The Chain Graph Markov Property

Author(s): Morten Frydenberg

Source: *Scandinavian Journal of Statistics*, Vol. 17, No. 4 (1990), pp. 333-353
Published by: Wiley on behalf of Board of the Foundation of the Scandinavian Journal of Statistics
Stable URL: http://www.jstor.org/stable/4616181

Scand J Statist 17: 333–353, 1990

The Chain Graph Markov Property

MORTEN FRYDENBERG
Århus University

ABSTRACT. A new class of graphs, chain graphs, suitable for modelling conditional independencies are introduced and their Markov properties investigated. This class of graphs, which includes the undirected and directed acyclic graphs, which includes the undirected and directed acyclic graphs, enables modelling recursive models with multivariate response variables. Results concerning the equivalence of different definitions of their Markov properties including a factorization of the density are shown. Finally we give a necessary and sufficient condition for two chain graphs to have the same Markov properties.

Key words: block-recursive models, causal models, conditional independence, Gibbs-factorization, graphical models, influence diagrams.

Introduction

The motivation for the present paper is the introduction of the graphical chain models by Lauritzen & Wermuth (1989).

A graphical model for a set of variables can shortly be defined as a model where the conditional independencies specified by the model are given by a graph. In the set-up of Lauritzen & Wermuth the variables are first divided into disjoint sets, which are ordered into a dependence chain in such a way that the resulting ordering reflects the known causal structure of the problem. The graph has the variables as vertices and edges are drawn after the following two rules: (i) Two variables are not connected by an edge if they are conditionally independent given all the other variables that do not come after both of the variables in the dependence chain. (ii) If two variables are to be connected then they are connected with an undirected edge if they belong to the same set in the dependence chain and with an arrow pointing from the first to the last if this is not the case. So in general the resulting graph will contain both undirected and directed edges and the missing edges will imply certain conditional independencies between pairs of variables. For a review of the use and ideas behind the graphical models the reader is referred to Wermuth & Lauritzen (1990).

Kiiveri, Speed & Carlin (1984) studied the Markov properties of recursive causal graphs. The causal models have the characteristics that they only allow one-dimensional response variables in each step of the recursive factorization of the model. In terms of the graph the vertex set is partitioned into two sets, a set of exogenous and a set of endogenous vertices, in such a way that all edges between two exogenous vertices are undirected, all edges between an exogenous and an endogenous vertex are directed towards the endogenous and all edges connecting two endogenous vertices are directed. In other words the causal models do not allow symmetric interaction between response variables. This is unfortunate, for example when groups of responses occur simultaneously.

The graphical association models do allow such simultaneous response variables, so their Markov properties are not covered by the results in Kiiveri, Speed & Carlin or in any other part of the literature about Markov random fields. As a complete description of the Markov properties is fundamental for the understanding, manipulation and interpretation of these models, new theory is needed. The purpose of the present paper is therefore to develop this theory for "block-recursive Markov random fields".

It turns out that the Markov properties of the graphical chain models are given as the

Markov properties of a special class of graphs, the chain graphs, which includes the undirected graphs, the directed acyclic graphs and the recursive causal graphs as special cases.

The structure of the paper is as follows.

The concept of chain graphs and the graph theoretical terminology are given in section 1. Section 2 gives a short review of Markov properties of undirected graphs. In the next section we show that, in analogy with the undirected case, it is possible to define a pairwise, a local and a global Markov property for chain graphs which are equivalent under similar assumptions as in the undirected case. Further we show that the chain Markov property used here is equivalent to the chain Markov property defined in Lauritzen & Wermuth (1989), implying that the Markov properties for the graphical chain models are covered by our results. A factorization result analogous to the Gibbs=Markov result for undirected graphs, see e.g. Speed (1979), is proved in section 4. As it is possible to specify the same Markov properties with different chain graphs it is of importance to know when two different chain graphs have the same Markov properties. This problem is treated in section 5, where we give a necessary and sufficient condition for two chain graphs to have the same Markov properties. Finally, the last section contains a short discussion of some loose ends and the relation to some recent results.

1. The chain graphs

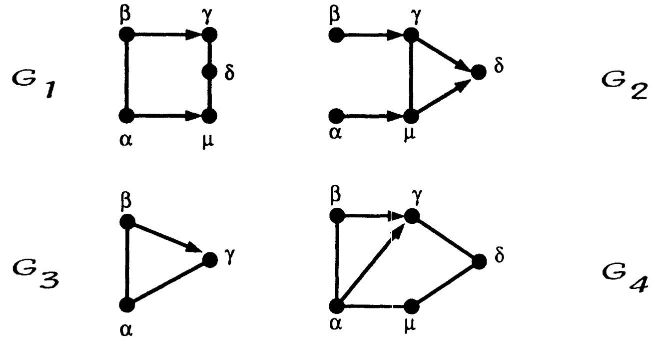
In this section we will introduce the concept of chain graphs and mention some of the properties which are needed in the following sections. A more detailed discussion of chain graphs is given in Frydenberg (1989).

By a *graph* we will understand a pair $G=(V,E)$, where V is a finite set of *vertices* and E is a set of *edges*, i.e. a subset of $E^*(V)=\{(\alpha,\beta)\,|\,\alpha\in V, \beta\in V \text{ and } \alpha\neq\beta\}$ the set of ordered pairs of distinct vertices. We will write $\alpha\rightarrow\beta$ if $(\alpha,\beta)\in E$, $\alpha\leftarrow\beta$ if $(\beta,\alpha)\in E$, $\alpha\leftrightarrow\beta$ if at least one of them holds, $\alpha\Leftrightarrow\beta$ if both are true and $\alpha=\beta$ if only the first is true. When drawing the graph we connect α to β with an arrow from α to β if $\alpha\Rightarrow\beta$ and with a line-segment if $\alpha\Leftrightarrow\beta$. If A is a subset of vertices the *induced subgraph* $G_A=(A,E_A)$ is given by $E_A=E\cap A\times A$ and we call A *complete* if $\alpha\leftrightarrow\beta$ for all pairs in A. The graph is called *undirected* if it does not contain any arrows or equivalently if $\alpha\rightarrow\beta$ implies $\alpha\leftarrow\beta$ and for any graph G we will denote the *underlying undirected graph* by $G^u=(V,E^u)$, i.e. $E^u=\{(\alpha,\beta)\,|\,(\alpha,\beta)\in E \text{ or } (\beta,\alpha)\in E\}$. Now for a set of vertices A we let $bd(A)=\{\beta\in V\backslash A\,|\,\beta\rightarrow\alpha \text{ for some } \alpha \text{ in } A\}$ and $cl(A)=bd(A)\cup A$ be respectively the *boundary* and the *closure* of A (note that the notation here differs somewhat from the one used in Lauritzen *et al.* (1990), where the boundary only consists of the vertices connected to A with undirected edges).

An ordered n-tuple $(\alpha_1,\ldots,\alpha_n)$ of distinct vertices is called a *path from* α_1 to α_n if $\alpha_i\rightarrow\alpha_{i+1}$ for $i=1,\ldots,(n-1)$. If $\alpha_i\Rightarrow\alpha_{i+1}$ for some i we call the path *directed* and if this is not the case, i.e. if $\alpha_i\Leftrightarrow\alpha_{i+1}$ for all i, we call it *undirected*. Finally, a *cycle* is a path with the modification that $\alpha_n=\alpha_1$.

In this set-up a graph is called a *chain graph* if it does not contain any directed cycles, i.e. if we start at a vertex α and move along in the graph respecting the directions of the arrows, then we cannot come back to α if we have passed an arrow. So the graphs G_1 and G_2 below are chain graphs, while G_3 and G_4 are not chain graphs, because $(\beta,\gamma,\alpha,\beta)$ is a directed cycle in G_3 and for example $(\alpha,\gamma,\delta,\mu,\alpha)$ is a directed cycle in G_4.

For a further characterization of the chain graph we need a bit more notation. For a pair α and β we will write $\alpha\leqslant\beta$ if there exist a path from α to β or $\alpha=\beta$ and $\alpha\sim\beta$ if both $\alpha\leqslant\beta$ and $\beta\leqslant\alpha$. It is easily seen that $\sim$ is a equivalence-relation on V and if G is a chain graph we will denote the induced set of equivalence classes by $\mathscr{T}(G)$, the *chain components*, and the unique chain



component containing α by $\tau(\alpha)$. An equivalent definition of the chain components of a chain graph is that they are the connected components in the graph with all arrows removed. Letting $\alpha<\beta$ stand for the statement "there exist a directed path from α to β", i.e. $\alpha<\beta$ if and only if $\alpha\leqslant\beta$ but not $\beta\leqslant\alpha$, we let the *future* of α be given by $\phi(\alpha)=\{\beta|\alpha<\beta\}$ and the *past* $\pi(\alpha)=\{\beta|\beta<\alpha\}$. If τ is a chain component we see that the future or the past for all the vertices in τ is the same so we can use the notation $\phi(\tau)$ and $\pi(\tau)$ for their common future and past. Further we call τ *terminal* if $\phi(\tau)$ is empty and *initial* if $\pi(\tau)$ is empty.

