The Annals of Statistics
1990, Vol. 18, No. 2, 790–805

MARGINALIZATION AND COLLAPSIBILITY IN GRAPHICAL INTERACTION MODELS

BY MORTEN FRYDENBERG

University of Aarhus

The behaviour of a graphical interaction model under marginalization is discussed. A graphical interaction model is called collapsible onto a set of variables if the class of marginal distributions is the same as that implied by the related subgraph. The necessary and sufficient condition for collapsibility is found and it is shown that collapsibility is equivalent to a range of other important statistical properties of the model.

0. Introduction. Recent years have seen an increase in the use of graphs in fields such as expert systems, decision analysis and statistics [see, e.g., Lauritzen and Spiegelhalter (1988), Pearl (1988), and Shachter (1986)]. One of the general features of these applications of graphs is that the graphs, which can be both directed and undirected, are used to represent knowledge or hypotheses concerning the association between different variables/decisions. In statistics one of the most recent developments in this direction is the introduction of the graphical association models by Lauritzen and Wermuth (1984, 1989).

The class of graphical association models is a class of models for association between variables that can be of discrete as well as of continuous type. Each model is specified by a graph where each variable is represented by a vertex, and two vertices are connected if there is a direct association between the corresponding variables, or, in other words, two variables are not connected if they are conditionally independent given some other variables. In general the graph contains both undirected edges, indicating a symmetric association, and directed edges, indicating an influence–response association. A review of the ideas behind and use of these models is given in Wermuth and Lauritzen (1990). Because the building blocks in these general models (the graphical chain models) are the symmetric models, where the graph is undirected, a deeper understanding of this latter class, termed interaction models, is important.

The graphical interaction model is a generalization to the mixed case of the graphical models for contingency tables [Darroch, Lauritzen and Speed (1980)] for purely discrete data and the covariance selection models [Dempster (1972)] for purely continuous data. In Frydenberg and Lauritzen (1989), the central concept of decomposition of a graphical interaction model was considered. In the present work we will look at another important concept: collapsibility. As in Asmussen and Edwards (1983), considering log-linear models for contin-

Received January 1988; revised June 1989.
AMS 1980 *subject classifications*. Primary 62H99; secondary 62J99.
Key words and phrases. Graphical models, decomposition, collapsibility, maximum likelihood estimate, contingency tables, covariance selection, marginalization, cut.

gency tables, we will say that a given graphical interaction model is *collapsible* onto a set of variables if the implied model for the marginal distributions for these variables is equal to the graphical interaction model given by the induced subgraph. The results shown in this paper cover the results found in the above paper for graphical models for contingency tables and the results shown in Porteous (1985) in the case of covariance selection models. As in the case of contingency tables, the graphical interaction models can be seen as a subclass of a larger class of symmetric interaction models, called the hierarchical interaction models [see Edwards (1990)], which also give an extension of some of the results presented here to this larger class.

Here we will only indicate a few of the many implications of collapsibility for inference, understanding and interpretation of both the symmetric interaction models and the more general graphical chain models; a more detailed discussion can be found in Asmussen and Edwards (1983). If a graphical interaction model is collapsible onto a set A of variables, then:

1. The maximum likelihood estimate in the full model of the marginal distribution of the variables in A is equivalent to the maximum likelihood estimate in the graphical interaction models only containing the variables in A.
2. Inference concerning removal of some edges can be performed in smaller graphical interaction models.
3. The model is equivalent to a graphical chain model with the variables in A as explanatory variables to the rest of the variables.

The structure of the paper is as follows. In Section 1, after having introduced the class of CG-distributions formally and recollected the concept of Markov properties given by an undirected graph, we define the graphical interaction models. The following section contains a proof of the fact that a graphical interaction model given by a graph G is Markov perfect, by which we mean that the model does not prescribe any other conditional independencies between sets of variables than those which can be read from the graph. Sections 3 and 4 are concerned with what happens under marginalization. First, assuming that $\mathscr{P}$ is a Markov perfect class of G-Markovian measures, we give a necessary and sufficient condition for all of the marginal distributions for a subset A of variables to be Markovian w.r.t. the subgraph G_A given by A. A special case in which this condition is satisfied is when the marginal distributions under the graphical interaction model fulfill the Markov properties w.r.t. the related subgraph.

Next, in Section 4, the question of when the marginal distribution of a CG-distribution is again a CG-distribution is discussed, and a necessary and sufficient condition for the marginal distributions for a graphical interaction model to be CG-distributions is given. In the last section we collect the results found in the first sections and establish in the main theorem the equivalence of collapsibility to five other properties of the model. Especially we show that a graphical interaction model is collapsible onto a set of variables if and only if it is equivalent to a graphical chain model with the specific set of variables as explanatory to the rest. Using this we can quite easily prove the general result concerning equivalence of graphical interaction models and graphical chain models. Finally, we show some implications of the main theorem followed by a brief discussion of the consequences of collapsibility for estimation problems in graphical association models.

1. Preliminaries and notation. As in Lauritzen and Wermuth (1989), we let V be an index set partitioned into Δ (the discrete indices) and Γ (the continuous indices), and suppose that to each $\alpha \in V$ there is associated a set $\mathscr{X}_\alpha$ in such a way that $\mathscr{X}_\alpha = \mathscr{I}_\alpha$, a finite set if $\alpha \in \Delta$ and $\mathscr{X}_\alpha = \mathbb{R}$ if $\alpha \in \Gamma$. We shall consider the product space $\mathscr{X} = \times_{\alpha \in V} \mathscr{X}_\alpha = \mathscr{I} \times \mathscr{Y}$, where $\mathscr{I} = \times_{\delta \in \Delta} \mathscr{I}_\delta$ and $\mathscr{Y} = \mathbb{R}^\Gamma$. A point in $\mathscr{X}$ will be denoted by $x = (i, y)$, where $i \in \mathscr{I}$ and $y \in \mathscr{Y}$. For $A \subset V$ we will let $\mathscr{X}_A = \times_{\alpha \in A} \mathscr{X}_\alpha$, $\mathscr{I}_A = \times_{\delta \in A \cap \Delta} \mathscr{I}_\delta$ and $\mathscr{Y}_A = \mathbb{R}^{\Gamma \cap A}$ and the projection of x onto $\mathscr{X}_A$ will be denoted by x_A, i.e., if $x = \{x_\alpha; \alpha \in V\}$, then $x_A = \{x_\alpha; \alpha \in A\}$. All densities will be with respect to the natural measure, i.e., the product measure of the counting measure on $\mathscr{I}$ and the Lebesgue measure on $\mathbb{R}^\Gamma$. A probability distribution P is a *CG-distribution* if it has a density p of the form

$$p(x) = p(i, y) = \exp\{g(i) + h(i)'y - \tfrac{1}{2}y'K(i)y\}, \qquad x \in \mathscr{X},$$

where for each $i \in \mathscr{I}$, $g(i) \in \mathbb{R}$, $h(i) \in \mathbb{R}^\Gamma$ and $K(i)$ is a symmetric positive definite $\Gamma \times \Gamma$ matrix, or equivalently that P has a positive density and the conditional distribution of Y given $I = i$ is Gaussian. A CG-distribution is called *homogeneous* if $K(i)$ or, equivalently, the conditional variance of Y given $I = i$, does not depend on i.

