Identifiability in Causal Bayesian Networks: A Gentle Introduction

Marco Valtorta
mgv@cse.sc.edu

Yimin Huang
huang6@cse.sc.edu

Department of Computer Science and Engineering
University of South Carolina

May 21, 2007

Abstract

In this paper we describe an important structure used to model causal theories and a related problem of great interest to semi-empirical scientists. A causal Bayesian network is a pair consisting of a directed acyclic graph (called a causal graph) that represents causal relationships and a set of probability tables, that together with the graph specify the joint probability of the variables represented as nodes in the graph. We briefly describe the probabilistic semantics of causality proposed by Pearl for this graphical probabilistic model, and how unobservable variables greatly complicate models and their application. A common question about causal Bayesian networks is the problem of identifying casual effects from nonexperimental data, which is called the identifability problem. In the basic version of this problem, a semi-empirical scientist postulates a set of causal mechanisms and uses them, together with a probability distribution on the observable set of variables in a domain of interest, to predict the effect of a manipulation on some variable of interest. We explain this problem, provide several examples, and direct the readers to recent work that provides a solution to the problem and some of its extensions. We assume that the Bayesian network structure is given to us and do not address the problem of learning it from data and the related statistical inference and testing issues.

1 Motivation

Flash back to the late 1950s. Evidence was mounting that smoking was bad for one's health. In particular, some researchers postulated a causal link between smoking and lung cancer. Such a model could be represented by the graph of Figure 1.

$$X \longrightarrow Y$$

Figure 1: A Simple Causal Model Relating Cigarette Smoking and Lung Cancer

The intuitive meaning of the model in Figure 1 is that there is a mechanism that relates cigarette smoking (node X) and lung cancer (node Y), in the sense that the probability of a person getting lung cancer is affected by that person smoking cigarettes. All that could be observed was a strong correlation between smoking and lung cancer, but the public health community of the 1950s suspected that the correlation was a manifestation of a causal link, and that therefore a manipulation of the smoking variable, namely by making smoking less pervasive, would lead to a reduction of the incidence of lung cancer in the population.
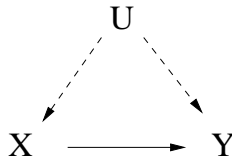
Figure 2: R.A. Fisher's Genotype Model Explaining the Correlation between Smoking and Lung Cancer

The model of Figure 1, however, was not universally accepted, and in a way that has a profound impact on the expected value of public health policy efforts directed at reducing smoking in the general population. The distinguished British statistician R.A. Fisher suggested that this model was a manifestation of "an error . . . of an old kind, in arguing from correlation to causation," and proposed that the correlation between smoking and lung cancer could be explained by a genetic predisposition to smoking and lung cancer, as described in Figure 10. Since the genotype (represented by node U in the figure) was not observable at the time of Fisher's work, it appeared impossible to conclude that reducing smoking would have a positive effect on the prevalence of lung cancer. (The links from U to X and Y are drawn dashed to emphasize that U is unobservable.) By accepting Fisher's model, we would have to conclude that the effect of smoking on lung cancer is *unidentifiable*, i.e. it cannot be determined from statistics derived solely from variables that are in the model and that can be observed (such as X and Y). Incidentally, studies involving

identical twins (who presumably would have the same genetic predisposition to both smoking and lung cancer) were carried out in attempt to identify the causal effect of smoking on lung cancer, but at least Fisher was unconvinced that they settled the matter [1].

Even in these very simple models, we can observe two important features of the causal graphs that we will formalize as causal Bayesian networks in the remainder of this paper. First, the models relate variables (genotype, smoking, lung cancer) represented as nodes in a directed acyclic graph through directed edges or links (such as the edge from smoking to lung cancer) indicating a causal influence. The links from U to X and Y are drawn dashed to emphasize that U is unobservable. Second, in the first model, the joint probability of the variables can be represented by the product of the prior probability of smoking times the conditional probability of lung cancer given smoking. In the second model, the joint probability of genotype, smoking, and lung cancer can be represented by the product of the prior probability of the genotype, times the conditional probability of smoking given the genotype, times the conditional probability of the genotype and smoking. This decomposition is allowed by the fundamental rule of probability in the case of a joint distribution of two and three variables.