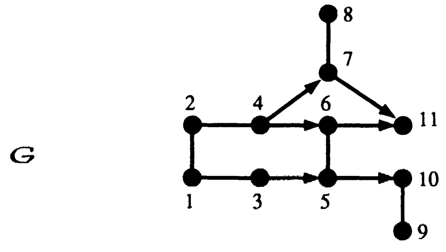
As special classes of chain graphs we note the undirected graphs, the directed acyclic graphs, which are chain graphs with chain components of size one, and the recursive causal graphs (see Kiiveri, Speed & Carlin (1984)) in which all chain components, which have size larger than one, are initial in the sense above.

It is fairly easy to see that the chain graphs have properties similar to the directed acyclic graphs, such as: any subgraph of a chain graph is again a chain graph and a chain graph has at least one terminal chain component. Further we have that the chain graph we get by deleting any chain component τ, has the same chain components as the original graph leaving out τ, and if τ is terminal in the original graph then the boundary of any subset in the new graph is the same as in the old graph and the future of any vertex is equal to the old future again leaving out τ.

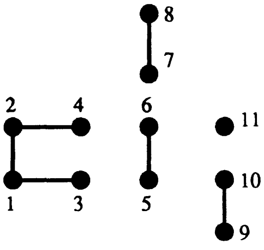
A chain graph can be generated from any undirected graph in the following simple way. Let $G=(V,E)$ be an undirected graph and $\mathscr{D}=\{V(1),V(2),\ldots,V(K)\}$ be a so-called *dependence chain* i.e. an ordered disjoint partitioning of V. We can then construct the graph $G^{\mathscr{D}}=(V,E^{\mathscr{D}})$ given by $E^{\mathscr{D}}=\{(\alpha,\beta)|(\alpha,\beta)\in E,\ \alpha\in V(i),\ \beta\in V(j)$ and $i\leqslant j\}$, i.e. the graph where we substitute a line between α and β in G with an arrow from α to β if α comes before β in the ordering. It is obvious from the construction of $G^{\mathscr{D}}$ that it cannot have any directed cycles and therefore it is a chain graph. In fact it is possible to generate all chain graphs in this way.

Before we turn to an example we need two more concepts. We will call a subset of vertices A an *anterior set* if it can be generated by stepwise removal of terminal chain components. The anterior sets are a generalization of the *ancestral sets*, that is the sets which have no boundary, as defined for directed acyclic graphs in Lauritzen *et al.* (1990). And finally we will define the *moral graph*, $G^m=(V,E^m)$, generated from G by $E^m=E^u\cup_{\tau\in\mathcal{T}(G)}E^*\{bd(\tau)\}$, i.e. the underlying undirected graph, where the boundary w.r.t. G of every chain component is made complete. The moral graphs are a natural generalization to chain graphs of the similar concept for directed acyclic graphs given in Lauritzen & Spiegelhalter (1988) and Lauritzen *et al.* (1990).

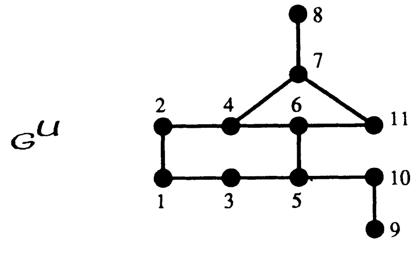
Example. In this example we will illustrate some of the concepts introduced above.
Consider the graph



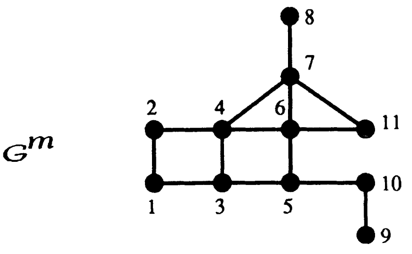
An inspection shows that the graph does not have any directed cycles hence it is a chain graph. If we delete all the arrows we get the graph



from this we see that the chain components are $\mathcal{T}(G)=\{\{1,2,3,4\}, \{5,6\}, \{7,8\}, \{9,10\}, \{11\}\}$ of which $\{9,10\}$ and $\{11\}$ are terminal. The generated undirected and moral graphs are

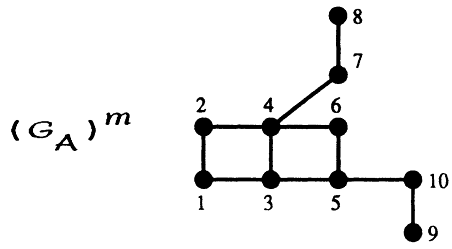


and



where the edges between 3 and 4 and 6 and 7 are added in G^m because $\{3,4\}\subseteq bd(\{5,6\})$ and $\{6,7\}\subseteq bd(\{11\})$. If we remove the terminal chain component $\{11\}$ we get the anterior set

$A = V \backslash \{11\}$ that has the moral graph



and we note that this has fewer edges than $(G^m)_A$.

Returning to G, we see that $\phi(5) = \{9, 10, 11\}$ as 5 only is connected to 9, 10 and 11 by directed paths (note that $6 \notin \phi(5)$ because the path between 5 and 6 is undirected). Finally we see that $B = \{1, 2, 3, 4, 5, 6, 9, 10\}$ is an anterior set, as it can be generated by first removing the terminal chain component $\{11\}$ and then removing the chain component $\{7, 8\}$, which is terminal in $G_{V \backslash \{11\}}$. □

2. The Markov properties for undirected graphs

Before we start investigating the Markov properties of the chain graphs we will give a short review of the Markov properties for the undirected graphs. In the rest of this work P will denote a probability measure on some product space $\mathscr{X} = \times_{\alpha \in V} \mathscr{X}_\alpha$. An element in $\mathscr{X}$ will be denoted by $x = (x_\alpha;\ \alpha \in V)$ and for a subset of vertices A, P_A is the marginal measure on $\mathscr{X}_A = \times_{\alpha \in A} \mathscr{X}_\alpha$ and $x_A = (x_\alpha;\ \alpha \in A)$. For three disjoint subsets A, B and C of V we write $A \perp B \mid C[P]$ if X_A and X_B are conditionally independent given X_C under P. A discussion of conditional independence can be found in Dawid (1980), where it is shown that the following holds for any probability measure:

(CI1) $A \perp B \mid C[P]$ implies $B \perp A \mid C[P]$
(CI2) $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid D[P]$
(CI3) $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid C \cup D[P]$
(CI4) ($A \perp B \mid D[P]$ and $A \perp C \mid D \cup B[P]$) implies $A \perp B \cup C \mid D[P]$

where A, B, C are D are disjoint sets.

In the literature there have been slightly different ways of naming the Markov properties for undirected graphs. Here we use the following notation. If G is an undirected graph then P is said to be:

(UP) *Pairwise G-Markovian* if $\alpha \perp \beta \mid V \backslash \{\alpha, \beta\}[P]$ whenever α and β are not adjacent
(UL) *Local G-Markovian* if $\alpha \perp V \backslash cl(\alpha) \mid bd(\alpha)[P]$ for all α
(UG) *Global G-Markovian* if $A \perp B \mid C[P]$ whenever C separates A and B,

where everything is w.r.t. G. And we say that α and β are *adjacent* if $\alpha \leftrightarrow \beta$ and C *separates* A and B if every path between A and B intersects C. As $bd(\alpha)$ separates α and $V \backslash cl(\alpha)$ we see that (UG) implies (UL), which again by (CI3) implies (UP). But none of the reverse implications hold without some regularity assumptions about P.

Several writers have shown that (UP), (UL) and (UG) all are equivalent if P is assumed to have a strictly positive density, p, w.r.t. some product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on $\mathscr{X}$.

If this is the case, i.e. P has a positive density, then P will fulfil:

(CI5) $(A \perp B \mid D \cup C[P]$ and $A \perp C \mid D \cup B[P])$ implies $A \perp B \cup C \mid D[P]$

The following elegant result, due to Pearl & Paz (1986), states that in fact suffices to assume that (CI5) holds

Theorem
(Pearl & Paz (1986).) *If P fulfils* (CI5) *and G is an undirected graph then the conditions* (UP), (UL) *and* (UG) *are equivalent and P is said to be G-Markovian if they hold.*

Proof. We need to show that if (UP) holds then (UG) holds, i.e. that if C separates A and B then $A \perp B \mid C$, where we for the sake of simplicity leave out the $[P]$. We do this by backwards induction on the number of vertices in C. If C has $(\# V-2)$ elements then both A and B consist of one vertex and the conditional independence follows directly from (UP). So suppose that C has r elements, where r is less than $(\# V-2)$, and that the result holds for all C's with more elements. Let us first assume that $A \cup B \cup C = V$ implying that at least one of the sets A and B, say A, has more than one element and let α be one of them. It is easy to see that $C \cup \{\alpha\}$ separates $A \setminus \{\alpha\}$ and B and that $C \cup A \setminus \{\alpha\}$ separates $\{\alpha\}$ and B, so by the induction hypothesis we have $A \setminus \{\alpha\} \perp B \mid C \cup \{\alpha\}$ and $\{\alpha\} \perp B \mid C \cup A \setminus \{\alpha\}$. Using (CI5) then gives $A \perp B \mid C$. If $A \cup B \cup C \neq V$, we can choose α in $V \setminus (A \cup B \cup C)$ and then $C \cup \{\alpha\}$ separates A and B implying $A \perp B \mid C \cup \{\alpha\}$. Further we see that either $C \cup A$ will separate $\{\alpha\}$ and B or $C \cup B$ will separate $\{\alpha\}$ and A. Assuming that the first holds we get $\{\alpha\} \perp B \mid C \cup A$ which together with $A \perp B \mid C \cup \{\alpha\}$ gives $A \perp B \mid C$ by use of (CI5). □

From this we get the following lemma that will make some of the proofs in the next sections more easy.

