Deriving and testing hypotheses in chain graph models

Svend Kreiner
Jørgen Holm Petersen
Volkert Siersma

Research Report 09/9
Department of Biostatistics
University of Copenhagen
Deriving and testing hypotheses in chain graph models

Svend Kreiner¹, Jørgen Holm Petersen¹, Volkert Siersma²

¹Dept. of Biostatistics, University of Copenhagen.
²Research Unit for general Practice in Copenhagen, University of Copenhagen

Abstract
A chain graph model is a block recursive model where some pairs of variables are conditionally independent given all current and prior variables. In high-dimensional models, testing these independencies directly is computationally infeasible. Instead, global Markov properties of chain graph models may be used to derive hypotheses of conditional independence that require conditioning with subsets of the variables. This paper provides a systematic account of how these properties may be used for statistical analysis with a focus on data reduction by marginalizing over variables, conditioning on outcomes of variables, and/or collapsing onto functions of subsets of variables.

Keywords: graphical models, chain graphs, global Markov properties, collapsibility

1 INTRODUCTION
In multivariate statistical analysis we distinguish between three problems. One problem is to develop a statistical model describing the probabilistic structure among variables. A second problem is to check that the model provides an adequate fit to the observed data. The third problem is to describe relationships between key variables of the model taking the confounding and mediating effect of the other variables into account. This provides answers to open research questions concerning dependence between variables. Statistical tests of independence may be involved in the solutions to all three problems. First, tests of independence may be required if subject matter knowledge is inadequate to completely define the probabilistic structure such that we have to engage in exploratory model search in order to complete the model. Second, model checking may require confirmatory tests of independence in order to support the model’s assumptions of both independence and dependence. Finally, research questions are often formulated as hypotheses of conditional independence to be assessed using the observed data.

Graphical models defined by conditional independence of variables provide a useful frame of inference for such analyses. In graphical models the independence structures are represented by mathematical networks – graphs – where a missing edge between variables means that variables are conditionally independent. In their original conception, graphical models were models of joint
distributions; specifically all edges were undirected because the relations of conditional independence were symmetric (Darroch, Lauritzen and Speed, 1982). Analysis of data by undirected graphical models is relatively straightforward using results on conditioning and marginalizing that have been worked out by Whittaker (1990) and Lauritzen (1996).

In an extension of the family of graphical models, temporal and/or causal order imposes structure on the model by making the edges of the graph directed – graphically represented by arrows. For example, the progression of time requires that yesterday’s blood pressure can affect today’s, but not the other way around. If all edges can be assigned a direction in such a way that no causal cycles occur, the graph is referred to as a directed acyclic graph (DAG). In the DAGs where the edges have causal interpretation epidemiological notions of mediation and confounding are well defined. Notably, relations of conditional independence are readily – but not always trivially – found from the DAG. Analysis of data by DAGs turns out simpler than by undirected graphical models since it may be reduced to a sequence of standard regression analyses.

In between the undirected graphical models and the DAGs we find the general class of chain graph models (CGMs) which also assume some causal and/or temporal order, but acknowledge that there are situations where it is impossible or unwanted to assign a causal direction to many of the edges in the model. A CGM makes two assumptions. The first is that the variables can be organized in groups, also called recursive blocks, in a meaningful – causal and/or temporal – way according to some recursive structure. The second assumption is that some variables are conditionally independent given conditioning sets of variables containing current and prior variables according to a graph with undirected edges within blocks and directed edges between blocks.

Even though there are a number of interesting applications of chain graph models the general impression is that CGMs, despite their versatility, have not found popularity in applications to the degree that for example DAGs have. One reason for this is that it is relatively difficult to navigate the CGM. If we have a CGM, devised by an expert and/or by empirical model estimation one of the aims will be to test a number of specific conditional independence assumptions. In high-dimensional CGMs, it is often unrealistic to test the conditional independence assumptions of the models directly because the numbers of variables in the conditioning sets are too large. Instead, the graph can be used to derive hypotheses referred to as global Markov property (GMP) hypotheses where the conditioning sets are small enough for tests of conditional independence to be feasible since GMP hypotheses all are true if the assumptions of the model are true. GMP hypotheses are
hypotheses that are derived from the global Markov properties of CGMs. The first purpose of the paper is to present a systematic account of these hypotheses and the way they are identified.

We consider three ways to reduce the number of variables in a model: by marginalizing over variables, by conditioning on outcomes of variables or by replacing a number of variables with a summary function of these variables. Primarily, the analysis of data by CGMs, is more complicated than by undirected graphical models or DAGs since results on marginalizing and conditioning in the former have not been developed to the same degree. The second purpose of the paper is to show how the global Markov properties and the GMP hypotheses may be used to construct reduced models in each of these ways, and how analysis of data by the reduced models can be used to address hypotheses in the complete model.

The remainder of the paper is organized as follows. Section 2 summarizes notions and definitions from graph theory of importance for the theory of CGMs. CGMs are defined and discussed in Section 3. Most of the content of Sections 2 and 3 is known from other sources (e.g. Lauritzen, 1996; Whittaker, 1990), but the definition of truncated and tapered chain graph models and the notion of GMP hypotheses appear to be new to this paper. Section 4 presents results on marginalization and conditioning in graphical models. Again, some of these results are well-known for special types of graphical models, but results relating to the general family of CGMs are rarely discussed in the literature on graphical models and are, to our knowledge, also new to this paper. Marginalization in graphical models is closely connected to graphical collapsibility as defined by Whittaker, (1990). Section 5 defines and discusses a notion of functional collapsibility of graphical models that may be useful, when graphical collapsibility suggests analysis of conditional independence in marginal models with an impractical number of variables. Finally, inference in CGMs is briefly discussed and illustrated in Section 6 using data from a longitudinal study of living conditions in Denmark in the period 1968 – 1992.

2 Graphs
This section reviews relevant graph theoretical results. Specifically, it focuses on the notions of separation and moralizing. A comprehensive treatment of graph theoretical results of importance for CGMs is given by Lauritzen (1996).
2.1 Graphs and subgraphs

A graph is a pair $G = (V, E)$, where $V$ is a finite set of nodes and $E \subseteq V \times V$ represents links between nodes. Graphs are often displayed as visual networks with dots representing nodes and edges or arrows representing links between nodes. Two nodes, $a$ and $b$, are joined by an (undirected) edge in the visual graph if $(a, b) \in E$ and $(b, a) \in E$ or by a (directed) arrow pointing from $a$ to $b$ if $(a, b) \in E$ while $(b, a) \notin E$. Hence, $E$ partitions into two subsets, $E = U \cup D$, where $U$ is the set of undirected edges $U = \{(a, b) \in E : (b, a) \in E\}$ and $D$ is the set of directed edges $D = \{(a, b) \in E : (b, a) \notin E\}$. The graph is undirected if $D$ is empty. A directed or undirected edge between two nodes is referred to as a direct link between the nodes.

A subset $W \subset V$ defines a subgraph $\text{Subgraph}(W, G) = (W, F)$ where $F = E \cap (W \times W)$. When it is obvious that $W$ is a subset of nodes of a graph $G$ we will refer to the subgraph as $G_W$.

2.2 Paths and separation in graphs

A path from node $a_0$ to node $a_k$ in $G$ is a sequence of distinct nodes – $a_0, a_1, \ldots, a_{k-1}, a_k$ – such that $(a_{i-1}, a_i) \in E$ for all $i = 1, \ldots, k$. We say that two nodes, $a$ and $b$, are connected in $G$ if there is a path from $a$ to $b$ and a path from $b$ to $a$. The two paths connecting $a$ and $b$ may be one and the same undirected path or two distinct paths. Connection defines a partition of $V$ into the connectivity components of the graph: $V = C_1 \cup \ldots \cup C_c$, where nodes in the same component are connected and nodes in different components are not.

