Effects of an unobserved confounder on a system with an intermediate outcome

Wermuth, Nanny

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

**Effects of an unobserved confounder
on a system with an intermediate outcome**
Nanny Wermuth

ZUMA
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim
Telefon: 0621-1246-155
Telefax: 0621-1246-100
E-Mail: wermuth@zuma-mannheim.de

Effects of an unobserved confounder on a system with an intermediate outcome *

Nanny Wermuth

Center of Survey Research (ZUMA), Mannheim, Germany

wermuth@zuma-mannheim.de

Abstract. In the theory of graphical Markov models in which relations between many variables are simplified via conditional independencies a special role is played by directed acyclic graphs. They can be used to represent statistical models in which data are generated in a stepwise fashion. Responses and intermediate variables may be event histories.

We discuss such a system with sequentially administered treatments and a confounder, that is a variable which affects both the final outcome and one of its explanatory variables. The effect of not observing the confounder is to obtain the final and an intermediate outcome as joint responses and leads to the important observation by Robins and Wasserman (1997) that any univariate conditional distribution for the final outcome will be inappropriate for analysis no matter whether the intermediate outcome is conditioned on or not.

It means in particular that the independence structure of the observed variables can no longer be fully described by a directed acyclic graph, that criteria for reading indepencencies off graphs have to be modified and that joint instead of univariate regression models are needed.

These modifications resolve directly the puzzling situation which has been discussed by the above authors for randomized clinical trials as a case in which a true hypothesis of no treatment effect is always falsely rejected. Joint response models provide an alternative route for avoiding this unpleasant situation.

Keywords: Directed acyclic graphs, conditional independence, conditioning, confounder, generating process, intermediate variable, joint response models, marginalizing, summary graph, univariate recursive regressions

1. Introduction

1.1 Generating processes and directed acyclic graphs

Sequences of univariate linear regression models have been introduced under the name of path analysis by geneticist Sewall Wright at the beginning of this century. He used them to describe hypotheses about how data might have been generated and to evaluate these hypothesis in the light of observations. His main goal was to gain insight into genetic processes. He attempted to 'trace paths' of development.

To a large extent he used directed graphs to represent these processes. Nodes indicate variables. Arrows denote direct dependencies which are strong enough to be of substantive interest. A path is an ordered set of distinct nodes having an edge present in the graph for each consecutive pair. An ordering of the variables results from the substantive context and often involves time. Of special interest are stepwise processes, in which a full set of data is generated from knowledge of only the direct dependencies for each of the ordered variables taken one at a time. Such graphs are fully directed because there are no joint responses and they are acyclic because no variable is taken to be explanatory for itself, i.e. it is impossible to start from any node, follow the direction of arrows and return to the same node.

Directed acyclic graphs are also mere mathematical objects used to characterize independence structures in probability distributions. An independence structure is a set of independence statements sufficient to capture all independencies that the joint distribution is to satisfy. A number of different sets of independence statements may describe the same structure because, typically, sequences of such statements lead to further independencies being satisfied as well. Accordingly, a number of different generating processes may give rise to the same independence structure.

If a full ordering of the variables is specified for a given directed acyclic graph then all edges, present or absent, have a precise conditioning set attached to them, i.e. to node i all nodes with higher indices are its potential ancestors. But, any edge present remains compatible with conditional independence of the corresponding variable pair. This differs from the meaning of an arrow present in the graph of a generating process, i.e. in a graph representing a substantive research hypothesis (Wermuth and Lauritzen,

1990).

It is therefore helpful to distinguish between the two types of graphs in the way illustrated in Figure 1. The graph with boxes describes the hypothesis of a stepwise generating process. As mentioned before, in it each arrow present corresponds to a nonvanishing conditional dependence of substantive interest and each edge absent in the graph has a specific conditional independence statement attached to it.
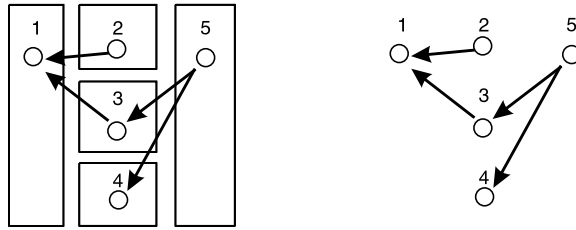

Figure 1: A generating graph representing a system of univariate recursive regressions (left) and the underlying directed acyclic graph (right)

According to the left graph of Figure 1 the joint distribution is generated by starting with the distribution of variable Y_5 and generating separately, the conditional distributions of Y_3 given Y_5 and of Y_4 given Y_5. The conditional distribution of Y_1 given Y_3 and the independently generated distribution of Y_2 completes the stepwise process. We use the terms univariate conditional distributions and univariate regressions exchangeably.

The graph obtained by deleting the boxes is directed and acyclic and is called the graph underlying the generating process. It captures the independence structure of the generating process. However, taken on its own without a prespecified ordering of the variables, it does not attach a unique conditioning set to each arrow present and it does not imply some strictly nonvanishing dependence for arrows present in the graph.

