An Introduction to
Causal Inference

Richard Scheines

In *Causation, Prediction, and Search* (*CPS* hereafter), Peter Spirtes, Clark Glymour and I developed a theory of statistical causal inference. In his presentation at the Notre Dame conference (and in his paper, this volume), Glymour discussed the assumptions on which this theory is built, traced some of the mathematical consequences of the assumptions, and pointed to situations in which the assumptions might fail. Nevertheless, many at Notre Dame found the theory difficult to understand and/or assess. As a result I was asked to write this paper to provide a more intuitive introduction to the theory. In what follows I shun almost all formality and avoid the numerous and complicated qualifiers that typically accompany definitions or important philosophical concepts. They can be all be found in Glymour's paper or in *CPS*, which are clear although sometimes dense. Here I attempt to fix intuitions by highlighting a few of the essential ideas and by providing extremely simple examples throughout.

The route I take is a response to the core concern of many I talked to at the Notre Dame conference. Our techniques take statistical data and output sets of directed graphs. Most everyone saw how that worked-but they could not easily assess the additional assumptions necessary to give such output a causal interpretation, that is an interpretation that would inform us about how systems would *respond to interventions*. I will try to present in the simplest terms the assumptions that allow us to move from probabilistic independence relations to the kind of causal relations that involve counterfactuals about manipulations and interventions. I first separate out the various parts of the theory: directed graphs, probability, and causality, and then clarify the assumptions that connect causal structure to probability. Finally, I discuss the additional assumptions needed to make inferences from statistical data to causal structure.

1. DAGs and d-separation

The theory we developed unites two pieces of mathematics and one piece of philosophy. The mathematical pieces are directed acyclic graphs (DAGs) and probability theory (with the focus on conditional independence), and the philosophy involves causation among variables.

A DAG is a set of vertices and a set of edges (arrows) that connect pairs of these vertices. For example, we might have a set of three vertices: $\{X_1, X_2, X_3\}$, and a set of two edges among these vertices: $\{X_1 \rightarrow X_2 , X_2 \rightarrow X_3\}$. We almost always represent DAGs with a picture, or path diagram, e.g., this DAG looks like: $X_1 \rightarrow X_2 \rightarrow X_3$

Prior to any interpretation, a DAG is a completely abstract mathematical object. In our theory, DAGs are given two distinct functions. In the first they represent sets of probability distributions and in the second they represent causal structures. The way they represent probability distributions is given by the Markov condition, which (in DAGs) turns out to be equivalent to a more generally useful graphical relation: d-separation (Pearl 1988).[1] D-separation is a relation between three disjoint sets of vertices in a directed graph. Although too complicated to explain or define here,[2] the basic idea involves checking whether a set of vertices Z blocks all connections of a certain type between X and Y in a graph G. If so, then X and Y are d-separated by Z in G. In the DAG on the left side of Fig. 1, for example, X_2 blocks the only directed path connecting X_1 and X_3, so X_1 and X_3 are d-separated by X_2 in this DAG. By choosing d-separation to connect DAGs to probability distributions, we assume that in all of the distributions P a DAG G can represent, if sets of vertices X and Y are d-separated by a set Z in the DAG G, then X and Y are independent conditional on Z in P. For example, applying d-separation to the DAG in Fig. 1 gives us: X_1 and X_3 are d-separated by X_2. We then assume that in all distributions this DAG can represent, X_1 is independent of X_3 conditional on X_2. We use a notation for independence introduced by Phil Dawid (1979); $X_1 \perp\!\!\!\perp X_3 \mid X_2$ means: X_1 and X_3 are independent conditional on X_2.

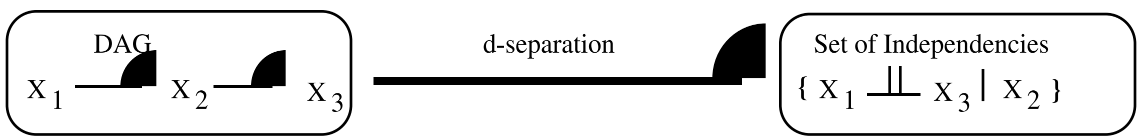

Fig. 1

It should be stressed that as long as we remain agnostic and give no interpretation to DAGs, then they are just mathematical objects which we can connect to probability distributions in any way we like. We could just as easily define and then use e-separation, or f-separation, or any graphical relation we please, as long as it produced consistent sets of independencies. When we give DAGs a causal interpretation, it then becomes necessary to argue that d-separation is the correct connection between a causal DAG and probability distributions. Let us put off that task for a few more pages, however.

There are often many distinct DAGs that represent exactly the same set of independence relations, and thus the same set of distributions. And just as one might want a procedure that computes d-separation for any graph, one might want an algorithm that computes all the DAGs that represent a given set of independence relations (Fig. 2).

