PROBABILITY DISTRIBUTIONS WITH SUMMARY GRAPH STRUCTURE

By Nanny Wermuth*

Mathematical Statistics at Chalmers/ University of Gothenburg, Sweden

A set of independence statements may define the independence structure of interest in a family of joint probability distributions. This structure is often captured by a graph that consists of nodes representing the random variables and of edges that couple node pairs. One important class are multivariate regression chain graphs. They describe the independences of stepwise processes, in which at each step single or joint responses are generated given the relevant explanatory variables in their past. For joint densities that then result after possible marginalising or conditioning, we use summary graphs. These graphs reflect the independence structure implied by the generating process for the reduced set of variables and they preserve the implied independences after additional marginalising and conditioning. They can identify generating dependences which remain unchanged and alert to possibly severe distortions due to direct and indirect confounding. Operators for matrix representations of graphs are used to derive these properties of summary graphs and to translate them into special types of path in graphs.

1. Motivation, some previous and some of the new results.

1.1. *Motivation.* Graphical Markov models are probability distributions defined for a $d_V \times 1$ random vector variable Y_V whose component variables may be discrete or continuous and whose joint density f_V satisfies the independence statements specified directly as well as those implied by an associated graph. The set of all such statements is the independence structure captured by the graph.

One such type of graph had been introduced for multivariate regression chains by Cox and Wermuth (1993, 1996) for which special results have been derived by Drton (2009), Kang and Tian (2009), Marchetti and Lupparelli (2010), Wermuth and Cox (2004), Wermuth, Wiedenbeck and Cox (2006), Wermuth, Marchetti and Cox (2009).

*Supported in part by the Swedish Research Society via the Gothenburg Stochastic Center and by the Swedish Strategic Fund via the Gothenburg Math. Modelling Center

AMS 2000 subject classifications: Primary 62H99; secondary 62H05, 05C50

Keywords and phrases: Concentration graph, Directed acyclic graph, Endogenous variables, Graphical Markov model, Independence graph, Multivariate regression chain, Partial closure, Partial inversion, Triangular system.

A multivariate regression chain graph consists of nodes in set, V, that represent random variables and of edges that couple node pairs such that a recursive order of the joint responses is reflected in the graph and each defining independence constraint respects the given ordering; see Marchetti and Lupparelli (2010). This distinguishes multivariate regression graphs from all other currently known types of chain graphs; see Drton (2009) for implications on associated discrete distributions.

Because of this property, multivariate regression chain graphs are particularly well suited for studies of effects of hypothesized causes on joint responses, see Cox and Wermuth (2004), and more generally for modeling developmental processes, such as in panel studies. These provide data on a group of individuals, termed the 'panel', collected repeatedly, say over years or decades. Often one wants to compare corresponding analyses with results in other studies that have core sets of variables in common, but that have omitted some of the variables or that were carried out for subpopulations.

It is the outstanding feature of multivariate regression chains that consequences of a model can be derived, for instance regarding implications after marginalizing over some variables, in set M, or after conditioning on others, in set C. In particular, graphs can be obtained for node set $N = V \setminus \{C, M\}$ which capture precisely the independence structure implied by a generating graph in node set V for the distribution of Y_N given Y_C, the distribution of interest for the variables in the reduced node set N.

Such graphs are named independence-preserving, when they can be used to derive the independence structure that would have resulted from the generating graph by conditioning on a larger node set $\{C, c\}$ or by marginalising over a larger node set $\{M, m\}$. Two types of such classes are a subclass of the much larger class of MC-graphs of Koster (2002) and the typically more compact MAGs (maximal ancestral graphs) of Richardson and Spirtes (2002). We speak of two corresponding graphs if they result from a given generating graph relative to the same sets C, M.

A third class of this type are the summary graphs of Wermuth, Cox and Pearl (1994). This class is presented in the current paper in simplified form together with proofs based on operators for binary matrix representations of the graphs. In contrast to a MAG, a corresponding summary graph can be used to identify those dependences of a given generating process for families of distributions of Y_V that remain unchanged and those generating dependences that may be severely distorted in the MAG model for Y_N given Y_C. This is especially helpful at the planning stage of new studies when several alternative sets M and C are considered that may result from a given, much larger generating graph. The annotated, undirected graphs of Paz (2007), defined for $C = \emptyset$, serve a similar purpose.

The warning signals provided by summary graphs are essential for understanding consequences of a given data generating process with respect to dependences in addition to independences. For this, some special properties of the types of generating graph will be introduced as well as specific requirements on the types of generating process. These lead to families of distributions that are said to be generated over parent graphs.

1.2. *Some notation and concepts.* Some definitions for graphs are almost self-explanatory. If pair $i \neq k$ of V is coupled by a directed edge such that an arrow starts at node k and points to node i, then k is named a parent of i and i the offspring of k. For two disjoint subsets α and β of V, an ik-arrow, $i \prec\!\!-\!\!-\, k$, is said to point from β to α if the arrow starts at a node k in β and points to a node i in α. For three or more nodes, an ik-path connects the path endpoint nodes i and k by a sequence of edges that couple nodes. Nodes other than the endpoint nodes are the inner nodes of a path; only the latter have to be distinct.

An edge is regarded as a path without inner nodes. Both a graph and a path, are called directed if all its edges are arrows. If all arrows of a directed ik-path point towards node i, then node k is an ancestor of i and i a descendant of k. Such a path is direction-preserving and called a descendant-ancestor path.

Directed acyclic graphs form an important subclass of MRC graphs. They arise from recursive stepwise generating processes of exclusively univariate response variables, see also Section 2 below. These graphs have no directed cycles, that is they have no descendant-ancestor ik-path such that $i = k$.

As we shall show, two different types of undirected graph are also subclasses of MRC graphs that can be derived from directed acyclic graphs. For joint Gaussian distributions, they give models for zero constraints on covariances and on concentrations, see e.g. Wermuth and Cox (1998) and (2.5), (2.16) below. To distinguish between them in figures, edges in concentration graphs are shown as full lines, $i \,\text{——}\, k$, and in covariance graphs by dashed lines, $i \text{---} k$.

Separation criteria provide what is called the global Markov property of a graph which in turn gives all independence statements that belong to the graph's independence structure.

Definition 1. *A graph, consisting of a node set and of one or more edge sets, is an independence graph if node pairs are coupled by at most one edge and each missing edge corresponds to at least one independence statement.*

MRC graphs and MAGs are independence graphs but, in general, summary and MC graphs are not, even with at most one edge for each node pair; see the discussion of Figure 3*b)* below.

The same graph theoretic notion of separation applies to both types of undirected graph. Let α and β be two nonempty, disjoint subsets of their node set V and let $\{\alpha, \beta, m, c\}$ partition V, then we write Y_α is conditionally independent of Y_β given Y_c compactly as $\alpha \perp\!\!\!\perp \beta | c$. In a concentration graph, α is separated by c from β if every path from α to β has a node in c, while in the covariance graph, α is separated by m from β if every path from α to β has a node in m. Given separation of α and β by set c, a concentration graph implies $\alpha \perp\!\!\!\perp \beta | c$; see Lauritzen (1996). Given separation of α and β by set m, a covariance graph implies $\alpha \perp\!\!\!\perp \beta | c$; see Kauermann (1996), who expresses the result in a different but equivalent way.

When a graph is directed or contains different types of edge then its separation criterion is more complex than the one for undirected graphs. For directed acyclic graphs, there are several different separation criteria that permit to obtain all independence statements implied by the graph, see Marchetti and Wermuth (2009) for proofs of equivalence.

The criterion due to Geiger, Verma and Pearl (1990), has been extended in almost unchanged form by Koster (2002) to the much larger class of MC-graphs. A path-based proof, due to Sadeghi (2008), is for the subclass of MC-graphs that is of interest here, the graphs that can be derived from a directed acyclic graph. For summary graphs, it is stated below as Lemma 1.

A list of independence statements associated with missing edges of an independence graph is the graph's pairwise Markov property. Whenever it defines the graph's independence structure, then the pairwise Markov property is said to be equivalent to the global Markov property.

For all disjoint subsets a, b, c, d of node set V, the following general definitions are relevant, respectively, for combining pairwise independences in covariance graph and in concentration graph models

DEFINITION 2. *the composition property:*

$$a \perp\!\!\!\perp b | d \text{ and } a \perp\!\!\!\perp c | d \text{ imply } a \perp\!\!\!\perp bc | d \,,$$

DEFINITION 3. *the intersection property:*

$$a \perp\!\!\!\perp b | cd \text{ and } a \perp\!\!\!\perp c | bd \text{ imply } a \perp\!\!\!\perp bc | d \,.$$

Given these properties, the independence structure of interest in a covariance or concentration graph model can be specified in terms of independence constraints on a set of variable pairs. For general searching discussions; see Dawid (1979), Pearl (1988), Lauritzen (1996), Studený (2005).

Necessary and sufficient conditions under which discrete and Gaussian distributions satisfy the intersection property, have been derived by San Martin, Mouchart and Rolin (2005). They show in particular that of the commonly specified sufficient conditions, some may be much too strong, for instance requiring exclusively positive probabilities for discrete distributions. For joint Gaussian distributions, a positive definite joint covariance matrix is sufficient. In both cases, no component of the involved random variables is degenerate.

Definition 4. *A family of joint distributions is said to vary fully if its random variables contain no degenerate components and it satisfies the intersection property.*

In distributions without the composition property, there may be subsets of variables with pairwise but no mutual independence. For families of joint distributions with the composition property that are associated with a MRC graph, the global and the pairwise Markov property are equivalent; see Kang and Tian (2009).

For a long time, only the family of Gaussian distributions was known to satisfy both the composition and the intersection property provided it varies fully. Under the same type of constraint, this is now known to hold for the special family of distributions in symmetric binary variables introduced by Wermuth, Marchetti and Cox (2009).

The notion of completeness had been introduced and studied in quite different contexts; see Lehmann and Scheffé (1955), Brown (1986) Theorem 2.12, and Mandelbaum and Rüschendorf (1987). It means that the joint family of distribution of vector variable Y is such that a zero expectation of any function $g(y)$ implies that the function itself is zero with probability one, that is almost surely (a.s.).

Definition 5. *Let $f(y)$ denote the density of a member of a complete family of distributions and $g(y)$ be some function of Y, then it holds for every $f(y)$ that*

$$\int g(y)f(y)\,dy = 0 \implies g(y) = 0 \; a.s. \,.$$

For any trivariate family of distributions with precisely two associated variable pairs, say (Y_1, Y_2) and (Y_1, Y_3), but $2 \perp\!\!\!\perp 3$, completeness of the joint distribution is sufficient to conclude that Y_2 is conditionally dependent on

Y_3 given Y_1. This follows from Corollary 3 of Wermuth and Cox (2004) and properties of completeness. In this situation, the generating graph

$$2 \longrightarrow 1 \longleftarrow 3$$

is inducing a 23-edge in the summary graph obtained by conditioning on node 1 and a non-vanishing conditional association for Y_2, Y_3 given Y_1.

In Section 2, we define parent graphs as directed acyclic graphs with special properties and corresponding types of stepwise generating processes such that edge-inducing paths are also association-inducing. The families of distributions generated over parent graphs are complete and each member of the families satisfies the intersection and the composition property in addition to the general laws of probability that govern independences in any joint family of distribution; for a discussion of the latter see Studený (2005).

1.3. *Definition and construction of summary graphs.* In contrast to MC-graphs and MAGs, MRC graphs are not closed under marginalising and conditioning, that is one can get with a MRC graph outside the given class after marginalizing and conditioning; as illustrated with Figure 3 below. But the graph resulting in this way from a MRC graph is always within the class of summary graphs. This explains partly why we consider the larger class of summary graphs.

DEFINITION 6. *A summary graph, G^N_{sum}, has node set N which consists of two disjoint subsets u, v, ordered as (u, v). The graph has a mixture of a directed acyclic graph and of a covariance graph within u and a concentration graph within v. Between u and v, only arrows point from v to u.*

The notions of parents, offsprings, ancestors and descendants remain unchanged in a summary graph compared to a directed acyclic graph. As will be shown, every summary graph in node set N can be generated from a directed acyclic graph in node set $V = \{\circ\}$ by conditioning on $C = \{\boxed{\circ}\}$ and marginalising over $M = \{\not\circ\}$ so that $N = V \setminus \{C, M\}$. This graph is denoted by $G^{V\setminus[C,M]}_{\mathrm{sum}}$, an associated density by $f_{N|C}$ which results from f_Y, the given density of the generating graph, which factorizes according to this graph; see (1.3) below.

The density $f_{N|C}$ may concern discrete, continuous or mixed variables, as implied by f_V. It has a factorization according to (u, v) which is written compactly in terms of node sets as

$$f_{N|C} = f_{u|vC} f_{v|C}\,. \tag{1.1}$$

In the larger generating graph in node set V, every node in v and no node in u is an ancestor of the conditioning set C. Thus, each component of Y_v has been generated before Y_u; see Figure 2 for an example.

Figures 1 to 3 illustrate how summary graphs may be generated. For this, the stepwise construction of a summary graph by marginalizing over $m = \{t\}$ or conditioning on $c = \{j\}$ in G^N_{sum} is given in Table 1.

If a node t is coupled to both of the nodes i and k then t is said to be their common neighbor. In two-edge paths, the inner node is named a collision node, s, for

$$\circ \!\longrightarrow\! s \!\longleftarrow\! \circ, \qquad \circ \!\longrightarrow\! s \,{-}{-}{-}\, \circ \qquad \circ \,{-}{-}{-}\, s \,{-}{-}{-}\, \circ,$$

and a transmitting node, t, otherwise. A path for which all inner nodes are collision nodes is a collision path and a path for which all inner nodes are transmitting nodes is a transmitting path.

TABLE 1

Types of induced edge when each of m or c contains a single node in G^N_{sum}.

Types of induced edge when marginalizing over the common neighbor node t

	$t \longrightarrow \circ$	$t \!-\!\!-\!\!-\! \circ$	$t \longleftarrow \circ$	$t \,{-}{-}{-}\, \circ$
$\circ \longleftarrow t$	$\circ \,{-}{-}{-}\, \circ$	$\circ \longleftarrow \circ$	$\circ \longleftarrow \circ$	$\circ \,{-}{-}{-}\, \circ$
$\circ \!-\!\!-\!\!-\! t$	·	$\circ \!-\!\!-\!\!-\! \circ$	$\circ \!-\!\!-\!\!-\! \circ$	$\circ \longrightarrow \circ$

and types of induced edge when conditioning on the common neighbor node s or on one of the descendants of s

	$s \longleftarrow \circ$	$s \,{-}{-}{-}\, \circ$
$\circ \longrightarrow s$	$\circ \!-\!\!-\!\!-\! \circ$	$\circ \longrightarrow \circ$
$\circ \,{-}{-}{-}\, s$	·	$\circ \,{-}{-}{-}\, \circ$

where the · notation indicates a symmetric entry.