In the next section we define formally the key notion of causal Bayesian network. In Section Three, we discuss the concept of intervention and define an identifiability problem. In Section Four, we present a proof of unidentifiability for a particular causal Bayesian network and use different versions of smoking-lung cancer causal models to show why the identifiability problem is interesting and how it can be solved mathematically. The conclusions, which consists mainly of a survey of the recent literature on the topic, are included in Section Five.

Disclaimers are in order. We do not claim that any of our specific examples are reflective of good domain knowledge. We are not interested in validating or repudiating causal assumptions specific to a domain. We assume that the domain knowledge is obtained somehow before our analysis. The framework described in this paper is limited to answering the question of whether a given set of assumptions is sufficient for quantifying causal effects from non-experimental data.

2   Causal Bayesian Networks

A *Bayesian network* (BN) is a graphical representation of the joint probability distribution of a set of discrete variables. The representation consists of a directed acyclic graph (DAG), prior probability tables for the nodes in the DAG that have no parents and conditional probabilities tables (CPTs) for the nodes in the DAG given their parents. As an example, consider the network in Figure 3.

More formally, a Bayesian network is a pair composed of: (1) a multivariate probability distribution over n random variables in the set $V = V_1, \ldots, V_n$,

A	A_1, A_2
B	B_1, B_2, B_3
C	C_1, C_2
D	D_1, D_2, D_3, D_4
E	E_1, E_2, E_3, E_4

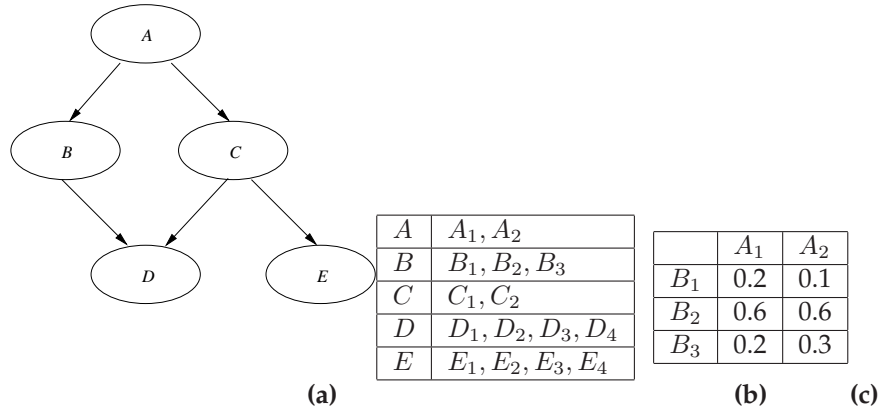|	A_1	A_2
B_1	0.2	0.1
B_2	0.6	0.6
B_3	0.2	0.3

(a) **(b)** **(c)**

Figure 3: (a) Example Bayesian Network, (b) Variable States, and (c) Conditional Probability Table for B Given A.

and (2) a directed acyclic graph (DAG) whose nodes are in one-to-one correspondence with $V_1, \ldots, V_n$. (Therefore, for the sake of convenience, we do not distinguish the nodes of a graph from variables of the distribution.)

Bayesian networks allow specification of the joint probability of a set of variables of interest in a way that emphasizes the qualitative aspects of the domain. The defining property of a Bayesian network is that the conditional probability of any node, given any subset of non-descendants, is equal to the conditional probability of that same node given the parents alone. The Chain rule for Bayesian networks [2] follows from the above definition: Let $P(V_i \mid \pi(V_i))$ be the conditional probability of V_i given its parents. (If there are no parents for V_i, let this be $P(V_i)$.) If all the probabilities involved are nonzero, then $P(V) = \prod_{v \in V} P(v \mid \pi(v))$.

Three features of Bayesian networks are worth mentioning. First, the directed graph constrains the possible joint probability distributions represented by a Bayesian network. For example, in any distribution consistent with the graph of Figure 3, D is conditionally independent of A given B and C. Also, E is conditionally independent of any subset of the other variables given C.

Second, the explicit representation of constraints about conditional independence allows a substantial reduction in the number of parameters to be estimated. In the example, assume that the possible values of the five variables are as shown in Figure 3(b). Then, the joint probability table $P(A, B, C, D, E)$ has $2 \times 3 \times 2 \times 4 \times 4 = 192$ entries. It would be very difficult to assess 191 independent parameters. However, the independence constraints encoded in the graph permit the factorization $P(A, B, C, D, E) = P(A) \times P(B|A) \times P(C|A) \times P(D|B, C) \times P(E|C)$ which reduces the number of parameters to be estimated to $1 + 4 + 2 + 18 + 6 = 31$. The second term in the sum is the table for the conditional probability of B given A. This probability is shown in Figure 3(c); note that there are only four independent parameters to be estimated since the

sum of values by column is one.

