Causal Structure among Measured Variables Preserved with Unmeasured Variables

by

Peter Spirtes, Clark Glymour

April 1990

Report CMU-PHIL-14

Carnegie Mellon

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

Causal Structure among Measured Variables Preserved with Unmeasured Variables

Peter Spirtes and Clark Glymour

Department of Philosophy
Carnegie Mellon University

April 25, 1990

In recent papers we have described a framework for inferring causal structure from relations of statistical independence among a set of measured variables. Using Pearl's notion of the perfect representation of a set of independence relations by a directed acyclic graph we proved :

> **Theorem I** (Verma and Pearl): If a directed acyclic graph G perfectly represents a distribution P over a set of variables V, then vertices A and B are adjacent in G if and only if for all subsets S of V containing neither A nor B, A, B are not independent conditional on S for distribution P.

> **Theorem 2** (Spirtes and Glymour): If directed acyclic graph G perfectly represents a distribution P over a set of variables V and in G there is an edge between vertices A and B and an edge between vertices B and C, but no edge between vertices A and C, then there is an edge from A to B and from C to B if and only if for all subsets S of V containing B but containing neither A nor C, A,C are not independent conditional on S for distribution P.

We conjectured that Theorems 1 and 2 characterize all and only the directed graphs that perfectly represent a distribution P, and we proved a restricted case of the conjecture. We used these theorems to implement in TETRAD II an algorithm that for normally distributed variables builds the corresponding class of directed acyclic graphs.

One of the important questions about using such procedures on real data concerns which conclusions about causal structure based on the measured sample remain valid if there are unmeasured variables at work. When, applying Theorems 1 and 2 to a set of measured variables, are concusions about the presence or absence of a directed path from measured variable A to measured variable B valid even if the true causal structure contains unmeasured variables that act as common causes of the measured variables? When, in other words, can we be confident of causal conclusions even though we may not know whether there are latent variables acting? Some answers to this question follow as corollaries to theorem recently proved by Verma and Pearl.

Our results build on recent theorems by Verma and Pearl. Using Theorems 1 and 2 (their Lemmas 1 and 2) they prove our first conjecture:

> **Theorem 3** (Verma and Pearl): If distribution P is perfectly represented by directed acyclic graph H, then graph G perfectly represents P if and only if
>
> (i) there is an edge between A and B in G if and only if for all subsets S of V containing neither A nor B, A, B are not independent conditional on S for distribution P; and
>
> (ii) if in G there is an edge between vertices A and B and an edge between vertices B and C, but no edge between vertices A and C, then there is an edge from A to B and from C to B if and only if for all subsets S of V containing B but containing neither A nor C, A,C are not independent conditional on S for distribution P.

Verma and Pearl prove two other theorems that will be essential in what follows. These results essentially characterize the causal structure induced on a subset of variables of a directed acyclic graph representing a distribution P by the marginal of P on the subset. They require some definitions:

From the equivalence class of a graph construct the partially directed graph with directed edges representing the edges whose direction is common to all members of the class. This partially directed graph is the **pattern** of the equivalence class and of each of its members. A **hybrid graph** contain may contain bidirected edges edges. Such graphs may be used to represented the marginal structure on a set of measured variables when unmeasured common causes have edges that collide (in Verma and Pearl's terminology, are "head to head") with edges between measured variables. **A o-> B** denotes that a pattern contains an edge between A, B with at least a head pointed into B.

Let D be a directed acyclic graph with vertex set U of which the subset O are observable. **The pattern P of D restricted to O** is the hybrid graph with fewest arrowheads that satisfies the following conditions:

(i) A,B are adjacent in P if and only if for all S included in O but containing neither A nor B, A and B are not independent conditional on S.
(2) A o-> B if there exists a C in O such that for all subsets S of O - {A,B,C} A is adjacent to B and B is adjacent to C in P, A is not adjacent to C in P, and A and C are not independent conditional on S U {B}.
(3) A o-> B if there exists a C in O - {A,B} such that C o-> A is in P and A -> B is in D.

