Directed Cyclic Graphs, Conditional Independence, and Non-Recursive Linear Structural Equation Models

by

Peter Spirtes

April 1993

Report CMU-PHIL-35

Carnegie Mellon

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

Directed Cyclic Graphs, Conditional Independence and Non-Recursive Linear Structural Equation Models

Peter Spirtes[1]

Department of Philosophy
Carnegie Mellon University
Pittsburgh, Pa. 15213
e-mail: ps7z@andrew.cmu.edu

Department of Philosophy Technical Report CMU-Phil-35

April 26, 1993

[1]I thank C. Glymour, R. Scheines, M. Drudzel and H. Simon for helpful conversations. Research for this paper was supported by the National Science Foundation through grant 9102169.

Abstract

Recursive linear structural equation models can be represented by directed acyclic graphs. When represented in this way, they satisfy the Markov Condition. Hence it is possible to use the graphical d-separation to determine what conditional independence relations are entailed by a given linear structural equation model. I prove in this paper that it is also possible to use the graphical d-separation applied to a *cyclic* graph to determine what conditional independence relations are entailed to hold by a given *non*-recursive linear structural equation model. I also give a causal intepretation to the linear coefficients in a non-recursive structural equation models, and explore the relationships between cyclic graphs and undirected graphs, directed acyclic graphs with latent variables, and chain independence graphs.

1. Introduction

Over the last decade, a number of investigators have contributed to our understanding of a class of statistical models, directed acyclic graph (or DAG) models, encoding independence and conditional independence constraints. (For a good introduction, see Pearl 1988. Henceforth we will simply say "conditional independence constraints" rather than "independence and conditional independence constraints".) As a consequence of this research, DAG models have acquired several useful features, including a relatively clear causal interpretation, easily computed maximum likelihood estimates for multinomial and other distribution families, efficient decision procedures for determining the statistical indistinguishability of DAGs, procedures for forming conditional distributions, reliable procedures for generating a class of DAG models from sample data and background knowledge, etc. The key to these developments was the formulation of a purely graphical condition for the conditional independence of variables in DAG models.

DAG models are almost generic; a variety of familiar statistical formalisms, such as linear structural equation models, various types of regression models, factor analytic models, path models, and discrete latent variable models can be represented as DAG models. But DAG models do exclude a kind of model familiar in engineering and economics. Processes with feedback are sometimes represented by simultaneous equations in which a variable X_1 is expressed as a function of X_2 and other variables, and X_2 is also a function of X_1 and other variables. Such models are naturally associated with directed *cyclic* graphs, and one would therefore like to have a theoretical understanding comparable to that now available for DAG models. The econometric literature has developed an estimation theory for linear "non-recursive" models, and the engineering literature contains algorithms that can be used to compute correlations in the linear case for non-recursive systems that are based on cyclic graphs. The first essential step in generalizing from acyclic to cyclic directed graphical models requires a purely graphical condition for conditional independence. The principal result of this paper is such a condition for linear models. The condition may generalize to the non-linear case, but proofs for the generalization have not yet been obtained.

2. DAG Models

A DAG is an ordered pair <**V**,**E**> where **V** is a set of random variables, and **E** is a set of directed edges (ordered pairs of vertices) between random variables. By definition, there are no directed cycles in a DAG, i.e. there are no directed paths from a vertex to itself. In a DAG model the directed acyclic graph can be used for two quite distinct purposes. On the one hand, each DAG *G* can be paired with any member *P* of families of probability distributions over variables represented by vertices in the graph. One fundamental condtions relates DAGs to distributions. The **Markov Condition** states that for admissible <*G*, *P*>, *X* is independent of its non-descendants in *G* given its parents in *G*. For a positive distribution, the Markov Condition entails that the joint density function over the variables in **V** can be factored according to the formula:

$$f(\mathbf{V}) = \prod_{V \in \mathbf{V}} f(V | \mathbf{Parents}(V))$$

where *A* is in **Parents**(*V*) if and only if there is a directed edge from *A* to *V* in *G*, and we use "*f*" to denote joint, marginal, and conditional density functions.

A DAG *G* can also be used to represent hypothetical causal relations between random variables. Under such an interpretation, an edge *A* -> *B* indicates that there is a direct influence of variable *A* on variable *B* not blocked by holding constant any other variables in the system.

Pearl, Geiger, and Verma have shown that given a DAG *G*, there is a graphical condition among disjoint sets of variables **A**, **B**, and **C** that holds if and only if for all distributions satisfying the Markov Condition for *G*, **A** is independent of **B** given **C**. In order to define d-separability, we need the following definitions. A directed edge from *A* to *B* is an ordered pair <*A*,*B*> in which *A* is the **tail** of the edge and *B* is the **head**; we say the edge is **out of** *A* and **into** *B*, and *A* is **parent** of *B* and *B* is a **child** of *A*. Let an **undirected path** *U* be a sequence of edges $<E_1,...,E_n>$ in *G* such that for $1 \leq i < n$ one of the endpoints of E_i equals one of the endpoints of E_{i+1}, and $E_i \neq E_{i+1}$[2] A path *U* is **acyclic** if

[2]In a number of articles and books, including Spirtes, Glymour and Scheines(1993), an undirected path is defined as a sequence of vertices rather than a sequence of edges. The two definitions are essentially equivalent for DAG models, because a pair of vertices can be identified with a unique edge in the graph.

no vertex occurring on an edge in the path occurs more than once. Let a **directed path** D be a sequence of edges $<E_1,...,E_n>$ in G such that for $1 \le i < n$ the head of E_i equals the tail of E_{i+1}, and $E_i \ne E_{i+1}$. If there is an acyclic directed path from A to B or $B = A$ then A is an **ancestor** of B, and B is a **descendant** of A. A vertex V is a **collider** on undirected path U if and only if two adjacent edges on U are into V. For a directed acyclic graph G, if X and Y are vertices in G, $X \ne Y$, and $\mathbf{W}$ is a set of vertices in G not containing X or Y, then X and Y are **d-separated** given $\mathbf{W}$ in G if and only if there exists no acyclic undirected path U between X and Y, such that (i) every collider on U has a descendent in $\mathbf{W}$ and (ii) no other vertex on U is in $\mathbf{W}$. We say that if $X \ne Y$, and X and Y are not in $\mathbf{W}$, then X and Y are **d-connected** given $\mathbf{W}$ if and only if they are not d-separated given $\mathbf{W}$. If $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ are disjoint sets of vertices in G then we say that $\mathbf{U}$ and $\mathbf{V}$ are **d-separated** given $\mathbf{W}$ if and only if every pair $<U,V>$ in the cartesian product of $\mathbf{U}$ and $\mathbf{V}$ is d-separated given $\mathbf{W}$. If $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ are disjoint sets of vertices in G then we say that $\mathbf{U}$ and $\mathbf{V}$ are **d-connected** given $\mathbf{W}$ if and only if $\mathbf{U}$ and $\mathbf{V}$ are not d-separated given $\mathbf{W}$.

D-Separation Theorem (See Pearl 1988): In a DAG G containing disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ if and only if in every distribution that satisfies the Markov Condition for G, $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$.

Linear structural equation models (which we will refer to as linear causal models or LCTs) can also be represented as DAG models. The following is an example of a an LCT, where a and b are real constants, ε_X ,ε_Y, and ε_Z are "error terms", and X, Y, Z, are random variables:

$$X = a \times Y + \varepsilon_X$$
$$Y = b \times Z + \varepsilon_Y$$
$$Z = \varepsilon_Z$$

ε_X, ε_Y, and ε_Z are jointly independent and normally distributed

An LCT consists of a set of linear equations relating random variables to each other and a unique exogenous "error term", and a joint distribution over the error terms. The set of equations can be represented by a directed graph, in which all of the non-error terms that appear in the equation for a given variable V are parents of V in the graph. (See figure 1.)

However, a cyclic graph may contain more than one edge between a pair of vertices. In that case it is no longer possible to identify a pair of vertices with a unique edge.

Figure 1

If the directed graph that represents the set of equations is acyclic, the model is said to be recursive.

We say that an LCT containing disjoint sets of variables **X**, **Y**, and **Z** **linearly entails** that **X** is independent of **Y** given **Z** if and only if **X** is independent of **Y** given **Z** for all values of the non-zero linear coefficients and all distributions of the exogenous variables in which they have positive variances. Similarly an LCT containing X, Y, and **Z**, where $X \neq Y$ and X and Y are not in **Z**, **linearly entails** that $\rho_{XY.\mathbf{Z}} = 0$ if and only $\rho_{XY.\mathbf{Z}} = 0$ for all values of the non-zero linear coefficients and all distributions of the exogenous variables in which they have positive variances. Kiiveri and Speed (1982) pointed out that if the error terms are jointly independent, then any distribution that is an LCT with an *acyclic* directed graph G satisfies the Markov Condition for G; one can therefore use the d-separation relation applied to the DAG in a recursive LCT to compute the conditional independencies and zero partial correlations linearly entailed by the LCT.