To illustrate the latter distinction in more detail, we take two special distributions corresponding to Figure 1. If each variable in the generating process represented by the left of Figure 1 is a Gaussian variable of mean zero then the regression equations are linear and have independent residuals between equations. They can be written in terms of conditional expectations as:

$$E(Y_1|Y_2, \ldots, Y_5) = \beta_{12.3}Y_2 + \beta_{13.2}Y_3$$

$$E(Y_2|Y_3, \ldots, Y_5) = 0, \quad E(Y_3|Y_4, Y_5) = \beta_{35}Y_5, \quad E(Y_4|Y_5) = \beta_{45}Y_5 \quad E(Y_5) = 0,$$

where these equations describe a process to generate a joint Gaussian distribution satisfying some independencies, described in more detail below. The generating graph in the left of Figure 1 indicates that each of the four regression coefficients corresponding to the four arrows present in the left graph $(\beta_{12.3}, \beta_{13.2}, \beta_{35}, \beta_{45})$ is strictly nonzero. By contrast, the directed acyclic graph in the right of Figure 1 is compatible with null-values of these four coefficients. In particular, it is also compatible with mutual independence of all five variables.

If instead each of the variables in Figure 1 is binary and each conditional distribution is logistic and if for variable 1 as response a two-factor interaction term is included then the joint distribution generated is a log-linear model with only independencies as restrictions. The regression coefficients are log-odds ratios. The independence statements satified by the joint distribution coincide with those in a Gaussian distribution generated over the same graph.

1.2 Independence structures

The defining independence structure can be read off either graph in Figure 1 as follows: each response i is independent of other potentially explanatory variables j given its direct influences. Written in terms of nodes the interpretation of any missing arrow from j to i is

$$i \perp\!\!\!\perp j \mid (\text{parents of } i),$$

where parents of a node i are the nodes from which an arrow points to i.

Descendants of a node i are all those nodes which can be reached from it by following the direction of the arrows. If i is a descendant of j then it is equivalent to say that j is a (proper) ancestor of i. Parents denote direct influences, i.e. directly explanatory variables, all other proper ancestors denote indirect influences, i.e. variables which are indirectly explanatory. In the case of a generating process, each parent j of i in the graph denotes, in addition, a variable which is of substantive importance in the process and for predicting Y_i given the remaining direct influences.

From the boxed graph of a generating process, the defining independence structure can also be read off in terms of the past, i.e. in terms of all nodes listed in boxes to the right of the node(s) considered:

$$i \perp\!\!\!\perp (\text{its past excluding parents of } i) \mid (\text{parents of } i)$$

and mutual conditional independence of variables shown in stacked boxes, i.e.

$$(i_1 \perp\!\!\!\perp i_2 \ldots \perp\!\!\!\perp i_s) \mid (\text{their past}).$$

For Figure 1 this gives with

$$Y_1 \perp\!\!\!\perp (Y_4, Y_5) \mid (Y_2, Y_3), \quad (Y_2 \perp\!\!\!\perp Y_3 \perp\!\!\!\perp Y_4) \mid Y_5 \quad \text{and} \quad Y_2 \perp\!\!\!\perp Y_5,$$

a slightly more compact way of writing the independence structure then by using the definition in terms of each response, taken one at a time.

An example for an independence statement implied by the graph of Figure 1 is $Y_2 \perp\!\!\!\perp (Y_3, Y_4, Y_5)$. This and other independencies may be derived by combining probability statements (see, for instance, Dawid, 1979), by using the separation criterion for directed acyclic graphs (Pearl, 1988), with the help of a matrix algorithm (Wermuth and Cox, 2000) or by using the generalized version of Pearl's path criterion stated below in Section 2.3, which applies to joint and univariate response models that may result by marginalizing over some nodes in a directed acyclic graph.

1.3 Some historical remarks

Wright (1921) used the notion of conditional independence only implicitly. Missing arrows in his fully directed graphs without directed cycles point at variable pairs for which the observed marginal correlations may deviate from the ones implied by the generating process. Thus, missing arrows, which correspond to conditional independences, are used to evaluate the generating process in the light of observations. It was shown much later (Wermuth, 1980) under which conditions the sum of differences between these observed and implied correlations defines a component of the likelihood-ratio-statistic, the general form of which was derived by Wilks (1938) to test the goodness-of-fit of a model.

By contrast, Andrej Andrejwich Markov used the notion of conditional independence explicitly (1912) to simplify multivariable structures. Markov chain models can be viewed as distributions defined over a special type of directed acyclic graph: over a graph which consists of a single direction-preserving path, with A_j depending directly only on A_{j+1} for $i = 1 \ldots, p - 1$, say. For example for five discrete variables A_j such a joint distribution is given by

$$\Pr(A_1, A_2, \ldots, A_5) = \Pr(A_1|A_2)\Pr(A_2 \mid A_3)\Pr(A_3|A_4)\Pr(A_4|A_5)\Pr(A_5),$$

i.e. at each stage only the most recent past is 'remembered' in the system.

The joint distribution defined over the graph of Figure 1 can be written as:

$$\Pr(A_1, A_2, \ldots, A_5) = \Pr(A_1|A_2, A_3)\Pr(A_2)\Pr(A_3|A_5)\Pr(A_4|A_5)\Pr(A_5).$$

If all variables are binary then each of the conditional distribution could be, for instance, logistic or probit regressions. Since, in general, some of the responses may be discrete, others continuous it has become a convention to use in the graphs dots for discrete and empty circles for continuous variables.