Lemma 2.1
If P fulfils (CI5), *then the following statements are equivalent for three subsets A, B and W of V such that $A \subseteq W$, $B \subseteq W$ and $A \cap B = \emptyset$:*

(*i*) $A' \perp B' \mid W \setminus (A \cup B)[P]$ *for all* $A' \subseteq A$ *and* $B' \subseteq B$
(*ii*) $A \perp B \mid W \setminus (A \cup B)[P]$
(*iii*) $\alpha \perp \beta \mid W \setminus \{\alpha, \beta\}[P]$ *for all* $\alpha \in A$ *and* $\beta \in B$
(*iv*) *P_W is $\tilde{G}$-Markovian, where* $\tilde{G} = (W, \tilde{E})$ *and* $\tilde{E} = E^*(W \setminus B) \cup E^*(W \setminus A)$.

Proof. (ii) is a special case of (i) and (iii) follows from (ii) by use of (CI3). (iii) says that P_W is pairwise $\tilde{G}$-Markovian as the only missing edges in $\tilde{G}$ are between A and B, so (iii) implies (iv). Finally we get the last implication by observing that $W \setminus (A \cup B)$ separates A and B in $\tilde{G}$. □

A well known and very useful property of the undirected graphs and their Markov properties is the fact that the Markov properties under some circumstances can be decomposed into Markov properties of subgraphs. If we as in Lauritzen & Wermuth (1989) call a subset of vertices *simplicial* in G if its boundary is complete, then one version of this decomposition result is

Proposition 2.2
Let G be an undirected graph, P fulfil (CI5) *and C be a subset of vertices. If C is simplicial in G then*

(*i*) *P is G-Markovian*

if and only if

(ii) $P_{V\setminus C}$ *is* $G_{V\setminus C}$*-Markovian,* $P_{cl(C)}$ *is* $G_{cl(C)}$*-Markovian and* $C \perp V\setminus cl(C) \mid bd(C)[P]$.

Proof. That (i) implies (ii) follows from the fact that $bd(C)$ separates C and $V\setminus cl(C)$ in G and that if α and β are not adjacent in $G_{V\setminus C}$ or $G_{cl(C)}$ then they are separated by $[V\setminus C]\setminus\{\alpha,\beta\}$ or $cl(C)\setminus\{\alpha,\beta\}$, respectively, in G. To show the reverse implication assume that (ii) holds and that α and β are not adjacent in G implying that at least one of them, say, α, does not belong to $bd(C)$. We then have to prove that $\alpha \perp \beta \mid V\setminus\{\alpha,\beta\}$. The proof is split into three cases. First consider the case where $\{\alpha,\beta\} \subseteq V\setminus C$ and let $D=[V\setminus C]\setminus\{\alpha,\beta\}$. From the first part of (ii) we get $\alpha \perp \beta \mid D$ from the last part of (ii) and (CI3) that $\alpha \perp C \mid D \cup \{\beta\}$. Combining these two statements we get using (CI4) $\alpha \perp \{\beta\} \cup C \mid D$ which by (CI3) implies $\alpha \perp \beta \mid D \cup C$, i.e. $\alpha \perp \beta \mid V\setminus\{\alpha,\beta\}$. Completely analogous arguments take care of the case where α and β both belongs to $C \cup bd(C)$. The last case, where say $\alpha \in V\setminus cl(C)$ and $\beta \in C$, follows from $C \perp V\setminus cl(C) \mid bd(C)$ by use of (CI3). □

3. The Markov properties for chain graphs

Now let G be a chain graph and recall that the future of a vertex α, $\phi(\alpha)$, consists of all the vertices that can be reached from α by a directed path. The probability measure P is then said to be

(P) *Pairwise G-Markovian* if $\alpha \perp \beta \mid [V\setminus\phi(\alpha)]\setminus\{\alpha,\beta\}[P]$ whenever $\beta \notin \phi(\alpha)$ and β and α are not adjacent

(L) *Local G-Markovian* if $\alpha \perp [V\setminus\phi(\alpha)]\setminus cl(\alpha) \mid bd(\alpha)[P]$ for all α

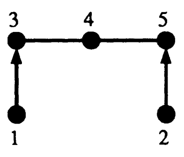
(G) *Global G-Markovian* if $A \perp B \mid C[P]$ whenever C separates A and B in $(G_{an(A\cup B\cup C)})^m$

where $(G_{an(A\cup B\cup C)})^m$ is the moral graph of the smallest anterior set containing $A\cup B\cup C$. Some remarks are needed here. First we see that if G is undirected then $\phi(\alpha)$ is empty, hence (P) and (UP) are identical and the same is true for (L) and (UL). That (G) and (UG) are equivalent for undirected graphs follows from the fact that one set separates two other sets in an undirected graph if and only if it separates them in the collection of connected components, i.e. chain components, containing them all.

Secondly we see that by (CI3) we have that (L) implies (P). To see that (G) implies (L), also without any assumption about P, is a little more complicated, but this follows from the fact that $V\setminus\phi(\alpha)$ is an anterior set and $bd(\alpha)$ separates α and $[V\setminus\phi(\alpha)]\setminus cl(\alpha)$ in $(G_{V\setminus\phi(\alpha)})^m$, as no edges containing α are added in this graph.

Example.

G



The pairwise Markov property for G states:

$1 \perp 2$
$1 \perp 4 \mid (2, 3, 5)$
$1 \perp 5 \mid (2, 3, 4)$
$2 \perp 4 \mid (1, 3, 5)$
$2 \perp 3 \mid (1, 4, 5)$
$5 \perp 3 \mid (1, 2, 4)$

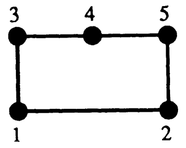
The local Markov property for G states:

$1 \perp 2$
$2 \perp 1$
$3 \perp (2, 5) \mid (1, 4)$
$4 \perp (1, 2) \mid (3, 5)$
$5 \perp (1, 3) \mid (2, 4)$

As there are four anterior sets in G, $\{1\}$, $\{2\}$, $\{1,2\}$ and $V=\{1,2,3,4,5\}$ we see that the global Markov property says that $1 \perp 2$, because the smallest anterior set containing $\{1,2\}$ is $\{1,2\}$ and $(G_{\{1,2\}})^m$ is

and $A \perp B \mid C$ if $A \cup B \cup C$ contains a vertex in $\{3,4,5\}$ and C separates A and B in

G^m



This yields all the conditional independencies given by the pairwise and the local properties plus the following list:

$1 \perp 5 \mid (2, 3)$
$1 \perp 5 \mid (2, 4)$
$2 \perp 3 \mid (1, 4)$
$2 \perp 3 \mid (1, 5)$
$1 \perp (4, 5) \mid (2, 3)$
$2 \perp (3, 4) \mid (4, 5)$.

Notice that none of the above statements says that $3 \perp 5 \mid 4$, which might have been expected from the graph. That the statement does not follow from, for instance, the pairwise Markov property can be seen by letting $X=(X_1, \ldots, X_5)$ follow a Gaussian distribuition with covariance matrix

$$\Sigma=\begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ -1 & -\frac{1}{4} & \frac{33}{16} & & \\ -\frac{1}{2} & -\frac{1}{2} & \frac{9}{8} & \frac{3}{2} & \\ -\frac{1}{4} & -1 & \frac{3}{4} & \frac{9}{8} & \frac{33}{16} \end{bmatrix}$$

having as inverse

$$\Sigma^{-1}=\begin{bmatrix} 2 & & & & \\ \frac{1}{4} & 2 & & & \\ 1 & 0 & \frac{4}{3} & & \\ 0 & 0 & -\frac{2}{3} & \frac{5}{3} & \\ 0 & 1 & 0 & -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

implying that X has a pairwise G-Markovian distribution. But the 3, 5 entry in the inverse of

$$\Sigma(3,4,5)=\begin{bmatrix} \frac{33}{16} & & \\ \frac{9}{8} & \frac{3}{2} & \\ \frac{3}{4} & \frac{9}{8} & \frac{33}{16} \end{bmatrix}$$

is $\frac{4}{63}$ so X_3 and X_5 are not conditionally independent given X_4. □

Because the results for the undirected case are included in this set-up we cannot show any of the reverse implications without some assumption about P, but as we will show in this section it will, also in this case, suffice to assume that (CI5) holds.