3. Non-Recursive LCTs and Cyclic Graphs

Non-recursive LCTs are commonly used in the econometrics literature to represent feedback processes that have reached equilibrium. Corresponding to a set of non-recursive linear equations is a cyclic graph, as the following example from Whittaker (1990) illustrates.

$$X_1 = \varepsilon_{X1}$$
$$X_2 = \varepsilon_{X2}$$
$$X_3 = \beta_{31} \times X_1 + \beta_{34} \times X_4 + \varepsilon_{X3}$$
$$X_4 = \beta_{42} \times X_2 + \beta_{43} \times X_3 + \varepsilon_{X4}$$

ε_{X1}, ε_{X2}, ε_{X3}, ε_{X4} are jointly independent and normally distributed

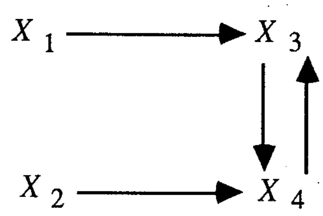


Figure 2

The graph is written with the convention that an edge does not appear if and only if the corresponding entry in the coefficient matrix is zero.

The use of non-recursive LCTs to represent feedback processes presents several problems of interpretation. The coefficients in a recursive LCT can be interpreted as regression coefficients, and a zero coefficient always linearly entails some conditional independence. In contrast, Haavelmo (1943) pointed out that the coefficients in a non-recursive LCT cannot be interpreted as regression coefficients. Whittaker (1990) pointed out that in a non-recursive LCT it is not possible to interpret the absence of an edge between vertices *A* and *B* as independence of *A* and *B* conditional on some subset of variables determined by the graphical structure. Indeed, there are non-recursive LCTs with zero coefficients (i.e. absent edges) that do not imply any conditional independencies at all. We will consider the problem of interpreting the linear coefficients both in terms of (i) the causal interpretation attached to them, and (ii) the implications of the linear coefficients for conditional independence.

In fact, the causal interpretation of the linear parameters is as clear in the cyclic as in the acyclic case, and as clear in linear systems as in nonlinear. In order to explain the causal interpretation of the coefficients we must distinguish between prediction in a population with a fixed probability distribution, and prediction when the probability distribution has been changed by an outside intervention, or manipulation of the variables. The difference between these two kinds of prediction was well described in Haavelmo's (1943) article:

> The economist may have two different purposes in mind when he constructs a model...
>
> First, he may consider himself in the same position as an astronomer; he cannot interfere with the actual course of events. So he sets up the system ... as a tentative description of the economy. If he finds that it fits the past, he hopes it will fit the

future. On that basis he wants to make predictions, assuming that no one will interfere with the game.

Next, he may consider himself as having the power to change certain aspects of the economy in the future. If then the system ... has worked in the past, he may be interested in knowing it as an aid in judging the effect of his intended future planning, because he thinks that certain elements of the old system will remain invariant.

For example, in the actual population there are many causes of smoking, such as peer pressure. However, one might intervene to ban all smoking, and if the ban is effective, then the value of smoking has been manipulated to zero, and peer pressure is no longer a cause of smoking. We must be careful to distinguish between the unmanipulated population (in which peer pressure causes smoking) and the unmanipulated population (in which peer pressure does not cause smoking.) In order to predict the effect of an intervention, one must modify the equations so that they describe the manipulated population, not the unmanipulated population. This can be done by setting variables whose values have been set by outside manipulation to constants. For example, consider the following non-recursive LCT.

$$X = \varepsilon_X$$
$$Y = a \times X + b \times Z + \varepsilon_Y$$
$$Z = c \times Y + \varepsilon_Z$$

ε_X, ε_Y, and ε_Z are jointly independent

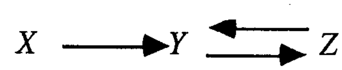


Figure 3

Suppose the variable Y in figure 3 is manipulated to a given value. In that case, the causes of Y in the pre-manipulation population (X, Z and ε_Y) are no longer causes of Y in the manipulated population. The manipulation can be represented by writing a new set of structural equations in which Y is simply equal to a constant:

$$X = \varepsilon_X$$
$$Y = d$$
$$Z = c \times Y + \varepsilon_Z$$

ε_X and ε_Z are independent

Figure 4

(This assumes of course that the only effect on the intervention was to set the value of *Y*, and that the intervention did not interfere with the other causal mechanisms described by the other equations.) The new set of equations can be represented by the graph in figure 4 in which all edges into *Y* have been broken, but all other edges have been left intact.

The coefficient *c* of *Y* in the equation for *Z* now has the following simple causal interpretation. If all of the other parents of *Z* are manipulated to fixed values, then a unit change in the value of *Y* produces a change of *c* in the value of *Z*. One consequence of this interpretation is that since the coefficient of *X* in the equation for *Z* is zero, then when all of the parents of *Z* (i.e. *Y*) are manipulated to have a constant value, a change in the value of *X* has no effect on the value of *Z*. Note that this does not imply that *X* is independent of *Z* conditional on *Y* in the unmanipulated population, although *X is* independent of *Z* given *Y* in the manipulated population. (A more general theory of prediction of the effects of manipulations is presented in Spirtes, Glymour, and Scheines (1993). Haavelmo (1943) presents an example of a manipulation in which his calculations seem to proceed from implicit assumptions similar to the ones described here.)

How can we use the directed graph to determine what conditional independence relations are linearly entailed by a directed graph, cyclic or acylic? There are suggestions that just as d-separation allows one to calculate the conditional independence relations and zero partial correlations linearly entailed by a recursive LCT, so d-separation allows one to calculate the conditional independence relations and zero partial correlations linearly entailed by a non-recursive LCT, and this result is conjectured in Spirtes, et al. (1993).Essentially this result is proved in Glymour, et al. (1987) for the special case of a separating set consisting of a single variable. Another special case follows from a result presented in Whittaker (1990). In a DAG *G*, *V* is a **common cause** of *A* and *B* if and only if there is a directed path from *V* to *A* that does not contain *B*, and a directed path from *V* to *B* that does not contain *A*. When the set **U** of all variables in a separating set and the separated variables together includes all common causes of members of **U**, the result follows by writing the error variable for each member of **U** as a function of variables in **U** and computing the inverse covariance matrix, as in Whittaker(1990), p. 302.

The general conjecture is correct:

Cyclic D-Separation Theorem 1: In an LCT L with (cyclic or acyclic) directed graph G containing disjoint sets of variables **X**, **Y** and **Z**, **X** is d-separated from **Y** given **Z** if and only L linearly entails that **X** is independent of **Y** given **Z**.

Cyclic D-Separation Theorem 2: In an LCT L with (cyclic or acyclic) directed graph G containing X, Y and **Z**, where $X \neq Y$ and **Z** does not contain X or Y, X is d-separated from Y given **Z** if and only L linearly entails that $\rho_{XY.\mathbf{Z}} = 0$.

The proofs are given in the Appendix, but the idea is as follows. Haavelmo (1943) noted that the non-error terms in an LCT can be derived from a variable transformation of the error terms, and vice-versa. Hence the joint distribution of the non-error terms can be derived from the joint distribution of the error terms by variable substitution. Let the set of non-error variables in an LCT L be **V**. We will denote the set of error terms for variables in subset $\mathbf{S} \subseteq \mathbf{V}$ as **Err(S)**. In an LCT, each non-error variable can be expressed as a linear function of the error variables. Because the system of equations is linear, each error variable can also be expressed as a linear function of non-error variables. In an LCT with graph G, for a variable X in **V**, let **Parents**(X) be the set of variables that are parents of X in graph G, and **Ancestors**(X) be the set of variables that are ancestors of X in graph G. (Recall that the graph contains only non-error variables.) For each X in **V**, ε_X (the error variable for X) is a linear function of X and the non-error parents of X in the graph, that is ε_X = g(X,**Parents**(X)).

By assumption

$$f(\mathbf{Err(V)}) = \prod_{\varepsilon \in \mathbf{Err(V)}} f(\varepsilon)$$

We can derive the density function for the set of variables **V** by replacing each ε_i in f(ε_i) by g(X_i,**Parents**(X_i)).

$$f(\mathbf{V}) = \prod_{X \in \mathbf{V}} f(g(X, \mathbf{Parents}(X)) \times J_{\mathbf{Err(V)} \rightarrow \mathbf{V}}$$

where $J_{\mathbf{Err(V)} \rightarrow \mathbf{V}}$ is the Jacobian of the transformation. Because the transformation is linear, the Jacobian is constant.

