J. R. Statist. Soc. B (1990)
52, *No.* 1, *pp.* 21–50

On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Chain Models

By NANNY WERMUTH† and STEFFEN LILHOLT LAURITZEN

University of Mainz, Federal Republic of Germany *Aalborg University, Denmark*

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, April 12th, 1989*, Professor A. P. Dawid *in the Chair*]

SUMMARY

Graphs consisting of points, and lines or arrows as connections between selected pairs of points, are used to formulate hypotheses about relations between variables. Points stand for variables, connections represent associations. When a missing connection is interpreted as a conditional independence, the graph characterizes a conditional independence structure as well. Statistical models, called graphical chain models, correspond to special types of graphs which are interpreted in this fashion. Examples are used to illustrate how conditional independences are reflected in summary statistics derived from the models and how the graphs help to identify analogies and equivalences between different models. Graphical chain models are shown to provide a unifying concept for many statistical techniques that in the past have proven to be useful in analyses of data. They also provide tools for new types of analysis.

Keywords: ANALYSIS OF VARIANCE; ASSOCIATION STRUCTURE; CONDITIONAL GAUSSIAN DISTRIBUTION; CONDITIONAL GAUSSIAN REGRESSION; COVARIANCE SELECTION; FACTOR ANALYSIS; LATENT CLASS MODEL; LINEAR STRUCTURAL EQUATIONS; LOG-LINEAR MODEL; MULTINOMIAL LOGIT MODEL; MULTIPLE LINEAR REGRESSION; PATH ANALYSIS

1. INTRODUCTION

Much research in the social sciences concerns determinants of the development and changes in human behaviour and reactions. Typically, cross-sectional studies are used to obtain empirical evidence on variables capturing characteristics, behaviour, abilities, attitudes of people or historical and environmental conditions. These variables are *properties of observational units.*

The primary objective of such studies is to improve knowledge about the association structure in a given set of such variables, in a so-called *system* of variables. Associations are relations between pairs of variables, and an *association structure* is a description of relations in a system such that all associations can be deduced.

Subject matter knowledge and theories lead to expectations regarding the type, strength or possibly the sign of some of the associations and to hypotheses about relations which are only indirect. Such expectations and hypotheses picturing actual properties of observational units are called *substantive research hypotheses* or simply research hypotheses. They contrast with hypotheses formulated with the sole purpose of falsifying them by actual observations.

†*Address for correspondence*: Johannes Gutenberg-Universität, Postfach 3980, D-6500 Mainz, Federal Republic of Germany.

 0035–9246/90/52021 $2.00

Direct and *only indirect* relations are key notions in studying association structures. One hopes that an understanding of the structure is enhanced if it can be well described in terms of a few direct relations, i.e. with a few strong associations, which account for the indirect relations in the system. To say that a relation between a pair of variables is indirect means that they have a substantial association given the information on one subset of variables, but not given the information on another subset of variables in the system. This implies, in particular, that when hypotheses about indirect relations are contemplated the system will typically not contain any subsets of variables which are completely unrelated.

Graphs may be used to formulate research hypotheses about indirect relations in an association structure. These graphs consist of points for variables, and of lines or arrows for each pair of variables having a direct relation. Such a hypothesis has two main components which specify

(a) that the direct relations are sufficient to understand all associations in the system and
(b) that the set of direct relations cannot be further reduced without destroying such an understanding.

The second aspect is essential for the distinction between a graph characterizing the research hypothesis about an association structure and the same graph identifying a corresponding statistical model for associations.

An example of a graph with both interpretations is given in Fig. 1. If viewed as a substantive research hypothesis it contains two types of properties of observational units: variables with a nominal scale, called categorical or *qualitative*, and variables for which numerical measurements are obtained, called *quantitative*. There are two types of direct associations, those called *directional associations* for variable pairs where one is regarded as a response variable and the other as an explanatory or influencing variable, and those called *symmetric associations* where no direction of dependence has been specified for the relation. Symmetric associations are used for variables considered to be on an equal footing: they are either all response variables only, like those in subset a of Fig. 1, or they are all just influences, like those in subset c, or they are intermediate variables in the sense of being both responses and influences, like the variables in subset b.

If, instead, the graph is viewed as characterizing a statistical model, then the model could be a graphical chain model as introduced by Lauritzen and Wermuth

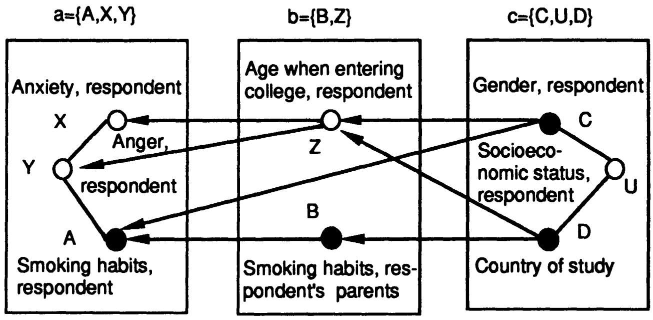


Fig. 1. Example of a research hypothesis about indirect relations among four qualitative variables (A, B, C, D) and four quantitative variables (X, Y, Z, U)

(1984, 1989). These models are defined for sets of both discrete and continuous random variables in terms of specific distributional assumptions and a set of conditional independence restrictions. The graph in Fig. 1 depicts the second aspect, in the sense that each missing direct connection for a variable pair corresponds to a specific conditional independence statement while the set of all independence restrictions represents a *conditional independence structure*. In contrast with the research hypothesis, the statistical model permits a variable pair to have no association, even if it is directly connected in the graph.

Graphical chain models tie up with the early proposal for path analysis by Wright (1921, 1923, 1934). He had suggested the use of graphs with arrows and lines to characterize a correlation structure of interval-scaled variables and to use linear equations, which are in one-to-one correspondence with the graph, to formulate a statistical model. In path analysis an indirect relation means that the simple correlation coefficient of the variable pair can be expressed, and therefore explained, in terms of the remaining simple correlations of those pairs which have direct connections in the graph. Graphical chain models can be viewed as extending this proposal in four directions:

(a) there may be more than one response variable to each set of influences;
(b) in addition to correlated quantitative variables the system can contain qualitative influences *and* qualitative responses;
(c) each indirect relation corresponds to a conditional independence statement;
(d) some of the variables may be latent, which means that information on them is obtained only indirectly with the help of other, directly observable variables.

Only the first and last of these properties are shared by a different extension of path analysis, by linear structural equations (Goldberger, 1964; Jöreskog, 1977). Discussion of the relations between the two approaches and of ambiguities associated with some specifications on linear structural equations have been given by Wermuth (1988). The second and the last property, but not the other two, apply to the extension of linear models to generalized linear models as discussed by McCullagh and Nelder (1983).

The primary objective of this paper is to enhance an understanding of the type of substantive research hypotheses which may be analysed with the help of graphical chain models. It is explained in Section 2 how the same type of graph may be derived either from subject matter considerations about properties of observational units or from conditional independence restrictions for random variables satisfying specific distributional assumptions. In Section 3 several examples of association structures illustrate how different the parametric consequences may be for the same type of conditional independence structure. Problems in evaluating a substantive research hypothesis in term of likelihood ratio tests and Studentized interaction parameters are discussed in Section 4. In Section 5 simple criteria are described to identify equivalent statistical models from graphs. Such knowledge is important, since it is not possible to discriminate between alternative substantive research hypotheses whenever they correspond to equivalent statistical models. Finally, in Section 6 standard statistical models which are either elements or special cases of graphical chain models are listed.

2. THREE DISTINCT MEANINGS OF THE SAME GRAPH

Graphs like Fig. 1 or similar ones can be used to formulate substantive research hypotheses, to study conditional independence structures or to characterize statistical models. In each of these situations the derivation or justification of the graph, its precise meaning and the associated main question differ. This is illustrated in the following.

2.1. *Substantive Research Hypotheses*

2.1.1. *Example of derivation of a research hypothesis*

Subject matter considerations about the variables in Fig. 1 are as follows. It is established knowledge in psychology that there are personality differences between occasional smokers, regular smokers, ex-smokers and people who have never smoked cigarettes. Therefore these classifications are treated as categories of a qualitative variable named smoking habits (A).

Such personality differences show up, for instance, in the emotions anxiety (X) and anger (Y). These emotions are viewed as dispositions, also called traits of a person. Questionnaires have been developed by Spielberger *et al.* (1970, 1983b) to obtain quantitative measurements of these variables.

The two traits are expected to correlate positively. Decisions on the direction of influence between emotions and smoking habits cannot be based on psychological theory: it is just as conceivable that smoking habits are influenced by personality characteristics as it is that personality characteristics are influenced by smoking habits. Consequently, the three variables are regarded as being on an equal footing. They are combined in one set $a = \{A, X, Y\}$. Within this set symmetric associations are of interest.

Smoking habits and emotions constitute the main set of *multiple responses*, with all the remaining variables in the system as potential influences. Therefore directional associations are wanted for all pairs in which one variable is one of the multiple responses (A, X, Y) and the second is one of the remaining variables (B, C, D, Z, U).

This latter set of potential influencing variables may be further subdivided into one containing background variables and another one regarded as potential responses to these background variables. The background variables are gender (C), socioeconomic status of respondent (U) and the country in which the study is conducted (D). They are seen as having been determined in the more distant past compared with the intermediate variables which are smoking habits of the respondent's parents (B) and the age of the respondent when entering college (Z). Of interest are again symmetric associations *within* the sets of responses and influences but directional associations *between* the two sets: $b = \{B, Z\}$ and $c = \{C, U, D\}$.

These arguments lead to a division of the system into three subsets indicated by boxes in the graph. Within boxes lines represent symmetric associations, between boxes arrows point from the potential influences to the responses. The three boxes in Fig. 1 are said to define a *dependence chain* with three sets of *concurrent variables.* Concurrent variables are those which are to be considered simultaneously for a precise description of the type of each partial association. The sets of concurrent variables are obtained by stepwise deletion of response sets from the dependence chain. In Fig. 1, the three sets of concurrent variables are $a \cup b \cup c$, $b \cup c$ and c.