Let us for a start show that removal of a terminal chain component leaves the pairwise Markov properties unchanged in the following sense:

Lemma 3.1
If τ is a terminal chain component in G and P fulfilling (CI5) *is pairwise G-Markovian then $P_{V\backslash\tau}$ is pairwise $G_{V\backslash\tau}$-Markovian.*

Proof. We have to show that for a pair α and β in $V\backslash\tau$ such that β does not belong to $\tilde{\phi}(\alpha)$, the future of α in $G_{V\backslash\tau}$, and α and β are not adjacent, we have that $\alpha\perp\beta\,|\,[V\backslash\tau]\backslash\tilde{\phi}(\alpha)\backslash\{\alpha,\beta\}$. Now $\tilde{\phi}(\alpha)=\phi(\alpha)\backslash\tau$, so if τ is a subset of $\phi(\alpha)$ the wanted result follows directly from the pairwise Markov property on G. The only thing left is then to show the result in the case where τ is not a subset of $\phi(\alpha)$ implying that they are disjoint and thereby that $\tilde{\phi}(\alpha)=\phi(\alpha)$. The pairwise Markov property on G used on α and β then gives $\alpha\perp\beta\,|\,[V\backslash\phi(\alpha)]\backslash\{\alpha,\beta\}$ and we need to show that it is enough to condition on $[V\backslash\phi(\alpha)]\backslash\{\alpha,\beta\}\backslash\tau$. To do this we observe that no pair δ and γ with $\delta\in\phi(\alpha)\cup\{\alpha\}$ and $\gamma\in\tau$ can be adjacent in G, so we have $\delta\perp\gamma\,|\,V\backslash\{\delta,\gamma\}$ for all such pairs. This implies by lemma 2.1 that $\alpha\perp\gamma\,|\,[V\backslash\phi(\alpha)]\backslash\{\alpha,\gamma\}$ for all γ in τ and by the above statement also for $\gamma=\beta$. We can then again apply lemma 2.1 (iii) $\Rightarrow$ (i) this time with $W=V\backslash\phi(\alpha)$, $A=\{\alpha\}$ and $B=\tau\cup\{\beta\}$ to get the needed result. □

The next thing we need before we can show the equivalence of the three chain graph Markov properties is a result connecting the pairwise Markov property with the Markov property of an undirected graph, the moral graph.

Lemma 3.2
If P fulfilling (CI5) *is pairwise G-Markovian, then P is G^m-Markovian.*

Proof. Induction on the number of chain components in G. If G has only one chain component then it is undirected and $G^m=G$, and the lemma is trivial. Now suppose the lemma holds for all chain graphs with less than r chain components. Let G be a chain graph with r chain components and choose τ terminal in G. We then have to prove for instance that $\alpha\perp\beta\,|\,V\backslash\{\alpha,\beta\}$ for all pairs (α,β) not in E^m, i.e. that P is local G^m-Markovian.

If α or β belongs to τ then we get the wanted result by applying the pairwise Markov property directly. If this is not the case, that is if both α and β belong to $V\backslash\tau$, we proceed as follows. First we note that the fact α and β are not adjacent in G^m implies that they are not adjacent in the moral graph generated by $G_{V\backslash\tau}$. So we can use lemma 3.1 and the induction hypothesis to get $\alpha\perp\beta\,|\,[V\backslash\tau]\backslash\{\alpha,\beta\}$. Then we observe that at least one of the vertices, say α, does not belong to the boundary of τ, because else α and β would be adjacent in G^m. So by the pairwise Markov property we have $\alpha \perp \gamma | V \backslash \{\alpha, \gamma\}$ for all γ in τ yielding $\alpha \perp \tau | [V \backslash \tau] \backslash \{\alpha\}$ by lemma 2.1. We can now use (CI4) to get $\alpha \perp \{\beta\} \cup \tau | [V \backslash \tau] \backslash \{\alpha, \beta\}$ and we are finished. □

We can now prove the main result in this section.

Theorem 3.3
The following four statements are equivalent for any P fulfilling (CI5) *and any chain graph G:*

(i) *P is global G-Markovian*
(ii) *P is local G-Markovian*
(iii) *P is pairwise G-Markovian*
(iv) *P_A is $(G_A)^m$-Markovian for every anterior set A of G.*

If the statements hold we say that P is G-Markovian.

Proof. We have already shown (i) ⇒ (ii) ⇒ (iii). To show that (iv) follows from (iii) we recall that an anterior set is a set which is generated by stepwise removal of terminal chain components. So by lemma 3.1 we have that P_A is pairwise G_A-Markovian and by lemma 3.2 we get (iv). Now if (iv) holds then (i) holds by the undirected global Markov property. □

We conjecture that the global Markov property given here is the strongest possible, that is if for three sets A, B and C, C does not separate A and B in $(G_{an(A \cup B \cup C)})^m$ then there exists a state space $\mathscr{X} = \times_{\alpha \in V} \mathscr{X}_\alpha$ and a G-Markovian probability measure on $\mathscr{X}$ such that A and B are not conditionally independent given C. This holds in the undirected case and in the case where G is a directed acyclic graph (see Frydenberg (1990) and Verma (1988)).

Theorem 3.3 does not say so much about how a G-Markovian probability measure looks like or how it can be constructed. A better understanding of this is given by the corollary below, which gives a recursive definition of the chain graph Markov properties.

Corollary 3.4
If P fulfils (CI5), *G is a chain graph and τ is a terminal chain component of G then*

(i) *P is G-Markovian,*
(ii) *$P_{V \backslash \tau}$ is $G_{V \backslash \tau}$-Markovian and P is G^m-Markovian*

and

(iii) *$P_{V \backslash \tau}$ is $G_{V \backslash \tau}$-Markovian, $P_{cl(\tau)}$ is $(G_{cl(\tau)})^m$-Markovian and $\tau \perp V \backslash cl(\tau) | bd(\tau)[P]$*

are equivalent.

Proof. That (i) implies (ii) is lemma 3.1 and 3.2. (iii) follows from (ii), the fact that $(G_{cl(\tau)})^m = (G^m)_{cl(\tau)}$ and proposition 2.2. That (iii) implies (i) is proved if we can prove that P_A is $(G_A)^m$-Markovian for all anterior sets A. The case where A is a subset of $V \backslash \tau$ follows directly from $P_{V \backslash \tau}$ is $G_{V \backslash \tau}$-Markovian. In the other case, i.e. $\tau \subseteq A$, $P_{A \backslash \tau}$ is $G_{A \backslash \tau}$-Markovian and hence $(G_{A \backslash \tau})^m$-Markovian, which again implies that $P_{A \backslash \tau}$ is $\{(G_A)^m\}_{A \backslash \tau}$-Markovian as this graph has more edges. Now as we also have that $\tau \perp A \backslash cl(\tau) | bd(\tau)$ and $(G_{cl(\tau)})^m = \{(G_A)^m\}_{cl(\tau)}$ proposition 2.2 does the rest. □

In their paper on graphical association models Lauritzen & Wermuth (1989) used a slightly different approach in the definition of a chain Markov property. Recall that a *dependence chain* $\mathscr{D}$ is an ordered partition of V into disjoint subsets $\mathscr{D} = \{V(1), V(2), \ldots, V(K)\}$ and introduce the notation $W(k) = \cup_{j \leqslant k} V(j)$ and $V(\alpha) = V(k)$ and $W(\alpha) = W(k)$ if $\alpha \in V(k)$. So in this notation the chain graph generated by the undirected graph G and $\mathscr{D}$ has the edges $E^{\mathscr{D}} = \{(\alpha, \beta) | (\alpha, \beta) \in E \text{ and } \alpha \in W(\beta)\}$.

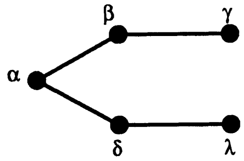
Lauritzen & Wermuth then defined the local Markov property given by the undirected graph G and the dependence chain $\mathscr{D}$ as follows: P is said to be

Local $[G, \mathscr{D}]$-*Markovian* if $\alpha \perp W(\alpha) \backslash cl(\alpha) \mid bd(\alpha) \cap W(\alpha) [P]$ for all α

i.e. if we only look at the vertices that come before or at the same time as α and condition on the vertices that are adjacent with α in G then α is independent of the rest.

Example.