The special CG-distribution which specifies complete independence of all the coordinates, constant density on $\mathscr{I}$ and for each $\gamma \in \Gamma$ a standardized Gaussian distribution for X_γ is denoted by P^0, i.e., $p^0(x) = p^0(i, y) = \text{const}\exp\{-\frac{1}{2}\sum_{\gamma \in \Gamma} y_\gamma^2\}$. For $A \subset V$ and arbitrary distribution P on $\mathscr{X}$ we will denote the marginal distribution on $\mathscr{X}_A$ by P_A. If P has positive density we will denote the marginal density on $\mathscr{X}_A$ by p_A and the conditional density on $\mathscr{X}$ given $\mathscr{X}_A$ by p^A, i.e., $p^A(x_{V \setminus A}|x_A) = p(x_V)/p_A(x_A)$, and if $\mathscr{P}$ is a class of such distributions we put $\mathscr{P}_A = \{p_A(\cdot); P \in \mathscr{P}]$ and $\mathscr{P}^A = \{p^A(\cdot \mid \cdot); P \in \mathscr{P}\}$. As in Barndorff-Nielsen (1978) X_A (or A) is called a *cut* in $\mathscr{P}$ if the mapping $\mathscr{P} \to \mathscr{P}_A \times \mathscr{P}^A$ sending p to (p_A, p^A) is onto $\mathscr{P}_A \times \mathscr{P}^A$, i.e., if for each pair (p_A, p^A) in $\mathscr{P}_A \times \mathscr{P}^A$ the product $p_A(x_A) \cdot p^A(x_{V \setminus A}|x_A)$ is a density of some measure in $\mathscr{P}$.

Now, let $E_V^* = \{(\alpha, \beta)|\alpha \in V, \beta \in V, \alpha \neq \beta\}$ be the set of all ordered pairs of distinct elements of V. Then a pair $G = (V, E)$, where $E \subset E_V^*$ is called a *graph* with *vertices* V and *edges* E. In this paper all graphs will be *undirected*, i.e., $(\alpha, \beta) \in E$ implies $(\beta, \alpha) \in E$. For a graph $G = (V, E)$ and a subset A of V, we will denote the induced *subgraph* on A by G_A, i.e., $G_A = (A, E \cap (A \times A))$, the *boundary* of A by $\text{bd}(A) = \{\beta \in V \setminus A | \exists\ \alpha \in A$ such that $(\beta, \alpha) \in E\}$ and the *closure* of A by $\text{cl}(A)$, i.e., $\text{cl}(A) = A \cup \text{bd}(A)$. A graph

$G = (V, E)$ is said to be *complete* if $E = E_V^*$ and a subset $A \subset V$ is called complete in G if G_A is complete. A n-tuple of n different vertices $(\alpha_1, \dots, \alpha_n)$ is called a *path* of length $(n - 1)$ from α_1 to α_n if $(\alpha_i, \alpha_{i+1}) \in E$ for $1 \le i < n$. A subset A of V is a *connected component* in G if every pair of vertices in A is connected by a path and none of the vertices outside A is connected to A. B' is called a connected component in $B \subset V$ if it is a connected component in the subgraph G_B. Finally $C \subset V$ is said to *separate* two subsets A and B of V if every path from $\alpha \in A$ to $\beta \in B$ contains at least one element from C.

The graphs we shall be considering are graphs with vertices of two types, $V = \Delta \cup \Gamma$, and when we draw a graph we will indicate the type of vertex by drawing the vertices in Δ (the discrete vertices) as dots and the vertices in Γ (the continuous vertices) as circles.

In the graphical models the graphs are used to specify the conditional independencies in the model. The concept of conditional independence is discussed in Dawid (1979, 1980), and following those papers we will here use the notation $X_A \perp X_B | X_C\ [P]$, or in short $A \perp B | C\ [P]$, if X_A and X_B are conditionally independent given X_C under P.

A distribution P is said to be *G-Markovian* if $A \perp B | C$ whenever C separates A and B in G. As proved by Pearl and Paz [see Pearl (1988)] this global Markov property is equivalent to the pairwise Markov property: "$\alpha \perp \beta | V \setminus \{\alpha, \beta\}$ for all nonadjacent vertices α and β" if P has a positive density.

Now having introduced the CG-distributions and the Markov properties given by a graph, we define the *graphical interaction model* $\mathscr{M}(G)$ specified by G as the class of all G-Markovian CG-distributions on $\mathscr{X}$.

Some equivalent definitions of $\mathscr{M}(G)$, such as a special factorization of the densities, can be found in Lauritzen and Wermuth (1989). For $A \subset V$, $\mathscr{M}(G_A)$ will denote the graphical interaction model on $\mathscr{X}_A$ specified by G_A. The main point of this paper is to find the necessary and sufficient condition for $\mathscr{M}(G)_A$ to be equal to $\mathscr{M}(G_A)$, i.e., for the class of marginal distributions of $\mathscr{M}(G)$ on a subset of variables to be equal to the graphical interaction model specified by the induced subgraph. As a start, we will here give the easy argument for the inclusion $\mathscr{M}(G_A) \subset \mathscr{M}(G)_A$. Suppose $P_{(A)} \in \mathscr{M}(G_A)$. Then $p(x) = p_{(A)}(x_A) \cdot p^0_{V \setminus A}(x_{V \setminus A})$ will be a density for a $P \in \mathscr{M}(G)$ and by definition $P_A = P_{(A)}$.

For $(x^{(1)}, \dots, x^{(n)}) \in \mathscr{X}^n$ and $A \subset V$, we let $\hat{P}_{[A]}$ denote the maximum likelihood estimate in $\mathscr{M}(G_A)$ (if it exists) based on $(x_A^{(1)}, \dots, x_A^{(n)})$ assuming $x_A^{(i)}$ are i.i.d. observations from some measures in $\mathscr{M}(G_A)$. We will write $\hat{P}$ for $\hat{P}_{[V]}$.

Finally let us notice that the graphical interaction models contain both the graphical models for contingency tables ($\Gamma = \varnothing$) and the covariance selection models ($\Delta = \varnothing$) as special cases.