The convention adopted for graphs with dependence chains is that each pairwise relation is a partial association given information on all the remaining concurrent variables. In Fig. 1, for instance, the direct relation between smoking habits (A) and anger (Y) means that there is a partial symmetric association given information on (Z, B, C, U, D) and on anxiety (X). The direct relation between smoking habits of the respondent (A) and of the parents (B) implies a partial directional association given information on the remaining potential influences (C, D, Z, U) and on the emotions (X, Y). The indirect relation between parents' smoking habits (B) and gender of respondent (C) means

(a) that there is no partial directional association given information on age (Z), socioeconomic status of respondent (U) and the country of study (D) and
(b) that its marginal association can be deduced from the direct relations in their set of concurrent variables, i.e. from the associations for pairs (C, U), (U, D), (Z, C), (Z, D) and (B, D).

The indirect relation between gender (C) and country of study (D) implies no partial symmetric association for this variable pair at fixed levels of socioeconomic status (U) and that its marginal association can be deduced from the two direct relations of socioeconomic status to gender and to country of study.

2.1.2. *Examples of indirect relations not expressible with chain graphs*

Neither arbitrary graphs for variables and associations nor arbitrary sets of indirect relations lead to graphs, like Fig. 1, which are called *chain graphs*.

Let us consider a graph with points representing variables and with variable pairs having at most one connection, either a line or an arrow. Such a graph is a chain graph if a dependence chain can be attached to it. This means that the set of all variables can be partitioned into subsets such that we obtain—after rearranging the variables—a graph displaying

(a) the subsets ordered in a horizontal row,
(b) as connections only arrows pointing in one direction between subsets and
(c) only lines within subsets, i.e. a chain of boxes.

Simple examples of incomplete graphs which are not chain graphs are shown in Fig. 2. Even though such association structures may be of subject matter interest it is not clear how to define corresponding appropriate statistical models such that an empirical evaluation becomes possible.

To give examples of hypotheses about indirect relations, which do not lead to chain graphs, assume that linear dependences among four quantitative variables (X, Y, Z, U) are of interest, i.e. correlation coefficients are appropriate measures of association. Assume further that four variable pairs have direct relations: the pairs (X, Y), (Y, Z),

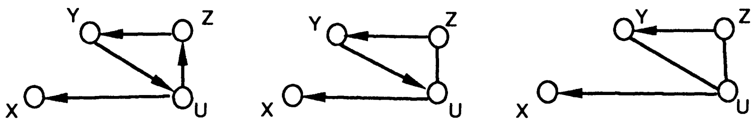


Fig. 2. Examples of association structures which do not correspond to chain graphs since the vertices cannot be arranged in a horizontal row of boxes containing lines within and arrows between boxes

(Z, U) and (U, X). Then the hypothesis $\rho_{xz} = \rho_{yu} = 0$ cannot be formulated with a chain graph. It is an example of 'linear structure in the covariance matrix' (Anderson, 1973). If, instead, the association structure is supported to have $\rho_{xz.u} = \rho_{yu.z} = 0$ then it cannot be described with the help of a chain graph either. It is a hypothesis in a multivariate regression model which may be evaluated by using a program for fitting linear structural equations (Jöreskog, 1977).

2.1.3. *Key questions associated with research hypotheses*

Empirical evidence on research hypotheses about indirect relations between variables is obtained with the help of statistical models. Such statistical models lead to particular measures of associations, the values of which may be derived from known values of the parameters in the model. For an empirical evaluation of the hypotheses the questions of primary concern are as follows.

(a) How is the notion of an indirect relation implemented in the statistical model?
(b) Which implications does a given research hypothesis have for estimates of parameters or of measures of association?
(c) How do the estimates help to discriminate between two alternative hypotheses for a given system, i.e. for the same set of variables?
(d) Which effects on the estimates can be expected when there are changes in the system?
(e) How much will an evaluation of the research hypothesis be affected when the statistical model for a given system is modified?

In this paper only the first three questions are partially answered: for statistical models which are *conditional Gaussian* (CG) *chain models*, i.e. for chain models in which distributions are of the conditional Gaussian type. Effects of measurement error, of transformations on variables or of changes in distributional assumptions have not yet been systematically treated. Effects of deleting variables from the system have been discussed in special situations as 'parametric collapsibility' or 'moderating effects' (Bishop, 1971; Whittemore, 1978; Wermuth, 1987, 1989a, b).

2.2. *Conditional Independence Structures*

The graph in Fig. 1 may be viewed as characterizing a conditional independence structure. In that case it is a mathematical object and may be appropriately described in terms of graph theoretic language: points are *vertices*, connections are *edges*, lines are *undirected edges* and arrows are *directed edges*. Vertices represent *random variables* and edges *associations* between these variables. The graph is *marked* since there are two types of vertices, circles for continuous and dots for discrete random variables. If the graph has an edge between all pairs of vertices, the graph is *complete*. Complete graphs do not imply any conditional independence restrictions.

The set V of all vertices may contain a subset Δ of discrete and a subset Γ of continuous variables, $V = \Delta \cup \Gamma$. To keep the notation simple, the variables are named by capital letters such as $\Delta = \{A, B, C, D\}$ and $\Gamma = \{X, Y, Z, U\}$. The graph is specified in terms of its set of vertices and its set of edges. It may have at most one edge for each pair of distinct vertices.

Such a graph is called a *chain graph* if a dependence chain can be attached to it in the way described in Section 2.1.2. A *dependence chain* is an ordered partitioning of the vertex set V into *chain elements* such that edges within chain elements are undirected and edges between chain elements are directed edges all pointing in the same direction. In Fig. 1 the dependence chain is $\mathscr{C} = (a, b, c)$. It partitions the vertex set $V = a \cup b \cup c$. In graph theoretic language a chain graph is characterized as a graph in which 'no subset of vertices induces a directed cycle' (Frydenberg, 1986). An *induced graph* is formed by a subset of vertices by keeping the given edges within the subset. In each of the three graphs in Fig. 2 a directed cycle is induced by three vertices.

In a chain graph a dependence chain determines

(a) response sets,
(b) sets of concurrent variables and
(c) the meaning of each missing edge.

Definitions of responses and of concurrent variables agree with those given in Section 2.1. A *response set* is a chain element to which directed edges can point, and a *set of concurrent variables* is the union of a chain element with chain elements from which directed edges could be pointing to it. In Fig. 1 the response sets are a and b, and the three sets of concurrent variables are $a \cup b \cup c$, $b \cup c$ and c. A missing edge means that the variable pair is conditionally independent given all its remaining concurrent variables.

In Fig. 1 this gives, for instance, for the following selected pairs

$$
\begin{array}{ll}
(A, X): & A \perp\!\!\!\perp X|(a \cup b \cup c\backslash\{A, X\}), \\
(A, D): & A \perp\!\!\!\perp D|(a \cup b \cup c\backslash\{A, D\}), \\
(B, C): & B \perp\!\!\!\perp C|(b \cup c\backslash\{B, C\}), \\
(C, D): & C \perp\!\!\!\perp D|(c\backslash\{C, D\}),
\end{array}
$$

where we have adapted the notation introduced by Dawid (1979).

Different dependence chains may be *compatible* with the same conditional independence structure in the sense of having the same *underlying chain graph*, i.e. the same vertices and edges between all pairs. An example is the structure displayed in Fig. 3 as a directed chain graph without and with two compatible dependence chains.

To understand the meaning of a research hypothesis, it is crucial to know the dependence chain since it assigns a specific meaning to each missing edge. This is not the case for the corresponding conditional independence structure since it depends only on the underlying chain graph (Frydenberg, 1986, 1989).

The proof of this important fact is based on the following. All compatible dependence chains are derivable from a unique partitioning of the vertex set called the

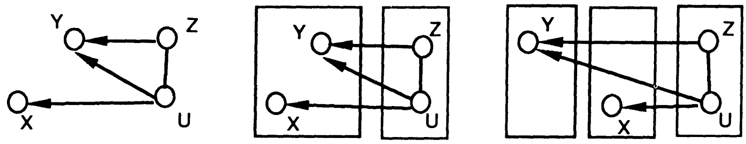


Fig. 3. Chain graph without and with two distinct compatible dependence chains, i.e. without and with specific meanings attached to each missing edge: the missing edge (X, Z) means $X \perp\!\!\!\perp Z|(Y, U)$ in the graph in the centre and $X \perp\!\!\!\perp Z|U$ in the graph on the right-hand side; a compact description of all independences is $X \perp\!\!\!\perp (Y, Z)|U$

minimal chain components of the chain graph. They are obtained from any given dependence chain by subdividing each chain element into *connected components*, i.e. into subsets which have no direct relations in the subgraph induced by their chain element. For instance in Fig. 1 the minimal chain components are the elements of $\{\{A, X, Y\}, \{B\}, \{Z\}, \{C, U, D\}\}$. The three dependence chains compatible with the underlying chain graph in Fig. 1 are $\mathscr{C} = (a, b, c)$, $\mathscr{C}' = (a, \{B\}, \{Z\}, c)$ and $\mathscr{C}'' = (a, \{Z\}, \{B\}, c)$.

To distinguish between a conditional independence structure and the graph which corresponds to a compatible substantive research hypothesis we speak of a *recursive conditional independence graph* or, in short, of a conditional independence graph whenever a chain graph has a specific dependence chain attached to it. The recursive conditional independence graph is specified in terms of its dependence chain, its set of vertices and its set of edges.

Three classes of conditional independence graphs will be important in the following: those with exclusively symmetric associations, those with only single responses and those with multiple responses in the system. To distinguish between the different situations we denote a general recursive conditional independence graph by G and speak of

(a) a multiple-response graph (G^{mr}) if at least one response set contains more than one variable,
(b) a single-response graph (G^{sr}) if all response sets contain one variable and
(c) a symmetric association graph (G^{a}) if there are no response sets.

Examples of the three types of graph are given in Figs 1, 4 and 5 respectively.

The main question for conditional independence structures is: what are the independences implied by the graph, i.e. what are its Markov properties? This has, for example, been addressed by Darroch *et al.* (1980), Kiiveri (1983), Kiiveri *et al.* (1984), Lauritzen and Wermuth (1984, 1989), Pearl (1986), Verma (1988), Geiger and Pearl (1988), Lauritzen *et al.* (1989) and Frydenberg (1989). It is not discussed in this paper.

2.3. *Graphical Conditional Gaussian Chain Models*

A graph such as Fig. 1 corresponds to a *graphical chain model* if the joint distribution is specified in terms of distributions involving the different sets of concurrent variables and a chain graph represents the conditional independence restrictions on the joint distribution.

