International Statistical Review
Revue Internationale de Statistique

International Statistical Review (2012), 80, 3, 415–438 doi:10.1111/j.1751-5823.2012.00195.x

Traceable Regressions

Nanny Wermuth

Department of Mathematics, Chalmers Technical University, University of Gothenburg, Sweden, and International Agency of Research on Cancer, Lyon, France
E-mail: wermuth@chalmers.se

Summary

In this paper, we define and study the concept of traceable regressions and apply it to some examples. Traceable regressions are sequences of conditional distributions in joint or single responses for which a corresponding graph captures an independence structure and represents, in addition, conditional dependences that permit the tracing of pathways of dependence. We give the properties needed for transforming these graphs and graphical criteria to decide whether a path in the graph induces a dependence. The much stronger constraints on distributions that are faithful to a graph are compared to those needed for traceable regressions.

Key words: Chain graphs; edge-matrix calculus; faithfulness of graphs; graphical Markov models; independence axioms; regression graphs.

1 Introduction and Motivation

1.1 Single and Joint Response Regressions

Sequences of regressions are arguably the most important statistical tool in observational and interventional studies for investigating pathways of dependences and hence development over time. In each regression, one distinguishes *response* variables and *regressor* variables; with responses depending on the regressors.

In applications, the substantive context determines which variable pairs are modelled by a conditional independence and which are taken to be dependent because they are needed in a generating process of the joint distribution. Suppose one regressor is a risk factor for a response, then quite different sizes of dependence strength will be relevant if this response is the occurrence of a common cold, or the infection with HIV virus or an accident in a nuclear plant, since the prevention of these risks is judged to be of quite different importance.

There may be single or joint responses, where only the latter permit to model simultaenously occurring effects of an intervention. Components of joint responses may be discrete or continuous random variables or be mixed of both types. Typically, a subset of variables is taken as given, possibly determined by study design, and its components are named *context variables* since they describe the context or background or the basic features of individuals under study.

The generated joint density factorizes into an ordered sequence of conditional densities of the responses, which we call shortly regressions, and into a joint marginal density of the context variables. Under mild conditions, estimation of sequences of regressions can be decomposed into separate tasks for each response component of the factorization, using well-developed tools such as linear or logistic regressions or conditional Gaussian regressions, which permit joint responses to be mixed of discrete and continuous component variables; see Lauritzen & Wermuth (1989), Edwards & Lauritzen (2001). Tailored to the requirements in many specific situations, special results are available to estimate the form and parameters of univariate and joint conditional distributions.

However, many consequences of sequences of regressions can already be derived if one does not know or estimate the involved parameters but just uses a generating graph and properties of graph transformations. Relevant, important results concerning independences in sequences of regressions have been obtained only recently; see Sadeghi & Lauritzen (2012) and Wermuth & Sadeghi (2012). The additional properties needed to draw conclusions about induced dependences are set out in this paper.

Sequences of regressions are an essential part of longitudinal studies, named also cohort or panel studies in medical, economic, and social science research. Prominent examples are the Framingham heart study, the European Community household panel or the Swiss HIV cohort study. By using regression graphs, it will become possible to simplify analyses and interpretations of sequences of regressions and to directly compare dependences arising in different types of sequences of regressions for the same set of variables, or in sequences of regressions for subsets of variables studied for subpopulations. The results in this paper prepare for these possibilities in applications.

1.2 Independences and Dependences Given by Regression Graphs

Sequences of univariate, that is of single-response regressions, have been represented by *directed acyclic graphs*. With regression graphs, directed acyclic graphs are extended by including two types of undirected graph, one for joint responses, the other for joint context variables. Nodes of the graph represent random variables. Distinct node pairs are coupled by at most one edge so that a regression graph is one type of what in graph theory are called *simple graphs*. Each missing edge of a regression graph corresponds to a conditional independence where the conditioning set depends on the type and position of the missing edge, the graph is therefore also one type of *independence graph*.

Properties or axioms for combining independence statements have been studied by Dawid (1979) and Pearl (1988). Their connections to graphs have been discussed and modified in information theory; see Studený (2005) and Lněnička & Matúš (2007). Different types of extensions have been proposed in the computer science literature; see Castillo *et al.* (1997), Flesch & Lucas (2007). But, for instance, by requiring a property called strong transitivity, one excludes even the whole family of regular joint Gaussian distributions. By contrast, regular Gaussian families form a subclass of what we introduce here as traceable regressions.

The *independence structure* of a graph is the set of all independence statements implied by the graph. These are well-studied for regression graphs, with important results obtained only recently. For instance, a proof by Sadeghi & Lauritzen (2012) implies equivalence of a *pairwise Markov property*, that is of the set of independences attached to the missing edges of a given regression graph, to the *global Markov property*, the criterion known to give all independence statements implied by the graph. For two regression graphs with identical node sets and with the same set of coupled node pairs but with different types of edge, there is a simple graphical criterion to decide whether the two graphs define nevertheless the same independence structure, that is whether they are *Markov equivalent*; see Wermuth & Sadeghi (2012).

1.3 Tracing Pathways of Dependence

Much less is known about the dependence structures that can be captured by graphs. Since graphs do not distinguish between additive and interactive effects of regressor variables on responses, nor between linear and nonlinear types of dependences, it has been argued by Wermuth & Lauritzen (1989) that graphs may represent *research hypotheses about dependent variable pairs needed to generate the joint distribution*. For this, each edge present in the graph indicates a conditional dependence, where the conditioning set depends on the type and position of the edge present, while the form of the dependence is not specified.

For tracing pathways of dependences, dependence-inducing sequences of edges of different type are the focus of interest, while independences just lead to simplified strengthened interpretations of the relevant dependences. In this paper, we set out the properties of traceable regressions and show, in particular that these properties impose mild constraints on the types of generated distribution. This contrasts with strong constraints required in general for faithful distributions. This notion was introduced by Spirtes *et al.* (1993) for distributions in which all independence statements hold that are implied by a graph and no others.

Tracing pathways of dependence goes back to the geneticist Sewall Wright (1889–1988), who introduced it in 1923 as *path analysis* for sequences of univariate linear regressions. He suggested to judge the goodness-of-fit of a research hypothesis, represented by a directed acyclic graph, by comparing observed correlations with those that are expected if the data had been generated over the graph. His rules for computing expected marginal correlations, trace all pathways that induce a dependence by marginalizing.

The extension of tracing pathways of dependences, when there is conditioning on variables in addition to marginalizing, became feasible after a first *separation criterion* had been formulated by Pearl (1988) and proven by Geiger *et al.* (1990) to give the global Markov property of directed acyclic graphs. When separation fails, then there is at least one path in the directed acyclic graph that may induce a dependence by marginalizing over one subset of variables and conditioning on another set. Here, such a path is said to be edge-inducing since it leads to a transformed graph.

1.4 Structure of the Paper

In Section 2, we introduce and discuss dependence base regression graphs and traceable regressions. Section 3 contains examples of tracing paths and of planning future follow-up studies on the same topic so that there are no paths distorting a generating dependence of interest. Small Gaussian families of distributions are used to illustrate independence properties of traceable regressions. In Section 4, several discrete families of distributions are given to show how the properties of traceable regressions can be violated. In Section 5, the known properties of an edge matrix calculus to transform graphs are collected first. These are used to derive new properties of transforming regression graphs and to distinguish traceable regressions from distributions that are faithful to regression graphs. A short discussion ends the paper.

2 Definitions and Terminology

2.1 Some Terminology for Graphs

Most of the following definitions are standard or evocative and listed for completeness. A graph consist of a *node set* $\mathbf{N} = \{1, \ldots d_N\}$ and of *edges* that couple node pairs. In simple

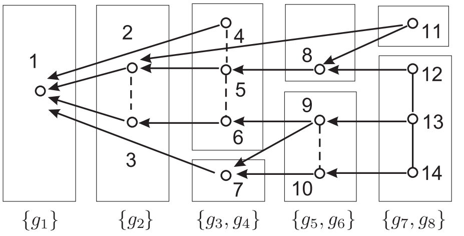


Figure 1. *A regression graph in 14 nodes and node set partitioned into 8 connected components; single responses in* g_1, g_4, g_5 *and joint responses in* g_2, g_3, g_6*; context variables in* g_7, g_8.

graphs, edges couple exclusively distinct node pairs by at most one edge so that the endpoints i and k of an *ik-edge* never coincide.

An *ik-path* connects the *path endpoints i* and *k* by a sequence of edges. An *ik*-path can be an edge, otherwise it has *distinct inner nodes* such that each edge visits an inner node once. For an *ik*-arrow, $i \prec\!\!-\!\!- k$, node k is commonly named the *parent* of node i. An *ik*-path is a *cycle* if its endpoints coincide, that is if $i = k$. A cycle is *directed* if it contains at least one arrow and one returns to the starting node by following the direction of the arrow(s).

In a regression graph, G^N_{reg}, such as in Figure 1, there are no directed cycles. There is a *split* of the node set, that is an ordered partitioning of N into two components, as $N = (u, v)$. Set u contains the response nodes, which are possibly coupled by dashed lines and each has possibly several parent nodes. Set v contains context nodes, which are possibly coupled by full lines and none has any parent node. There is an additional partial set ordering of u, defined by the direction of the arrows in the graph.

A regression graph has *three types of edge sets*, $E_{\prec\!-\!-}$ for directed dependences of responses on their regressors, $E_{- -}$, for undirected dependences among components of a joint response, and $E_{\_\_}$ for undirected dependences among context variables.

Nodes that are connected by an undirected path in G^N_{reg} are named *concurrent*. The undirected edges within any set of concurrent nodes of G^N_{reg}, are always of one type. Arrows may point to any node in u but never to a node in v. An arrow may couple nodes from two different sets of concurrent responses but only in such a way that no directed cycle results; see again Figure 1.