Table 1 is taken from Wermuth, Cox and Pearl (1994); see also Appendix A here. It implies that a collision node is edge-inducing by conditioning on it while a transmitting node is edge-inducing by marginalising over it.

Let now a summary graph, G^N_{sum}, be given and nodes $j \neq t$ of N be selected. Suppose one intends to marginalize over node t and to condition on node j and d_j denotes the ancestors of j within u of G^N_{sum}. Then, a new summary graph in node set $N' = N \setminus \{j, t\}$ results by use of Proposition 1. It has its concentration graph in $v' = v \setminus \{j, t\}$ for both nodes in v, in $v' = v \setminus \{j\}$ for only j in v, in $v' = \{v \setminus \{t\}, d_j\}$ for only t in v and in $v' = \{v, d_j\}$ for both nodes in u.

Proposition 1. Generating a summary graph from G^N_{sum} by operating on at most two nodes. *From G^N_{sum}, the independence-preserving summary graph $G^{N\setminus[j,t]}_{\text{sum}}$ is generated, with t the marginalising node and j the conditioning node, by inducing edges as prescribed in Table 1*
(1) first for the neighbors of t, second for the neighbors of both j and of all of its ancestors (ignoring in the second step additional edges with t) or, reversing the first and the second step (and thereby ignoring in the second step additional edges involving j),
(2) changing each edge present within v' into a full line and each edge present between u' and v' into an arrow pointing from v' to u',
(3) keeping for each node pair of several edges that are of the same kind just one and deleting all nodes and edges involving j or t.

See Section 3 for proofs in terms of operators for matrix representations of graphs. The proofs imply for any node subset $\{m, c\}$ of N that $G^{N\setminus[\emptyset,m]}_{\text{sum}}$ may be derived before conditioning on set c, or $G^{N\setminus[c,\emptyset]}_{\text{sum}}$ before marginalizing over set m and that within sets c or m any order of the nodes can be chosen.

The matrix formulations lead more directly to $G^{N\setminus[c,m]}_{\text{sum}}$, but Proposition 1 gives an algorithm for operating on one node at a time. It is also helpful for small graphs as illustrated below with Figures 1 to 3. Proposition 1 implies that no coupled pair gets ever uncoupled and that the two types of path which may occur when constructing a summary graph are replaced in $G^{N\setminus[j,t]}_{\text{sum}}$:

$$\circ \longrightarrow \circ \;\text{——}\; \circ \quad\text{by}\quad \circ \;\text{——}\; \circ \;\text{——}\; \circ,$$

$$\circ \;\text{- - -}\; \circ \;\text{——}\; \circ \quad\text{by}\quad \circ \longleftarrow \circ \;\text{——}\; \circ.$$

The starting summary graph of Figure 1 is in 1*a)*. For $j = 5$ and $t = 4$, Figure 1*b)* shows the edges induced by operating first on j, Figure 1*c)* those induced by operating first on t and Figure 1*d)* displays $G^{N\setminus[5,4]}_{\text{sum}}$.

By construction, a summary graph contains no directed cycle, but possibly mixed directed cycles. These are direction-preserving ik-paths with $i = k$ that contain some undirected edges; see Figure 3*b)* for examples.

Corollary 1. MRC and summary graphs. *A multivariate regression chain graph is a summary graph without mixed directed cycles.*

In contrast to a summary graph, a MRC graph is an independence graph graph which has at most one edge coupling any node pair; compare Figures 2*b)* and 3*b)*. Figure 2*b)* shows a MRC graph generated from a directed acyclic graph and Figure 3*b)* a summary graph with mixed cycles.

FIG 1. *a) A summary graph with node 4 to be marginalized over and node 5 to be conditioned on, b) the graph of a) including edges induced for conditioning on node 5, c) the graph of a) including edges induced for marginalising over node 4, d)* $G_{\mathrm{sum}}^{N\setminus[5,4]}$.

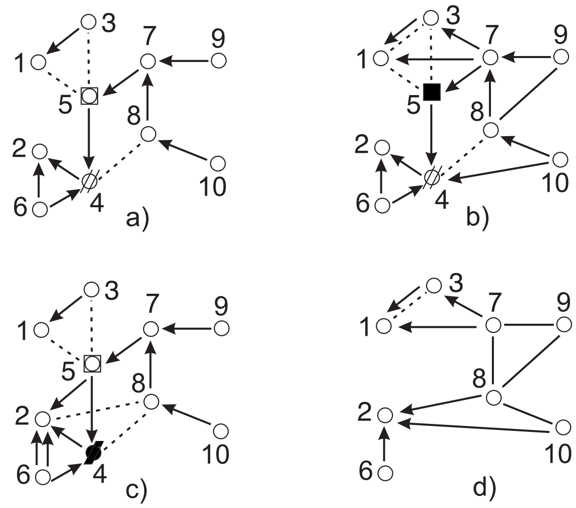
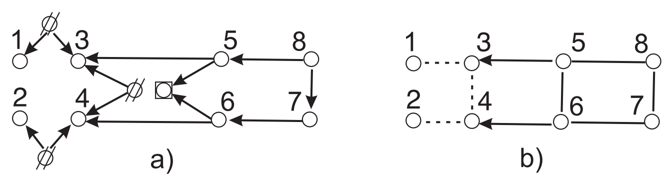
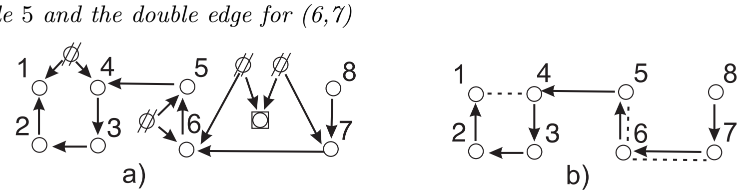


FIG 2. *a) A directed acyclic graph generating b) a summary graph without mixed cycles;* $u = \{1,2,3,4\}$ *and* $v = \{5,6,7,8\}$.



By replacing each dashed ik-edge by an ik-path $i \longleftarrow \not\!\emptyset \longrightarrow k$, every summary graph has a virtual generating directed acyclic graph for the nodes within u even though a dashed line might actually have been generated by over-conditioning, i.e. by including an offspring into the conditioning set of two of its parents; see for example $\not\!\emptyset \longrightarrow \boxed{\circ} \longleftarrow \not\!\emptyset$ as the inner nodes of the 6,7-path in Figure 3*a)*.

FIG 3. *a) A directed acyclic graph generating b) a summary graph with v the empty set and several mixed directed cycles; the 4,4-path with inner nodes* 1, 2, 3*, the 6,6-path via inner node* 5 *and the double edge for (6,7)*



Similarly, each chordless cycle within v may be generated by including additional nodes, $\boxed{\circ}$ and $\not\!\emptyset$, in appropriate ways; see Cox and Wermuth

(2000). The summary graph is uniquely defined if generated from a directed acyclic graph in node set V for given sets M, C, but typically many different directed acyclic graphs in node set V, or having nodes added to V, may lead to the same summary graph.

1.4. *Independence interpretation of summary graphs.* A criterion to decide whether a given summary graph, $G^{V\setminus[C,M]}_{\mathrm{sum}}$, implies $\alpha \perp\!\!\!\perp \beta | c\, C$ is given next. For this, the node set N is partitioned as $N = \{\alpha, \beta, c, m\}$ where only subsets c or m may be empty, for $a \subset V$.

LEMMA 1. Koster (2002), Sadeghi (2008). Path criterion for the global Markov property. *The graph $G^{V\setminus[C,M]}_{\mathrm{sum}}$ implies $\alpha \perp\!\!\!\perp \beta | c\, C$ if and only if it has no ik-path between α and β such that of its inner nodes every collision node is in c or has a descendant in c and every other node is outside c.*

In addition to the directly described path, Lemma 1 specifies implicitly many special types of forbidden path. We name a path of $n > 2$ nodes an a-line path if all inner nodes are within set a. The marginalising set is defined by $m = N \setminus \{\alpha, \beta, c\}$. Then, in $G^{V\setminus[C,M]}_{\mathrm{sum}}$ there should be for node i in α and node j in β no ik-edge, no m-line transmitting ik-path, no c-line collision ik-path, no ik-path with all inner transmitting nodes in m and all inner collision nodes in c.

COROLLARY 2. Active ik-paths. *An ik-path in G^{N}_{sum} is active relative to $[c, m]$ if and only if it is an ik-edge or every inner transmitting node is in m and every inner collision node is in c or has a descendant in c.*

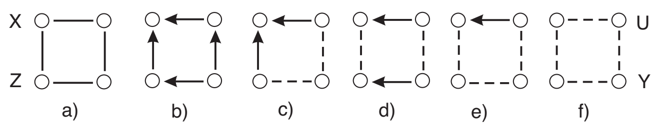
If an active ik-path relative to $[c, m]$ has uncoupled endpoints, the path is closed by an ik-edge in $G^{N\setminus[c,m]}_{\mathrm{sum}}$. If an active ik-path has coupled endpoints, the path is edge-inducing in the construction process of $G^{N\setminus[c,m]}_{\mathrm{sum}}$. Thus, we often replace 'active' by the more concrete term 'edge-inducing'. Figure 2*b)* represents a multivariate regression chain graph, hence each missing edge corresponds to at least one independence statement. This contrasts with Figure 3*b)*, which has three mixed directed cycles with an arrow starting at nodes 4, 5 and 6, respectively, and no independence statement implied for pairs (1,5), (5,7), (5,8), (6,8). For pair $(1, 5)$, we give more detailed arguments.

In the graph of Figure 2*b)*, node 3 has no descendants and is an inner collision node in every path connecting 1 and 5. Hence, when node 3 is marginalized over, $1 \perp\!\!\!\perp 5 | C$ is implied. In the graph of Figure 3*b)*, pair $(1, 5)$ is connected by a descendant-ancestor path with inner nodes in $\{2, 3, 4\}$.

Therefore, a 1,5-edge is induced by marginalizing over nodes 2,3,4 and hence $1 \perp\!\!\!\perp 5|C$ is not implied. An 1,5-edge is induced by conditioning on node 4 or on any of its descendants in $\{1, 2, 3\}$ so that $1 \perp\!\!\!\perp 5|cC$ is not implied, $c \neq \emptyset$.

The following Figure 4 shows special cases of summary graphs, noting that C and one of u, v may be empty sets. Figure 4 shows that summary

FIG 4. *Important special cases of summary graphs. The two pairs* X, Y *and* Z, U *are constrained given* Y_C*; with* $X \perp\!\!\!\perp Y|ZU$ *in a),b),c), with* $X \perp\!\!\!\perp Y|U$ *in d),e), and with* $X \perp\!\!\!\perp Y$ *in f); with* $Z \perp\!\!\!\perp U$ *in c),e), f), with* $Z \perp\!\!\!\perp U|Y$ *in b),d) and with* $Z \perp\!\!\!\perp U|XY$ *in a).*



graphs cover all six possible combinations of independence constraints on two non-overlapping pairs of four variables X, Z, U, Y. Substantive research examples with well-fitting data to linear models of Figure 4 have been given by Cox and Wermuth (1993) to the concentration graph in Figure 4*a)*, the directed acyclic graph in 4*b)*, the graph of seemingly unrelated regression graph in 4*d)* and the covariance graph in 4*f)*.

1.5. *Markov equivalence.* The notion of Markov equivalence is important, because for any given set of data, one cannot distinguish between two Markov equivalent graph models on the basis of goodness-of-fit tests.

DEFINITION 7. *Two different graphs in node set N are Markov equivalent if they capture the same independence structure.*

Since a different set of two independence statements is associated with each of the graphs in Figure 4, none of the six graphs are Markov equivalent.

Known conditions, under which a concentration graph or a covariance graph is Markov equivalent to a directed acyclic graph, may be proven by orienting the graphs that is by changing each edge present into an arrow. The same type of argument can be extended to other independence graphs such as to a MRC graphs; see also Proposition 2 below. For this, we need a few more definitions for graphs.

For $a \subset N$, the subgraph induced by a is obtained by keeping of the graph all nodes in a and all edges coupling nodes in a. Subgraphs induced by three nodes are named V-configurations if they have two edges. A path is said to be chordless if each inner node forms a V-configuration with its two neighbors.

For the V-configurations of a MRC graph that are collision paths with endpoints i and k, the inner node is excluded from the conditioning set of any independence statement for Y_i, Y_k implied by the graph. In contrast, for V-configurations of a MRC graph that are transmitting paths, the inner node is included in the conditioning set of any independence statement for Y_i, Y_k implied by the graph. Thus, the independence structure of the graph is changed whenever any collision-oriented V-configuration is exchanged with a transmitting-oriented V-configuration.

A concentration graph with a chordless 4-cycle, as in Figure 4*a)*, or with any larger chordless cycle, is not Markov equivalent to a directed acyclic graph; see Dirac (1961) and Lauritzen (1996). The reason is that it is impossible to orient the graph without obtaining either a directed cycle or at least one collision-oriented V-configuration.

Similarly, a covariance graph that contains a chordless path in four or more nodes, i.e. a n-chain with $n \geq 4$, is not Markov equivalent to a directed acyclic graph; see Pearl and Wermuth (1994). The reason is that is impossible to orient each edge without obtaining at least one transmitting-oriented V-configuration.

For four nodes, there are three types of chordless collision paths in a multivariate regression graph:

$$\circ\!\longrightarrow\!\circ\;{-}{-}{-}\;\circ\!\longleftarrow\!\circ, \qquad \circ\;{-}{-}{-}\;\circ\;{-}{-}{-}\;\circ\!\longleftarrow\!\circ, \qquad \circ\;{-}{-}{-}\;\circ\;{-}{-}{-}\;\circ\;{-}{-}{-}\;\circ.$$

The following result explains why in general three types of edge are needed after marginalising and conditioning in a directed acyclic graph.