2. The Markov properties of $\mathscr{M}(G)$. It follows from the definition of $\mathscr{M}(G)$ that if C separates A and B in G and $P \in \mathscr{M}(G)$, then $A \perp B | C$ under P. The purpose of this section is to show that the opposite is also true in the sense that if $A \perp B | C$ under *every* P in $\mathscr{M}(G)$, then C must separate A and B

in G; in other words, that the model $\mathscr{M}(G)$ does not satisfy any conditional independencies other than those given by G.

DEFINITION 2.1. A class $\mathscr{P}$ of distributions on $\mathscr{X}$ is said to be *Markov perfect* w.r.t. G if $A \perp B | C$ under every P in $\mathscr{P}$ implies that C separates A and B in G.

(We will say a G-Markovian class is Markov perfect if it is Markov perfect w.r.t. G.)

Before we start proving that $\mathscr{M}(G)$ is Markov perfect, we may note that if $\mathscr{P}$ is Markov perfect w.r.t. G and A is a subset of V, then $\mathscr{P}_A$ is Markov perfect w.r.t. G_A. In fact, suppose that C, A_1 and A_2 are subsets of A and that $A_1 \perp A_2 | C$ under every $P_A \in \mathscr{P}_A$ or equivalently under every $P \in \mathscr{P}$. If $\mathscr{P}$ is Markov perfect w.r.t. G, then C must separate A_1 and A_2 in G, which implies that C separates A_1 and A_2 in G_A.

In the following lemma, for every discrete vertex δ, we let $\mathscr{I}_\delta$ be divided in two nonempty sets A_δ and B_δ, $A_\delta \cap B_\delta = \varnothing$, and we define the functions h_δ and w_δ by setting $h_\delta(x_\delta)$ equal to -1 if $x_\delta \in A_\delta$ and equal to 1 if $x_\delta \in B_\delta$, and $w_\delta(x_\delta) = |A_\delta|^{-1}$ if $x_\delta \in A_\delta$ and $w_\delta(x_\delta) = |B_\delta|^{-1}$ if $x_\delta \in B_\delta$. That is, the possible values of X_δ are partitioned into two groups which are assigned the interaction parameters -1 and 1 and the weights 1 divided by the number of values in each group. For a continuous vertex γ we let $h_\gamma(x_\gamma) = x_\gamma$ and $w_\gamma(x_\gamma) = 1$.

LEMMA 2.2. *Let $G = (V, E)$ where $V = \{\alpha_l\}_{l=1}^n$ and $E = \{(\alpha_k, \alpha_m); |k - m| = 1\}$. Then the distribution with the density*

$$p(x) = \text{const} \times \prod_{l=1}^{n} w_l(x_l) \exp\left\{-\tfrac{1}{2}\left[h_l(x_l)^2 - h_l(x_l) h_{l+1}(x_{l+1})\right]\right\}$$

[setting $h_{n+1}(x_{n+1}) \equiv 0$] is a G-Markovian homogeneous CG-*distribution with the property that $h_k(X_k)$ and $h_m(X_m)$ are not marginally independent for any pair (k, m) (and therefore X_k and X_m are not independent either).*

PROOF. First we observe that $p(x) = \exp\{g(i) + h(i)'y - \frac{1}{2}y'Ky\}$ for suitable choice of g and h, and with K being the positive definite symmetric matrix with diagonal elements 1 and off-diagonal elements $-\frac{1}{2}$ if the corresponding continuous variables are adjacent and otherwise 0. From this we can see that p has a finite integral and that it is a homogeneous CG-density. That P also is G-Markovian follows from the factorization of the density.

For simplicity we will here only show that $h_k(X_k)$ and $h_m(X_m)$ are dependent in the case where $I_\delta = \{-1, 1\}$, $A_\delta = \{-1\}$ and $B_\delta = \{1\}$ for every $\delta \in \Delta$. The proof in the general case then follows, working with the transformed variables $(h_l(x_l))$. Now, if $I_\delta = \{-1, 1\}$, then $h_l(x_l) = x_l$ and $w_l(x_l) = 1$ for all

l, implying that $p(x)$ is of the form

$$p(x) = \text{const} \times \prod_{l=1}^{n} \exp\left(-\tfrac{1}{2}\{x_l^2 - x_l x_{l+1}\}\right), \qquad x \in \mathscr{X}$$

$(x_{n+1} = 0)$.

Next, we will show that:

(i) $p(x_l) = p(-x_l) > 0$.

(ii) $p(x_k, x_m) = p(-x_k, -x_m) > 0$, $1 \le k < m \le n$, $x_k \in \mathscr{X}_k$, $x_m \in \mathscr{X}_m$.

(iii) $p(x_k, x_m) > p(x_k, -x_m)$, $1 \le k < m \le n$, $x_k \in \mathscr{X}_k$, $x_m \in \mathscr{X}_m$, $x_k x_m > 0$.

This implies that x_k and x_m are dependent.

(i) and (ii) follow easily from the symmetry of the density.

An induction argument now gives, by repeated integration, that for $k < m$,

$$p(x_k, \dots, x_m) = \mathscr{S}_k^1(x_k)\left[\prod_{l=k}^{m-1} \exp\left(-\tfrac{1}{2}\{x_l^2 - x_l x_{l+1}\}\right)\right]\mathscr{S}_m^n(x_m),$$

where $\mathscr{S}_k^1(x_k) = \mathscr{S}_k^1(-x_k)$ and $\mathscr{S}_m^n(x_m) = \mathscr{S}_m^n(-x_m)$. Especially we have that $p(x_l, x_{l+1}) = \mathscr{S}_l^1(x_l)\exp(-\frac{1}{2}\{x_l^2 - x_l x_{l+1}\})\mathscr{S}_{l+1}^n(x_{l+1})$, yielding $p(x_l, x_{l+1}) = e^{x_l x_{l+1}} p(x_l, -x_{l+1})$ and thus (iii) for $m = k + 1$. The proof of (iii) in general is carried out by induction in $m - k$. Suppose (iii) holds for any pair k, m such that $0 < m - k < r \le n$. Now let $m = k + r$ and put $l = m - 1$. The Markov property implies that $X_k \perp X_m | X_l$ and, by induction, (iii) holds for $p(x_k, x_l)$ and $p(x_l, x_m)$. Let $x_k \in \mathscr{X}_k$ and $x_m \in \mathscr{X}_m$ both be positive. Then using the conditional independence, we have