2.3.1. *Joint distribution in chain model*

As for the definition of a conditional independence graph we assume that a given set V of *random variables* may contain a subset Δ of *discrete* and a subset Γ of

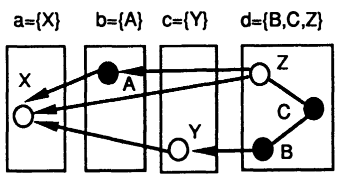


Fig. 4. Example of a graph with only single responses G^{sr}, having dependence chain $\mathscr{C} = (a, b, c, d)$

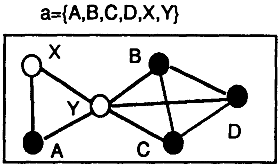


Fig. 5. Example of a symmetric association graph G^a, having dependence chain $\mathscr{C} = (a)$

continuous variables $V = \Delta \cup \Gamma$. They are to model qualitative and quantitative variables respectively. To repeat, in Fig. 1 the two sets of variables are $\Delta = \{A, B, C, D\}$ and $\Gamma = \{X, Y, Z, U\}$, the dependence chain is $\mathscr{C} = (a, b, c)$, the three sets of concurrent variables are $a \cup b \cup c$, $b \cup c$ and c, and the response sets are a and b.

A dependence chain plays two different roles in the statistical model: it determines a way of obtaining the joint density of all variables in the system and it defines a specific conditional independence restriction for each variable pair having a missing edge in the graph.

If the dependence chain contains T elements then a factorization of the joint density f_V of all variables in the system is given as a product of $T - 1$ conditional densities and one marginal density. For example, in Fig. 1 we have

$$f_V = f_{a|bc} f_{b|c} f_c.$$

Here, $f_{a|bc}$ denotes the conditional density of responses in set a given all variables in $b \cup c$, $f_{b|c}$ the conditional density of those in b given c and f_c denotes the marginal density of the variables in c. With a dependence chain having T elements, T distributions have to be specified corresponding to T simultaneous recursive analyses of the sets of concurrent variables.

In principle, different types of distribution may be appropriate for each set of concurrent variables. However, in CG chain models all $T - 1$ conditional densities as well as the marginal density are of the CG type, i.e. of the CG type described in more detail in the next subsection.

We reserve the symbol g for densities of the CG type so that for instance in Fig. 1 the joint distribution of a CG chain model is understood to be

$$f_V = g_{a|bc} g_{b|c} g_c.$$

Given any recursive conditional independence graph and distributional assumptions of this specific type the graph can be viewed as characterizing a graphical CG chain model. Such a model is called saturated if it is given by a complete graph and non-saturated if it is given by an incomplete graph. A *saturated model* is a statistical model defined just by the distributional assumptions, while a *non-saturated model* is obtained from a saturated model by imposing restrictions on some of its parameters.

2.3.2. *Conditional Gaussian distribution*

Here, we define a density of the CG type for a symmetric association model, i.e. a CG density corresponding to an undirected graph or to a dependence chain having one element.

Suppose that the set of all variables ($V = \Gamma \cup \Delta$) contains q continuous variables. A CG distribution for V is defined by a conditional joint Gaussian distribution of the continuous variables given the discrete variables and by positive probabilities for each level combination of the discrete variables. The joint density of all variables can be expressed with the help of *moment characteristics*. These are the probabilities π_l, the means μ_l and the conditional covariance matrices Σ_l, where $l = 1, \ldots, L$ denote the level combinations of the discrete variables. When the conditional covariance matrix does not depend on the level combinations of the discrete variables, i.e. when $\Sigma_l = \Sigma$, the CG distribution is said to be *homogeneous*. The joint density is a product of conditional Gaussian densities $g_{\Gamma|\Delta}$, and of the marginal probability function $g_\Delta = \pi_l$

$$g_V = g_{\Gamma|\Delta} g_\Delta = \left[\left\{ \frac{1}{\sqrt{(2\pi)}} \right\}^q |\Sigma_l|^{-1/2} \exp\{-\tfrac{1}{2}(x - \mu_l)^{\mathrm{T}} \Sigma_l^{-1} (x - \mu_l)\} \right] \pi_l,$$

where x contains the realizations of the continuous variables. Occasionally, $K_l = \Sigma_l^{-1}$ is referred to as a concentration matrix having concentrations as off-diagonal elements and precisions along the diagonal.

Equivalently, the logarithm of the density may be written in terms of canonical characteristics as

$$\log g_V = d_l + h_l^{\mathrm{T}} x - \tfrac{1}{2} x^{\mathrm{T}} K_l x,$$

where the discrete, linear and quadratic canonical characteristics are denoted by d_l, h_l^{T} and K_l respectively.

The relations between the two sets of characteristics (d_l, h_l, K_l) and (π_l, μ_l, Σ_l) are

$$d_l = \log \pi_l - \tfrac{1}{2}\{q \log(2\pi) + \log|\Sigma_l| + \mu_l^{\mathrm{T}} \Sigma_l^{-1} \mu_l\},$$

$$h_l = \Sigma_l^{-1} \mu_l, \qquad K_l = \Sigma_l^{-1}.$$

Some of the properties of CG distributions are summarized from Lauritzen and Wermuth (1989) as follows. Symmetric association models based on these distributions lead to exponential families. The sufficient statistics are all functions of familiar data summaries: of counts, means and covariance matrices. The conditional independence restrictions determine for which level combination of the discrete variables these data summaries are to be computed, or, to put it differently, which statistics form the minimal set of sufficient statistics.

If a conditional density is computed from g_{ab}, i.e. from a joint CG density for variables in $a \cup b$, it stays in the family of CG distributions. However, a marginal distribution obtained from g_{ab} may or may not be a CG distribution. An illustration of this last fact is already possible with a mixed bivariate density for the variables X and A, which we write as g_{AX}.

$$g_{AX} = g_{X|A} g_A = g_{A|X} f_X.$$

The marginal distribution X is a mixture of two univariate normal or Gaussian distributions. Such a mixture is typically not a CG distribution. This illustrates in which sense the multivariate family of CG distributions is more complex than the family of Gaussian distributions. The latter is closed under conditioning as well as marginalizing (see, for example, Anderson (1958)).

2.3.3. *Interactions of conditional Gaussian distributions*

An important feature of a CG distribution for evaluating research hypotheses about indirect relations is its parameterization with interactions. It has been proved

(Lauritzen and Wermuth (1989), proposition 3.1) that in CG distributions a variable pair is conditionally independent given the remaining variables if and only if all interaction terms containing this variable pair are zero. Thus, formulating a hypothesis with missing edges in a symmetric association graph G^a is equivalent to specifying a particular set of interactions to be zero. Whenever such a research hypothesis is correct, this will be reflected in estimates of the interaction parameters: estimates of all interactions corresponding to missing edges will be close to zero. The notion of interaction in a CG distribution is related to, but distinct from, the notion of interaction used in the context of analysis of variance models (Cox, 1984).

The parameterization of a CG distribution with interactions is illustrated here for $V = \{A, B, X, Y\}$ and $\mathscr{C} = (V)$, more precisely for $\Delta = \{A, B\}$ with A, B having categories $i = 1, \ldots, I$ and $j = 1, \ldots, J$ respectively and $\Gamma = \{X, Y\}$ having realizations (x, y). In this case the CG density can be written in terms of canonical characteristics as

$$\log g(i, j, x, y) = d_{ij} + h^x_{ij} x + h^y_{ij} y - \tfrac{1}{2} k^x_{ij} x^2 - \tfrac{1}{2} k^y_{ij} y^2 - k^{xy}_{ij} xy \tag{1}$$

and in terms of interactions as

$$\begin{aligned}\log g(i, j, x, y) = {} & \lambda + (\lambda^A_i + \lambda^B_j + \lambda^{AB}_{ij}) + (\eta^X + \eta^{AX}_i + \eta^{BX}_j + \eta^{ABX}_{ij})x \\ & + (\eta^Y + \eta^{AY}_i + \eta^{BY}_j + \eta^{ABY}_{ij})y \\ & - \tfrac{1}{2}(\psi^X + \psi^{AX}_i + \psi^{BX}_j + \psi^{ABX}_{ij})x^2 \\ & - \tfrac{1}{2}(\psi^Y + \psi^{AY}_i + \psi^{BY}_j + \psi^{ABY}_{ij})y^2 \\ & - (\psi^{XY} + \psi^{AXY}_i + \psi^{BXY}_j + \psi^{ABXY}_{ij})xy.\end{aligned}$$

Constraints, such as symmetric constraints, have to be adopted to assure uniqueness of the interaction terms. These give for example $0 = \Sigma_i \lambda^A_i = \Sigma_i \eta^{ABX}_{ij} = \Sigma_i \psi^{ABY}_{ij}$.

To relate this interaction representation to more familiar ones, we may further specialize it. For only two discrete variables, we obtain the log-linear representation (Birch, 1963) of probabilities in a two-way contingency table. For only two continuous variables, we obtain the exponential family representation of a bivariate normal distribution (Dempster, 1972). *Mixed interactions* involving both discrete and continuous variables, such as η^{AY}_i, ψ^{AY}_i, do not seem to have appeared explicitly in other statistical models. In a homogeneous CG distribution the mixed ψ interactions are all equal to zero.

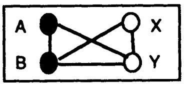
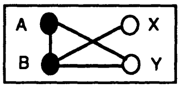
Because of their importance for the interpretation of models we emphasize the following results from Lauritzen and Wermuth (1989).

Fact 1. For each pair from a set of variables $\Delta \cup \Gamma$, in a symmetric association graph G^a, the following statements are equivalent:

(a) the variable pair is conditionally independent given all the remaining variables;
(b) all interactions involving the variable pair are equal to zero in the interaction parameterization of the CG distribution;
(c) the edge of the variable pair is missing in the symmetric association graph.

The two graphs in Table 1 provide an illustration.

TABLE 1
Examples to illustrate fact 1

Case	*Symmetric association graph*	*Equivalent independences*	*Equivalent sets of zero interactions*
1		$A \perp\!\!\!\perp X \mid (B, Y)$	$S_1 = \{\eta_i^{AX}, \eta_{ij}^{ABX}\} \cup \{\psi_i^{AX}, \psi_{ij}^{ABX}\} \cup \{\psi_i^{AXY}, \psi_{ij}^{ABXY}\}$
2		$A \perp\!\!\!\perp X \mid (B, Y)$ and $X \perp\!\!\!\perp Y \mid (A, B)$	$S_2 = S_1 \cup \{\psi^{XY}, \psi_j^{BXY}\}$

2.3.4. *Conditional Gaussian regression*

Conditional densities in a CG chain model are all *CG regressions*. A CG regression is a conditional distribution which looks just as if the joint distribution of the variables in its response and influence set would follow a joint CG distribution (Lauritzen and Wermuth (1989), proposition 2.4). The density of this conditional distribution is of the CG type since it may be expressed in terms of discrete, linear and quadratic canonical characteristics, just as for a CG distribution. The difference is that the parameters of a CG regression may depend on the values of the influencing variables.