An essential extension of Sewell Wright's method of tracing paths became possible with Judea Pearl's criterion for reading all independencies directly off the graph for distributions of any type defined over directed acyclic graphs. Conditions under which a lack of independence can be interpreted positively as the presence of an association have been given for quasi-linear systems (Wermuth and Cox, 1998). However more work is needed for general types of distributions generated over graphs to better understand the type of the resulting association models.

In 1943 Trygve Haavelmo noted an important limitation of univariate linear recursive equations. His result motivated the development of joint response models with cyclic dependences. He showed that two linear equations with each response having a direct dependence on the other response and – at the same time – independent residuals between equations are incompatible with a definition of equation parameters in terms of conditional expectations. For a simplified version of his argument see Cox and Wermuth (1993). This has led to the development of simultaneous equation models in econometrics and to linear structural relation models in psychometrics. A quite

different approach to joint response models for discrete and continuous variables has led to the graphical Markov models in which joint distributions are formulated which satisfy independence restrictions. Systems of linear recursive regressions, such as those described for Figure 1, represent an important subclass within either formulation of two types of model classes (see e.g. Wermuth, 1992; Koster 1999)

1.4 Objectives

The main aims in the present paper are twofold. We first derive the independence graphs that result from marginalizing over nodes in directed acyclic graphs, classify the types of models which can result, and give the corresponding separation theorem to read directly off the graph all independencies implied for the variables remaining after marginalization. These results do not depend on the type of variables or distributions involved.

Next we apply the results to a problem described by Robins and Wasserman (1997) for randomized clinical trials in which treatments are administered sequentially and there is no treatment effect given information on the health status of the patient prior to entering the trial, see Figure 12 below. The authors show that if the health status is unobserved then a naive use of regression models and of standard parametrizations can lead to rejecting a true hypothesis of no treatment effect with probability approaching one as the sample size increases. We show here, in particular, that such a naive use of regression models can be avoided by deriving the proper independence graph for the observed variables. We also point at alternative standard parametrizations which do not share the deficiencies of conditional Gaussian distributions noted by the authors in the context of their example.

2. Marginalizing over nodes in directed acyclic graphs

We take a joint distribution generated over a directed acyclic graph, G_{dag}^V, having vertex set, i.e. node set, V and derive the independence graph implied for the distribution of Y_S, where S is the selected subset of nodes remaining after marginalizing over a subset of nodes M of V, i.e. $S = V \setminus M$. The resulting graph is called the summary graph, G_{sum}^S, for the distribution of Y_S. It may contain three types of edge (Wermuth and

Cox, 2000):

$$\leftarrow \quad \text{---} \quad \underleftarrow{\text{---}}.$$
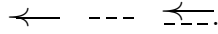
Independence structures in graphs of this type have also been studied by Koster (1999), Spirtes et al (1998), Richardson (1999) and Wermuth et al. (2000).

For the edge with two components there are two different paths between the node pair. The graph is without directed cycles, i.e. it is impossible to start from a node, follow a direction-preserving path of arrows and return to the starting node. However, partially directed cycles may occur, the smallest configuration of this type is the edge with two components.

In Section 2.1 we show how to modify such a graph by marginalizing over some of its nodes and give some simple examples. In Section 2.2 we classify the types of models which can be derived in this way for Gaussian distributions and we give a criterion to read off all independencies specified with a summary graph.

2.1 A summary graph derived by marginalizing

In pictures of graphs we point at the nodes to be marginalized over by a double crossing of the nodes, $\emptyset$. We indicate that edges have been inserted due to marginalizing by blacked in crossings such as in Figure 2, second row. Since it is important to be able to do marginalizing in any order of the nodes and obtain the same summary graph we give directly the effects of marginalizing over nodes in a summary graph in the table below. Marginalizing over any single node in a directed acyclic graph is a special case.

To marginalize over nodes $m = \{\emptyset\}$ in G_{sum}^S: an edge i, j is inserted within $s = S \setminus m$ for a common neigbour node t which is an element of $\{\emptyset\}$ as

	$t \leftarrow j$	$t \text{ --- } j$	$t \underleftarrow{\text{---}} j$	$t \rightarrow j$
$i \rightarrow t$	no	no	no	$\rightarrow$
$i \text{ --- } t$	no	no	no	---
$i \underrightarrow{\text{---}} t$	no	no	no	$\underrightarrow{\text{---}}$
$i \leftarrow t$	$\leftarrow$	---	$\underleftarrow{\text{---}}$	---

After edges have been inserted accordingly for each node in the marginalizing set the nodes and edges of $\{\emptyset\}$ are deleted.

8

Figure 2 shows the effects of marginalizing in three-node directed acyclic graphs over the common neighbour node. The common neighbour node is a transition node in the left graph of Figure 2, a source node in the middle and a collision or sink node in the right graph of Figure 2. For the transition node the inserted edge is an arrow, i.e. marginalizing corresponds to a shortening of the direction-preserving path. For the common source node the inserted edge is undirected. Finally, by marginalizing over a common sink node, i.e by ignoring the common response, no edge is inserted.
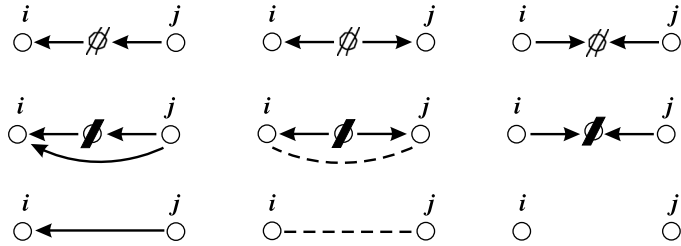