[1] If directed graphs have cycles, or chains of arrows that lead from a variable back to itself, then this equivalence breaks down.

[2] We try to explain it in CPS, pp. 71-74.

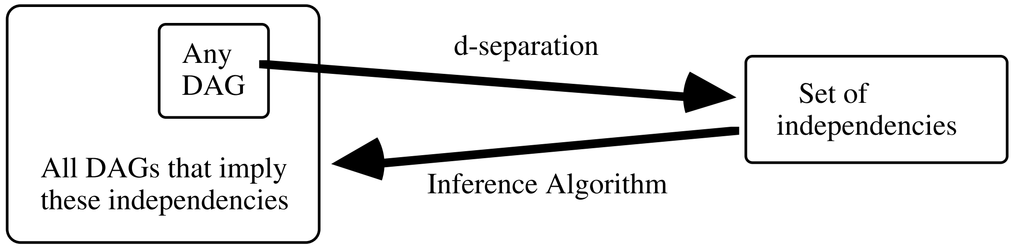

Fig. 2

We have developed several such algorithms, one of which is called the PC algorithm and is computed by the TETRAD II program.[3] Its input is a set of independence relations over a set of variables and its output is a set of DAGs over these variables that are d-separation, or Markov, equivalent.[4] Applying the PC algorithm to the same set of independence relations shown on the right side of Fig. 1, you can see (in Fig. 3) that there are two other DAGs that are d-separation equivalent to the DAG in Fig. 1. PC is known to be complete in the sense that its output contains all and only those DAGs that are d-separation equivalent.

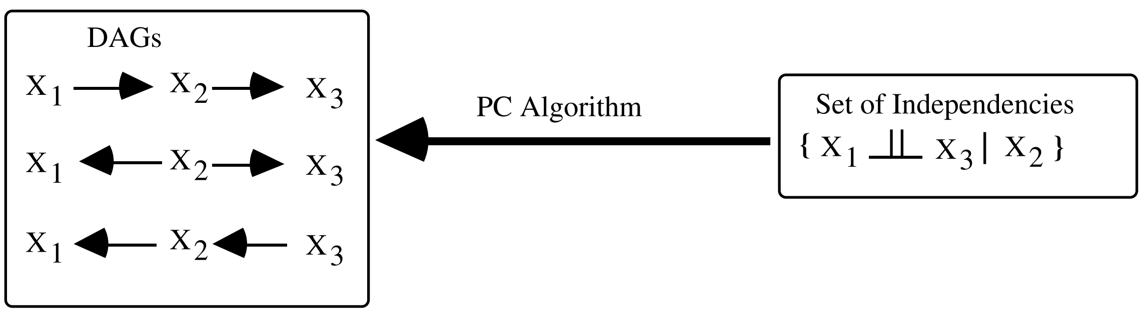

Fig. 3

2. Causal Graphs

If taken no further, d-separation and the Markov condition are just mathematics connecting DAGs and probability distributions and need not involve causation at all.[5] One might be content to use this mathematical theory solely to produce compact and elegant representations of independence structures,[6] or one might take a further step by assuming that when *DAGs are*

---

[3]For those interested in getting acquainted with the ideas as they are embodied in the program that computes many of the discovery algorithms presented in CPS, the TETRAD II program is available from Lawrence Erlbaum Associates, Hillsdale, NJ.

[4] Sometimes there are no DAGs that can represent a given set of independence relations.

[5] In fact, Judea Pearl originally developed the theory connecting DAGs and probability in order to afford robots or other AI agents an efficient way to store and use probability distributions which represented the agent's uncertainty over states of the world.

[6] In fact we have often heard just such a purpose endorsed explicitly in public by able statisticians, but in almost every case these same people over beer later confess their heresy by concurring that their real ambitions are

interpreted causally the Markov condition and d-separation are in fact the *correct* connection between causal structure and probabilistic independence. We call the latter assumption the Causal Markov condition, and it is a stronger assumption than the Markov condition.

DAGs that are interpreted causally are called *causal graphs*. There is an arrow from X to Y in a causal graph involving a set of variables **V** just in case X is a direct cause of Y relative to **V**. For example, if S is a variable that codes for smoking behavior, Y a variable that codes for yellowed, or nicotaine stained, fingers, and C a variable that codes for the presence of lung cancer, then the following causal graph (Fig. 4) represents what I believe to be the causal structure among these variables.

