On association models defined over independence graphs

NANNY WERMUTH[1]* and DAVID R. COX[2]

[1]*Johannes Gutenberg-Universität, Mainz 55099, Germany. e-mail: wermuth@animal.sowi.uni-mainz.de*
[2]*Nuffield College, Oxford OX1 1NF, UK. e-mail: david.cox@nuffield.oxford.ac.uk*

Conditions on joint distributions are given under which two variables will be conditionally associated whenever an independence graph does not imply a corresponding conditional independence statement. To this end the notions of parametric cancellation, of stable paths and of quasi-linear models are discussed in some detail.

Keywords: conditional independence; graphical Markov model; quasi-linear dependence; separation criteria; triangular matrix decomposition

1. Introduction

In the recent extensive research on the use of independence or Markov graphs to study associations and dependences in multivariate systems, each variable is represented by a node; some pairs of nodes are joined by edges, sometimes directed and sometimes undirected and a key role is played by conditional independence. The absence of an edge between two nodes means that the corresponding variables are conditionally independent, the conditioning set depending on the nature of the graph. See Lauritzen and Wermuth (1989), Cox and Wermuth (1993, 1996), Edwards (1995), Lauritzen (1996) and Wermuth (1997). For a description of various kinds of special dependence which may be so represented.

The following qualitative distinction is important in applications but rather difficult to capture formally in probabilistic theory. In a statistical model corresponding to a given graph an edge that is present typically corresponds to a free parameter, e.g. a correlation or regression coefficient, which may take any value in the relevant parameter space, including values at zero, the value for a particular independence. On the other hand in a substantive research hypothesis (Wermuth and Lauritzen 1990) an edge that is present is to represent a dependence large enough to be of subject-matter interest. Now the magnitude of such an effect depends on the context. All that we can require for a general discussion is to characterize situations in which rules for specifying and reading off a given graph conditional independences and conditional associations do not overlook some of these

---

*To whom correspondence should be addressed.

relations. We stress that often the representation of dependences is at least as important as representing independences and that the stronger interpretation can be essential.

We are mainly concerned with applications of independence graphs to observational studies in the social sciences. In that context it is typical that the number of variables is not very large, that associations are of moderate strength and that for the variables analysed no event occurs with probability one.

In the present paper we first summarize some terminology for independence graphs and state known so-called separation criteria and some of their implications. A separation criterion answers one central question: are variables corresponding to two sets *A*, *B* conditionally independent given variables in set *C*? Here *A*, *B*, *C* are arbitrary disjoint subsets of the set of all nodes *V* and the independence statements refer to the family of joint distribution associated with the given graph.

Second, we introduce and discuss the notion of parametric cancellation. It characterizes situations in which one special member of a family of distribution satisfies more independence statements than those implied by an associated independence graph, that is more than those given by the appropriate separation criterion. In particular, we give examples with few variables and present some sufficient conditions under which there is no parametric cancellation.

Third, we prove a theorem for relations induced in joint Gaussian distributions generated over a directed acyclic graph in such a way that every edge present in a directed graph corresponds to a particular, strictly non-zero regression coefficient and that in addition there is no parametric cancellation relative to a particular set of variables conditioned on.

The theorem gives several reformulations of the separation criterion for directed acyclic graphs to answer the question: does the generating process imply two variables corresponding to nodes *i* and *j* to be conditionally dependent or independent given the variables in some other subset *C* of *V*? In this formulation the primary aim is to assert dependence whereas the separation criteria had been formulated to derive independences implied by a graph.

Finally, we introduce the notion of quasi-linearity to extend the usefulness of the theorem to some systems containing interactive effects and nonlinear dependences.

Note that independences connected with a given Markov graph do not require specification of the form of distribution involved. Moreover joint distributions are here typically defined only implicitly via a sequence of (conditional) distributions and the independences specified by the graph. In any particular application, particular families of distributions are used. When we discuss parametric cancellation we are concerned with additional independences holding for very special members of the fuller family. Similarly, induced dependences are relevant only for the special members corresponding to an assumed physical generating process.

2. Terminology, separation criteria and induced graphs

2.1. Graph terminology

We consider graphs of *p nodes* $V = \{1, \ldots, p\}$ indicating random variables $Y_1, \ldots, Y_p$ and at most one *edge i*, *j* between each pair of nodes *i* and *j*. Edges represent conditional

association parameters in the distribution of Y_V, assumed here to be non-degenerate. An edge may be directed and then drawn as an *arrow*, or it may be undirected and then drawn as a *line*. If a graph has only lines, then it is an *undirected graph*. The graph is said to be *fully directed* if all its edges are arrows, and *partially directed* if some of its edges are lines and some are arrows. In undirected and in partially directed graphs, two types of edge, *dashed* or *full*, may be present, indicating different conditioning sets for associations.

An edge i, j has *no orientation* if it is a line or it has one of two possible *orientations* when it is an arrow, either *pointing from j to i*, i.e. having the arrow point at i, or pointing from i to j. A *path* of length $n - 1$ nodes is a succession of edges connecting nodes $i_1, \ldots, i_n$, irrespective of the orientation of the edges. Note that a path of length 1 is an edge present in the graph. The graph obtained from any given one by ignoring type and orientation of edges is called the *skeleton* graph.

Two nodes i, j are said to be *adjacent* or *neighbours* if they are connected by an edge and are said to have a *common neighbour t*, if t is adjacent to both i and j. Three types of common neighbour node t of i, j can be distinguished along any fully directed path as shown in Figure 1. Two arrows point to i and to j from a *source node t* (Figure 1a); a *transition node t* has one incoming arrow, say from j, and one outgoing arrow (Figure 1b); a *sink node t* has two arrows pointing at it one from each of i and j (Figure 1c). Because two arrows meet head on at a sink node, it is also called a *collision node* or a node having *converging* arrows. With several paths passing through a node, the same node may take on different roles along different paths.

A path containing a collision node is a *collision path* and a path is said to be *collisionless* otherwise. A path of arrows pointing to i from j via transition nodes is called a *direction-preserving* path; an example is given in Figure 1b. In such a path, i is named a *descendant* of node j and node j is called an *ancestor* of i.

A graph constructed from a given one by keeping nodes and edges present within a selected subset S of nodes is an *induced subgraph*. The subgraph preserves type and orientation of the edges. If the induced subgraph of three nodes i, t, j has exactly two edges it is named a V-*configuration* and if it is one of the paths of Figure 1a, b and c, it is called *source*, *transition* and *sink oriented*, respectively.