We speak of *multiple-response* or of *block* regressions and of *univariate* regressions, depending on whether the response set contains several variables or only a single variable. A univariate CG regression is a *linear regression* if the response is continuous and a *logistic regression* with linear and quadratic dependence on the influencing variables if the response is discrete. A homogeneous CG regression is derived from a homogeneous CG distribution. In this case the linear regressions are parallel and the logistic regressions depend linearly on the influencing variables.

For example from equation (1) the linear regression of Y given A, B and X is given by

$$E(Y|x) = \alpha_{ij} + \beta_{ij}x, \quad \operatorname{var}(Y|x) = 1/k_{ij}^y, \quad \alpha_{ij} = h_{ij}^y/k_{ij}^y, \quad \beta_{ij} = -k_{ij}^{xy}/k_{ij}^y.$$

The logistic regression of A given B, X and Y for a binary response is

$$\log(\pi_{1|jxy}/\pi_{2|jxy}) = d_{1j} - d_{2j} + (h_{1j}^x - h_{2j}^x)x + (h_{1j}^y - h_{2j}^y)y - \tfrac{1}{2}(k_{1j}^x - k_{2j}^x)x^2 - \tfrac{1}{2}(k_{1j}^y - k_{2j}^y)y^2 - (h_{1j}^{xy} - h_{2j}^{xy})xy.$$

For the CG regression having density $g_{a|b}$ we can speak of an *attached joint CG distribution* for the variables in $a \cup b$. The reason is that each CG regression can be viewed as being derived from a joint CG distribution. The attached CG distribution is not unique unless the parameters in the marginal distribution of the influences are specified. For instance, in Fig. 1 the CG regression with density $g_{a|bc}$ is a conditional distribution of $a = \{A, X, Y\}$ given all the remaining variables just as if $\{X, Y, Z, U\}$ had a Gaussian distribution given $\{A, B, C, D\}$. The CG regression with density $g_{b|c}$ is a conditional distribution of $b = \{B, Z\}$ given $c = \{C, U, D\}$ just as if $\{Z, U\}$ had a Gaussian distribution given $\{B, C, D\}$.

2.3.5. *Joint distribution in a conditional Gaussian chain model*

In a CG chain model it is assumed that

(a) the $T - 1$ conditional densities in the joint distribution are all CG regressions and
(b) the marginal distribution of the set of variables containing no responses is a CG distribution.

A CG chain model with exclusively single responses is said to be a *univariate recursive regression model* given by the graph G^{sr}, while a model with multiple responses is called a *block recursive regression model* given by G^{mr}.

Although all distributions of a CG chain model are based on CG distributions the joint distribution of it need not be a CG distribution itself. The simplest example is a mixed bivariate density for the variables X and A defined for the dependence chain $\mathscr{C} = (\{A\}, \{X\})$. The joint density is

$$f_{AX} = g_{A|X} g_X,$$

which is typically not of the CG type.

Two graphical chain models are *equivalent* if they have the same joint distribution and the same conditional independence structure. As was shown by Frydenberg (1986) not only the conditional independence structure but also the joint distribution of a CG chain model is completely determined by the underlying chain graph.

In general, complete graphs with different dependence chains define different saturated models for the same collection of variables since they may have different distributional implications. Two such *saturated models* or complete systems are equivalent if their joint distributions coincide. Fig. 6 shows chain graphs for saturated models which are not equivalent.

In the special case where all variables in a given system are of the same type, i.e. either discrete or continuous, all possible saturated models are equivalent. The reasons are that, as mentioned before, the family of Gaussian distributions is closed under conditioning and marginalizing and that probability functions with positive cell probabilities stay positive after conditioning or marginalizing.

As we shall see next, each graphical chain model can be interpreted with the help of T symmetric association graphs: with those implicitly attached to each given dependence chain. We recall from the previous section that each CG regression has an implicitly attached CG distribution. For a CG regression in a graphical chain model we can, similarly, speak of an *attached symmetric association graph* of the

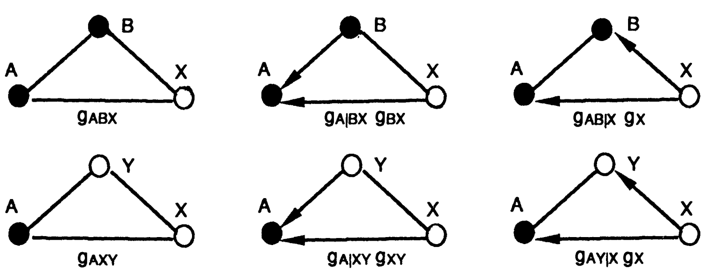


Fig. 6. Two sets of three complete chain graphs which have the same vertices but a different joint distribution

G^a attached to the concurrent variables of set a={A,X,Y} of Figure 1

G^a attached to the concurrent variables of set b={B,Z} of Figure 1

G^a attached to the concurrent variables of set c={C,U,D} of Figure 1

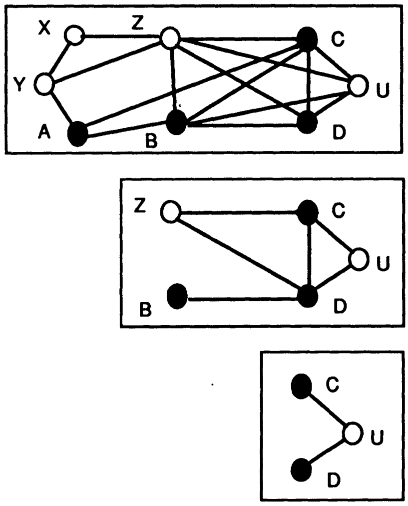


Fig. 7. Symmetric association graphs attached to the three sets of concurrent variables in Fig. 1, i.e. to $a \cup b \cup c$, $b \cup c$ and c

concurrent variables. Such an attached graph is obtained from the subgraph induced by a given set of concurrent variables by changing each arrow to a line and by adding a line between all pairs of influencing variables that were not yet connected. Thus, there can be edges missing for pairs involving variables in the response set but not within the set of influencing variables. Fig. 7 displays the three symmetric association graphs attached to the graph in Fig. 1. These considerations permit the following results which are analogous to fact 1.

Fact 2. For each pair from a set of variables $\Delta \cup \Gamma$ in a general recursive conditional independence graph G, the following statements are equivalent:

(a) the variable pair is conditionally independent given all its remaining concurrent variables;
(b) the variable pair is conditionally independent in any joint CG distribution attached to its concurrent variables;
(c) all interaction terms involving the variable pair are equal to zero in the interaction parameterization of any attached CG distribution;
(d) the edge of the variable pair is missing in the attached symmetric association graph G^a of its concurrent variables;
(e) the edge of the variable pair is missing in the chain graph G.

The implications from (a) to (e)—in that order—and from (e) to (a) follow from the definitions, and for (b) to (e) from fact 1.

2.3.6. *Likelihood estimation and tests in conditional Gaussian chain models*

Key issues associated with CG chain models are the investigation of distributional assumptions and the fitting of models. These will not be discussed at any length in the present paper.

However, it should be noted that each of the CG regressions and the marginal CG distribution involved in the model specification can be investigated and fitted separately (Lauritzen and Wermuth, 1989). In many situations the tasks will reduce to familiar problems as pointed out in Section 6.

In some situations the CG regression can be fitted by deriving estimates from the fitted values in the attached joint CG distribution (Lauritzen, 1989). This follows from results by Frydenberg and Lauritzen (1989), based on the notion of a decomposition of a marked graph (Leimer, 1989). Techniques from logistic regression models (Cox, 1970; McCullagh and Nelder, 1983) may be used whenever the CG regression cannot be fitted directly with the help of the attached joint CG distribution. Many models with multiple response may be fitted by properly exploiting equivalences of models such as those reported in Section 5.

A suitable algorithm for maximum likelihood estimation in a general graphical CG chain model is not available at present. Algorithms developed in Frydenberg and Edwards (1989) and implemented in a program by Edwards (1987) may be used if it is appropriate to fit the attached joint CG distribution.

A discussion of factorizations of likelihood ratio tests in symmetric association models has been given by Goodman (1971), Haberman (1974), Andersen (1974), Sundberg (1975), Wermuth (1976a, b), Porteous (1985a, b) and Frydenberg and Lauritzen (1989).

3. TWO DISTINCT TYPES OF VARIABLES IN THE SAME CONDITIONAL INDEPENDENCE STRUCTURE

In this section we look more closely at one particular conditional independence structure. It corresponds to symmetric association graphs which differ in the number or position of discrete and continuous variables. The purpose is twofold: to illustrate that a given set of conditional independence statements may correspond to quite different graphical chain models and that not all parameterizations of a given model are equally well suited to disclose the independences of the association structure.

Our selected structure has four vertices with three lines, all touching one of the vertices. Its interpretation is that three variables are mutually conditionally independent given the fourth variable. The fourth variable is called the *conditioning variable.* The four graphs in Fig. 8 differ, however, in the way that the vertices are marked and therefore with respect to the type of variables involved.