G



If $\mathscr{D} = (\{\alpha\}, \{\beta, \gamma\}, \{\delta, \lambda\})$ then the local $[G, \mathscr{D}]$-Markov property states:

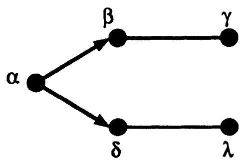
$\lambda \perp (a, \beta, \gamma) \mid \delta$
$\delta \perp (\gamma, \beta) \mid (\alpha, \lambda)$
$\gamma \perp \alpha \mid \beta$

If $\mathscr{D}' = (\{\alpha\}, \{\beta, \gamma, \delta, \lambda\})$ then the local $[G, \mathscr{D}']$-Markov property states:

$\delta \perp (\gamma, \beta) \mid (\alpha, \lambda)$
$\lambda \perp (\alpha, \beta, \gamma) \mid \delta$
$\gamma \perp (\alpha, \delta, \lambda) \mid \beta$
$\beta \perp (\delta, \lambda) \mid (\alpha, \gamma)$

It is easily checked that P is local $[G, \mathscr{D}]$-Markov if and only if it is local $[G, \mathscr{D}']$-Markov and that the chain graphs $G^{\mathscr{D}}$ and $G^{\mathscr{D}'}$ both are equal to

$\tilde{G}$



which again have the same Markov properties. □

As indicated above the local $[G, \mathscr{D}]$-Markov property only depends on the generated chain graph $G^{\mathscr{D}}$ as formulated by the following theorem, which follows from corollary 3.4 by an induction argument.

Theorem 3.5
If P fulfils (CI5), *G is an undirected graph and $\mathscr{D}$ is a dependence chain on its vertices then P is local $[G, \mathscr{D}]$-Markovian if and only if it is $G^{\mathscr{D}}$-Markovian.*

4. A factorization of the density

As shown in the beginning of the previous section the global chain Markov property is the strongest of the three Markov properties. But an even stronger condition is the condition of factorization of the density.

So assume that P has a density, p, w.r.t. some product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on $\mathscr{X}$. We will then say that P or p has a *Gibbs-factorization* w.r.t. the undirected graph G if there exist functions ψ^C such that $p(x) = \Pi_{C \in \mathscr{C}} \psi^C(x_C)$, where $\mathscr{C}$ denotes the set of cliques, i.e. the maximal complete sets, in G. It is easy to see that this factorization implies that P is global G-Markovian. But again the reverse implication does not hold without some further assumptions about P and it seems that the assumption needed here is that the density p should be strictly positive. That this in fact is enough, i.e. that

(A1): *If G is undirected, P is global G-Markovian and has a strictly positive density then P has Gibbs-factorization w.r.t. G.*

is true, has been proved in many special cases, e.g. assuming the $\mathscr{X}_\alpha$ finite or countable or assuming that P is Gaussian (see e.g. Speed (1979) and Isham (1981)), but a general proof of (A1) seems not to be published.

In the following we will show that the chain graph Markov property can be characterized by a similar factorization of the density and that this in fact can be proved using only the above factorization result for the undirected Markov property. Note that if P has a positive density then (CI5) holds so the results from the last section can be applied. We will use the standard notation, that is for a set of vertices A, $p_A(\cdot)$ will denote the marginal density on $\mathscr{X}_A = \times_{\alpha \in A} \mathscr{X}_\alpha$, i.e. $p_A(x_A) = \int p(x_A, x_{V \setminus A}) \mu_{V \setminus A}(dx_{V \setminus A})$ and further if $B \subseteq V \setminus A$, $p_{B|A}(\cdot|\cdot)$ will denote the conditional density of X_B given X_A i.e. $p_{B|A}(x_B|x_A) = p_{A \cup B}(x_A, x_B)/p_A(x_A)$.

Theorem 4.1

Assuming that (A1) *holds the following four statements are equivalent for any P with a positive density and any chain graph G:*

(*i*) *P is G-Markovian*

(*ii*) $p(x) = \prod_{\tau \in \mathscr{T}} p_{\tau|bd(\tau)}(x_\tau | x_{bd(\tau)})$ *where $P_{cl(\tau)}$ is $(G_{cl(\tau)})^m$-Markovian for all $\tau \in \mathscr{T}$*

(*iii*) *p can be factorized* $p(x) = \prod_{\tau \in \mathscr{T}} \prod_{C \in \mathscr{C}_\tau} \psi_\tau^C(x_C)$ *such that and* $\int_{\mathscr{X}_\tau} \prod_{C \in \mathscr{C}_\tau} \psi_\tau^C(x_C) \mu_\tau\,(dx_\tau) \equiv 1$

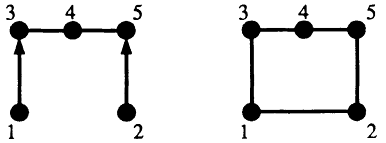
for all τ in $\mathscr{T}$ where $\mathscr{T}$ denotes the set of chain components in G and $\mathscr{C}_\tau$ denote the cliques in $(G_{cl(\tau)})^m$

(*iv*) *If A is an anterior set then p_A has a Gibbs factorization w.r.t. $(G_A)^m$.*

Proof. (i) implies (ii) by corollary 3.4 and induction on the number of chain components. As $p_{\tau|bd(\tau)} = p_{cl(\tau)}/p_{bd(\tau)}$ we get (iii) from (ii) by using (A1) on $P_{cl(\tau)}$ and observing that $bd(\tau)$ is complete in $(G_{cl(\tau)})^m$. (iv) follows from (iii) by stepwise integration over the chain components not in A. Finally we get (i) from (iv) by observing that P_A is $(G_A)^m$-Markovian for all anterior sets A. □

Example.

G



G^m

If P has a positive density then it is G-Markovian if and only if it is G^m-Markovian and the density can be factorized:

$$p(x) = p_1(x_1) \cdot p_2(x_2) \cdot p_{345|12}(x_3, x_4, x_5 | x_1, x_2)$$

or equivalently if P has a density on the form

$$p(x)=\psi_1^1(x_1)\psi_2^2(x_2)\psi_{345}^{12}(x_1,x_2)\psi_{345}^{13}(x_1,x_3)\psi_{345}^{25}(x_2,x_5)\psi_{345}^{34}(x_3,x_4)\psi_{345}^{45}(x_4,x_5)$$

where

$$\psi_{345}^{12}(x_1,x_2)^{-1}=\int\psi_{345}^{13}(x_1,x_3)\psi_{345}^{25}(x_2,x_5)\psi_{345}^{34}(x_3,x_4)\psi_{345}^{45}(x_4,x_5)\mu_{345}(dx_3\,dx_4\,dx_5)$$

for all x_1 and x_2. □

5. Markov equivalence of chain graphs

It is well known that the Markov properties which can be read off a directed acyclic graph and its underlying undirected graph are the same if and only if the boundary of every vertex is complete, implying that the undirected graph is triangulated, see e.g. Wermuth (1980) or Kiiveri, Speed & Carlin (1984). But no other results of this type seem to be shown, not even for the directed acyclic graphs.

In this section we will say that two chain graphs G and $\tilde{G}$ have the same Markov properties if P is G-Markovian if and only if P is $\tilde{G}$-Markovian for all P fulfilling (CI1) to CI5). The proofs in this section are essentially consequences of proposition 2.2, which shows how the Markov properties for an undirected graph can be decomposed into the Markov properties of two subgraphs and corollary 3.4, which gives a recursive description of the Markov properties of chain graphs.

For a start we show a little result concerning the Markov properties of a chain graph and the chain graph in which we have split a terminal chain component into two.

Lemma 5.1
Let $\tilde{G}$ be a chain graph and G be the chain graph, which differs from $\tilde{G}$ in the only way that for τ, a connected subset of $\tilde{\tau}$ a terminal chain component in $\tilde{G}$ all edges in $\tilde{G}$ between $\tilde{\tau}\backslash\tau$ and τ are changed into arrows towards τ in G. If τ is simplicial in $\tilde{G}^m$ or equivalently if $(\tilde{G}_{cl(\tau)})^m=(G_{cl(\tau)})^m$ then G and $\tilde{G}$ have the same Markov properties.

Proof. First we can use corollary 3.4 twice, noting that τ is a terminal chain component in G and $\tilde{\tau}\backslash\tau$ is a terminal chain component in $G_{V\backslash\tau}$, yielding that P is G-Markovian if and only if P is G^m-Markovian, $P_{V\backslash\tau}$ is $(G_{V\backslash\tau})^m$-Markovian and $P_{V\backslash\tilde{\tau}}$ is $G_{V\backslash\tilde{\tau}}$-Markovian. Secondly it follows from the construction of G and the fact that τ is simplicial in $\tilde{G}$ that $G^m=\tilde{G}^m$ and $(G_{V\backslash\tau})^m=(\tilde{G}_{V\backslash\tau})^m=(\tilde{G}^m)_{V\backslash\tau}$. Hence P is G-Markovian if and only if P is $\tilde{G}^m$-Markovian, $P_{V\backslash\tau}$ is $(\tilde{G}^m)_{V\backslash\tau}$-Markovian and $P_{V\backslash\tilde{\tau}}$ is $\tilde{G}_{V\backslash\tilde{\tau}}$-Markovian. Proposition 2.2 used on τ and $\tilde{G}^m$ states that $P_{V\backslash\tau}$ is $(\tilde{G}^m)_{V\backslash\tau}$-Markovian if P is $\tilde{G}^m$-Markovian. Using this we see that P is G-Markovian if and only if P is $\tilde{G}^m$-Markovian and $P_{V\backslash\tilde{\tau}}$ is $\tilde{G}_{V\backslash\tilde{\tau}}$-Markovian. Finally we can use corollary 3.4 again, this time on $\tilde{G}$, yielding that P is G-Markovian if and only if it is $\tilde{G}$-Markovian. □

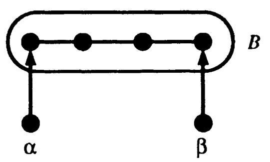
If we say $\tilde{G}=(V,\tilde{E})$ is *larger than* $G=(V,E)$ if $E\subseteq\tilde{E}$, we get the following more general result.