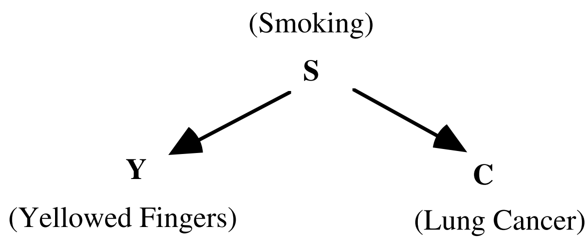

Fig. 4

Causal graphs are assumed to be complete in one sense and not in another. They are incomplete in that they do not necessarily include *all* of the causes of each variable in the system. Thus many of the causes of lung cancer have been left out, e.g., asbestos inhalation, genetic factors, etc. They also leave out many variables that might lie in between a specified cause and its effect, e.g., cillia trauma in the bronchial lining might lie on the "true" causal pathway from smoking to lung cancer. But a causal graph is assumed to be complete in the sense that all of the *common causes* of specified variables have been included. For example, if there is some variable that is a cause of both smoking behavior and lung cancer, e.g., a genetic factor, then the causal graph above is not an accurate depiction of the causal structure among these three variables. The causal graph is also assumed to be complete in the sense that *all* of the causal relations among the specified variables are included in the graph. For example, the graph in Fig. 4 has no edge from Y to S, so it is only accurate if the level of nicotine stains does not in any way cause smoking behavior.

The semantics of a causal graph involve ideal manipulations and the changes in the probability distribution that follow such manipulations. Such an account is circular, because to manipulate *is* to cause. Our purpose, however, is not to provide a reductive definition of causation, but rather to connect it to probability in a way that accords with scientific practice and allows a systematic investigation of causal inference.

---

causal and their public agnosticism is a prophylactic against the abuse of statistics by their clients or less careful practitioners.

To manipulate a variable ideally is to change it in a way that, at least for the moment, leaves every other variable undisturbed. Such a manipulation must directly change only its target and leave changes in the other variables to be produced by these targets or not at all. For example, suppose we are attempting to experimentally test hypotheses concerning the causal relations between athletic performance and confidence. Suppose we intervene to inhibit athletic performance by administering a drug that blocks nutrient uptake in muscle cells, but that this drug also imitates the chemical structure of neurotransmitters that inhibit feelings of insecurity and anxiety, thus serving to directly increase anxiety and lower confidence. This intervention provides little help in trying to reason about the sort of causal relation that exists between athletic performance and confidence, because it directly alters both variables. It is an intervention with a "fat hand."[7] Ideal interventions are perfectly selective in the variables they *directly* change.

The causal graph tells us, for any ideal manipulation we might consider, which other variables we would expect to change in some way and which we would not. Put simply, the only variables we can hope to change must be causally "downstream" of the variables we manipulated. Although we can make inferences upstream, that is from effects to their causes, we cannot manipulate an effect and hope to change its other causes. In Fig. 4, for example, after an ideal manipulation of the level of nicotine stains, nothing at all would happen to the probabilities of smoking and lung cancer. They would take on the same values they would have if we had done no manipulation at all. If we could manipulate the lung cancer level of an individual without directly perturbing his or her smoking behavior or finger stains, then again, we would not expect to change the probability of smoking or of finger stains. If, however, we could manipulate smoking behavior in a way that did not directly perturb any other variable, then (for at least some of these manipulations) we would perturb the probability of the other variables through the direct causal route from smoking to the other variables.

If the causal graph changes, so does the set of counterfactuals about ideal manipulations. If, for example, the causal graph is as I picture it in Fig. 5 (absurd as it may seem), then only the statement concerning manipulations of lung cancer remains unchanged. Any ideal manipulation of smoking will result in no change in Y, but some will result in a change in C's probability, and (some) manipulations of Y will result in changes in S's probability.

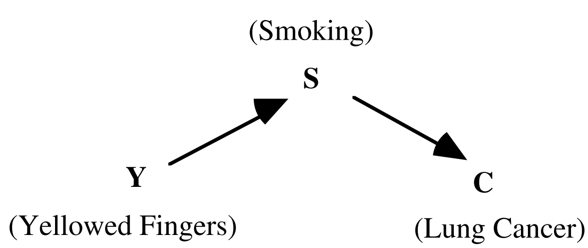

Fig. 5

[7] Kevin Kelly suggested this nomenclature.

Ignoring all sorts of subtleties, the point should be clear: the sort of causation we are after involves the response of a system to interventions.

3. The Causal Markov Condition

The Causal Markov assumption can be stated simply:

> A variable X is independent of every other variable (except X's effects) conditional on all of its direct causes.

Applying this to each variable in the causal graph in Fig. 4 yields the following independence relations:[8]

For Y: Y is independent of C conditional on S
For S: All of the other variables are S's effects, so the condition is vacuous
For C: C is independent of Y conditional on S

By probability theory, the first and last of these independences are equivalent, so this causal graph entails one independence by the Causal Markov assumption. You can see that Fig. 5 implies the same independence relations as does Fig. 4, even though it is different causally and thus entails different counterfactuals about interventions.