Thirdly, the Bayesian network representation allows a substantial (usually, dramatic) reduction in the time needed to compute marginals for each variable in the domain. The explicit representation of constraints on independence relations is exploited to avoid the computation of the full joint probability table in the computation of marginals both prior and conditioned on observations. Limitation of space prevents the description of the relevant algorithms; see [3] for a discussion of the justly famous junction tree algorithm.

The most common operation on a Bayesian network is the computation of marginal probabilities both unconditional and conditioned upon evidence. Marginal probabilities are also referred as *beliefs* in the literature [4]. This operation is called probability updating, belief updating, or belief assignment.

A link between two nodes in a Bayesian network is often interpreted as a causal link. However, this is not necessarily the case. When each link in a Bayesian network is causal, then the Bayesian network is called a *causal Bayesian network* or *Markovian model*. Markovian models are popular graphical models for encoding distributional and causal relationships. To summarize, a Markovian model consists of a DAG G over a set of variables $V = \{V_1, \ldots, V_n\}$, called a *causal graph* and a probability distribution over V, which has some constraints on it that will be specified precisely below. We use $V(G)$ to indicate that V is the variable set of graph G. If it is clear in the context, we also use V directly. The interpretation of such kind of model consists of two parts. The probability distribution must satisfy two constraints. The first one is that each variable in the graph is independent of all its non-descendants given its direct parents. The second one is that the directed edges in G represent causal influences between the corresponding variables. A Markovian model for which only the first constraint holds is called a *Bayesian network*, and its DAG is called a *Bayesian network structure*. This explains why Markovian models are also called causal Bayesian networks. As far as the second condition is concerned, some authors prefer to consider equation 3 (below) as definitional; others take equation 3 as following from more general considerations about causal links, and in particular the account of causality that requires that, when a variable is set, the parents of that variable be disconnected from it. A full discussion of this is beyond the scope of this paper, but see [5] and [6].

In this paper, capital letters, like V, are used for variable sets; lower-case letters, like v, stand for the instances of variable set V. Capital letters like X, Y and V_i are also used for single variables, and their values can be x, y and v_i. Normally, we use $F(V)$ to denote a function on variable set V. An instance of this function is denoted as $F(V)(V = v)$, or $F(V)(v)$, or just $F(v)$. Each variable is in one-to-one correspondence to one node in the causal graph.

We use $Pa(V_i)$ to denote parent node set of variable V_i in graph G and $pa(V_i)$ as an instance of $Pa(V_i)$. $Ch(V_i)$ is V_i's children node set; $ch(V_i)$ is an instance of $Ch(V_i)$.

Based on the probabilistic interpretation, we get that the joint probability

function $P(v) = P(v_1, \ldots, v_n)$ can be factorized as

$$P(v) = \prod_{V_i \in V} P(v_i | pa(V_i)) \tag{1}$$

From the joint probability, all marginal prior and posterior probabilities can be obtained, by marginalizing and conditioning. The notion of conditional probability is a well-defined and accepted one. The conditional probability of an event S given an event D is defined as

$$P(S \mid D) = \frac{P(S, D)}{P(D)}$$

This definition is actually an axiom of probability, which can be shown to hold in all useful interpretation of probability, including the subjective Bayesian one [2]. It is a tenet of applied Bayesian reasoning that beliefs are updated by conditioning when new knowledge is gained. There are, however, two kinds of updating in causal Bayesian networks, viz. updating by conditioning (also known as updating by observation) and updating by intervention.

Updating by conditioning is well defined and understood, both in principle and algorithmically [7] [8]. There are free and commercial software packages, like Hugin[1], that perform update by conditioning on Bayesian network in a very efficient manner in most practical cases. In the next section, we explain updating by intervention.