When determining whether **X** and **Y** are independent given **Z** we are actually interested in transforming only some of the error variables. It follows from a result in Lauritzen et. al. (1990) that in the case of distributions that satisfy the Markov Condition for an acyclic directed graph *G*,

$$f(\mathbf{Ancestors(X \cup Y \cup Z)}) = \prod_{V \in \mathbf{Ancestors(X \cup Y \cup Z)}} f(g(V, \mathbf{Parents}(V)) \times c$$

and this result extends to linear structural equations represented by cyclic graphs also.

If *X* is not in **Z** then let *W* be a member of **Samefactor**(*X*,**Y**,**Z**) just when *W* is not a member of **Z** and there is some factor in

$$\prod_{R \in \mathbf{Y}} f(g(R, \mathbf{Parents}(R))$$

that contains both *W* and *X*. Let **Samefactor***(*X*,**Y**,**Z**) be the transitive closure of **Samefactor**(*X*,**Y**,**Z**), i.e. *W* is in **Samefactor***(*X*,**Y**,**Z**) if and only if *W* is in **Samefactor**(*X*,**Y**,**Z**) or there is some *M* in **Samefactor***(*X*,**Y**,**Z**) such that *W* is in **Samefactor**(*M*,**Y**,**Z**). Let *W* be in **Samefactor*(X,Y,Z)** if and only if *W* is in **Samefactor***(*X*,**Y**,**Z**) for some *X* in **X**.

The Appendix contains a proof that in an LCT with directed graph *G*, if **Y** is d-separated from **X** given **Z** then no member of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**, and if no member of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)** then it is possible to partition **Ancestors(X ∪ Y ∪ Z)** into two sets **A** and **B** that overlap only in **Z**, and such that the joint density of **Ancestors(X ∪ Y ∪ Z)** is equal to *h*(**A**)*h*'(**B**). The latter fact entails that **X** is independent of **Y** given **Z**.[3]

[3]There is a graphical relationship that corresponds to **Samefactor**(*X*,**Y**,**Z**) which we can call d-adjacency. In graphical terms, a variable *V* is d-adjacent to *X* given **Y** and **Z** (and hence in **Samefactor**(*X*,**Y**,**Z**)) if and only if in the subgraph of *G* containing only members of **Y**, *V* is not in **Z** and is either a child of *X*, a parent of *X*, or *V* and *X* have a common child. Just as there is an undirected path between *X* and *Y* if and only if *Y* is in the transitive closure of the variables adjacent to *X*, so **X** is d-connected to **Y** given **Z** if and only if **Y** is in the transitive closure of the variables d-adjacent to **X** given **Ancestors(X ∪ Y ∪ Z)** and **Z**. This observation is closely related to the characterization of d-separation in Lauritzen et. al. (1990).

It is interesting to note that although the d-separation relation allows one to calculate the conditional independencies linearly entailed by either acyclic or cyclic LCTs, in some sense d-separation plays a fundamentally different role in cyclic graphs than it does in acyclic graphs. In acyclic graphs, d-separation is ususally interpreted as a device for calculating the consequences of the Markov Condition. However, it does not play this role for cyclic graphs. Consider, for example the cyclic graph depicted in figure 5.

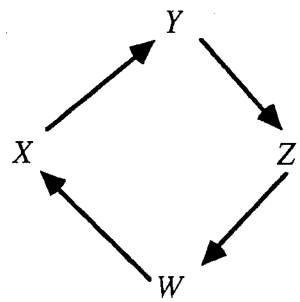


Figure 5

The LCT with the graph in figure 5 linearly entails two conditional independence relations: X and Z are independent conditional on Y and W, and Y and W are independent conditional on X and Z. However, a straightforward extension of the Markov Condition to cyclic graphs does not entail any conditional independence relations (because every vertex is a descendant of every other vertex.)

4. Cyclic Graphs and Latent Variables

Consider a directed acyclic graph G representing a causal process, and any associated probability distribution P, where $<G,P>$ satisfy the Markov condition. Suppose that only a proper subset **O** of variables in the graph are measured or recorded. What conditional independence relation among variables in **O** is required by the Markov condition applied to G? What graphical object represents those marginal conditional independence relations and also represents information about G? An answer to both these questions is given in Spirtes, Glymour, and Scheines (1993), based on results given in Verma and Pearl (1990). Verma and Pearl introduced the notion of the *inducing path graph* for G which contains only measured variables in G. In Spirtes, Glymour, and Scheines (1993) it is proved that the inducing path graph encodes all of the marginal conditional independence relations G entails (by the Markov condition) and includes some of the causal information represented in G.

An undirected path U between X and Y is an *inducing path* over (or given) **O** in G if and only if (i) every member of **O** on U except for the endpoints is a collider on U, and (ii) for every vertex V that is a collider on U, there is an acyclic directed path from V to X or from V to Y. In Verma and Pearl(1990) it was proved that there is an inducing path between X and Y in G over **O** if and only if X and Y are dependent conditional on every subset of $\mathbf{O}\backslash\{X,Y\}$. For variables X, Y in **O**, in the inducing path graph H for G over **O**, X <-> Y in H if and only if there is an inducing path between X and Y over **O** in G that is directed into X and also directed into Y; there is an edge X -> Y in G if and only if there is no edge X <-> Y in H, and there is an inducing path between X and Y over **O** in G that is out of X and into Y. (It is easy to show that if G is acyclic then there are no inducing paths connecting X and Y in G over **O** that are neither directed into X nor into Y.) The two kinds of edges in an inducing path graph H have a straightforward causal interpretation: A directed edge X -> Y occurs in H only if there is a directed path from X to Y in G, i.e. X is a cause of Y; a double headed edge X <-> Y occurs in H only if there is an unmeasured T that is a common cause of X and Y.

In general, it is not possible to construct a unique inducing path graph from a given distribution. However, the FCI algorithm (Spirtes, Glymour, and Scheines 1993) constructs a partially oriented inducing path graph (i.e. an inducing path graph in which some of the orientations of the edges are not determined) from a distribution

It is easy to extend the relationship between conditional dependence and inducing paths to cyclic graphs:

Theorem 3: If L is an LCT with directed graph G (cyclic or acyclic) containing a subset of variables **O**, and X and Y are in **O**, that X and Y are dependent conditional on all subsets of $\mathbf{O}\backslash\{X,Y\}$ if and only if there is an inducing path between X and Y given **O** in G.

However, there are several important differences between inducing paths in an directed acyclic graph and in a directed cyclic graph. First, it is possible that in a directed cyclic graph that there are inducing paths between X and Y that are out of X and out of Y. Second, it is possible that a directed cyclic graph over a set of variables **V** contains an inducing path between X and Y given the entire set of variables **V**, whereas this cannot happen in a directed acyclic graph. For example, in figure 2 the path consisting of the edges $<X_1,X_3>$ and $<X_4,X_3>$ is an inducing path between X_1 and X_4 given $\mathbf{V} = \{X_1, X_2, X_3, X_4\}$ that is out of X_1 and out of X_4. The existence of inducing paths given the entire set of variables in directed cyclic graphs but not in directed acyclic graphs is the reason that the absence of an edge between X and Y in a directed acyclic graph always entails that X and Y are independent conditional on some subset of variables in the graph, whereas the absence of an edge between X and Y in a directed cyclic graph does not always linearly entail some that X and Y are independent conditional on some subset of variables in the graph.

Sometimes it is not possible to determine from a given probability distribution whether it was generated by a non-recursive set of equations, or a recursive set of equations. For example, the graph in figure 3 linearly entails the same (empty) set of conditional independence relations as any complete directed acyclic graph containing just the variables X, Y, and Z.

However, there are cases where it is possible to determine from a given probability distribution whether it was generated by a non-recursive set of equations, or a recursive set of equations. The cyclic graph in figure 2 does not linearly entail the same set of conditional independence relations as any directed acyclic graph, nor (contrary to a remark in Spirtes, Glymour and Scheines, 1993) is there any directed acyclic graph with latent variables that entails the same set of conditional independence relations over $\mathbf{O}$ as the cyclic graph in figure 2 does.[4]

5. Cyclic Directed Graphs and Chain Independence Graphs

A chain graph consists of a set of variables connected by directed edges and undirected edges (Wermuth and Lauritzen 1990). It is supposed that the variables can be partitioned into blocks which are completely ordered. Variables within a block are not ordered. Variables from different blocks have the same order as the blocks they are in. Given a partial ordering of the variables of this type, a chain graph G represents a distribution P if and only if

[4]This is easily determined by application of the FCI algorithm to the set of conditional independence relations entailed by figure 2, which yields a POIPG which entails the wrong set of conditional independence relations. This can happen only when the distribution is not the marginal of any distribution represented by a directed acyclic graph.