The independence relations entailed by applying the Causal Markov assumption to a causal graph is the same as those obtained from applying d-separation to a causal graph, but it is simpler to justify the connection between causal graphs and probability when stated in a Markov form. The intuition behind the Causal Markov assumption is simple: ignoring a variable's effects, all the relevant probabilistic information about a variable that can be obtained from a system is contained in its direct causes. In a Markov process, knowing a system's current state is relevant to its future, but knowing how it got to its current state is completely irrelevant. Hans Reichenbach (1956) was the first philosopher to explicitly discuss the Markov properties of causal systems, but variants have been discussed by Nancy Cartwright (1989), Wesley Salmon (1984), Brian Skyrms (1970), Patrick Suppes (1970), and many other philosophers.

How does such an assumption capture the asymmetry of causation? For systems of two variables it cannot. The two causal graphs in Fig. 6 imply the same independencies by the Causal Markov condition, and are thus indistinguishable solely on the basis of probabilistic independence.

---

[8] A variable X is always independent of Y conditional on Y, so in this list I do not include the trivial independences between each variable and its direct causes when we condition on these direct causes.

$$\mathbf{X} \longrightarrow \mathbf{Y} \qquad\qquad \mathbf{X} \longleftarrow \mathbf{Y}$$

Fig. 6

But leaping from this simple indistinguishability to the conclusion that probabilities can never give us information about causal structure is patently fallacious. As Hausman (1984) and Papineau (1985) realized, adding a third variable changes the story entirely. The two graphs in Fig. 7 are *not* Markov or d-separation equivalent, and the difference between their independence implications underlies the connection between independence and causal priority more generally.

$$X \longrightarrow Y \longleftarrow Z \qquad\qquad X \longleftarrow Y \longleftarrow Z$$

Independence Relations Entailed by d-separation

$$X \perp\!\!\!\perp Z \qquad\qquad X \perp\!\!\!\perp Z \mid Y$$

Fig. 7

We see three main lines of justification for the Causal Markov assumption, although surely there are others. First, versions of the assumption are used, perhaps implicitly, in making causal inferences from controlled experiments. Second, philosophical treatments of probabilistic causality embrace it (Suppes, 1970; Reichenbach, 1956), and third, structural equation models (Bollen, 1989), which are perhaps the most widely used class of statistical causal models in social science are Causally Markov. Elaborating on the first two lines of support are beyond the scope of this paper; they are covered in *CPS* or in Glymour's paper. Here I will try to make the connection between structural equation models and the Causal Markov assumption a little more explicit.

In a structural equation model, each variable is equal to a linear function of its direct causes plus an "error" term. Thus the causal graph in Fig. 4 would translate into the following structural equation model:

$$Y = \beta_1 S + \varepsilon_y$$
$$C = \beta_2 S + \varepsilon_c$$

where β_1 and β_2 are real valued coefficients and ε_c and ε_y are error terms with strictly positive variance. If the system in Fig. 4 is complete as specified, that is, its causal graph is complete with respect to common causes, then a structural equation modeller would assume that ε_c and ε_y are independent of each other and of S. Indeed, in structural equation models in which all of the common causes are included the error terms are assumed to be independent. It turns out that such models *necessarily satisfy* the Causal Markov assumption (Kiiveri and Speed, 1982). Spirtes (1994) has generalized the result to models in which each effect is an *arbitrary* function of its immediate causes and an independent error.[9] The nature of the function connecting cause and effect is not so important as the independence of the error terms.

The connection between structural equation models and causation (as it involves the response of a system to interventions) arises through the connection between independent error terms and ideal manipulations. Although ideal manipulations provide the semantic ground for causal claims, such manipulations are sometimes only ideal and cannot be practically realized. For example, although poverty may cause crime, we cannot ethically intervene to impoverish people. In such situations we resort to collecting data passively. Since experimental science specializes in creating arrangements in which ideal manipulations exist and are subject to our will, it is no surprise that critics of causal inference from statistical data insist that experiments are the only means of establishing causal claims. But besides: “I can’t imagine how it can be done,” what is their argument?

In the first place, there might well be ideally selective sources of variation that exist in nature but which we cannot now or ever hope to control. For example, the moon’s position exerts a direct effect on the gravitational field over the oceans, which causes the tides. But though the moon is a source that we cannot control, at least we can measure it.