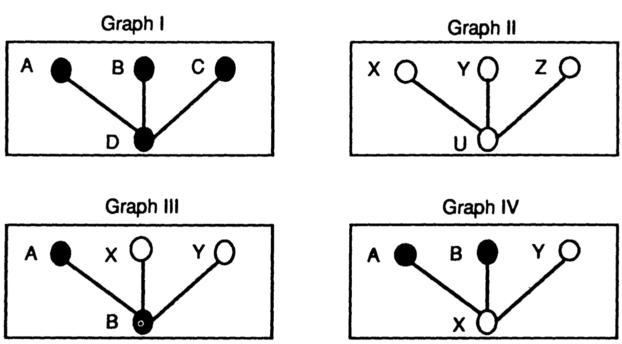


Fig. 8. Examples of distinct symmetric association graphs each with four variables having the same type of independence structure: three variables are mutually conditionally independent given the fourth variable

TABLE 2

Numerical example of an association structure for graph I in Fig. 8

Levels of ABCD (ijkl)	*Moment characteristics ($2000\pi_{ijkl}$)*	*Conditional risks with level 2 as basis for* $A = 1$	$B = 1$	$C = 1$	$D = 1$	*Interactions at level 1 for each variable* *Type*	*Value*	*Conditional odds ratios* *Pair*	*Value*
1111	216	0.3	0.9	0.8	0.86	—	−3.66		
2111	504		0.9	0.8	0.99	A	0.34		
1211	24	0.3		0.8	0.14	B	0.20		
2211	56			0.8	0.78	AB	0	(A, B)	1.00
1121	54	0.3	0.9		0.27	C	0.00		
2121	126		0.9		0.89	AC	0	(A, C)	1.00
1221	6	0.3			0.01	BC	0	(B, C)	1.00
2221	14				0.18	ABC	0		
1112	36	0.9	0.2	0.2		D	0.07		
2112	4		0.2	0.2		AD	−0.76	(A, D)	0.05
1212	144	0.9		0.2		BD	0.90	(B, D)	36.00
2212	16			0.2		ABD	0		
1122	144	0.9	0.2			CD	0.69	(C, D)	16.00
2122	16		0.2			ACD	0		
1222	576	0.9				BCD	0		
2222	64					$ABCD$	0		

Since the graphs are to represent graphical CG chain models it is implicitly understood that the distributions are an arbitrary probability function ($\pi_{ijkl} > 0$) for graph I, a four-dimensional Gaussian distribution for graph II, and mixed CG distributions for graphs III and IV having two discrete and two continuous variables each. The conditioning variable is discrete in graph III but continuous in graph IV.

3.1. *Discrete Variables Only*

Table 2 shows an association structure of discrete variables with conditional independences given by graph I in Fig. 8.

Neither the moment characteristics, which are mean counts, nor the canonical characteristics, which are logarithms of probabilities, are helpful in recognizing the indirect and direct relations in the structure. Instead, these essential features of the association structure are reflected in measures of associations such as interactions involving more than one variable, in conditional relative risks or in conditional odds ratios. A variable pair is conditionally independent given all the remaining variables if all interactions involving this pair are zero, if the conditional odds ratios of this pair are all equal to unity or if the conditional relative risks of this pair are all equal to unity.

In Table 2 all interactions involving pairs (A, B), (A, C) and (B, C) are zero. In a structure with four variables this implies that all higher order interactions are zero as well. The two-factor interactions corresponding to edges in the graph are all substantial. The conditional odds ratios are constant at fixed levels of the remaining variables since all interactions of third and higher order are zero. For instance, for levels (1, 1) and (2, 1) of (A, B) the odds ratios of pair (C, D) are

$$\frac{\pi_{1111}\pi_{1122}}{\pi_{1121}\pi_{1112}} = \frac{216 \times 144}{54 \times 36} = 16$$

TABLE 3
Numerical example of an association structure for graph II in Fig. 8

Variable	*Moment characteristics: Σ Variables*				*Variable*	*Canonical characteristics: Σ^{-1} Variables*			
	X	Y	Z	U		X	Y	Z	U
X	0.50	0.12	0.16	0.20	X	2.381	0	0	−0.952
Y		0.50	0.24	0.30	Y		3.125	0	−1.875
Z			0.50	0.40	Z			5.556	−4.444
U				0.50	U				7.062

and

$$\frac{\pi_{2111}\pi_{2122}}{\pi_{2121}\pi_{2112}} = \frac{504 \times 16}{126 \times 4} = 16$$

respectively, indicating a strong partial association. Relative risks are ratios of risks, which, in turn, are conditional probabilities. For instance, the conditional risk for $A = 1$ given levels (1, 1, 1) of (B, C, D) and the relative risk for $A = 1$ comparing levels 1 and 2 of B given levels (1, 1) of (C, D) are

$$\pi_{1|111}^{ABCD} = \frac{\pi_{1111}}{\pi_{1111} + \pi_{2111}} = \frac{216}{216 + 504} = 0.3$$

and

$$\pi_{1|111}^{ABCD}/\pi_{1|211}^{ABCD} = 0.3/0.3 = 1.$$

The meaning of values different from unity is more easily interpretable for relative risks than for odds ratios.

3.2. *Continuous Variables Only*

Table 3 shows an association structure of continuous variables with conditional independences given by graph II in Fig. 8. As had been the case for only discrete variables, the moment characteristics are not well suited to recognize the independences of the association structure. Instead the interactions in the CG distribution are appropriate. The moment characteristics for continuous variables are means, variances and covariances. The two-factor interactions are concentrations; their standardized counterparts are partial correlation coefficients. Gaussian distributions have, by definition, only two-factor interactions.

In Table 3, for instance, the partial correlation coefficient

$$\rho_{zu.yx} = \frac{-\sigma^{zu}}{\sqrt{(\sigma^{zz}\sigma^{uu})}} = \frac{4.444}{\sqrt{(5.556 \times 7.062)}} = 0.71$$

indicates a strong partial association for pair (Z, U), while $\rho_{xy.zu} = 0$ informs about the conditional independence of (X, Y) given the remaining variables.

3.3. *Continuous and Discrete Variables*

Tables 4 and 5 show association structures of mixed variables. Displayed in the tables are moment and canonical characteristics. As we shall see, with a discrete

TABLE 4

Numerical example of an association structure for graph III in Fig. 8

		Moment characteristics Variables and levels					*Canonical characteristics Variables and levels*			
n_{ij}		*B = 1*		*B = 2*		d_{ij}	*B = 1*		*B = 2*	
$A = 1$		100		40		$A = 1$	−16.43		−19.51	
$A = 2$		40		80		$A = 2$	−17.35		−18.81	
μ_{ij}		*B = 1*		*B = 2*		h_{ij}	*B = 1*		*B = 2*	
		X	*Y*	*X*	*Y*		*X*	*Y*	*X*	*Y*
$A = 1$		6	10	10	5	$A = 1$	1.20	1.67	2.50	0.63
$A = 2$		6	10	10	5	$A = 2$	1.20	1.67	2.50	0.63
Σ_{ij}		*B = 1*		*B = 2*		Σ_{ij}^{-1}	*B = 1*		*B = 2*	
		X	*Y*	*X*	*Y*		*X*	*Y*	*X*	*Y*
$A = 1$	*X*	5	0	4	0	$A = 1$	0.20	0	0.25	0
	Y		6		8			0.17		0.13
$A = 2$	*X*	5	0	4	0	$A = 2$	0.20	0	0.25	0
	Y		6		8			0.17		0.13

TABLE 5

Numerical example of an association structure for graph IV in Fig. 8

		Moment characteristics Variables and levels					*Canonical characteristics Variables and levels*			
n_{ij}		*B = 1*		*B = 2*		d_{ij}	*B = 1*		*B = 2*	
$A = 1$		520		5		$A = 1$	−146.71		−84.86	
$A = 2$		195		280		$A = 2$	−94.53		−32.67	
μ_{ij}		*B = 1*		*B = 2*		h_{ij}	*B = 1*		*B = 2*	
		X	*Y*	*X*	*Y*		*X*	*Y*	*X*	*Y*
$A = 1$		31.08	33.60	20.77	20.72	$A = 1$	12.47	−3.00	10.39	−3.00
$A = 2$		34.24	37.55	15.15	13.69	$A = 2$	8.53	−3.00	6.46	−3.00
Σ_{ij}		*B = 1*		*B = 2*		Σ_{ij}^{-1}	*B = 1*		*B = 2*	
		X	*Y*	*X*	*Y*		*X*	*Y*	*X*	*Y*
$A = 1$	*X*	3.57	4.46	3.13	3.91	$A = 1$	1.17	−0.71	1.21	−0.71
	Y		7.32		6.64			0.57		0.57
$A = 2$	*X*	7.16	8.94	5.58	6.98	$A = 2$	1.03	−0.71	1.07	−0.71
	Y		12.93		10.47			0.57		0.57

conditioning variable, as in graph III, the independences of the association structure can be read off the values of both sets of characteristics. With a continuous conditioning variable, as in graph IV, the moment characteristics disclose none of the independences, the canonical characteristics reveal parts and only the values of interactions disclose all independences.

For an understanding of the parametric implications it is helpful to decompose independences. The mutual conditional independence $(A \perp\!\!\!\perp X \perp\!\!\!\perp Y)|B$ given by

TABLE 6
Interactions for Tables 4 and 5, displayed at levels 1 only

Type of interaction	*Structure with $(A \perp\!\!\!\perp X \perp\!\!\!\perp Y)\|B$ Involved discrete variables* None	*A*	*B*	*AB*	*Structure with $(A \perp\!\!\!\perp B \perp\!\!\!\perp Y)\|X$ Involved discrete variables* None	*A*	*B*	*AB*
λ	−18.02	0.06	1.14	0.40	−80.69	−26.09	−30.93	0
η^X	1.85	0	−0.65	0	9.46	1.97	1.04	0
η^Y	1.15	0	0.52	0	−3.00	0	0	0
ψ^X	0.23	0	−0.03	0	1.12	0.07	−0.02	0
ψ^Y	0.15	0	0.02	0	0.57	0	0	0
ψ^{XY}	0	0	0	0	−0.71	0	0	0

graph III may be expressed in terms of $X \perp\!\!\!\perp Y|(A, B)$, $X \perp\!\!\!\perp A|B$ and $Y \perp\!\!\!\perp A|B$. The first implies zero covariances and zero concentrations at all level combinations of A and B, while $Y \perp\!\!\!\perp A|B$ and $X \perp\!\!\!\perp A|B$ imply the same means *and* variances for both Y and X at fixed levels of B but changing levels of A.

Similarly, the mutual conditional independence $(A \perp\!\!\!\perp B \perp\!\!\!\perp Y)|X$ given by graph IV may be split up as $Y \perp\!\!\!\perp A|(B, X)$, $Y \perp\!\!\!\perp B|X$ and $A \perp\!\!\!\perp B|X$, or, more compactly, as $Y \perp\!\!\!\perp (A, B)|X$ and $A \perp\!\!\!\perp B|X$. The first part implies that regression equations of Y on X are identical if computed at fixed level combination of A, B. In Table 5 the intercepts and regression coefficients are for all i, j

$$\alpha_{yx}(i, j) = \frac{h^x_{ij}}{k^y_{ij}} = \frac{3.00}{0.571} = 5.25, \qquad \beta_{yx}(i, j) = \frac{-k^{xy}_{ij}}{k^y_{ij}} = \frac{0.714}{0.571} = 1.25.$$

The independence $A \perp\!\!\!\perp B|X$ describes a fairly complicated relationship, the simplest type of a *non-decomposable* mixed association model. This means that the model cannot be completely decomposed into simple components for which we know how to obtain direct, non-iterative maximum likelihood estimates. This independence cannot be directly identified from values of canonical characteristics.