(i) there is a directed edge from X to Y in G whenever X is before Y in the partial ordering, and X is dependent on Y given all variables except X and Y that are not after Y in the partial ordering;
(ii) there is an undirected edge between X and Y in G whenever X is not before Y and Y is not before X in the partial ordering, and X is independent of Y given all variables except X and Y that are not after Y in the partial ordering.

It is sometimes suggested (e.g. Whittaker (1990)) that in a chain independence graph a directed edge between X and Y represents X causes Y, and an undirected edge between X and Y represents a feedback process between X and Y. If that is the case, then the proper representation of the causal structure that we have depicted in figure 2 is the block chain graph in figure 6.

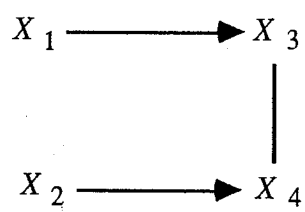


Figure 6

The chain graph in figure 6 does not entail the same set of conditional independence relations as the set of linear equations associated with figure 2. Figure 6 does not entail that X_1 and X_2 are independent given X_3 and X_4, as the cyclic graph in figure 2 and the non-recursive set of equations associated with figure 2 do. Clearly, either the set of linear equations associated with figure 2 do not represent any kind of feedback process, or this chain independence graph does not represent the kind of feedback process represented by the equations associated with figure 2. Whittaker seems to conclude that the latter is the case.

However, the use of non-recursive sets of linear equations to represent feedback is common not only in econometrics and sociology, it is also a well-established and well tested practice in engineering (where "flowgraph analysis" of circuits with amplifiers and various signal systems with feedback is common.) In the absence of any convincing argument that *all* feedback processes ought to be represented by block chain graphs, I see no reason to abandon the use of non-recursive sets of linear equations as a model of feedback.

6. Cyclic Directed Graphs and Undirected Graphs

Undirected graphs are also used to represent probability distributions. Whittaker(1990) gives the following definition (where X_k is a set of random variables):

> The *conditional independence graph* of *X* is the undirected graph $G = (K,E)$ where $K = \{1,2,\ldots,k)$ and (i,j) is not in the edge set *E* if and only if $X_i \perp\!\!\!\perp X_j | X_{k \setminus \{i,j\}}$.

A DAG G satisfies the Wermuth Condition if and only if no pair of non-adjacent vertices have a common child. (See Whittaker 1990.) No DAG that fails to satisfy the Wermuth Condition entails the same set of conditional independence relations as any undirected graph. However, there are directed *cyclic* graphs which fail to satisfy the Wermuth Condition and that linearly entail the same set of conditional independence relations as an undirected graph. For example, the cyclic graph in figure 3 and a complete undirected graph both entail no conditional independence relations. The following theorem states a necessary condition for a directed graph to linearly entail the same set of conditional independence relations as some undirected graph.

Theorem 4: A (cyclic or acyclic) directed graph *G* over **V** does not linearly entail the same set of conditional independence relations as some undirected graph *G'* if there is a pair of vertices *X* and *Y* in *G* such that *X* and *Y* have a common child *Q* and there is no inducing path between *X* and *Y* over **V** in *G*.

An undirected graph *G'* **moralizes** a directed graph *G* when *G* and *G'* contain the same vertices, and *A* and *B* are adjacent in *G'* if and only they are adjacent in *G* or they have a common child in *G*. In the case of directed acyclic graphs, the undirected graph that moralizes *G* entails a subset of the conditional independence relations entailed by *G*. The same is true of cyclic graphs.

Theorem 5: If G_M moralizes a (cyclic or acyclic) directed graph *G*, G_M entails a subset of the conditional independence relations linearly entailed by *G*.

An undirected graph is **chordal** if and only if it contains no undirected cycle of length four or more that has no edge joining two non-consecutive vertices. If an undirected graph is not chordal, then it is does not entail the same set of conditional independence relations as any DAG. However, there are non-chordal undirected graphs that entail the same set of conditional independence relations as are linearly entailed by a cyclic graph. For example, the undirected graph in figure 7 entails the same conditional independence relations as the directed graph in figure 5.

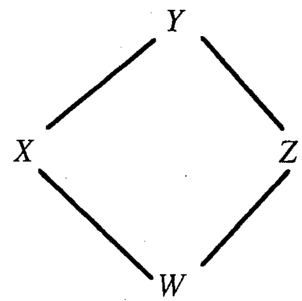


Figure 7

I conjecture that the undirected graph in figure 8 entails a set of conditional independence relations that is not equal to the set of conditional independence relations linearly entailed by any directed graph, cyclic or acyclic (because in any directed graph with the same edges the Wermuth Condition is violated.)

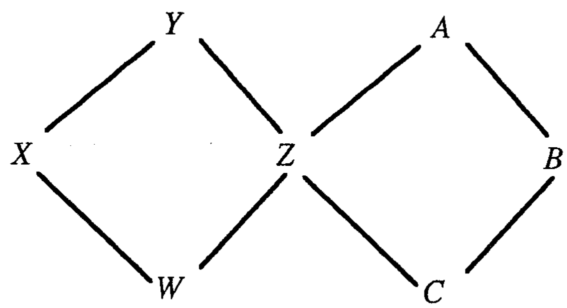


Figure 8

6. Some Open Questions

The following questions remain open.

1. Is it possible to extend the d-separation theorems to non-linear structural equation models? The assumption of linearity is used in the proofs only to show that the Jacobean of the transformation is constant. If it could be shown that the Jacobean of a wider class of transformations of the error terms factors in the correct way, the proof could be extended to non-linear structural equation models. Is it possible to extend the d-separation theorems to discrete distributions?

2. There are polynomial algorithms (Verma and Pearl 1990, Fryedenberg 1990) for determining when two arbitrary directed acyclic graphs entail the same set of conditional independence relations. Is there a polynomial algorithm for determining when two arbitrary directed graphs (cyclic or acyclic) linearly entail the same set of conditional independence relations? There are polynomial algorithms (Spirtes and Verma 1992) for determining when two arbitrary directed acyclic graphs entail the same set of conditional independence relations over a common subset of variable **O**. Is there a polynomial algorithm for determining when two arbitrary directed graphs (cyclic or acyclic) linearly entail the same set of conditional independence relations over a common subset of variables **O**?

3. There are polynomial algorithms for reliably inferring features of (sparse) directed acyclic graphs from a probability distribution when there are no latent common causes (see Spirtes and Glymour 1991, Cooper and Herskovitz 1992, and Wedelin 1993). Are there polynomial algorithms for reliably inferring features of directed graphs (cyclic or acyclic) from a probability distribution when there are no latent common causes?

There are algorithms for reliably inferring features of directed acyclic graphs from a probability distribution even when there may be latent common causes (see Spirtes, 1992 and Spirtes, Glymour and Scheines 1993). Are there reliable algorithms for inferring features of directed graphs (cyclic or acyclic) from a probability distribution even when there may be latent common causes?

Appendix

Let the set of non-error variables in an LCT L be **V**. We will denote the set of error terms for variables in subset $\mathbf{S} \subseteq \mathbf{V}$ as **Err(S)**. In an LCT, each non-error variable can be expressed as a linear function of the error variables. Because the system of equations is linear, each error variable can also be expressed as a linear function of non-error variables. For each X in **V**, ε_X (the error variable for X) is a linear function of X and the ·non-error) parents of X in the graph, that is $\varepsilon_X = g(X,\mathbf{Parents}(X))$.

By assumption

$$f(\mathbf{Err(V)}) = \prod_{\varepsilon \in \mathbf{Err(V)}} f(\varepsilon)$$

We can derive the density function for the set of variables **V** by replacing each ε_i in $f(\varepsilon_i)$ by $g(X_i,\mathbf{Parents}(X_i))$.

$$f(\mathbf{V}) = \prod_{X \in \mathbf{V}} f(g(X,\mathbf{Parents}(X)) \times J_{\mathbf{Err(V)}->\mathbf{V}}$$

where $J_{\mathbf{Err(V)}->\mathbf{V}}$ is the Jacobian of the transformation. Because the transformation is linear, the Jacobian is a constant.