In other systems, such ideal sources of variation might exist, but be both beyond our control and unobservable. In fact a natural interpretation of the error terms in structural equation models gives them precisely these properties. There is a unique error term ε_x for each specified variable X, and in systems which include all the common causes ε_x is assumed to be a source of X’s variation that *directly* affects only X. And although we cannot measure them or control them, structural equation modellers assume that such error terms exist. It is this assumption that makes such systems Causally Markov.[10]

It is tempting to think that even if we do interpret error terms as unobservable but ideal manipulations, then we are then stopped dead for purposes of causal inference just because we cannot observe them. But this is a fallacy. It is true that we cannot learn as much about the causal structure of such systems as we could if the error terms were observable (the experimenters world), but by no means does it follow that we can learn nothing about them. In fact this is precisely where causal inference starts, with systems that are assumed to be Causally Markov.

---

[9] In fact it is only necessary for the proof that the error term is a function of the variable for which it is an error and that variable's immediate causes_.

[10] In certain contexts the detrimental effect on causal inference of violating this assumption is well undersood. For example, in a regression model in which some of the regressors are correlated with the error term, then the result is a bias in estimating the causal effect of these regressors.

4. Inference

Accepting the Causal Markov assumption, I now turn to the subject of inference: moving from statistical data to conclusions about causal structure. Beginning with statistical data and background knowledge, we want to find all the possible causal structures that might have generated these data. The fewer assumptions we make constraining the class of possible causal structures, the weaker our inferential purchase. I should note, however, that it is not the job of our theory to dictate which assumptions a practicing investigator should endorse, but only to characterize what can and cannot be learned about the world given the particular assumptions chosen.

In this section I discuss a few of the assumptions that we have studied. There are many others that we are now studying or that would be interesting to study. An enormous class of problems I will not deal with at all involves statistical inference about independence: inferring the set of independence relations in a population from a sample. In what follows I assume that the data are statistically ideal and that in effect the population lies before us, so that any probabilistic independence claim can be decided with perfect reliability.

Faithfulness

The first assumption I will discuss is *Faithfulness*. By assuming that a causal graph is Causally Markov, we assume that any population produced by this causal graph has the independence relations obtained by applying d-separation to it. It does not follow, however, that the population has exactly these and no additional independencies. For example, suppose Fig. 8 is a causal graph that truly describes the relations among exercise, smoking, and health, where the + and - signs indicate positive and inhibitory relations respectively.[11]

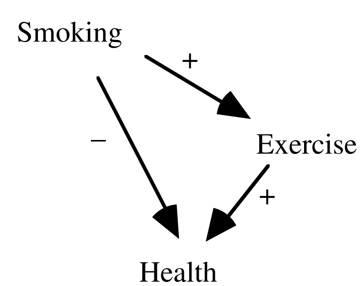

Fig. 8

[11] This example is from Cartwright (1983).

In this case the Causal Markov assumption alone puts no constraints on the distributions that this structure could produce, because we obtain no independencies whatsoever from applying d-separation or the Markov condition to its DAG. But in some of the distributions that this structure could produce, Smoking might be independent of Health "by coincidence." If Smoking has a negative direct effect on Health, but Smoking has a positive effect on Exercise (absurd as this may seem) and Exercise has a positive effect on Health, then Smoking serves to directly inhibit Health and indirectly improve it. If the two effects happen to exactly balance and thus cancel, then there might be no association at all between Smoking and Health. In such a case we say that the population is *unfaithful* to the causal graph that generated it.

If there are any independence relations in the population that are not a consequence of the Causal Markov condition (or d-separation), then the population is unfaithful. By assuming Faithfulness we eliminate all such cases from consideration. Although at first this seems like a hefty assumption, it really isn't. Assuming that a population is Faithful is to assume that whatever independencies occur in it arise not from incredible coincidence but rather from structure. Some form of this assumption is used in every science. When a theory cannot explain an empirical regularity save by invoking a special parameterization, then scientists are uneasy with the theory and look for an alternative that can explain the same regularity with structure and not luck. In the causal modeling case, the regularities are (conditional) independence relations, and the Faithfulness assumption is just one very clear codification of a preference for models that explain these regularities by invoking structure and not by invoking luck. By no means is it a guarantee; nature might indeed be capricious. But the existence of cases in which a procedure that assumes Faithfulness fails seems an awfully weak argument against the possibility of causal inference. Nevertheless, critics continue to create unfaithful cases and display them (see, for example, David Freedman's long paper in this volume).

Assuming Faithfulness seems reasonable and is widely embraced by practicing scientists. The inferential advantage gained from the assumption in causal inference is enormous. Without it, all we can say on the basis of independence data is that whatever causal structure generated the data, it cannot imply any independence relations by d-separation that are not present in the population. With it, we can say that whatever structure generated the data, it implies by d-separation *exactly* the independence relations that are present in the population. For example, suppose we have a population involving three variables X_1, X_2, X_3, and suppose the independence relations in this population are as given below.