Figure 2: Effects of marginalizing over the common neighbour node which is a transition node (left), a source node (middle), a sink or collision node (right). Top: starting graphs; middle: edges inserted due to marginalizing; bottom: summary graphs of the bivariate distributions

Figure 3 shows the summary graph obtained by marginalizing over the common collision and a common source node in the graph of Figure 1, while Figure 4 shows the summary graph resulting by marginalizing over a single node which is a common source for each pair of four other nodes. In both cases the resulting graphs are covariance graphs, i.e. an edge represents the marginal pairwise association of a variable pair (Cox and Wermuth, 1993, 1996, 2000).
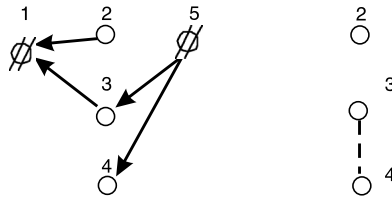

Figure 3: Marginalizing over a common sink and a common source node (left); the summary graph is an incomplete covariance graph (right)
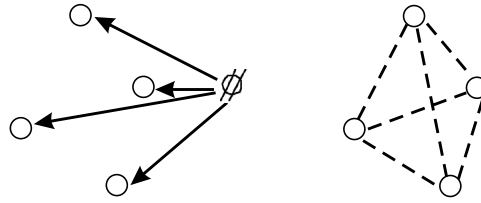
Figure 4: Marginalizing over a source node common to all other variable pairs (left); the summary graph is an complete covariance graph (right)

Figure 5 shows the summary graph resulting by marginalizing over a transition node and over a node which is not a common neighbour in Figure 1. The resulting graph is then again directed and acyclic.
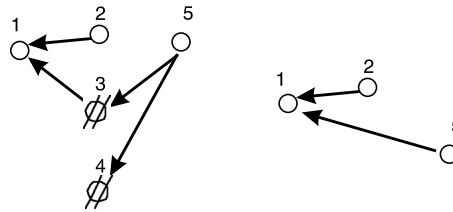

Figure 5: Marginalizing over a transition node and a response node (left); the summary graph is directed acyclic (right)

2.2 Types of model generated by marginalizing

Four different model subclasses can be identified which arise by marginalizing over nodes in a directed acyclic graph and which are - in the case of a joint Gaussian distribution - also within the class of linear structural relation models:

– univariate recursive systems with independent residuals,

– multivariate regression chains, including seemingly unrelated regressions,

– covariance graph models,

– univariate recursive systems with correlated residuals .

However, only models in the first two of these classes can be reinterpreted as generating processes in those variables which correspond to nodes of the given graph. This may permit simplified estimation and interpretation.

As noted previously a stepwise generating process in the given observed variables is specified by univariate recursive regression systems with independent residuals. The corresponding independence graph contains only arrows, i.e. it is fully directed, and it is acyclic, in addition.

A direct interpretation as generating process is also possible for multivariate regression chains (Cox and Wermuth, 1993; 1996). The corresponding independence graphs have as edges arrows and dashed lines for joint responses. They are directed acyclic in joint responses. This means that there are no fully directed and no partially directed cycles, i.e. it is impossible to start on a path with an arrow on it and return to the starting node without meeting an arrow head along the path.

Otherwise, no direct generating processes are specified after marginalizing. Two broad model classes of this type are discussed in the literature for joint Gaussian distributions: covariance graph and noncyclic structural equation models. The models with a pattern of zeros in marginal correlations had been introduced as linear in covariances models by Anderson (1973) and have more recently be called covariance graph models. Their independence graph is undirected with exclusively dashed lines as edges. Gaussian noncyclic structural equation models are the most general type of summary graph models: in the econometric literature some of these are discussed as linear univariate recursive system with correlated residuals (Goldberger, 1964) and as sequences of seemingly unrelated regressions (Zellner, 1962).
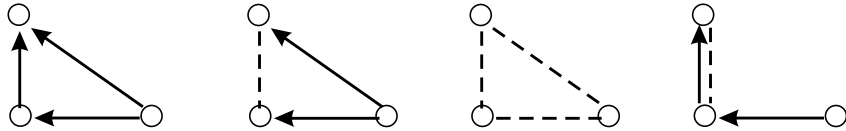
Figure 6: Four types of summary graphs. For a joint Gaussian distribution each represents a saturated model: univariate recursive regressions with independent errors (first), multivariate regression (second), a covariance graph model (third), univariate recursive regressions with correlated errors (fourth)

For each of the four types of model classes Figure 6 shows the graphs for saturated models in three variables, i.e for a joint distribution without any independencies. In a Gaussian distribution a marginal independence for Y_i, Y_j holds if and only if the