Consider three disjoint subsets, $A \subset V$, $B \subset V$ and $S \subset V$. We say that $S$ separates $A$ from $B$ in $G$ if all paths from nodes in $A$ to nodes in $B$ and vice versa contain at least one node in $S$. Hence, if $S$ separates $A$ and $B$, the subgraph $G_{V\setminus S}$ has no paths from nodes in $A$ to nodes in $B$. Note that if $S$ separates $A$ and $B$ and $S \subset T \subset V \setminus \{A, B\}$ then $T$ separates $A$ and $B$.

2.3 Parents, neighbors, boundaries and ancestral sets

Graph theory defines a number of different subsets of nodes related to a node: the parents of $a$: $\text{pa}(a) = \{ b \in V : (b, a) \in D \}$, the neighbors of $a$: $\text{ne}(a) = \{ b \in V : (a, b) \in U \}$, and the boundary of $a$: $\text{bd}(a) = \text{pa}(a) \cup \text{ne}(a)$. These definitions easily generalize to definitions for sets of nodes. If $A$ is a set of nodes then $\text{pa}(A) = \bigcup_{a \in A} \text{pa}(a)$, $\text{ne}(A) = \bigcup_{a \in A} \text{ne}(a) \setminus A$, and $\text{bd}(A) = \text{pa}(A) \cup \text{ne}(A)$. Finally, $A \subseteq V$ is ancestral, if $\text{bd}(a) \subseteq A$ for all $a \in A$. For any $A \subseteq V$ there will always be a smallest ancestral set, $\text{An}(A)$, containing $A$. 
2.4 Chain graphs, truncated chain graphs and tapered chain graphs

Chain graphs are graphs where, on top of the structure inherent in the graph, the set of nodes is partitioned into a number of disjoint subsets or recursive blocks, \( V = V_1 \cup \ldots \cup V_t \); nodes in the same block may only be joined by undirected edges while nodes from different blocks may only be joined by arrows pointing from the node in the block with the higher index to the node in the block with the lower index. The graphs in Figure 1a and 1c are chain graphs whereas Figures 1b and 1d are undirected graphs.

The number of recursive blocks defining the chain graph is called the length of the chain. Associated with a chain graph \( G \) of length \( t \), we define a number of truncated chain graphs, 
\[
\text{Trunc}(G,s) = G_{\mathcal{W}}, \quad \text{where} \quad \mathcal{W} = V_s \cup \ldots \cup V_t \quad \text{for} \quad 1 \leq s < t.
\]
Truncated chain graph are ancestral in \( G \). The graph in Figure 1c is a truncated version of the chain graph in Figure 1a (for both possibilities of recursive structure).

Connectivity components of chain graphs are called chain components. The nodes of a chain component always belong to the same block, but a block may contain more than one chain component. Subgraphs of chain components are always undirected. The chain graph in Figure 1a has four chain components, \{A,B\}, \{C,D\}, \{E\} and \{F,G\}.

Assume finally, that the first recursive block consists of \( k \) terminal chain components, \( V_1 = C_1 \cup \ldots \cup C_k \). We will refer to a subgraph of \( G \) without some, but not all terminal chain components, as a tapered chain graph. For instance, \( G_{V \setminus T} \), where \( T = C_2 \cup \ldots \cup C_k \), is a tapered chain graph.

2.5 Moralized graphs

The moral graph \( G^M \) of a chain graph \( G \) is a graph with the same nodes as \( G \) but with only undirected edges. Two nodes, \( a \) and \( b \), are joined by an undirected edge in \( G^M \) if one of the following two conditions is satisfied:

1) \( (a,b) \in E \) or \( (b,a) \in E \),
2) \( a \) and \( b \) have children in the same chain component \( C_i \): \( (a,b) \in \text{pa}(C_i) \).

Separation in moralized graphs is important for CGMs. It is illustrated in Figure 1 presenting moralized versions of a complete and a truncated chain graph. D and E are separated by \{C,F,G\} in the moralized graph, Figure 1b, and by \{C,G\} in the moralized truncated graph, Figure 1d.
Figure 1 Chain graphs. Dotted edges in (b) and (d) have been added by moralization of (a) and (c) respectively.

3 Chain graph models

Chain graph models (CGMs) are multivariate statistical models defined by two sets of assumptions: assumptions of recursive structure and assumptions of conditional independence.

3.1 Recursive structure

The assumption of recursive structure partitions the variables of the model into disjoint subsets of variables \( V_1, \ldots, V_t \) that are assumed to be naturally ordered, e.g. by temporal and/or causal structure. We refer to \( V_1, \ldots, V_t \) as recursive blocks or recursive levels and write \( [V_1] \leftarrow [V_2] \leftarrow [V_3] \leftarrow \ldots \leftarrow \)}
[V_t] when we need to spell out the recursive order of the blocks. Note, that we adopt notation where blocks with lower numbers follow after blocks with higher order in the recursive sequence. The recursive structure defines a statistical model where the joint distribution of all variables is given by the product of conditional probabilities or densities of variables in specific blocks given all the variables in prior blocks, \( P(V) = \prod_{i=1}^{t-1} P(V_i|V_{i+1},...,V_t)P(V_t) \).

The recursive structure of the model defines legitimate conditioning sets of variables for statistical analysis of conditional association among pairs of variables. Legitimate conditioning sets for \( \{X_i,X_j\} \subseteq V \) where \( i<j \) consists of variables that are concurrent with or prior to \( X_i \) according to the recursive structure. If \( X_i \in V_k, S \) is a legitimate conditioning set for \( (X_i,X_j) \) if \( S \subseteq \cup_{j=k+1}^{t} V_j \setminus \{X_i,X_j\} \).

### 3.2 Conditional independence

Stochastic variables A and B are conditionally independent given C if \( P(A,B|C) = P(A|C)P(B|C) \) or, alternatively \( P(A|B,C) = P(A|B) \). Research questions can often be framed as hypotheses of conditional independence. We adopt the notation of David (1979) writing \( A \perp B | C \) for conditional independence of A and B given C. Note that conditional independence may be a trivial consequence of the way variables are defined. If, for instance, A is a function of C, \( A = h(C) \), it is trivial that \( A \perp B | C \) since \( P(A=a | C=c, B) = 1 \) if \( a=h(c) \) and \( P(A=a | C=c, B) = 0 \) if \( a \neq h(c) \) irrespective of the outcome on \( B \). The second set of assumptions defining a CGM assumes that some pairs of variables are non-trivially conditionally independent given the largest of all legitimate conditioning sets.

**Example 1:** A CGM containing seven variables and organized in three recursive blocks, \( [A,B] \leftarrow [C,D] \leftarrow [E,F,G] \) is shown in Figure 1a. The recursive structure rewrites the joint distribution of the seven variables as \( P(A,B,C,D,E,F,G) = P(A,B|C,D,E,F,G)P(C,D|E,F,G)P(E,F,G) \) and defines legitimate conditioning sets for each pair of variables. The largest legitimate conditioning set for an analysis of the conditional association between A and E consists of all remaining variables of the model, whereas the largest legitimate conditioning set for an analysis of the association between D and E is \( \{C,F,G\} \).

Model 1, below is a CGM defined by the recursive structure suggested above and a set of 12 conditional independence assumptions, whereas Model 2 is an undirected graphical model for the (same) joint distribution as in Model 1. The conditional independence assumptions of Model 2 are derived from Model 1. Most of the assumptions are trivial but two (c.i. 13 and 14) are non-trivial.
consequences implied by the independence assumptions of Model 1. The arguments leading from Model 1 to Model 2 require the results on global Markov properties as summarized in Proposition 1 below.