PROPOSITION 2. Lack of Markov equivalence. *If a MRC graph has a chordless collision path for $n \geq 4$ nodes or a chordless cycle in $n \geq 4$ nodes within v then it is not Markov equivalent to any directed acyclic graph in the same node set.*

PROOF. It is impossible to orient the graph with any one of the chordless collision paths for $n \geq 4$ nodes into edges of a directed acyclic graph without switching between the two types of inner nodes in at least one V-configuration, that is between a collision and a transmitting node. And, for the chordless cycle for $n \geq 4$, the above result due to Dirac applies. $\square$

Currently, one knows how to generate three types of independence-preserving graph from a given directed acyclic graph in node set V for the same disjoint subsets M and C of V. In a MC-graph, four types of edge may occur in combination. The summary graph may have up to three types of edge and one type of double edge, while the maximal ancestral graph is an independence graph with up to three types of edge. For proofs of Markov equivalence of the three corresponding types see Sadeghi (2009).

1.6. *Families of distribution generated over parent graphs.* A distribution and its joint density f_V is said to be generated over a directed acyclic graph whenever f_V factorizes recursively into univariate conditional densities that satisfy the independence constraints specified with the graph. Any full ordering of V is compatible with a given directed acyclic graph if, for each node i, all ancestors of i are in $\{i+1,\dots,d_V\}$. The set of parent nodes of i is denoted by par_i.

For $V = (1,\dots d_V)$ specifying a compatible ordering of node set V, a defining list of constraints for a directed acyclic graph is

$$f_{i|i+1,\dots,d_V} = f_{i|\mathrm{par}_i} \iff i \perp\!\!\!\perp \{i+1,\dots,d_V\} \setminus \mathrm{par}_i | \mathrm{par}_i \ , \tag{1.2}$$

the factorization of the density generated over the graph is

$$f_V = \textstyle\prod_{i=1}^{d_V} f_{i|\mathrm{par}_i}. \tag{1.3}$$

To generate f_V recursively, one can take any compatible ordering of V.

Definition 8. *For a recursive generating process of f_V, one starts with the marginal density f_{d_V} of Y_{d_V}, proceeds with the conditional density of Y_{d_V-1} given Y_{d_V} continues to $f_{i|i+1,\dots,d_V}$ and ends with the conditional density of Y_1 given $Y_2,\dots,Y_{d_V}$*

To let a directed acyclic graph represent one of such recursive generating processes, the graph is to capture both, independences and dependences.

Definition 9. *A directed acyclic graph, with a given compatible ordering of V, is edge-minimal for f_V generated over it if*

$$f_{i|\mathrm{par}_i} \neq f_{i|\mathrm{par}_i \setminus l} \text{ for each } l \in \mathrm{par}_i \,.$$

Under this condition of edge-minimality of the generating graph for f_V, all relevant explanatory variables are included for each Y_i and no edge can be removed from the graph without changing the independence statements satisfied by Y_i given its past, $\mathrm{pst}_i = \{i+1,\dots,d_V\}$.

An edge-minimal graph may represent a research hypothesis in a given substantive context. For such a hypothesis, those dependences are considered that are strong enough to be of substantive interest while others are translated into independence statements; see Wermuth and Lauritzen (1990).

Definition 10. *A recursive generating process of f_V in the order $V = (1,\dots,d_V)$ is said to consist of freely chosen components Y_i if each Y_i can* *be discrete or continuous, the form of the family of distribution of Y_i given Y_{pst_i} may be of any type, and parameters of $f_{i|\mathrm{pst}_i}$ are variation independent of those of f_{pst_i}.*

For exponential families of distributions, variation-independent factorizations of $f_{i,\mathrm{pst}_i} = f_{i|\mathrm{pst}_i} f_{\mathrm{pst}_i}$ coincide with the notion of a cut given by Barndorff-Nielsen (1978), p. 50. These types of factorization imply that the overall likelihood function can be maximized by maximizing each factor $f_{i|\mathrm{pst}_i}$ separately.

In families of distribution with f_V consisting of freely chosen components that satisfy the defining independences (1.2) of the given graph, some further constraints on each $f_{i|\mathrm{par}_i}$ are possible such as no-higher-order interactions or such as requiring Y_i to have dependences of equal strength on several of its explanatory variables, that is on several components of Y_{par_i}. Excluded are for instance constraints across conditional distributions, such as dependences of Y_i on some of Y_{par_i} to be equal to those of Y_k on some of Y_{par_k}.

Freely chosen components Y_i are in general incompatible with distributions that are to be faithful to a generating directed acyclic graph. The notion was introduced by Spirtes, Glymour and Scheines (1993). It means that the independence structure of f_V coincides with the independence structure captured by the graph and it leads in general to complex constraints on the parameter space for distributions generated over parent graphs; see Figure 1 of Wermuth, Marchetti and Cox (2009) for a simple example with three binary variables. In contrast, variation independence permits special constellations of parameter values that may lead to independencies in f_V that are additional to those implied by the graph.

For research hypotheses, defined in terms of recursive constraints on the independence structure and on dependences of f_V, appropriate specifications and resulting properties can now be given. For this, only connected graphs are considered, those with each node pair connected by at least one path.

DEFINITION 11. *A connected directed acyclic graph is named a parent graph, G^V_{par}, if the order of its node set $V = (1, \ldots, d_V)$ is given by the recursive generating process of f_V and it is edge-minimal for f_V.*

DEFINITION 12. *A family of distributions is said to be generated over a given parent graph if it varies fully and each component $f_{i|\mathrm{pst}_i}$ of f_V is freely chosen in the recursive generating process of f_V.*

PROPOSITION 3. General properties of families of distribution generated over G^V_{par}. *A family of distributions generated over G^V_{par} is complete and each* *member satisfies the intersection and the composition property.*

PROOF. Freely chosen components of Y_i contradict incompleteness of each family of distributions f_{i,pst_i} and joint families of distributions defined recursively in terms of complete families of distributions are complete. Independences implied by G^V_{par} combine in f_V generated over G^V_{par} as in a non-degenerate Gaussian distribution; see Lemma 1 of Marchetti and Wermuth (2009). □

COROLLARY 3. Association-inducing paths in G^V_{par}. *In a family f_V generated over a parent graph, every ik-path present in G^V_{par}, that induces an ik-edge by marginalising or conditioning, is association-inducing for Y_i, Y_k.*

PROOF. Completeness of f_V generated over G^V_{par} is sufficient for uniqueness of the independence statement attached to each V-configuration. □

Impossible is then for instance, for an uncoupled node pair i, k with V-configuration $i \prec\!\!-\!\!-\, j \prec\!\!-\!\!-\, k$ and $\gamma \subseteq \mathrm{pst}_k$, that

$$\int f_{ij|\gamma} f_{jk|\gamma} / f_{j|\gamma}\, dy_j = f_{i|\gamma} f_{k|\gamma}, \text{ or equivalently } \int (f_{i|j\gamma} - f_{i|\gamma}) f_{j|k\gamma}\, dy_j = 0.$$

1.7. *Using summary graphs to detect distortions of generating dependences.* An ik-dependence in a MAG model may differ qualitatively from the generating dependence of Y_i on Y_k in f_V, in particular it may change the sign but stay a strong dependence. If this remained undetected, one would come to qualitatively wrong conclusions when interpreting the parameters measuring conditional dependence of Y_i on Y_k in $f_{u|vC}$.

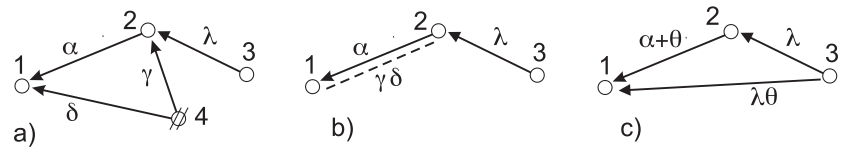
The summary graph corresponding to a MAG detects, whether and for which of the generating dependences, $i \prec\!\!-\!\!-\, k$, having both of i, k within u, such distortions can occur due to direct or indirect confounding; see Wermuth and Cox (2008) and Corollary 4, Lemma 1 below. We illustrate here direct confounding with Figure 5 and indirect confounding with Figure 6.

For a joint Gaussian distribution, the distortions are compactly described in terms of regression coefficients for variables Y_i standardized to have mean zero and variance one. For Figure 5*a)*, the generating equations be

$$Y_1 = \alpha Y_2 + \delta\, Y_4 + \varepsilon_1, \quad Y_2 = \lambda Y_3 + \gamma Y_4 + \varepsilon_2, \quad Y_3 = \varepsilon_3, \quad Y_4 = \varepsilon_4, \tag{1.4}$$

With residuals ε_i assumed to have zero means and to be uncorrelated, the equations of the summary graph model that result from (1.4) for Y_4

FIG 5. *a) Generating graph for Gaussian relations in standardized variables, leading for variable Y_4 unobserved to b) the summary graph and c) the maximal ancestral graph for the observed variables; with the generating dependences as attached to the arrows in a), implied are as simple correlations* $\rho_{12} = \alpha + \gamma\delta$, $\rho_{13} = \alpha\lambda$, $\rho_{23} = \lambda$ *and* $\theta = \gamma\delta/(1-\lambda^2)$.



unobserved, have one pair of correlated residuals

$$Y_1 = \alpha Y_2 + \eta_1, \quad Y_2 = \lambda Y_3 + \eta_2, \quad Y_3 = \eta_3,$$

$$\eta_1 = \delta\, Y_4 + \varepsilon_1, \quad \eta_2 = \gamma Y_4 + \varepsilon_2, \quad \eta_3 = \varepsilon_3, \quad \mathrm{cov}(\eta_1, \eta_2) = \gamma\delta.$$

The equation parameters to the standardized Gaussian associated with the MAG of Figure 5*c)* are instead defined via

$$E(Y_1|Y_2 = y_2, Y_3 = y_3), \quad E(Y_2|Y_3 = y_3),$$

with all residuals in the recursive equations being uncorrelated. The generating dependence α is retained in the summary graph model.

The parameter for the dependence of Y_1 on Y_2 in the MAG model, expressed in terms of the generating parameters of Figure 5*a)*, is $\alpha+\gamma\delta/(1-\lambda^2)$. The summary graph is in Figure 2*b)* a graphical representation of the simplest type of an instrumental variable model, used in econometrics, see Sargan (1958), to separate a direct confounding effect, here $\gamma\delta$, from the dependence of interest, here α.

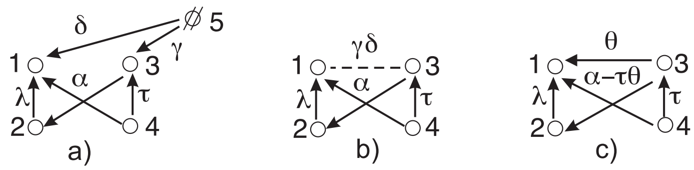
In general, possible distortions due to direct confounding in parameters of dependence in MAG models, are recognized in the corresponding summary graph by a double edge $i \overset{\prec\!\!-\!\!-}{- - -} k$. In the following example for Gaussian standardized variables, there is no direct confounding of the generating dependence α but there is indirect confounding of α while λ remains undistorted.

To simplify the figures, the coefficient attached to $2 \prec\!\!\!-\!\!\!- 3$ is not displayed in any of three graphs of Figure 6. The generating graph in Figure 6*a)* is directed and acyclic so that the corresponding linear equations in standardized Gaussian variables, defined implicitly by Figures 6*a)* have uncorrelated residuals. The example is adapted from Robins and Wasserman (1997). The summary graph in Figure 6*b)* shows with a dashed line the induced association for pair Y_1, Y_3 that results by marginalising f_V over Y_5.

The equations of the summary graph model, obtained for Y_5 unobserved, have precisely one pair of correlated residuals, $\mathrm{cov}(\eta_1, \eta_3) = \gamma\delta$ and

$$Y_1 = \lambda Y_2 + \alpha Y_4 + \eta_1, \quad Y_2 = \rho_{23} Y_3 + \eta_2, \quad Y_3 = \tau Y_4 + \eta_3, \quad Y_4 = \eta_4.$$

FIG 6. *a) Generating graph for linear relations in standardized variables, leading for variable Y_5 unobserved to b) the summary graph and c) the maximal ancestral graph for the observed variables; with the generating dependences as attached to the arrows in the a), implied are $\theta = \gamma\delta/(1-\tau^2)$, generating dependence λ undistorted in both models to the graphs b), c); generating dependence α preserved with b), distorted with c).*



The summary graph model preserves both λ and α as equation parameters.

In the corresponding MAG model, represented by the graph in Figure 6*c)*, the equation parameters associated with arrows present in the graph are unconstrained linear least squares regression coefficients. These coefficients, expressed in terms of the generating parameters of Figure 6*a)*, are shown next to the arrows in Figure 6*c)*. Thus, the generating coefficient λ is preserved, while α is changed into $\alpha - \tau\theta$, with $\theta = \gamma\delta/(1-\tau^2)$.

Direct confounding of a generating dependence of Y_i on Y_k is avoided in intervention studies, such as experiments and controlled clinical trials, by randomized allocation of individuals to the levels of Y_k, but severe indirect confounding may occur nevertheless; see Wermuth and Cox (2008).

Let the set of ancestors of node i in G^V_{par} be denoted by anc_i. Then, the set ancestors of node i in $G^{V\setminus[C,M]}_{\text{sum}}$ within u is $c_i = u \bigcap \text{anc}_i$ since no additional ancestor of i is ever generated within u. Then, by conditioning Y_i on Y_v and Y_{c_i}, one marginalises implicitly over the nodes in set $m_i = \{\{1,\dots,i\},\{u \cap \text{pst}_i \setminus c_i\}\}$ and indirect confounding may result.

COROLLARY 4. Lack of confounding in measures of conditional dependence. *A generating dependence $i \prec\!\!-\, k$ present in G^V_{par} is undistorted in the* MAG *model in nodes $V \setminus \{C,M\}$ (1) by direct confounding if in G^V_{par} there is no active ik-path relative to $\{C,M\}$ and (2) by indirect confounding if in $G^{V\setminus[C,M]}_{\text{sum}}$ there is no active ik-path relative to $\{c_i, m_i\}$.*

In distributions generated over G^V_{par}, every active path is association-inducing, hence a generating dependence will be confounded unless the distortion is cancelled by other edge-inducing paths. When a distortion is judged to be severe depends on the subject matter context. To detect indirect confounding, we name k a forefather of i if it is an ancestor but not a parent of i and three dots indicate more edges and nodes of the same type.