In a *directed acyclic graph* G_{dag}^V, all edges are directed and there is no direction-preserving path from a node back to itself. Given the set of ancestors adjacent to a node i, this graph defines Y_i to be conditionally independent of the remaining ancestors of i.

In the contexts that we are mainly concerned with, a complete ordering of the variables is typically specified from subject-matter knowledge about the variables. Given such an order we have a *univariate recursive regression system* that is a sequence of conditional distributions with Y_1 regressed on $Y_2, \ldots, Y_p$, with Y_2 regressed on $Y_3, \ldots, Y_p$, and so on up to Y_{p-1} regressed on Y_p. Each *response* Y_i has then *potentially explanatory* variables
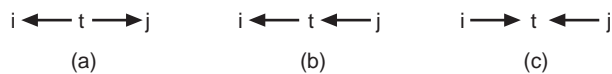

Fig. 1. The three distinct types of common neighbour node t along paths in a fully directed graph: (a) a source node; (b) a transition node; (c) a sink or collision node.

$Y_{i+1}, \ldots, Y_p$ and we assume that its conditional dependence on Y_j given its remaining potentially explanatory variables can be captured by a set of parameters, the null values of which imply the corresponding independence statement: $Y_i \perp\!\!\!\perp Y_j | Y_{\{i+1,\ldots,p\}\setminus\{j\}}$, for $i < j$.

For instance, in Gaussian systems these parameters are partial regression coefficients; for binary quadratic exponential distributions (Cox and Wermuth 1994) they are constant partial log-odds-ratios, for general logistic regressions and analyses of variance models they are main- and higher-order effects. Not covered by this type of specification are most regressions for binary responses which are linear in probabilities, since conditional independence shows there in several parameters being equal; see Cox and Wermuth (1992) for some further discussion of this point.

The graphical representation of such a system is sometimes called a *univariate recursive regression graph*. It is a directed acyclic graph with two additional features: each edge present represents a specific non-vanishing conditional dependence and each edge absent represents one particular conditional independence statement for the variable pair involved. We then say that the joint distribution is generated over the given G_{dag}^V or that G_{dag}^V is a *generating graph* because it is intended to represent a process by which the data could have been generated.

For the same directed acyclic graph the nodes can often be numbered in more than one way without changing the independences implied by a directed acyclic graph (Frydenberg 1990). Therefore, the recursive ordering of the generating regressions is often indicated by drawing a chain of boxes around nodes $1, \ldots, p$. Whenever there are several conditionally independent responses given their common ancestors, then the nodes of this set are drawn in *stacked boxes*; such variables are mutually independent given the variables in all boxes to the right. Examples are in Figures 2 and 3 below.

We now turn to systems whose properties are conveniently specified via covariances or concentrations, as is especially appropriate for Gaussian systems. Given a selected triplet of nodes $S = \{i, t, j\}$, a set 'm' of nodes denoting a (vector) variable X_m over which marginalization has taken place, a set 'c' referring to the (vector) variable X_c on which conditioning has occurred, and an ordering $V = (m, S, c)$, we can write the covariance and the concentration matrix of the set S given the set c as

$$\Sigma_{SS.c} = \begin{pmatrix} \sigma_{ii.c} & \sigma_{it.c} & \sigma_{ij.c} \\ . & \sigma_{tt.c} & \sigma_{tj.c} \\ . & . & \sigma_{jj.c} \end{pmatrix}, \qquad \Sigma_{SS.c}^{-1} = \Sigma^{SS.m} = \begin{pmatrix} \sigma^{ii.m} & \sigma^{it.m} & \sigma^{ij.m} \\ . & \sigma^{tt.m} & \sigma^{tj.m} \\ . & . & \sigma^{jj.m} \end{pmatrix},$$

and, for instance, the partial correlation coefficient $\rho_{ij.c}$ can be expressed in terms of linear regression coefficients, covariances and concentrations (Wermuth 1976) via

$$\rho_{ij.c} = \beta_{ij.c} \left(\frac{\sigma_{jj.c}}{\sigma_{ii.c}} \right)^{1/2} = -\sigma^{ij.m} (\sigma^{ii.m} \sigma^{jj.m})^{-1/2},$$

$$\beta_{ij.c} = \frac{\sigma_{ij.c}}{\sigma_{jj.c}} = -\frac{\sigma^{ij.m}}{\sigma^{ii.m}},$$

where in a non-degenerate system all variances $\sigma_{ii.c}, \ldots,$ and all precisions $\sigma^{ii.m}, \ldots$ are positive.

The dot notation denotes conditioning for covariances and marginalizing for concentrations. It is particularly useful to express recursion relations such as

$$\beta_{ij.c} = \beta_{ij.tc} + \beta_{it.jc}\beta_{tj.c},$$

$$\sigma_{ij.c} = \sigma_{ij.tc} + \frac{\sigma_{it.c}\sigma_{jt.c}}{\sigma_{tt.c}},$$

$$\sigma^{ij.m} = \sigma^{ij.tm} + \frac{\sigma^{it.m}\sigma^{jt.m}}{\sigma^{tt.m}}$$

(see Cochran (1938), Anderson (1984) and Dempster (1969), respectively).

With the help of these, the important consequences of a fixed order can be emphasized: each edge missing and each edge present has a precise interpretation for the involved variable pair. Let the selected triplet S refer to V-configurations of a large generating graph, let m denote descendants of S and c ancestors of S. Then for Gaussian systems the source-oriented V-configuration of Figure 1a specifies $0 = \rho_{ij.tc}$, $\beta_{it.c} \neq 0$, $\beta_{jt.c} \neq 0$ and implies that $\rho_{ij.c} \neq 0$. The transition-oriented V-configuration in Figure 1b specifies $0 = \beta_{ij.tc}$, $\beta_{it.jc} \neq 0$, $\beta_{tj.c} \neq 0$ and implies that $\rho_{ij.c} \neq 0$. The sink-oriented V-configuration of Figure 1c specifies $0 = \sigma_{ij.c}$, $\beta_{ti.jc} \neq 0$, $\beta_{tj.ic} \neq 0$ and implies that $\rho_{ij.tc} \neq 0$. In each case the implied non-zero correlation for the missing edge i, j results because the configuration corresponds to exactly one zero element in $\Sigma_{SS.c}$ or $\Sigma_{SS.c}^{-1}$ and hence to a non-zero element in the same position of the corresponding inverse matrix.