3   Interventions and the Identifiability Problem

The causal interpretation of Markovian model enables us to predict intervention effects. Here, intervention means some kind of modification of factors in product (1). The simplest kind of intervention is fixing a subset $T \subseteq V$ of variables to some constants t, denoted by $do(T = t)$ or just $do(t)$. Then, the post-intervention distribution

$$P_T(V)(T = t, V = v) = P_t(v) \tag{2}$$

is given by:

$$P_t(v) = P(v|do(t)) = \begin{cases} \prod_{V_i \in V \backslash T} P(v_i | pa(V_i)) & v \text{ consistent with } t \\ 0 & v \text{ inconsistent with } t \end{cases} \tag{3}$$

To stress the distinction between observation and intervention, we present a simple example based on the sneezing model of Figure 4, originally presented in [9], in which wiping one's nose (W) is caused by sneezing (S), which in turn is caused by either a cold (C) or hay fever (F), or both. To complete this causal Bayesian network, we give the probabilities in Table 1 and $P(Cold) = (.2, .8)$, $P(HayFever) = (.1, .9)$, and $P(WipingOne'sNose \mid Sneezing = y) = .9$.
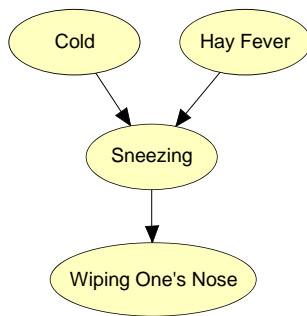
Figure 4: A Causal Graph for Sneezing

	H	
C	y	n
y	.9	.8
n	.7	.1

Table 1: Table for $P(Sneezing = y \mid Cold, HayFever)$
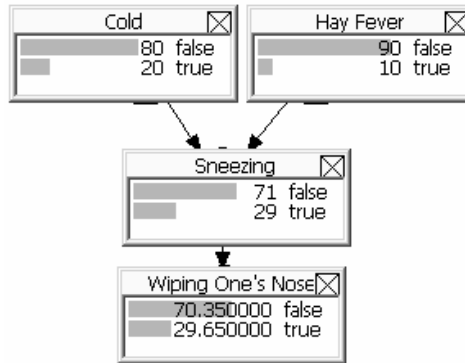
Figure 5: Initial Marginal Probabilities for the Sneezing Model

We use this causal graph to compare the notions of observation and intervention. The initial probabilities for the four variables are shown in Figure 5. Suppose that it is observed that sneezing occurs. The probabilities of each variable in the network are updated as shown in Figure 6.

Suppose now that we intervene and force sneezing. The connections between node Sneezing and its parents are cut, as indicated in the model of Figure 7. Unlike the situation in which sneezing is observed, the posterior probabilities of Cold and Hay Fever are unchanged.

We call interventions of the simple kind described so far in this section, which consist in fixing a subset of variables to some constants, *crisp* interventions. Referring to the sneezing causal graph, a simple example is setting the variable Sneezing to the value *true* (i.e., forcing sneezing to occur), by the administration of a perfectly effective sneezing powder. More complicated interventions can be described using the *intervention graph* or *augmented model* [10], in the way originally described in [11] and [12]. The intervention graph is formed by adding a parent to each node representing a variable where intervention is contemplated. The brief discussion here follows the excellent presentation of [6, Section 3.3.2] very closely. The interested reader should consult that reference for more detail.

The effect of a crisp intervention $do(X_i = x_i)$ can be encoded by adding to G a link $F_i \rightarrow X_i$, 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. The new parent set of X_i in the augmented network is $Pa'(X_i) = Pa(X_i) \cup \{F_i\}$, and it is related to X_i by the conditional probability.
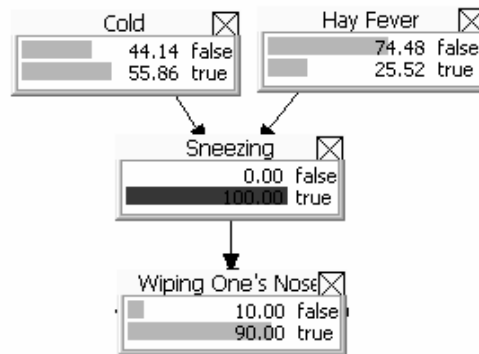
---

[1]http://www.hugin.com/

Figure 6: Marginal Probabilities for the Sneezing Causal Bayesian Network, after Updating by Conditioning on the Evidence of Sneezing
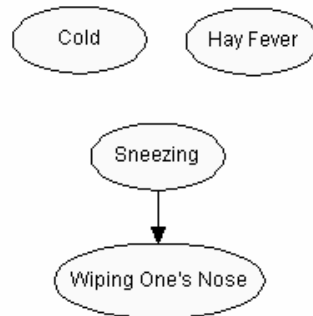