LEMMA 2. Wermuth and Cox (2008). A graphical criterion. *For $i \prec\!\!-\, k$*

of G^V_{par}, indirect confounding in the absence of direct confounding is generated in the MAG *model by marginalising over $M = \{l > k, l+1, \ldots, d_V\}$ if and only if in the corresponding summary graph $G^{V\setminus[\emptyset,M]}_{\text{sum}}$ which is without double edges, associations for Y_i, Y_k do not cancel that result by conditioning on all ancestors of node i, that is from the following types of collision ik-path that have as inner nodes only forefathers of i and are edge-inducing:*

$$i \,{-}{-}{-}\, \boxed{\circ} \ldots \boxed{\circ} \,{-}{-}{-}\, \boxed{\circ} \,{-}{-}{-}\, k, \qquad i \,{-}{-}{-}\, \boxed{\circ} \ldots \boxed{\circ} \,{-}{-}{-}\, \boxed{\circ} \prec\!\!\!\longrightarrow k. \tag{1.5}$$

An example of such a path of indirect confounding is given with the above Fig. 6*b)*, where for $1 \prec\!\!\!\longrightarrow 4$, it is the path $1 \,{-}{-}{-}\, 3 \prec\!\!\!\longrightarrow 4$.

In the following two sections, we give further preliminary results and those proofs of new results for which we use more technical arguments.

2. Further preliminary results. The edge matrix $\mathcal{A}$ of a parent graph is a $d_V \times d_V$ unit upper-triangular matrix, i.e. a matrix with ones along the diagonal and with zeros in the lower triangular part, such that for $i < k$, element $\mathcal{A}_{ik}$ of $\mathcal{A}$ satisfies

$$\mathcal{A}_{ik} = 1 \text{ if and only if } i \prec\!\!\!\longrightarrow k \text{ in } G^V_{\text{par}}. \tag{2.1}$$

Because of the triangular form of the edge matrix $\mathcal{A}$ of G^V_{par}, a density f_V generated over a given parent graph, has also been called a triangular system of densities.

2.1. *Linear triangular systems.* A linear triangular system is given by a set of recursive linear equations for a mean-centred random vector variable Y of dimension $d_V \times 1$ having $\text{cov}(Y) = \Sigma$, i.e. by

$$AY = \varepsilon, \tag{2.2}$$

where A is a real-valued $d_V \times d_V$ unit upper-triangular matrix, given by

$$E_{\text{lin}}(Y_i | Y_{i+1} = y_{i+1}, \ldots, Y_{d_V} = y_{d_V}) = -A_{i,\text{par}_i} y_{\text{par}_i},$$

and $E_{\text{lin}}(\cdot)$ denotes a linear predictor. The random vector ε of residuals has zero mean and $\text{cov}(\varepsilon) = \Delta$, a diagonal matrix. A Gaussian triangular system of densities is generated if the distribution of each residual ε_i is Gaussian, the corresponding joint Gaussian family varies fully if $\Delta_{ii} > 0$ for all i.

The covariance and concentration matrix of Y are, respectively,

$$\Sigma = A^{-1}\Delta(A^{-1})^{\mathrm{T}}, \qquad \Sigma^{-1} = A^{\mathrm{T}}\Delta^{-1}A. \tag{2.3}$$

Linear independences that constrain the equations (2.2) are defined by zeros in the triangular decomposition, (A, Δ^{-1}), of the concentration matrix. For joint Gaussian distributions

$$A_{ik} = 0 \iff i \perp\!\!\!\perp k | \mathrm{par}_i \text{ for } k \in \mathrm{pst}_i \setminus \mathrm{par}_i.$$

The edge matrix $\mathcal{A}$ of G^V_{par} coincides for Gaussian triangular systems generated over G^V_{par} with the indicator matrix of zeros in A, i.e. $\mathcal{A} = \mathrm{In}[A]$, where $\mathrm{In}[\cdot]$ changes every nonzero entry of a matrix into a one. Furthermore, since the parent graph in node set V is edge-minimal for f_V, we have

$$A_{ik} = 0 \iff \mathcal{A}_{ik} = 0.$$

Edge matrices expressed in terms of components of a set of given generating edge matrices are called induced. Simple examples of edge matrices induced by $\mathcal{A}$ of (2.1) are the overall covariance and the overall concentration graph, see Wermuth and Cox (2004). These two types of graph have as induced edge matrices, respectively,

$$\mathcal{S}_{VV} = \mathrm{In}[\mathcal{A}^-(\mathcal{A}^-)^{\mathrm{T}}], \text{ and } \mathcal{S}^{VV} = \mathrm{In}[\mathcal{A}^{\mathrm{T}}\mathcal{A}], \tag{2.4}$$

where $\mathcal{A}^-$ has all ones of $\mathcal{A}$ and an additional one in position (i, k) if and only if k is a forefather of node i in G^V_{par}. In the graph with edge matrix $\mathcal{A}^-$, every forefather k of i is turned into a parent, that $i \prec\!\!\!-\!\!\!- k$ is inserted.

By writing the two matrix products in (2.4) explicitly, one sees that for an uncoupled node pair i, k in the parent graph, there is an additional edge in the induced concentration graph of Y_V if and only if the pair has a common offspring in G^V_{par}. With a zero in position i, k of $\mathcal{A}^-$, there is an additional ik-edge in the induced covariance graph if and only if an uncoupled pair has a common parent in the directed graph with edge matrix $\mathcal{A}^-$.

Both of these induced matrices are symmetric. The covariance and the concentration matrix, implied by a linear triangular system and given in (2.3), contain all zeros present in the corresponding induced edge matrices, but possibly more. This happens for (i, k) whenever the associations induced for Y_i, Y_k cancel that are due to several edge-inducing ik-paths. Then there are particular parametric constellations; see Wermuth and Cox (1998) for examples in Gaussian distributions generated over parent graphs.

By contrast, the induced edge matrices capture consequences of the generating independence structure, they contain structural zeros, those that occur for all permissible parametrizations, or, expressed differently, that occur for each member of a family f_V generated over a given G^V_{par}.

For distributions generated over parent graphs, a zero in position (i,k) of $\mathcal{S}_{VV}$ and of $\mathcal{S}^{VV}$ means, respectively, that

$$i \perp\!\!\!\perp k, \quad i \perp\!\!\!\perp k | V \setminus \{i,k\} \tag{2.5}$$

is implied by G^V_{par}. Thus, in contrast to the global Markov property, the induced graphs answer all queries concerning sets of these two types of independence statements at once.

More complex induced edge matrices arise for instance in MRC-graphs and in summary graphs derived from $\mathcal{A}$. For transformations of linear systems, we use the operator named partial inversion introduced next; for proofs and discussions see Wermuth, Wiedenbeck and Cox (2006), Marchetti and Wermuth (2009), Wiedenbeck and Wermuth (2010).

2.2. *Partial inversion.* Let F be a square matrix of dimension d_V with principal submatrices that are all invertible. This holds for instance for every A of (2.2) and for every covariance matrix of a Gaussian distribution which varies fully, so that $\text{cov}(Y)$ is positive definite, i.e. Y has no degenerate component.

For any subset a of V and $b = V \setminus a$, by applying the operator named partial inversion to the linear equations $FY = \eta$, say, these are modified into

$$\text{inv}_a F \begin{pmatrix} \eta_a \\ Y_b \end{pmatrix} = \begin{pmatrix} Y_a \\ \eta_b \end{pmatrix}. \tag{2.6}$$

By applying partial inversion to b of V in equation (2.6), one obtains $Y = F^{-1}\eta$. Thus, full inversion is decomposed into two steps of partial inversion.

Partial inversion extends the sweep-operator for symmetric, invertible F

$$\text{inv}_a F = \begin{pmatrix} F_{aa}^{-1} & -F_{aa}^{-1}F_{ab} \\ F_{ba}F_{aa}^{-1} & F_{bb.a} \end{pmatrix} \text{ with } F_{bb.a} = F_{bb} - F_{ba}F_{aa}^{-1}F_{ab}. \tag{2.7}$$

LEMMA 3. Wermuth, Wiedenbeck and Cox (2006). Some properties of partial inversion. *Partial inversion is commutative, can be undone and is exchangeable with selecting a submatrix. For V partitioned as $V = \{a, b, c, d\}$*

(1) $\text{inv}_a \,\text{inv}_b F = \text{inv}_b \,\text{inv}_a F,$
(2) $\text{inv}_{ab} \,\text{inv}_{bc} F = \text{inv}_{ac} F,$
(3) $[\text{inv}_a F]_{J,J} = \text{inv}_a F_{JJ}$ *for* $J = \{a, b\}$.

In contrast, the sweep operator cannot be undone; see Dempster (1972). Example 1 shows how the triangular equations (2.2) are modified by partial inversion on a, where a consists of the first d_a components of Y. Instead of the full recursive order $V = (1, \dots d_V)$ with uncorrelated residuals, a block-recursive order $V = (a, b)$ results, where residuals within a are correlated, but uncorrelated with the unchanged residuals within b.

EXAMPLE 1. Partial inversion applied to a linear triangular system (2.2) with an order-respecting split of V. *For* $a = \{1, \dots, d_a\}$, $b = \{d_a+1, \dots, d_V\}$

$$\mathrm{inv}_a A = \begin{pmatrix} A_{aa}^{-1} & -A_{aa}^{-1} A_{ab} \\ 0 & A_{bb} \end{pmatrix} \textit{ gives with } Y_a = -A_{aa}^{-1} A_{ab} Y_b \ + A_{aa}^{-1} \varepsilon_a,$$

the implied form of linear least-squares regression of Y_a *on* Y_b, *where*

$$E_{\mathrm{lin}}(Y_a | Y_b = y_b) \ = \Pi_{a|b} y_b,\ Y_{a|b} = Y_a \ - \Pi_{a|b} Y_b\ ,\ \mathrm{cov}(Y_{a|b}) = \Sigma_{aa|b}, \textit{ and}$$

$$\Sigma_{aa|b} = A_{aa}^{-1} \Delta_{aa} A_{aa}^{-\mathrm{T}} \textit{ where } F^{-\mathrm{T}} = (F^{-1})^{\mathrm{T}} \textit{ and } F^{\mathrm{T}} \textit{ is } F \textit{ transposed.}$$

Example 2 shows how the triangular equations (2.2) are modified by partial inversion on b, where $V = (a, b, c)$ so that b consists of intermediate components of Y. To use directly the matrix formulation in (2.7), one sets $b := (a, c)$, $a := b$, leaves components within a and within b unchanged to obtain $\tilde{A}$ which is not block-triangular in (a, b). After partial inversion of $\tilde{A}$ on a, the original order is restored for the results presented in Example 2.

EXAMPLE 2. Partial inversion applied to a linear triangular system (2.2) for an order-respecting partitioning $V = (a, b, c)$. *With* $a = \{1, \dots, d_a\}$, $b = \{d_a + 1, \dots, (d_a + d_b)\}$ *and* $c = \{(d_a + d_b) + 1, \dots, d_V\}$,

$$\mathrm{inv}_b A = \begin{pmatrix} A_{aa} & A_{ab} A_{bb}^{-1} & A_{ac.b} \\ 0 & A_{bb}^{-1} & -A_{bb}^{-1} A_{bc} \\ 0 & 0 & A_{cc} \end{pmatrix} \textit{ gives } Y_a = -A_{aa}^{-1} A_{ac.b} Y_c \ + \eta_a,$$

the implied form of the linear least-squares regression of Y_a *on* Y_c, *with*

$$\eta_a = A_{aa}^{-1} \varepsilon_a + \Pi_{a|b.c} A_{bb}^{-1} \varepsilon_b, \quad \Pi_{a|bc} = (\Pi_{a|b.c},\ \Pi_{a|c.b}) = -A_{aa}^{-1} (A_{ab},\ A_{ac}).$$

For $\Pi_{a|c}$, *a special form of Cochran's recursive definition of regression coefficients results; see also Wermuth and Cox (2004),*

$$\Pi_{a|c} = \Pi_{a|c.b} + \Pi_{a|b.c} \Pi_{b|c} = -A_{aa}^{-1} (A_{ac} - A_{ab} A_{bb}^{-1} A_{bc}) = -A_{aa}^{-1} A_{ac.b}.$$

For $\mathrm{cov}(Y_{a|c})$, *Anderson's recursive definition of covariance matrices results*

$$\Sigma_{aa|c} = A_{aa}^{-1} \Delta_{aa} A_{aa}^{-\mathrm{T}} + \Pi_{a|b.c} (A_{bb}^{-1} \Delta_{bb} A_{bb}^{-\mathrm{T}}) \Pi_{a|b.c}^{\mathrm{T}} = \Sigma_{aa|bc} + \Sigma_{ab|c} \Sigma_{bb|c}^{-1} \Sigma_{ba|c}.$$

For b, c, the result in Example 2 are is as in Example 1. For Y_a, the original recursive regressions given Y_b, Y_c are modified into recursive regressions given only Y_c. The residuals between Y_a, Y_b are correlated since $\mathrm{cov}(Y_{a|c}, Y_{b|c}) = \Sigma_{ab|c}$ but remain uncorrelated from those in c. In the modified equations, Y_b can be removed without affecting any of the other remaining relations.

For a more detailed discussion of the three different types of recursion relations of linear association measures due to Cochran, Anderson and Dempster; see Wiedenbeck and Wermuth (2010).

For Example 3, one starts with equations (2.2) premultiplied by $A^{-\mathrm{T}}\Delta^{-1}$ and obtains linear equations in which the equation parameter matrix, Σ^{-1}, coincides with the covariance matrix of the residuals that is with

$$\Sigma^{-1}Y = A^{-\mathrm{T}}\Delta^{-1}\varepsilon. \tag{2.8}$$

EXAMPLE 3. Partial inversion with any split of V applied to Σ^{-1}. *The covariance matrix Σ and the concentration matrix Σ^{-1} of Y are written, partitioned according to (a, b) for a any subset of V, as*

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ . & \Sigma_{bb} \end{pmatrix}, \qquad \Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} \\ . & \Sigma^{bb} \end{pmatrix},$$

where the . notation indicates symmetric entries. Partial inversion of Σ^{-1} on a leads to three distinct components, $\Pi_{a|b}$, the population coefficient matrix of Y_b in linear least squares regression of Y_a on Y_b, the covariance matrix $\Sigma_{aa|b}$ of $Y_{a|b}$ and the marginal concentration matrix $\Sigma^{bb.a}$ of Y_b

$$\mathrm{inv}_a\Sigma^{-1} = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma^{bb.a} \end{pmatrix}, \tag{2.9}$$

where the $\sim$ notation denotes entries that are symmetric except for the sign.

Since (2.6) *gives directly* $\mathrm{inv}_a\Sigma^{-1} = \mathrm{inv}_b\,\Sigma$, *several well known dual expressions for the three submatrices in* (2.9) *result*

$$\begin{pmatrix} (\Sigma^{aa})^{-1} & -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ \sim & \Sigma^{bb} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ab} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} & \Sigma_{ab}\Sigma_{bb}^{-1} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix},$$

where $\Sigma^{bb.a}$ is Dempster's recursive definition of concentration matrices.