When determining whether **X** and **Y** are independent given **Z** we are actually interested in a partial transformation of the error variables. Suppose that **R** is a subset of **V** such that each ancestor in **V** of a member of **R** is also in **R**. Then none of the variables in **R** is a function of any error term in **Err(V\R)**. (This follows from Mason's rule. See Heise1975.)

$$f(\mathbf{R} \cup \mathbf{Err}(\mathbf{V} \setminus \mathbf{R})) = \prod_{X \in \mathbf{R}} f(g(X,\mathbf{Parents}(X)) \times \prod_{\varepsilon \in \mathbf{Err(V\setminus R)}} f(\varepsilon) \times J_{\mathbf{Err(V)}->\mathbf{R}\cup\mathbf{Err(V\setminus R)}}$$

Lemma 1: In an LCT with directed graph G, if **V** is the set of non-error variables, and **R** is a subset of **V** such that every ancestor of a member of **R** is also in **R**, then

$$f(\mathbf{R}) = \left(\prod_{R \in \mathbf{R}} f(g(R,\mathbf{Parents}(R)) \right) \times c,$$

where c is a constant.

Proof. Let $\mathbf{S} = \mathbf{Err}(\mathbf{V} \backslash \mathbf{R})$. Let $d\mathbf{S}$ be

$$\prod_{\varepsilon \in \mathbf{S}} d\varepsilon$$

and integration over $d\mathbf{S}$ be multiple integration over the product of the $d\varepsilon$.

$$f(\mathbf{R} \cup \mathbf{Err}(\mathbf{V} \setminus \mathbf{R})) = \prod_{R \in \mathbf{R}} f(g(R, \mathbf{Parents}(R)) \times \prod_{\varepsilon \in \mathbf{S}} f(\varepsilon) \times J_{\mathbf{Err(V)->R \cup S}}$$

$$\int_{-\infty}^{\infty} f(\mathbf{R} \cup \mathbf{Err}(\mathbf{V} \setminus \mathbf{R}))\, d\mathbf{S} = \int_{-\infty}^{\infty} \prod_{R \in \mathbf{R}} f(g(R, \mathbf{Parents}(R)) \times \prod_{\varepsilon \in \mathbf{S}} f(\varepsilon) \times J_{\mathbf{Err(V)->R \cup S}}\, d\mathbf{S} =$$

$$\prod_{R \in \mathbf{R}} f(g(R, \mathbf{Parents}(R)) \times J_{\mathbf{Err(V)->R \cup S}} \times \int_{-\infty}^{\infty} \prod_{\varepsilon \in \mathbf{S}} f(\varepsilon)\, d\mathbf{S} =$$

$$\prod_{R \in \mathbf{R}} f(g(R, \mathbf{Parents}(R)) \times J_{\mathbf{Err(V)->R \cup S}}$$

because none of the variables in **R** is a function of any error term in **S**. Hence, we can write the density function for **R** as

$$(1)\ f(\mathbf{R}) = \left(\prod_{R \in \mathbf{R}} f(g(R, \mathbf{Parents}(R)) \right) \times c \therefore$$

If **Z** is any subset of **V**, and X is not in **Z** then let W be a member of **Samefactor**(X,**Y**,**Z**) just when W is not a member of **Z** and there is some factor in

$$\prod_{R \in \mathbf{Y}} f(g(R, \mathbf{Parents}(R))$$

that contains both W and X. Let **Samefactor***(X,**Y**,**Z**) be the transitive closure of **Samefactor**(X,**Y**,**Z**), i.e. W is in **Samefactor***(X,**Y**,**Z**) if and only if W is in **Samefactor**(X,**Y**,**Z**) or there is some M in **Samefactor***(X,**Y**,**Z**) such that W is in **Samefactor**(M,**Y**,**Z**). Let W be in **Samefactor***(**X**,**Y**,**Z**) if and only if W is in **Samefactor***(X,**Y**,**Z**) for some X in **X**.

We will prove that in an LCT with directed graph G, if **Y** is d-separated from **X** given **Z** then no member of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**, and if no member of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)** then it is possible to partition **Ancestors(X ∪ Y ∪ Z)** into two sets **A** and **B** that overlap only in **Z**, and such that the joint density of **Ancestors(X ∪ Y ∪ Z)** is equal to h(**A**)h'(**B**). The latter fact entails that **X** is independent of **Y** given **Z**.

For a given set of factors **T** in the product on the right hand side of (1), let **Varin(T)** be the set of variables that appear in those factors (i.e. for a single factor $f(g(R,$**Parents**$(R))$, **Varin**$(f(g(R,$**Parents**$(R)))$ is just $\{R\}$ ∪ **Parents**(R). For a given subset **S** of **R**, let **Factors(S)** be the set of factors in the product on the right hand side of (1) that contain some variable in **S**. If U is an acyclic path between variables X and Y that contains M and N, we will denote the subpath of U from M to N by $U(M,N)$.

Lemma 2: In an LCT L with directed graph G, if **X**, **Y** and **Z** are disjoint, and **X** is d-separated from **Y** given **Z** in G, no member of Y is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**.

Proof. We will prove the contrapositive. Suppose that some member Y of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**. It follows that there is a sequence of factors in (1) $t'_1,\ldots,t'_m$ such that t'_1 contains some X in **X**, t'_m contains Y, and for each pair of factors t'_i and t'_{i+1} adjacent in the sequence, there is a variable M in (**Varin**(t'_i) ∩ **Varin**(t'_{i+1}))**Z**. If there is such a sequence of factors, then there is a sequence of factors $t_1,\ldots,t_n$ with the same properties that is minimal in the sense that if any subsequence of the factors in the sequence is removed, then there is a pair of factors adjacent in the new sequence such that **Varin**(t_i) ∩ **Varin**$(t_{i+1}) = \varnothing$. Let R_i be some variable in **Varin**(t_i) ∩ **Varin**(t_{i+1}). Then corresponding to the sequence of factors $t_1,\ldots,t_m$ is a sequence of variables $X, R_1,\ldots, R_m,Y$, where none of the R_i are in **Z**. For each t_i, **Varin**(t_i) consists of a variable together with its parents. Hence for each pair of variables R_i, R_{i+1} in the sequence, either R_i is a parent of R_{i+1}, R_{i+1} is a parent of R_i, or R_i and R_{i+1} are both parents of some other variable Q appearing in t_i. For each pair of variables in the sequence satisfying only the latter condition, insert Q into the sequence between R_i and R_{i+1}. This new sequence of variables corresponds to an undirected path U in G. It is acyclic because otherwise the sequence of factors $t_1,\ldots,t_n$ is not minimal. No non-collider on the path is in **Z** because none of the R_i are in **Z**. Because each ancestor of a member of **Ancestors(X ∪ Y ∪ Z)** is a member of **Ancestors(X ∪ Y ∪ Z)**, the only variables in the equation for an error term of a variable in **Ancestors(X ∪ Y ∪ Z)** are members of **Ancestors(X ∪ Y ∪**

Z). Hence all members of **Samefactor***(**X**,**Ancestors**(**X** ∪ **Y** ∪ **Z**),**Z**) are in **Ancestors**(**X** ∪ **Y** ∪ **Z**). Each collider on the path has a descendant in either **X, Y**, or **Z**. If all of the colliders on the path have a descendant in **Z**, then *U* is a path that d-connects **X** and **Y** given **Z**. Suppose then that some collider on *U* does not have a descendant in **Z**.

Let *M* be the collider on *U* that has no descendant in **Z**, and is closest to *X* on *U*. Because *M* is a member of **Ancestors**(**X** ∪ **Y** ∪ **Z**) but has no descendant in **Z**, there is (i) either a directed path *D* from *M* to *X'* in **X** that contains no member of **Z** ∪ **Y** or (ii) a directed path from *M* to *Y'* in **Y** that contains no member of **Z** ∪ **X**. (If *D* contains members of both **X** and **Y**, then some subpath contains only members of **X** or only members of **Y**.) Suppose (i) is the case. Either *X'* is on *U* or it is not. Suppose first that *X'* is on *U*. Every collider on *U*(*Y*,*X'*) is a collider on *U* and hence has a descendant in **Z**. Every non-collider on *U*(*Y*,*X'*) with the possible exception of *X'* is a non-collider on *U*, and hence not in **Z**. By hypothesis, *X'* is not in **Z**. Hence *U*(*Y*,*X'*) d-connects **X** and **Y** given **Z**. Suppose then that *X'* is not on *U*. Let *W* be the last vertex on *D* that is also on *U*(*M*,*Y*). Let *U'* be the concatenation of *U*(*Y*,*W*) and *D*(*W*,*X'*). See figure 9 (in which for ease of illustration *X'* = *X*.) .

