Independence Properties of Directed Markov Fields

S. L. Lauritzen
Aalborg University, Aalborg, Denmark
A. P. Dawid
University College London, London, United Kingdom
B. N. Larsen
A/S Dumex, Copenhagen, Denmark
H.-G. Leimer
Hoechst AG, Frankfurt am Main, Federal Republic of Germany

We investigate directed Markov fields over finite graphs without positivity assumptions on the densities involved. A criterion for conditional independence of two groups of variables given a third is given and named as the directed, global Markov property. We give a simple proof of the fact that the directed, local Markov property and directed, global Markov property are equivalent and — in the case of absolute continuity w.r.t. a product measure — equivalent to the recursive factorization of densities. It is argued that our criterion is easy to use, it is sharper than that given by Kiiveri, Speed, and Carlin and equivalent to that of Pearl. It follows that our criterion cannot be sharpened.

1. INTRODUCTION

The present paper has been motivated partly by attempting to give simple rules for interpreting recursive models such as studied, for example, in [6, 15, 16], and partly by recent developments in expert system research, see, for example, [7, 9], where it also is of crucial importance to have effective criteria for deciding which information or facts bear relevance upon others in the light of present knowledge.

Models of this type have been known for a long time, in connection with path analysis, see, for example, [17]. We shall refer the reader to Kiiveri et al. [6] for a small survey. This reference also gives a basic discussion of these Markov properties, but it becomes apparent that the results are incomplete on some points that are critical to the situations mentioned:

- The restrictive assumption of existence and positivity of densities is, as we shall see, unnecessary.

NETWORKS, Vol. 20 (1990) 491–505
© 1990 John Wiley & Sons, Inc.
CCC 0028-3045/90/050491-015$04.00

- The criterion Kiiveri et al. named the recursive, global Markov property is not strong enough to read off *all* conditional independencies directly.
- The criterion is complicated to use, and the proof of its correctness is difficult.

The purpose of the present paper is to remedy these shortcomings. We first give some graphtheoretic terminology before we review a few basic and known facts about Markov fields over undirected graphs. Then we show the main result, using the trick of forming the moral graph of a directed, acyclic graph, described by Lauritzen and Spiegelhalter [7]. Further, we compare our criterion to that of Kiiveri et al. [6] and Pearl [9], showing that the latter is equivalent to ours, although our criterion can be considerably easier to check in some cases. Finally, the basic result is extended to the case where existence of densities is not assumed.

2. GRAPHTHEORETIC TERMINOLOGY

A *graph*, as we use it throughout this paper, is a pair $\mathcal{G} = (V,E)$, where V is a finite set of *vertices* and the set of *edges* E is a subset of the set $V \times V$ of ordered pairs of distinct vertices. Thus, our graphs are *simple*, i.e., there are no multiple edges and they have no loops.

Edges $(\alpha,\beta) \in E$ with both (α,β) and (β,α) in E are called *undirected*, whereas an edge (α,β) with its opposite (β,α) not being contained in E is called *directed*.

A graph is represented by a picture, using a dot for a vertex, a *line* joining α to β for an undirected edge, whereas an *arrow* from α, pointing toward β, is used for a directed edge (α,β) with $(\beta,\alpha) \notin E$. If the graph has only undirected edges (lines), it is an *undirected* graph, and if all edges are directed (arrows), the graph is said to be *directed*. Our graphs are either directed or undirected.

We use the notation

$$\alpha \rightarrow \beta \quad \alpha \sim \beta$$
$$\alpha \not\rightarrow \beta \quad \alpha \not\sim \beta$$

to signify that

$$(\alpha,\beta) \in E \qquad (\alpha,\beta) \in E \wedge (\beta,\alpha) \in E$$
$$(\alpha,\beta) \notin E \qquad (\alpha,\beta) \notin E \wedge (\beta,\alpha) \notin E\,.$$

Note that $\alpha \not\sim \beta$ then means that there is no arrow between α and β either and that, for example, $\alpha \rightarrow \beta \wedge \beta \rightarrow \alpha \Leftrightarrow \alpha \sim \beta$.