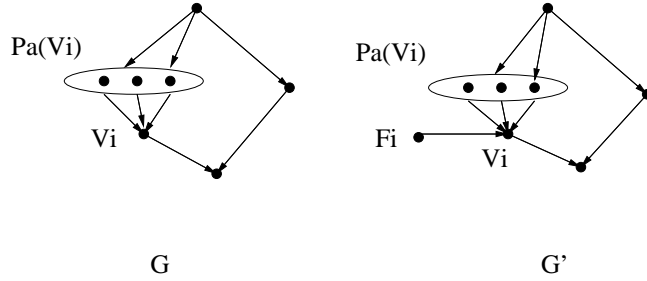Figure 7: The Sneezing Causal Graph, after a Crisp Intervention that Forces Sneezing to Occur

Figure 8: Intervention Graph

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

The effect of the intervention $do(x_i')$ is to transform the original probability function $P(X_1, \ldots, X_n)$ into a new probability function $P_{X_i=x_i'}(X_i, \ldots, X_n)$, given by

$$P_{X_i=x_i'}(X_1, \ldots, X_n) = P'(X_1, \ldots, X_n|F_i = do(x_i')) \tag{5}$$

where P' is the distribution specified by the augmented network $G' = G \cup \{F_i \rightarrow X_i\}$ and (4), with an arbitrary prior distribution on F_i. In general, by adding a hypothetical intervention link $F_i \rightarrow X_i$ to multiple nodes in G, we can construct an augmented probability function $P'(x_1, \ldots, x_n; F_1, \ldots, F_n)$, which allows for interventions beyond the setting of a subset of variable to constants. For a simple example related to the causal graph of Figure 4, imagine administering an imperfectly effective sneezing power, which causes sneezing with probability p. This would be modeled by setting the prior probability of the value *true* for the forcing variable $F_{Sneezing}$ in the model of Figure 9 to p.

In many cases, an empirical scientist faces the following question: Can we estimate the post-intervention distribution under crisp intervention from non-experimental data? When all the variables in the model are observable, the answer for the question above is positive. But when some variables in V are unobservable, things are much more complex, and the answer may be either positive or negative, depending on the structure of the causal Bayesian network and the relative location of observable and unobservable variables. We give a detailed proof of unindentifiability for the key example of Figure 10 in the next section.
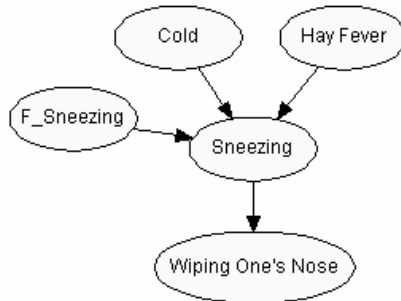
Figure 9: An Intervention Graph for the Sneezing Model of Figure 4

4 An Unidentifiable Model

A simple example of unidentifiable model is R.A. Fisher's genotype model of the relation between smoking and lung cancer [12], which we briefly discussed in the first section of this paper. R.A. Fisher suggested that the observed correlation between smoking(X) and lung cancer(Y) can be explained by some sort of carcinogenic genotype(U) that involves inborn craving for nicotine.
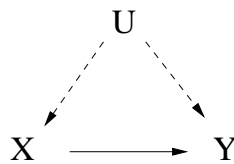

Figure 10: R.A. Fisher's Genotype Model Explaining the Correlation between Smoking and Lung Cancer (Figure 2, repeated here for the reader's convenience)

The carcinogenic genotype is presented by Fisher as a concept that has not been observed in nature, and it is therefore modeled as an unobservable variable. Intuitively, the effect of smoking on lung cancer is unidentifiable because we are not sure whether the observed response (lung cancer) is due to our action (smoking) or to the confounder event (genotype) that triggers the action and simultaneously causes the response.