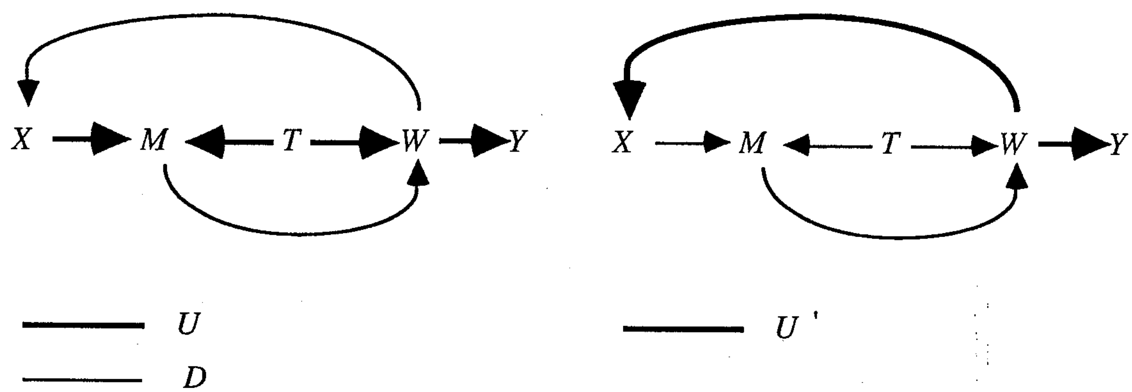


Figure 9

U' is acyclic because the only member of *U*(*Y*,*W*) on *D*(*W*,*X'*) is *W*. Every vertex on *D*(*W*,*X'*) is a non-collider on *U'* and not in **Z**. Every non-collider on *U*(*Y*,*W*) is not in **Z**, because every non-collider on *U*(*Y*,*W*) with the possible exception of *W* is a non-collider on *U*, and *W* is not in **Z**. Hence every non-collider on *U'* is not in **Z**, and *U'* has at least one fewer collider that does not have a descendant in **Z** than *U* does. This process can be repeated until a path *U''* that has no colliders that do not have a descendant in **Z** is formed. *U''* d-connects some *X* in **X** and *Y* in **Y** given **Z**.

Suppose (ii), i.e. there is a directed path D from M to some member Y' of **Y** that contains no member of $\mathbf{Z} \cup \mathbf{X}$. Let M be the collider on U that has no descendant in **Z**, and is closest to X on U. Either Y' is on $U(X,M)$ or it is not. If Y' is on $U(X,M)$ then $U(X,Y')$ d-connects X and Y' given **Z**, and hence it d-connects X and **Y** given **Z**. If Y' is not on $U(X,M)$ then let W be the last vertex on D that is also on $U(X,M)$. Let U' be the concatention of $U(X,W)$ and $D(W,Y')$. U' is acyclic because $U(X,W)$ and $D(W, Y')$ intersect only at W. No vertex on $D(W,Y')$ is a collider on U' or a member of **Z**. Every vertex that is not a collider on $U(X,W)$ is not a member of **Z** and is not a collider on U'. Every collider on $U(X,W)$ with the possible exception of W is a collider on U' and has a descendant in **Z**. W is not a collider on U' and is not a member of **Z**. U' d-connects X and Y' given **Z**, and hence it d-connects **X** and **Y** given **Z** ∴

Lemma 3: In an LCT with graph G, if **Y** is not in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)** then there is a partition of the factors in the product on the right hand side of (1) into two sets **A** and **B** such that every member of **X** is in **Varin(A)** and no member of **X** is in **Varin(B)**, and every member of **Y** is in **Varin(B)** and no member of **Y** is in **Varin(A)**, and **Varin(A) ∩ Varin(B) ⊆ Z**.

Proof. Let **A** = **Factor(Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z))** and **B** equal the set of other factors in

$$\prod_{R \in \mathbf{Ancestors}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z})} f(g(R \quad ,\mathbf{Parents}(R))$$

First we will show that **Varin(A) ∩ Varin(B) ⊆ Z**. Suppose that there is some variable W not in **Z** that is in **Varin(A)** and **Varin(B)**. By definition, if W is in **Varin(A)**, then it appears in the same factor as some variable V in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)** . By definition, then, W is also in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**. But then each factor containing W is in **Varin(A)**, and not in **Varin(B)**. Hence W is not in **Varin(B)**. This is a contradiction.

By hypothesis no member of **Y** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)**, so no member of **Y** is in **Varin(A)**. It follows that every member of **Y** is in **Varin(B)**.

Every member of **X** is in **Samefactor*(X,Ancestors(X ∪ Y ∪ Z),Z)** by definition, so every member of **X** is in **Varin(A)**. Because no member of **X** is in **Z**, no member of **X** is in **Varin(B)**. ∴

Lemma 4: In an LCT L with graph G, if **X**, **Y**, and **Z** are disjoint sets of variables and there is a partition of the factors in the product on the right hand side of (1) into two sets **A** and **B** such that every member of **X** is in **Varin**(**A**) and no member of **X** is in **Varin**(**B**) and every member of **Y** is in **Varin**(**B**) and no member of **Y** is in **Varin**(**A**), and **Varin**(**A**) $\cap$ **Varin**(**B**) $\subseteq$ **Z**, then L linearly entails that **X** is independent of **Y** given **Z**.

Proof. We can rewrite the product on the right hand side of (1) as

$$f(\mathbf{Ancestors}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z})) = \prod_{t \in \mathbf{A}} t \times \prod_{t \in \mathbf{B}} t$$

where each t is of the form f(g(V,**Parents**(V)). We can then rewrite

$$\prod_{t \in \mathbf{A}} t$$

as a function h(**Varin**(**A**)), and

$$\prod_{t \in \mathbf{B}} t$$

as a function h'(**Varin**(**B**)). Hence f(**Ancestors**(**X** $\cup$ **Y** $\cup$ **Z**)) = h(**Varin**(**A**)) $\times h'$(**Varin**(**B**)), where **Varin**(**A**) $\cap$ **Varin**(**B**) $\subseteq$ **Z**. It follows that **Varin**(**A**)**Z** is independent of **Varin**(**B**)**Z** given **Z**. Because **X** is included in **Varin**(**A**)**Z** and **Y** is included in **Varin**(**B**)**Z**, **X** is independent of **Y** given **Z**. $\therefore$

Lemma 5: In an LCT L with directed graph G, if **X**, **Y**, and **Z** are disjoint sets of variables, and **X** is d-separated from **Y** given **Z**, then L linearly entails that **X** is independent of **Y** given **Z**.

Proof. This follows from lemmas 1, 2, 3, and 4. $\therefore$

Lemma 6: In an LCT with directed graph G, if **X**, **Y**, and **Z** are disjoint sets of variables, and **X** is d-connected to **Y** given **Z** in G, then **X** is d-connected to **Y** given **Z** in an acyclic directed subgraph of G.

Proof. Suppose **X** is d-connected to **Y** given **Z**. Then for some X in **X** and Y in **Y**, X is d-connected go Y given **Z** by path U in G. Each collider C_i on U has a descendant in **Z**. Let D_i be a shortest directed path from C_i to a member of **Z**; we will call D_i the directed acyclic path associated with C_i. U is acyclic by definition, and so are each of the D_i, so if there are any directed cycles in G they contain edges from more than one of these paths.

We will show that **X** is d-connected to **Y** given **Z** in an acyclic directed subgraph by modifying the path U and the associated directed paths from colliders on U to members of **Z** in such a way that no subset of the edges on the modified paths form a directed cycle.

First we will form a path U' that d-connects X and Y given **Z**, such that for each collider C_i' on U' there is a directed path from C_i' to a member of **Z** that does not contain an edge into a vertex on U'. Ultimately this will guarantee that no edge in U' is in a directed cycle in the sugraph that we will form. See figure 10.