2.2. Separation criteria for the directed acyclic graph G_{dag}^V

As mentioned before, conditions which help to decide which independence statements are implied by any given graph have been called *separation criteria*. Such separation criteria were given for two types of undirected graph by Darroch *et al.* (1980) and by Kauermann (1996) and for partially directed graphs by Frydenberg (1990). We restate here two different separation criteria for directed acyclic graphs. These have been shown to be equivalent and for all disjoint subsets A, B, C of V one may conclude under very general conditions that Y_A is independent of Y_B given Y_C provided that the criterion is satisfied (Lauritzen *et al.* 1990).

Separation Criterion 1 for G_{dag}^V (Pearl 1988, p. 117; Pearl and Verma 1988). *If A, B, C are three disjoint subsets of nodes in a directed acyclic graph, then C is said to d-separate A from B if there is no path between a node in A and a node in B along which the following conditions hold: (1) every node with converging arrows is in C or has a descendant in C; (2) every other node is outside C.*

Separation Criterion 2 for G_{dag}^V (Lauritzen et al. 1990). *A and B are separated by C in the directed acyclic graph, whenever in the moral graph formed from the smallest ancestral set containing $A \cup B \cup C$ every path from A to B has a node in C.*

This moral graph is constructed in three steps: (1) obtain from G_{dag}^V the subgraph induced by nodes of the union of A, B, C and their ancestors; (2) join in it every sink-oriented V-

configuration by a line; (3) replace every arrow in the resulting graph by a line. Then the separation criterion for undirected graphs is used to decide on separation in the directed graph.

Figure 2 illustrates an application to a generating graph with seven nodes.

2.3. Induced covariance and concentration graphs

Whenever a joint distribution is generated over a given directed acyclic graph G_{dag}^V, the above separation criteria may be used to derive further features of the joint distribution, for instance those represented by the following two types of undirected graph.

The overall *covariance graph* G_{cov}^V is an undirected graph of dashed lines. In it an edge i, j is absent if G_{dag}^V implies that the variables Y_i and Y_j are marginally independent $(Y_i \perp\!\!\!\perp Y_j)$ and an edge is present otherwise. The overall *concentration graph* G_{con}^V is an undirected graph of full lines. In it an edge between nodes i, j is absent if G_{dag}^V implies that the variables Y_i and Y_j are conditionally independent given all other $p - 2$ variables $(Y_i \perp\!\!\!\perp Y_j | Y_{V \setminus \{i,j\}})$. In a multivariate normal distribution the absence of an edge in the overall covariance graph means that there is a zero in a corresponding position of the covariance matrix Σ, while the absence of an edge in the overall concentration graph is equivalent to there being a zero in a corresponding position of the concentration matrix, i.e. in the inverse covariance matrix Σ^{-1}.

The following formulation of results derived by Pearl and Wermuth (1994) is a direct consequence of the above Separation Criterion 1. A given generating graph G_{dag}^V induces an edge i, j in the overall covariance graph G_{cov}^V if and only if in the generating graph there is a collisionless path between the two nodes. Further, it induces an edge between nodes i and j in the overall concentration graph G_{con}^V if and only if in the generating graph either i, j is an edge or nodes i and j have a common collision node. Figure 3 shows the two undirected graphs G_{con}^V (Figure 3b) and G_{cov}^V (Figure 3c) induced by the directed acyclic graph of Figure 3a.

(a)   (b)   (c)
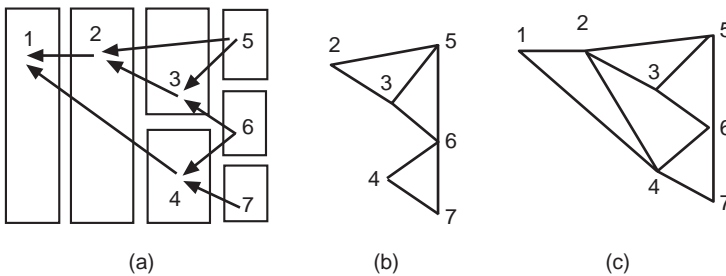
Fig. 2. (a) A generating graph in seven nodes; the corresponding moral graphs for the same sets $A = \{2\}$ and $B = \{7\}$ but different conditioning sets C: (b) $C = \{3, 4, 5\}$ separates A and B in the undirected graph since every path from A to B has a node in C; (c) $C = \{1, 3, 4\}$ does not separate A and B, since the path 2,5,6,7 is outside C.
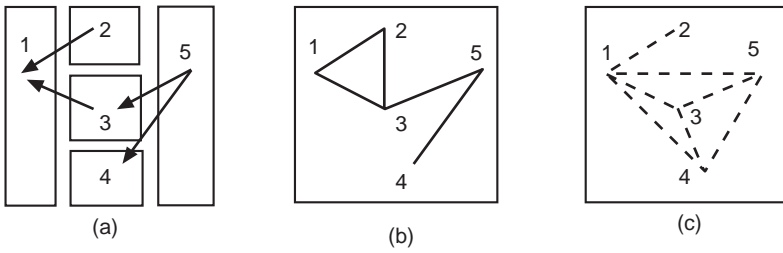
Fig. 3. (a) A generating graph G_{dag}^V with five nodes; (b) the overall concentration graph G_{con}^V induced by it; (c) the overall covariance graph G_{cov}^V induced by it.

It follows that covariance and concentration graphs have at least as many edges as the generating directed graph from which they are induced. Thus, neither can imply more independence statements than the generating graph G_{dag}^V. Two graphs are *independence equivalent* if they imply the same set of independence statements, i.e. whenever every independence statement derived in one graph can also be derived via the appropriate separation criterion for the other graph. If two graphs are independence equivalent, then they have identical skeleton graphs.

A given generating graph G_{dag}^V is independence equivalent to its induced concentration graph G_{con}^V if and only if there is no sink-oriented V-configuration in G_{dag}^V (Wermuth 1980; Wermuth and Lauritzen 1983; Kiiveri *et al.* 1984; Frydenberg 1990).

Similarly, for Gaussian distributions a generating graph G_{dag}^V is independence equivalent to its induced covariance graph G_{cov}^V if and only if in G_{dag}^V there is no source- or transition-oriented V-configuration (Pearl and Wermuth, 1994).

