ share a C-component in G'^w, no node in $\mathbf{Z}$ has a bidirected path to $\mathbf{Y}$ in G'^w. This implies, by Lemma 1 in [Shpitser *et al.*, 2009], that $P(\mathbf{y}, w, x|do(\mathbf{z})) = P(\mathbf{y}|\mathbf{z}, w, x)P(w, x)$.

Since $\mathbf{Z}$ intercepts all frontdoor paths from X to $\mathbf{Y}$ (by the Frontdoor criterion), $P(\mathbf{y}, w|do(\mathbf{z}, x)) = P(\mathbf{y}, w|do(\mathbf{z})) = \sum_x P(\mathbf{y}|\mathbf{z}, w, x)P(w, x)$.

We conclude that $P(\mathbf{y}, w|do(x))$ is equal to $\sum_{\mathbf{z}} P(\mathbf{z}|x) \sum_x P(\mathbf{y}|\mathbf{z}, w, x)P(w, x)$. Since $P(w|do(x)) = P(w)$ in G'^w, $P(\mathbf{y}, w|do(x)) = \sum_{\mathbf{z}} P(\mathbf{z}|x) \sum_x P(\mathbf{y}|\mathbf{z}, w, x)P(x|w)$.

Finally, recall that W is just a copy of X, and X is observed to attain value x' in the "natural" world. This implies that our expression simplifies to $\sum_{\mathbf{z}} P(\mathbf{z}|x)P(\mathbf{y}|\mathbf{z}, x')$, which proves our result. □

If neither the Backdoor nor the Frontdoor criteria hold, we must invoke general causal effect identification algorithms from the previous section. However, in the case of ETT of a single variable, there is a simple complete graphical criterion which works.

Theorem 9 ([Shpitser and Pearl, 2009]) *$P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is identifiable from $P(\boldsymbol{v})$ if and only if there is no bidirected path from X to a child of X in $G_{an(\boldsymbol{y})}$. Moreover, if there is no such bidirected path, the estimand for $P(\boldsymbol{Y}_x = \boldsymbol{y}|X = x')$ is obtained by multiplying the estimand for $\sum_{an(\boldsymbol{y})\setminus(\boldsymbol{y}\cup\{x\})} P(an(\boldsymbol{y}) \setminus x|do(x))$ (which exists by results in [Tian and Pearl, 2002a]) by $\frac{Q[S^x]'}{P(x')\sum_x Q[S^x]}$, where S^x is the C-component in G containing X, and $Q[S^x]'$ is obtained from the expression for $Q[S^x]$ by replacing all occurrences of x with x'.*

7 Conclusion

In this paper we described the state of the art in identification of causal effects and related quantities in the framework of graphical causal models. We have shown how this framework, developed over the period of two decades by Judea Pearl and his collaborators, and presented in Pearl's seminal work [Pearl, 2000], can sharpen causal intuition into mathematical precision for a variety of causal problems faced by scientists.

References

[Balke and Pearl, 1994a] A. Balke and J. Pearl. Counterfactual probabilities: Computational methods, bounds, and applications. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*, pages 46–54. Morgan Kaufmann, San Mateo, CA, 1994.

[Balke and Pearl, 1994b] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume I, pages 230–237. MIT Press, Menlo Park, CA, 1994.

[Galles and Pearl, 1995] D. Galles and J. Pearl. Testing identifiability of causal effects. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 185–195. Morgan Kaufmann, San Francisco, 1995.

[Glymour and Cooper, 1999] C. Glymour and G. Cooper, editors. *Computation, Causation, and Discovery*. MIT Press, Cambridge, MA, 1999.

[Heckerman and Shachter, 1995] D. Heckerman and R. Shachter. Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research*, 3:405–430, 1995.

[Heckman, 1992] J.J. Heckman. Randomization and social policy evaluation. In C. Manski and I. Garfinkle, editors, *Evaluations: Welfare and Training Programs*, pages 201–230. Harvard University Press, 1992.

[Huang and Valtorta, 2006a] Y. Huang and M. Valtorta. Identifiability in causal bayesian networks: A sound and complete algorithm. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1149–1154, Menlo Park, CA, July 2006. AAAI Press.

[Huang and Valtorta, 2006b] Y. Huang and M. Valtorta. Pearl's calculus of interventions is complete. In R. Dechter and T.S. Richardson, editors, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*. AUAI Press, July 2006.

[Kuroki and Miyakawa, 1999] M. Kuroki and M. Miyakawa. Identifiability criteria for causal effects of joint interventions. *Journal of the Japan Statistical Society*, 29(2):105–117, 1999.

[Lauritzen, 2000] S. Lauritzen. Graphical models for causal inference. In O.E. Barndorff-Nielsen, D. Cox, and C. Kluppelberg, editors, *Complex Stochastic Systems*, chapter 2, pages 67–112. Chapman and Hall/CRC Press, London/Boca Raton, 2000.

[Pearl and Robins, 1995] J. Pearl and J.M. Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 444–453. Morgan Kaufmann, San Francisco, 1995.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 1993] J. Pearl. Comment: Graphical models, causality, and intervention. *Statistical Science*, 8:266–269, 1993.

[Pearl, 1995] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82:669–710, December 1995.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, NY, 2000.

[Robins *et al.*, 2006] James M. Robins, VanderWeele Tyler J., and Thomas S. Richardson. Comment on causal effects in the presence of non compliance: a latent variable interpretation by antonio forcina. *METRON*, LXIV(3):288–298, 2006.

[Robins, 1986] J.M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

[Robins, 1987] J.M. Robins. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases*, 40(Suppl 2):139S–161S, 1987.

[Robins, 1997] J.M. Robins. Causal inference from complex longitudinal data. In *Latent Variable Modeling with Applications to Causality*, pages 69–117. Springer-Verlag, New York, 1997.

[Shpitser and Pearl, 2006a] I. Shpitser and J. Pearl. Identification of conditional interventional distributions. In R. Dechter and T.S. Richardson, editors, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 437–444. AUAI Press, July 2006.

[Shpitser and Pearl, 2006b] I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1219–1226, Menlo Park, CA, July 2006. AAAI Press.

[Shpitser and Pearl, 2007] Ilya Shpitser and Judea Pearl. What counterfactuals can be tested. In *Twenty Third Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2007.

[Shpitser and Pearl, 2008] I. Shpitser and J. Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9:1941–1979, 2008.

[Shpitser and Pearl, 2009] Ilya Shpitser and Judea Pearl. Effects of treatment on the treated: Identification and generalization. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, volume 25, 2009.

[Shpitser *et al.*, 2009] Ilya Shpitser, Thomas S. Richardson, and James M. Robins. Testing edges by truncations. In *International Joint Conference on Artificial Intelligence*, volume 21, pages 1957–1963, 2009.

[Spirtes *et al.*, 2001] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search (2nd Edition)*. MIT Press, Cambridge, MA, 2001.

[Tian and Pearl, 2002a] J. Tian and J. Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI)*, pages 567–573, Menlo Park, CA, 2002. AAAI Press/The MIT Press.

[Tian and Pearl, 2002b] J. Tian and J. Pearl. On the testable implications of causal models with hidden variables. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002.

[Tian and Pearl, 2003] J. Tian and J. Pearl. On the identification of causal effects. Technical Report R-290-L, Department of Computer Science, University of California, Los Angeles, 2003.

[Tian, 2004] J. Tian. Identifying conditional causal effects. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2004.