There is an *a-line path*, if all its inner nodes are in subset a of N. A path of only arrows is *direction-preserving* if all its arrows point in the same direction. For a, b arbitrary disjoint subsets of N, one says there is an *ik-path between b and a* if node i is in a and node k is in b or vice versa. For a direction-preserving *ik*-path that starts with node k in set b and ends at node i in set a, we say there is a *path from b to a*. In a direction-preserving *ik*-path, node k is named an *ancestor* of i and node i a *descendant* of k.

A *subgraph, induced by a subset a* of the node set N, consists of the nodes within a and of the edges present in the graph within a. A special type of induced subgraph, needed in this paper, consists of three nodes and two edges. It is named a V-*configuration* or just a V. Thus, a three-node path forms a V if its induced subgraph has two edges.

In a *complete graph*, every node pair is coupled by an edge. In a *connected subgraph*, every node can be reached by a path. The unique *set of connected components* of G^N_{reg} consists of single nodes and the undirected, connected subgraphs that remain when all arrows of G^N_{reg} are removed but all nodes and all other edges are retained. Whenever there is no arrow between two or more connected components g_j of G^N_{reg}, the graph alone implies only a partial ordering of its connected components. Then, there is more than one full *compatible ordering* such that, for $g_1 < g_2 < \ldots g_J$, arrows may start from a node in any g_j except g_1, but never point to a node in

$g_{>j} = g_{j+1} \cup g_{j+2}, \ldots, \cup g_J$. An important consequence is that different compatible orderings for G^N_{reg} define the same independence structure.

By convention, we number nodes and components g_j of G^N_{reg} first from top to bottom, then from left to right. In Figure 1, $g_3 = \{4, 5, 6\}$ and $g_8 = \{12, 13, 14\}$ contain three nodes, each of g_2 and g_6 contain two nodes, all others contain a single node; the connected components g_7 and g_8 are in v, all others are in u.

The connected components of two or more elements, that contain concurrent variables, indicate in statistical terminology which subsets of the variables are joint responses and which are joint context variables. Single responses correspond in the statistical model to univariate regressions, joint responses to multivariate regressions, including the seemingly unrelated regressions of Zellner (1962). In Figure 1, seemingly unrelated regressions belong to the subgraphs induced by each of the three node sets {2, 3, 5, 6}, {5, 6, 8, 9}, {9, 10, 13, 14}.

2.2 Generating Sequences of Regressions and Graphs

We consider now joint densities f_N of a d_N-dimensional, real-valued random vector variable Y_N, which may have discrete or continuous components or be a mixture of the two types. The density f_N is defined relative to the product distribution; see for instance Lauritzen & Wermuth (1989) for a more formal discussion. The variables have labels in node set N. In the following, an element i of N is not distinguished from the singleton $\{i\}$ and the union sign for combining subsets of N is often omitted.

For i, k a node pair and $c \subset N \setminus \{i, k\}$, we write $i \perp\!\!\!\perp k | c$ for Y_i, Y_k conditionally independent given Y_c. In terms of a joint conditional density f_{ikc}, this is equivalent to the following constraints on conditional densities:

$$i \perp\!\!\!\perp k | c \iff (f_{i|kc} = f_{i|c}) \iff f_{ik|c} = (f_{i|c} f_{k|c}).$$

It has become common to say that a joint family of densities f_N can be *generated over a chain graph* if it factorizes according to a full set ordering of the nodes, called a chain, and f_N satisfies all independences implied by the graph. Different types of chain graph and corresponding models for discrete variables are discussed by Drton (2009).

When independence structures are the focus of interest, one starts traditionally with the chain graph in any one compatible ordering of its connected components g_j. As stated before, regression graphs G^N_{reg} in node set N have three types of edge sets, $E_{\prec\!\!-}$, E_{--}, and $E_{—}$ and there is a split, $N = (u, v)$, so that response nodes are in u and context nodes in v.

The undirected subgraph induced by v has edges i —— k and is commonly called *a concentration graph*. The undirected subgraph induced by any g_j within u has edges i - - - k and is a *conditional covariance graph* given $g_{>j}$. With $g_{>J} = \emptyset$, the *basic factorization of* f_N generated over G^N_{reg} is

$$f_N = f_{u|v} f_v \text{ with } f_{u|v} = \prod\nolimits_{g_j \subseteq u} f_{g_j | g_{>j}} \text{ and } f_v = \prod\nolimits_{g_j \subseteq v} f_{g_j}. \tag{1}$$

The independence structure of any given G^N_{reg} can be derived in terms of this factorization and the constraints on the given densities implied by the missing edges.

When tracing of pathways is of main interest, one starts instead with a stepwise generating process of f_N for which $N = (u, v)$ and *one full ordering of* g_j is given. In this process, the density of variables in g_J is generated first, the one of g_{J-1} given g_J next, up to the density of g_1 given $g_{>1}$. Then, variable pairs needed to generate f_N define the edge set of G^N_{reg}; the factorization in equation (1) and pairwise independences result as a byproduct of the generating process.

For a *variable pair Y_i,Y_k needed in the generating process of f_N*, we say it is conditionally dependent given Y_c for some $c \subset N \setminus \{i, k\}$ and write $i \pitchfork k|c$; so that in particular, $i \perp\!\!\!\perp k|c$ does not hold for such a pair. A regression graph is said to be *edge-minimal* for a distribution generated over it, if every missing edge in the graph corresponds to a conditional independence statement and every edge present to a dependence. A family of densities f_N, generated over an edge-minimal graph, changes if any one edge is removed from the graph since then an additional independence holds in f_N.

Definition 1. **Defining pairwise dependences of G^N_{reg}.**
An edge-minimal regression graph specifies with $g_1 < \ldots < g_J$ a generating process for f_N, where the dependences

$$\begin{array}{lll} i\,\text{-}\,\text{-}\,\text{-}\,k : i \pitchfork k|g_{>j} & \text{for} & i, k \text{ concurrent response nodes in } g_j \text{ of } u, \\ i \prec\!\!\!\text{---}\, k : i \pitchfork k|g_{>j} \setminus \{k\} & \text{for} & \text{response node } i \text{ in } g_j \text{ of } u \text{ and node } k \text{ in } g_{>j}, \\ i \text{ --- } k : i \pitchfork k|v \setminus \{i, k\} & \text{for} & i, k \text{ concurrent context nodes in } g_j \text{ of } v, \end{array} \qquad (2)$$

define the **edges present** *in G^N_{reg}. The meaning of each* **edge missing** *in G^N_{reg} results with the dependence sign $\pitchfork$ replaced by the independence sign $\perp\!\!\!\perp$.*

Thus, for the given order of the components g_j, the graph implies for each variable pair i, k either conditional dependence or conditional independence given the same conditioning set, with $i\,\text{-}\,\text{-}\,\text{-}\,k$ for two response nodes, with $i \prec\!\!\!\text{---}\, k$ for i a response node in g_j and k a node in the past of g_j, with $i \text{ --- } k$ for two context nodes. Notice that each pair of concurrent responses, Y_i, Y_k with $i, k \in g_j$, is exclusively conditioned on variables that are in $g_{>j}$. This permits to model simultaneously occurring effects on responses with an intervention variable Y_l, for $l \in g_{>j}$. These types of joint effect cannot be modelled with a directed acyclic graph or with another type of chain graph than G^N_{reg}.

2.3 Compatible versus Covering Models

When different generating processes lead to the same regression graph, we say that there is another *compatible model* for the generated density f_N. Results on Markov equivalence prove this, but an intuitive argument is as follows. When there are several compatible orderings of the connected components, then some components, $g_j, g_{j+1}, \ldots, g_t$, say of G^N_{reg}, have an interchangeable labeling because they induce, taken jointly, disconnected undirected subgraphs. Such components are displayed in Figure 1 within stacked boxes.

In a connected G^N_{reg}, stacked response components $g_j, \ldots, g_t$ have the nodes in $g_{>t}$ as their *common past* and nodes in $g_{<j} = g_1 \cup g_2, \ldots, \cup g_{j-1}$ as their *common future*. Thus, for a given generating process, each arrow starting from a node in $g_j, g_{j+1}, \ldots, g_t$, points to response nodes in the common future but never to a node in the common past. For instance in Figure 1 above, the partial order implied by the arrows in $E_{\prec\!\text{---}}$ of G^N_{reg} remains unchanged if just the order of g_3 and g_4 is interchanged that is of the two disconnected undirected subgraphs induced by the set {4, 5, 6, 7}. The constraints on the joint distribution remain then also unchanged so that a compatible ordering of the g_j defines another process to generate the same joint family of distributions.

Recall that connected components of G^N_{reg} are uniquely obtained as the connected subgraphs that remain after deleting all arrows from the regression graph and keeping the undirected edges and all nodes. Thus, for any given graph, it is not necessary to show the stacked boxes, but they are sometimes included to reflect the *first ordering of N*, the prior knowledge about a sequence of joint and single responses and about context variables. The corresponding basic factorization is then for a f_N with a *complete regression graph* which has $N = (u', v')$, a single connected component v' and no independences.

A density f_N generated over a given G^N_{reg} can be regarded as a *reduced model*, as discussed by Cox & Wermuth (1990), to a partly completed regression graph as *covering model* whenever the ordering implied by the arrows of G^N_{reg} is retained. A small covering model is for instance obtained for Figure 1 by replacing g_3 and g_4 with a single dashed-line complete graph in node set $\{4, 5, 6, 7\}$. Thereby, the independence structure implied by G^N_{reg} is changed, but the partial order implied by the arrows, present in G^N_{reg} , is unchanged. Replacing for instance the subgraph of a seemingly regression by a complete graph, leads for a joint Gaussian distribution to simplified estimation in the covering model compared to the reduced model.

To each regression graph with $g_1 < \ldots < g_J$ used in the generating process with an edge-minimal G^N_{reg} , there exists a unique, fully *completed graph* which has the same set of context nodes and no missing edges and which respects the ordering given by $E_{\prec\!\!-\!\!-}$ of G^N_{reg}. It is obtained by inserting a full line for every missing edge within the context node set v, a dashed line for every missing edge within each set of stacked response components (with $g_j \subseteq u$) and arrows for all remaining missing edges pointing in the direction defined by $E_{\prec\!\!-\!\!-}$. The completed regression graph of Figure 1 has five connected components: $g_1' = g_1$, $g_2' = g_2$, $g_3' = g_3 \cup g_4$, $g_4' = g_5 \cup g_6$, $g_5' = g_7 \cup g_8$.