More generally, covariance and concentration graphs of a subset of variables Y_S may be considered conditionally given another subset Y_C (Wermuth 1995). Such a covariance graph $G_{\text{cov}}^{S.C}$ will have an (undirected dashed line) edge i, j if $Y_i \perp\!\!\!\perp Y_j | Y_C$ is not implied by the generating graph and such a concentration graph $G_{\text{con}}^{S.C}$ will have an (undirected full line) edge i, j if $Y_i \perp\!\!\!\perp Y_j | Y_{C \cup S \setminus \{i,j\}}$ is not implied by the generating graph. For a multivariate normal distribution of Y_V the absence of an edge in the covariance graph $G_{\text{cov}}^{S.C}$ means that in the conditional distribution of Y_S given Y_C there is a zero in a corresponding position of the covariance matrix, in $\Sigma_{SS.C}$, while the absence of an edge in the concentration graph $G_{\text{con}}^{S.C}$ is equivalent to there being a zero in a corresponding position of the concentration matrix, in $\Sigma_{SS.C}^{-1}$.

We say that a generating graph G_{dag}^V is *completed by conditioning on C* if it is modified by joining with a line the non-adjacent nodes of every sink-oriented V-configuration having a descendant in C.

Construction of an induced covariance graph. *The covariance graph $G_{\text{cov}}^{S.C}$ induced by G_{dag}^V has an edge for nodes i, j if and only if in the generating graph, completed by conditioning on C, there is a collisionless path between i and j outside C.*

Construction of an induced concentration graph. *The concentration graph $G_{\text{cov}}^{S.C}$ induced by*

G_{dag}^V has an edge for nodes i, j if and only if in the generating graph, completed by conditioning on $S \cup C$, there is an edge i, j or there is a collisionless path between i and j outside $S \cup C$.

Both construction criteria are direct adaptations of Separation Criterion 1 to the special cases considered.

3. Parametric cancellations and stable paths

A *parametric cancellation* is a very special constellation among parameters such that an independence statement holds even though it is not implied by the generating graph, i.e. even though it cannot be derived from the separation criteria. Then the specific numerical values of the parameters are such that an independence arises that does not hold in general for structures associated with the given graph. Spirtes *et al.* (1993) assume this feature to hold when they speak of a 'faithful graph' and Pearl (1988) uses the term 'accidental cancellation'.

3.1. Examples of parametric cancellation

We assume again that G_{dag}^V is a generating graph, so that each edge present corresponds to a non-vanishing conditional dependence in a system of univariate recursive regressions and distinguish parametric cancellations for edges present and those for edges absent in the generating graph.

3.1.1. Examples for edges present in G_{dag}^V

The graph in Figure 4a defines marginal independence for variable pair (B, C) and implies no other independence statements.

In the corresponding joint distributions of three binary variables, each taking values -1 and 1 with probability $\frac{1}{2}$, there is parametric cancellation for both pairs (A, B) and (A, C) if the joint probabilities are $\log P(A = i, B = j, C = k) = \mu + \alpha ijk$, because these special
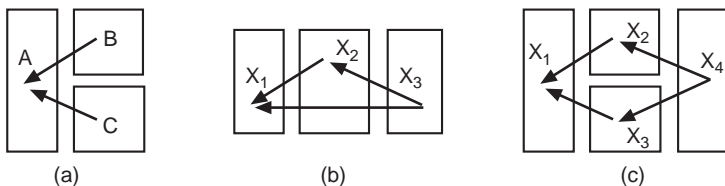

| (a) | (b) | (c) |

Fig. 4. Simple situations in which parametric cancellation could occur: (a) a binary response A directly dependent on two binary variables B, C being marginally independent; (b) a complete Gaussian system in three variables; (c) an incomplete Gaussian system in four variables.

probabilities imply marginal independence for the two pairs even though they have an edge in the generating graph.

The graph in Figure 4b implies no independence statement in the system. For Gaussian variables of mean zero it is such that

$$E\{X_1|(X_2, X_3)\} = \beta_{12.3}X_2 + \beta_{13.2}X_3, \qquad E\{X_2|X_3\} = \beta_{23}X_3.$$

For joint distributions with the special constellation $\beta_{12.3} = -\beta_{13.2}\beta_{23}$, marginal independence results for pair X_1, X_2, so that there is parametric cancellation.

3.1.2. Examples for edges absent in G_{dag}^V

The graph in Figure 4c specifies $X_1 \perp\!\!\!\perp X_4|(X_2, X_3)$, $X_2 \perp\!\!\!\perp X_3|X_4$ and implies no marginal independence for pair X_1, X_4 and no conditional independence for X_2, X_3 given X_1. This means for Gaussian variables of mean zero that

$$E\{X_1|(X_2, X_3, X_4)\} = \beta_{12.3}X_2 + \beta_{13.2}X_3,$$
$$E\{X_2|X_4\} = \beta_{24}X_4, \qquad E\{X_3|X_4\} = \beta_{34}X_4,$$

and, in particular that $\sigma_{23} = \beta_{24}\beta_{34}\sigma_{44}$. However, for Gaussian distributions following this given system with $\sigma_{23} = \sigma_{12}\sigma_{13}/\sigma_{11}$ the independence $X_2 \perp\!\!\!\perp X_3|X_1$ holds even though it is not implied by the graph. If, instead, we have $\beta_{12.3}\beta_{24} = -\beta_{13.2}\beta_{34}$, then the effects (defined below) of the two paths present between nodes 1 and 4 cancel so that $\beta_{14} = 0$ and hence $X_1 \perp\!\!\!\perp X_4$ holds in such a special Gaussian system to Figure 4c.

An example of near-parametric cancellation for an edge absent is a Markov chain in p Gaussian variables in which p becomes large. The generating graph G_{dag}^V consists in that case of a single direction-preserving path with node p, say, denoting the last ancestor of node 1 and the implied marginal correlation ρ_{1p}^* is

$$\rho_{1p}^* = \rho_{12}\rho_{23} \ldots \rho_{p-1,p},$$

where ρ_{rs} denotes the marginal correlation of X_r, X_s with edge r, s being present in the graph. It results by taking repeatedly expectations of conditional means in the recursive system or by the matrix argument given at the end of Appendix 1. Provided that most of the ρ_{rs} are not close to one, the implied correlation ρ_{1p}^* tends to zero rapidly with increasing p.

For observations on any given set of variables it can be checked directly whether a parametric cancellation occurs. For such a check, point estimates may be regarded as parameters of an assumed family of distributions and the implied strength of the association of interest can be computed under that model. If (near-)parametric cancellation were observed and especially if repeated in independent studies, a rational explanation in subject-matter terms would be called for. For general discussion and as guidance in specific applications it is helpful to know that there are some general conditions under which parametric cancellation is impossible.

We could show that special Gaussian and other distributions can be generated over directed acyclic graphs such that parametric cancellations do not occur, no matter which conditioning sets are chosen. However, these systems have very specialized features.