Formally, we show that the effect of smoking on lung cancer is unidentifiable in the following way. We show that, with a given observational distri-

bution $P(x, y)$, it is possible to find two different causal Bayesian networks M_1 and M_2 that share the graph of Figure 10, and such that $P^{M_1}(x, y) = P^{M_2}(x, y)$. As an extreme example, we can construct two drastically different models following the explanation that smoking is the cause of lung cancer (the part a in Figure 11) or carcinogenic genotype is the only cause of this fatal disease (the part b in Figure 11). Both of them satisfy the graph in Figure 10. That is, Figure 10 generalizes both graphs in Figure 11, or, equivalently, the graphs in Figure 11 are special cases of the graph in Figure 10.
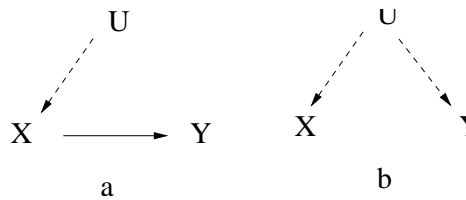

Figure 11: Different Models Compatible with the Model of Figure 10

Our question about the smoking and lung cancer model of Figure 10 is: If we intervene on variable X, which means we control the smoking behavior, is unexperimental observational knowledge about smoking and lung-cancer (i.e., $P(x, y)$) sufficient to determine the probability of lung cancer? Mathematically, this problem can be explained as a question on the Markovian model of Figure 10: If we know $P(x, y)$, which is the joint probability on observable variables X, Y, can we calculate $P_x(y)$ for all (x, y)? Unfortunately, the answer to this question is negative.

The fact is that it is possible to create a model compatible with Figure 11 part a and another model compatible with Figure 11 part b, both of them satisfying $P(x, y)$, but with different $P_x(y)$. We cannot get the same $P_x(y)$ from these two models because for the first model, the intervention will change the probability of lung cancer but if the second model is correct, the behavior of smoking has no effect on lung cancer at all.

We now carry out the calculations in detail to show that the effect of smoking on lung cancer is unidentifiable for the causal Bayesian network of Figure 10. All variables are binary, and their states are denoted as 0 and 1. For variable U, we assume $P(U = 0) = P(U = 1) = 1/2$. The conditional probability tables of variable X and Y in model M_1 are defined as below:

$$
\begin{array}{cc|c}
x & u & P^{M_1}(x|u) \\
\hline
0 & 0 & 0.6 \\
0 & 1 & 0.4
\end{array}
\tag{6}
$$

$$
\begin{array}{ccc|c}
y & x & u & P^{M_1}(y|x,u) \\
\hline
0 & 0 & 0 & 0.7 \\
0 & 0 & 1 & 0.2 \\
0 & 1 & 0 & 0.2 \\
0 & 1 & 1 & 0.7
\end{array}
\tag{7}
$$

The conditional probability tables of variable X and Y in model M_2 are defined as:

$$
\begin{array}{cc|c}
x & u & P^{M_2}(x|u) \\
\hline
0 & 0 & 0.7 \\
0 & 1 & 0.3
\end{array}
\tag{8}
$$

$$
\begin{array}{ccc|c}
y & x & u & P^{M_2}(y|x,u) \\
\hline
0 & 0 & 0 & 0.65 \\
0 & 0 & 1 & 0.15 \\
0 & 1 & 0 & 0.15 \\
0 & 1 & 1 & 0.65
\end{array}
\tag{9}
$$

Note that, since $P(X,Y) = \sum_U P(Y|x,U)P(X|U)P(U)$, for both models M_1 and M_2, we obtain:

$$
\begin{array}{cc|c}
y & x & P^{M_i}(y,x) \\
\hline
0 & 0 & 0.25 \\
0 & 0 & 0.25 \\
0 & 1 & 0.25 \\
0 & 1 & 0.25
\end{array}
\tag{10}
$$

We also have $P_X(Y) = \sum_U P(Y|X,U)P(U)$ for both two models, and for M_1, $P_{X=0}^{M_1}(Y=0) = 0.45$, but for M_2, we have $P_{X=0}^{M_2}(Y=0) = 0.40$.

We conclude that $P_X(Y)$ is unidentifiable in this causal graph.