In the case of a large sample size n compared to d_N and prior information on an ordering, this provides a justification to start from a generating process for a complete regression graph and to search for simplifying independences that eventually lead to an edge-minimal graph.

2.4 General and Special Properties of Probability Distributions

For i, h, k single, distinct indices, a, b, c, d disjoint subsets of index set N, where only d may be empty, there are the common independence properties (i) to (iv) which are satisfied by all probability distributions. The discussed properties (v) to (viii) constrain distributions, but they permit the use of just the graph to derive different types of consequences for families of distributions f_N generated over G^N_{reg} .

(i) **symmetry:** $a \perp\!\!\!\perp b|c \iff b \perp\!\!\!\perp a|c$,
(ii) **contraction:** $(a \perp\!\!\!\perp b|cd \text{ and } b \perp\!\!\!\perp c|d) \iff ac \perp\!\!\!\perp b|d$,
(iii) **decomposition:** $a \perp\!\!\!\perp bc|d \Rightarrow (a \perp\!\!\!\perp b|d \text{ and } a \perp\!\!\!\perp c|d)$,
(iv) **weak union:** $a \perp\!\!\!\perp bc|d \Rightarrow (a \perp\!\!\!\perp b|cd \text{ and } a \perp\!\!\!\perp c|bd)$.

Joint distributions, for which the reverse implications of (iii) and of (iv) hold, have as additional properties, respectively,

(v) **composition:** $(a \perp\!\!\!\perp b|d \text{ and } a \perp\!\!\!\perp c|d) \Rightarrow a \perp\!\!\!\perp bc|d$,
(vi) **intersection:** $(a \perp\!\!\!\perp b|cd \text{ and } a \perp\!\!\!\perp c|bd) \Rightarrow a \perp\!\!\!\perp bc|d$.

Properties (v) and (vi) are needed to derive the independence structure implied by G^N_{reg}. Two further types of properties are to be considered for tracing pathways of dependence,

(vii) **set transitivity:** $(a \perp\!\!\!\perp b|d \text{ and } a \perp\!\!\!\perp b|cd) \Rightarrow (a \perp\!\!\!\perp c|d \text{ or } b \perp\!\!\!\perp c|d)$,

(viii) **singleton transitivity:** $(i \perp\!\!\!\perp k|d \text{ and } i \perp\!\!\!\perp k|hd) \Rightarrow (i \perp\!\!\!\perp h|d \text{ or } k \perp\!\!\!\perp h|d)$.

Distributions that satisfy set transitivity are also singleton-transitive, since a set c may contain only one element, but distributions that are singleton-transitive need not be set-transitive. For a conditional independence of Y_i, Y_k given only Y_d and given both Y_h and Y_d to hold, singleton transitivity requires that there is at least one additional independence given Y_d involving Y_h, the additional single variable in the conditioning set. For set transitivity, the single variable Y_h is replaced by a vector variable Y_c.

For Proposition 3 below, we shall show that with transformations of G^N_{reg}, by which no edge of the starting graph gets removed, set transitivity, (vii), is implicitly used, while for Proposition 1 only singleton transitivity, (viii), is needed in addition to (i) to (vi) to decide for a given edge-minimal G^N_{reg}, which Vs along a path are inducing a dependence for their endpoints.

Singleton transitivity is a feature of what we define below as traceable regressions. So far, it had been known to be common to all positive binary distributions where, for instance, for ($1 \pitchfork 2$ and $1 \pitchfork 3$) either $2 \perp\!\!\!\perp 3$ can hold or $2 \perp\!\!\!\perp 3|1$ but not both; see Simpson (1951). It also holds in all regular Gaussian distributions; see for instance Studený (2005), Corollary 2.5 in Section 2.3.6.

On the other hand, set transitivity imposes stronger constraints on distributions; see for instance Figure 1 for a trivariate binary distribution in Wermuth *et al.* (2009). It is confusing that, in the literature, the term "weak transitivity" has sometimes been used for property (vii) and sometimes for (viii). Set transitivity excludes some regular Gaussian families.

2.5 Regular Gaussian Families Violating Set Transitivity

For $N = (u, v)$, let Y_u and Y_v be mean-centered vector variables of equal dimension, $d_u = d_v$, having a regular joint Gaussian distribution. The components of Y_v be mutually independent with equal variances $\omega > 0$ so that $\text{cov}(Y_v) = \Sigma_{vv}$ equals the diagonal matrix ωI_{vv}, where I denotes an identity matrix. Let further $\Sigma_{uu} = \kappa I_{uu}$ with $\kappa\omega > 1$ and all elements in the correlation matrix of Y_u, Y_v be nonzero so that every component of Y_u is dependent on every component in Y_v. When in addition, $\Sigma_{uv} = \text{cov}(Y_u, Y_v)$ is orthogonal, so that $\Sigma_{uv}\Sigma_{vu} = I_{uu}$, the marginal independences specified with

$$\text{cov}(Y_u) = \Sigma_{uu} \text{ diagonal}, \quad \text{cov}(Y_v) = \Sigma_{vv} \text{ diagonal}.$$

carry over to conditional independences with

$$\text{cov}(Y_u|Y_v) = \Sigma_{uu|v} \text{ diagonal}, \quad \text{cov}(Y_v|Y_u) = \Sigma_{vv|u} \text{ diagonal},$$

and set transitivity is always violated for $d = \emptyset$, a split $v = (a, b)$ and $c = \{1, \dots, d_u\}$. This family extends the example in equation (8) of Cox & Wermuth (1993).

2.6 Some Important Properties of G^N_{reg} and f_N

Two basic types of Vs in G^N_{reg} need to be distinguished. There are ***collision* Vs**:

$$i \text{ - - -} \circ\!\!\longleftarrow k, \quad i \longrightarrow\!\circ\!\longleftarrow k, \quad i \text{ - - -} \circ \text{- - -} k,$$

and ***transmitting*Vs**:

$$i \longleftarrow \circ \longleftarrow k, \quad i \longleftarrow \circ \text{ —— } k, \quad i \text{ —— } \circ \text{ —— } k, \quad i \longleftarrow \circ \longrightarrow k, \quad i \longleftarrow \circ \text{- - -} k.$$

Recall that two different graphs in the same node set are Markov equivalent if they define the same independence structure, the set of all independences implied by the graph. The *skeleton* of a graph results by replacing each edge present by a full line.

LEMMA 1. **Markov equivalence**.
(Wermuth & Sadeghi, 2012). *Two regression graphs with the same skeleton are Markov equivalent if and only if their sets of collisionVs are identical.*

A more compact characterization of the pairwise independences in Definition 1 is based on the notion of anterior paths. Recall first that with $N = (u, v)$, there are only undirected full-line paths within v and there are only arrows pointing from v to u. An *anterior ik-path* is either a descendant-ancestor ik-path, or a context nodes ik-path, or a descendant-ancestor iq-path with a context-nodes qk-path attached to it,

$$i \prec\!\!-\, \overbrace{\circ \prec\!\!-\, \circ, \ldots, \circ \prec\!\!-\, q}^{\text{ancestors of } i} \underbrace{ -\!\!-\, \circ, \ldots, \circ -\!\!-\, k}_{} .$$

$$\underbrace{\circ \prec\!\!-\, \circ, \ldots, \circ \prec\!\!-\, q -\!\!-\, \circ, \ldots, \circ -\!\!-\,}_{\text{anteriors of } i}$$

The joint set of anteriors of nodes i and k is $\text{ant}_{ik} = \{\text{ant}_i \cup \text{ant}_k\} \setminus \{i, k\}$. For any subset c of N, the anterior set of nodes within c is defined similarly and denoted by ant_c.

The intersection (vi) and the composition property (v) are needed for the following Lemma 2 and Lemma 3. By using them, the independences attached to the missing edges of G^N_{reg} in Definition 1 reduce to the more compact statements $i \perp\!\!\!\perp k|\text{ant}_{ik}$ and this leads to the definition of an active path in G^N_{reg} and to the global Markov property proven by Sadeghi (2009) for a more general class of graphs.

Let $\{a, b, c, m\}$ partition N, where c denotes a conditioning set of interest for a, b and m the set of nodes to be ignored that is to be marginalized over. Only c, m or both may be empty sets. A *path in* G^N_{reg} *is active given* c if of its inner nodes, every collision node is in $c \cup \text{ant}_c$ and every transmitting node is in m. For graph transformation, such a path is also said to be *edge-inducing*.

LEMMA 2. **Global Markov property of $\mathbf{G^N_{reg}}$.**
(Sadeghi, 2009). *The regression graph* G^N_{reg} *implies* $a \perp\!\!\!\perp b|c$ *if and only if there is no active path in* G^N_{reg} *between* a *and* b *given* c.

LEMMA 3. **Equivalence of the pairwise and the global Markov property**.
(Sadeghi & Lauritzen, 2012). *The independence structure of* G^N_{reg} *is equivalently defined by its lists of the three types of missing edges and by its global Markov property.*

To derive dependence-inducing Vs, we note first for three-node graphs that by Definition 1, the inner node of each collision V is excluded from the defining conditioning set for its endpoints, while the inner node of each transmitting V is included in it. This observation is generalized with the following Lemma 4 that results from Lemma 2.

LEMMA 4. **Conditioning sets in $\mathbf{G^N_{reg}}$.**
The conditioning set of any independence statement implied by G^N_{reg} *for the endpoints of any of its Vs, includes the inner node if it is a transmitting V and excludes the inner node and all its descendants if it is collision V.*

Let a V in an edge-minimal G^N_{reg} have endpoints i, k and inner node o. Then by Definition 1 and Lemma 4, there is at least one c with $c \subseteq N \setminus \{i, k, \text{o}\}$ such that $i \perp\!\!\!\perp k|c$ is implied by G^N_{reg} if (i, o, k) is a collision V and $i \perp\!\!\!\perp k|\text{o}c$ if (i, o, k) is a transmitting V. Now, we want to express formally that properties (i) to (vi), needed already to derive implied independences, together with singleton transitivity assures for each V of an edge-minimal G^N_{reg}, that this graph implies a dependence for nodes i, k, when conditioning sets are changed with respect to the inner node of the V.