Therefore we consider properties of paths instead, such that they imply a conditional association whenever with the separation criterion no corresponding conditional independence can be deduced.

3.2. Stable paths in Markov graphs

Let D denote a subset of V without any nodes along a given path between nodes i and j. Then the path is said to be *stable relative to D* if every edge r, s along it indicates a strictly non-zero conditional association for X_r and X_s after conditioning on X_D.

Thus, in a stable path of the generating graph G_{dag}^V the property of strictly non-zero associations corresponding to each edge present is preserved after conditioning on D, where for each node along the path D may contain some of its ancestors as well as some of its descendants.

The property may again be directly checked in applications. It has the consequence that relative to D there is no parametric cancellation for edges along the given path. Also, whenever there are several paths between i and j in G_{dag}^V which are all stable relative to D, then parametric cancellation can only occur if the effects of the different paths cancel.

As direct consequences we have that there can be no parametric cancellation between i and j relative to a set D if in G_{dag}^V (i) there is just one path between i and j and it is stable relative to D or (ii) there are several paths between i and j which are all collisionless and stable relative to D and the conditional associations corresponding to each edge along the paths can be described as having all the same sign conditionally given D.

Note that in Gaussian systems which are totally positive of order two (Karlin and Rinott 1983) the association of each edge present in the generating graph is positive and never becomes negative, no matter which conditioning set is considered (Bolviken 1982).

4. Relations induced by univariate recursive regressions

In this section we prove the main result, which builds on the available separation criterion for directed acyclic graphs but is formulated to see more directly the stronger statements about associations needed for applications to observational studies. First, however, we define a path to be active because it is correlation inducing in Gaussian systems provided that the path is also stable.

4.1. Active paths and stable active paths

A path between nodes i and j, $i \leqslant j$ in a directed acyclic graph G_{dag}^V is *active relative to C*, $C \subseteq V \backslash \{i, j\}$, if either of the following holds.

 (i) It is collisionless with every node along it outside C.
 (ii) A collisionless path wholly outside C is generated from it by conditioning on C, i.e. by completing with a line the non-adjacent nodes of every sink-oriented V-configuration along it having a descendant in C.

The definition is illustrated in Figure 5. If a generating graph consists just of one of the paths shown in Figure 5a–c, it induces in each of the three cases an identical-looking concentration graph for nodes outside C considered conditionally given C: a single undirected path between i and j.

Note that a path is collisionless if i is a descendant of j (Figure 5a) so that Y_j is directly or indirectly explanatory for Y_i or if there is a source node t which is an ancestor to both i and j (Figure 5b) so that Y_i and Y_j have Y_t as a common explanatory variable. Furthermore, if a collisionless path outside C is induced as illustrated in Figure 5c, then either path end-point variable becomes indirectly explanatory for the other end-point variable conditionally given C. This provides a qualitative justification for the following result.

Let C, $C \subseteq V \setminus \{i, j\}$, be again a conditioning set of interest and consider a path present in G_{dag}^V between i and j which is active relative to C. Further, for all sink-oriented V-configurations along the given path, let C_p denote the set of first descendants which are in C. We say that there is a *stable active path relative to C* if the edges along the given path and along its relevant direction-preserving paths leading into C are stable relative to $C \setminus C_p$. In Figure 5c we have $C = C_p = \{c_1, c_2\}$.

Effect of a stable active path relative to C. *In a system of linear regressions in Gaussian variables generated over G_{dag}^V a stable active path between i and j relative to C provides a strictly non-zero contribution to the correlation coefficient $\rho_{ij.C}$.*

Note that with several stable active paths the contributions all add to the covariance of X_i on X_j given X_C.

Proof. By the assumption of a stable active path between i and j relative to $C \setminus C_p$ there is a non-zero (partial) correlation corresponding to each edge along it and along the paths leading into C after conditioning on the variables in $C \setminus C_p$. We note that C_p is empty if the given path is collisionless and treat the three types of active path in turn.

In the active collisionless path in which i is a descendant of j a non-zero contribution is
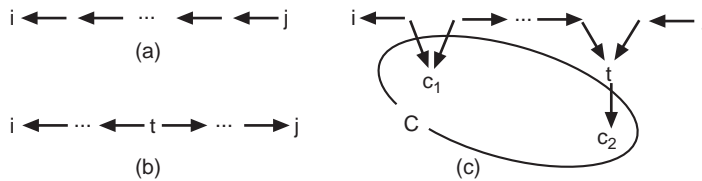

Fig. 5. Examples of the three types of active path, each correlation inducing for pair (i, j) given C provided that it is a stable active path relative to C: (a) collisionless with i descendant of j; (b) collisionless with i and j descendants of t; both paths having all nodes along it outside C; (c) a collision path (with collision nodes c_1 and t) from which a collisionless path wholly outside C is generated by conditioning on C, i.e. by joining with a line the non-adjacent nodes of every sink-oriented V-configuration having a descendant in C.

introduced to $\beta_{ij.C}$ and hence to $\rho_{ij.C}$ by marginalizing over all nodes along the path by the argument of tracing paths given in Appendix 1.

In the active collisionless path with i and j having common source node t a non-zero contribution results by the same argument both in $\beta_{it.C}$ and in $\beta_{jt.C}$. With these two coefficients being non-zero and $\rho_{ij.tC} = 0$, it follows, as explained at the end of Section 2.1, that $\rho_{ij.C} \neq 0$.

For the active collision path we consider the path itself and the relevant paths leading into C conditionally given $D = C \backslash C_p$ so that all nodes along these paths are outside the set D. By marginalizing over the transition nodes along these paths we introduce for each node h of C_p and a pair of nodes (r, s), satisfying $\rho_{rs.D} = 0$, two non-zero correlations $\rho_{hr.D}$ and $\rho_{hs.D}$. This configuration implies in turn, that $\rho_{rs.hD}$ is non-zero. Hence, after conditioning further on all variables of C_p a stable collisionless path is created between i and j relative to C so that by tracing the path a non-zero $\rho_{rs.C}$ is implied.   □

Note that absence of any active path is the condition stated in Section 2.2 as Separation Criterion 1 of Pearl and Verma.

4.2. Relations induced in Gaussian systems generated over G^V_{dag}

In the context of the present paper the key point of the following theorem is the equivalence of (i) and (ii); the relations with (iii) and (iv) are included to provide links with other work and with other graphical representations.