In both cases interactions reflect the independences in appropriate zero patterns. Interactions corresponding to the canonical characteristics in Tables 4 and 5 are shown in Table 6.

4. SPECIAL ASPECTS OF EVALUATING RESEARCH HYPOTHESES ABOUT INDIRECT RELATIONS

Complete graphs do not imply hypotheses about indirect relations in the association structure. In that case the primary purpose of an empirical investigation is to estimate the associations. A research hypothesis about indirect relations is evaluated by comparing at least two types of estimates, for instance by comparing associations estimated under the non-saturated chain model and estimated under the *corresponding saturated model*. The complete graph of the latter is obtained from the incomplete graph of the former by adding lines within boxes and arrows between boxes. Some aspects of such an evaluation are described in this section.

The hypothesis specified with the graph in Fig. 9 concerns the joint dependence of qualitative and quantitative variables on an influence. Even though such a hypothesis

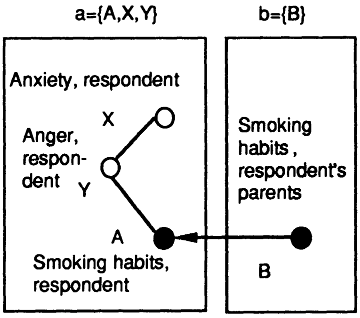


Fig. 9. Research hypothesis about indirect relations among variables with data summaries displayed in Table 7

is familiar in social science theory, no statistical models had been developed, so far, to permit proper empirical investigations. These are partially feasible with graphical chain models, but the models do not directly disclose which measures of conditional and marginal associations permit simple interpretations. Further research is needed here.

The hypothesis in Fig. 9 is regarded as a subhypothesis of the hypothesis in Fig. 1 since

(a) it concerns a new system which is a subset of variables in the larger system,
(b) the graph of the new system coincides with the subgraph induced by the new system in the larger graph and
(c) the remaining variables in the larger system, which have direct relations to a variable in the new system, either have fixed values or are assumed to have no substantial variation for the observational units of the new system.

We report data summaries obtained by personal communication from C. Spielberger. The data are for 384 female students in Florida having no older siblings who smoke. This subgroup of respondents, which is homogeneous with respect to role models that older siblings provide, was chosen to avoid possible confounding effects. The reasons are as follows. It is known from previous studies and analyses (Spielberger *et al.*, 1983a; Wermuth, 1987) that the risk of smoking is increased if a student has older siblings who smoke. Consequently, it is expected that the variable 'role model of older siblings' may moderate the association structure in an unknown way, i.e. neglecting this variable in an analysis could produce confounding effects.

The two qualitative variables are student's smoking status (A) having categories smoker, quit smoking, never smoked and parents' smoking habits (B) with categories neither parent smoked, one parent smoked, both parents smoked. The two quantitative variables are trait anxiety (X) and trait anger (Y).

The sufficient statistics of the saturated symmetric association model corresponding to Fig. 9 are displayed in Table 7. The reported tests are asymptotic likelihood ratio tests (Wilks, 1938). The global test statistic for all independences formulated with the research hypothesis in Fig. 9 indicates a reasonable fit: the value of 40.46 for a chi-squared statistic on 36 degrees of freedom corresponds to a p value of 0.28. However, a more detailed analysis reveals that this judgment is not justifiable, that the

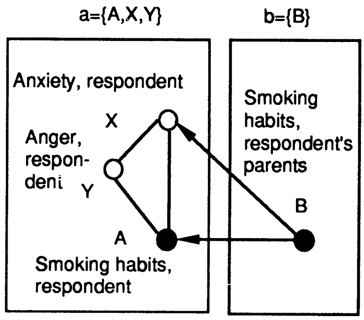


Fig. 10. Modified hypothesis as a result of an analysis

TABLE 7
Sufficient statistics of the saturated symmetric association model in Fig. 9

Sufficient statistics	*Student's smoking status*	*Variables and levels* — *Parents' smoking habits* — *Neither smoked*		*One smoked*		*Both smoked*	
Counts	Smoker	11		20		60	
	Quit smoking	12		21		43	
	Never smoked	47		91		79	
		Anxiety	*Anger*	*Anxiety*	*Anger*	*Anxiety*	*Anger*
Means	Smoker	19.45	25.27	23.05	24.60	22.00	25.80
	Quit smoking	19.17	23.58	22.05	25.33	22.26	25.74
	Never smoked	17.81	19.53	19.56	20.69	19.86	21.48
Covariance matrices	Smoker	46.61	20.94	32.40	10.95	29.10	21.35
			48.58		28.53		46.55
	Quit smoking	10.16	15.18	34.06	28.88	35.99	21.43
			32.89		61.72		39.26
	Never smoked	19.82	9.85	28.52	13.53	27.32	11.11
			26.46		30.35		31.30

hypothesis has to be modified such as shown in Fig. 10. This just illustrates a frequently encountered phenomenon with high degrees of freedom tests.

One way to detect a hidden poor fit of a hypothesis about several indirect relations is to split the global test statistic into sequences of tests for pairwise independences and to look at *studentized interactions* which are estimates of interactions divided by their asymptotic standard deviations. The use of studentized interactions had been suggested in the context of contingency table analyses by Goodman (1970) and was further discussed by Haberman (1978), Section 4. Studentized interactions in CG chain models can be computed after exploiting results by Dempster (1973) for the multinomial logit model. Details will be discussed elsewhere. For our data in Table 7 the largest studentized interactions are

$$\hat{\psi}^{XY} = -4.3, \quad \hat{\psi}^{ABX}_{13} = 2.7, \quad \hat{\psi}^{ABX}_{11} = -2.4, \quad \hat{\eta}^{ABX}_{11} = -2.4.$$

They indicate a poor fit of the observations to the hypotheses of only indirect relations between pairs (X, Y), (A, B), (A, X) and (B, X).

5. EQUIVALENCE OF MODELS

As mentioned previously, distinct research hypotheses defined in terms of two different conditional independence graphs may correspond to equivalent statistical models. Here we state some results on how to read off such equivalences from the graphs

(a) to deal with uncertainty in specifying the dependence chain and
(b) to derive an alternative interpretation of the association structure.

5.1. *Reading Equivalences from Graphs*

Two graphical chain models for the same set of variables are equivalent if their distributional specifications as well as their conditional independences coincide. Consequences of such an equivalence are that the parameters of two CG chain models as well as their maximum likelihood estimates are related by one-to-one transformations, and that corresponding research hypotheses cannot be distinguished by an analysis of data.

Conditions for the equivalence of CG chain models have been studied by several researchers. Situations may be distinguished as comparisons of a single-response graph with a symmetric association graph (Wermuth, 1980; Lauritzen, 1982; Asmussen and Edwards, 1983; Wermuth and Lauritzen, 1983; Kiiveri, 1983; Porteous, 1985b), as comparisons of a single- with a single- or with a multiple-response graph (Lauritzen and Wermuth, 1989; Frydenberg, 1989) and as comparisons of two multiple-response graphs (Frydenberg, 1990). The three situations described here correspond to propositions 8.1 and 8.2 of Lauritzen and Wermuth (1989).

Two CG chain models given by the chain graphs G_1 and G_2 cannot be equivalent if one of the following conditions holds.

Condition 1.

(a) The variables in the graphs do not coincide.
(b) The two sets of missing edges in G_1 and G_2 are not identical.
(c) The distributional implications do not agree since in one but not in both graphs there is a continuous influence having either a discrete response or a continuous response which is connected to a discrete variable within the same response set.

Condition 1(c) does not depend on the dependence chain, since the statement 'variables connected within the same response set' just means that their edges are lines. The condition implies that the joint distribution of some concurrent variables differ in the two models. Fig. 6 in Section 2, and Figs 11 and 12 illustrate condition 1(c).

We are now ready to turn to comparisons of special graphical CG chain models.

Situation 1: Comparing a symmetric association model given by the graph G^a and a univariate recursive regression model given by the graph G^{sr}. Suppose that none of conditions 1(a)–1(c) applies to the two graphs. Then, the models are equivalent if and only if no response in G^{sr} depends directly on two influences without direct relation, i.e. no induced subgraph in G^{sr} forms a three-vertex chain with arrows meeting head to head (i.e. can be arranged like ($x \rightarrow x \leftarrow x$) where x denotes a variable of either type). Thus, the three models in Fig. 13 are equivalent.

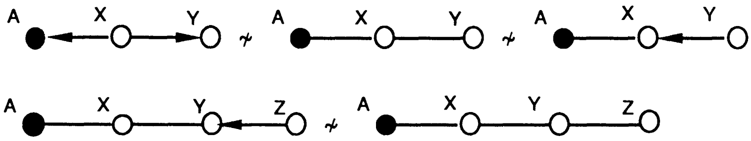


Fig. 11. Examples of sets of incomplete chain graphs in which the graphs imply the same independences but different joint distributions

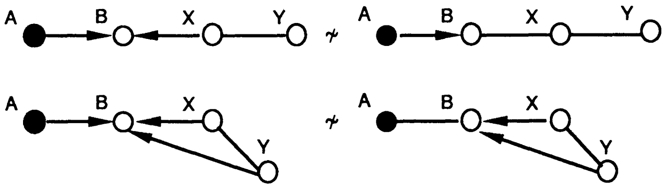


Fig. 12. Examples of pairs of incomplete chain graphs in which the two graphs imply different independences and different joint distributions as well

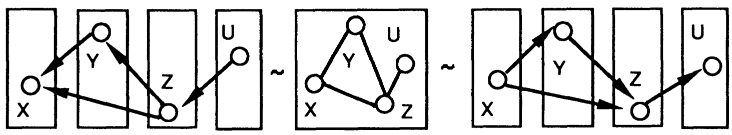


Fig. 13. Examples of conditional independence graphs which imply the equivalence of two distinct univariate recursive regression models to a symmetric association model

Situation 2: Comparing a symmetric association model given by the graph G^{a} and a block recursive regression model given by the graph G^{mr}. Suppose that none of conditions 1(a)–1(c) applies to the two graphs. Then, the models are equivalent if and only if no single response nor any two responses connected in the same response set depend directly on two influences without direct relation, i.e. no induced subgraph in G^{mr} forms a directed chain starting at both sides with arrow ends (i.e. can be arranged like ($x \rightarrow x \leftarrow x$) or ($x \rightarrow x - x \leftarrow x$) or ($x \rightarrow x \ldots x \leftarrow x$)). Thus, the models shown in Fig. 14 are equivalent while those in Fig. 15 are not.