For this, we take an edge-minimal regression graph of Definition 1, assume properties (i) to (vi) and (viii) and say that G^N_{reg} forms then a *dependence base*.

PROPOSITION 1. **Dependence-inducing Vs.**
For (i, o, k) *any* V *of a dependence base* G^N_{reg} *and each* $c \subseteq N \setminus \{i, k, \mathrm{o}\}$ *for which this regression graph implies one of* $i \perp\!\!\!\perp k|c$ *or* $i \perp\!\!\!\perp k|\mathrm{o}c$, *the following two equivalent statements are implied by* G^N_{reg} :

$$- (i, \mathrm{o}, k) \text{ forms a collision V } \iff (i \perp\!\!\!\perp k|c \Rightarrow i \pitchfork k|\mathrm{o}c)$$
$$- (i, \mathrm{o}, k) \text{ forms a transmitting V } \iff (i \perp\!\!\!\perp k|\mathrm{o}c \Rightarrow i \pitchfork k|c)\,.$$

Proof. For three-node graphs, collision Vs are Markov equivalent and transmitting Vs are Markov equivalent by Lemma 1.

For $c = \emptyset$ and G^N_{reg} a dependence base, both edges of a V indicate a conditional dependence for pairs i,o and k,o and by Definition 1, $i \perp\!\!\!\perp k$ holds in f_N for an inner collision node and $i \perp\!\!\!\perp k|\mathrm{o}$ for an inner transmitting node. Including the inner node of a collision V into the conditioning set, or excluding the inner node of a transmitting V from the conditioning set, generates an active path by Lemma 2. Such a path induces a dependence unless singleton transitivity is violated which contradicts an assumption.

Similarly, for $c \neq \emptyset$, an independence is implied by G^N_{reg} if there is no active path between i and k given c by Lemma 4, and an active path that implies $i \pitchfork k|\mathrm{o}c$ is generated just as for $c = \emptyset$. □

We can now define sequences of regressions that permit the tracing of pathways of dependence for f_N when a, b, c, d denote disjoint subsets of N and only d may be empty.

Definition 2. **Traceable regressions.**
Sequences of regressions are traceable if their joint density f_N *is generated over an edge-minimal regression graph and properties (v), (vi), and (viii) of Section 2.4 hold*.

Thus, a density f_N that results with the generating process of Definition 1 satisfies in addition to the general properties (i) to (iv) also composition (v), intersection (vi), and singleton transitivity (viii); for sufficient conditions of each of (v), (vi) and (viii) see Section 4 below. Here, we summarize the characterizing properties.

COROLLARY 1. **Characterizing properties of traceable regressions.**
Traceable regression generated over an edge-minimal G^N_{reg} *have for disjoint subsets* a, b, c, d *of* N

- *three equivalent decompositions of the joint independence* $b \perp\!\!\!\perp ac|d$*:*

$$(b \perp\!\!\!\perp a|cd \text{ and } b \perp\!\!\!\perp c|d), (b \perp\!\!\!\perp a|d \text{ and } b \perp\!\!\!\perp c|d), (b \perp\!\!\!\perp a|cd \text{ and } b \perp\!\!\!\perp c|ad), \textit{and}$$

- *edge-inducing* V*s of the graph are dependence-inducing for* f_N.

The three decompositions of Corollary 1 combine the previously discussed properties (ii) to (vi) and property (i), symmetry of independences, holds trivially as in all probability distributions. Undirected edges correspond to symmetric dependence statements. For each arrow $i \prec\!\!\!- k$ in G^N_{reg}, symmetry of dependence holds in the following weak sense. From Definition 1 for i in g_j, there is some $c \subseteq g_{>j} \setminus k$ with $f_{i|kc} \neq f_{i|c}$ used in the generating process. Then, for Y_k regressed instead on Y_i, Y_c, also $f_{k|ic} \neq f_{k|c}$.

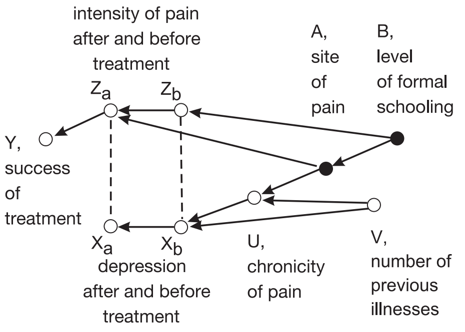


Figure 2. *Regression graph, well compatible with the data and resulting from statistical analyses. Binary variables are indicated by dots, variables treated as continuous by circles.*

Notice that traceable regression behave like regular Gaussian families generated over an edge-minimal G^N_{reg}. Therefore, for traceable regressions, a violation of set transitivity can occur only when there are at least two paths connecting the same node pair; see the family of regular Gaussian distributions given above that violates set transitivity and for further examples Wermuth & Cox (1998). We call these special types of parametric constellations *path cancellations* as they result for a pair i, k after combining dependences induced by active ik-paths in such a way that the joint contributions of all paths cancel.

3 Applications and Illustrations of Traceable Regressions

3.1 Tracing Paths

Whenever a pathway of dependence is traced in terms of an edge-minimal graph, one uses implicitly that every edge present represents a dependence that is strong enough to be of interest in the given substantive context and that every edge-inducing V along a path is dependence-inducing for its endpoints.

Figure 2 shows a well-fitting regression graph for nine features observed for patients. The regression graph represents a research hypothesis on the sets of regressors needed for each response to generate the joint distribution. In this example, we use data of Kappesser (1997) on 201 chronic pain patients, where variable descriptions and detailed statistical analyses are given in Wermuth & Sadeghi (2012), not in this paper.

The graph does not contain any information on the types of the dependence, but supplemented by estimates for the dependences, one can use the graph to interpret pathways of dependences. For instance the path Y, Z_a, A, B becomes active by marginalizing over its two inner transmitting nodes. This leads, together with the parameters estimated with linear and logistic models, to the following interpretations.

Patients with a higher level of formal schooling are more likely to have head or neck pain than back pain. For patients with head or neck pain, the intensity of pain is better reduced after treatment than for the back pain patients. For lower pain intensity scores after treatment, treatment is the more successful the lower the pain intensity. For higher pain intensity scores after treatment, there are no systematic changes in success of treatment.

The graphs in Figure 3 are consequences of the generating graph in Figure 2. Here, we want to discuss mainly the implications for success of treatment, Y, when some of the variables are not observed. The graphs are best derived using Proposition 2 in Section 5 below with, for Figure 3(a), a general marginalizing set $m_o = \{Z_a, X_a, Z_b, X_b\}$ and for Figure 3(b) the set $m_o = \{Z_a, X_a, Z_b, X_b, A\}$.

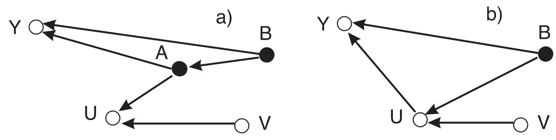


Figure 3. *The graph of Figure 2 transformed, preserving the original ordering for the remaining variables by (a) marginalizing over symptoms before and after treatment, X_a, Z_a, Z_b, X_b; (b) marginalizing over symptoms before and after treatment and, in addition, over site of pain, A.*

Alternatively, the new graphs may be derived by just applying Lemma 2 and Proposition 1 to an ordering given in terms of the arrows present in Figure 2 for the remaining responses. Thus for instance in Figure 3(b) for node $i = Y$ and another node $k \in \{U, B, V\}$, the conditioning set c contains, for each k, the remaining two nodes in this set and $m = m_o$. For node $i = U$ and one of its remaining potential regressors $k \in \{B, V\}$, the conditioning set c is, for each k, the single remaining node and $m = m_o \cup Y$.

An arrow, $i \prec\!\!- k$ is inserted in Figure 3(b) whenever there is an active path in Figure 2 for a given node pair (i, k) and the new conditioning set c for i. Thus for instance, the path Y, Z_a, A, U is active by marginalizing over Z_a, A and conditioning on B, V. One may proceed in a similar way to construct the graph in Figure 3(a).

Figure 3(a) implies that site of pain, A, would show a direct effect on Y if the two symptoms of chronic pain before and after treatment were either not measured or just omitted from the list of potentially important regressors. Similarly, chronicity of pain, U, would show a direct effect on Y if, in addition, site of pain, A, is omitted in 3b).

To derive and interpret transformed graphs larger than those in Figure 3, involving both marginalizing and conditioning, one is better off to use the general properties of transforming regression graphs given below in Section 5. For the analysis of data, one also has to realize that even a dependence induced by a two-edge path need not be strong. For instance, for two correlations of 0.4 associated with a transmitting V in the graph, the induced correlation given the inner node will only be of size 0.16 and hence would seem small if looked at alone, even though such a pathway of dependence may be important in many subject-matter contexts.

3.2 Planning Future Follow-up Studies

To show how tracing of active paths may lead to an improved planning of follow-up studies, we use the generating process, represented by the graph in Figure 4, adapted from Robins & Wasserman (1997), and assume that a strong dependence corresponds to each edge present in the graph.