Theorem 4: (Verma and Pearl): Let O be a subset of U, D a DAG over U, and P the pattern of P restricted to O. Two variables A, B in O are adjacent in P if and only if there exists a path p between A and B in D satisfying the following two condiitons:

(1) Every observable node on p must be a head-to-head node along p
(2) Every head to head node along p must be a shieldable ancestor of either A or B.

where an ancestor S of A is shieldable if and only if every directed path from S to A contains an observable other than A.

Theorem 5 (Verma and Pearl): For any pattern P, A o-> B if and only if there is a node C such that either (1) C is adjacent to B and not A (in P) and both edges A-B and B-C were induced by paths (of D) which ended pointing at B, or (2) C o-> A in P and B is a descendent of A in D.

x-y denotes x and y are adjacent in a pattern P. We will write x -> y in P iff x o-> y in P, and not y o-> x in P, and x <-> y if x o-> y and y o-> x. In order to state our results, we need the following definition.

Definition 1: In a DAG or a pattern, x, y, and z form a **triangle** if x is adjacent to y and z, and z is adjacent to y.

Definition 2: In a pattern P, an **undirected path u from x to y** is a sequence of vertices in P starting with x and ending with y such that each pair of adjacent vertices in the sequence is a pair of adjacent vertices in P. The **edge x-y is in u** iff x and y are adjacent in the sequence.

Definition 3: In a pattern P, a **semi-directed path p from x to y** is an undirected path from x to y such that there is no edge a -> b in p such that b occurs before a in u.

Definition 4: In a DAG G, a **trek between distinct vertices x and y,** is either a directed path from x to y, a directed path from y to x, or a pair of directed paths from a common source c to x any y respectively, that intersect only at c.

We will show the following;

Theorem 6: Let G be a graph over a set of vertices U, and O be a subset of U containing x and y, and P the pattern of G over O. If there exists a directed path p from x to y in G then P contains an undirected path u from x to y, and if P does not contain a triangle containing any two adjacent vertices on u, then u is a semi-directed path from x to y in P.

Theorem 6 tells us that if the probability distribution on the set of measured variables is not compatible with a directed path p from x to y that contains no triangles involving adjacent edges on p, then if that distribution is the marginal of a distribution perfectly represented by a graph on a larger vertex set, the larger graph also contains no directed path from x to y.

The following theorem states under what conditions it is possible to infer from a pattern the existence of a directed path in a graph G.

Theorem 7: Let O be a subset of vertices of G containing x and z, and let the pattern of G for O contain a directed edge x -> z, no triangle containing x and z, and a variable c such that c o-> x. Then in G there is a directed path from x to z.

Furthermore:

Corollary 1: Let O be a subset of vertices of G containing x and z, and let the pattern of G for O contain a vertex c such that c o-> x, and a directed path p from x to z such that no adjacent pair u, w on p is in a triangle in P. Then in G there is a directed path from x to z.

Lemma 1: Let G be a graph over a set of vertices U, O be a subset of U containing x and y, and P be the pattern of G restricted to O. If there is a directed path in G from x to y that induces an edge in P, then there is no edge y -> x in P.

Proof. Suppose that the antecedent is true, but the consequent is false. We will show that if the orientation y o-> x is induced by a path in G, then the orientation x o-> y is also induced by G.

Suppose y o-> x is true in the pattern P. Hence either (i) there is a c such that c is adjacent to x and c is not adjacent to y in P and the edges x-y and x-c are induced by paths pointing into x or (ii) there is a c such that c o->y and x is a descendent of y in G. (ii) is impossible because it implies the existence of a cycle in G. So assume (i) is the case.

(i) implies that the edges y-x and x-c are induced by undirected paths in G both pointing into x, and c is adjacent to x in G. This in turn implies that c o-> x in P. Clause (ii) of Theorem 5 now implies that x o-> y in P.