11

covariance σ_{ij} equals zero and a conditional independence, say for Y_i, Y_j given Y_k, holds if and only if the partial covariance $\sigma_{ij.k}$ equals zero. Edges in the first three graphs in Figure 6 correspond to modelling two, one and no conditional association or, equivalently one, two and three marginal associations as follows

$$\text{first}: \beta_{12.3} = \sigma_{12.3}/\sigma_{22.3}, \quad \beta_{13.2} = \sigma_{13.2}/\sigma_{22.3}, \quad \beta_{23} = \sigma_{23}/\sigma_{33};$$

$$\text{second}: \beta_{12} = \sigma_{12}/\sigma_{22}, \quad \beta_{23} = \sigma_{23}/\sigma_{33}, \quad \sigma_{12.3};$$

$$\text{third}: \sigma_{12}, \quad \sigma_{13}, \quad \sigma_{23}.$$

In general, whenever the graph represents a recursive system with some correlated errors, it may be interpreted as confounding of some direct or indirect dependencies. Then, a missing edge need no longer indicate an independency such as in the fourth graph in Figure 6. This makes graphical representations somewhat less attractive. In addition, different parametrizations are compatible in such situations. If we write the relations corresponding to the fourth graph in Figure 6 for mean-centered Gaussian variables as

$$Y_1 = \gamma_{12}Y_2 + \eta_1, \quad Y_2 = \gamma_{23}Y_3 + \eta_2, \quad Y_3 = \eta_3, \quad cov(\eta_1, \eta_2) \neq 0$$

then one parametrization corresponds to a structural equation model with $\gamma_{12} = \sigma_{13}/\sigma_{12}$ and $\gamma_{23} = \sigma_{23}/\sigma_{33}$, another has equation parameters as in a system without correlated residuals, i.e. with $\gamma_{12} = \sigma_{12}/\sigma_{22}$ and $\gamma_{23} = \sigma_{23}/\sigma_{33}$. In the former Y_3 acts like an 'instrumental variable' for the relation between Y_1, Y_2, in the latter the residual correlation is regarded as a secondary feature of a system generated essentially over a directed acyclic graph. Generalized systems of the second kind have recently been investigated by John van Briezen-Raz (personal communication).

2.3 Separation criterion in summary graphs

Independencies implied by a given summary graph may be read directly off the graph by using the following criterion.

Separation criterion in summary graphs, G_{sum}^S: Let a, b, c be nonoverlapping subsets of S, then $X_a \perp\!\!\!\perp X_b \mid X_c$ if every path from a node in a to one in b breaks by conditioning on nodes of c.

A path in G_{sum}^S breaks by conditioning on c if along the path there is

(*i*) a noncollision node in c, or

(*ii*) a collision node – together with all its descendants – is marginalized over since they are not in c.

There are three types of *collision nodes t* in a summary graph which are said to

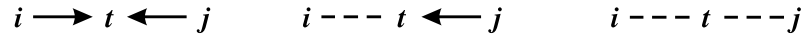$$i \longrightarrow t \longleftarrow j \qquad i --- t \longleftarrow j \qquad i --- t ---j$$

Figure 7: The three types of collision nodes in a summary graph

have visible or hidden arrow heads pointing to them. The heads are either both visible (left), or one is hidden (middle) or both are hidden (right) due to marginalizing. The reason is that every dashed line is generated by marginalizing over all nodes along a common source path present in the generating graph, i.e. via a path which had arrow heads at both path ends.

We give examples of three types of paths typical for directed acyclic graphs in Figures 8 to 10. The path in Figure 8 is a collisionless, descendant-ancestor path. Node i is the descendant of node j and node j is the ancestor of node i. The path is direction-preserving. It breaks if any node along the path is in c.

$$i \longleftarrow \bigcirc \longleftarrow \bigcirc \longleftarrow \bigcirc \longleftarrow_{\cdots} - \bigcirc \longleftarrow \bigcirc \longleftarrow j$$
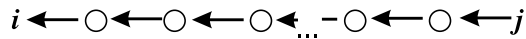
Figure 8: A noncollision, descendant-ancestor path

The path in Figure 9 is a collisionless, common source path. Nodes i and j have a common source along the path. This is a node from which one direction-preserving path leads to node i and another to node j. The path breaks if any node along it is in c.

$$i \longleftarrow \bigcirc \longleftarrow \bigcirc \longrightarrow \bigcirc -_{\cdots} \longrightarrow \bigcirc \longrightarrow \bigcirc \longrightarrow j$$
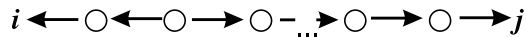
Figure 9: A noncollision, common-source path

The path in Figure 10 is of the most general type possible in a directed acyclic graph. It is a collision path because it contains collision nodes, but there are transition nodes

13

and a source node as well. This path breaks like the previous ones if any noncollision node is conditioned on. But it breaks also if a single collision node and all of its descendants are marginalized over.
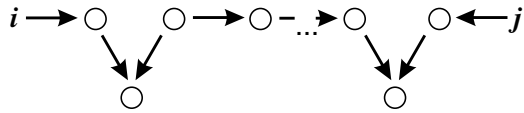

Figure 10: A collision path

The separation criterion for summary graphs may be applied to variables and distributions of any type, distributions may even be degenerate. This becomes different if we want to conclude from a and b not being separated by c that there is some strictly nonvanishing dependence between X_a and X_b given X_c. In general, a single path between a and b which does not break relative to c only means that $X_a \perp\!\!\!\perp X_b \mid X_c$ is not implied by the generating process (but that it may still hold under very special parametric constellations sometimes called parametric cancellations).