Suppose that in the planned study, it will be possible to observe all variables of Figure 4 except for U, because the tools needed to diagnose the health status before treatment, U, will not be

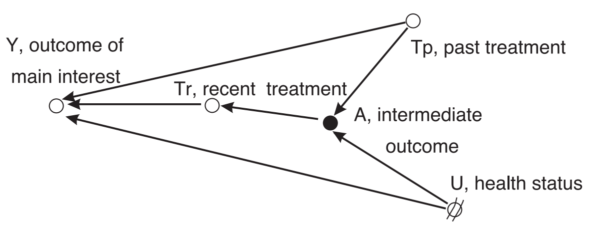


Figure 4. *Generating process in five variables, missing edge for (T_{p}, U) due to full randomized allocation of individuals to treatments, and missing edges for (T_{r}, U) and $(T_{\text{r}}, T_{\text{p}})$ due to randomization conditionally on A; U expected to be unobserved in a follow-up study.*

available. Marginalizing over U is indicated in Figure 4 by a crossed out node, $\not\!\phi$. With U excluded from any conditioning set for Y, the main response of interest, all remaining possible conditioning sets are explored. In general, whenever no active path is generated, one may proceed safely with estimating an effect, a dependence of main interest, directly in the follow-up study. Hence for Figure 4, the object is to find marginalizing sets for which there is no active path for (Y, T_r) or (Y, T_p).

With U unobserved, the dependence of Y on the past treatment T_p will always be modified, since by excluding also the intermediate outcome, A, and recent treatment, T_r from the list of regressors, one generates the active path Y, T_r, A, T_p, while by including either T_r or A or both as regressors for Y, one generates the active path Y, U, A, T_p; see Lemma 2. The former is an example of an overall effect deviating from a conditional effect and the latter is an *example of indirect confounding*.

If on the other hand, the dependences of Y on the recent treatment, T_r, is of main interest, then T_p is a common ancestor and the path Y, T_p, A, T_r becomes active by marginalizing over the inner nodes; an *example of direct confounding*. But no active path is generated between Y and T_r when A and T_p are regressors in addition to T_r, so that the conditional dependence of Y on T_r given A, T_p can be directly estimated.

Even though it may in principle be possible to recover the generating dependence given some distributional assumptions; see, e.g. Wermuth & Cox (2008), one needs to obtain very precise estimates to make any correction worthwhile since poorly estimated parameters may also lead to bad corrections.

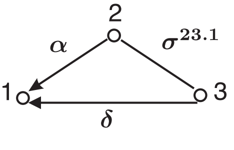
Both types of confounding can also be detected using graphical criteria on transformed graphs in reduced node sets, named summary graphs; see Wermuth (2011). For constructing summary graphs by removing repeatedly single nodes, one needs to take into account that any given node can be a collision node on one path and a transmitting node on another path. This contrasts with the graph transformations given below in Section 5, where different types of active paths are closed in sequence.

3.3 Examples of Small Gaussian Regression Graph Models

We illustrate next the intersection and the composition property by describing two different types of complete regression graphs in three nodes and the associated saturated models in the special case of regular families of Gaussian distributions for variables standardized to have zero mean and unit variance. Parameters are attached to the edges of the graphs. Example I shows in particular that the intersection property is implicitly used with backward selections of important regressors in multiple regressions and Example II how the composition property is relevant for selecting important regressors in multivariate regressions.

3.3.1 Example I: a complete single response graph with two context variables

The following complete graph in nodes 1, 2, 3



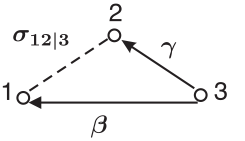
defines implicitly for standardized Gaussian variables, Y_1, Y_2, Y_3 three nonzero parameters measuring dependence in

$$E(Y_1|Y_2, Y_3) = \alpha Y_2 + \delta Y_3 \quad E(Y_2 Y_3) = \rho_{23} \quad \sigma^{23.1} = -\rho_{23}/\left(1 - \rho_{23}^2\right),$$

where ρ_{23} denotes the marginal correlation of Y_2, Y_3 and $\sigma^{23.1}$ the concentration in their bivariate distribution obtained by marginalizing over Y_1. For this complete graph, $\alpha \neq 0$ means $1 \pitchfork 2|3$, $\delta \neq 0$ means $1 \pitchfork 3|2$, and $\sigma^{23.1} \neq 0$ means $2 \pitchfork 3$. With $\alpha = \delta = 0$, one requires $1 \perp\!\!\!\perp 2|3$ and $1 \perp\!\!\!\perp 3|2$ and removes the 12-edge and the 13-edge from the complete graph so that node 1 remains isolated from 2 —— 3. For the resulting graph, the seemingly obvious interpretation $1 \perp\!\!\!\perp (2, 3)$ requires the intersection property, where in the statement of property (vi), we now have $a = 1$, $b = 2$, $c = 3$ and $d = \emptyset$.

3.3.2 *Example II: a complete joint response graph with a single regressor*

The following complete graph



defines for standardized Gaussian variables three non-vanishing parameters, β, γ, $\sigma_{12|3}$, in

$$\mathrm{E}(Y_1|Y_3) = \beta Y_3 \quad \mathrm{E}(Y_2|Y_3) = \gamma Y_3 \quad \mathrm{cov}(Y_1 Y_2|Y_3) = \sigma_{12|3}\,.$$

Here, $\sigma_{12|3} \neq 0$ means $1 \pitchfork 2|3$, $\beta \neq 0$ means $1 \pitchfork 3$, and $\gamma \neq 0$ means $2 \pitchfork 3$. With $\beta = \gamma = 0$, one requires $1 \perp\!\!\!\perp 3$ and $2 \perp\!\!\!\perp 3$ and removes the 13-edge and the 23-edge from the complete graph so that node 3 remains isolated from 1 - - - 2. For the resulting graph, the interpretation $(1, 2) \perp\!\!\!\perp 3$ requires the composition property, where in the statement of property (v), we now have $a = 3$, $b = 1$, $c = 2$, and $d = \emptyset$.

3.4 *Standard Properties for Combining Independences*

Properties (ii) to (iv) that are common to all probability distributions with a given density, are illustrated next by using the directed acyclic graphs in the three ordered nodes (1, 2, 3) shown in Figure 5, again for standardized Gaussian distributions.

3.4.1 *Example III: a complete directed acyclic graph*

The complete graph in nodes 1, 2, 3 of Figure 5(a) gives for standardized Gaussian variables three nonzero parameters, α, δ, γ, measuring dependence in

$$\mathrm{E}(Y_1|Y_2, Y_3) = \alpha Y_2 + \delta Y_3,\, \mathrm{E}(Y_2|Y_3) = \gamma Y_3,\, \mathrm{E}(Y_3) = 0\,,$$

where $\alpha \neq 0$ means $1 \pitchfork 2|3$, $\delta \neq 0$ means $1 \pitchfork 3|2$, and $\gamma \neq 0$ means $2 \pitchfork 3$.

The interpretation of δ changes with $\alpha = 0$; it then means $1 \pitchfork 3$ in Figure 5(b) where $1 \perp\!\!\!\perp 2|3$ is implied by the graph. This reflects that a different family of distributions is generated when

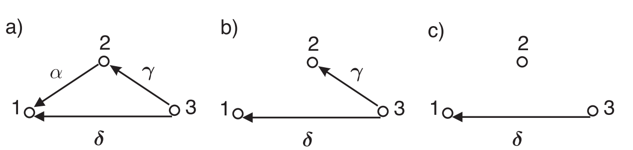


Figure 5. *Directed acyclic graphs in three nodes with parameters in standardized Gaussian distributions attached to the edges; (a) the complete graph, (b) the graph implying* $1 \perp\!\!\!\perp 2\,|\,3$*, (c) the graph implying* $2 \perp\!\!\!\perp (1, 3)$*.*

the 12-edge is removed. The graphs define implicitly the factorizations of f_N in equation (1), respectively, as

$$f_{123} = f_{1|23} f_{2|3} f_3, \quad (f_{123} = f_{1|3} f_{2|3} f_3) \Rightarrow 1 \perp\!\!\!\perp 2|3, \quad (f_{123} = f_{1|3} f_2 f_3) \Rightarrow 2 \perp\!\!\!\perp (1, 3)\,.$$

The factorization of a joint density as specified with a complete directed acyclic graph is formally always possible. Independence constraints imposed in sequence on two consecutive factors of f_{123} generated as in Figure 5(a), such as $1 \perp\!\!\!\perp 2|3$ constraining $f_{1|23} = f_{1|3}$, changes the triangle in the graph of Figure 5(a) to a V in Figure 5(b) and $2 \perp\!\!\!\perp 3$, constraining $f_{2|3} = f_2$, creates next an isolated node 2 and $1 \prec\!\!\!-\!\!\!- 3$, in Figure 5(c).

The removal of the two arrows gives one direction of the contraction property, starting from the factorization to Figure 5(c) gives the other direction. Given the factorization of any density to Figure 5(c), marginalizing over Y_3 leaves $f_{12} = f_1 f_2$ and marginalizing over Y_1 gives directly $f_{23} = f_2 f_3$ that is decomposition, while conditioning on Y_2, Y_3 leaves directly $f_{1|23} = f_{1|3}$ and conditioning on Y_1, Y_2 gives $f_{3|12} = f_{3|1}$ that is weak union. Also in more complex situations, these three properties, (ii), (iii), (iv), common to all probability distributions, can be derived by transforming factorized densities.

4 Violating Properties of Traceable Regressions

Some small discrete families of distribution are given that are not traceable regressions. These may be extended and many similar families may be constructed.

4.1 Violation of Singleton Transitivity

As mentioned before, singleton transitivity is satisfied in all regular Gaussian distributions and in all binary distributions. But the discrete family of distributions for a $2 \times 2 \times 3$ contingency table in Table 1, violates singleton transitivity. It is adapted from Birch (1963), equation (5.4). We write π_{ijk} for the joint probabilities of variables A, B, C at levels i, j, k and for instance $\pi_{+jk} = \sum_i \pi_{ijk}$. The conditional probabilities for A given B, C are $\pi_{i|jk} = \pi_{ijk}/\pi_{+jk}$.

Here, the conditional odds ratios being 1 imply that $A \perp\!\!\!\perp B|C$ and the marginal probabilities of A, C and of B, C show that $A \pitchfork C$ and $B \pitchfork C$. Nevertheless, also $A \perp\!\!\!\perp B$ since

$$\textstyle\sum_k \pi_{i+k}\, \pi_{+jk}/\pi_{++k} = \pi_{i++}\pi_{+j+},$$

a very special constellation discussed first by Darroch (1962) and generalized by Wermuth & Cox (2004), Section 7, to general types of distributions that are also not dependence-inducing. Though one can construct families of distributions with such peculiar parametric constraints, it is difficult to imagine that they could capture a structure of interest in any substantive context when studying sequences of regressions.