Theorem 6: Let G be a graph over a set of vertices U, and O be a subset of U containing x and y, and P the pattern of G over O. If there exists a directed path p from x to y in G then P contains an undirected path u from x to y, then u is a semi-directed path from x to y in P.

Proof. Break the path p in G into a series of subpaths such that only the endpoints of the subpath are in O. Let a be the source and b be the sink of some such arbitrary subpath. The edge a-b exists in P by Theorem 4. a is prior to b on p. The concatenation of the edges induced by the subpaths are an undirected path u from x to y in P. By Lemma 1, it is not the case that b -> a. By definition of semi-directed path, u is a semi-directed path from x to y.

Theorem 7: Let O be a subset of vertices of G containing x and z, and let the pattern P of G for O contain a directed edge x -> z, no triangle containing x and z, and a variable c such that c o-> x. Then in G there is a directed path from x to z.
Proof. Since x-z in P there is a path p in G between x and z such that every observable node on p is head to head and every head to head node on p is a shieldable ancestor of x or z.

If x -> z in pattern P arises because of clause (2) of Theorem 5, we are done because z is a descendant of x in G. So suppose x -> z is oriented by condition (1) of Theorem 5. Then there is a path p in G that induces x -> z and p is into z in G. If p contains no head to head node and is not into x, then p is a directed path from x to z, and we are done. Otherwise there are two cases: p contains a head to head node, or p is into x.

First, we consider the case where p is into x. Then there is a path between x and z that induces an edge in P, and is into x. By assumption there is a vertex d in P such that d o-> x. By Theorem 5, either there is a vertex c in P such that c is adjacent to x and not d, and both edges c-x and d-x were induced by paths of G which ended pointing at x, or c o-> d in P and x is a descendant of d in G.

Suppose that the first disjunct is true. In that case, either c is adjacent to z or it isn't. If it is adjacent to z, then there is a triangle in P containing x and z. If c is not adjacent to z, then by clause (2) of Theorem 5, z o-> x, contrary to our assumption.

Suppose now that the second disjunct is true. Since x is a descendant of d in G, there is a directed path from some variable e in O to x that does not contain any variables in O other than x and e. This path induces an edge e-x in P. If e is adjacent to x in P, then P contains a triangle containing x and z; if e is not adjacent to x in P, then by clause (2) of Theorem 5, z o-> x, contrary to our assumption.

We now consider the case where p is not into x, but there is a head to head vertex on p. Let k be the first head to head node on p after x. If p is not into k, then there is a path from x to k, and hence no path from k to x. So by clause (2) of Theorem 4, k is a shieldable ancestor of z. The concatenation of the paths from x to k and from k to z is a directed path from x to z, and we are done.

Corollary 1: Let O be a subset of vertices of G containing x and z, and let the pattern of G for O contain a vertex c such that c o-> x, and a directed path p from x to z such that for no adjacent pair u, w on p is there a triangle in P containing both u and w. Then in G there is a directed path from x to z.

Proof. Since there is a directed path from x to z in P, there is a sequence of edges x -> a -> b ... -> z in P. By Theorem 8, there is a directed path from x to a in G. Since x -> a, Theorem 8 can next be applied to a -> b, to show that there is a directed path from a to b in G. Repeating this process in turn for each edge on p, implies that there is a directed path from x to z in G.

References

J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988.

P. Spirtes, C. Glymour and R. Scheines, "Causality from Probability" Proceedings of the Conference on Advanced Computing for the Social Sciences, Williamsburg, Va. 1990.

P. Spirtes, C. Glymour and R. Scheines, "Causality from Probability" in G. McKee, ed. *Evolving Knowledge in Natural and Artificial Intelligence*, Pitman, 1990.

P. Spirtes, C. Glymour and R. Scheines, "From Probability to Causality" *Philosophical Studies*, forthcoming.

T. Verma and J. Pearl, "On Equivalence of Causal Models" Technical Report, R-150, Department of Computer Science, University of California at Los Angeles, April 1990.