The undirected graph obtained from $\mathcal{G}$ by substituting lines for arrows everywhere is denoted by $\mathcal{G}^{\sim}$ and called the *undirected graph corresponding to* $\mathcal{G}$.

If $A \subseteq V$ is a subset of the vertex set, it induces a subgraph $\mathcal{G}_A = (A,E_A)$, where the edge set $E_A = E \cap (A \times A)$ is obtained from $\mathcal{G}$ by keeping edges with both endpoints in A.

A graph is *complete* if all vertices are joined by an arrow or a line. A subset is *complete* if it induces a complete subgraph. A complete subset that is maximal (w.r.t. $\subseteq$) is called a *clique*.

If $\alpha \to \beta$, α is said to be a *parent* of β and β a *child* of α. The set of parents of β is denoted by $\mathrm{pa}(\beta)$ and the set of children of α by $\mathrm{ch}(\alpha)$.

If $\alpha \sim \beta$, α and β are said to be *adjacent* or *neighbors*. The *boundary* $\mathrm{bd}(A)$ of a subset A of vertices is the set of vertices in $V \backslash A$ that are neighbors to vertices in A. The *closure* of A is $\mathrm{cl}(A) = A \cup \mathrm{bd}(A)$.

A *path* of length n from α to β is a sequence $\alpha = \alpha_0, \ldots, \alpha_n = \beta$ of distinct vertices such that $\alpha_{i-1} \to \alpha_i$ for all $i = 1, \ldots, n$. If there is a path from α to β, we say that α *leads to* β and write $\alpha \mapsto \beta$. The *descendants* $\mathrm{de}(\alpha)$ of α are the vertices β such that $\alpha \mapsto \beta$. The *nondescendants* are $\mathrm{nd}(\alpha) = V \backslash (\mathrm{de}(\alpha) \cup \{\alpha\})$. The vertices α that lead to β are the *ancestors* of β, denoted by $\mathrm{an}(\beta)$.

A subset $A \subseteq V$ is an *ancestral set* if it contains its own ancestors, i.e., if

$$\mathrm{an}(\alpha) \subseteq A \quad \text{for all} \quad \alpha \in A.$$

Clearly, if A and B are ancestral sets, so is $A \cap B$, and, therefore, we can to any subset A of V assign the set $\mathrm{An}(A)$ as the smallest ancestral set containing A. Indeed, $\mathrm{An}(A) = A \cup (\cup_{\alpha \in A} \mathrm{an}(\alpha))$.

A vertex α of a directed graph is said to be *terminal* if it has no children.

A *chain* of length n from α to β is a sequence $\alpha = \alpha_0, \ldots, \alpha_n = \beta$ of distinct vertices such that $\alpha_{i-1} \to \alpha_i$ or $\alpha_i \to \alpha_{i-1}$ for all $i = 1, \ldots, n$.

A subset S is said to *separate A from B* if all chains from vertices $\alpha \in A$ to $\beta \in B$ intersect S.

A *cycle* is a path with the modification that $\alpha = \beta$, i.e., it begins and ends in the same point. A graph is *acyclic* if it contains no cycles. An easy argument yields that *a directed, acyclic graph has at least one terminal vertix.*

For a directed acyclic graph $\mathcal{G}$, we define as in [7] its *moral graph* $\mathcal{G}^m$ as the undirected graph with the same vertex set but with α and β adjacent in $\mathcal{G}^m$ if and only if either $\alpha \to \beta$ or $\beta \to \alpha$ or if there is a γ such that $\alpha \to \gamma$ and $\beta \to \gamma$. In other words, the moral graph is obtained from the original graph by "marrying parents" with a common child and then dropping directions on arrows.