$$\begin{aligned} &p(x_k, x_m) - p(x_k, -x_m) \\ &\quad = \int_{-\infty}^{\infty} p(x_k, x_l) p(x_l, x_m)/p(x_l)\mu(dx_l) \\ &\qquad - \int_{-\infty}^{\infty} p(x_k, x_l) p(x_l, -x_m)/p(x_l)\mu(dx_l) \\ &\quad = \int_{-\infty}^{\infty} p(x_k, x_l)[p(x_l, x_m) - p(x_l, -x_m)]/p(x_l)\mu(dx_l), \end{aligned}$$

where $\int_{-\infty}^{\infty} \mu(dx_l)$ is either the Lebesgue integral over the whole line or the sum $\sum_{x_l = -1, 1}$. With this in mind and using (ii) on $p(x_l, x_m)$ and (i) on $p(x_l)$, we get

$$\begin{aligned} &\int_{-\infty}^{0} p(x_k, x_l)[p(x_l, x_m) - p(x_l, -x_m)]/p(x_l)\mu(dx_l) \\ &\quad = \int_{0}^{\infty} p(x_k, -x_l)[p(-x_l, x_m) - p(-x_l, -x_m)]/p(-x_l)\mu(dx_l) \\ &\quad = \int_{0}^{\infty} p(x_k, -x_l)[p(x_l, -x_m) - p(x_l, x_m)]/p(x_l)\mu(dx_l) \\ &\quad = -\int_{0}^{\infty} p(x_k, -x_l)[p(x_l, x_m) - p(x_l, -x_m)]/p(x_l)\mu(dx_l), \end{aligned}$$

implying

$$p(x_k, x_m) - p(x_k, -x_m)$$
$$= \int_0^{\infty} [p(x_k, x_l) - p(x_k, -x_l)][p(x_l, x_m) - p(x_l, -x_m)]/p(x_l)\mu(dx_l).$$

Finally, we use the induction hypothesis on $p(x_k, x_l)$ and $p(x_l, x_m)$ to conclude that the two terms in square brackets are positive, yielding a strictly positive integral, and (iii) is shown for $x_k > 0$ and $x_m > 0$. The result for $x_k < 0$ and $x_m < 0$ follows from (ii). □

We are now ready to prove the main result of this section.

THEOREM 2.3. *$\mathscr{M}(G)$, the class of G-Markovian* CG-*distributions, is Markov perfect.*

PROOF. We have to prove that if $A \perp B|C$ under every P in $\mathscr{M}(G)$, then C separates A and B in G, or equivalently that if C does not separate A and B in G, then there exists a P in $\mathscr{M}(G)$ such that A and B are not conditionally independent given C.

So suppose C does not separate A and B in G. Then there exists a path, say $(\alpha_1, \dots, \alpha_n)$, from A to B not intersecting C. Setting $\tilde{V} = \{\alpha_l\}_{l=1}^n$ and P equal to the distribution in $\mathscr{M}(G)$ with density

$$p(x) = p_{\tilde{V}}(x_{\tilde{V}}) \cdot p^0_{V \setminus \tilde{V}}(x_{V \setminus \tilde{V}}),$$

where $p_{\tilde{V}}(x_{\tilde{V}})$ is the density given in Lemma 2.2, we see that the marginal distribution on $\mathscr{X}_{\tilde{V}}$ is of the form in Lemma 2.2, which implies that α_1 and α_n are not marginally independent. This together with the fact that $\tilde{V} \perp C$ under P gives that α_1 and α_n are not conditionally independent given C. Thus A and B are not conditionally independent given C under P. □

Another way of formulating Theorem 2.3 is that

$$C \text{ separates } A \text{ and } B \text{ in } G$$

if and only if

$$A \perp B|C \text{ under every } P \text{ in } \mathscr{M}(G).$$

It should be noted that the p constructed in the above proof is a homogeneous CG-distribution, so we have also proved that the class of G-Markovian homogeneous CG-distributions is Markov perfect.

3. The Markov properties under marginalization. In this section we will establish the necessary and sufficient condition for a Markov perfect G-Markovian class of distributions to have marginal distributions on $\mathscr{X}_A$ which are G_A-Markovian.

DEFINITION 3.1. Let $G = (V, E)$ be an undirected graph. A subset $B \subset V$ is called *simplicial* in G if $\mathrm{bd}(B)$ is complete and a *simplicial collection* in G if every connected component in B is simplicial in G.

In the proof of Corollary 2.5 in Asmussen and Edwards (1983) it is shown that $V \setminus A$ is a simplicial collection in G if and only if the following is true: If C separates A_1 and A_2 in G_A, then C separates A_1 and A_2 in G, too. Using this and the definition of the Markov property, we get the result:

PROPOSITION 3.2. *Let P be any G-Markovian probability distribution on $\mathscr{X}$ and let $A \subset V$. If $V \setminus A$ is a simplicial collection, then P_A is G_A-Markovian.*

On the other hand, suppose we have a G-Markovian class $\mathscr{P}$ of distributions on $\mathscr{X}$ and suppose that $\mathscr{P}_A$ is G_A-Markovian, too. Furthermore, let C, A_1 and A_2 be any three subsets of A such that C separates A_1 and A_2 in G_A. It follows that $A_1 \perp A_2 | C$ under any P_A in $\mathscr{P}_A$ or equivalently under any P in $\mathscr{P}$. Now, if $\mathscr{P}$ is Markov perfect, then C must separate A_1 and A_2 in G, and as a consequence $V \setminus A$ must be a simplicial collection in G.

THEOREM 3.3. *Let $\mathscr{P}$ be a Markov perfect G-Markovian class of distributions on $\mathscr{X}$ and let $A \subset V$. Then the following are equivalent:*

(i) *$V \setminus A$ is a simplicial collection in G.*
(ii) *$\mathscr{P}_A$ is a G_A-Markovian class.*

In the previous section we showed that $\mathscr{M}(G)$ is Markov perfect, so for the graphical interaction models we have:

COROLLARY 3.4. *Let $A \subset V$. Then $\mathscr{M}(G)_A$ is a G_A Markovian class if and only if $V \setminus A$ is a simplicial collection in G.*

Noting that if $V \setminus A$ is a simplicial collection in G, then $B \setminus A$ is a simplicial collection in G_B, we obtain the following result concerning graphical association models specified by subgraphs.

COROLLARY 3.5. *Let A and B be subsets of V. If $\mathscr{M}(G)_A$ is G_A-Markovian, then $\mathscr{M}(G_B)_{A \cap B}$ is $G_{A \cap B}$-Markovian.*

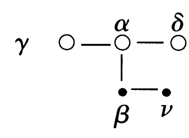
We have now clarified the behaviour of the Markov properties under marginalization, and in the following section we will investigate the similar behaviour of the distributional assumptions. But before we start on this, some remarks about the two pure cases, i.e., $\Gamma = \varnothing$ and $\Delta = \varnothing$. In both cases, i.e., the contingency tables and the covariance selection models, the distributional assumptions are always preserved under marginalization, so we have the following result.