**Model 1**

Recursive structure: [A,B]←[C,D]←[E,F,G]

Pairwise conditional independencies:

\[(c.i. 1): \quad A \perp D \mid BCEFG, \quad (c.i. 8): \quad C \perp G \mid DEF, \]
\[(c.i. 2): \quad A \perp E \mid BCDFG, \quad (c.i. 9): \quad D \perp E \mid CFG, \]
\[(c.i. 3): \quad A \perp G \mid BCDEF, \quad (c.i. 10): \quad D \perp F \mid CEG, \]
\[(c.i. 4): \quad B \perp C \mid ADEFG, \quad (c.i. 11): \quad E \perp F \mid G, \]
\[(c.i. 5): \quad B \perp E \mid ACDFG, \quad (c.i. 12): \quad E \perp G \mid F, \]
\[(c.i. 6): \quad B \perp F \mid ACDEG, \]
\[(c.i. 7): \quad B \perp G \mid ACDEF, \]

**Model 2** – the joint distribution, \(P(A,B,C,D,E,F,G)\)

\([A,B,C,D,E,F,G]\)

Trivial conditional independencies derived from Model 1:

\[(c.i. 1): \quad A \perp D \mid BCEFG, \quad (c.i. 8): \quad C \perp G \mid DEF, \]
\[(c.i. 2): \quad A \perp E \mid BCDFG, \quad (c.i. 9): \quad D \perp E \mid CFG, \]
\[(c.i. 3): \quad A \perp G \mid BCDEF, \quad (c.i. 10): \quad D \perp F \mid CEG, \]
\[(c.i. 4): \quad B \perp C \mid ADEFG, \quad (c.i. 11): \quad E \perp F \mid G, \]
\[(c.i. 5): \quad B \perp E \mid ACDFG, \quad (c.i. 12): \quad E \perp G \mid F, \]
\[(c.i. 6): \quad B \perp F \mid ACDEG, \]
\[(c.i. 7): \quad B \perp G \mid ACDEF, \]

Non-trivial conditional independencies derived from the global Markov properties discussed below

\[(c.i. 13): \quad C \perp G \mid ABDEF \quad (c.i. 14): \quad D \perp E \mid ABCFG \]

### 3.3 Markov graphs

The assumptions of the CGMs are encoded in chain graphs where variables are represented by nodes with a recursive structure representing the recursive structure of the variables. Variables are joined in the graph, unless they are conditionally independent according to the assumptions of the model. Edges between variables are undirected if the variables are posited in the same block and directed if they are in different blocks.
The graphs are often called independence graphs as the conditional independence assumptions of the model can be read from the graph. They are also termed Markov graphs (Frank and Strauss, 1986) because a number of Markov properties of the statistical model may be revealed by graph theoretical analyses of the graphs (Lauritzen, 1996). In recent years, the terminology of graphical models has become somewhat indistinct. Different types of graphs with different meanings and implications are called independence graphs (Cox and Wermuth, 1996) and a models are sometimes referred to as “graphical” models if they in some way can be represented or illustrated using a visual graph (Cox and Wermuth, 1996, Cowell et al. 2007, Drton and Richardson, 2008). Cox and Wermuth, in particular, define a number of independence graphs, each with their own implications for the statistical models. For this reason, we prefer to refer to the graphs encoding the conditional independence assumptions of CGMs as Markov graphs. This emphasizes that the type of graphs discussed here are the particular type from which certain Markov properties discussed in the next subsection may be inferred.

The graph in Figure 1a is the Markov graph of Model 1, whereas the moralized graph, Figure 1b, is the Markov graph of Model 2.

The assumed recursive structure defines more or less complicated structures. Undirected graphical models are models making no assumptions of recursive structure at all while models for DAGs assume that each recursive block contains exactly one variable. In this paper we do not distinguish between these different types of CGMs but are concerned with properties that apply for all these types of models.

3.4 Truncated and tapered chain graph models
It is convenient to be able to distinguish between two types of chain graph models that may be derived from a given CGM.

Truncated CGMs are defined in the same way as truncated chain graphs, that is by the distributions, \( P(V_s, V_{s+1}, \ldots, V_t) = \prod_{i=s, s+1, \ldots, t} P(V_i | V_{i+1}, \ldots, V_t) P(V_t) \). A truncated CGM is itself a CGM with a Markov graph which is equal to the truncated graph, \( \text{Trunc}(G, s) \) (Lauritzen, 1996, Lemma 3.31).

Tapered CGMs are CGMs defined by marginalizing over some but not all terminal chain components. The Markov graph of a tapered CGM is equal to the subgraphs of the complete CGM without the terminal chain components that have been removed from the model (Lauritzen, 1996, Lemma 3.32).
Truncated CGMs and tapered CGMs are naturally viewed as models defined by marginalizing over some of the variables of a CGM. General results on marginalization in CGMs are given in Section 4.

3.5 Global Markov properties

A particular convenient feature of graphical models is that results on conditional independence in marginal models may be read directly off the graphs, due to the Markov properties (Lauritzen, 1996 page 55-57). The most important of these is

The global Markov property of CGMs: Let W be the smallest ancestral set, W = An(A∪B∪S), of three disjoint subsets of nodes of a Markov graph, G, of a CGM. Let H = G_W be the subgraph containing W and let H^M be the moralized version of H. If S separates A from B in H^M then A⊥B|S.

When a CGM contains so many variables that direct tests of the independence assumptions becomes impractical, the global Markov property is the key to testing the assumptions. To use the global Markov property, we first have to identify separating sets, S, satisfying the Markov property for each pair of conditionally independent variables, {A,B}. Second, we test (the smaller) conditional independence hypothesis: A⊥B|S. Note, however, that the number of potentially separating sets is very large in high-dimensional models. For the global Markov property to be useful for model checking and construction, we need an easy way to identify subsets of variables satisfying the global Markov property. The following proposition, which is a direct consequence of the global Markov property, provides such a way.

Proposition 1: Let G be the Markov graph of a CGM where A, B and S are three disjoint subsets. If S separates A from B in G^M, then A⊥B|S.

Proof: This follows from Lauritzen (1996, Lemma 3.33 and Theorem 3.7). Let W be the smallest ancestral set, W = An(A∪B∪S) for A,B and S in Proposition 1 and let H = G_W. Since all paths from nodes in A to nodes in B in H^M are also paths in G^M and since S⊆W separates A from B in G^M it follows that S also separates A from B in H^M. Hence, according to the global Markov property, A⊥B|S.
Proposition 1 also applies to truncated and tapered CGMs. The graph, Figure 1a, may for example define a CGM with three recursive blocks while Figure 1c defines a truncated model. Both models have moralized graphs with separation properties corresponding to global Markov properties. The definition of the graph in Figure 1a assumes that $D \perp F | C, E, G$. The moral graph, Figure 1b, shows that it cannot be assumed that $D \perp F | A, B, C, E, G$ because $\{A, B, C, E, G\}$ does not separate $D$ and $F$ in the moralized graph, Figure 1b. Since $\{C, G\}$ separates $D$ and $F$ in the moralized truncated graph, Figure 1d, we see, however that the model defined by Figure 1a implies that $D \perp F | C, G$. Note that the latter conditional independence statement is stronger than the original statement of the assumption; notably, inference for this relationship is more efficient in the latter situation.

3.6 Global Markov Property hypotheses

The various conditional independence statements defined by Proposition 1 for the complete, truncated and tapered models are called global Markov property (GMP) hypotheses.

Assume that $S_1$ separates $A$ and $B$ in the moral graph and that $S_2$ satisfies $S_1 \subseteq S_2$, $A \cap S_2 = \emptyset$ and $B \cap S_2 = \emptyset$. From this it follows that $S_2$ also separates $A$ from $B$ such that $A \perp B | S_2$. This suggests the following definitions:

**Definition 1**: Let $H$ be the complete, truncated or tapered chain graph, let $H^M$ be the moralized version of $H$, and let $S_A = \text{ne}(A)$ and $S_B = \text{ne}(B)$ in $H^M$. $S_A$ and $S_B$ obviously separate $A$ from $B$ in $H^M$ if there is no edge between $A$ and $B$ in $H^M$. The corresponding GMP hypotheses, $A \perp B | S_A$ and $A \perp B | S_B$ are referred to as local GMP hypotheses.

If $A$ belongs to the first block of $H$, $S_A$ may be easily identified without moralization of $H$ since moralization cannot add edges to $A$ in $H^M$. Some of the local GMP hypotheses are therefore easy to identify if $A$ or $B$ belongs to the first recursive block of $H$.