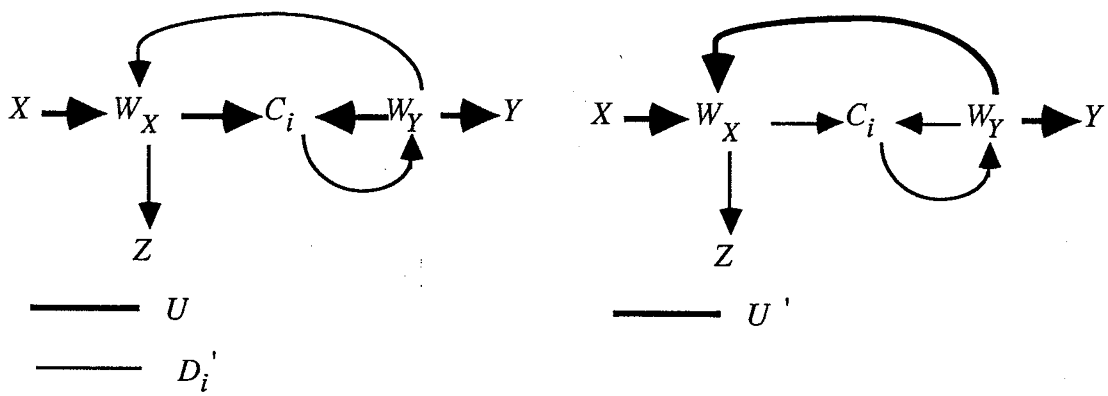
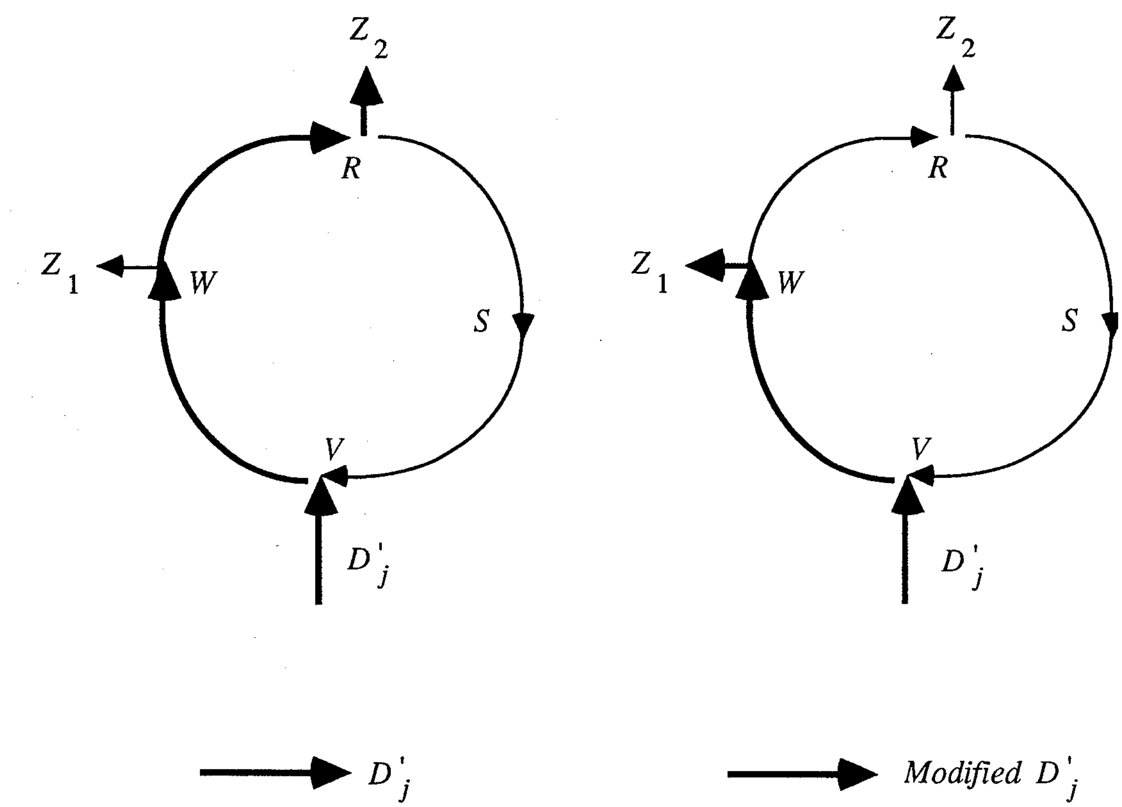


Figure 10

Form the path U' in the following way. If D_i intersects U at a vertex other than C_i then let W_X be the vertex on D_i and U that is closest to X on U, and W_Y be the vertex on D_i and U that is closest to Y on U. Suppose without loss of generality that W_X is after W_Y on D_i. Let U' be the concatenation of $U(X,W_X)$, $D_i(W_Y,W_X)$, and $U(W_Y,Y)$. U' is acyclic because there is no point of intersection of D_i and U between X and W_X, and no point of intersection of D_i and U between W_Y and Y. Every non-collider on $U'(W_Y,Y)$ with the possible exception of W_Y is a non-collider on U, and hence not in **Z**. W_Y is not a collider on U' because the edge on $D_i(W_Y,W_X)$ containing W_Y is out of W_Y. W_Y is not in **Z** because it is on D_i and not an endpoint of D_i. Every non-collider on $U'(X,W_X)$ with the possible exception of W_X is a non-collider on U, and hence not in **Z**. If W_X is not a collider on U', then W_X is not a collider on U, and hence not in **Z**. If W_X is in **Z**, then W_X is a collider on U, and hence $U(X,W_X)$ is into W_X. D_i is into W_X, and hence W_X is a collider on U'; in that case let $D_i(W_X,Z)$ be the associated directed path of W_X. It follows that U' d-connects X and Y given $\mathbf{Z}$, but C_i is not a collider on U'. The only vertex on U' that may be a collider on U' but not on U is W_X, but there is a directed path from W_X to a member of $\mathbf{Z}$ that does not intersect U' except at W_X. Hence the number of colliders on U' that are not in $\mathbf{Z}$, but whose associated directed paths are into a vertex on U' is at least one less than the number of colliders on U that are not in $\mathbf{Z}$, but whose associated directed path are into a vertex on U. Repeat this process until there are no colliders on U' that are not in $\mathbf{Z}$ but whose associated directed paths are into a vertex on U'.

For each C'_i on U' let D'_i be the directed acyclic path associated with C'_i . There is no directed cycle in G containing edges from U' and some of the D'_i because no D'_i contains an edge into a vertex on U'. If there is a directed cycle in G that contains only edges from some combination of the associated directed paths then modify the associated directed paths in the following way. Suppose C is a cyclic directed path such that each edge occurs on some associated directed path and that D'_i intersects C. Let W be the vertex on D'_i and on C that is closest to the sink Z of D'_i. We will show how to remove the edge out of W on C from every directed path associated with a collider on U'. Consider an arbitrary D'_j that contains the edge out of W on C, and suppose the first vertex on D'_j that intersects C is V. Let the new D'_j be the concatenation of $D'_j(C'_j,V)$, $C(V,W)$, and $D'_i(W,Z)$. Every cyclic directed path between any pair of vertices A and B contains an acyclic directed subpath between A and B. So if D'_j is a cyclic path remove the cycles and make D'_j acyclic. D'_j is an acyclic path that does not contain the edge on C out of W. D'_j does not have an edge into U' because it contains only edges that were already in some associated directed path and none of the associated directed paths contained any edges into U'. Modify each associated path that contains the edge out of W on C in this way. See figure 11.



Original D'_j **Contains Edge Out of** W **Modified** D'_j **Contains No Edge Out of** W

Figure 11

After this modification, no associated directed path contains an edge out of W, and hence cycle C can no longer be formed out of edges occurring on associated directed paths. Hence the number of cycles that can be formed from the directed acyclic paths associated with colliders on U' has been reduced by at least one. Repeat this process until there are no cycles that can be formed from the directed acyclic paths associated with collider on U'. Let G' be the subgraph of G consisting of the union of the vertices occurring on U', some D'_i, **X**, **Y**, and **Z**, and containing only edges that occur on U', or some associated paths D'_i. G' is an acylic subgraph of G in which **X** is d-connected to **Y** given **Z**. ∴

Lemma 7: In an LCT L with directed graph G, if **X**, **Y**, and **Z** are disjoint sets of variables, and **X** is d-connected to **Y** given **Z**, then L does not linearly entail that **X** is independent of **Y** given **Z**.

Proof. By lemma 6, if in an LCT with directed graph G, **X** is d-connected to **Y** given **Z**, then there is an acyclic subgraph G' of G in which **X** is d-connected to **Y** given **Z**. In Spirtes et. al. (1993) it was proved that if **X** is d-connected to **Y** given **Z** in an LCT with an acyclic directed graph, then there are values of the parameters for which **X** is not independent of **Y** given **Z**. If the coefficients of edges that are in G but not in G' are set to zero, and the coefficients of edges that are in G' are set to values that make **X** dependent on **Y** given **Z**, then the LCT has been parameterized in such a way that **X** is dependent on **Y** given **Z**. ∴

Theorem 1: In an LCT L with (cyclic or acyclic) directed graph G containing disjoint sets of variables **X**, **Y** and **Z**, **X** is d-separated from **Y** given **Z** if and if only L linearly entails that **X** is independent of **Y** given **Z**.
Proof. This follows from lemmas 5 and 7. ∴

Theorem 2: In an LCT L with (cyclic or acyclic) directed graph G containing X, Y and **Z**, where $X \neq Y$ and **Z** does not contain X or Y, X is d-separated from Y given **Z** if and only L linearly entails that $\rho_{XY.\mathbf{Z}} = 0$.
Proof. (This proof for cyclic or acyclic graphs is based on the proof for acyclic graphs in Verma and Pearl 1990.) Suppose in an LCT L with directed graph G,X is d-separated from Y given **Z** in G. Let L' be an LCT with the same directed graph G and that is the same as L except that the exogenous variables are normally distributed with the same variances as the corresponding variables in L. By theorem 1, L' linearly entails that X is independent of Y given **Z**. Hence for all values of the linear coefficients and all joint normal distributions over the exogenous variables in which the exogenous variables have positive variance, $\rho_{XY.\mathbf{Z}} = 0$. Because the value of a partial correlation in an LCT depends only on the values of the linear coefficients and the variances of the exogenous variables, L' linearly entails $\rho_{XY.\mathbf{Z}} = 0$, and hence L linearly entails that $\rho_{XY.\mathbf{Z}} = 0$.