COROLLARY 3.6. *If* $\Delta = \emptyset$ *or* $\Gamma = \emptyset$, *then* $\mathscr{M}(G)_A = \mathscr{M}(G_A)$ *if and only if* $V \setminus A$ *is a simplicial collection in* G.

This result was first proved in Asmussen and Edwards (1983) for the purely discrete case and in Porteous (1985) for the purely continuous case.

EXAMPLE 3.7.

G



Using Corollary 3.4 we see that $\mathscr{M}(G)_{\{\alpha,\beta\}}$ is a $G_{\{\alpha,\beta\}}$-Markovian class and that $\mathscr{M}(G)_{\{\delta,\nu\}}$ is not a $G_{\{\delta,\nu\}}$-Markovian class.

4. The marginal distributions. As shown in Lauritzen and Wermuth (1989), the class of CG-distributions is closed under conditioning but in general not under marginalization. In the following we will give the necessary and sufficient condition for the marginal distribution of a CG-distribution to be a CG-distribution.

LEMMA 4.1. *Let* $A \subset V$, $B = V \setminus A$ *and suppose that* $X = (I_A, I_B, Y_A, Y_B)$ *has a* CG-*distribution. Then* $X_A = (I_A, Y_A)$ *has a* CG-*distribution if and only if* $Y_A \perp I_B | I_A$.

PROOF. The if part follows from Propositions 2.1 and 2.2 in Lauritzen and Wermuth (1989), but the short proof will be repeated here.

First, we observe that the marginal density for X_A is always positive because X has a positive density. So X_A has a CG-distribution if and only if the conditional distribution of Y_A given $I_A = i_A$ alone is Gaussian. By assumption, the conditional distribution of Y_A given $(I_A, I_B) = (i_A, i_B)$ is Gaussian with mean and variance depending on (i_A, i_B). Now, if Y_A is conditionally independent of I_B given I_A, then the distribution of Y_A given $I_A = i_A$ is equal to that of Y_A given $(I_A, I_B) = (i_A, i_B)$ for any value of i_B, i.e., it is Gaussian. On the other hand, suppose Y_A is not conditionally independent of I_B given I_A, i.e., there exists an $i_A \in I_A$ such that the conditional mean or variance of Y_A given $(I_A, I_B) = (i_A, i_B)$ does depend on i_B. Then the conditional distribution of Y_A given $I_A = i_A$ is a discrete mixture of Gaussian distributions which do not have common means and variances, and consequently this cannot be Gaussian [see, e.g., Yakowitz and Spragins (1968)]. □

The condition $Y_A \perp I_B | I_A$ could also be written $(\Gamma \setminus B) \perp (B \cap \Delta) | \Delta \setminus B$. Turning to $\mathscr{M}(G)$ we want to translate this statement into some properties of B in G.

DEFINITION 4.2. A subset B of V is called *strong* in G if $B \subset \Gamma$ or $\mathrm{bd}(B) \subset \Delta$, and a *strong collection* in G if every connected component in B is strong in G.

Suppose $\delta \in \Delta \cap B$, $\gamma \in \Gamma \setminus B$ and γ and δ are connected with a path which must contain at least one vertex in $\mathrm{bd}(B_t)$, where B_t is the connected component in B containing δ. Now if B is a strong collection, then $\mathrm{bd}(B_t) \subset \Delta \setminus B$, so the path between γ and δ must contain a vertex in $\Delta \setminus B$, and hence this set separates $\Gamma \setminus B$ and $B \cap \Delta$. On the other hand, if $\Gamma \setminus B$ and $B \cap \Delta$ are separated by $\Delta \setminus B$ and $B_t \cap \Delta \neq \emptyset$, then $\mathrm{bd}(B_t)$ cannot contain any continuous vertices, because if it did we would have a path running in $B_t \cup \Gamma$ connecting a continuous vertex outside B with a discrete vertex in B. We can therefore conclude that B is a strong collection if and only if $\Delta \setminus B$ separates $\Gamma \setminus B$ and $B \cap \Delta$. The latter is then, again by Theorem 2.3, equivalent to $(\Gamma \setminus B) \perp (B \cap \Delta) | \Delta \setminus B[P]$ for every P in $\mathscr{M}(G)$. Finally, using Lemma 4.1, we have:

THEOREM 4.3. *Let $A \subset V$. Then $\mathscr{M}(G)_A$ is a class of CG-distributions if and only if $V \setminus A$ is a strong collection in G.*

COROLLARY 4.4. *Let G be an undirected graph and let A and B be two subsets of V. If $\mathscr{M}(G)_A$ is a class of CG-distributions, then $\mathscr{M}(G_B)_{A \cap B}$ is a class of CG-distributions.*

PROOF. $V \setminus A$ is a strong collection in G implies that $B \setminus A$ is a strong collection in G_B. □

EXAMPLE 4.5 (Example 3.7 continued). $\mathscr{M}(G)_{\{\alpha, \beta\}}$ is a class of CG-distributions and $\mathscr{M}(G)_{\{\gamma, \alpha\}}$ is not a class of CG-distributions.

5. Collapsibility of $\mathscr{M}(G)$. Let again A be a subset of vertices and consider the marginal model $\mathscr{M}(G)_A$. In Section 3 we saw that the Markov properties are preserved under marginalization onto A if and only if $V \setminus A$ is a simplicial collection in G, and from Section 4 we know that the assumption about CG-distributions is preserved if and only if $V \setminus A$ is a strong collection in G. Combining this we see that $\mathscr{M}(G)_A \subset \mathscr{M}(G_A)$ if and only if $V \setminus A$ is a strong and simplicial collection in G. As $\mathscr{M}(G_A)$ always is a subset of $\mathscr{M}(G)_A$, we see that $\mathscr{M}(G)_A \subset \mathscr{M}(G_A)$ is equivalent to $\mathscr{M}(G)_A = \mathscr{M}(G_A)$ and, as in Asmussen and Edwards (1983), we will say that $\mathscr{M}(G)$ is collapsible onto A if this is true.

DEFINITION 5.1. $\mathscr{M}(G)$ (or G) is said to be *collapsible* onto $A \subset V$ if $\mathscr{M}(G)_A = \mathscr{M}(G_A)$.

Before we turn to the main theorem concerning necessary and sufficient conditions for collapsibility, we prove a lemma, which is the keystone in connecting the undirected models treated here and the more general graphical chain models defined in Lauritzen and Wermuth (1984, 1989).