We emphasize, again, that the models we use are not intended to be correct representation of domain knowledge: we use them simply to illustrate how they can be used in representing causal modeling assumptions, which have different bearing on the identifiability of causal effects. In that spirit, we conclude the section by describing two more models that may match some researchers' understanding of the relation between smoking and lung cancer.
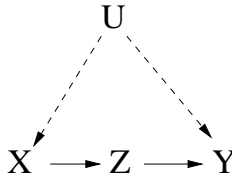

Figure 12: Smoking (X) Affects Lung Cancer (Y) through Tar Deposits (X)

It is observed that smoking causes tar deposit on lung. So, with the evidence that genotype may cause lung cancer and smoking behavior, a researcher may establish a causal model about smoking (X) and lung cancer (Y) with tar (Z). See Figure 12. Incidentally, R.A. Fisher himself seems to hint at such a model [1]. This causal graph assumes that smoking cigarettes has no effect on the production of lung cancer except as mediated through tar deposits and genotype has no effect on the amount of tar in the lungs except indirectly through cigarette smoking.

Finally, the causal graph of Figure 13 adds to the causal graph of Figure 12 the assumption that the production of tar deposits in the lung (Z) is affected by pollution (U_1), which also affects the propensity towards smoking (X).
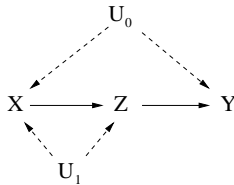

Figure 13: Air Pollution (U_1) Affects Tar Deposits (Z) and the Propensity to Smoke (X)

The causal effect of smoking on lung cancer ($P_X(Y)$) is identifiable in the causal graph of Figure 12, but unidentifiable in the causal graph of Figure 13. We do not attempt to prove this claim in this paper, and refer the reader to the references given in the next section for the algorithms needed to establish this claim and, in the case of Figure 12, compute the value of the causal effect.

5 Conclusion

This paper provides an introduction to the problem of inferring the strength of cause-and-effect relationships from a causal Bayesian network, emphasizing conceptual foundations and examples. In this concluding section, we provide a roadmap to the literature where proofs and algorithms are provided.

A causal Bayesian network consists of a causal graph, an acyclic directed graph expressing causal relationships, and a probability distribution respecting the independence relation encoded by the graph. Because of the existence of unmeasured variables, the following identifiability questions arise: "Can we assess the strength of causal effects from nonexperimental data and casual relationships? And if we can, what is the total causal effect in terms of estimable quantities?"

The questions just given could partially be answered using graphical approaches due to Pearl and his collaborators. More precisely, graphical conditions have been devised to show whether a causal effect, that is, the joint

response of any set S of variables to interventions on a set T of action variables, denoted $P_T(S)$ [2] is identifiable or not. Those results are summarized in [6]. For example, "back-door" and "front-door" criteria and do-calculus [15]; graphical criteria to identify $P_T(S)$ when T is a singleton [16]; graphical conditions under which it is possible to identify $P_T(S)$ where T and S are, possibly non-singleton, sets, subject to a special condition called Q-identifiability [17]. Further study can be also found in [18] and [19].

More recently, Tian and Pearl published a series of papers related to this topic [20, 13, 14]. Their new methods combined the graphical character of causal graph and the algebraic definition of causal effect. They used both algebraic and graphical methods to identify causal effects. The basic idea is, first, to transfer causal graphs to semi Markovian graphs [13], then to use Algorithm 2 in [14] (the *Identify* algorithm) to calculate the causal effects we want to know.

Tian and Pearl's method was a great contribution to this study area. But there were still some problems left. First, even though we believe, as Tian and Pearl do, that the semi Markovian models obtained from the transforming Projection algorithm in [13] are equal to the original causal graphs, and therefore the causal effects should be the same in both models, still, to the best of our knowledge, there was no formal proof for this equivalence. Second, the completeness question of the Identify algorithm in [14] was still open, so that it was unknown whether a causal effect was identifiable if that Identify algorithm failed.

In a series of papers, Huang and Valtorta [21, 22, 23] and, independently, Shpitser and Pearl [24, 25] solved the open questions and several related ones. In particular, following Tian and Pearl's work, Huang and Valtorta [21] solved the second question. They showed that the Identify algorithm Tian and Pearl used on semi-Markovian models is sound and complete. In [23], they followed the ideas Tian and Pearl presented in [14], but instead of working on semi-Markovian models, they focused on general causal graphs directly, and their proofs showed, that Algorithm 2 in [14] can also be used in general causal models, and that the algorithm is sound and complete, which means a causal effect is identifiable if and only if the given algorithm runs successfully and returns an expression that is the target causal effect in terms of observable quantities.