**Definition 2**: $A \perp B | S$ is a minimal GMP hypothesis if there are no smaller subsets, $T \subset S$, that separates $A$ and $B$ in either the moralized chain graph or in one of the moralized truncated graphs.
**Definition 3:** A ⊥ B | S is *a smallest possible GMP hypothesis* if there are no other minimal GMP hypotheses, A ⊥ B | T, where the number of variables in T is smaller than the number of variables in S.

Testing all GMP hypotheses in order to check the adequacy of a high-dimensional graphical model is not practical due to the size of the tables and the loss of power of the test statistics. Instead one may test local, minimal or smallest GMP hypotheses where the conditioning sets may be small enough to handle without problems and without loss of power. Local GMP hypotheses have the advantage that they are relatively easy to identify, but are in most cases neither minimal nor smallest GMP hypotheses. If the technical facilities for separation in graphs are available, addressing a smallest possible GMP hypothesis is to be preferred.

**4 MARGINALIZATION AND CONDITIONING IN CHAIN GRAPH MODELS**

In high-dimensional data, the number of variables will in itself often be prohibitive. To facilitate analysis we therefore have to search for ways to partition the analysis into a series of subanalyses where the separate subanalyses only consider subsets of variables. There are three obvious ways to reduce data for subanalyses: 1) by marginalizing over some of the variables, 2) by conditioning over variables, and 3) by collapsing over a set of variables onto a summary function of the variables. In all three cases, the question to be addressed is whether hypotheses of conditional independence transfer from the original to the smaller model so that rejections of hypotheses in the reduced model can be taken as evidence that the purported independence structure of the complete model is incorrect. Marginalizing and conditioning will be discussed in this section while the discussion of collapsibility onto functions of variables will be deferred to the next section.

*Example 1 revisited:* Consider first the model defined by Figure 1a and the question of what happens if we marginalize over C or condition on C=c.

Some of the consequences are trivial. First, the variables remaining in the reduced model must have the same recursive structure as in the original model, [AB] ← [D] ← [EFG]. In this structure it is convenient to refer to the [AB], [D] and [EFG] blocks as respectively the future, the present and the past of C.
Second, it is also trivial that parts of the independence structure remain in the reduced model. Marginalizing over C, for instance, has no effect on the independence structure of the past since C is not a legitimate conditioning variable for this part of the model. Conditioning on C=c, on the other hand, does not alter the independence structure of the present and future variables since, for instance, the $\text{A} \perp \text{D} | \text{B,C,D,E,F,G}$ assumption of the complete model implies that $\text{A} \perp \text{D} | \text{B,D,E,F,G}$, C=c.

These observations are general. Marginalizing over a variable does not change the past of the variable, but may change the independence structure of the future and the present if the marginalized variable is a mediator of indirect links in the complete model. Conditioning, on the other hand, will not change the present and the future, but may cause association among the variables of the past due to selection. The non-trivial effects of marginalizing and conditioning has to do with the fact that parts of the independence structure of the present and future may remain unchanged under marginalization and that parts of the structure of the past may remain under conditioning. To identify the parts that do transfer to the reduced models we need to check the global Markov properties of the complete model.

Consider first the association between A and D when we marginalize over C. Since $\{\text{D,E,F}\}$ and therefore also $\{\text{B,D,E,F}\}$ separates A from G in Figure 1b it follows that $\text{A} \perp \text{G} | \text{BDEF}$ which is a legitimate assumption of conditional independence for a CGM defined on top of the recursive $[\text{AB}] \leftarrow [\text{D}] \leftarrow [\text{EFG}]$ structure. There will consequently be no need for an arrow pointing from G to A in the Markov graph of the marginal model. The same arguments apply for (B,E) and (B,F) since $\{\text{A,D}\}$ is a minimal separating set for both pairs of variables. However, directed edges $\text{A} \leftarrow \text{D}$, $\text{A} \leftarrow \text{E}$, $\text{D} \leftarrow \text{E}$, and $\text{D} \leftarrow \text{F}$ appear, since C is a member of all separating sets in Figure 1b for the four pairs of variables.

To see whether edges connecting D to E and F are required after marginalizing over C, we have to consider the truncated CGM in Figure 1c, since missing edges from E, F and/or G to D in the marginal model require assumptions of conditional independence where the conditioning sets only include variables from the [EFG] block. The moralized truncated graph is shown in Figure 1d. In this graph, all the sets of variables separating D from E and F contain C so that we have no results justifying an assumption that D and one of the variables in the [EFG] block are conditionally independent given the remaining two variables from this block. The Markov graph of the marginal model therefore must include edges connecting D to E and F as shown in Figure 2a.
To examine the effect of conditioning on outcomes of $C$ on the independence structure of the $[EFG]$ block we consider the truncated model, Figure 1c, marginalized over $D$. Since marginalizing over $D$ create an arrow from $G$ to $C$ in the Markov graph of the marginal $[C] \leftarrow [EFG]$ model, it follows that the moralized graph is saturated since $E$, $F$, and $G$ are all parents of $C$ in the model without $D$. From this it follows that the independence structure of Figure 1a does not imply that $E$ is conditionally dependent of $F$ and $G$ given $C=c$. Therefore, the Markov graph associated with the conditional distribution of $ABDEFG$ given $C=c$ includes undirected edges connecting $E$ to $F$ and $G$ as shown in Figure 2b.

![Figure 2](image)

**Figure 2.** Reduced chain graph models derived from the model defined in Figure 1a: (a) after marginalizing over $C$ and (b) after conditioning on outcomes of $C$. Bold lines are lines required by marginalizing or conditioning.

The following two corollaries extend the arguments above to the class of CGMs in general. In these corollaries, $G = (V,E)$ is the Markov graph of a CGM with variables in $k$ recursive blocks, $V = \bigcup_{i=1}^{k} V_i$. $A$ and $B$ are two variables, $A \in V_r$, $B \in V_s$, $r \leq s$ that are not connected by an edge in $G$ and $C \in V_t$, is a third variable. The problem to be addressed in this section is whether marginalizing over $C$ or conditioning on $C=c$ induces an edge between $A$ and $B$ in the reduced models.

### 4.2 Marginalizing

$H_{\text{marg}(C)}$ is the Markov graph of the marginal model after marginalizing over $C$. Corollary 1 provides sufficient conditions under which there will be no edge between $A$ and $B$ in $H_{\text{marg}(C)}$. 


Corollary 1. There is no edge connecting A to B in \( H_{\text{marg}(C)} \) if either

1) \( t < r \), or

2) there is a minimal GMP hypothesis, \( A \perp B \mid S \) derived from \( \text{Trunc}(G,r) \) where \( C \notin S \).

Proof: C belongs to the future of A if \( t < r \). Marginalizing over C will therefore have no implication for the conditional independence of A and B since C is not a legitimate conditioning variable.

If \( r \leq t \), \( C \in \bigcup_{i=r}^k V_i \setminus \{A,B\} \). \( A \perp B \mid S \) is a minimal GMP hypothesis in \( \text{Trunc}(G,r) \). All paths from A to B in \( \text{Trunc}(G,r) \) therefore pass through S and therefore also through \( \bigcup_{i=r}^k V_i \setminus \{A,B,C\} \supseteq S \) since \( C \notin S \). Therefore, according to the global Markov property of G, \( A \perp B \mid \bigcup_{i=r}^k V_i \setminus \{A,B,C\} \) which corresponds to an assumption of a missing edge between A and B in \( H_{\text{marg}(C)} \).

4.3 Conditioning

\( H_{\text{cond}(C=c)} \) is the Markov graph of the marginal model after conditioning on \( C = c \). Corollary 2 provides sufficient conditions under which there will be no edge between A and B in \( H_{\text{cond}(C=c)} \).

Corollary 2. There is no edge connecting A to B in \( H_{\text{cond}(C=c)} \) if either

1) \( r \leq t \), or

2) if \( t < r \) and either \( A \notin \text{pa}(C) \) or \( B \notin \text{pa}(C) \) in the Markov graph of \( P(C, \bigcup_{i=r}^k V_i) \).