A more complex key result is that for any block-triangular system of linear equations for Y, with equation parameter matrix H and with possibly correlated residuals obtained from $W = \mathrm{cov}(HY)$, the implied form of $\mathrm{inv}_a\Sigma^{-1}$ can be expressed in terms of partially inverted matrices H and W.

Linear equations in a mean-centred vector variable Y are block-triangular in two ordered blocks (a,b) with a positive definite $\Sigma^{-1} = H^{\mathrm{T}}W^{-1}H$ if

$$HY = \eta, \text{ with } H_{ba} = 0, \quad E(\eta) = 0, \quad \mathrm{cov}(\eta) = W \text{ positive definite,} \tag{2.10}$$

For $K = \mathrm{inv}_a H$ and $Q = \mathrm{inv}_b W$, direct computations give

$$\mathrm{inv}_a(H^{\mathrm{T}}W^{-1}H) = \begin{pmatrix} K_{aa}Q_{aa}K_{aa}^{\mathrm{T}} & K_{ab} + K_{aa}Q_{ab}K_{bb} \\ \sim & H_{bb}^{\mathrm{T}}Q_{bb}H_{bb} \end{pmatrix}. \tag{2.11}$$

A simple special case is the triangular linear system (2.2). Example 4 shows how a multivariate regression chain in blocks (a,b) results from it.

EXAMPLE 4. *For* (2.10) *with* $H = A$ *of* (2.2)*,* $W = \Delta$ *diagonal and* $a = 1, \ldots, d_a$*,*

$$\mathrm{inv}_a(H^{\mathrm{T}}\Delta^{-1}H) = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix} = \begin{pmatrix} K_{aa}\Delta_{aa}K_{aa}^{\mathrm{T}} & K_{ab} \\ \sim & A_{bb}^{\mathrm{T}}\Delta_{bb}^{-1}A_{bb} \end{pmatrix}$$

Other special cases of linear block-triangular systems (2.10) are Gaussian summary graph models; see Section 3.

2.3. *Partial closure.* Let $\mathcal{F}$ be a binary edge matrix for node set $V = \{1, \ldots, d_V\}$ associated with F. The operator called partial closure transforms $\mathcal{F}$ into $\mathrm{zer}_a\mathcal{F}$ so that in the corresponding graph a-line paths of special type become closed. For instance, applied to $\mathcal{A}$, every a-line ancestor of node i is turned into a parent of i and, applied to the edge matrix of an undirected graph, such as $\mathcal{S}^{VV}$, every a-line path is closed. Zeros in the new binary matrix $\mathrm{zer}_a\mathcal{F}$ are the structural zeros that remain of $\mathrm{inv}_a F$.

In matrix form, with $n-1 = d_a$ and $\mathcal{I}_{aa}$ a $d_a \times d_a$ identity matrix,

$$\mathrm{zer}_a\mathcal{F} = \mathrm{In}[\begin{pmatrix} \mathcal{F}_{aa}^{-} & \mathcal{F}_{aa}^{-}\mathcal{F}_{ab} \\ \mathcal{F}_{ba}\mathcal{F}_{aa}^{-} & \mathcal{F}_{bb.a} \end{pmatrix}] \text{ with } \mathcal{F}_{bb.a} = \mathrm{In}[\mathcal{F}_{bb} + \mathcal{F}_{ba}\mathcal{F}_{aa}^{-}\mathcal{F}_{ab}], \tag{2.12}$$

$$\mathcal{F}_{aa}^{-} = \mathrm{In}[(n\,\mathcal{I}_{aa} - \mathcal{F}_{aa})^{-1}]. \tag{2.13}$$

The inverse in (2.13) assures nonnegative entries in $\mathcal{F}_{aa}^{-}$ and is a type of regularization; see Tikhonov (1963). It generalizes limits of scalar geometric series; see Neumann (1884), p. 29.

LEMMA 4. Wermuth, Wiedenbeck and Cox (2006). Some properties of partial closure. *Partial inversion is commutative, cannot be undone and is exchangeable with selecting a submatrix. For V partitioned as $V = \{a, b, c, d\}$*

(1) $\mathrm{zer}_a\,\mathrm{zer}_b F = \mathrm{zer}_b\,\mathrm{zer}_a F$,
(2) $\mathrm{zer}_{ab}\,\mathrm{zer}_{bc} F = \mathrm{zer}_{abc} F$,
(3) $[\mathrm{zer}_a F]_{J,J} = \mathrm{zer}_a F_{JJ}$ *for* $J = \{a, b\}$.

Given Gaussian parameter matrix components after partial inversion, such as in equation (2.11), the corresponding induced edge matrices are obtained using the following Lemma 5, provided each component matrix belongs to the model of the starting graph and the expressions are minimal, that is condensed in such a way that they do not contain any parameter matrices that cancel, as for instance $A_{aa}A_{aa}^{-1}$ would.

LEMMA 5. Marchetti and Wermuth (2009). Edges induced by a starting graph obtained with minimal matrix expressions of Gaussian parameter matrices. *Edge matrices replace corresponding parameter matrices after*
(1) changing each negative sign to a positive sign,
(2) replacing in the resulting expressions each diagonal matrix by an identity matrix or deleting it if it arises within a matrix product,
and then applying the indicator function.

For instance, the matrix formulation of partial inversion in (2.12) can be viewed as arising from (2.7) by use of Lemma 5.

EXAMPLE 1 continued. *Let $\mathcal{K}_{aa} = \mathcal{A}_{aa}^-$ and $\mathcal{K}_{ab} = \mathcal{A}_{aa}^-\mathcal{A}_{ab}$. After partial closure in G^V_{par} on a, there are two induced edge matrix components. For directed edges, it is $\mathrm{zer}_a\mathcal{A}$, and for undirected dashed line edges, it is $\mathcal{S}_{aa|b}$*

$$\mathrm{zer}_a\mathcal{A} = \mathrm{In}[\begin{pmatrix} \mathcal{K}_{aa} & \mathcal{K}_{ab} \\ 0 & \mathcal{A}_{bb} \end{pmatrix}],\ \mathcal{P}_{a|b} = \mathrm{In}[\mathcal{K}_{ab}],\ \mathcal{S}_{aa|b} = \mathrm{In}[\mathcal{K}_{aa}\mathcal{K}_{aa}^{\mathrm{T}}]\ .$$

The induced graph of two components is a multivariate regression graph.

EXAMPLE 2 continued. *By marginalising over the intermediate node set b of $V = (a, b, c)$ in G^V_{par}, a directed, acyclic graph results. The induced Gaussian parameter and edge matrices are for $N = V \setminus b$, respectively,*

$$[\mathrm{inv}_b A]_{N,N} = \begin{pmatrix} A_{aa} & A_{ac.b} \\ 0 & A_{cc} \end{pmatrix},\ [\mathrm{zer}_b\mathcal{A}]_{N,N} = \mathrm{In}[\begin{pmatrix} \mathcal{A}_{aa} & \mathcal{A}_{ac.b} \\ 0 & \mathcal{A}_{cc} \end{pmatrix}].$$

EXAMPLE 3 continued. *A concentration graph has for joint Gaussian distributions Σ^{-1} as parameter matrix and $\mathcal{S}^{VV}$ as edge matrix. By partial* *closure on a of $\mathcal{S}^{VV}$ given any split $V = \{a, b\}$, every a-line path is closed. Three edge matrix parts result $\mathcal{S}_{aa|b}, \mathcal{P}_{a|b}$ and $\mathcal{S}^{bb.a}$. They give the structural zeros in the corresponding parameter matrices $\Sigma_{aa|b}, \Pi_{a|b}$ and $\Sigma^{bb.a}$. In general, the edge matrix $\mathcal{S}^{bb.a}$ is for the marginal concentration graph of Y_b.*

When the generating graph is G^V_{par}, then a concentration graph is induced for the node set which contains ancestors of C outside C. In Example 4, the three components of $\mathrm{inv}_a\Sigma^{VV}$ are directly expressed in terms of the triangular decomposition (A, Δ^{-1}).

EXAMPLE 4 continued. *For the order-respecting split, $V = (a, b)$, and $\mathcal{K}_{aa} = \mathcal{A}^-_{aa}$ and $\mathcal{K}_{ab} = \mathcal{A}^-_{aa}\mathcal{A}_{ab}$, a parent graph G^V_{par} induces a* MRC *graph for $f_{a|b}$ and f_b with the following three edge matrix components*

$$\begin{pmatrix} \mathcal{S}_{aa|b} & \mathcal{P}_{a|b} \\ . & \mathcal{S}^{bb.a} \end{pmatrix} = \mathrm{In}[\begin{pmatrix} \mathcal{K}_{aa}\mathcal{K}^{\mathrm{T}}_{aa} & \mathcal{K}_{ab} \\ . & \mathcal{A}^{\mathrm{T}}_{bb}\mathcal{A}_{bb} \end{pmatrix}]. \tag{2.14}$$

The result combines the one in (2.4) in slightly modified form with the above continuation of Example 1 by considering the consequences of a given parent graph for the distributions of Y_a given Y_b and of Y_b.

For the more complex generating graphs connected with block-triangular linear systems (2.10) and given edge matrices $\mathcal{H}, \mathcal{W}$, the three edge matrix components in the induced MRC graph of just two components, are with

$$\mathcal{K} = \mathrm{zer}_a\mathcal{H}, \quad \mathcal{Q} = \mathrm{zer}_b\mathcal{W},$$

$$\begin{pmatrix} \mathcal{S}_{aa|b} & \mathcal{P}_{a|b} \\ . & \mathcal{S}^{bb.a} \end{pmatrix} = \mathrm{In}[\begin{pmatrix} \mathcal{K}_{aa}\mathcal{Q}_{aa}\mathcal{K}^{\mathrm{T}}_{aa} & \mathcal{K}_{ab} + \mathcal{K}_{aa}\mathcal{Q}_{ab}\mathcal{K}_{bb} \\ . & \mathcal{H}^{\mathrm{T}}_{bb}\mathcal{Q}_{bb}\mathcal{H}_{bb} \end{pmatrix}]. \tag{2.15}$$

From (2.15) for $a = \{\alpha, \delta\}$, the edge matrices induced by G^V_{par} for $f_{\alpha|b}$ are

$$\mathcal{S}_{\alpha\alpha|b} = [\mathcal{S}_{aa|b}]_{\alpha,\alpha}, \quad \mathcal{P}_{\alpha|b} = [\mathcal{P}_{\alpha|b}]_{\alpha,b},$$

and with a split of b as $\{\beta, \gamma\}$, the edge matrix induced for $f_{\beta|\gamma}$ and for the dependence of $Y_{\alpha|\gamma}$ given $Y_{\beta|\gamma}$ are

$$\mathcal{S}^{\beta\beta.a} = [\mathcal{S}^{bb.a}]_{\beta,\beta}, \text{ and } \mathcal{P}_{\alpha|\beta.\gamma} = [\mathcal{P}_{a|b}]_{\alpha,\beta}.$$

In general, the induced graphs of (2.14) or (2.15) with dashed lines for $\mathcal{S}_{aa|b}$, arrows for $\mathcal{P}_{a|b}$ and full lines for $\mathcal{S}^{bb.a}$ will not be independence-preserving graphs. In both graphs, the global Markov property of Lemma 1 implies the meaning of a missing ik-edge as

$$i \perp\!\!\!\perp k|b \text{ in } \mathcal{S}_{aa|b}, \quad i \perp\!\!\!\perp k|b \setminus k \text{ in } \mathcal{P}_{a|b}, \quad i \perp\!\!\!\perp k|b \setminus \{i, k\} \text{ in } \mathcal{S}^{bb.a}. \tag{2.16}$$

Whenever every edge-inducing path is association-inducing, conditional dependences correspond to edges present in the graph, that is in $f_{a|b}$ and f_b, unless the associations due to several edge-inducing paths cancel.

3. Summary graphs and associated models.

3.1. *Gaussian summary graph models.* Starting from a Gaussian triangular system (2.2) generated over a parent graph in node set V, marginalising over M and conditioning on C gives a linear system of equations for $Y_{N|C}$ for $N = (u, v) = V \setminus \{C, M\}$ of the following form, where for the equations in the ancestors v of C that are outside of C, the equation parameter matrix and the covariance matrix coincide with a concentration matrix, as in (2.8).

DEFINITION 13. Gaussian summary graph model. *A Gaussian summary graph model is a system of equations $HY_{N|C} = \eta$, that is a block-triangular and orthogonal in (u, v) with*

$$\begin{pmatrix} H_{uu} & H_{uv} \\ 0 & \Sigma^{vv.uM} \end{pmatrix} \begin{pmatrix} Y_{u|C} \\ Y_{v|C} \end{pmatrix} = \begin{pmatrix} \eta_u \\ \zeta_v \end{pmatrix}, \qquad \mathrm{cov} \begin{pmatrix} \eta_u \\ \zeta_v \end{pmatrix} = \begin{pmatrix} W_{uu} & 0 \\ . & \Sigma^{vv.uM} \end{pmatrix}, \tag{3.1}$$

where H_{uu} is unit upper-triangular, W_{uu} and $\Sigma^{-1}_{vv|C} = \Sigma^{vv.uM}$ are symmetric, and each of η_u and ζ_v have freely varying joint Gaussian distributions. The independence structure is given by a summary graph in node set N; see Definition 6 and Section 3.2 below.

For $Y_{v|C}$, equation (3.1) specifies a Gaussian concentration graph model. These models had been studied under the name of covariance selection by Dempster (1972); see also Speed and Kiiveri (1986). For each member of the this family of models, the likelihood function has a unique maximum.

With $W_{uv} = 0$, the residuals of $Y_{u|C}$ and $Y_{v|C}$ are uncorrelated, therefore the system of equations (3.1) is said to be orthogonal in (u, v). Because of this orthogonality, $\Pi_{u|v.C} = -H_{uu}^{-1} H_{uv}$ is the population least-squares regression coefficient matrix in linear regression of $Y_{u|C}$ on $Y_{v|C}$; compare Example 1 above. In econometrics, the equation in $Y_{u|c}$ resulting by premultiplication with H_{uu}^{-1} from the first equation of (3.1) is called the reduced form.