All Possible[12] Independences among X_1, X_2, X_3	In Population	Not In Population
$X_1 \perp\!\!\!\perp X_2$	√	
$X_1 \perp\!\!\!\perp X_3$		√
$X_2 \perp\!\!\!\perp X_3$		√
$X_1 \perp\!\!\!\perp X_2 \mid X_3$		√
$X_1 \perp\!\!\!\perp X_3 \mid X_2$		√
$X_2 \perp\!\!\!\perp X_3 \mid X_1$		√

Even if we assume that all the Causally Markov graphs that might have produced data with these independencies involve only X_1, X_2, and X_3, then there are still nine such graphs. Their only shared feature is that each has some direct connection between X_1 and X_3 and between X_2 and X_3. Adding Faithfulness reduces the set of nine to a singleton (Fig. 9).

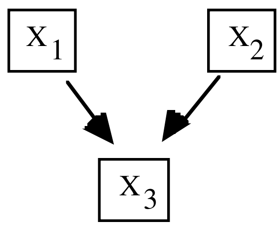

Fig. 9

Causal Sufficiency

In this example we have managed to infer that both X_1 and X_2 are direct causes of X_3 from a single marginal independence between X_1 and X_2. This gives many people pause, as it should. We have achieved such enormous inferential leverage in this case not only by assuming Faithfulness, but also by assuming Causal Sufficiency, which I noted above by writing: "all the Causally Markov graphs that . . . involve only X_1, X_2, and X_3."

The assumption of Causal Sufficiency is satisfied if we have *measured* all the common causes of the measured variables. Although this sounds quite similar to the assumptions about the completeness of causal graphs, it is not exactly the same thing. When we assume that a causal graph is complete with respect to common causes, it is in service of being clear about what sorts of systems we are representing with such graphs. In the inferential case we are making two assumptions: one involves the existence of some causal graph that is complete with respect to common causes and that is Causally Markov, and the other is an assumption about the variables we have measured as opposed to those we have not. For example, we might build a

[12] All possible non-trivial independences, that is.

model in which we specify a variable called Intelligence which we cannot directly or perfectly measure, but four of whose effects we can measure, say test scores X_1 - X_4 (Fig. 10).

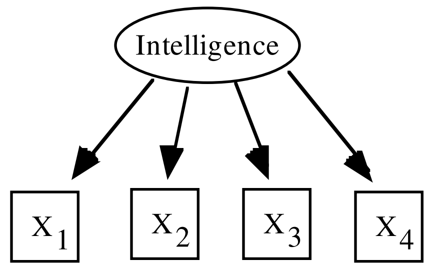

Fig. 10

Supposing that the causal graph among Intelligence and X_1 - X_4 is complete with respect to common causes, and that it is Causally Markov and Faithful to whatever population it produces over {Intelligence, X_1 - X_4}, then the following list of the independence relations will hold in this population.

$$X_1 \perp\!\!\!\perp X_2 \mid \text{Intelligence}$$
$$X_1 \perp\!\!\!\perp X_3 \mid \text{Intelligence}$$
$$X_1 \perp\!\!\!\perp X_4 \mid \text{Intelligence}$$
$$X_2 \perp\!\!\!\perp X_3 \mid \text{Intelligence}$$
$$X_2 \perp\!\!\!\perp X_4 \mid \text{Intelligence}$$
$$X_3 \perp\!\!\!\perp X_4 \mid \text{Intelligence}$$

Since Intelligence is unmeasured, however, our data will only include independencies that do not involve it, which in this case is the empty set. Thus the causal graph involving Intelligence and X_1 - X_4 is complete with respect to common causes, but the measured variables X_1 - X_4 are not Causally Sufficient. To summarize, the Causal Markov assumption, although it involves a representational form of causal sufficiency, is an assumption about the way causation and probability are connected, while Causal Sufficiency is an assumption about what we have managed to measure. I have so far discussed three different assumptions:

> 1) the Causal Markov assumption: upon accurately specifying a causal graph G among some set of variables V (in which V includes all the common causes of pairs in V), *at least* the independence relations obtained by applying d-separation to G hold in the population probability distribution over V.
>
> 2) the Faithfulness assumption: *exactly* the independence relations obtained by applying d-separation to G hold in the probability distribution over V.
>
> 3) the Causal Sufficiency assumption: the set of *measured* variables M include all of the common causes of pairs in M.

In the example concerning Faithfulness, we managed to infer the unique causal structure in Fig. 9 from the single marginal independence X_1 _||_ X_2 by making all three assumptions. It is still possible to make inferences about the structure(s) underlying the data without the Causal Sufficiency assumption, but of course we can learn less.