We report a further result only for *comparably arranged graphs* meaning that no direction of influence, i.e. no arrow, in one graph appears reversed in the other.

Situation 3: Comparing a block recursive regression model given by the graph G^{mr} and a univariate recursive regression model given by the comparably arranged graph G^{sr}. Suppose that none of conditions 1(a)–1(c) applies to the two graphs. Suppose further that any discrete variable with a continuous influence in G^{sr} has no other influences in G^{sr} that are variables in its own response set in G^{mr}. Then, the two models are equivalent if any two not directly related influences in G^{sr} are also influencing variables in G^{mr}, i.e. both arrows of an induced three-vertex chain in G^{sr} are arrows in G^{mr} as well.

We believe that this condition is also necessary. The graphs in Fig. 16 provide an illustration.

Further equivalences are possible for two block recursive regression models given by comparably arranged graphs G_1^{mr} and G_2^{mr} such as in Fig. 17 or given by two general

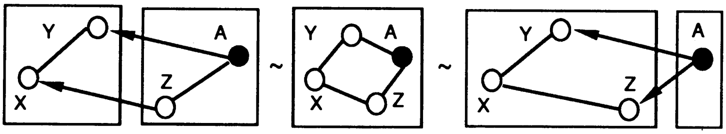


Fig. 14. Examples of conditional independence graphs with equivalence of block recursive regression models to a symmetric association model

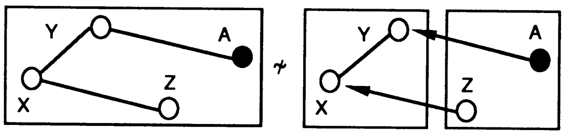


Fig. 15. Examples of conditional independence graphs where the block recursive regression model is not equivalent to the corresponding symmetric association model

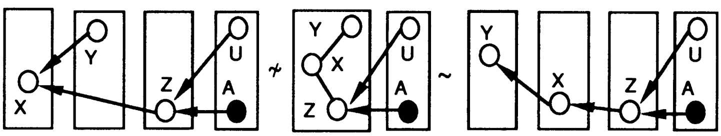


Fig. 16. Examples of conditional independence graphs where the block recursive regression model is equivalent to one but not to the other corresponding univariate recursive regression model

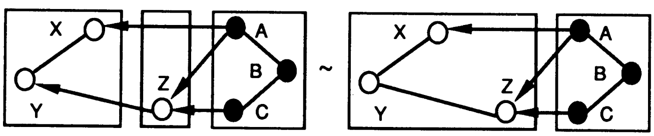


Fig. 17. Example of two comparably arranged conditional independence graphs with equivalence of the block recursive regression models

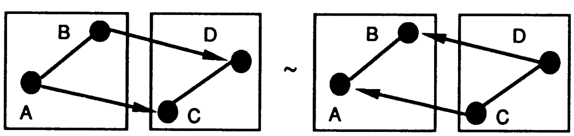


Fig. 18. Example of two conditional independence graphs with equivalence of block recursive regression models which have reversals in the directions of influence

chain graphs such as in Fig. 18. Frydenberg (1989) has given the necessary and sufficient condition for the corresponding conditional independence structures to be identical, but the general condition for equivalence of CG chain models still waits to be formulated and proven.

5.2. *Applications of Equivalences*

The following examples show how equivalences can be exploited in the interpretation of association structures.

A path analysis of five macroeconomic variables has been published by von der Lippe (1977). The variables are employment (X), investment (Y), capital gains (Z), consumption (U) and exports (V). With a univariate recursive regression model a

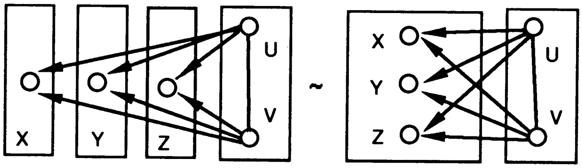


Fig. 19. Example of a path analysis model in which an equivalence shows that the recursive ordering of the responses is not relevant for the evaluation of the research hypothesis

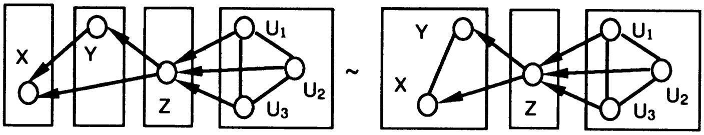


Fig. 20. Example of a path analysis model in which an equivalence leads to a more condensed description of the association structure

dependence chain is implicitly defined as $\mathscr{C} = (\{X\}, \{U\}, \{Z\}, \{U, V\})$. In general this specification is problematic as it permits, for instance, capital gains to be a potential influence for investment but not capital gains to be a potential response to the amount of investment. However, one of the main hypotheses can be expressed with a single-response graph which is equivalent to the multiple-response graph in Fig. 19. Thus, the assumed recursive ordering of the responses is irrelevant for the research hypothesis of interest.

Goldberg (1971) reported a path analysis for a system of variables with an undisputable recursive ordering of the responses. The responses are respondent's vote for president in the USA (X), respondent's satisfaction with politics before the vote (Y) and respondent's party affiliation (Z). There are three variables in the system which are influences only (U_1, U_2, U_3). The implicitly defined dependence chain is $\mathscr{C} = (\{X\}, \{Y\}, \{Z\}, \{U_1, U_2, U_3\})$. A well-fitting model is given by a single-response graph which is equivalent to the multiple-response graph in Fig. 20.

Using this equivalence gives the following interpretation of the structure. Whenever information on respondent's party affiliation is given, the prediction of respondent's satisfaction with politics before the vote and of his actual vote is not improved by information on the remaining potential influences. Thus, a plausible and rather condensed description of the many reported correlation coefficients becomes possible.

In some descriptions of path analysis it is claimed that the analysis permits conclusions on the direction of an influence, which in turn is interpreted as evidence for some causal mechanism. However, the equivalence of models with reversals in the directions of influence such as in Fig. 16 make it evident that such causal interpretations of well-fitting path analysis models are not possible. Any direction of influence is prespecified by the researcher. It is not implied by models for associations nor can it be deduced from corresponding analyses of data.

6. STANDARD MODELS AS ELEMENTS AND SPECIAL CASES OF GRAPHICAL CONDITIONAL GAUSSIAN CHAIN MODELS

Some but not all the association structures discussed in the earlier sections could be described in terms of known and well-studied models, an exception being, for instance, the model of Fig. 9 in Section 4.

We now turn to a more systematic account of this aspect by listing known relations for symmetric association graphs G^a (case 1) for single-response graphs G^{sr} (cases 2–4) and for multiple-response graphs G^{mr} (cases 5 and 6).

Case 1. A symmetric association model given by the graph G^a is

(a) a *covariance selection model* if all variables are continuous and
(b) a *graphical log-linear model* if all variables are discrete.

Covariance selection models have been proposed by Dempster (1972) and studied by Wermuth (1976a, b, 1980), Kiiveri (1983), Porteous (1985a, b) and Speed and Kiiveri (1986). Applications have been described by Hodapp and Wermuth (1983), Kiiveri and Speed (1982), Hodapp (1984) and Edwards (1989). A numerical example is given in Table 3, Section 3.2.

Graphical log-linear models were defined as a model class by Darroch *et al.* (1980). Special aspects of these models have been studied by Goodman (1970), Andersen (1974), Haberman (1974), Sundberg (1975), Wermuth (1976a, b), Whittaker (1982), Edwards and Kreiner (1983), Kiiveri (1983), Edwards and Havránek (1985) and Porteous (1985b). A numerical example is given in Table 2, Section 3.1.

By adapting the nomenclature given by Fienberg (1977) we can state case 2.

Case 2. A univariate CG regression is

(a) a set of *I one-way analysis of variance models* if the response is continuous, the influences are all discrete and have *I* level combinations,
(b) a (*multiple*) *logistic regression with linear and quadratic dependences* if the response is discrete and the influences are all continuous,
(c) a set of *I* (*multiple*) *linear regression models* if the response is continuous, influences are both discrete and continuous and the discrete influences have *I* level combinations,
(d) a (*multiple*) *linear regression model* if the response is continuous and influences are all continuous and
(e) a (*multiple*) *logit model* if the response is discrete and influences are all discrete.

Case 3. A univariate homogeneous CG regression is

(a) an *analysis of variance model* if the response is continuous and all influences are discrete,
(b) a (*multiple*) *linear logistic regression* if the response is discrete and all influences are continuous and
(c) a (*multiple*) *analysis of covariance model* if the response is continuous and influences are of both types. This is also called a model with parallel regressions.

If influences and responses are either all continuous or all discrete, the homogeneous CG distributions are identical with the non-homogeneous distributions, so that we have the same situation as in case 2, items (d) and (e).

Discussions of these models can be found in statistical text-books such as, for continuous responses, Draper and Smith (1966), Kerlinger and Pedhazur (1973), Weisberg (1980) and, for discrete responses, Cox (1970), Bishop *et al.* (1975), Haberman (1978), Andersen (1980), Plackett (1981) and McCullagh and Nelder (1983).

Case 4. A univariate recursive regression model given by a single-response graph G^{sr} having a complete subgraph induced by variables which are influences only specifies

(a) a *univariate path analysis model* if all variables are continuous, i.e. a set of *linear recursive equations with independent residuals*, and
(b) a univariate *recursive graphical model* for contingency tables if all variables are discrete.

Path analysis had been proposed by Wright (1921, 1923, 1934) including suggestions for estimating associations and testing for the goodness of fit of a model. However, conditions under which the estimates and tests have desirable properties were not specified. Such results were provided by Tukey (1954), Wold (1954) and Wermuth (1980). Applications have mainly been reported in the social science literature (Goldberger and Duncan, 1973; Blalock, 1971).

The analogous class of recursive models for contingency tables was defined and studied by Wermuth and Lauritzen (1983) by utilizing previous results of Birch (1963) and Goodman (1973).

Extensions to models with constraints on the endogenous variables and to models with vector variables have been suggested by Kiiveri (1983), Kiiveri and Speed (1982) and Kiiveri *et al.* (1984). They constitute subclasses of CG chain models.