Table 1
A family of distributions that violates singleton transitivity.

	$4\pi_{ijk}(1+\alpha+\alpha^2)$, $\alpha > 1$					
	$C: k = 1$		$k = 2$		$k = 3$	
A/B :	$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 1$	$j = 2$
$i = 1$	α^2	α	α	1	1	α^2
$i = 2$	α	1	α^2	α	1	α^2
Odds-ratio	1		1		1	

Table 2
A family of distributions that violates the intersection property.

$3\pi_{ijk}, 0 < \alpha \neq \beta < 1, 2\alpha + \beta < 1$						
	$C : k = 1$		$k = 2$		$k = 3$	
A/B :	$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 1$	$j = 2$
$i = 1$	α	0	α	0	0	β
$i = 2$	$1 - \alpha$	0	$1 - \alpha$	0	0	$1 - \beta$

In a generating process of f_N, singleton transitivity can be achieved when the individual regressions are permitted to vary independently of the other response components and of their common past. This is reached, in particular, when the family corresponding to a completed graph has a rich enough parametrization and only the independence constraints of Definition 1 are imposed on G^N_{reg}.

4.2 *Violation of the Intersection Property*

The intersection property is satisfied in positive distributions and in all regular Gaussian distributions; the known necessary and sufficient conditions are less restrictive; see San Martin *et al.* (2005).

The discrete family of distributions in Table 2 for a $2 \times 2 \times 3$ contingency table violates the intersection property. This violation occurs whenever a pair of variables shares some common information. For three binary variables, violation of the intersection property coincides with the degenerate case of two variables being identical.

In the family shown in Table 2, $A \perp\!\!\!\perp B|C$ and $A \perp\!\!\!\perp C|B$, since

$$\pi_{i|jk} = \pi_{i|k} \text{ and } \pi_{i|jk} = \pi_{i|j}$$

but $A \pitchfork BC$. More precisely, $A \pitchfork B$ since $\pi_{i|j} \neq \pi_i$ and $A \pitchfork C$ since $\pi_{i|k} \neq \pi_i$. The marginal joint distribution of B, C shows the type of common information shared by variables B and C. Variable B taking on level 1 coincides with C taking on value 1 or 2 and B being at level 2 coincides with C being at level 3.

Thus, when the joint distribution of B, C had been generated by first knowing the distribution of variable C and then generating the conditional distribution of B given C, the levels of variable B are not permitted to vary freely and thereby lead to the violation of the intersection property.

4.3 *Violation of the Composition Property*

The composition property is always satisfied in regular Gaussian distributions and in multivariate symmetric binary distributions generated over directed acyclic graphs; see Wermuth *et al.* (2009). On the other hand, it is always violated when pairwise independences do not imply mutual independence.

The binary family of distributions in Table 3 for a $2 \times 2 \times 2$ contingency table also violates the composition property. In this family, there is a log-linear three-factor interaction since the conditional odd-ratios for A, B differ at the two levels of C.

More precisely, at level 2 of C, the variables A, B are independent while the dependence of this pair is strong at level 1 of C whenever α is large. At the same time, the marginal AC and BC tables reveal that $A \perp\!\!\!\perp C$ and $B \perp\!\!\!\perp C$.

Thus, when regressing the two components of a joint response AB separately on C, one sees no separate effects, but the conditional dependence of A on B changes with the levels of C.

Table 3
A family of distributions that violates the composition property.

	$8\pi_{ijk}$, $0 < 2\alpha < 1$			
	$C: k = 1$		$k = 2$	
A/B :	$j = 1$	$j = 2$	$j = 1$	$j = 2$
$i = 1$	$1 + 2\alpha$	$1 - 2\alpha$	1	1
$i = 2$	$1 - 2\alpha$	$1 + 2\alpha$	1	1
odds-ratio	$\{(1 + 2\alpha)/(1 - 2\alpha)\}^2$		1	

This type of structure could in particular not be generated by a single unobserved common explanatory variable or if all sets of variables with higher-order effects also have main effects in the regressions, or equivalently, have two-factor interactions in the joint distribution of the set of responses and its regressors, when the higher-order interactions are ignored.

With a pragmatic strategy for model selection in which one checks for higher-order interactions only when there are also main effects, one may overlook such structures that could be of substantive interest. For sequences of discrete joint responses, the violation will be detected when using the parametrization suggested by Marchetti and Lupparelli (2011). In general, the graphical checks for nonlinearities and interactions, as proposed by Cox and Wermuth (1994), provide some protection, but only for effects that are detectable also in marginal trivariate distributions.

5 Transforming Regression Graphs

The transformations of regression graphs to be introduced, are based on binary matrix representations of G^N_{reg}. Our notation for these edge matrices mimics the one for parameter matrices in Gaussian sequences of regressions generated over the graph. There are one-to-one correspondences between a zero in an edge matrix, a vanishing parameter in the regular Gaussian family of distributions and a conditional independence statement.

5.1 Linear Sequences of Regressions

For a mean-centered vector variable Y_N with a regular Gaussian distribution generated over G^N_{reg} with a split $N = (u, v)$, the matrix of equation parameters, denoted by H_{NN}, is upper block-triangular and

$$H_{NN} Y_N = \eta_N \text{ with } W_{NN} = \text{cov}(\eta_N) \text{ block-diagonal in the sizes of } g_j,$$

where the submatrix of H_{uu} in rows g_j and columns $g_{>j}$ is $-\Pi_{g_j|g_{>j}}$, the negative of the population least-squares coefficient matrix obtained when regressing Y_{g_j} on $Y_{>g_j}$. The square diagonal submatrices in the sizes of g_j are identity matrices. The submatrix H_{vv} is the marginal concentration matrix of Y_v, denoted by $\Sigma^{vv.u}$. This implies $W_{vv} = \Sigma^{vv.u}$. The square submatrices of W_{uu} are $\Sigma_{g_j g_j|g_{>j}}$, the conditional covariance matrices of Y_{g_j} given $Y_{>g_j}$. For just two connected response components a, b the parameter matrices are

$$H_{NN} = \begin{pmatrix} I_{aa} & -\Pi_{a|b.v} & -\Pi_{a|v.b} \\ 0_{ba} & I_{bb} & -\Pi_{b|v} \\ 0_{va} & 0_{vb} & \Sigma^{vv.ab} \end{pmatrix} W_{NN} = \begin{pmatrix} \Sigma_{aa|bv} & 0_{ab} & 0_{av} \\ 0_{ba} & \Sigma_{bb|v} & 0_{bv} \\ 0_{va} & 0_{vb} & \Sigma^{vv.ab} \end{pmatrix},$$

where we use a Yule-Cochran notation. The regression coefficient matrix when Y_a is regressed on Y_b, Y_v is denoted by $\Pi_{a|bv}$, and the coefficient matrix of Y_b in this regression by $\Pi_{a|b.v}$. For instance 0_{ba} denotes a matrix of zeros, and I_{bb} an identity matrix.

For a new split $N = (a, b)$, to obtain $f_{a|b} f_b$ we let $c = a \cap u$, $d = b \cap u$, and get

$$K_{NN} = \begin{pmatrix} H_{aa}^{-1} & -H_{aa}^{-1} H_{ab} \\ H_{ba} H_{aa}^{-1} & H_{bb} - H_{ba} H_{aa}^{-1} H_{ab} \end{pmatrix} Q_{uu} = \begin{pmatrix} W_{cc} - W_{cd} W_{dd}^{-1} W_{dc} & W_{dd}^{-1} W_{dc} \\ -W_{dd}^{-1} W_{dc} & W_{dd}^{-1} \end{pmatrix},$$

by partial inversion of H_{NN} with respect to a and by partial inversion of W_{uu} with respect to d; see for instance Marchetti & Wermuth (2009), Appendix 1.

LEMMA 5. **Orthogonalized linear equations**
(Wermuth & Cox, 2004, Theorem 1, and Wermuth, 2011, equation 2.11). *The Gaussian density* $f_N = f_{u|v} f_v$ *generated over* G_{reg}^N *is for any split* $N = (a, b)$ *transformed into* $f_N = f_{a|b} f_b$ *with* $E(Y_a|Y_b) = \Pi_{a|b}$, $\text{cov}(Y_a|Y_b) = \Sigma_{aa|b}$, $\text{con}(Y_b) = \Sigma^{bb.a}$ *with*

$$\Pi_{a|b} = K_{ab} + K_{aa} Q_{ab} K_{bb}, \tag{3}$$

$$\Sigma_{aa|b} = K_{aa} Q_{aa} K_{aa}^{\mathrm{T}},\ \Sigma^{bb.a} = H_{bb}^{\mathrm{T}} Q_{bb} H_{bb}. \tag{4}$$

5.2 The Edge Matrices of Regression Graphs

Edge matrices are binary matrix representations of graphs. They are symmetric for undirected graphs, upper block-triangular for arrows in a generating G_{reg}^N and upper-triangular for directed acyclic graphs. The essential change compared to the more traditionally used adjacency matrices is that ones are added along the diagonal of each square matrix. This has the effect that sums of matrix products are well defined and can represent the closing of special types of path in graphs; such as in equations (8) and (9) below.