Proof: To prove the first part note that \( C \in \bigcup_{i=r}^k V_i \setminus \{A,B\} \) when \( r \leq t \). The missing edge between A and B in G means that \( A \perp B \mid \bigcup_{i=r}^k V_i \setminus \{A,B\} \) so that \( A \perp B \mid \bigcup_{i=r}^k V_i \setminus \{A,B,C\}, C = c \) for all c.

The second part follows from the global Markov property of the marginal CGM for \( (C, \bigcup_{i=r}^k V_i) \). An edge between A and B appears in the moralized version of the marginal graph only if \( A \in \text{pa}(C) \) and \( B \in \text{pa}(C) \) since A belongs to the recursive block immediately below C in \( P(C, \bigcup_{i=r}^k V_i) \).
For undirected graphical models, marginalization and conditioning is particularly simple. The Markov graph under conditioning on \( C=c \) is simply the subgraph of \( G \) without \( C \), \( H_{\text{Cond}(C=c)} = G_{C} \). Marginalization over \( C \), on the other hand, adds edges between variables \( A \) and \( B \) if \( A \) and \( B \) have neighbours in common, \( \text{ne}(A) \cap \text{ne}(B) \neq \emptyset \).

Notice finally, that Corollaries 1 and 2 only tell us what can be deduced about \( H_{\text{Marg}(C)} \) and \( H_{\text{Cond}(C=c)} \) from knowledge of the global Markov property of \( G \) and that conditional independence relationships that are not related in any way to the Markov properties of \( G \) may appear after both marginalization and conditioning. To identify such relationships one either has to analyze data or use arguments based on substance matter arguments.

5 Functional collapsibility in graphical models

Marginalizing in such a way that some of the independence properties of a graphical model reappear in marginal models is referred to as graphical collapsibility by Whittaker (1990, page 394). In high-dimensional datasets with a complicated independence structure, the data reduction implied by graphical collapsibility may not be enough to make the analysis feasible. The conditioning sets in the GMP hypotheses still contain too many variables. In such cases, the notion of functional collapsibility may be considered.

**Definition 3 (functional collapsibility).** Let \( X \) and \( Y \) be vectors of random variables distributed according to a CGM with Markov graph \( G_{XY} \), and let \( S \) be a function of \( Y \), \( S = f(Y) \). We say that the model is functionally collapsible onto \((X,S)\) if \( X \perp Y | S \).

The condition defining functional collapsibility states that knowing \( S \), reading \( Y \) is irrelevant for reading \( X \), or, in other words, that all information from \( Y \) on \( X \) is collected in \( S \). The question of the association between \( X \) and \( Y \) can therefore be addressed by an analysis of the association between \( X \) and \( S \) that ignores the response patterns among \( Y \) variables.

There are many natural examples of applications of functions of variables, where functional collapsibility is necessary. Scales summarizing responses to a set of items is one obvious situation. This situation is often discussed in item response theory, where \( S \) purports to be a proxy measure for a latent variable, but the notion of functional collapsibility is more general than that and is here introduced without reference to latent variables. Classification rules summarizing classifying
outcomes on a number of variables into a small set of nominal or ordinal classes is another. Finally, category collapsibility where the categories of a variable are collapsed onto a smaller set of categories is a special case relative to classification of outcomes to more than one variable, but is sometimes discussed as a special topic in connection with loglinear models for contingency tables (Jackson et al., 2008).

Lauritzen (1996, p. 29) refers to two results concerning conditional independence involving functions of variables that are useful in connection with a discussion of the implications of functional collapsibility. They are

(a) if \( X \perp Y \mid Z \) and \( U = f(X) \), then \( U \perp Y \mid Z \)
(b) if \( X \perp Y \mid Z \) and \( U = f(X) \), then \( X \perp Y \mid Z, U \)

If \( Y \) and \( X \) are vectors, it follows from (a) together with functional collapsibility, \( X \perp Y \mid S = f(Y) \), that \( X_i \perp Y_j \mid S \) for all pairs of variables consisting of one \( X \)-variable and one \( Y \)-variable. Tests of hypotheses of this kind are household tests of no differential item functioning in item response models (Holland and Thayer, 1988), but are here seen as a consequence of the more general notion of functional collapsibility.

The following Lemma 1 summarizes the consequences of functional collapsibility on the joint distributions of \((X,S,Y)\) and \((S,X)\)

**Lemma 1.** If \( X \perp Y \mid S = f(Y) \) and \( X_i \perp X_j \mid X_A, Y \) where \( X_A \subseteq X \setminus \{X_i, X_j\} \) then \( X_i \perp X_j \mid X_A, S \)

**Proof:** The conditional independence of \( X_i \) and \( X_j \) implies that

\[
P(X_i, X_j, X_A, Y) = \frac{P(X_i, X_A, Y)P(X_j, X_A, Y)}{P(X_A, Y)} = \frac{P(X_i, X_A, Y \mid S)P(S)P(X_j, X_A, Y \mid S)P(S)}{P(X_A, Y \mid S)P(S)}
\]

which may be rewritten as
because $X$ and $Y$ are conditionally independent given $S$ and $P(Y|S)P(S) = P(Y,S) = P(Y)$ since $S$ is a function of $Y$. Hence,

$$P(X_i, X_j, X_A, Y) = P(X_i | X_A, S)P(X_j | X_A, S)P(X_A | S)P(Y)$$

so that

$$P(X_i, X_j, X_A, S) = \sum_{Y,F(Y)=S} \left( P(X_i | X_A, S)P(X_j | X_A, S)P(X_A | S)P(Y) \right)$$

$$= P(X_i | X_A, S)P(X_j | X_A, S)P(X_A | S)P(S) = P(X_i | X_A, S)P(X_j | X_A, S)P(X_A, S)$$

From this it follows that

$$P(X_i, X_j | X_A, S) = P(X_i | X_A, S)P(X_j | X_A, S)$$

which was to be proven.

Lemma 1 implies that $X_i \perp X_j | X_A, S, Y_B$, where $Y_B \subseteq Y$ when $X \perp Y | S = f(Y)$ and $X_i \perp X_j | X_A, Y$. To see this, we just have to use Lemma 1 together with (B) and the function $g(Y) = (f(Y), Y_B)$.

**Proposition 2 (functional collapsibility).** If the distribution of $(X,Y)$ is functionally collapsible onto $(X,S)$ and if all variables of $Y$ belongs to one and the same recursive block then it follows that the joint distribution of $(X,S,Y)$ satisfies all the global Markov property implied by $G_{XSY}$ defined in the following way:

1) The recursive structure of variables in $G_{XSY}$ is the same as in $G_{XY}$ with $S$ in the same recursive block as $Y_1$ (assuming that $Y$ variables are ordered according to the recursive structure of the model).

2) The set of nodes for $(Y,S)$ is complete.

3) There is no edge between $X_i$ and $X_j$ in $G_{XSY}$ if there is no edge or arrow between the variables in $G_{XY}$

4) There is no edge between $X_i$ and $S$ in $G_{XSY}$ if there are no edges between $X_i$ and each of $Y_j \in Y$
Functional collapsibility consequently justifies describing $G_{XYS}$ as the Markov graph of $P(X,S,Y)$ even though the requirement of positive probabilities or densities of all combinations of outcomes of the variables for deriving the Global Markov property are violated by $P(X,S,Y)$.

**Proof:** To prove Proposition 2 we consider all minimal separating sets in $G^M_{XYS}$ for variables that are not connected by edges in $G^M_{XYS}$ and then prove that these variables are conditionally independent give the separating set and arbitrary subsets of additional variables from the model. It is convenient to consider three different situations. The proof in the first case is straightforward. The second and third case require some remarks on the difference between $G^M_{XY}$ and $G^M_{XYS}$ to make it apparent how the GMP hypotheses defined by $G^M_{XYS}$ may be derived from the global Markov properties of $G_{XY}$. These remarks follow after the discussion of the first case.