Suppose that L does not linearly entail that $\rho_{XY.\mathbf{Z}} = 0$. Then there is an LCT L' with graph G such that $\rho_{XY.\mathbf{Z}} \neq 0$. Let L'' be an LCT with the same directed graph G and that is the same as L' except that the exogenous variables are normally distributed with the same variances as the corresponding variables in L'. In L'', $\rho_{XY.\mathbf{Z}} \neq 0$, because $\rho_{XY.\mathbf{Z}} \neq 0$ in L'. Hence in L'', X is not independent of Y given **Z**. It follows that X is not d-separated from Y given **Z** in G. ∴

Theorem 3: If L is an LCT with (cyclic or acyclic) directed graph G containing a subset of variables **O**, and X and Y are in **O**, then X and Y are dependent conditional on all subsets of $\mathbf{O} \setminus \{X,Y\}$ if and only if there is an inducing path between X and Y given **O** in G.

Proof. Suppose first that X and Y are dependent conditional on every subset of **O**. Then X and Y are d-connected given (**Ancestors**($A \cup B$) $\cap$ **O**)\{A,B}. Hence there is a path U that d-connects A and B given (**Ancestors**($A \cup B$) $\cap$ **O**)\{A,B}. Every collider on U is an ancestor of a member of (**Ancestors**($A \cup B$) $\cap$ **O**)\{A,B}, and hence an ancestor of A or B. Every vertex on U is an ancestor of either A or B or an ancestor of a collider on U, and hence every vertex on U is an ancestor of A or B. If U d-connects A and B given (**Ancestors**($A \cup B$) $\cap$ **O**)\{A,B}, then every member of (**Ancestors**($A \cup B$) $\cap$ **O**)\{A,B} that is on U, except for the endpoints, is a collider. Since every vertex on U is in **Ancestors**($A \cup B$), every member of **O** that is on U, except for the endpoints, is a collider. Hence U is an inducing path between A and B given **O**.

Next suppose that there is an inducing path U between X and Y given **O** in G. For each arbitrary subset **Z** of **O** we will construct a path that d-connects X and Y given **Z**. If every collider on U is an ancestor of a member of **Z**, then U d-connects X and Y given **Z**.

Suppose then that some vertex on U does not have a descendant in **Z**, and let W be the closest such vertex to X on U. Because U is an inducing path between X and Y given **O**, W is an ancestor of either X or Y. Suppose first that it is an ancestor of X. Then there is a directed path D from W to X, and because W is not an ancestor of a member of **Z**, D does not contain any member of **Z**. Let Q be the point of intersection of D and U closest to the sink of D. Let U' be the concatenation of $U(Y,Q)$ and $D(Q,X)$. U' is acyclic because the only vertex on $U(Y,Q)$ that is also on D is Q. No vertex on $D(Q,X)$ is a collider on U' or a member of **Z**. Every non-collider on $U(Y,Q)$ is not in **Z** because U is an inducing path between X and Y given **O**, and **Z** is a subset of **O**. U' contains one fewer collider that does not have a descendant in **Z** than U does.

If W is an ancestor of Y, then a path U' that contains one fewer collider that does not have a descendant in **Z** than U does can be formed in an analogous manner.

Repeat this process until a path U'' is formed that contains no colliders that do not have a descendant in **Z**. U'' d-connects X and Y given **Z**. $\therefore$

Theorem 4: A (cyclic or acyclic) directed graph G over **V** does not linearly entail the same set of conditional independence relations as some undirected graph G' if there is a pair of vertices X and Y in G such that X and Y have a common child Q and there is no inducing path between X and Y over **V** in G.

Proof. Suppose that a directed graph G contains a pair of vertices X and Y in G such that X and Y have a common child Q and there is no inducing path between X and Y given **V**. Because X is adjacent to Q in G, X is d-connected to Q given every subset of $\mathbf{V}\backslash\{X,Q\}$. It follows that in every undirected graph that entails the same set of conditional independence relations as G, X is adjacent to Q. Similarly, in every undirected graph that entails the same set of conditional independence relations as G, Y is adjacent to Q. Because there are edges from X to Q and from Y to Q, X and Y are d-connected given $\mathbf{V}\backslash\{X,Y\}$. Hence there is an edge between X and Y in every undirected graph G' that entails the same set of conditional independence relations as G. But then G' does not entail that X and Y are independent given any subset of **V**. However, because there is no inducing path between X and Y given **V** in G, G does entail that X and Y are independent given some subset of **V**. This is a contradiction. ∴

In an undirected graph G, A and B are **separated** given **Z** if every undirected path between A and B contains a member of **Z**. If A and B are not separated given **Z**, they are **connected** given **Z**. If G is a conditional independence graph of a distribution P, then G entails that A and B are independent given **Z** if and only if A and B are separated given **Z**.

Theorem 5: If G_M moralizes a (cyclic or acyclic) directed graph G, G_M entails a subset of the conditional independence relations linearly entailed by G.

Proof. For each undirected path U in G, let U' be the corresponding undirected path in G_M. We will show that if X and Y are d-connected given **Z** in G, then they are connected given **Z** in G_M. Suppose that X and Y are d-connected given **Z** by a path U in G. If U does not contain any member of **Z** then the corresponding path U' in G_M connects X and Y given **Z**. Suppose then that U does contain members of **Z**. Because U d-connects X and Y given **Z**, each such member of **Z** is a collider on U. Let Z be an arbitrary member of **Z** on U, and A and B be the vertices adjacent to **Z** on U. By hypothesis, there is an edge between A and B in G_M. A and B are not in **Z** because they are not colliders on U. If the subpath of U' that contains the edges between A and Z, and Z and B is replaced by the edge between A and B, then U' now contains one fewer member of **Z**. Repeat this process until U' contains no member of **Z**. U' now connects X and Y given **Z** in G_M. ∴

Bibliography

Cooper, G. and Herskovitz, E. (1992). A Bayesian method for the induction of probabilistic networks from data. Machine Learning.

Frydenberg, M. (1990). The chain graph Markov property. Scand. J. Statist, 17, pp. 333-353.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. Econometrica, 11, pp. 1-12.

Heise, D. (1975). Causal Analysis. Wiley, New York.

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

Lauritzen, S., Dawid, A., Larsen, B., and Leimer, H. (1990). Independence properties of directed Markov fields. Networks, 20, pp. 491-505.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan and Kaufman, San Mateo.

Spirtes, P., Glymour, C., and Scheines, R. (1991). An algorithm for fast recovery of sparse causal graphs. Social Science Computer Review, vol. 9, 1991, pp. 62-72.

Spirtes, P., Glymour, C., and Scheines, R. (1993). Causation, Prediction, and Search. Springer-Verlag Lecture Notes in Statistics 81. New York, N.Y.

Spirtes, P. (1992). Building causal graphs from statistical data in the presence of latent variables", forthcoming in Proceedings of the IX International Congress on Logic, Methodology, and the Philosophy of Science, B. Skyrms, ed., Uppsala, Sweden, 1991.

Spirtes, P., and Verma, T. (1992) Equivalence of Causal Models with Latent Variables. Philosophy, Methodology and Logic Technical Report 33, Carnegie-Mellon University.

Verma, T. and Pearl, J. (1990). Equivalence and Synthesis of Causal Models. Proc. Sixth Conference on Uncertainty in AI. Association for Uncertainty in AI, Inc., Mountain View, CA.

Wedelin, D. (1993) Discovering causal structure from data, preprint, Department of Computer Science, Chalmers University of Technology, Sweden.

Wermuth, N. and Lauritzen, S. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. J. Roy. Statist. Soc. Ser. B 52, 21-50.

Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics. Wiley, New York.