Let $\mathscr{D} = (V_1, V_2, \dots, V_T)$ be a *dependence chain* that is an ordered partition of V into disjoint subsets and let $W_t = \bigcup_{s \le t} V_s$ be the concurrent vertices at time t. P is then said to be pairwise $[G, \mathscr{D}]$-Markovian if for every nonadjacent pair α and β such that $\alpha \in V_t$ and $\beta \in W_t$ we have $\alpha \perp \beta | W_t \setminus \{\alpha, \beta\}$. The *graphical chain model* $\mathscr{M}(G, \mathscr{D})$ specified by the undirected graph G and $\mathscr{D}$ is then given as the class of distributions with the properties: P is $[G, \mathscr{D}]$-Markovian, $p(x) = \prod_t p_{V_t|W_{t-1}}(x_{V_t}|x_{W_{t-1}})$ and $p_{V_t|W_{t-1}}(\cdot|\cdot)$ is a CG-*regression* for all t, that is, $p_{V_t|W_{t-1}}$ is a conditional density derived from some CG-distribution on W_t. Notice that the definition may be phrased as: $p \in \mathscr{M}(G, (V_1, V_2, \dots, V_T))$ if and only if $p_{W_{T-1}} \in \mathscr{M}(G_{W_{T-1}}, (V_1, V_2, \dots, V_{T-1}))$, $p_{V_T|W_{T-1}}$ is a CG-regression, and $\alpha \perp \beta | V \setminus \{\alpha, \beta\}$ if α and β are nonadjacent and at least one of them belongs to V_T.

As an example, let $\mathscr{D} = (A, B)$, where $B = V \setminus A$. In this case $P \in \mathscr{M}(G, \mathscr{D})$ if for all nonadjacent pairs (α, β) we have

$$\alpha \perp \beta | A \setminus \{\alpha, \beta\} \text{ if both } \alpha \text{ and } \beta \text{ belong to } A$$

and

$$\alpha \perp \beta | V \setminus \{\alpha, \beta\} \text{ if at least one of the vertices belongs to } B$$

and the distributional assumptions are that p_A is a CG-density and there exists a CG-density $\tilde{p}$ on V such that $p_{B|A} = \tilde{p}_{B|A} = \tilde{p}/\tilde{p}_A$.

LEMMA 5.2. *Let $A \subset V$ and $B = V \setminus A$. If B is a strong and simplicial collection in G, then $\mathscr{M}(G) = \mathscr{M}(G, \mathscr{D})$ where $\mathscr{D} = (A, B)$.*

PROOF. $\subset$: If $P \in \mathscr{M}(G)$, then $p_{B|A}$ is by definition a CG-regression. Furthermore, if B is a strong collection, then by Theorem 4.3 p_A is a CG-density on A, so the distributional assumptions in $\mathscr{M}(G, \mathscr{D})$ are satisfied. Moreover the $[G, \mathscr{D}]$-Markov properties hold because $\alpha \perp \beta | V \setminus \{\alpha, \beta\}$ for all nonadjacent α and β and P_A is G_A-Markovian, as B is a simplicial collection.

$\supset$: Suppose $p = p_A p_{B|A} \in \mathscr{M}(G, \mathscr{D})$ and let $p_{B|A} = \tilde{p}/\tilde{p}_A$, where $\tilde{p}$ is a CG-density on V. Since the second set of the conditional independence requirements above puts constraints on $p_{B|A}$, we see that $\tilde{p}$ is $\tilde{G}$-Markovian, where $\tilde{G}$ is obtained from G by connecting all vertices in A. Thus we get, as B also is a strong collection in $\tilde{G}$, that $\tilde{p}_A$ is a CG-density and hence $p = p_A \tilde{p}/\tilde{p}_A$ is a CG-density because it is of the right form. To see that P also is G-Markovian, it suffices to show that $\alpha \perp \beta | V \setminus \{\alpha, \beta\}$ for all nonadjacent pairs α and β. We consider three different cases. It is obviously true in the case where α or β belongs to B. If $\alpha \in A \setminus \mathrm{bd}(B)$ and $\beta \notin B$, we first consider $\tilde{P}$. As $\mathrm{bd}(B)$ separates B and $A \setminus \mathrm{bd}(B)$ in $\tilde{G}$, we have that $A \setminus \mathrm{bd}(B) \perp B | \mathrm{bd}(B)$ under $\tilde{P}$, but this conditional independency in fact only concerns $\tilde{p}_{B|A}$, so it must also be true under P. If we combine this with $\alpha \perp \beta | A \setminus \{\alpha, \beta\}$ we get the desired conditional independence. The last case where both α and β belongs to $\mathrm{bd}(B)$ is more tricky. We divide B into B_1 and B_2, where B_1 is the union of all connected components B' of B for which $\alpha \in \mathrm{bd}(B')$ and $B_2 = B \setminus B_1$. As B is a simplicial collection in $\tilde{G}$ we have that $A \setminus \{\alpha, \beta\}$ separates B_1 and β and that $B_1 \cup A \setminus \{\alpha, \beta\}$ separates α and B_2 in $\tilde{G}$. This implies the two conditional independencies $\alpha \perp B_1 | A \setminus \{\alpha, \beta\}$ and $\alpha \perp B_2 | B_1 \cup A \setminus \{\alpha, \beta\}$ are true under $\tilde{P}$ and hence also under P. Again we combine with $\alpha \perp \beta | A \setminus \{\alpha, \beta\}$ and get the factorization of the density:

$$\begin{aligned} p &= p_{A \setminus \{\alpha, \beta\}} p_{\{\alpha, \beta\} | A \setminus \{\alpha, \beta\}} p_{B_1 | A} p_{B_2 | B_1 \cup A} \\ &= p_{A \setminus \{\alpha, \beta\}} p_{\alpha | A \setminus \{\alpha, \beta\}} p_{\beta | A \setminus \{\alpha, \beta\}} p_{B_1 | A \setminus \{\beta\}} p_{B_2 | B_1 \cup A \setminus \{\alpha\}}. \end{aligned}$$

As no factor contains both α and β, we have that $\alpha \perp \beta | V \setminus \{\alpha\}$. □

LEMMA 5.3. *Let $\{P^n\}_{n=1}^{\infty}$ be a sequence of* CG-*distributions on $\mathscr{X}$. If P^n converges weakly to P and P has positive density, then P is a* CG-*distribution on $\mathscr{X}$. If furthermore P^n is G-Markovian for each n, i.e., $\{P^n\}_{n=1}^{\infty} \subset \mathscr{M}(G)$, then P is G-Markovian, i.e., $P \in \mathscr{M}(G)$.*