(i) All pairs of variables $(Y_i,X_j)$ are separated by $S$ in the $G_{XYS}$ and therefore also by $(S,X_A,Y_B)$ where $Y_A \subseteq Y \setminus Y_i$ and $X_B \subseteq X \setminus X_j$. Again, according to Lauritzen (1996, page 29), $X \perp Y|Z$ and $U=h(X)$ implies that $X \perp Y|(Z,U)$. Apply this relationship twice to show that $Y \perp X|S,Y_A,X_B$ from which it follows that $Y_i \perp X_i|S,Y_A,X_B$.

Assume next that there is no edge between $X_i$ and $X_j$ in $G^M_{XYS}$, and that a minimal separating set for the two variables has been identified. Since $Y$ is only connected to $X$ through $S$ it follows that the separating set either consists of nothing but $X$ variables, $X_C \subseteq X \setminus \{X_i,X_j\}$ or that it consists of $S$ and a subset of $X$ variables, $\{S,X_C\}$. In both cases, $X_C$ may of course be empty. The two different situations are treated as different cases below, even though the arguments are very similar.

To derive GMP hypotheses corresponding to separation properties of $G^M_{XYS}$ from the global Markov property of $G_{XY}$ we must compare separation in $G^M_{XYS}$ to the separation in $G^M_{XY}$. All variables belonging to $\text{pa}(S)$ in $G_{XYS}$ will be connected in $G^M_{XYS}$. If $Y$ contains more than one chain component, $G^M_{XY}$ may have fewer edges that $G^M_{XYS}$ since edges generated by moralization only adds edges between variables that are parents to nodes in the same chain component. From this it follows that all paths in $G^M_{XY}$ are also paths in $G^M_{XYS}$. If $X_C$ separates $X_i$ and $X_j$ in $G^M_{XYS}$, $X_C$ will also be a separating set in $G^M_{XY}$. This leads directly to the first of the two remaining cases.
(ii) Assume, that \((X_i,X_j)\) are separated by a minimal separating set, \(X_C \subseteq X \setminus \{X_i,X_j\}\) in \(G_{XYS}\), that \(X_A \subseteq X \setminus \{X_i,X_j,X_C\}\) and that \(Y_B \subseteq Y\). From the arguments above, \(X_C\) separate \((X_i,X_j)\) in \(G_{XY}^M\). For this reason, \(X_C,X_A,Y\) also separate \((X_i,X_j)\) in \(G_{XY}^M\). It therefore follows from the global Markov properties that \(X_i \perp X_j \mid X_C,X_A,Y\). From this and Proposition 2 it follows that \(X_i \perp X_j \mid X_C,X_A,S,Y_B\).

(iii) Assume finally, that a minimal separating set for \((X_i,X_j)\) in \(G_{XYS}^M\) contains \(S\) in addition to \(X_C\). Since \(S\) is included in a minimal separating set, there must be at least one path from \(X_j\) to \(X_j\) passing through \(S\). This path contains two nodes directly connected by edges to \(S\) in \(G_{XYS}^M\) and therefore also – according to the definition of \(G_{XYS}\) – a subset of nodes, \(Y_D \subseteq Y\), in \(G_{XY}^M\). From this it follows that \((X_i,X_j)\) are separated by \(\{X_C,Y_D\}\) and therefore also by \(\{X_A,X_C,Y\}\) in \(G_{XY}^M\). From this and Proposition 2 it once again follows that \(X_i \perp X_j \mid X_C,X_A,S,Y_B\).

Note that Proposition 2 also holds for all truncated versions of the CGM down to and including the truncated model in which the \(Y\)-variables belong to the ultimate set of response variables of the model.

Proposition 2 implies that \(G_{XYS}\) is a Markov graph for \(P(X,Y,S)\) and that all global Markov properties hold under functional collapsibility even though the chain component containing \((Y,S)\) contains structural zeroes and that \(G_{XS} = \text{Subgraph}(\{X,S\},G_{XSY})\) is a Markov graph for \(P(X,S)\). The following corollary, which is a trivial consequence of Propositions 2 and 3 implies that there will be no edge connecting \(X_i\) to \(S\) in \(G_{XS}\) if there are no edges connecting \(X_i\) to \(Y\)-variables in \(G_{XY}\).

**Corollary 3** If \((X,Y)\) is functionally collapsible onto \((X,S)\) it follows that

1) If \(X_i \perp X_j \mid (X,Y)\) then \(X_i \perp X_j \mid (X,S)\)

2) If \(X_i \perp Y_j \mid (X,Y)\) for all \(Y_j \in Y\) then \(X_i \perp S \mid X \setminus X_i\)

### 6 INFERENCE IN CHAIN GRAPH MODELS

With categorical data, it is straightforward to calculate overall deviances comparing observed to expected counts for both the graphical regression models of \(P(V_i \mid V_i,..,V_k)\) and for the recursive loglinear chain graph model as a whole (Fienberg,1979) where the overall deviance is equal to the sum of deviances for the different regression models. Such deviances are, however, not useful for
model checking because the overall evidence against a high-dimensional model is too imprecise to pinpoint what is wrong with the model. Further, the calculation of the number of estimable parameters and the degrees of freedom for the deviances is not straightforward. Even if calculated, the asymptotic approximation of the exact distribution of the deviance for large sparse tables is poor, rendering the result useless.

Directly addressing the different independence assumptions underlying the model by tests of conditional independence is a much better approach since such tests disclose accurate evidence of model problems and have exact distributions that may easily be approximated by standard Monte Carlo procedures (Kreiner, 1987). The one persevering problem in this approach, is that the power of conventional chi squared and likelihood ratio tests of conditional independence of pairs of variables given a large conditioning set of variables is poor. Instead, we suggest that tests of minimal or smallest GMP hypotheses is a better approach, even though the evidence against the model may be less accurate than evidence against the assumptions of pairwise conditional independence defining the model.

The GMP hypotheses for this purpose may be derived in two different ways: from the different graphical regression models comprising the chain graph models or from the complete and truncated chain graph models. Deriving GMP hypotheses from the regression models is easier, if one does not have access to specialized programs addressing separation in moralized graphs, but the GMP hypotheses based on regression models often have larger conditioning sets than GMP hypotheses derived from moralized chain graphs. The example in Section 7 below illustrates these points.

7 AN EXAMPLE: a longitudinal study of living conditions in Denmark

The data used for this example originated in a longitudinal study where a representative sample of children in the 8th grade of the Danish public school in 1968 was followed for 25 years. We will use results from an analysis of this data set to investigate the claims of Herrnstein and Murray (1994) that intelligence is the most important factor of importance for success in life. Their claim were founded on a large number of logistic regression analyses with a number of dichotomized success criteria where the effect of intelligence on late success in life for white American non-Latino males was controlled for age and Socio Economic Status (SES).

The major difference between our analysis by CGMs and the HM analysis is that we consider the joint distribution of success criteria instead of addressing the dependence of the various success
criteria on intelligence and parental SES in separate analyses. Another difference is in the way that information on education is included in the analyses. The HM analysis considers the effect of intelligence on outcome variables in three different ways: 1) marginally without consideration of the education, 2) conditionally given an education below college level, and 3) conditionally given a college education. In our setting, using CGMs, education is included in a joint analysis allowing for a parsimonious model from which the results from the different HM approaches can be discussed.

The analysis performed here illustrates that it is possible to distinguish between the effect of education and the effect of intelligence. However, we will only use the results on marginalization and conditioning in CGMs to be able to see what the model without information on education would tell us about the effect of intelligence on the response variables, what the conditional model for subjects with a specific educational level should be would tell us, and to consider a model which is functionally collapsed over the information on education onto a variable classifying education into persons with or without a college education and what that model would tell us about the effect of intelligence.