Recall that regression graphs have three types of edge sets, $E_{\prec\!\!-\!\!-}$, E_{---}, and $E_{\_\_\_}$. The edge matrix components of G_{reg}^N are a $d_N \times d_N$ upper block-triangular matrix $\mathcal{H}_{NN} = (\mathcal{H}_{ik})$ such that

$$\mathcal{H}_{ik} = \begin{cases} 1 \ \text{if and only if} \quad i \prec\!\!-\!\!-k \quad \text{or} \quad i \text{ ---}k \text{ in } G_{\text{reg}}^N \text{ or } i = k, \\ 0 \ \text{otherwise,} \end{cases} \tag{5}$$

and a $d_u \times d_u$ symmetric matrix $\mathcal{W}_{uu} = (\mathcal{W}_{ik})$ such that

$$\mathcal{W}_{ik} = \begin{cases} 1 \ \text{if and only if} \quad i\text{ - - - }k \text{ in } G_{\text{reg}}^N \quad \text{or} \quad i = k, \\ 0 \ \text{otherwise,} \end{cases} \tag{6}$$

where, E_{---} corresponds to $\mathcal{W}_{uu}$, $E_{\_\_\_}$ to $\mathcal{H}_{vv}$, and $E_{\prec\!\!-\!\!-}$ to $\mathcal{H}_{uN}$.

Every regression graph G_{reg}^N can be represented by its edge matrices given in equations (5) and (6). Every dependence base G_{reg}^N defines in particular a corresponding family of Gaussian regressions in which each edge present can be identified by a single non-vanishing parameter, an off-diagonal element of H_{NN} or W_{uu}.

5.3 Partial Closure of Paths

Partial closure, introduced by Wermuth *et al.* (2006), is a matrix operator, denoted by $\text{zer}_a(\cdot)$ which acts on row and columns a of a binary matrix. It is applied to edge matrix representations of a starting graph in node set N to give the edge matrix representations of a new graph in which there is an additional ik-edge for a pair i, k that is in the starting graph uncoupled but connected by a specific type of edge-inducing a-line path.

With partial closure, the set of nodes, node labels, and edges present in the starting graph, are preserved in the transformed graph so that the mappings are graph homomorphisms; for this

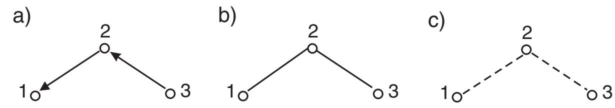


Figure 6. *Dependence base, 3-node graphs: a* V *in a (a) directed acyclic, (b) concentration, (c) covariance graph; an active path (1,2,3) induces in (a) and (b)* $1 \pitchfork 3$ *and in (c)* $1 \pitchfork 3|2$.

notion see Hell & Nešetřil (2004), for corresponding reparametrizations of exponential families see Wiedenbeck & Wermuth (2010).

LEMMA 6. **Basic properties of partial closure.**
(Wermuth *et al.*, 2006). *Partial closure is (i) commutative, (ii) cannot be undone and (iii) is exchangeable with selecting a submatrix.*

By property (i), it is enough, for some purposes, to show how the operator acts on a single node. By property (ii), independences can be removed but never reintroduced. This property is essential to understand that these transformations use implicitly set transitivity. Property (iii) justifies node and edge reductions since closing edge-inducing a-line paths in a large graph and then selecting a square submatrix for a subset containing a, gives the same result as selecting the square submatrix first and then closing the a-line paths. This property is needed for Proposition 2.

Because of property (i), one can always permute the matrix $\mathcal{F}$ into $\tilde{\mathcal{F}}$ and start partial closure with node i corresponding to position (1,1) of $\tilde{\mathcal{F}}$. With a set of nodes, say a, to be operated on to give $\text{zer}_a\, \tilde{\mathcal{F}}$ as a final output, the matrix $\text{zer}_i\, \tilde{\mathcal{F}}$ is – at least conceptually – permuted first back to the original order and next so that another element of a is chosen as index i, to be placed in row 1 and column 1 of the new matrix $\tilde{\mathcal{F}}$, and so on.

For $b = N \setminus \{i\}$, one gets as the transformed edge matrix at each step

$$\text{zer}_i\, \tilde{\mathcal{F}} = \text{In}[\begin{pmatrix} 1 & \mathcal{F}_{ib} \\ \mathcal{F}_{bi} & \mathcal{F}_{bb} + \mathcal{F}_{bi}\mathcal{F}_{ib} \end{pmatrix}]. \tag{7}$$

This says that particular Vs in the graph are closed which have node i as inner node. In the three small examples of Figure 6, an edge for node pair 1, 3 is induced with $i = 2$.

Applying zer_i to the edge matrix of a directed acyclic graph, covariance graph or concentration graph mimics, respectively, the recursion relation for regression coefficients, covariances, and concentrations; discussed for instance in Wermuth & Cox (1998).

By letting the edge induced by the three Vs in Figure 6, "remember the type of edge at the path endpoints," the induced edges become, respectively,

$$\text{a) } 1 \prec\!\!\!-\!\!\!-\, 3, \;\; \text{b) } 1 \;—\; 3, \;\; \text{c) } 1 \text{ - - - } 3.$$

The transformation $\text{zer}_a \mathcal{F}$ means that all Vs along a-line paths represented by the edge matrix $\mathcal{F}$ are closed by an edge. The basic property (i) implies that the nodes in a may be chosen for this in any order. This requires in particular that the inner nodes of the paths of $\mathcal{F}$ are of the same type, either all are collision nodes to form *collision paths*, or all are transmitting nodes.

For two graph transformations, we take the edge matrices of G^N_{reg}, $\mathcal{H}_{NN}$ of (5) and $\mathcal{W}_{uu}$ of (6), and again, as for Lemma 5, a new split $N = (a, b)$ and $c = a \cap u$, $d = b \cap u$.

LEMMA 7. **Partial closure applied to $\mathbf{G}^{\mathbf{N}}_{\mathbf{reg}}$.**
The transformation $\mathcal{K}_{NN} = \text{zer}_a\mathcal{H}_{NN}$ *closes each a-line anterior path and* $\mathcal{Q}_{uu} = \text{zer}_d\mathcal{W}_{uu}$ *each dashed, d-line collision path.*

Proof. Each anterior path in G^N_{reg} and no other type of path is represented by $\mathcal{H}_{NN}$ and each dashed-line path in G^N_{reg} and no other type of path is represented by $\mathcal{W}_{uu}$. Applying partial closure of 7 to $\mathcal{H}_{NN}$ on all nodes of a and to $\mathcal{W}_{uu}$ on all nodes of d, leads to the closing of the stated a-line and d-line paths, as required by Proposition 1. Remembering the type of edge at the endpoints of each V on an a-line path of $\mathcal{H}_{NN}$ leads to the same induced edge for the endpoints of the path, irrespective of the order in choosing single nodes of a. □

5.4 *Closing Active Paths in Regression Graphs*

For directed acyclic graphs, it is known that the path criterion on the starting graph for separation of α from β given c can be reduced to an edge criterion after transforming first the generating graph in terms of partial closure and closing next the remaining paths that are relevant for deciding whether $\alpha \perp\!\!\!\perp \beta | c$ is implied; see Marchetti & Wermuth (2009). This approach is now extended to regression graphs and to dependences in traceable regressions. For this, we take the partitioning $N = \{\alpha, \beta, c, m\}$ of the node set of G^N_{reg}, $a = \alpha \cup m$, $b = \beta \cup c$, and

$$\mathcal{K}_{NN} = \text{zer}_a\mathcal{H}_{NN},\ \mathcal{Q}_{uu} = \text{zer}_b\mathcal{W}_{uu},\ \mathcal{Q}_{uv} = 0,\ \mathcal{Q}_{vv} = \mathcal{K}_{vv}.$$

PROPOSITION 2. **Induced edge matrices for $\mathbf{f_{a|b}f_b}$.**
Sequences of regressions with graph G^N_{reg} *in node set* $N = (u, v)$ *and generating edge matrices* H_{NN} *and* W_{uu} *imply for* $f_{a|b}f_b$, *as induced edge matrices of the regression graph* $G^{N-a|b}_{\text{reg}}$:

$$\mathcal{P}_{a|b} = \text{In}[\mathcal{K}_{ab} + \mathcal{K}_{aa}\mathcal{Q}_{ab}\mathcal{K}_{bb}], \tag{8}$$

$$\mathcal{S}_{aa|b} = \text{In}[\mathcal{K}_{aa}\mathcal{Q}_{aa}\mathcal{K}^{\text{T}}_{aa}],\ \mathcal{S}^{bb.a} = \text{In}[\mathcal{H}^{\text{T}}_{bb}\mathcal{Q}_{bb}\mathcal{H}_{bb}]. \tag{9}$$

Proof. Partial closure mimics transformations of partial inversion such that all elements of the induced matrices are non-negative. The zero entries in equations (3) and (4) coincide with those in (8) and (9), non-zero entries in the former correspond to ones in the latter; see Lemma 3 of Marchetti & Wermuth (2009) for more detail. □

Of the active paths, defined for Lemma 2 and needed to decide for uncoupled pairs i, k of G^N_{reg} whether they are coupled in $G^{N-a|b}_{\text{reg}}$, some remain uncoupled after applying $zer_a\mathcal{H}_{NN}$ and $\text{zer}_d\mathcal{W}_{uu}$ in Lemma 7, but get closed with the non-negative sums of edge matrix products in (8) and (9). As with partial closure, no edges get ever removed with the latter types of graph transformations so that set transitivity is used implicitly.

For $N = (a, b)$ as for Proposition 2, let o_a denote nodes in a and o_b nodes in b.

COROLLARY 2. *For* i, k *the endpoints of paths that are edge-inducing for* $G^{N-a|b}_{\text{reg}}$, *three types of* ik*-path, still uncoupled in the graph having edge matrices* $\mathcal{K}_{NN}$ *and* $\mathcal{Q}_{uu}$,

$$i \prec\!\!\!-\!\!\!-\, \text{o}_a - - - \text{o}_b \prec\!\!\!-\!\!\!-\, k, \quad i \prec\!\!\!-\!\!\!-\, \text{o}_a - - - \text{o}_a \,-\!\!\!-\!\!\!\succ k, \quad i \,-\!\!\!-\!\!\!\succ \text{o}_b - - - \text{o}_b \prec\!\!\!-\!\!\!-\, k,$$

are closed with the induced edge matrices $\mathcal{P}_{a|b}$, $\mathcal{S}_{aa|b}$, $\mathcal{S}^{bb}$, *respectively, in (8) and (9).*

After remembering the types of edge at the path endpoints, we have with $\mathcal{P}_{a|b}$ an induced bipartite graph of arrows pointing from b to a, with $\mathcal{S}_{aa|b}$ an induced conditional covariance graph, and with $\mathcal{S}^{bb.a}$ an induced concentration graph.