Proposition 5.2
Let $G=(V,E)$ and $\tilde{G}=(V,\tilde{E})$ be two chain graphs with the same undirected graph and $\tilde{G}$ is larger than G, i.e. G might have arrows where $\tilde{G}$ has undirected edges. If all chain components τ in G are simplicial in $(\tilde{G}_{cl(\tau)})^m$ then G and $\tilde{G}$ have the same Markov properties.

Proof. Note that for any $W \subset V$, a terminal chain component in $\tilde{G}_W$ will contain a terminal chain component in G_W. Hence the chain components in G can be ordered, $\tau_1, \tau_2, \ldots, \tau_T$, such that τ_t is terminal in $G_{W(t)}$ and a subset of a terminal chain component in $\tilde{G}_{W(t)}$, where $W(t) = \cup_{s \leq t} \tau_s$. As $V = W(T)$ we are finished if we can show that $G_{W(t)}$ and $\tilde{G}_{W(t)}$ have the same Markov properties for all $1 \leq t \leq T$. This is obviously true for $t=1$, as $G_{\tau_1} = \tilde{G}_{\tau_1}$. In general we consider $\tilde{G}_{W(t)}$ and using lemma 5.1 we see that this graph has the same Markov properties as the graph $\hat{G}$ on $W(t)$, with only arrows pointing towards τ_t from $W(t-1)$. Now following corollary 3.4 P is $\hat{G}$-Markovian if and only if $P_{W(t-1)}$ is $\hat{G}_{W(t-1)}$-Markovian, $\tau_t \perp W(t) \backslash c\hat{l}(\tau_t) \mid b\hat{d}(\tau_t)$ and $P_{cl(\tau_t)}$ is $(\hat{G}_{c\hat{l}(\tau_t)})^m$-Markovian. So as $b\hat{d}(\tau_t) = bd(\tau_t)$, $(\hat{G}_{c\hat{l}(\tau_t)})^m = (G_{cl(\tau_t)})^m$ and $\hat{G}_{W(t-1)} = \tilde{G}_{W(t-1)}$ we see again by corollary 3.4 that if $G_{W(t-1)}$ and $\tilde{G}_{W(t-1)}$ have the same Markov properties then $G_{W(t)}$ and $\tilde{G}_{W(t)}$ have the same Markov properties, too and the wanted result follows by induction. □

We can now check whether two chain graphs have the same Markov properties, but only if they are nested. To get a result on how to check two arbitrary chain graphs it is convenient to translate the conditions in proposition 5.2 into some other property of the chain graphs.

We will call a triple (α, B, β) a *complex in G* if B is a connected subset of a chain component, τ, and α and β are two non-adjacent vertices in $bd(\tau) \cap bd(B)$. Further we will call (α, B, β) a *minimal* complex if $B = B'$ whenever B' is a subset of B and (α, B', β) is a complex. It is easily seen that (α, B, β) is a minimal complex in G if and only if $G_{B \cup \{\alpha, \beta\}}$ looks like:



The concept of complexes in chain graphs can be seen as a generalization of configurations [>] in recursive causal graphs (see Wermuth (1980) and Kiiveri, Speed & Carlin (1984)). In those papers a triple (α, γ, β) of vertices in a recursive causal graph is said to be in configuration [>] if $(\alpha, \{\gamma\}, \beta)$ is a (minimal) complex.

It is easily seen that if G and $\tilde{G}$ are two chain graphs with the same underlying undirected graph and $\tilde{G}$ is larger than G, then G and $\tilde{G}$ have the same minimal complexes if and only if any minimal complex in G is a minimal complex in $\tilde{G}$. Using this we can now prove.

Proposition 5.3

Let $G=(V, E)$ and $\tilde{G}=(V, \tilde{E})$ be two chain graphs with the same undirected graph and $\tilde{G}$ is larger than G, i.e. G might have arrows where $\tilde{G}$ has undirected edges. Then

(*i*) *all chain components τ in G are simplicial in $(\tilde{G}_{cl(\tau)})^m$*

if and only if

(*ii*) *G and $\tilde{G}$ have the same minimal complexes.*

Proof. Assume that (i) holds, (α, B, β) is a minimal complex in G and let $B \subseteq \tau \subseteq \tilde{\tau}$, where τ and $\tilde{\tau}$ are chain components in G and $\tilde{G}$ respectively. We note that $\{\alpha, \beta\} \subseteq bd(\tau)$ and that they are not adjacent in $G^u = \tilde{G}^u$. As (i) holds α and β are adjacent in $(\tilde{G}_{cl(\tau)})^m$ so α and β must also be connected to τ with arrows in $\tilde{G}$ and hence (α, B, β) must be a minimal complex in $\tilde{G}$. On the other hand let τ be a chain component in G and α and β be two non-adjacent vertices in $bd(\tau)$ implying (α, B, β) is a minimal complex in G for some $B \subseteq \tau$. So if (ii) holds then (α, B, β) is a minimal complex in $\tilde{G}$, too, and hence in $\tilde{G}_{cl(\tau)}$, but then α and β will be adjacent, when we moralize this graph. □

So now proposition 5.2 can be read: if G and $\tilde{G}$ have the same underlying undirected graph, the same minimal complexes and G is larger than $\tilde{G}$, then they have the same Markov properties.

But what about the case where $G=(V, E)$ and $\tilde{G}=(V, \tilde{E})$ are two arbitrary chain graphs? First we can observe that if we change all the arrows in the graph $G \cup \tilde{G} := (V, E \cup \tilde{E})$, (this might not be a chain graph), which are part of a directed cycle, into undirected edges, then we get a chain graph, which we will denote $G \vee \tilde{G}$ (the details are given in Frydenberg (1989)). Next we can use the following graph theoretical result (the proof is given at the end of this section).

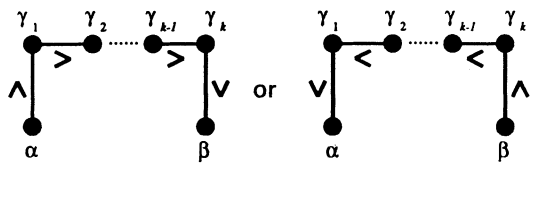
Proposition 5.4
Two chain graphs $G=(V, E)$ and $\tilde{G}=(V, \tilde{E})$ with the same underlying undirected graph have the same minimal complexes if and only if they both have the same minimal complexes as $G \vee \tilde{G}$, the unique smallest chain graph larger than both of them.

We can now conlude.

Proposition 5.5
Two chain graphs have the same Markov properties if they have the same underlying undirected graph and the same minimal complexes.

Observe that the proposition as it stands only gives sufficient conditions for Markov equivalence of chain graphs. To show that the condition is also necessary requires construction of counter-examples. So let us consider the state space $\mathscr{X} = \{-1, 1\}^V$ and note that (A1) and hence theorem 4.1 hold for this state space.

Let G and $\tilde{G}$ be two chain graphs and suppose that α and β are adjacent in $\tilde{G}$ but not in G. It is then easily seen that α and β are not adjacent in the moral graph of the smallest anterior set containing α and β, $an(\alpha, \beta)$, in G, implying that α and β are conditionally independent given the rest of the vertices in $an(\alpha, \beta)$ for all G-Markovian P. But if P is the probability measure on $\mathscr{X}$ with density $p(x) = \text{constant} \times \exp\{x_\alpha x_\beta\}$, then we see that P is $\tilde{G}$-Markovian, because it factorizes with respect to $\tilde{G}$, and that α and β are not independent no matter what we condition on, so P is not G-Markovian. So if G and $\tilde{G}$ have the same Markov properties then they must have the same underlying undirected graph. So assume that this is the case, but the graphs do not have the same minimal complexes. We can then without loss of generality assume that (α, B, β) is a minimal complex in G and that $\tilde{G}_{B \cup \{\alpha, \beta\}}$ does not contain any complexes. Again we see that $\alpha \perp \beta \mid an(\alpha, \beta) \setminus \{\alpha, \beta\}$ for any G-Markovian P, as α and β are not adjacent in $(G_{an(\alpha, \beta)})^m$. Further we see that $an(\alpha, \beta)$ does not contain any vertex from B as all the vertices in this set comes after both α and β. Now $\tilde{G}_{B \cup \{\alpha, \beta\}}$ look likes,



where a > off an edge indicates a possible direction of the edge. If we let $\gamma_0=\alpha$ and $\gamma_{k+1}=\beta$ and consider the density $p(x)=\text{constant}\times\Pi_0^k \exp\{x_{\gamma_i}x_{\gamma_{i+1}}\}$ we see that P is $\tilde{G}$-Markovian as it factorizes with respect to $\tilde{G}$. Further we see that P is not G-Markovian as α and β are only conditionally independent under P if we condition with at least one vertex in B. So proposition 5.4 can be strengthened to

Theorem 5.6 (*Markov equivalence of chain graphs*)
Two chain graphs have the same Markov properties if and only if they have the same underlying undirected graph and the same minimal complexes.