Theorem (induced relations in Gaussian systems). *Let the joint distribution in Gaussian variables $Y_1, \ldots, Y_p$ be generated by linear regressions over G^V_{dag} such that relative to C, $C \subseteq V \backslash \{i, j\}$, every active path is a stable active path and effects of different such paths do not cancel, then the following statements are equivalent.*

   (i) *Y_i depends on Y_j given Y_C.*
  (ii) *In the generating graph G^V_{dag} there is at least one active path between i and j relative to C.*
 (iii) *In the concentration graph of the union of i, j, C, and their ancestors there is outside C at least one path between i and j which is stable relative to C.*
 (iv) *In covariance graphs which include nodes i and j and are considered conditionally given C there is a stable edge i, j relative to C.*

Proof. First, (i) implies (ii). As mentioned before, this has been proven to hold in fact in much greater generality than for Gaussian distributions (Lauritzen *et al*. 1990) by showing that the absence of any active path relative to C implies conditional independence: $Y_i \perp\!\!\!\perp Y_j | Y_C$.

For completeness we give a proof for the special case considered here, for non-degenerate Gaussian systems. Suppose that there is no active path between i and j relative to C. Then there is no edge i, j in the generating graph and the paths possibly connecting i,

j in G_{dag}^V can be characterized as follows: (1) there is at least one collision node t along the path, which is not in C and does not have a descendant in C, or (2) there is at least one transition node t along the path and it is in C, or (3) there is at least one source node t along the path and it is in C.

The first implies for consideration of nodes i, j and set C that we marginalize in a collision path over one collision node and over all its descendants, but then the path end-points become disconnected and no dependence can be induced. If this case is excluded, there remain paths between i and j with at least one transition or source node which is in C. By conditioning on such a node the conditional independence of its direct neighbors is preserved and the path endpoints become again disconnected so that no dependence can be induced for them.

The latter may alternatively be derived from the properties of triangular decompositions of covariance matrices and their inverses or, equivalently, from the rule of tracing paths in linear systems, which in its simplest form is due to Wright (1923) and which is illustrated here in Appendix 1. The rule says that by marginalizing over collisionless paths present in G_{dag}^V a non-zero contribution to $\rho_{ij.C}$ can result only from those paths which do not have a node in C. Hence none of the above-mentioned collisionless paths can introduce a non-zero contribution to $\rho_{ij.C}$. Finally, from $\rho_{ij.C} = 0$ it follows for Gaussian distributions that Y_i and Y_j are independent given Y_C.

Second, to show that (ii) implies (i), we note that a stable active path has a strictly non-zero contribution to $\rho_{ij.C}$. If there are several such paths, their effects combine additively and hence, in the absence of cancellation of effects of different paths, $\rho_{ij.C} \neq 0$ and Y_i and Y_j are dependent given Y_C.

Third, (ii) implies both (iii) and (iv) by construction; see Section 2.3. Then relative to C each of the relevant undirected edges is stable relative to C because it corresponds either to a directed edge present in G_{dag}^V with that property or because it is generated from a stable active path after conditioning on C.

Fourth, (iii) implies (iv). Suppose that in the concentration graph of the union of nodes i, j, C and their ancestors there is a path between i and j outside C which is stable relative to C. Then after marginalizing over all nodes outside C a non-zero contribution to the partial correlation $\rho_{ij.C}$ results so that dependence of Y_i and Y_j given Y_C is implied and there is in any conditional covariance graph given C an edge i, j which is stable relative to C.

Fifth, (iv) implies (i). Suppose that in the induced covariance graph there is an edge i, j which is stable relative to C, then by definition (i) follows and this completes the proof. $\square$

The proof that the absence of any active path implies conditional independence ((i) implies (ii)) does not use parametric cancellation and, indeed, the more general results on d separation show that this notion is irrelevant for deducing conditional independences.

The induced concentration graph of variables in the union of i, j, C as well as their ancestors is the 'moral graph' used by Lauritzen *et al*. (1990) to show separation in the directed graph with the help of separation in an undirected full-line graph; see Separation Criterion 2 in Section 2.2. Lack of separation of non-adjacent i and j by C in such an induced concentration graph is equivalent to the presence of a path between i and j outside

C. Note that the induced overall concentration graph may, in general, not be used in a similar way to judge the absence or presence of separation. The reason is that it can include paths which have a collision node in the generating graph with all its descendants outside *C*. Such a path gets disconnected, i.e. removed, by marginalizing over nodes outside *C* and hence it would not be correlation inducing relative to *C*.

The covariance graph of all variables given *C* can be interpreted as the graphical representation of marginal correlations implied to be zero and non-zero by the generating process for fixed conditions specified with *C*. Thus, if data are in fact generated over G_{dag}^V, then marginal correlations observed under such conditions *C* should be near zero for edges absent in the covariance graph and non-zero otherwise, unless there is parametric cancellation.

4.3. Relations induced in quasi-linear systems generated over G_{dag}^V

To extend the usefulness of the above theorem we introduce the notion of quasi-linear dependence. For this we use the result that any conditional or marginal dependence can be written as a linear least-squares regression plus an error term, uncorrelated with the explanatory variables; this contrasts strongly with the assumption of linear regression with an independent error term which is a major restriction on the form of the dependence.

By *quasi-linear dependence* we mean (1) that any dependence present has a linear component, i.e. the vanishing of a partial least-squares regression coefficient implies exactly or to a close approximation a conditional independence statement about the conditional mean, and (2) a conditional independence statement implies that the corresponding partial correlation coefficient does not deviate much from zero. By this we essentially exclude dependences that are so curved or involve such high-order interactions that they lead to vanishing correlation and systems in which a strong partial correlation is present in spite of conditional independence (for an example of the latter situation, see Wermuth and Cox (1998)).

As a special case take a regression of *Y* on explanatory variables *X* and *Z*, where *X* is a scalar and *Z* a vector random variable and assume that $E(Y) = g\{\beta X, h(Z, \gamma)\}$ where $g(.)$ is strictly increasing in its first argument, $h(.)$ is arbitrary and γ is a vector of parameters. We want the partial correlation of *Y* with *X* given *Z* to have the same sign as β and to vanish if and only if $\beta = 0$. It is proved in Appendix 2 that this holds if given *Z* the expected value of either *Y* or *X* is linear in *Z*.