PROOF. $P^n \to_{n \to \infty}^{w} P$ implies that $P_\Delta^n(i) \to_{n \to \infty} P_\Delta(i)$ for each $i \in I$ and that the conditional distribution of Y given $I = i$ under P^n converges weakly to that under P, for each i such that $P_\Delta(i) > 0$. The latter again implies that the conditional distribution of Y given $I = i$ under P is Gaussian if $P_\Delta(i) > 0$. Now, if P has positive density, then $P_\Delta(i) > 0$ for every $i \in I$, and the conditional distribution of Y given $I = i$ must be regular Gaussian, i.e., P is a CG-distribution. It can also be shown that for every $A \subset V$ the density $\tilde{p}_A$, given by $\tilde{p}_A(x_A) = \lim_{n \to \infty} p_A^n(x_A)$, is a density for P_A. So if $A \perp B | C\ [P^n]$, that is

$$p_{A \cup B \cup C}^n(x_A, x_B, x_C) = p_{A \cup C}^n(x_A, x_C) p_{B \cup C}^n(x_B, x_C) / p_C^n(x_C),$$

then

$$\tilde{p}_{A \cup B \cup C}(x_A, x_B, x_C) = \tilde{p}_{A \cup C}(x_A, x_C) \tilde{p}_{B \cup C}(x_B, x_C) / \tilde{p}_C(x_C),$$

or equivalently $A \perp B | C\ [P]$. Thus the last statement in the lemma is shown. □

Now we can formulate the main theorem, recollecting that $\hat{P}_{[A]}$ denotes the maximum likelihood estimate in $\mathscr{M}(G_A)$.

THEOREM 5.4. *Let $G = (V, E)$ be an undirected graph and let $A \subset V$. Then the following are equivalent*:

(i) *G is collapsible onto A,* i.e., $\mathscr{M}(G)_A = \mathscr{M}(G_A)$.
(ii) *$V \setminus A$ is a strong and simplicial collection in G.*
(iii) (a) *If $A_1 \perp A_2 | C\ [\mathscr{M}(G)]$ and $A_1, A_2 \subset A$, then $A_1 \perp A_2 | C \cap A\ [\mathscr{M}(G)]$ and* (b) $\Delta \setminus A \perp \Gamma \cap A | \Delta \cap A\ [\mathscr{M}(G)]$.
(iv) *$\mathscr{M}(G)$ is equal to the chain model $\mathscr{M}(G, \mathscr{D})$, where $V_1 = A$, $V_2 = V \setminus A$ and $\mathscr{D} = \{V_1, V_2\}$.*
(v) *X_A is a cut in $\mathscr{M}(G)$.*
(vi) *If $\hat{P}$ exists, then $\hat{P}_A = \hat{P}_{[A]}$.*

PROOF. The theorem will be proved as follows (i) $\Leftrightarrow$ (ii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (i), (ii) $\Leftrightarrow$ (iii) and [(i) and (v)] $\Rightarrow$ (vi) $\Rightarrow$ (i).

Using Corollary 3.4 and Theorem 4.3 we see that (i) $\Leftrightarrow$ (ii). Lemma 5.2 yields (ii) $\Rightarrow$ (iv) and by the recursive definition of the chain models we see that (iv) implies (v). So now suppose X_A is a cut in $\mathscr{M}(G)$, i.e., $\mathscr{M}(G) \approx \mathscr{M}(G)_A \times \mathscr{M}(G)^A$. We know that $P^0 \in \mathscr{M}(G)$ and as a consequence $P^0_{V \setminus A} \in \mathscr{M}(G)^A$. This implies that $P_A \in \mathscr{M}(G)_A$ if and only if $p_A(x_A) \cdot p^0_{V \setminus A}(x_{V \setminus A})$ belongs to $\mathscr{M}(G)$, which is easily seen to be true if and only if P_A is a G_A-Markovian CG-distribution. This proves (v) $\Rightarrow$ (i). The equivalence of (ii) and (iii) is a consequence of the equivalence of separation in G and conditional independence under $\mathscr{M}(G)$; the details are given in Sections 3 and 4. The only thing left is to show [(i) and (v)] $\Rightarrow$ (vi) $\Rightarrow$ (i).

First assume that X_A is a cut in $\mathscr{M}(G)$ and that $\mathscr{M}(G)_A = \mathscr{M}(G_A)$. Then the likelihood function factorizes as $L(p) = L_A(p_A) \cdot L^A(p^A)$, where L_A is the marginal likelihood function for $\mathscr{M}(G_A)$ and L^A is a conditional likelihood function. So if $\hat{p}$ maximizes $L(\cdot)$, then, due to the variation independence of p_A and p^A, $\hat{p}_A$ will maximize L_A, i.e., $\hat{P}_A = \hat{P}_{[A]}$ and [(i) and (v)] $\Rightarrow$ (vi) is shown.

Finally, assume (vi) is true and let $P \in \mathscr{M}(G)$. If we can show that $P_A \in \mathscr{M}(G_A)$, we are finished. Due to the fact that $\mathscr{M}(G)$ is a regular exponential family [see, e.g., Barndorff-Nielsen (1978)], there exists a sequence $\{x^{(\nu)}\}_{\nu=1}^{\infty} \subset \mathscr{X}^{\infty}$ and an n_0 such that the maximum likelihood estimate $\hat{P}^{(n)}$ based on $\{x^{(\nu)}\}_{\nu=1}^{n}$ exists for $n \geq n_0$ and $\hat{P}^{(n)}$ converges weakly to P [see Andersen (1969)]. As a consequence $\hat{P}_A^{(n)} = \hat{P}_{[A]}^{(n)}$ converges weakly to P_A, which has a positive density, and using Lemma 5.3 on $\{\hat{P}^n_{[A]}\}$ and P_A we see that $P_A \in \mathscr{M}(G_A)$. This concludes the proof. $\square$

If C is a complete set, then the question whether G is collapsible onto C is only the question whether $\mathscr{M}(G)_C$ is a class of CG-distributions, because G_C does not prescribe any conditional independencies. So in the pure cases, i.e., $\Gamma = \emptyset$ or $\Delta = \emptyset$, where the distributional type is preserved under marginalization, $\mathscr{M}(G)$ is collapsible onto all the complete subsets. In the mixed case this is not true in general. If for instance G is connected and C only contains continuous vertices, then $V \setminus C$ is not a strong and simplicial collection, implying that $\mathscr{M}(G)_C$ is not a class of CG-distributions on $\mathscr{X}_C$ (multivariate Gaussian distributions).

COROLLARY 5.5. *Let A and B be subsets of V. If G is collapsible onto A, then G_B is collapsible onto $A \cap B$.*

PROOF. It is easily checked that if $V \setminus A$ is a strong and simplicial collection in G, then $B \setminus A$ is a strong and simplicial collection in G_B. $\square$

The equivalence of (ii) and (iv) in Theorem 5.4 shows that the sufficient condition given in Lemma 5.2 for equivalence of an undirected model and a chain model with two chain elements is also necessary. Using the recursive definition of the graphical chain model, an induction argument yields the following general results, first proved in Lauritzen and Wermuth (1984).