LEMMA 8. **Edge matrices induced by $\mathbf{G}^{\mathbf{N}}_{\mathbf{reg}}$ for $\mathbf{f}_{\alpha\beta|c}$.**
The subgraph induced by nodes $\alpha \cup \beta$ in $G^{N-a|b}_{\text{reg}}$ captures the independence implications of G^N_{reg} for $f_{\alpha|\beta c} f_{\beta|c}$.

Proof. By the interpretation of the edge matrix components $\mathcal{P}_{a|b}, \mathcal{S}_{aa|b}, \mathcal{S}^{bb.a}$, no edges are induced by taking

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

Jointly, these edge submatrices define the subgraph induced by $\alpha \cup \beta$ in $G^{N-a|b}_{\text{reg}}$. □

The induced graphs in node set $\alpha \cup \beta$ and $G^{N-a|b}_{\text{reg}}$ in node set N, are examples of independence-predicting graphs in contrast to independence-preserving graphs such as the ribbonless graphs of Sadeghi & Lauritzen (2012) and their different types of Markov-equivalent graphs, such as summary graphs. With *independence-preserving graphs*, one can derive effects of additional marginalizing and conditioning in the starting graph while *independence-predicting graphs* can, in general, only be used to decide on edges present or missing in the induced graph.

PROPOSITION 3. **Edge criteria for implied independences and dependences**
A dependence base G^N_{reg} implies $\alpha \perp\!\!\!\perp \beta|c$ if $\mathcal{P}_{\alpha|\beta.c} = 0$ and it implies $\alpha \pitchfork \beta|c$ if $\mathcal{P}_{\alpha|\beta.c} \neq 0$.

Proof. The statement results with Lemma 7, equation (8) and Lemma 8. □

Proposition 3 states when a derived edge matrix implies a dependence or an independence. These are statements for induced families of distributions when one is starting from a family of traceable regressions. Recall that paths in an edge-minimal G^N_{reg} are traceable whenever the necessary and sufficient conditions of Corollary 1 hold or the set of sufficient conditions for properties (v), (vi), (viii) of Section 2.4 given in Section 4.

For many recently obtained theoretical results, it has been assumed that the studied distributions are faithful to a graph so that they satisfy precisely the independences implied by a graph and no others. By Propositions 1, 2, and 3, this means for G^N_{reg} that generated sequences of regressions have to be traceable and satisfy in addition set transitivity. Faithful distributions can now be equivalently characterized with the following Corollary 3. However, testable criteria for faithfulness are still unknown.

5.5 Distributions Satisfying All and Only the Independences Captured by G^N_{reg}

A given distribution is said to be faithful to a graph if every of its independence constraints is captured by a given independence graph; see Spirtes *et al.* (1993). For a distribution to be faithful to G^N_{reg}, it has to satisfy the properties needed for the graph transformations of Proposition 3, that is properties (i) to (vii) of Section 2.4.

COROLLARY 3. **Distributions that are faithful to $\mathbf{G}^{\mathbf{N}}_{\mathbf{reg}}$**

For a distribution with density f_N generated over a dependence base G^N_{reg}, the following statements are equivalent

(i) *the distribution is faithful to G^N_{reg},*
(ii) *every independence and every dependence statement implied by G^N_{reg} holds for f_N,*
(iii) *f_N satisfies as additional properties: composition, intersection and set transitivity,*
(iv) *f_N can be generated as a traceable regression without any path cancellations.*

In general, faithfulness imposes an additional strong constraint on traceable sequences of regressions. Exceptions are directed acyclic graphs in which each response has only one parent. But the most common situation in observational and in interventional studies is to have two or more regressors influencing a response. Therefore, for using regression graphs to interpret such structures or to plan future studies with a subset of the variables in a subpopulation, it is not sensible to assume that a given distribution is faithful to this graph. For a faithful distribution to G^N_{reg}, traceable regressions are generated such that no path cancellations occur. Thus, faithfulness is a sufficient but not a necessary condition for tracing pathways of dependence.

6 Discussion

Sequences of regressions in joint responses permit to model changes in several response components occurring at the same time when there is an intervention. This contrasts with interventions in sequences of regressions in only single responses and in other types of chain graph models.

We have identified properties of sequences of regressions in essentially arbitrary joint and single response variables and named them traceable regressions. A corresponding regression graph, G^N_{reg} is a dependence base of the joint distribution in addition to capturing the independences in the regressions. One knows now that the independence structure of such traceable regressions can differ from the implications derived in terms of its generating regression graph only when there are path cancellations.

The consequences derivable with a graph give changes in structure that result in families of distributions generated over the graph while one may not be able to generalize to this family from the structure that one can see for a distribution with one given set of parameters, for instance as estimated in a sample.

Sequences of traceable regressions and a given G^N_{reg} have implications for a regression of Y_a on Y_b and dependences of Y_b alone when these are based on a reordered node set $N = (a, b)$ that can be expressed with transformed edge matrix components of G^N_{reg}. By modifying G^N_{reg} with a marginalizing set $a = \alpha \cup m$ and a conditioning set $b = \beta \cup c$, the specific implications of G^N_{reg} for the conditional densities $f_{\alpha|\beta c}$ and $f_{\beta|c}$ can now be derived with a subgraph induced by $\alpha \cup \beta$ in this transformed graph. An edge matrix criterion gives the global Markov property of G^N_{reg} and it detects, in addition, when all Vs on selected paths induce dependences in the generated families of traceable regressions.

Many new questions have opened up. These include different types of sufficient conditions on a given distribution under which it represents a traceable regression, conditions on independence-predicting graphs which assure that they are also independence-preserving, applications such as the special details needed to improve existing methods for meta-analyses, or computational aspects, such as conditions under which one type of several equivalent graph transformations becomes computationally much less intensive than others.

Acknowledgements

The work reported in this paper was undertaken during the tenure of a Senior Visiting Scientist Award by the International Agency of Research on Cancer. I thank G. Byrnes, D.R. Cox, G.M. Marchetti, K. Sadeghi, I. Solis-Trapala, and the referees for their most constructive comments.

References

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Stat. Soc. B*, **25**, 220–233.

Castillo, E., Gutiérrez, J.M. & Hadi, A.S. (1997). *Expert Systems and Probabilistic Network Models.* New York: Springer.

Cox, D.R. & Wermuth, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika*, **77**, 747–761.

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

Cox, D.R. & Wermuth, N. (1994). Tests of linearity, multivariate normality and adequacy of linear scores. *J. Roy. Stat. Soc. C*, **43**, 347–355.

Darroch, J.N. (1962). Interactions in multi-factor contingency tables. *J. Roy. Stat. Soc. B*, **24**, 251–263.

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

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

Edwards, D. & Lauritzen, S.L. (2001). The TM algorithm for maximising a conditional likelihood function. *Biometrika*, **88**, 961–972.

Flesch, I. & Lucas, P. (2006). Markov equivalence in Bayesian networks. In *Advances in Probabilistic Graphical Models*, Eds. P. Lucas, J. Gámez & E. Samerón, pp. 3–38, Berlin: Springer.

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

Hell, P. & Nešetřil, J. (2004). *Graphs and Homomorphisms.* Oxford: Oxford University Press.

Kappesser, J. (1997). *Bedeutung der Lokalisation für die Entwicklung und Behandlung chronischer Schmerzen.* Thesis, Department of Psychology, University of Mainz.

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

Lněnička, R. & Matúš, F. (2007). On Gaussian conditional independence structures. *Kybernetika*, **43**, 323–342.

Marchetti, G.M. & Lupparelli, M. (2011). Chain graph models of multivariate regression type for categorical data. *Bernoulli*, **17**, 827–844.

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

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* San Mateo: Morgan Kaufmann.

Robins, J. & Wasserman, L. (1997). Estimation of effects of sequential treatments by reparametrizing directed acyclic graphs. In *Proc. 13th Ann. Conf., UAI*, Eds. D. Geiger & O. Shenoy, pp. 409–420. San Mateo: Morgan and Kaufmann.

Sadeghi, K. (2009). Representing modified independence structures. *Transfer Thesis. Oxford, UK: Oxford University.*.

Sadeghi, K. & Lauritzen, S.L. (2012). Markov properties of mixed graphs. Available at http://arxiv.org/abs/1109.5909.

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

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

Studený, M. (2005). *Probabilistic Conditional Independence Structures*. London: Springer.

Wermuth, N. (2011). Probability models with summary graph structure. *Bernoulli*, **17**, 845–879.

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

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

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

Wermuth, N., Cox, D.R. & Marchetti, G.M. (2006). Covariance chains. *Bernoulli*, **12**, 841–862.

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

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

Wermuth, N. & Sadeghi, K. (2012). Sequences of regressions and their independences (with discussion). *TEST*, **21**, 215–279.

Wermuth, N., Wiedenbeck, M. & Cox, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Math.*, **46**, 883–901.
Wiedenbeck, M. & Wermuth, N. (2010). Changing parameters by partial mappings. *Statist. Sinica*, **20**, 823–836.
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.*, **57**, 348–368.

Résumé

Dans cet article, nous définitions et étudions le concept de traçabilité des régressions et l'appliquer á quelques exemples. Régressions traçables sont des séquences de distributions conditionnelles dans les réponses individuelles ou conjointes pour lesquelles un graphe correspondant capte non seulement une structure indépendance, mais représente, en outre, dépendances conditionnelles qui permettent le traçage des voies de la dépendance. Nous donnons les propriétés nécessaires pour transformer ces graphes et des critères graphiques de décider si un chemin dans le graphe induit une dépendance. Les contraintes beaucoup plus fortes sur les distributions qui sont fidèles á un graphe sont comparés á ceux nécessaires pour les régressions traçables.

[*Received November 2011, Accepted July 2012*]