There are several consequences of an assumption of quasi-linear dependence. If in the directed acyclic graph there is an arrow to *i* from *j*, and no other edge to *i*, then $Y_i = \beta_{ij} Y_j + \epsilon_{ij}$, where $\beta_{ij} \neq 0$ and the error ϵ_{ij} is uncorrelated with Y_j (but in general not independent of Y_j) and with variables directly or indirectly explanatory to Y_j. Further, we may reverse the direction of a single arrow and write $Y_j = \beta_{ji} Y_i + \epsilon_{ji}$, where the error is uncorrelated with Y_i and with the variables directly or indirectly explanatory to Y_j. Finally, suppose that two or more arrows point to *i*, say from *j*, *j'*, *j''* and so on. In general the effects would combine nonlinearly, but in a quasi-linear distribution we may take a linear

least-squares regression with non-zero coefficients of the contributing variables and with uncorrelated error.

As mentioned before, in actual applications it may be checked directly whether such paths are stable relative to a conditioning set of interest. If they are, then every stable active path between i and j relative to C induces a dependence between Y_i and Y_j given Y_C just as in a Gaussian system.

Appendix 1: Tracing paths present in the generating graph

We summarize here some known results for triangular decompositions (A, D) of a concentration matrix Σ^{-1} and (B, T) of a covariance matrix Σ and show, in addition, how in linear systems generated over G_{dag}^V all elements of B may be read off the generating directed acyclic graph. This involves tracing collisionless paths in G_{dag}^V relative to a given conditioning set C and generalizes Wright's (1923) result for 'implied correlations' which involves tracking collisionless paths in the case that C is the empty set.

Every joint distribution generated with linear main-effect regressions over G_{dag}^V in responses ordered as $(Y_1, \ldots, Y_{p-1})$ has a covariance matrix Σ and a triangular decomposition of its inverse $\Sigma^{-1} = A^{\text{T}} D A$, such that every edge i, j $(i < j)$ missing in the generating graph G_{dag}^V corresponds to a zero element a_{ij} of A and every edge present corresponds to a strictly non-zero off-diagonal element of A. The reason is that the linear independence statement connected to the missing edge i, j implies a zero partial regression coefficient $\beta_{ij.i+1,\ldots,j-1,j+1,\ldots,p}$ and element a_{ij} of A is a multiple of this: $a_{ij} = -\beta_{ij.i+1,\ldots,j-1,j+1,\ldots,p}$ (Wermuth 1980).

An element of B in the triangular decomposition of the covariance matrix $\Sigma = BTB^{\text{T}}$, where $B = A^{-1}$ and $T = D^{-1}$, has a less direct interpretation although it is a particular linear least-squares regression coefficient as well (Dempster 1969; Wermuth 1989, 1992). In column j of B we have for rows $h = 1, \ldots j - 1$ coefficients obtained by regressing X_h on X_j conditional on $X_{j+1}, \ldots X_p$, i.e. $b_{hj} = \beta_{hj.j+1,\ldots,p}$ for $h = 1, \ldots, j - 1$. The ith element of the diagonal matrix T is the residual variance $\sigma_{ii.i+1,\ldots,p}$ corresponding to the regression equation of Y_i specified implicitly with row i of A.

Example. For $p = 4$ variables and a complete generating graph, the elements of A and B have the following explicit form in terms of regression coefficients:

$$
A = \begin{pmatrix} 1 & -\beta_{12.34} & -\beta_{13.24} & -\beta_{14.23} \\ 0 & 1 & -\beta_{23.4} & -\beta_{24.3} \\ 0 & 0 & 1 & -\beta_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & \beta_{12.34} & \beta_{13.4} & \beta_{14} \\ 0 & 1 & \beta_{23.4} & \beta_{24} \\ 0 & 0 & 1 & \beta_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

If we simplify the notation to denote the generating regression coefficients in A by $\gamma_{ij} = \beta_{ij.i+1,\ldots,j-1,j+1,\ldots,p}$, then direct computations show that the elements of B can be expressed as follows:

$$B = \begin{pmatrix} 1 & \gamma_{12} & \gamma_{13} + \gamma_{12}\gamma_{23} & \gamma_{14} + \gamma_{12}\gamma_{24} + \gamma_{13}\gamma_{34} + \gamma_{12}\gamma_{23}\gamma_{34} \\ 0 & 1 & \gamma_{23} & \gamma_{24} + \gamma_{23}\gamma_{34} \\ 0 & 0 & 1 & \gamma_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In general, the conditioning set of a regression coefficient in position (i, j) of B, for $i < j$ is the set $\{j + 1, \ldots, p\}$.

Tracing paths to compute the inverse B of the triangular matrix A. *Element b_{ij} of B, for $j > i = 1, \ldots, p - 1$, results by tracing all collisionless paths present in G_{dag}^V between i and j outside $\{j + 1, \ldots, p\}$.*

For instance, position $(1, 3)$ of the above matrix B contains the regression coefficient $\beta_{13.4}$. Its defining conditioning set is $\{4\}$ and, in a complete generating graph of four nodes, there are two additive contributions to the regression coefficient $\beta_{13.4}$: one is γ_{13} from the edge $(1, 3)$ present; the other is from the collisionless path from 1 to 3 via node 2 not touching node 4. Position $(1, 4)$ of B contains the marginal coefficient β_{14}. Its defining conditioning set is empty and in a complete generating graph there are additive contributions of four collisionless paths between 1 and 4.

If the system is incomplete in such a way that all elements $a_{ij} = 0$, for $j > i + 1$, i.e. if there is just one path along nodes $1, 2, \ldots, p$, then there is still a non-zero contribution to correlation ρ_{1p} provided that each element $a_{i,i+1}$ of A for $i = 1, \ldots, p - 1$ is non-zero, i.e. provided that every edge along the path has a corresponding non-zero partial correlation coefficient associated with it.

Appendix 2: Increasing dependence and partial correlation

Here we show that in a quasi-linear system of special nonlinear form, in which β captures the conditional dependence of Y on X given Z, the partial correlation of Y with X given Z has the same sign as β and vanishes if and only if $\beta = 0$.

Consider the regression of Y on explanatory random variables X and Z, where X is a scalar and Z is a vector. We assume that

$$E(Y|X, Z) = g\{\beta X, h(Z, \gamma)\},$$

where $g(\cdot)$ is strictly increasing in its first argument, $h(\cdot)$ is arbitrary and γ is a vector of parameters.

Write $R_{Y.Z}$, $R_{X.Z}$ for the residuals in the linear least-squares regressions of Y on Z and of X on Z, respectively.