The model considered here contains eleven variables organized in 6 recursive blocks:

Block 1: SRH: Self reported health (5 categories)
Block 2: ChronDis: One or more chronic diseases (2 categories)
         Unempl: Information on experience of unemployment in the past five years (2 categories)
Block 3: VocEduc: Attained level of academic or vocational education (5 categories)
Block 4: School: Attained level of school education (4 categories)
Block 5: Intellig: Intelligence measured at the beginning of the 8th year in school (5 categories)
Block 6: FamSES: Parental SES (5 categories)
         FamEDUC: Parental level of education (6 categories)
         Urbaniza: Urbanization during childhood (3 categories)
         Sex: Gender of the individual

Figure 3 shows the Markov graph of a model obtained by a two-step model search procedure (Kreiner, 1986 and 2003 and Klein et.al., 1995) consisting of an opening screening for a initial chain graph followed by standard stepwise model search procedures using tests of GMP hypotheses. According to this model, intelligence during adolescence has a direct effect on income 25 years
later in addition to an indirect effect mediated through education. However, the evidence of the
direct effect is not overwhelming and depends to a large degree on the way the test is performed.
The purpose of the example therefore is to take a closer look at the analysis of this effect.

Data on Income and Intelligence is available for 2342 persons. Intelligence and Income is
marginally associated (Goodman and Kruskal’s $\gamma = 0.20$, p < 0.0005). However, from the point of
the CGM the hypothesis to address is that Intelligence and Income are conditional independent,
H0: Income $\perp$ Intellig | SRH, ChronDis, Unempl, VocEduc, School, FamSES, FamEDUC, Urbaniza, Sex

Testing this hypothesis is possible even though the table is both extremely large and extremely
sparse. The hypothesis is comfortably accepted both by the $\chi^2$ test ($\chi^2 = 546.7$, df = 406, p = 0.782)
and by the partial $\gamma$ coefficient ($\gamma = 0.07$, p = 0.313). Compared to the number of strata in the 11-
dimensional table, the degrees of freedom is very small due to the fact that there are no degrees of
freedom at all in the majority of strata. Since data in strata with df = 0 cannot be included, the test
of H0 suffers from a considerable loss of power. The acceptance of H0 may therefore not be the
last to be said about the direct effect of intelligence on income.

7.1 GMP hypotheses
We consider now the GMP hypotheses motivated by the independence structure of the chain graph
in Figure 3. The local GMP hypotheses defined by the neighbours and parents of Income may be
read directly off Figure 3:

GMP1: Income $\perp$ Intelligence | ChronDis, Unempl, VocEduc, School, Sex

A test of GMP1 also accepts conditional independence ($\chi^2 = 1174.2$, df = 1040, p = 0.297, $\gamma = 0.07$,
p = 0.064).

To identify other GMP hypotheses we turn to the moralized graph under H0 (without an edge
between intelligence and income) shown in Figure 4. In this graph it is easily seen that the other
local GMP hypothesis is

GMP2: Income $\perp$ Intelligence | VocEduc, School, FamSES, FamEDUC, Sex
GMP2 is nominally rejected by the partial $\gamma$ coefficient, but not by the $\chi^2$ test ($\chi^2 = 1118.9$, df = 1004, p = 0.918, $\gamma = 0.09$, p = 0.020). After adjusting for multiple testing by the Benjamini-Hochberg procedure, the single significant test result can not be regarded as convincing evidence against conditional independence of Intelligence and Income.

GMP1 and GMP2 both have conditioning sets with five variables. Neither of the hypotheses are minimal GMP hypotheses. A final test of H0 therefore considers the smallest possible GMP hypothesis identified by the moral graph, Figure 4. In this graph, Income and Intelligence are separated by VocEduc, School and Sex. The smallest possible GMP hypothesis defined by the global Markov property of the model defined by Figure 3 together with H0 is therefore

GMP3: Income $\perp$ Intelligence $|$ VocEduc, School, Sex

This hypothesis is also accepted by $\chi^2$ but rejected by the partial $\gamma$ ($\chi^2 = 575.7$, df = 548, p = 0.375, $\gamma = 0.08$, p = 0.009), which, after Benjamini-Hochberg adjustment for multiple testing, provides evidence of a weak, but significant, direct effect of intelligence on income.

Figure 3. The CGM for inference about the effect of the intelligence measured in 1968 on subsequent 1992 income. The evidence of a direct effect of intelligence on Income after controlling for relevant confounders and intermediate variables is weak.
Marginalization over education

Marginalizing over VocEduc and School not only adds an edge between Income and Intelligence, but also a number of other edges and arrows as shown in Figure 5. This marginal distribution can therefore not be used for a check of the missing direct effect of Intelligence on Income in a frame of inference defined by the complete CGM in Figure 3.

The fact that conditional independence of intelligence and income in the marginal model cannot be inferred from the complete model does not imply that the two variables are conditionally associated. This will have to be tested specifically by means of the GMP hypothesis in the marginal model framework. We therefore include a test of conditional independence in the marginal model. To derive GMP hypotheses of conditional independence of Income and Intelligence we therefore remove the edge between the two variables in Figure 5 and then moralize the graph. The smallest GMP hypothesis under conditional independence in the marginal model is

\[
\text{GMP4: Income} \perp \text{Intelligence} \mid \text{SRH, ChronDis, Unempl, FamSES, FamEduc, Sex}
\]

This hypothesis is once again accepted by the $\chi^2$ test but strongly rejected by the partial $\gamma$ coefficient ($\chi^2 = 1193.9$, df = 1061, $p = 0.837$, $\gamma = 0.19$, $p < 0.0001$). So without taking into account the attained educational level, the test documents an indisputable (although moderate) effect of
Intelligence on Income. The difference in interpretation between this result and the previous, taking into education into account, is noticeable.

### 7.4 Conditioning on education

Figure 6 shows the Markov graph for the conditional distribution of variables given 10 years of public schooling followed by three years of vocational training. Conditioning on educational variables generates selection bias among background variables and intelligence. Moralization of the graph does not add edges to the structure, and it is easily seen that the smallest GMP hypothesis is Income $\perp$ Intelligence | Sex. A test of this hypothesis supports the existence of a direct effect of intelligence on Income ($\chi^2 = 48.9$, df = 48, $p = 0.443$, $\gamma = 0.10$, $p < 0.023$).

**Figure 5.** The Markov graph of the marginal model after collapse over educational variables. Bold edges have been added by marginalization.
7.5 Functional collapsibility on College education

The variable measuring school education distinguishes between persons with less than nine years, nine years, ten years and 12 years in the Danish public school, where 12 years is the number of years required to finish high school. The variable measuring vocational education distinguishes between no formal education after the public school, and three, four, five or more years of formal education or training. However, the question is whether this detailed description of educational level provides information of any impact on the analysis of the effect of intelligence above or beyond the information provided by a more compact classification according to whether or not the person has concluded a college education. In the HM analyses education was included, distinguishing between males with or without a college education. To conclude the example in this paper, we therefore examine whether the model is functionally collapsible over the educational variables onto a classification into two classes roughly similar to the College variable of HM requiring a high school education followed by four or more years of formal education to be counted as a college education. According to this rule 16.6% of the persons included in the survey finished a college education in the period from 1968 to 1992. The CGM assuming functional collapsibility – that is the $G_{XYS}$ graph defined in Proposition 2 – is shown in Figure 7. The smallest GMP hypothesis of conditional independence of Intelligence and Income is
GMP5: Income ⊥ Intelligence | College, Sex

GMP5 is clearly rejected by both the $\chi^2$ test and the partial $\gamma$ ($\chi^2 = 158.9$, df = 96, $p = 0.002$, $\gamma = 0.15$, $p = 0.001$).

The relatively large difference between the $\gamma$ coefficient calculated under functional collapsibility and the $\gamma$ coefficient calculated for the test of GMP3 hypothesis is a result of residual confounding if the model is not functionally collapsible unto College. Before we take the test results at face value we therefore test the GMP hypotheses of conditional independence of the two educational variables and the remaining variables of the model given the College variable. These tests that are summarized in Tables 1 and 2 provide overwhelming evidence against functional collapsibility. For this reason, the evidence against GMP5 is not counted as evidence against H0.