It is our hope that the reader will be motivated to study, implement, refine, and apply the algorithmic framework to causal modeling that Pearl pioneered and that, the authors believe, is ready to be put to the test of deployment in actual applications.

---

[2]Pearl and Tian used notation $P(s|do(t))$ and $P(s|\hat{t})$ in [6] and $P_t(s)$ in [13], [14].

References

[1] R. Fisher, "Cancer and smoking," *Nature*, vol. 182, pp. 596–597, August 1958.

[2] R. E. Neapolitan, *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. New York, NY: John Wiley and Sons, 1990.

[3] F. Jensen, *Bayesian Networks and Decision Graphs*. Springer Verlag, 2001.

[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan-Kaufmann Publishers, 1988.

[5] S. Lauritzen, "Causal inference from graphical models," in *Complex Stochastic Systems* (O. Barndorff-Nielsen and C. Klueppelberg, eds.), pp. 63–107, London/Baton Rouge: Chapman and Hall/CRC, 2001.

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

[7] G. Cooper, "The Computational Complexity of Probabilistic Inference Using Bayesian Networks," *Artificial Intelligence*, vol. 42, no. 2-3, pp. 393–405, 1990.

[8] S. L. Lauritzen, *Graphical Models*. Oxford, United Kingdom: Clarendon Press, 1996.

[9] J. Pearl and T. Verma, "A theory of inferred causation," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, (San Mateo, CA), pp. 441–452, Morgan Kaufman, April 1991.

[10] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*. Boca Raton, FL USA: Chapman Hall/CRC Press, 2003.

[11] J. Pearl, "Graphical models, causality, and intervention." comments on: 'Linear Dependencies Represented by Chain Graphs' by D. Cox and N. Wermuth, and 'Bayesian Analysis in Expert Systems' by D.J. Spiegelhalter, A.P. Dawid, S.L. Lauritzen, and R.G. Cowell, 1993. In Statistical Science, Vol. 8, 266-269.

[12] P. Spirtes, G. Glymour, and R. Scheines, *Causation, Prediction and Search*. New York, USA: Springer-Verlag, 1993.

[13] J. Tian and J. Pearl, "On the testable implications of causal models with hidden variables," in *Proceedings of the Eighteenth Annual Conference on Uncertainty in Artificial Intelligence(UAI-02)*, pp. 519–527, 2002.

[14] J. Tian and J. Pearl, "On the identification of causal effects, Technical report 290-L," tech. rep., Cognitive Systems Laboratory, University of California at Los Angeles, 2003. extended version available at http://www.cs.iastate.edu/ jtian/r290-L.pdf.

[15] J. Pearl, "Causal diagrams for empirical research," *Biometrika*, vol. 82, pp. 669–710, 12 1995.

[16] D. Galles and J. Pearl, "Testing identifiability of causal effects," in *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence(UAI-95)*, pp. 185–195, 1995.

[17] J. Pearl and J.M.Robins, "Probabilistic evaluation of sequential plans from causal models with hidden variables," in *Proceedings of the Eleveth Annual Conference on Uncertainty in Artificial Intelligence(UAI-95)*, pp. 444–453, 1995.

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

[19] J. Robins, "Causal inference from complex longitudinal data," in *Latent Variable Modeling with Applications to Causality, Volume 120 of Lecture Notes in Statistics* (M. Berkane, ed.), pp. 69–117, New York: SpringerVerlag, 1997.

[20] J. Tian and J. Pearl, "A general identification condition for causal effects," in *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, pp. 567–573, 2002.

[21] Y. Huang and M. Valtorta, "On the completeness of an identifiability algorithm for semi-Markovian models," tech. rep., University of South Carolina Department of Computer Science, 1 2006. Preprint available at http://www.cse.sc.edu/ mgv/reports/tr2006-001.pdf; journal version to appear.

[22] Y. Huang and M. Valtorta, "Pearl's calculus of intervention is complete," in *Proceedings of the Twenty-second Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 217–224, 2006.

[23] Y. Huang and M. Valtorta, "Identifiability on causal bayesian networks: A sound and complete algorithm," in *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, pp. 1149–1154, 2006.

[24] I. Shpitser and J. Pearl, "Identification of conditional interventional distributions," in *Proceedings of the Twenty-second Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 437–444, 2006.

[25] I. Shpitser and J. Pearl, "Identification of conditional interventional distributions," in *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, pp. 1217–1226, 2006.