The partial correlation of Y and X given Z is proportional to the simple covariance of the residuals $R_{Y.Z}$ and $R_{X.Z}$ which we can write as

$$\mathrm{cov}\,(R_{Y.Z}, R_{X.Z})\} = \mathrm{E}_Z\mathrm{cov}\,(R_{Y.Z}, R_{X.Z}|Z) + \mathrm{cov}_Z\,\{\mathrm{E}(R_{Y.Z}), \mathrm{E}(R_{X.Z})|Z\}.$$

The second term vanishes if either Y or X has linear regression on Z alone, making one

of the expectations involved zero. More generally, it vanishes if the nonlinearities in the dependence of Y on Z and of X on Z are uncorrelated. We assume one of these conditions holds.

To calculate the first term, we assume further that we can write

$$g\{\beta X, \, h(Z, \gamma)\} = g\{\beta\{X - \mathrm{E}(X|Z)\}, \, h(Z, \gamma^*)\},$$

where γ^* is a new value of the parameter vector. This will hold if, for example, $g(u, v) = g(u + \alpha v)$ and $\mathrm{E}(X|Z)$ is linear in Z and $h(Z, \gamma)$ includes the family of linear functions as possible special cases.

Then, because $\mathrm{cov}(R_{Y, Z}, R_{X, Z}) = \mathrm{cov}(Y, X|Z)$

$$\mathrm{E}_Z \, \mathrm{cov}\,(Y, \, X|Z) = \mathrm{E}_Z \mathrm{E}_{Y, X|Z}\{Y\{X - \mathrm{E}(X|Z)\}|Z\}$$

$$= \mathrm{E}_Z \mathrm{E}_{X|Z} \, g\{\beta\{X - \mathrm{E}(X|Z)\}, \, h(Z, \gamma^*)\}\{X - \mathrm{E}(X|Z\}.$$

The inner expectation vanishes at $\beta = 0$ and its partial derivative with respect to β, being the expectation of

$$\{X - \mathrm{E}(X|Z)\}^2 \frac{\partial g}{\partial \beta},$$

is strictly positive. Therefore the conditional covariance of Y and X given Z and the partial correlation, $\rho_{yx.z}$, vanish if and only if $\beta = 0$.

Note that in this discussion the monotonic function involved may be a logistic function. Suppose for example that Y is a binary random variable, which we denote by A, taking values $1, -1$ and having logistic regression on two scalar random variables B, Z where B is binary and Z is continuous. If there are no interactive and no nonlinear effects we may write

$$E(A|B, \, Z) = g\{\beta B + (\gamma_0 + \gamma_1 Z)\},$$

where $g(u) = \tanh{(u/2)}$. The model is that of two binary regressions parallel on the logistic scale. Suppose further that B has on Z a regression that is linear in the logistic scale. Then, the condition that $\mathrm{E}(B|Z)$ is linear in Z can be exactly true only if Z has finite range. It holds approximately if $E(A|B, Z)$ depends only on Z; then the partial correlation between A and B given X deviates from 0 by at most 0.01 provided that the logistic regression coefficients of A on Z and of B on Z are less than 1, corresponding to correlation coefficients of less than 0.41 for variables standardized to have means zero and variances one.

Acknowledgements

References

Anderson, T.W. (1984) *Introduction to Multivariate Statistical Analysis*, 2nd edn. New York: Wiley.

Bolviken, E. (1982) Probability inequalities for the multivariate normal with nonnegative partial correlations. *Scand. J. Statist.*, **9**, 49–58.

Cochran, W.G. (1938) The omission or addition of an independent variate in multiple linear regression. *J. Roy. Statist. Soc., Suppl.*, **5**, 171–176.

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

Cox, D.R. and Wermuth, N. (1993) Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.*, **8**, 204–218; 247–277.

Cox, D.R. and Wermuth, N. (1994) A note on the binary quadratic exponential distribution. *Biometrika*, **81**, 403–408.

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

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980) Markov fields and log-linear models for contingency tables. *Ann. Statist.*, **8**, 522–539.

Dempster, A.P. (1969) *Elements of Continuous Multivariate Analysis*. Reading, Massachusetts: Addison-Wesley.

Edwards, D. (1995) *Introduction to Graphical Modelling*. New York: Springer-Verlag.

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

Karlin, S. and Rinott, Y. (1983) *M*-matrices as covariance matrices of multinormal distributions. *Linear Algebra Applic.*, **52–53**, 419–438.

Kauermann, G. (1996) On a dualization of graphical Gaussian models. *Scand J. Statist.*, **23**, 105–116.

Kiiveri, H.T., Speed, T.P. and Carlin, J.B. (1984) Recursive causal models. *J. Austral. Math. Soc. A*, **36**, 30–52.

Lauritzen, S.L. (1996) *Graphical Models*. Oxford: Oxford University Press.

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

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

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.

Pearl, J. and Verma, T. (1988) The logic of representing dependencies by directed graphs. *Proceedings of the American Association of Artificial Intelligence Conference*, Seattle, Washington.

Pearl, J. and Wermuth, N. (1994) When can association graphs admit a causal interpretation? In P. Cheseman and W. Oldford (eds), *Models and Data, Artificial Intelligence and Statistics IV*, pp. 205–214. New York: Springer-Verlag.

Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction and Search*. New York: Springer-Verlag.

Wermuth, N. (1976) Analogies between multiplicative models for contingency tables and covariance selection. *Biometrics*, **32**, 95–108.

Wermuth, N. (1980) Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.*, **75**, 963–997.

Wermuth, N. (1989) Moderating effects in multivariate normal distributions. *Methodika*, **3**, 74–93.

Wermuth, N. (1992) On block-recursive linear regression equations (with discussion). *REPRABE (Brazilian J. Probab. Statist.)*, **6**, 1–56.

Wermuth, N. (1995) On the interpretation of chain graphs. *Proceedings of the 50th Session of International Statistical Institute*, Beijing 1995, Book 1, pp. 415–429.

Wermuth, N. (1997) Graphical Markov models. In S. Kotz, C. Read and D. Banks (eds), *Encyclopedia of Statistical Sciences*. New York: Wiley. To appear.

Wermuth, N. and Cox, D.R. (1998) Statistical dependence and independence. In P. Armitage and T. Colton (eds), *Encyclopedia of Biostatistics*. New York: Wiley. To appear.

Wermuth, N. and Lauritzen, S.L. (1983) Graphical and recursive models for contingency tables. *Biometrika*, **70**, 537–552.

Wermuth, N. and Lauritzen, S.L. (1990) On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. Roy. Statist. Soc. B*, **52**, 21–72.

Wright, S. (1923) The theory of path coefficients: a reply to Niles' criticism. *Genetics*, **8**, 239–255.