The equation in $Y_{u|C}$ of (3.1) can equivalently be written as a recursive system in endogenous variables $Y_{u|vC} = Y_{u|C} - \Pi_{u|v.C} Y_{v|C}$

$$H_{uu} Y_{u|vC} = \eta_u \text{ with } \mathrm{cov}(\eta_u) = W_{uu}, \tag{3.2}$$

where the equation parameter matrix H_{uu} is, as in the linear triangular system (2.2), of unit upper-triangular form, but some of the residuals η_u are correlated. For estimation, one speaks in econometrics of the endogeneity problem; see Drton, Eichler and Richardson (2009) for a recent discussion.

Identification is an issue for estimating the equation parameters H_{uu} in (3.2). No necessary and sufficient condition is known yet; see for instance Kang and Tian (2009). One general sufficient condition is the absence of any double edge in the summary graph; see Brito and Pearl (2002). This says that for any pair i, k within u, either $H_{ik} = 0$, or $W_{ik} = 0$, or both hold.

However, some models with double edges in the G^N_{sum} correspond to identified instrumental variable models; see the above example to Figure 5*b)*. For the identifiability of latent variable models, which arise here via larger hypothesized generating processes, the notion of completeness is again relevant; see San Martin and Mouchart (2007).

3.2. *Generating* $G^{V\setminus[C,M]}_{\text{sum}}$ *from* G^V_{par}. The summary graph $G^{V\setminus[C,M]}_{\text{sum}}$ has four edge matrix components. With $\mathcal{S}^{vv.uM}$ a concentration graph results in node set v, with $\mathcal{H}_{uu}$ a directed acyclic graph within u, with $\mathcal{W}_{uu}$ a covariance graph of the residuals η_u, and with $\mathcal{H}_{uv}$ a bipartite graph for dependence of $Y_{u|C}$ on $Y_{v|C}$.

Starting from a Gaussian triangular system in (2.2) with parent graph G^V_{par}, the choice of any conditioning set C leads to an ordered split $V = (O, R)$, where we think of $R = \{C, F\}$ as the nodes to the right of O, see equation (3.3). Every node in F is an ancestor of a node in C outside C, so that we call F the set of foster nodes of C. No node in O has a descendant in R so that O is said to contain the outsiders of R. Equations, orthogonal and block-triangular in (O, R), are in unchanged order

$$\begin{pmatrix} A_{OO} & A_{OR} \\ 0 & A_{RR} \end{pmatrix} \begin{pmatrix} Y_O \\ Y_R \end{pmatrix} = \begin{pmatrix} \varepsilon_O \\ \varepsilon_R \end{pmatrix}. \tag{3.3}$$

After conditioning on Y_C and marginalising over Y_M, the resulting system preserves block-triangularity and orthogonality with $u \subseteq O$, $v \subseteq F$.

PROPOSITION 4. Linear equations obtained from $AY = \varepsilon$ after conditioning on Y_C and marginalising over Y_M. *Given a Gaussian triangular system* (2.2) *generated over* G^V_{par}*, conditioning set* C*, marginalising set* $M = (p, q)$ *with*

$$p = O \setminus u, \qquad q = F \setminus v,$$

and partially inverted parameter matrices arranged in the appropriate order,

$$D = \text{inv}_p\, \tilde{A}, \quad \text{inv}_q \tilde{\Sigma}^{FF.O} = \begin{pmatrix} \Sigma_{qq|vC} & \Pi_{q|v.C} \\ \sim & \Sigma^{vv.qO} \end{pmatrix},$$

the induced linear equations (3.1) *in* $Y_{N|C}$ *have equation parameters*

$$H_{uu} = D_{uu}, \quad H_{uv} = D_{uv} + D_{uq}\Pi_{q|v.C}, \quad \Sigma^{vv.uM} \tag{3.4}$$

and covariance matrices

$$W_{uu} = (\Delta_{uu} + D_{up}\Delta_{pp}D_{up}^{\mathrm{T}}) + (D_{uq}\Sigma_{qq|vC}D_{uq}^{\mathrm{T}}), \quad \Sigma^{vv.uM}. \tag{3.5}$$

PROOF. Equations (3.3) in Y are first modified into equations for $Y_{O|C}$ and $Y_{F|C}$. As for Example 3 above, one takes $\zeta_R = A_{RR}\,\Delta_{RR}^{-1}\varepsilon_R$. After noting that

$$\Sigma_{FF|C}^{-1} = [\Sigma^{RR.O}]_{F,F} = \Sigma^{FF.O}$$

and by the orthogonality in (O, R), these equations can be written as

$$A_{OO}Y_{O|C} + A_{OF}Y_{F|C} = \varepsilon_O, \qquad \Sigma^{FF.O}Y_{F|C} = \zeta_F.$$

Partial inversion on $M = (p, q)$ gives, after appropriately ordering,

$$\mathrm{inv}_M \begin{pmatrix} \tilde{A}_{OO} & \tilde{A}_{OF} \\ 0 & \tilde{\Sigma}^{FF.O} \end{pmatrix} \begin{pmatrix} \varepsilon_p \\ Y_{u|C} \\ \zeta_q' \\ Y_{v|C} \end{pmatrix} = \begin{pmatrix} Y_{p|C} \\ \varepsilon_u \\ Y_{q|C} \\ \zeta_v' \end{pmatrix}, \tag{3.6}$$

where after deleting the equations in $Y_{M|C}$, the uncorrelated residuals are

$$\eta_u = (\varepsilon_u - D_{up}\varepsilon_p) - D_{uq}\Sigma_{qq|vC}\zeta_q, \qquad \zeta_v = \zeta_v' + \Pi_{q|v.C}^{\mathrm{T}}\zeta_q'$$

Thus, the equation parameter matrices of (3.4), the covariance matrices (3.5) result, where $\Sigma_{vv|C}^{-1} = \Sigma^{vv.qO} = \Sigma^{vv.uM}$. □

It is instructive to check the relations of the parameter matrices in (3.4), (3.5) to multivariate regression coefficients and to conditional covariance matrices. With $\Pi_{u|R} = -D_{uu}^{-1}(D_{uv},\ D_{uq},\ D_{uC})$, one may write

$$-D_{uu}\Pi_{u|v.C} = D_{uv} + D_{uq}\Pi_{q|v.C}, \qquad D_{uu}(Y_{u|C} - \Pi_{u|v.C}Y_{v|C}) = D_{uu}Y_{u|vC},$$

and for W_{uu} defined in (3.2) and specialized in (3.5)

$$D_{uu}^{-1}W_{uu}D_{uu}^{-T} = \Sigma_{uu|vqC} + \Pi_{u|q.vC}\Sigma_{qq|vC}\Pi_{u|q.vC}^{\mathrm{T}} = \Sigma_{uu|vC},$$

so that the required covariance matrix of $Y_{u|vC}$ is obtained.

The summary graph in node set N, induced by the generating parent graph in node set V, results now directly with Lemma 5 applied to equations (3.4) and (3.5) as is stated in Corollary 5.

COROLLARY 5. *With the partially closed edge matrices corresponding to Proposition 4 and arranged in the appropriate order*

$$\mathcal{D} = \mathrm{zer}_p \tilde{\mathcal{A}}, \quad \mathrm{zer}_q \, \tilde{\mathcal{S}}^{FF.O} = \begin{pmatrix} \mathcal{S}_{qq|vC} & \mathcal{P}_{q|v.C} \\ . & \mathcal{S}^{vv.qO} \end{pmatrix}$$

the induced edge matrix components of the summary graph $G^{V\setminus[C,M]}_{\mathrm{sum}}$ *are*

$$\mathcal{H}_{uu} = \mathcal{D}_{uu}, \quad \mathcal{H}_{uv} = \mathrm{In}[\mathcal{D}_{uv} + \mathcal{D}_{uq}\mathcal{P}_{q|v.C}], \quad \mathcal{S}^{vv.uM}, \tag{3.7}$$

$$\mathcal{W}_{uu} = \mathrm{In}[(\mathcal{I}_{uu} + \mathcal{D}_{up}\mathcal{D}^{\mathrm{T}}_{up}) + (\mathcal{D}_{uq}\mathcal{S}_{qq|vC}\mathcal{D}^{\mathrm{T}}_{uq})]. \tag{3.8}$$

3.3. *Non-Gaussian models associated with summary graphs.* As noted before, the density $f_{N|C}$ of Y_N given Y_C is well-defined since it is obtained from a density of Y_V generated over a parent graph by marginalising over Y_M and conditioning on Y_C. As we have seen, this leads to the factorization of $f_{N|C}$ into $f_{u|vC}$ and $f_{v|C}$. The independence structure of Y_v given Y_C is captured by a concentration graph.

Corresponding models for discrete and continuous random variables have been studied by Lauritzen and Wermuth (1989), extending the Gaussian covariance selection models and the graphical, log-linear interaction models for discrete variables. Maximum-likelihood estimation is considerably simplified for variation-independent parameters; see Frydenberg and Lauritzen (1989).

For a joint Gaussian density f_V, the induced density $f_{u|vC}$ is again Gaussian, but in general, the form and parametrization of the density $f_{u|vC}$ induced by f_V may be complex. Nevertheless, we conjecture that the parameters associated with $G^{V\setminus[C,M]}_{\mathrm{sum}}$ may often be obtained via the notional stepwise generating process described in Section 1.3 that is by introducing latent variables that are mutually independent and independent of Y_v, Y_C.

If the additional latent variables are taken to be discrete and to have a large number of levels, then it should be possible to generate, or at least to approximate closely enough, any association corresponding to i - - - k. This follows for discrete variables by Theorem 1 of Holland and Rosenbaum (1989) and otherwise presumably by using Proposition 5.8 of Studený (2005), but a proof is pending.

3.4. *Generating a summary graph from a larger summary graph.* Let a summary graph be given, where the corresponding model, actually or only notionally, arises from a parent graph model by conditioning on Y_c and by marginalising over variables Y_m.

Then, the starting linear parent graph model is the triangular system of equation (2.2) in a mean-centred Gaussian variable Y where

$$AY = \varepsilon, \quad \operatorname{cov}(\varepsilon) = \Delta \text{ diagonal}, \quad A \text{ unit upper-triangular}.$$

With Proposition 4, one obtains for $V \setminus \{c, m\} = (\mu, \nu)$ the following equations in $Y_{\mu|c}$, $Y_{\nu|c}$, which coincide in form with equations (3.1),

$$\begin{pmatrix} B_{\mu\mu} & B_{\mu\nu} \\ 0 & \Sigma^{\nu\nu.\mu m} \end{pmatrix} \begin{pmatrix} Y_{\mu|c} \\ Y_{\nu|c} \end{pmatrix} = \begin{pmatrix} \eta'_\mu \\ \zeta_\nu \end{pmatrix}, \quad \operatorname{cov} \begin{pmatrix} \eta'_\mu \\ \zeta_\nu \end{pmatrix} = \begin{pmatrix} W'_{\mu\mu} & 0 \\ . & \Sigma^{\nu\nu.\mu m} \end{pmatrix}, \tag{3.9}$$

With added conditioning on a set $c_\nu \subseteq \nu$, no additional ancestors of c_ν are defined, since every node in ν is already an ancestor of c. But, with added conditioning on $c_\mu \subseteq \mu$, the set $\mu \setminus c_\mu$ is split into foster nodes f_μ of c_μ and into outsiders o of $\{r, \nu\}$, where $r = \{c_\mu, f_\mu\}$.

The equations for Y_μ are always block-triangular in (o, r). But, by contrast to the split of V into (O, R) in equation (3.3), these equations are not orthogonal in (o, r) so that conditioning on c_μ in the summary graph is more complex than conditioning directly on a set in the parent graph.

PROPOSITION 5. Linear equations obtained from (3.9) after conditioning on Y_{c_μ}, Y_{c_ν} and marginalising over Y_h, Y_l. *Given* (3.9) *to* $G^{V\setminus\{c,m\}}_{\text{sum}}$*, where* o *contains all outsiders of* $\{c_\mu, f_\mu, \nu\}$*, equations for* Y_μ *are block-triangular in*

$$\mu = (o, r) \text{ where } r = \{c_\mu, f_\mu\}.$$

The additional conditioning set $\{c_\mu, c_\nu\}$*, and additional marginalising sets* $h \subseteq o$ *and* $l \subseteq \{f_\mu, \nu \setminus c_\nu\}$ *give* $C = \{c, c_\mu, c_\nu\}$ *and* $M = \{m, h, l\}$*. With* $\psi = (r, \nu)$*, the new equations are block-triangular and orthogonal in* (u, v)*, where*

$$u = o \setminus h, \quad \phi = \psi \setminus \{c_\mu, c_\nu\}, \quad v = \phi \setminus l.$$

With orthogonalised residuals $\xi_o = \eta'_o - Q_{or}\eta'_r$*, orders* $\mu = (h, u, r)$*,* $\phi = (l, v)$

$$Q_{\mu\mu} = \operatorname{inv}_r \tilde{W}'_{\mu\mu}, \quad C_{o\psi} = B_{o\psi} - Q_{or}B_{r\psi}, \quad K = \operatorname{inv}_{hl} \begin{pmatrix} \tilde{B}_{oo} & \tilde{C}_{o\phi} \\ 0 & \tilde{\Sigma}^{\phi\phi.om} \end{pmatrix},$$

the linear summary graph model to $G^{V\setminus[C,M]}_{\text{sum}}$ *is*

$$\begin{pmatrix} K_{uu} & K_{uv} \\ 0 & \Sigma^{vv.uM} \end{pmatrix} \begin{pmatrix} Y_{u|C} \\ Y_{v|C} \end{pmatrix} = \begin{pmatrix} \eta_u \\ \zeta_v \end{pmatrix}, \quad \eta_u = \xi_u - K_{uh}\xi_h - K_{ul}\Sigma_{ll|vC}\zeta_l, \tag{3.10}$$

and coincides with the linear model obtained from the triangular system (2.2) *by directly conditioning on* Y_C *and marginalising over* Y_M*.*

PROOF. The conditioning set c_μ splits the set of nodes μ into (o, r), where o is without any descendant in $r = \{c_\mu, f_\mu\}$ and where every node in f_μ has a descendant in c. This implies a block-triangular form of $B_{\mu\mu}$ in (o, r) in the equations of $Y_{\mu|\nu c}$, however with correlated residuals η'_o and η'_r.

For $\psi = (r, \nu)$, block-orthogonality with respect to (o, ψ) in the equations in $Y_{o|c}$ and $Y_{\psi|c}$ is achieved by subtracting from η'_o the value predicted by linear least-squares regression of η'_o on η'_r and ζ_ν. This reduces, because of the orthogonality of the equations in (μ, ν), to subtracting $Q_{or}\eta'_r$ from η'_o.