We call a path active if it does not break. In linear and quasi-lineear systems an active path introduces an association for i, j given c if some special additional conditions hold (Wermuth and Cox, 1998). A more direct definition of an active path is as follows. *A path in G_{sum}^{S} is active relative to c* if along the path

(i) every collision node is in c or is an ancestor of a node in c, and

(ii) every noncollision node is marginalized over since it is not in c.

To illustrate this definition we use a symbol for conditioning as $c = \{\square\}$, in addition to the one for marginalizing $\emptyset\!\!\!/$. Figure 11 displays conditions under which the path of Figure 10 becomes active.

3. No treatment effect of sequential treatments

We now use the results summarized in the previous section to investigate properties of models used to study effects of sequentially administered treatments in randomized clinical trials.

Robins and Wasserman (1997) describe the following situation in which a naive use of regression models leads to false conclusions. It is a clinical trial in which AIDS pa-
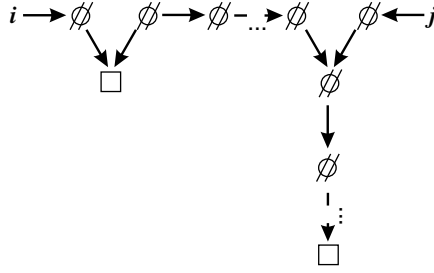
Figure 11: An active collision path: every collision node is in c, or is an ancestor of a node in c, every other node along the path is marginalized over.

tients have received AZT treatment twice. At both times dose of treatment is assigned at random. Randomization probabilities for the recent treatment dose, T_r, are however dependent on the previous treatment dose, T_p, and the effect this treatment had on an intermediate variable, on anemia of the patient, L. Of primary interest is the overall outcome, Y, measured as HIV-viral load at the end of a follow-up period. Hidden, i.e. unobserved, is the patient's immune function, U, an indicator of the patient's general health status. Figure 12 shows an ordering corresponding to such a generating process of the data and a directed acyclic graph which represents among other independencies the hypothesis of no treatment effect, i.e. $Y \perp\!\!\!\perp (T_r, T_p) \mid U$.

3.1 Defining and implied independencies

From the definitions in Section 1.2 the defining independencies in the graph of Figure 12 are:

$$Y \perp\!\!\!\perp (T_r, T_p, L) \mid U, \quad T_r \perp\!\!\!\perp U \mid (T_p, L), \quad T_p \perp\!\!\!\perp U.$$

In this graph of Figure 12 the edges (T_p, U) and (T_r, L) are missing by design, i.e. because treatment doses are assigned at random. Randomization assures independence of treatments and potential confounders, be they observed or not. Edges (Y, T_r) and (Y, T_p) are missing because the graph represents the null hypothesis of no treatment effect given U. The edge (Y, L) is missing to simplify exposition.

Since the patient's underlying immune function U is not observed the independencies after marginalizing over U are those of interest. They can be determined for every pair with a missing edge by using the separation criterion of Section 2.3. We note first
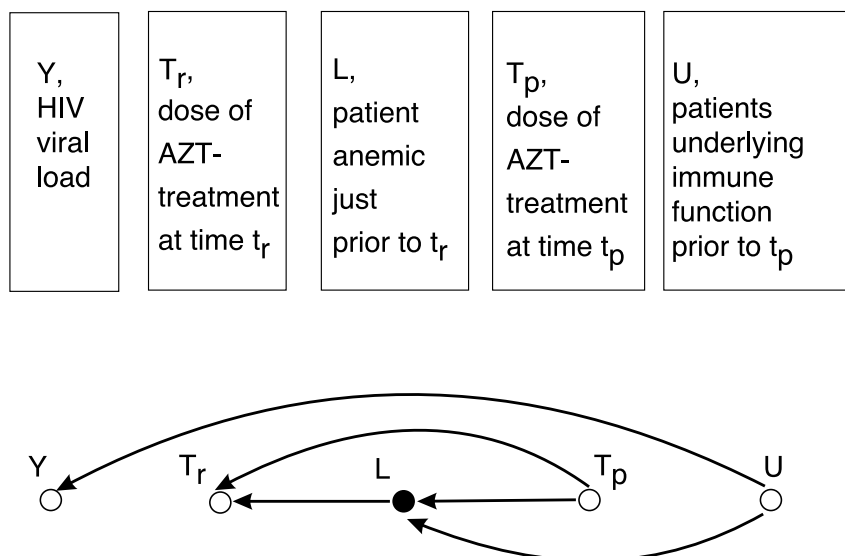
15

Figure 12: Ordering of variables in a clinical trial with two sequentially administered treatments; the corresponding directed acyclic graph reflecting randomized assignment of treatments (edges (T_p, U) and (T_r, U) are missing) and no treatment effect (edges (Y, T_p) and (Y, T_r) are missing) and an additional simplification (edge (Y, L) is missing).

that by marginalizing over the common source U, i.e. with U outside c, a path via U does not break. Then we look at paths for pairs (Y, T_p) and (Y, T_r), in turn.

There are two paths between Y and T_p. The path (Y, U, L, T_p) breaks iff, i.e. if and only if L and T_r are both marginalized over, since L is a collision node along this path and T_r is its descendant. Path (Y, U, L, T_r, T_p) breaks iff either the collision node T_r is marginalized over or the noncollision node L is in c. Thus, with U outside c, both paths break iff L and T_r are both marginalized over. This means that $Y \perp\!\!\!\perp T_p$ is implied, while $Y \perp\!\!\!\perp T_p \mid L$, and $Y \perp\!\!\!\perp T_p \mid T_r$ and $Y \perp\!\!\!\perp T_p \mid (L, T_r)$ are not implied by the generating process.