When we do not assume Causal Sufficiency, we still assume that there *is some* structure involving the measured variables (and perhaps other variables) that is complete with respect to common causes and that satisfies the Causal Markov assumption, but we must acknowledge that we might not have measured all the common causes. So whatever algorithm we use to move from independence relations to all the causal graphs that might have produced these independence relations, the set of graphs must include members that have common causes we have not measured. In the example in Fig. 9, we have measured X_1-X_3, and observed a single independence: X_1 _||_ X_2. If we assume Causal Markov and Faithfulness, but not Causal Sufficiency, the set of 10 causal graphs that would produce exactly this independence appears below (Fig. 11), where the T variables in circles are the common causes that we might not have measured.

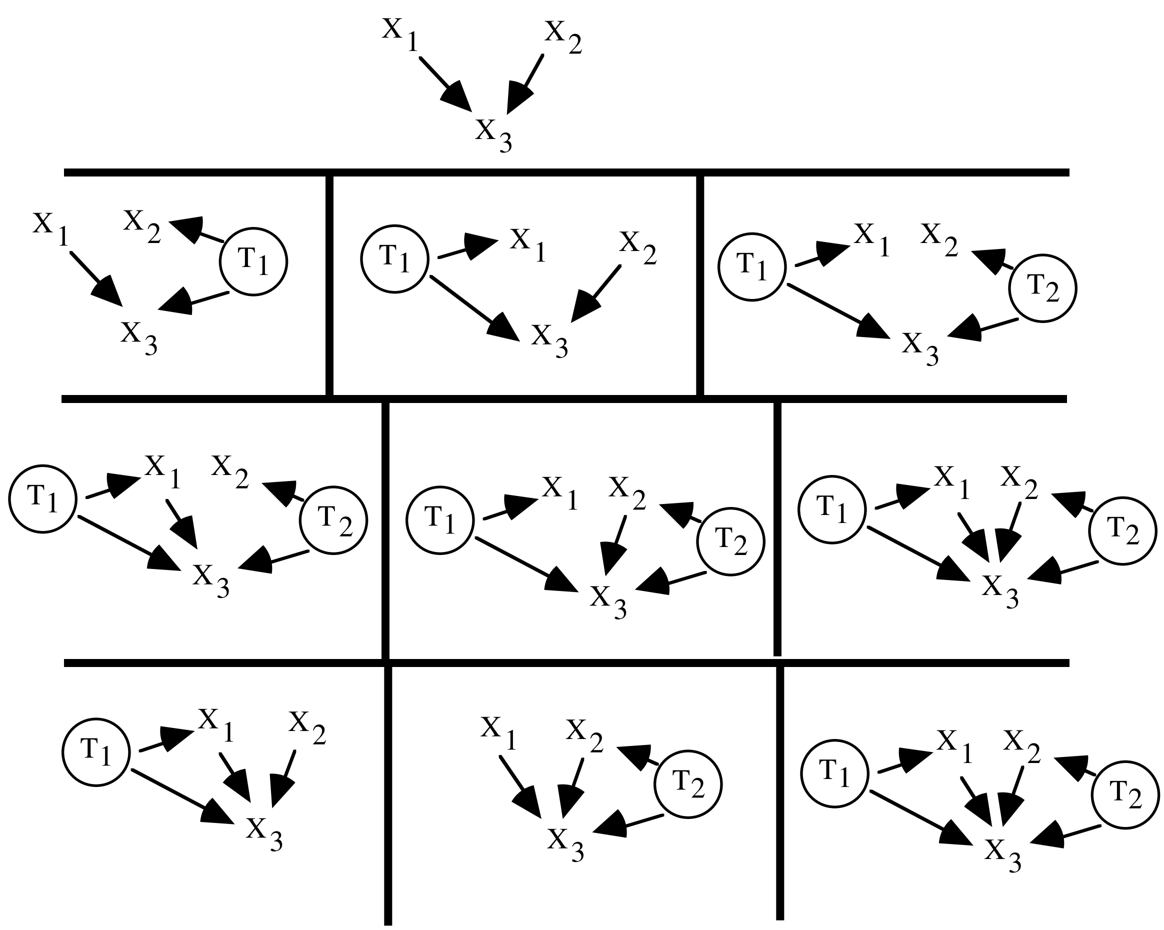

Fig. 11

In fact this set is still too small, for where we have specified a single unmeasured common cause of two variables (such as X_1 and X_3) and named it "T_1," in actuality there might be any number of distinct unmeasured common causes of X_1 and X_3. So wherever T_1 appears as a common cause (of say X_1 and X_3) this is really an abbreviation for: there exists some unmeasured common cause of X_1 and X_3.