The matrix of equation parameters of $Y_{\psi|c}$ coincides with the concentration matrix of $Y_{\psi|c}$ given by

$$\Sigma^{\psi\psi.om} = \Sigma^{-1}_{\psi\psi|c} = \begin{pmatrix} B^{\mathrm{T}}_{rr} Q_{rr} B_{rr} & B^{\mathrm{T}}_{rr} Q_{rr} B_{r\nu} \\ . & \Sigma^{-1}_{\nu\nu|c} + B^{\mathrm{T}}_{r\nu} Q_{rr} B_{r\nu} \end{pmatrix}. \tag{3.11}$$

By the block-triangularity and orthogonality in (o, ψ), the equations in $Y_{o|c}$ can be replaced by equations in $Y_{o|C}$. For the equations in $Y_{\phi|C}$, the matrix of equation parameters is $\Sigma^{-1}_{\phi\phi|C} = [\Sigma^{-1}_{\psi\psi|c}]_{\phi,\phi} = \Sigma^{\phi\phi.om}$. The resulting equations give the Gaussian linear model to the summary graph in node set $V \setminus \{C, m\} = (o, \phi)$.

In the linear model to $G^{V\setminus[C,m]}$, marginalizing over $Y_{h|C}$, where $h \subseteq o$, and over $Y_{l|C}$, where $l \subseteq \phi$, is achieved with partial inversion on h, l of the block-triangular matrix of equation parameters, just as in equation (3.6), and keeping only the equations in $Y_{u|C}$ and $Y_{v|C}$.

In the resulting equations (3.10), one knows by the commutativity and exchangeability of partial inversion for $m = (g, k)$, $p = \{g, h\}$, $q = \{k, l\}$ that

$$K_{uu} = [\mathrm{inv}_h \mathrm{inv}_g A]_{u,u} = [\mathrm{inv}_p A]_{u,u},$$

so that $K_{uu} = D_{uu}$, where D is defined for Proposition 4. Furthermore, by the properties of reduced form equations

$$-K_{uu}\Pi_{u|v.C} = K_{uv} = D_{uv} + D_{uq}\Pi_{q|v.C},$$

so that the parameter matrices of $Y_{u|C}$ and $Y_{v|C}$ given in (3.10) coincide with those in (3.4), (3.5) of Proposition 4 that is they give the Gaussian linear model to the summary graph in node set $V \setminus \{C, N\} = (u, v)$. $\square$

Since partial closure has the same exchangeability property as partial inversion and both operators are commutative, the same type of proof holds for the edge matrix expression corresponding to (3.10).

COROLLARY 6. *For $c \subset C$ and $m \subset M$, edge matrix components of the summary graph $G^{V\setminus[C,M]}$ result from the edge matrix components $\mathcal{B}_{\mu\mu}$, $\mathcal{B}_{\mu\nu}$, $\mathcal{W}'_{\mu mu}$ and $\mathcal{S}^{\nu\nu.\mu m}$ of $G^{V\setminus[c,m]}$ by using the transformed edge-matrices*

$$\mathcal{Q}_{\mu\mu} = \mathrm{zer}_r \tilde{\mathcal{W}}'{}_{\mu\mu}, \quad \mathcal{C}_{o\psi} = \mathrm{In}[\mathcal{B}_{o\psi} + \mathcal{Q}_{or}\mathcal{B}_{r\psi}], \quad \mathcal{K} = \mathrm{zer}_{hl}\begin{pmatrix} \tilde{\mathcal{B}}_{oo} & \tilde{\mathcal{C}}_{o\phi} \\ 0 & \tilde{\mathcal{S}}^{\phi\phi.om} \end{pmatrix}$$

to obtain $\mathcal{K}_{uu}$, $\mathcal{K}_{uv}$ directly, $\mathcal{S}^{vv.uM}$ as the edge matrix to (3.11)*, and*

$$\mathcal{W}_{uu} = \mathrm{In}[\mathcal{Q}_{uu} + \mathcal{K}_{uh}\mathcal{Q}_{hh}\mathcal{K}_{uh}^{\mathrm{T}} + \mathcal{K}_{ul}\mathcal{S}_{ll|vC}\mathcal{K}_{ul}^{\mathrm{T}}]. \tag{3.12}$$

3.5. *Path results derived from edge matrix transformations.* If one starts with the summary graph $G_{\mathrm{sum}}^{V\setminus[c,m]}$ and conditions by using Corollary 6, edges are induced by r-line collision paths, where we let $r = \{c_\mu, f_\mu\} = \{\ \boxed{\circledcirc}\ \}$:

(a) $\circ_\mu - - - \circ_\mu$ results with $\circ_\mu - - - \boxed{\circledcirc} \dots \boxed{\circledcirc} - - - \circ_\mu$,

(b) $\circ_\psi \;\text{——}\; \circ_\psi$ results with $\circ_\psi \longrightarrow \boxed{\circledcirc} - - - \boxed{\circledcirc} \dots \boxed{\circledcirc} - - - \boxed{\circledcirc} \longleftarrow \circ_\psi$,

(c) $\circ_\mu \longleftarrow \circ_\psi$ results with $\circ_\mu - - - \boxed{\circledcirc} \dots \boxed{\circledcirc} - - - \boxed{\circledcirc} \longleftarrow \circ_\psi$.

The corresponding relevant edge matrix expressions are, respectively, $Q_{\mu\mu} = \mathrm{zer}_r W_{\mu\mu}$, $\mathrm{In}[B_{r\psi}^{\mathrm{T}} Q_{rr} B_{r\psi}]$, and $\mathrm{In}[\mathcal{Q}_{or}\mathcal{B}_{r\psi}]$. For each pair, one keeps one edge of several of the same kind. The subgraph induced by nodes (o, ϕ) is $G^{V\setminus[C,m]}$.

By marginalising next over $m' = (h, l) = (\not\!\not\!\circ_h, \not\!\not\!\circ_l)$ in the graph $G_{\mathrm{sum}}^{V\setminus[C,m]}$, three types of edges are induced when closing m'-line transmitting paths:

(d) $\circ_\phi \;\text{——}\; \circ_\phi$ results with $\circ_\phi \;\text{——}\; \not\!\not\!\circ_l \dots \not\!\not\!\circ_l \;\text{——}\; \circ_\phi$,

(e) $\circ_o \longleftarrow \circ_o$ results with $\circ_o \longleftarrow \not\!\not\!\circ_h \dots \not\!\not\!\circ_h \longleftarrow \circ_o$,

(f) $\circ_o \longleftarrow \circ_\phi$ results with $\circ_o \longleftarrow \not\!\not\!\circ_h \dots \not\!\not\!\circ_h \longleftarrow \not\!\not\!\circ_l \;\text{——}\; \not\!\not\!\circ_l \dots \not\!\not\!\circ_l \;\text{——}\; \circ_\phi$,

(g) $\circ_u - - - \circ_u$ results with $\circ_u \longleftarrow \not\!\not\!\circ_h - - - \not\!\not\!\circ_h \longrightarrow \circ_u$,

(h) $\circ_u - - - \circ_u$ results with $\circ_u \longleftarrow \not\!\not\!\circ_l \;\text{——}\; \not\!\not\!\circ_l \longrightarrow \circ_u$.

The corresponding relevant edge matrix expressions are, respectively, $\mathcal{K}_{\phi\phi}$, $\mathcal{K}_{oo}$, $\mathcal{K}_{o\phi}$, $\mathrm{In}[\mathcal{K}_{uh}\mathcal{Q}_{hh}\mathcal{K}_{uh}^{\mathrm{T}}]$, and $\mathrm{In}[\mathcal{K}_{ul}\mathcal{S}_{ll|vC}\mathcal{K}_{ul}^{\mathrm{T}}]$. After keeping just one edge of several of the same kind, the subgraph induced by nodes (u, v) is $G^{V\setminus[C,M]}$.

Notice that the effect of the indicator function is to reduce several edges the same kind to just one. The closed form expressions of the edge matrix results imply that some of the paths are to be closed in the given order.

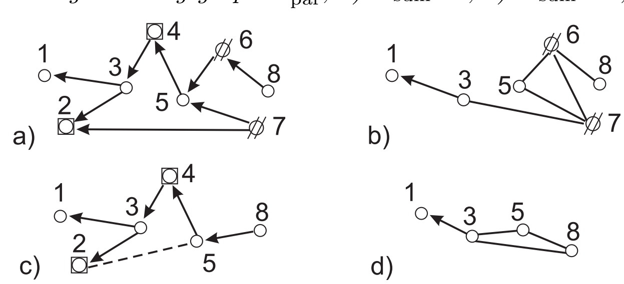
The edge matrices $\mathrm{In}[\mathcal{Q}_{or}\mathcal{B}_{r\psi}]$ and $\mathcal{K}_{o\phi}$ correspond in a Gaussian summary graph model to orthogonalising that is to removing some residual correlations. By the associated steps, (c) or (f), ik-arrows may be generated for which node k is not an ancestor of i in the generating graph.

In contrast, for the outsiders of the conditioning set, such as set o in the summary graph in nodes (o, ϕ), there is an ik-arrow if and only if k is a parent or a forefather of node i in the larger generating parent graph because the only arrow-inducing paths for the subset o are those in (e).

Since a summary graph results after conditioning with steps (a) to (c) and also after marginalising with steps (d) to (h), summary graphs are said to be closed under marginalising and conditioning and one may reverse the order of conditioning and marginalising. The following example illustrates such reversed stepwise constructions.

EXAMPLE 5. Path constructions of $G^{V\setminus[C,M]}_{\rm sum}$ for $M = q$ and $p = \emptyset$. *The node set of the parent graph is* $V = (1, \dots, 8)$. *The conditioning set* $C = \{2, 4\}$, *the marginalising set is* $M = \{6, 7\}$. *The foster nodes of* C, *are in* $F = \{3, 5, 6, 7, 8\}$ *and* $u = O = \{1\}$, $v = \{3, 5, 8\}$.

FIG 7. *a) The generating graph* $G^V_{\rm par}$, *b)* $G^{V\setminus[C,\emptyset]}_{\rm sum}$, *c)* $G^{V\setminus[\emptyset,M]}_{\rm sum}$, *d)* $G^{N}_{\rm sum}$.



In this example with graphs in Figure 7, the summary graph model is equivalent to a triangular system in $N = (1, 3, 5, 8)$ even though $G^{V\setminus[\emptyset,M]}_{\rm sum}$ is not Markov equivalent to any directed acyclic graph since it contains the chordless collision path $3 \!\longrightarrow\!\!\!\succ 2 \,{-}\,{-}\,{-}\, 5 \prec\!\!\!\longleftarrow 8$. It is typical, that further marginalizing or conditioning may again lead to simpler graphs and models.

With just one node in m', the paths (d) to (h) have just two edges. In addition, by the properties of partial inversion and partial closure, the paths (a) to (c) can be closed by operating on one node at the time and in any order. This leads to operating on one node at a time in any order; see Table 1 and the appendix, Table 1 and Proposition 1.

3.6. *The MAG corresponding to* $G^{V\setminus[C,M]}_{\rm sum}$ *and local Markov properties.* The keys to deriving the MAG corresponding to $G^{V\setminus[C,M]}_{\rm sum}$ are the definition of the variables in the Gaussian MAG model and the result (2.15). For Y_v, the summary graph and the MAG specify the same concentration graph and dependences to arrows pointing from v to u also coincide.

A full order of the nodes in u of $G^{V\setminus[C,M]}_{\rm sum}$ may sometimes be given by the arrows, such as in Figure 3*b)*, sometimes there is none as in Figure 2*b)*, more often there is a partial order, such as in Figures 1*d)* or 7*c)*. Then one may take any compatible full ordering of the nodes in u in which the ancestors within u of each node i in $G^{V\setminus[C,M]}_{\rm sum}$ are in the past of i, that is in $\{i+1,\dots,d_u\}$.

For each node i, we let $c_i \subseteq \{i+1,\dots,d_u\}$ denote the ancestors of i in $G^{V\setminus[C,M]}_{\rm sum}$ and $\bar{c}_i = \{i+1,\dots,d_u\}\setminus c_i$. Next, we derive for each node pair i,k with k in c_i and each node pair i,l with l in $\bar{c}_i$, the edges in the MAG corresponding to $G^{V\setminus[C,M]}_{\rm sum}$ by applying (2.15) to equations (3.2).

For $a = (1,\dots,i,\bar{c}_i)$ and $b = c_i$, the vector $\mathcal{P}_{i|b} = \mathrm{In}[\mathcal{K}_{ib} + \mathcal{Q}_{ib}\mathcal{K}_{bb}]$ gives zeros and ones for the dependence of Y_i on Y_{c_i} given Y_v, Y_C and

(3.13) in the MAG, $i \prec\!\!\!-\!\!\!- k$ for $\mathrm{In}[\mathcal{P}_{i|k.b\setminus k}] = 1$, $\;i,k$ uncoupled, otherwise.

Similarly, for i,l we let $e_{il} = c_i \cup c_l$ and $\bar{e}_{il} = \{i+1,\dots,d_u\}\setminus e_{il}$, take $a = (1,\dots,i,l,\bar{e}_{il})$ and $b = e_{il}$. With $\mathcal{S}_{aa|b} = \mathrm{In}[\mathcal{K}_{aa}\mathcal{Q}_{aa}\mathcal{K}_{aa}]$ of (2.15), $\mathcal{K}_{il} = 0$ and $\mathcal{W}_{uu}$ the edge matrix of the covariance graph of $G^{V\setminus[C,M]}_{\rm sum}$:

(3.14) in the MAG, i - - - l for $\mathrm{In}[\mathcal{W}_{il.b}] = 1$, $\;i,l$ uncoupled, otherwise.

The corresponding MAG results after inserting or replacing edges in $G^{V\setminus[C,M]}_{\rm sum}$ according to (3.13), (3.14) and keeping just one of several same edges.