Similarly, there are two paths between Y and T_r. Path (Y, U, L, T_r) breaks iff the transition node L is in c. Path (Y, U, L, T_r, T_p) breaks iff the source node T_p is in c or the collision node L is marginalized over. Note that L is a transition node along the first path but a collision node along the second path. Both paths break iff both, T_p and L, are in c. Thus, $Y \perp\!\!\!\perp T_r \mid (T_p, L)$ is implied by the graph of Figure 12 for the observed variables, but no other independence statement involving pair (Y, T_r) is implied by the
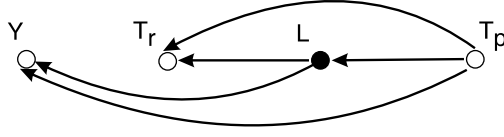
16

Figure 13: A directed acyclic graph for the observed variables corresponding to a system of univariate recursive regressions.

generating process.

Therefore the hypothesis of no treatment effect incorporated in the above generating process of all five variables $Y \perp\!\!\!\perp (T_r, T_p) \mid U$ to together with $Y \perp\!\!\!\perp L \mid U$ imply that

$$Y \perp\!\!\!\perp T_r \mid (T_p, L) \text{ and } Y \perp\!\!\!\perp T_p$$

for the four observed variables but no other independencies hold in this system of observed variables.

This means, in particular, that the independence statements $Y \perp\!\!\!\perp (T_r, T_p) \mid L$ and $Y \perp\!\!\!\perp (T_r, T_p)$ are both incompatible with joint distributions of the observed variables. To put it differently, tests of independence of the final outcome variable Y of both treatments simultaneously will be rejected for large numbers of observations, no matter whether we condition on L or not. This is what Robins and Wasserman observe. They use G-computation for the correct analysis.

The standard regression model for Y discussed by Robins and Wasserman is derived from the directed acyclic graph in Figure 13 for the observed variables. This graph keeps the ordering of the variables in the generating graph of Figure 12 and there is an arrow whenever the corresponding independence statement is not implied by the graph in Figure 12. Thus, in Figure 13 the edge for overall outcome Y is missing to T_r since its absence means $Y \perp\!\!\!\perp T_r \mid (T_p, L)$ which is implied by the generating graph. The arrow from T_p to Y is present, since its absence would mean $Y \perp\!\!\!\perp T_p \mid L$ and this independence is not implied by the generating graph.

Thus, the independence $Y \perp\!\!\!\perp T_p$ implied by the generating process is not reflected in the directed acyclic graph of Figure 13 and it cannot be captured by removing the arrow pointing to Y from T_p. Instead, the proper independence graph for the observed variables, i.e. their summary graph, is a graph in which the intermediate outcome, L,
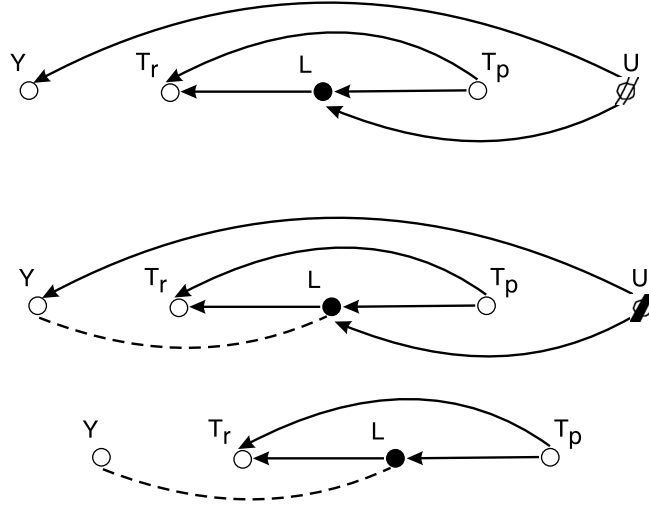
17

Figure 14: Derivation of the summary graph for the observed variables. Top: starting graph, directed acyclic, with node U to be marginalized over; middle: inserted edge due to marginalizing; bottom: summary graph: a joint response graph

and the final outcome, Y, occur as joint responses. Effects of this are discussed in the following Section.

3.2 The summary graph for the observed variables

To obtain the summary graph implied for the observed variables by the generating process to Figure 12 we need to marginalize over node U. Node U is a common source for nodes Y and L unconnected in Figure 12. Nodes Y and L become connected by a dashed line in the summary graph (see Table 1 and Figure 14). No other edges are induced. The corrresponding model is a joint response model (Cox and Wermuth, 1993; 1996) which reflects correctly $Y \perp\!\!\!\perp T_r \mid (L, T_p)$ and $Y \perp\!\!\!\perp T_p$, the independencies implied by the generating process for the observed variables.

The separation criterion of Section 2.3 may again be used to read these independencies directly off the summary graph in Figure 14 as follows. There are two paths between Y and T_r. Path (Y, L, T_r) breaks iff the noncollision node L along the path is in c. Path (Y, L, T_p, T_r) breaks iff the collision node L on this path is marginalized over or the source node T_p is in c. Hence, both paths break iff both L and T_p are in c, so that $Y \perp\!\!\!\perp T_r \mid (T_p, L)$ is implied.