**Figure 7.** The chain graph model assuming functional collapsibility onto a binary College variable. College is a function of the educational variables. Hence edges between College and these variable has been drawn as bold lines.
Table 1. Tests of conditional independence of VocEduc and other variables given College. Exact conditional p-values are estimated by Monte Carlo methods using random samples of 10,000 tables for each test

<table>
<thead>
<tr>
<th>Variables</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>asymp</th>
<th>exact</th>
<th>gamma</th>
<th>asymp</th>
<th>exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>D: Income</td>
<td>132.4</td>
<td>24</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.24</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>C: SRH</td>
<td>18.9</td>
<td>12</td>
<td>0.092</td>
<td>0.093</td>
<td>0.11</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>A: ChronDis</td>
<td>15.0</td>
<td>4</td>
<td>0.005</td>
<td>0.005</td>
<td>0.15</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>B: Unempl</td>
<td>77.0</td>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.33</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>I: Intellig</td>
<td>145.5</td>
<td>16</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.30</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>J: Urbaniza</td>
<td>16.1</td>
<td>12</td>
<td>0.189</td>
<td>0.186</td>
<td>0.08</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>K: FamSES</td>
<td>114.5</td>
<td>16</td>
<td>0.000</td>
<td>0.001</td>
<td>0.26</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>L: FamEduc</td>
<td>71.5</td>
<td>20</td>
<td>0.000</td>
<td>0.001</td>
<td>0.21</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>M: Sex</td>
<td>41.3</td>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.06</td>
<td>0.132</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 2. Tests of conditional independence of School education and other variables given College. Exact conditional p-values are estimated by Monte Carlo methods using random samples of 10,000 tables for each test

<table>
<thead>
<tr>
<th>Variables</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>asymp</th>
<th>exact</th>
<th>gamma</th>
<th>asymp</th>
<th>exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>D: Income</td>
<td>24.7</td>
<td>18</td>
<td>0.132</td>
<td>0.129</td>
<td>0.08</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>C: SRH</td>
<td>16.3</td>
<td>9</td>
<td>0.062</td>
<td>0.062</td>
<td>-0.11</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>A: ChronDis</td>
<td>11.9</td>
<td>3</td>
<td>0.008</td>
<td>0.008</td>
<td>-0.07</td>
<td>0.092</td>
<td>0.092</td>
</tr>
<tr>
<td>B: Unempl</td>
<td>28.8</td>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.17</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>I: Intellig</td>
<td>343.5</td>
<td>12</td>
<td>0.000</td>
<td>0.001</td>
<td>0.44</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>J: Urbaniza</td>
<td>34.8</td>
<td>9</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.15</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>K: FamSES</td>
<td>213.8</td>
<td>12</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.36</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>L: FamEduc</td>
<td>192.6</td>
<td>15</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.38</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>M: Sex</td>
<td>68.1</td>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.28</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

8 DISCUSSION

Despite the fact that the family of CGMs has been known for more than 25 years – Wermuth and Lauritzen (1983) is an early reference – relatively little has been written on statistical inference in such models. This paper attempts to rectify this.

So what is new in this paper?

First, the definition of CGMs differs in a minor and perhaps inconsequential way from the way the models are defined in most texts on this topic. To us, a CGM being a statistical model is defined by a recursive structure and conditional independencies. The model is not defined by the Markov graph only since several equivalent CGMs share the same Markov graph. Instead, we see the Markov graph as a second order mathematical model; as a model of the probabilistic model that captures some, but not all important features of the statistical model. The point may appear to be
moot since all equivalent models share the same joint distribution, but it becomes important the moment we have to modify the model in terms of evidence against some of the independence assumptions of the model. Consider, for instance the models defined by Figure 3. According to these models A and D are marginally independent. If a test of $A \perp D$ rejects the hypothesis we have to change the model. The most obvious way to do this would be to add an edge between A and D to the model. Should this edge be undirected, directed from A to D, or directed from D to A. A less obvious way to change the model would be to add edges between other variables such that $A \perp D$ can be regarded as a GMP hypothesis. And again, should the edges be undirected or directed?

Second, the notions of GMP hypotheses and the notions of truncated and tapered CGMs are new, but it can be claimed that they are already inherent parts of the theory of CGMs and that we are just developing a little bit of terminology. They have nevertheless been included here, since it is convenient to have names for the specific types of hypotheses if you want to talk about them and since the notion of truncated and tapered models is practical, for a systematic approach to derivation of GMP hypotheses. The results on marginalization and conditioning in CGMs are, to our knowledge, new in this paper. Again, one may claim that they are inherent in the definition of and results on global Markov properties of CGMs, but they need to be stated, if it is correct that they have not been described elsewhere. Finally, the notion of functional collapsibility is new to this paper. It is related to the theory of graphical IRT models and graphical Rasch models (Kreiner and Christensen, 2002, 2004, 2006, 2008). A graphical IRT model is a chain graph model with a latent variable and items depending on the latent variable. A graphical Rasch models is a graphical IRT model which is functionally collapsible onto the sum of item responses, where the graph defined in Proposition 2 is referred to as a Rasch graph. The notion of functional collapsibility is more general than that, however. Functional collapsibility does not require latent variables and applies to all types of functions of variables. The role of the global Markov properties during item analysis by graphical Rasch models has been discussed in Kreiner and Christensen (2008).

Deriving GMP hypotheses from regression graphs is not difficult if the number of variables in the recursive blocks is limited. These hypotheses may not be the smallest or even minimal GMP hypotheses that may be derived from the moralized graphs, such that the tests may lose power and, in the worst case, even be infeasible. Determining separation properties in high-dimensional graphical models is often difficult, requiring specialized software to be identified. In Section 6, the program SCD/DIGRAM (Kreiner, 2003) was used both for the analysis of the graphs and for the analysis of data.
The purpose of the paper was to illustrate the way GMP hypotheses of conditional independence in chain graph models may be derived and used during analysis of high-dimensional data. For this reason, we have not addressed a number of other important issues related to inference in CGMs, e.g. the relationships between model estimates obtained by exploratory model search on one hand and estimates and tests for particular important relationships among variables on the other. We recognize that these issues are important and that a lot remains to be discussed, but abstain from doing so in order to focus on the role the global Markov properties of the models play during the analysis.

REFERENCES


Klein, J. Keiding, N. Kreiner S (1995) Graphical Models for Panel Studies, Illustrated on Data from the Framingham Heart Study, Statistics in Medicine, 14, 1265-1290


08/01 Siersma, V. & Kreiner, S. Monte Carlo Evaluation of Model Search in Graphical Models for Ordinal Data.

08/02 Andersen, P.K. & Pohar Perme, M. Inference for outcome probabilities in multi-state models.

08/03 Scheike, T.H. & Zhang, M. Flexible competing risks regression modelling and goodness-of-fit.


08/05 Gerds, T.A., Qvist, V., Strub, J.R. & Keiding, N. Survival analysis in dental research: The typical observational patterns.

08/06 Carstensen. B. Limits of agreement: How to use the regression of differences on averages.


08/08 Martinussen, T. & Scheike, T.H.. Covariate selection for the semiparametric additive risk model.


08/10 Martinussen, T. & Scheike, T.H. The Aalen additive hazards model with high-dimensional regressors.


08/12 Andersen, P.K., Pohar Perme, M. Pseudo-observations in survival analysis.
Frydman, H., Gerds, T., Groen, R., Keiding, N. Non-parametric estimation in an “illness-death” model when the transition times are interval-censored and one transition is not observed.

Didelez, V., Kreiner, S., Keiding, N. On the use of graphical models for inference under outcome dependent sampling.

Scheike, T.H., Martinussen, T., Silver, J. Estimating haplotype effects for survival data.


Nielsen, L.H., Keiding, N. Describing episodes of drug treatment from joint observation of a prescription registry and a cross-sectional survey.

Nielsen, L.H., Løkkegaard, E., Keiding, N. Importance of effect measures in the evaluation of the effect of post menopausal hormone therapy on myocardial infarction.

Cortese, G. & Andersen, P.K. Competing risks and time-dependent covariates.

Kreiner, S. & Christensen, K.B. Item screening in graphical loglinear Rasch models.

Andersen, P.K. & Væth, M. Survival analysis.

Gerster, M., Ejrnæs, M. & Keiding, N. The effect of educational attainment on ultimate fertility – does feedback play a role?

Kreiner, S., Petersen, J.H. & Siersma, V. Deriving and testing hypotheses in chain graph models.