PROPOSITION 6. Local Markov properties of summary graphs. *Let the edge matrix components, H_{uN}, W_{uu} and $\mathcal{S}^{vv.uM}$ of $G^{V\setminus\{C,M\}}_{\rm sum}$ be given from Corollary 5. Let node l and sets c_i, e_{il} be defined as above, but their subscripts dropped. Let further β denote subsets of nodes uncoupled to node i, then*

(1) $i \perp\!\!\!\perp \beta | Cv\setminus\{i,\beta\} \iff \mathcal{S}^{i\beta.uM} = 0$ *for $i\in v$ and $\beta \subset v$,*

(2) $i \perp\!\!\!\perp \beta | Cv\setminus\beta \iff \mathcal{H}_{i\beta.c} = 0$ *for $i \in u$ and $\beta\subset v$,*

(3) $i \perp\!\!\!\perp l | Cve \iff (\mathcal{W}_{il} = 0$ *and* $\mathcal{W}_{ie}\mathcal{W}^{-}_{ee}\mathcal{W}_{el} = 0)$ *for $i\in u$, and $l \in \bar{c}$.*

(4) $i \perp\!\!\!\perp \beta | Cvc\setminus\beta \iff (\mathcal{H}_{i\beta} = 0$ *and* $\mathcal{W}_{ic}\mathcal{W}^{-}_{cc}\mathcal{H}_{c\beta} = 0)$ *for $i\in u$ and $\beta\subset c$.*

Notice that pairwise independences result if β's contain single elements.

PROOF. The independences in (1) within v are those of a concentration graph; see also (2.16) in Example 4. The independences in (2) are those obtained when regressing $Y_{i|C}$ on $Y_{v|C}$; see also Example 2. The independences in (3) and (4) are reformulations of (3.14), (3.13), respectively. $\square$

4. Discussion. The common attractive feature of a maximal ancestral graph and of the corresponding summary graph is that they elucidate consequences of a possibly much larger generating graph regarding independences. The smaller graphs capture the independence structure implied by the generating graph and they can be used to understand additional consequences of the generating graph for independences that result after additional marginalising and conditioning.

An advantage of the MAG is that each edge corresponds a conditional association, each missing edge to a conditional independence. A disadvantage of a MAG is that a dependence, say to $i \prec\!\!\!-\!\!\!- k$, may be severely distorted compared to the dependence to $i \prec\!\!\!-\!\!\!- k$ in the generating process. With the corresponding summary graph one can identify which of the conditional dependences in the MAG remain undistorted and which do not.

Given the summary graph, the corresponding MAG is derived in a few steps. But in general, one cannot obtain from a given MAG the corresponding summary graph and also not the information about distortions. Both types of graph may contain semi-directed cycles. These are typically of interest only in connection with a larger generating process.

In contrast, their common subclass of multivariate regression chain graphs gives a substantial and much needed enlargement of the types of research hypotheses that can be formulated with directed acyclic graphs. They model stepwise generating processes not only in univariate but also in joint responses. This leads to a corresponding recursive factorization of the joint density in these vector variables.

In addition, every independence constraint for a component of a joint response is conditional on variables in the past of the joint response. This is an important distinction from all other types of currently known chain graphs and in line with research in many substantive fields where the study of dependences on past variables is judged to be more fruitful than those of associations and of independences among variables arising at the same time.

For Gaussian multivariate regression chains, properties of estimators and test statistics have been quite well understood for a considerable time. For discrete random variables all multivariate regression chains are smooth; see Drton (2009). Such smooth models are curved exponential families, see e.g. Cox (2007) Section 6.8, so that they have desirable properties regarding estimation and asymptotic properties of tests.

Much less is known for joint responses of discrete and continuous random components. Thus, though we now can derive important consequences of any type of multivariate regression chain, more results on equivalence, identification, estimation and goodness-of-fit criteria are needed.

However, if the multivariate regression chain can be generated, as discussed, via special types of hidden variables in a larger parent graph model, then its independence structure is defined by a list of independence statements for variable pairs. This permits local fitting with univariate generalised linear models, with checks for linearity, interaction and conditional independence based on observed associations of variable pairs and triples.

This requires no knowledge about the form of the joint distribution and it permits to formulate research hypotheses that are well-compatible with a given set of data and that are to be investigated in further empirical studies.

APPENDIX A: TWO-EDGE PATHS OF SUMMARY GRAPHS

The following arguments show that the types of induced edge of Table 1 are self-consistent. A node to be marginalised over by indicated $\not\circ$ and a node to be conditioned on by $\boxed{\circ}$.

The three types of edge-inducing, two-edge paths (1) to (3) in a parent graph that have as inner node a transition, a source or a sink node, respectively, are defined to generate the following three different types of edge:

(1) $\circ\!\leftarrow\!\!\!-\, \not\circ \leftarrow\!\!\!-\!\circ \;\Longrightarrow\; \circ\!\leftarrow\!\!\!-\!\circ$,

(2) $\circ\!\leftarrow\!\!\!-\, \not\circ -\!\!\!\rightarrow\!\circ \;\Longrightarrow\; \circ\, \text{- - -}\, \circ$,

(3) $\circ -\!\!\!\rightarrow \boxed{\circ} \leftarrow\!\!\!-\!\circ \;\Longrightarrow\; \circ \!-\!\!\!-\! \circ$,

The arrow has one, the dashed line two and the full line no edge endpoints that define a collision node when the edge is mirrored at the same node. Dashed lines denote edges in covariance graphs and full lines in concentration graphs. Closing paths in such graphs is defined to preserve the type of edge:

(4) $\circ\, \text{- - -}\, \boxed{\circ}\, \text{- - -}\, \circ \;\Longrightarrow\; \circ\, \text{- - -}\, \circ$,

(5) $\circ \!-\!\!\!-\! \not\circ \!-\!\!\!-\! \circ \;\Longrightarrow\; \circ \!-\!\!\!-\! \circ$.

The next two paths (6) and (7) are equivalent and both induce an arrow:

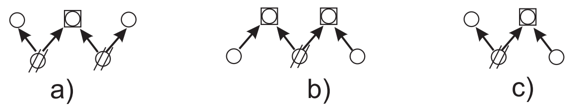
(6) $\circ\, \text{- - -}\, \boxed{\circ} \leftarrow\!\!\!-\!\circ \;\Longrightarrow\; \circ\!\leftarrow\!\!\!-\!\circ$,

(7) $\circ\!\leftarrow\!\!\!-\, \not\circ \!-\!\!\!-\! \circ \;\Longrightarrow\; \circ\!\leftarrow\!\!\!-\!\circ$,

Paths (4) to (7) arise from active alternating paths in a parent graph for which inner source nodes in $\{\not\circ\}$ alternate with inner sink nodes in $\{\boxed{\circ}\}$:

The two-edge paths (4) to (7) result from Figure 8 as follows: path (4) from a) by only marginalising, path (5) from b) by only conditioning, path (6) from c) by only marginalising and path (7) from c) by only conditioning.

FIG 8. *Active alternating paths that generate two-edge paths a) of type (4) inducing* $\circ{-}{-}{-}\circ$, *b) of type (5) inducing* $\circ$ —— $\circ$, *c) of type (6) or (7) inducing* $\circ\!\prec\!\!-\!\!-\circ$.



The paths a), b), c) of Figure 8 generalise paths (2), (3) and (1), respectively.

The three remaining edge-inducing paths of two edges in $G^{V\setminus[C,M]}_{\text{sum}}$ are

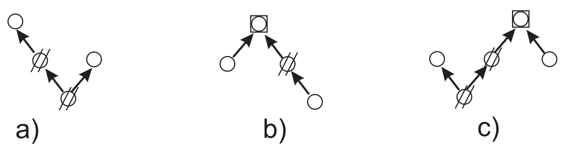
(8) $\circ\!\prec\!\!-\!\!-\, \boxed{\circ}\, {-}{-}{-}\circ \implies \circ{-}{-}{-}\circ$,

(9) $\circ$ —— $\not\!\!\!\circ\!\!\not\,\prec\!\!-\!\!-\circ \implies \circ$ —— $\circ$,

(10) $\circ{-}{-}{-}\, \boxed{\circ}$ —— $\circ \implies \circ\!\prec\!\!-\!\!-\circ$.

The three active paths of Figure 9 result by substituting the undirected edges in (8) to (10) by the appropriate generating components (2) or (3).

FIG 9. *Active paths that generate two-edge paths a) of type (8) inducing* $\circ{-}{-}{-}\circ$, *b) of type (9) inducing* $\circ$ —— $\circ$, *and of type (10) inducing* $\circ\!\prec\!\!-\!\!-\circ$.



By marginalising over the transition node in Figures 9*a)* to 9*c)*, one generates, respectively, path (2), path (3) and the path in Figure 8*c)*.

The construction of the summary graph simplifies considerably for special types of parent graphs, for instance for the graphs to the lattice conditional independence models, studied by Andersson et al. (1997), and for the graphs corresponding to labeled trees, studied by Castelo and Siebes (2003).

Acknowledgement. I thank D.R. Cox, M. Drton, G.M. Marchetti and K. Sadeghi for their insightful and constructive suggestions, the referees for their searching questions, and P. Albin and M. Studený for sharing with me some of their knowledge about families of complete distributions.

REFERENCES

ANDERSSON, S.A., MADIGAN D., PERLMAN M.D. AND TRIGGS, C.M. (1997). A graphical characterization of lattice conditional independence models. *Ann. Math. Artif. Intell.* **21**, 27-50.

BARNDORFF-NIELSEN, O.E. (1978). *Information and exponential families in statistical theory.* Wiley, Chichester.

BRITO, C. AND PEARL, J. (2002). A new identification condition for recursive models with correlated errors. *Struct. Equ. Model.* **9**, 459–474.

BROWN, L.D. (1986). Fundamentals of statistical exponential families with applications in statistical decision theory, LNMS **9**, Inst. Math. Statist., USA.

Castelo , R. and Siebes. A. (2003). A characterization of moral transitive acyclic directed graph Markov models as labeled trees. J. Stat. Plan. Inf, 115, 235–259.

Cox, D.R.(2007). *Principles of statistical inference.* Cambridge University Press. Cambridge.

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

Cox, D.R. and Wermuth, N. (1996). *Multivariate dependencies: models, analysis, and interpretation.* London: Chapman and Hall.

Cox, D.R. and Wermuth, N. (2000). On the generation of the chordless 4-cycle. *Biometrika*, **87**, 206–212.

Cox, D.R. and Wermuth, N. (2004). Causality: a statistical view. *Internat. Statist. Review* **72**, 285–305.

Dawid (1979). Some misleading arguments involving conditional independence *J. Roy. Statist. Soc. B*, **41**, 249–252.

Dempster, A.P. (1972). Covariance selection. *Biometrics* **28**, 157–175.

Dirac, G.A. (1961). On rigid circuit graphs. *Abhandl. Math. Seminar Hamburg* **25**, 71–76.

Drton, M. (2009). Discrete chain graph models. *Bernoulli* **15**, 736–753.

Drton, M., Eichler, M. and Richardson, T.S. (2009). Computing maximum likelihood estimates in recursive linear models. *J. Mach. Learn. Res.* **10**, 2329–2348.

Frydenberg, M. and Lauritzen, S.L. (1989). Decomposition of maximum likelihood in mixed interaction models. *Biometrika* **76**, 539–555.

Geiger, D., Verma, T.S. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20**, 507–534.

Holland, Paul W. and Rosenbaum, Paul R. (1989). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14**, 1523–1543.

Kang C. and Tian, J. (2009) Markov properties for linear causal models with correlated errors. *J. Mach. Learn. Res.* **10**, 41–70.

Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scan. J. Statist.* **23**, 115–116.

Koster (2002). Marginalising and conditioning in graphical models. *Bernoulli* **8**, 817–840.

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

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

Lehmann, E.L. and Scheffé, H. (1955). Completeness, similar regions and unbiased estimation. *Sankya* **14**, 219–236.

Mandelbaum, A. and Rüschendorf, L. (1987). Complete and symmetrically complete families of distributions. *Ann. Statist.* **15**, 1229–1244.

Marchetti, G.M. and Lupparelli, M. (2010). Chain graph models of multivariate regression type for categorical data. *Submitted.*

Marchetti, G.M. and Wermuth, N. (2009). Matrix representations and independencies in directed acyclic graphs. *Ann. Statist.* **47**, 961–978.

Neumann, C.G. (1884). *Vorlesungen über Riemann'sche Theorie der Abel'schen Integrale. 2nd ed.* Teubner, Leipzig.

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

Pearl, J. and Wermuth, N.(1994). When can association graphs admit a causal interpretation? In: *Models and data, artificial intelligence and statistics IV.* 205-214. P. Cheeseman and W. Oldford (eds.). Springer, New York.

Paz, A. (2007). The annotated graph model for representing DAG-representable relations–

algorithms approach. *Research Report, Technion, Haifa.*

Richardson, T.S. and Spirtes, P. (2002). Ancestral Markov graphical models. *Ann. Statist.* **30**, 962–1030.

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

Sadeghi, K. (2008). Graph theoretical proofs of separation criteria. *Mast. Scien. thesis, Chalmers University of Technology*

Sadeghi, K. (2009). Representing modified independence structures. *Transfer thesis, Oxford University*

San Martin E., Mochart M. (2007). On joint completeness: sampling and Bayesian versions, and their connections. *Sankya* **69**, 780–807.

San Martin E., Mochart M. and Rolin, J.M. (2005). Ignorable common information, null sets and Basu's first theorem. *Sankya* **67**, 674–698.

Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica.* **26**, 393–415.

Studený, M. (2005). *Probabilistic conditional independence structures.* Springer, London.

Speed, T.P. and Kiiveri, H.T. (1986). Gaussian Markov distributions over finite graphs. *Ann.Statist.* **14**, 138–150.

Spirtes, P., Glymour C. and Scheines R. (1993). *Causation, prediction and search.* Springer, New York.

Tikhonov, A.N. (1963). Solution of ill-posed problems and the regularization method. (Russian) *Dokl. Akad. Nauk SSSR* **153**, 49–52.

Wermuth, N. and Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli* **4**, 477–495.

Wermuth, N. and Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J.R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 687-717.

Wermuth, N. and Cox, D.R. (2008). Distortions of effects caused by indirect confounding. *Biometrika* **95**, 17–33.

Wermuth, N., Cox, D.R. and Pearl (1994). Explanantions for multivariate structures derived from univariate recursive regressions. *Ber. Stoch. verw. Geb., Univ. Mainz* **94-1**, ISSN-0098, Mainz.

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**, 21–75.

Wermuth, N., Marchetti, G.M. and Cox, D.R. (2009). Triangular systems for symmetric binary variables. *Electr. J. Statist.* **3**, 932–955.

Wermuth, N., Wiedenbeck, M. and Cox, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Mathematics* **46**, 883–901.

Wiedenbeck, M. and Wermuth, N. (2010). Changing parameters by partial mappings. *Statistica Sinica* **20, 2**, available online.

Mathematical Statistics
Chalmers/ University of Gothenburg
Chalmers Tvärgata 3
41296 Göteborg, Sweden
E-mail: wermuth@chalmers.se
URL: http://www.math.chalmers.se/~wermuth