PROPOSITION 5.6. *Let $\mathscr{D} = \{V_t\}_{t=1}^T$ be a dependence chain. Then $\mathscr{M}(G) = \mathscr{M}(G, \mathscr{D})$ if and only if for every t, V_t is a strong and simplicial collection in G_{W_t}.*

One of the many consequences of collapsibility is that it leads to a factorization of the maximum likelihood estimate. This is due to a close connection between the concept of strong decomposition of G and strong and simplicial collections in G.

DEFINITION 5.7 [Leimer (1989)]. A triplet (A, B, C) of disjoint subsets of V is called a *decomposition* of G if

(i) $V = A \cup B \cup C$.
(ii) C separates A and B in G.
(iii) C is complete.

It is called a *strong decomposition* if $B \subset \Gamma$ or $C \subset \Delta$.

THEOREM 5.8 [Frydenberg and Lauritzen (1989)]. *If (A, B, C) is a strong decomposition of G, then $\hat{P}$ exists if and only if $\hat{P}_{[A \cup C]}$ and $\hat{P}_{[B \cup C]}$ both exist, and in that case*

$$\hat{p}(x) = \frac{\hat{p}_{[A \cup C]}(x_{A \cup C}) \cdot \hat{p}_{[B \cup C]}(x_{B \cup C})}{\hat{p}_{[C]}(x_C)}.$$

Now let $A \subset V$ and let $B_1, \dots, B_T$ be the connected components of $V \setminus A$. We see that $V \setminus A$ is a strong and simplicial collection in G if and only if $(V \setminus \mathrm{cl}(B_t), B_t, \mathrm{bd}(B_t))$ is a strong decomposition for each $1 \le t \le T$.

THEOREM 5.9. *If G is collapsible onto A, then*

$$\hat{P} \text{ exists if and only if } \hat{P}_{[A]} \text{ and } \hat{P}_{[\mathrm{cl}(B_t)]},\ 1 \le t \le T \text{ exist}$$

and in that case

$$\hat{p}(x) = \hat{p}_{[A]}(x_A) \cdot \prod_{t=1}^{T} \frac{\hat{p}_{[\mathrm{cl}(B_t)]}(x_{\mathrm{cl}(B_t)})}{\hat{p}_{[\mathrm{bd}(B_t)]}(x_{\mathrm{bd}(B_t)})}.$$

PROOF. Setting $V_0 = V$ and $V_t = V_{t-1} \setminus B_t$ for $1 \le t \le T$, it can be seen that $(V_t \setminus \mathrm{bd}(B_t), B_t, \mathrm{bd}(B_t))$ is a strong decomposition of $G_{V_{t-1}}$. This implies that

$$\hat{P}_{[V_{t-1}]} \text{ exists if and only if } \hat{P}_{[V_t]} \text{ and } \hat{P}_{[\mathrm{cl}(B_t)]} \text{ exist}$$

and that $\hat{p}_{[V_{t-1}]} = \hat{p}_{[V_t]} \cdot (\hat{p}_{[\mathrm{cl}(B_t)]}/\hat{p}_{[\mathrm{bd}(B_t)]})$. Using this recursively and, finally, noting that $V_T = A$, gives the wanted results. □

EXAMPLE 5.10 (Example 3.7 continued). $\mathscr{M}(G)$ is not collapsible onto $\{\gamma, \alpha)$ or $\{\delta, \nu\}$ but onto $\{\alpha, \beta\}$. Using Theorem 5.9 we have that $\hat{p}$ exists if and only if $\hat{p}_{[\gamma, \alpha]}$, $\hat{p}_{[\alpha, \beta]}$, $\hat{p}_{[\alpha, \delta]}$ and $\hat{p}_{[\beta, \nu]}$ exist, and that

$$\hat{p}(i_\nu, i_\beta, y_\delta, y_\alpha, y_\gamma) = \frac{\hat{p}_{[\gamma, \alpha]}(y_\gamma, y_\alpha)\hat{p}_{[\alpha, \beta]}(y_\alpha, i_\beta)\hat{p}_{[\alpha, \delta]}(y_\alpha, y_\delta)\hat{p}_{[\beta, \nu]}(i_\beta, i_\nu)}{\hat{p}_{[\alpha]}(y_\alpha)^2 \hat{p}_{[\beta]}(i_\beta)}.$$

One should note that the questions of existence and calculation of the maximum likelihood estimate is reduced to the simple question of maximum likelihood estimation in four saturated models.

The result in Theorem 5.9 can be used to show that the likelihood ratio test for removing some edges in G can sometimes be performed in marginal models. A discussion of this in the discrete case is given in Asmussen and Edwards (1983).

Finally, it should be noted that all the stated results hold, if we only consider the *homogeneous graphical models*, i.e., the class of G-Markovian homogeneous CG-distributions. This follows because we have proved that this class is Markov perfect (see the remark after Theorem 2.3) and the fact that if the marginal distribution of a homogeneous CG-distribution is a CG-distribution, then it is homogeneous too.

Acknowledgments. I would like to thank Professor Ole E. Barndorff-Nielsen and Professor Steffen L. Lauritzen for their discussion of this paper, which essentially is based upon material in Frydenberg (1986).

REFERENCES

ANDERSEN, A. H. (1969). Asymptotic results for exponential families. *Bull. Inst. Internat. Statist.* **43** 241–242.

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

BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

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

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

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

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

EDWARDS, D. (1990). Hierarchical interaction models. *J. Roy. Statist. Soc. Ser. B.* To appear.

FRYDENBERG, M. (1986). Blandede interaktionsmodeller, kausale modeller, kollapsibilitet og estimation. Thesis, Statistiske Interna nr. 42, Aarhus Univ. (in Danish).

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

LAURITZEN, S. L. and 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. and WERMUTH, N. (1984). Mixed interaction models. Report R 84-8, Inst. of Elec. Syst., Univ. Aalborg.

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

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

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, Calif.

PORTEOUS, B. T. (1985). Properties of log-linear and covariance selection models. Ph.D. thesis, Cambridge Univ.

SHACHTER, R. D. (1986). Evaluating influence diagrams. *Oper. Res.* **34** 871–882.

WERMUTH, N. and LAURITZEN, S. L. (1990). On the interpretation and analysis of data using conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc. Ser. B.* To appear.

YAKOWITZ, S. J. and SPRAGINS, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39** 209–214.

DEPARTMENT OF THEORETICAL STATISTICS
INSTITUTE OF MATHEMATICS
UNIVERSITY OF AARHUS
DK-8000 AARHUS C
DENMARK