Although dropping the assumption of Causal Sufficiency has reduced our inferential power considerably, it has not completely eliminated it. Notice that in none of the structures in Fig. 11 is X_3 a cause of any other variable. So we have learned something about what causal relations do not exist: X_3 is not a cause of X_1 or of X_2, even though it is associated with both. In other words, we have inferred from independence data the following: if we were to ideally manipulate X_3, then we would do nothing to alter X_1 or X_2.

Can we ever gain knowledge about what causal relations do exist without assuming Causal Sufficiency? Yes, but not unless we either measure at least four variables or make additional assumptions. For example, if the following independencies are observed among X_1-X_4, then assuming Causal Markov and Faithfulness we can conclude that in every graph that could possibly have generated this data, X_3 is a cause of X_4.

$$X_1 \perp\!\!\!\perp X_2$$
$$X_1 \perp\!\!\!\perp X_4 \mid X_3$$
$$X_2 \perp\!\!\!\perp X_4 \mid X_3$$

That is: if Causal Markov and Faithfulness are satisfied, then from these independence relations we can conclude that a manipulation of X_3 would change the probability of X_4.

Adding other sorts of knowledge often improves the situation, e.g., knowledge about the time order of the variables. In the following case from James Robins, for example, we can obtain knowledge that one variable is a cause of another when we have only measured three variables, and we do so by assuming Causal Markov, Faithfulness, but not Causal Sufficiency. If we know that X_1 occurs before X_2 and X_2 before X_3, and we know that in the population $X_1 \perp\!\!\!\perp X_3 \mid X_2$, then under these assumptions we can conlcude that X_2 is a cause of X_3. We can also conclude that there is no unmeasured common cause of X_2 and X_3.

5. Conclusion

Contrary to what some take to be our purpose, we are not about trying to magically pull causal rabbits out of a statistical hat. Our theory of causal inference investigates what can and cannot be learned about causal structure from a set of assumptions that seem to be made commonly in scientific practice. It is thus a theory about the inferential effect of a variety of assumptions far more than it is an endorsement of particular assumptions. There are situations in which it is unreasonable to endorse the Causal Markov assumption (e.g., in quantum mechanical settings), Causal Sufficiency rarely seems reasonable, and there are certain situations where one might not want to assume Faithfulness (e.g., if some variables are completely determined by others). In the Robins case above, for example, we inferred that there was no unmeasured common cause, or "confounder," of X_2 and X_3. Robins believes that in epidemiological contexts there are always unmeasured confounders, and thus makes an informal Bayesian argument in which he decides that his degrees of belief favor giving up Faithfulness before accepting the conclusion that in this case it forced.

If our theory has any good effect on practice, it will be as much to cast doubt on pet theories by making it easy to show that reasonable and equivalent alternatives exist than it will be to extract causal conclusions from statistical data. If it succeeds in clarifying the scientific rationale that underlies causal inference, which is our real goal, then its most important effect will be to change the way studies are designed and data is collected.

References

Bollen, K. 1989. Structural equations with latent variables. Wiley, New York.

Cartwright, N. 1983. How the Laws of Physics Lie. Oxford University Press, New York.

Cartwright, N. 1989. Nature's capacities and Their Measurement. Clarendon Press, Oxford.

Dawid, A. 1979. “Conditional independence in statistical theory (with discussion).” Journal of the Royal Statistical Society B, 41, 1-31.

Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. 1987. Discovering causal structure. Academic Press, San Diego, CA.

Hausman, D. (1984). “Causal priority.” Nous 18, 261-279.

Kiiveri, H. and Speed, T. 1982. "Structural analysis of multivariate data: A review." Sociological Methodology, Leinhardt, S. (ed.). Jossey-Bass, San Francisco.

Papineau, D. 1985. “Causal Asymmetry.” British Journal of Philosophy of Science, 36: 273-289.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems. Morgan and Kaufman, San Mateo CA.

Reichenbach, H. 1956. The direction of time. Univ. of California Press, Berkeley, CA.

Salmon, W. 1984. Scientific explanation and the causal structure of the world. Princeton Univ. Press, Princeton, NJ.

Scheines, R., Spirtes, P., Glymour, C., and Meek, C. 1994. TETRAD II: User's Manual, Lawrence Erlbaum and Associates, Hillsdale, NJ.

Skyrms, B. 1980. Causal necessity: a pragmatic investigation of the necessity of laws. Yale University Press, New Haven.

Spirtes, P., Glymour, C., and Scheines, R. 1993. Causation, prediction, and search. Lecture Notes in Statistics, V. 81, Springer-Verlag, New York.

Spirtes, P. 1994. "Conditional Independence in Directed Cyclic Graphical Models for Feedback." Technical Report CMU-PHIL-54, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.

Suppes, P. 1970. A probabilistic theory of causality. North-Holland, Amsterdam.