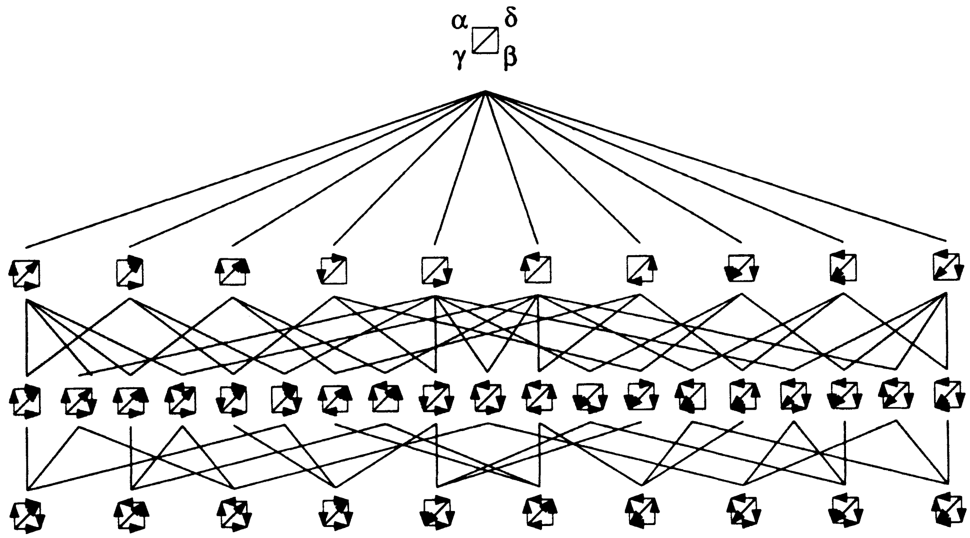
As an undirected graph does not have any complexes, the theorem states that a chain graph has the same Markov properties as its underlying undirected graph if and only if the chain graph does not have any complexes or, equivalently, if the boundary of every chain component is complete.

One more thing worth noting is that using the theorem and proposition 5.4 we can get:

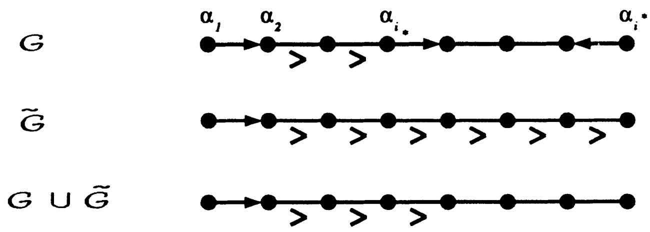
Proposition 5.7
For any chain graph G there exists a unique largest chain graph, i.e. a chain graph with at least as many undirected edges as G, with the same Markov properties as G.

Proof. Just observe that there only exist finitely many chain graphs with vertex set V and that if G_1 and G_2 are two different chain graphs larger than G and with the same underlying undirected graph and the same minimal complexes as G then, by proposition 5.4, $G_1 \vee G_2$ has the same properties. □

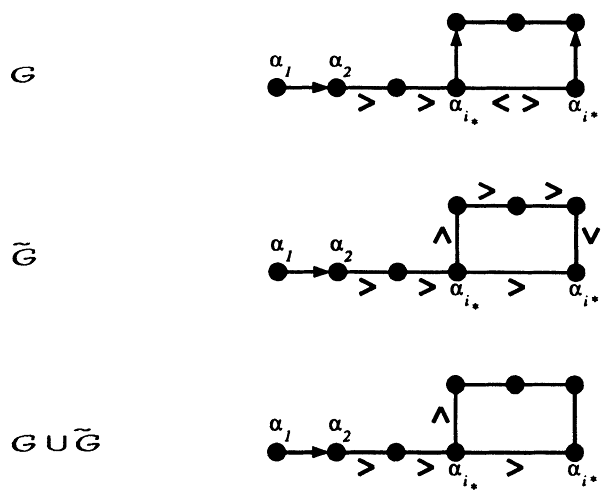
Example. The figure gives all the chain graphs which have the same Markov properties, $\alpha \perp \beta \mid (\gamma, \delta)$. The rows contain graphs with the same number of chain components and a line from one graph to another indicates that it can be derived by splitting or joining chain components.



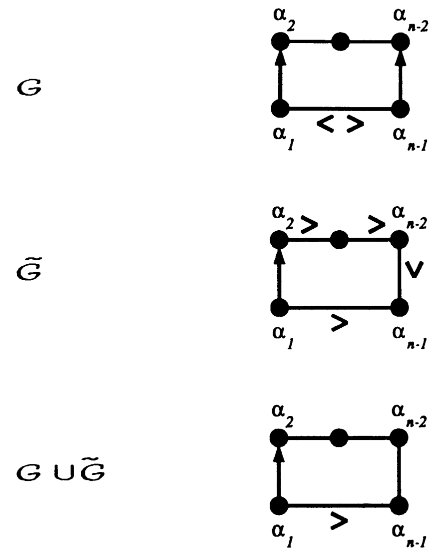
Proof of proposition 5.4. Proposition 5.4 is proved if we can prove that if (α, B, β) is a minimal complex, with $B=\{\gamma_i\}_{i=1}^k$, in G and $\tilde{G}$ then $\alpha \Rightarrow \gamma_1$ is not a part of a directed cycle in $G\cup\tilde{G}=(V, E\cup\tilde{E})$. Because this implies that (α, B, β) also is a minimal complex in $G\vee\tilde{G}$. Now suppose $\alpha \Rightarrow \gamma_1$ is a part of a directed cycle $(\alpha_1, \alpha_2, \dots, \alpha_n)$ in $G\cup\tilde{G}$, with $\alpha_n=\alpha_1=\alpha$ and $\alpha_2=\gamma_1$ and let the cycle be chosen as short as possible. We choose i^* so that $\alpha_{i^*-1} \Leftarrow \alpha_{i^*}$ is the first $\Leftarrow$ arrow in G or $\tilde{G}$, say in G. There are three things to notice here. Firstly that such an i^* exist, because otherwise $(\alpha_1, \alpha_2, \dots, \alpha_n)$ would be a directed cycle in both G and $\tilde{G}$. Secondly we see that α_{i^*-1} and α_{i^*} cannot be connected by a $\Leftarrow$ arrow in $\tilde{G}$, too, as this would mean that $(\alpha_1, \alpha_2, \dots, \alpha_n)$ also would have the same arrow in $G\cup\tilde{G}$, which cannot be. Thirdly we notice that i^* is smaller or equal to $n-1$, because else $(\alpha_1, \alpha_2, \dots, \alpha_n)$ would be a directed cycle in $\tilde{G}$. Further we choose i_* so $\alpha_{i_*} \Rightarrow \alpha_{i_*+1}$ is the last $\Rightarrow$ arrow in G before i^*. Here we note that this is possible as $\alpha_1 \Rightarrow \alpha_2$ in G. So we have a picture looking like this



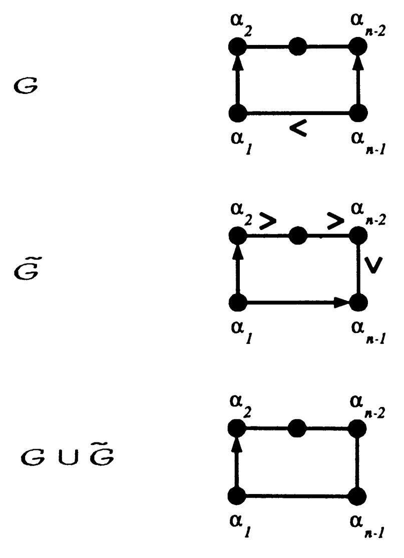
where a $<$ or $>$ off an edge indicates a possible direction of the edge and where there might be other edges than those drawn. We see that if α_{i_*} and α_{i^*} are not adjacent then $(\alpha_{i_*}, \hat{B}, \alpha_{i^*})$ would be a minimal complex in G for some subset $\hat{B}$ of $\{\alpha_i\}_{i_*<i<i^*}$. But in $\tilde{G}$ all arrows in $\{\alpha_i\}_{i_*\leq i\leq i^*}$ will be $\Rightarrow$ arrows so $(\alpha_{i_*}, \hat{B}, \alpha_{i^*})$ cannot be a minimal complex in $\tilde{G}$ and we can conclude that α_{i_*} and α_{i^*} are indeed adjacent. This leads us to