Case 5. A block recursive regression model given by the multiple-response graph G^{mr} defines simultaneous linear equations, in which each equation parameter is proportional to a partial regression coefficient and in which no problems of identification occur (Wermuth, 1988).

The type of hypotheses that is treatable with block recursive regression equations may, but need not, differ from those treatable with linear structural equations as discussed in econometrics (Goldberger, 1964) and in psychometrics.

Case 6. A multiple-response regression model which has unobservable influences and is given by a multiple-response graph G^{mr} with a single-response set, which induces a subgraph of unconnected vertices, defines

(a) a *factor analysis model* (Lawley and Maxwell, 1971) if all variables are continuous and
(b) a *latent class model* (Andersen, 1980) if all variables are discrete.

This fact is in essence contained in Kiiveri (1983), who also points explicitly at unresolved estimation problems that occur for such unobservable influences. Numerical examples of either case are implicitly given in Tables 2 and 3: if variable D in Table 2, Section 3.1, is unobservable, then the marginal table of variables A, B and C is an example of a table obtained under latent class model assumptions. Similarly, if variable U in Table 3, Section 3.2, is unobservable, then the marginal covariance matrix of variables X, Y and Z is an example for a matrix obtained under the assumptions of a factor analysis model with one factor.

ACKNOWLEDGEMENTS

We gratefully acknowledge special financial support by the Johannes Gutenberg-Universität for parts of this research. We thank David Edwards for letting us use his computer program, Ole Barndorff-Nielsen for encouraging us to prepare the paper, Morten Frydenberg for his detailed comments leading to correction, the referees and others for their suggestions on how to improve the presentation.

REFERENCES

Andersen, A. H. (1974) Multidimensional contingency tables. *Scand. J. Statist.*, **1**, 115–127.

Andersen, E. B. (1980) *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland.

Anderson, T. W. (1958) *An Introduction to Multivariate Statistical Analysis*, theorems 2.4.3, 2.5.1. New York: Wiley.

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

Asmussen, S. and Edwards, D. (1983) Collapsibility and response variables in contingency tables. *Biometrika*, **70**, 567–578.

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

Bishop, Y. M. M. (1971) Effects of collapsing multidimensional contingency tables *Biometrics*, **27**, 545–562.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: Massachusetts Institute of Technology Press.

Blalock, H. M., Jr (ed.) (1971) *Causal Models in the Social Sciences*. Chicago: Aldine–Atherton.

Cox, D. R. (1970) *Analysis of Binary Data*. London: Methuen.

—— (1984) Interaction. *Int. Statist. Rev.*, **52**, 1–31.

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

Dawid, A. P. (1979) Conditional independence in statistical theory. *J. R. Statist. Soc.* B, **41**, 1–31.

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

—— (1973) Aspects of the multinomial logit model. In *Proc. 3rd Symp. Multivariate Analysis, Dayton, 1972*, pp. 129–142. New York: Academic Press.

Draper, N. R. and Smith, H. (1966) *Applied Regression Analysis*. New York: Wiley.

Edwards, D. (1989) Hierarchical mixed interaction models. *J. R. Statist. Soc.* B, **52**, 3–20.

—— (1987) A guide to MIM. *Research Report 1*. Statistical Research Unit, Copenhagen.

Edwards, D. and Havránek, T. (1985) A fast procedure for model search in multidimensional contingency tables. *Biometrika*, **72**, 339–351.

Edwards, D. and Kreiner, S. (1983) The analysis of contingency tables by graphical models. *Biometrika*, **70**, 553–562.

Fienberg, S. E. (1977) *The Analysis of Cross Classified Categorical Data*. Cambridge: Massachusetts Institute of Technology Press.

Frydenberg, M. (1986) Blandede interaktions modeller, kausale modeller, kollapsibilitet og estimation. *Statistiske Interna 42*. Aarhus University.

—— (1989) The chain graph Markov property. *Research Report 186*. Department of Theoretical Statistics, Aarhus University.

—— (1990) Marginalization and collapsibility in graphical interaction models. *Research Report 166*. Department of Theoretical Statistics, Aarhus University.

Frydenberg, M. and Edwards, D. (1989) A modified iterative proportional scaling algorithm for estimation in regular exponential families. *Comput. Statist. Data Anal.*, to be published.

Frydenberg, M. and Lauritzen, S. L. (1989) Decomposition of maximum-likelihood in mixed interaction models. *Biometrika*, **76**, in the press.

Geiger, D. and Pearl, J. (1988) On the logic of influence diagrams. In *Proc. 4th Workshop Uncertainty and Artificial Intelligence*, pp. 136–147. Minneapolis.

Goldberg, A. S. (1971) Discerning a causal pattern among data on voting behaviour. In *Causal Models in the Social Sciences* (ed. H. M. Blalock, Jr), pp. 33–48. Chicago: Aldine–Atherton.

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

Goldberger, A. S. and Duncan, O. D. (eds) (1973) *Structural Equation Models in the Social Sciences*. New York: Seminar.

Goodman, L. A. (1970) The multivariate analysis of qualitative data. *J. Am. Statist. Ass.*, **65**, 226–256.

—— (1971) Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *J. Am. Statist. Ass.*, **66**, 339–344.

—— (1973) The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, **60**, 179–192

Haberman, S. (1974) *The Analysis of Frequency Data*. Chicago: University of Chicago Press.

—— (1978) *Analysis of Qualitative Data*, vol. 1. New York: Academic Press.

Hodapp, V. (1984) *Analyse Linearer Kausalmodelle*. Bern: Huber.

Hodapp, V. and Wermuth, N. (1983) Decomposable models: a new look at interdependence and dependence structures in psychological research. *Multiv. Behav. Res.*, **18**, 361–390.

Jöreskog, K. G. (1977) Structural equation models in the social sciences: specification, estimation and testing. In *Applications of Statistics* (ed. P. R. Krishnaiah), pp. 267–287. Amsterdam: North-Holland.

Kerlinger, F. N. and Pedhazur, E. J. (1973) *Multiple Regression in Behavioral Research*. New York: Rinehart and Winston.

Kiiveri, H. T. (1983) A unified theory of causal models. *PhD Thesis*. University of Western Australia, Perth.

Kiiveri, H. T. and Speed, T. P. (1982) Structural analysis of multivariate data: a review. In *Sociological Methodology* (ed. S. Leinhardt), pp. 209–289. San Francisco: Jossey Bass.

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

Lauritzen, S. L. (1982) *Lectures on Contingency Tables*, 2nd edn. Aalborg: Aalborg University Press.

—— (1988) Mixed graphical association models. *Scand. J. Statist.*, **16**, in the press.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer H.-G. (1988) Independence properties of directed Markov fields. *Research Report R-88-32*. Institute of Electronic Systems, Aalborg University.

Lauritzen, S. L. and Wermuth, N. (1984) Mixed interaction models. *Research Report R-84-8*. Aalborg University.

—— (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31–54.

Lawley, D. N. and Maxwell, A. E. (1971) *Factor Analysis as a Statistical Method*, 2nd edn. London: Butterworth.

Leimer, H.-G. (1989) Triangulated graphs with marked vertices. *Ann. Discr. Math.*, **41**, 311–324.

von der Lippe, P. (1977) Beschäftigungswirkungen durch Umverteilung. *WSI-Mitt.*, **8**, 505–512.

McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models*. London: Chapman and Hall.

Pearl, J. (1986) A constraint propagation approach to probabilistic reasoning. In *Uncertainty and Artificial Intelligence* (eds L. M. Kanal and J. Lemmer), pp. 357–370. Amsterdam: North-Holland.

Plackett, R. L. (1981) *The Analysis of Categorical Data*. London: Griffin.

Porteous, B. T. (1985a) Improved likelihood ratio statistics for covariance selection models. *Biometrika*, **72**, 473–475.

—— (1985b) Properties of log-linear and covariance selection models. *PhD Thesis*. University of Cambridge.

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

Spielberger, C. D., Gorsuch, R. L. and Lushene, R. E. (1970) *Manual for the State-trait Anxiety Inventory*. Palo Alto: Consulting Psychologists Press.

Spielberger, C. D., Jacobs, G. A., Crane, R. S. and Russell, S. F. (1983a) On the relation between family smoking habits and the smoking behaviour of college students. *Int. Rev. Appl. Psychol.*, **32**, 54–69.

Spielberger, C. D., Jacobs, G., Russell, S. and Crane, R. (1983b) Assessment of anger: the state-trait anger scale. In *Advances in Personality Assessment* (eds J. N. Butcher and C. D. Spielberger), vol. 2, pp. 159–187. Hillsdale: Erlbaum.

Sundberg, R. (1975) Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. *Scand. J. Statist.*, **2**, 71–79.

Tukey, J. W. (1954) Causation, regression, and path analysis. In *Statistics and Mathematics in Biology* (eds O. Kempthorne, T. A. Bancroft, J. W. Gowen and J. L. Lush), pp. 35–66. Ames: Iowa State College Press.

Verma, T. (1988) Causal networks: semantics and expressiveness. In *Proc. 4th Workshop Uncertainty and Artificial Intelligence*, pp. 352–359. Minneapolis.

Weisberg, S. (1980) *Linear Regression Analysis*. New York: Wiley.

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

—— (1976b) Model search among multiplicative models. *Biometrics*, **32**, 153–163.

—— (1980) Linear recursive equations, covariance selection and path analysis. *J. Am. Statist. Ass.*, **75**, 963–972.

—— (1987) Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. R. Statist. Soc.* B, **49**, 353–364.

—— (1988) Block-recursive linear regression equations. *Berichte zur Stochastik und Verwandten Gebieten, 88-4.* Mainz: Johannes Gutenberg-Universität.

—— (1989a) Moderating effects of subgroups in linear models. *Biometrika*, **76**, 81–92.

—— (1989b) Moderating effects in multivariate normal distributions. *Methodika*, **3**, 1–20.

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

Whittaker, J. (1982) Glim syntax and simultaneous tests for graphical log-linear models. *Lect. Notes Statist.*, **14**, 98–108.

Whittemore, A. S. (1978) Collapsibility of multidimensional contingency tables. *J. R. Statist. Soc.* **B**, **40**, 328–340.

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

Wold, H. O. (1954) Causality and econometrics. *Econometrica*, **22**, 162–177.

Wright, S. (1921) Correlation and causation. *J. Agricult. Res.*, **20**, 557–585.

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

—— (1934) The method of path coefficients. *Ann. Math. Statist.*, **5**, 161–215.