There are also two paths between Y and T_p. Path (Y, L, T_p) breaks iff the collision node L and its descendant T_r are both marginalized over. Path (Y, L, T_r, T_p) breaks iff L is in c or T_r is marginalized over. Hence both paths break iff both L and T_r are marginalized over, so that $Y \perp\!\!\!\perp T_p$ holds.

3.3 Alternative mixed parametrizations

So far, we did not need information on the type of variables involved. To obtain a joint distribution satisfying the defining independencies of Figure 12 with the intermediate response L being binary, one standard parametrization is in terms of Conditional Gaussian regressions (Lauritzen and Wermuth, 1989), i.e. linear regressions for continuous responses and logistic regresssions for binary responses.

This is a parametrizations discussed by Robins and Wasserman. However, if the main hypotheses of interest involve marginalization over a discrete intermediate response, here L, then the properties of mixed distribution have to be taken into account. A CG-distribution is closed under conditioning but not necessarily under marginalizing (Frydenberg, 1989). A joint Gaussian distribution for which some variables are dichotomized or, more generally, categorized, is closed under marginalizing but not necessarily under conditioning (Cox and Wermuth, 1993; 1999).

In particular, if the joint distribution of (Y, L, T_p) for L binary is defined in terms of CG-regressions for Figure 14, then a complicated marginal distribution for Y, T_p results which involves the parameters of the logistic regression of L on T_p. If however the joint distribution of (Y, L, T_p) is taken to be partially dichotomized Gaussian, then the marginal joint distribution of Y, T_p is Gaussian and, consequently, the test of $Y \perp\!\!\!\perp T_p$ reduces to a standard procedure.

3.4 Summary and open questions

We have discussed the situation of a randomized clinical trial introduced by Robins and Wasserman in which there is no treatment effect given information on the health status of the patient. Treatments are administered sequentially and the health status of the patient is not measured. It is an important example which shows that a univariate conditional distribution may be inappropriate to analyze the possible influences of a

final outcome no matter whether the intermediate outcome is included, i.e. conditioned on, or excluded from i.e. marginalized over, in the regression analysis.

We have shown, in particular, how naive use of univariate regression models can be avoided by deriving the proper independence structure for the observed variables and by noting that this summary graph is not directed acyclic but a joint response graph. These results apply to any type of joint distribution generated over the given graph and they provide an alternative approach to a correct analysis than the one suggested by Robins and Wasserman. We have also pointed at a standard parametrization for the joint distribution of observed variables in the case in which marginalizing over a disrete variable leads to a joint Gaussian distribution.

We have not discussed the problem of estimating the treatment effects or the situation in which the intermediate outcome, L, has a direct effect on the final outcome, Y. In the latter case the summary graph at the bottom of Figure 14 would have an arrow pointing from L to Y, in addition to the dashed line edge, indicating that there is some confounding effect. If such a situation can be anticipated early on, a different design of the study might be helpful.

Acknowledgement

References

Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135–141.

Cox, D.R. & Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika*, **79**, 441-461.

Cox, D.R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, **8**, 204-218; 247-277.

Cox, D.R. & Wermuth, N. (1996). *Multivariate Dependencies – Models, Analysis and Interpretation.* London: Chapman & Hall

Cox, D.R. & Wermuth, N. (1999). Likelihood factorizations for mixed discrete and continuous variables. *Scand. J. Stat.*, **25**, 209-220.

Dawid, A.P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc.* B **41**, 1-31.

Frydenberg, M. (1990). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* **18**, 790–805.

Goldberger, A.S. (1964). *Econometric Theory.* New York: Wiley.

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

Koster, J. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Statist,*, **26**, 413-431.

Lauritzen, S.L.& Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31-57.

Markov, A.A. (1912). *Wahrscheinlichkeitsrechnung.* (German translation of 2nd Russian edition: Markoff, A.A., 1908). Teubner, Leipzig.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufmann.

Richardson T.S. (1999). Markov Properties for acyclic directed mixed graphs. Technical Report 353, Department of Statistics, University of Washington.

Robins, J. and Wasserman, L. (1997). Estimation of effects of sequential treatments by reparametrizing directed acyclic graphs. In: D. Geiger and O. Shenoy (eds.) *Proccedings, 13th Annual Conference on Uncertainty in Artificial Intelligence.* 409-420. San Francisco: Morgan and Kaufmann.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., Glymour, C. (1998). Using path diagrams as a structural equation modelling tool. *Sociological Methods and Research*, **27**, 182-225.

Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.*, **75**, 963-97.

Wermuth, N. (1992). On block-recursive linear regression equations (with discussion). *RE-BRAPE (Brazilian Journal of Probability and Statistics)*, **6**, 1-56.

Wermuth, N. & Cox, D.R. (1998). On dependence models defined over independence graphs. *Bernoulli.* **4,** 477-496.

Wermuth,N. & Cox, D.R. (2000). A sweep operator for triangular matrices and its statistical applications. Research Report (ZUMA, 00-04)

Wermuth, N., Cox, D.R. & Pearl J. (2000). Explanations for multivariate structures derived from univariate recursive regressions. (under revision for *Bernoulli*)

Wermuth, N.& Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discusssion). *J. Royal Statist. Soc. B*, **52**, 21-75.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60–62.

Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20**, 162–177.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57**, 348–368.