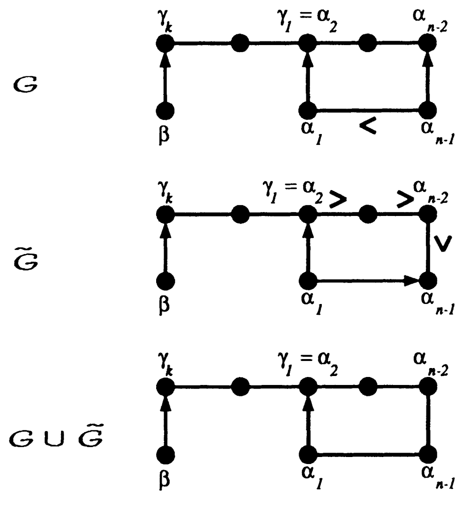
Further we see that, as at least $(\alpha_{i_*}, \alpha_{i^*}) \in \tilde{E}$, we have that $(\alpha_{i_*}, \alpha_{i^*}) \in E\cup\tilde{E}$. So if $i_*>1$ then $(\alpha_1, \alpha_2, \dots, \alpha_{i_*}, \alpha_{i^*}, \dots, \alpha_n)$ will be a directed cycle in $G\cup\tilde{G}$ which is shorter than the original. This is not possible so we conclude that $i_*=1$. The same argument will give that $i^*=n-1$. And the situation is now



As $\tilde{G}$ cannot have $\alpha_1 \Leftrightarrow \alpha_{n-1}$, because this would create a directed cycle in $\tilde{G}$, we see that $\alpha_1 \Rightarrow \alpha_{n-1}$ in $\tilde{G}$. Further we see that because $(\alpha_1, \alpha_2, \ldots, \alpha_1)$ is a cycle in $G \cup \tilde{G}$ we must have $\alpha_1 \Leftrightarrow \alpha_{n-1}$ in this graph, which implies that we cannot have a $\alpha_1 \Rightarrow \alpha_{n-1}$ arrow in G. So the picture now looks like this:

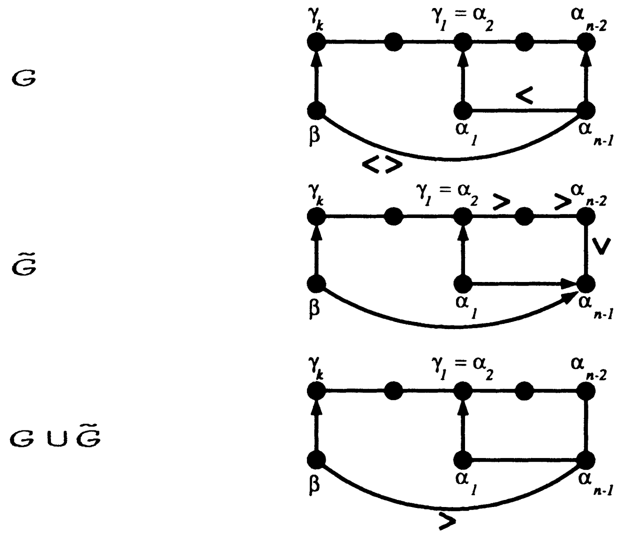


We can now incorporate the knowledge that we know that $(\alpha_1, \alpha_2)=(\alpha, \gamma_1)$ is an arrow in a minimal complex in G and $\tilde{G}$.



Here the illustration above might be a little misleading, because $\{\gamma_m\}_{m=2}^{k}$ and $\{\alpha_i\}_{i=2}^{n-2}$ might not be disjoint, but the main things are that β and α_{n-1} cannot be identical as α_{n-1} is adjacent to $\alpha_1=\alpha$ and that γ_k and α_{n-2} are connected in G with an undirected path.

If β and α_{n-1} are not adjacent then $(\beta, \grave{B}, \alpha_{n-1})$ is a minimal complex in G for some subset $\grave{B}$ of $\{\gamma_1, \ldots, \gamma_k, \alpha_2, \ldots, \alpha_{n-2}\}$. By looking at $\tilde{G}$ we see that this cannot be a minimal complex in $\tilde{G}$, and we conclude that β and α_{n-1} are adjacent. Further we see that in $\tilde{G}$ we have $\beta \Rightarrow \alpha_{n-1}$ and we have now reached the final picture:



As β and $\alpha=\alpha_1$ are not adjacent, because (α, B, β) is a minimal complex in G and $\tilde{G}$, we see that $(\alpha, \{\alpha_{n-1}\}, \beta)$ is a minimal complex in $\tilde{G}$. But as α and α_{n-1} cannot be connected $\alpha \Rightarrow \alpha_{n-1}$ in G this is not a minimal complex in G and we have a contradiction because we know that G and $\tilde{G}$ have the same minimal complexes. We must therefore conclude that (α, γ_1) cannot be a part of a directed cycle in $G \cup \tilde{G}$. □

6. Discussion

It is important to note that in the entire paper, except for section 4 and the constructed counter examples in section 5, we only use (CI1) to (CI5), so the results given here do not only hold for any probability measure but for any relation on triplets, which meets (CI1) to (CI5). This set-up is inspired by the papers Pearl & Paz (1986), Pearl & Verma (1987), Verma (1988), Geiger & Pearl (1988) and Smith (1989).

As mentioned in section 3, we are convinced that the global condition (G) is the strongest possible, i.e. that if C does not separate A and B in $(G_{an(A \cup B \cup C)})^m$ then there exists a G-Markovian probability measure such that A and B are not conditionally independent given C. So to prove that this in fact is the case we need the construction of a general counter-example, typically in the Bernoulli case $\mathscr{X} = \{0, 1\}^V$ or the Gaussian case $\mathscr{X} = \mathbb{R}^V$. Such counter-examples have been constructed in the Gaussian case where the graph is directed and acyclic (see Verma (1988) and Geiger & Pearl (1988)). The construction of the counter-examples seems to require a characterization of when C separates A and B in $(G_{an(A \cup B \cup C)})^m$ in terms of properties of the paths between A and B in the undirected graph, as given in Pearl & Verma (1987) by their d-separation criteria for the directed acyclic graphs. A similar separation criterion has not yet been found for the chain graphs (an attempt is given in Verma (1988)).

It would be nice to know a procedure that for a given chain graph constructs the largest chain graph with the same Markov properties, but such a procedure is not known.

Finally let us remark that Lauritzen *et al.* (1990) investigated the Markov properties for directed acyclic graphs and showed, among other things, that the equivalence of the local and the global Markov properties in this case holds, without assuming (CI5).

Acknowledgment

The author is indebted to Steffen Lauritzen for many fruitful and encouraging discussions of earlier versions of this paper.

References

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8**, 598–617.

Geiger, D. & Pearl, J. (1988). On the logic of influence diagrams. *Proceedings of 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, Minn. USA, August 1988, 136–147.

Frydenberg, M. (1990). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* **18** (to appear).

Frydenberg, M. (1989). Chain graphs and some of their properties. Manuscript.

Isham, V. (1981). An introduction to spatial point processes and Markov random fields. *Internat. Statist. Rev.* **49**, 21–43.

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

Lauritzen, S. L. (1989). Mixed interaction models. *Scand. J. Statist.* **16**, 273–306.

Lauritzen, S. L., Dawid, A. P., Larsen, B. & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks* (to appear).

Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50**, 157–224.

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

Pearl, J. & Paz, A. (1986). Graphoids. A graph-based logic for reasoning about relevancy relations. *Proceedings of 7th European Conference on Artificial Intelligence*, Brighton, United Kingdom, June 1986.

Pearl, J. & Verma, T. (1987). The logic of representing dependencies by directed graphs. *Proc. Amer. Assoc. Art. Intell. Conference*. Seattle, Washington.

Smith, J. Q. (1989). Influence diagrams for statistical modelling. *Ann. Statist.* **17,** 654–672.

Speed, T. P. (1979). A note on nearest-neighhbour Gibbs and Markov probabilities. *Sankhyā A* **41,** 184–197.

Verma, T. (1988). Causal networks. Semantics and expressiveness. *Proceedings of 4th Workshop on Uncertainty in Artificial Inteligence*, Minneapolis, Minn. USA, August 1988, 352–359.

Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *J. Amer. Statist. Assoc.* **75,** 963–972.

Wermuth, N. & Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. R. Statist. Soc. B* **52,** 21–50.

Received February 1989, in final form April 1990

Morten Frydenberg, Department of Theoretical Stastistics, Institute of Mathematics, Aarhus University, DK-8000 Aarhus, Denmark.