3. MARKOV FIELDS OVER UNDIRECTED GRAPHS

For an undirected graph $\mathcal{G} = (V, E)$ as in Section 2, we consider a random field with V as index set, i.e., a collection of random variables $(X_v), v \in V$ taking values in probability spaces $\mathcal{X}_v$, $v \in V$. The probability spaces can be quite general, just sufficiently well behaved to ensure existence of regular conditional probabilities. For A being a subset of V, we let $\mathcal{X}_A = \times_{v \in A} \mathcal{X}_v$ and further $\mathcal{X} = \mathcal{X}_V$. Typical elements of $\mathcal{X}_A$ are denoted by $x_A = (x_v)_{v \in A}$ and so on. Then, a probability measure P on $\mathcal{X}$ is said to *factorize* according to $\mathcal{G}$ if there exists nonnegative functions ψ_A defined on $\mathscr{X}_\alpha$ for only complete subsets A, and a product measure $\mu = \otimes_{v \in V} \mu_v$ on $\mathscr{X}$, such that

$$P = f \cdot \mu, \quad \text{where} \quad f(x) = \prod_A \psi_A(x_A).$$

The functions ψ_A are referred to as *factor potentials* of P. These are not uniquely determined. There is arbitrariness in the choice of μ, but also groups of functions ψ_A can be multiplied together or split up in different ways. In fact, one can without loss of generality assume — although this is not always practical — that only cliques (maximal complete subsets) appear as the sets A, i.e., that

$$f = \prod_{C \in \mathscr{C}} \psi_C, \tag{1}$$

where $\mathscr{C}$ is the set of cliques of $\mathscr{G}$.

Associated with this structure there is a range of Markov properties that, in general, can be different. A probability measure P on $\mathscr{X}$ is said to obey

(L) *the local Markov property*, relative to $\mathscr{G}$, if for any vertex $\alpha \in V$,

$$\alpha \perp\!\!\!\perp V \setminus \mathrm{cl}(\alpha) \mid \mathrm{bd}(\alpha);$$

(P) *the pairwise Markov property*, relative to $\mathscr{G}$, if for any pair(α,β) of nonadjacent vertices.

$$\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha,\beta\};$$

(*G*) *the global Markov property*, relative to $\mathscr{G}$, if for any triple (A,B,S) of disjoint subsets of V such that S separates A from B in $\mathscr{G}$,

$$A \perp\!\!\!\perp B \mid S.$$

Here we have used the notation "$A \perp\!\!\!\perp B \mid S$", etc., as short for "$X_A$ is conditionally independent of X_B given X_S". See [1,2] for detailed properties of conditional independence.

It is obvious that we have

$$(\mathrm{G}) \Rightarrow (\mathrm{L}) \Rightarrow (\mathrm{P}), \tag{2}$$

but, in general, the three properties are different, see [13] for discussion. In fact, if P admits a density w.r.t. μ that is strictly positive, the properties are equivalent, as shown, for example, in [10]. We shall not be concerned so much with this but rather with the following.

Proposition 1. *If P factorizes according to $\mathscr{G}$, it also obeys the global Markov property.*

Proof. From standard properties of conditional independence, it is enough to show this under the assumption that $A \cup B \cup S = V$. The cliques can be partitioned into those that have nonempty intersection with $A, \mathscr{C}_A$ say, and the remaining cliques. But since S separates A from B, we must have $B \cap C = \emptyset$ for all C in $\mathscr{C}_A$. From the factorization (1), we obtain that

$$f(x) = \prod_{C \in \mathscr{C}_A} \psi_C(x_C) \prod_{C \in \mathscr{C} \setminus \mathscr{C}_A} \psi_C(x_C) = f_1(x_{A \cup S}) f_2(x_{B \cup S}),$$

whereby the result follows. ■

The global Markov property (G) is important because it gives a general criterion for deciding when two groups of variables A and B are conditionally independent given a third group of variables S. This criterion can, apart from degenerate cases, not be improved, see, for example, [4], in the sense that if A and B are not separated by S, then there will be factorizing probabilities P, such that $A \perp\!\!\!\perp B | C$ will not hold.

In the special case where P admits a strictly positive density w.r.t. μ, each of the equivalent Markov properties imply that P factorizes, a result known as the Hammersley–Clifford theorem (see [13] for a discussion of the problems involved).

In this article, we shall primarily be concerned with factorizing probability measures, and Proposition 1 combined with (2) says that the factorization is the strongest of all the requirements.

4. MARKOV FIELDS OVER DIRECTED GRAPHS

Consider the same setup as in the previous section except that now the graph $\mathscr{G}$ is assumed to be directed and acyclic.

We say that P admits a *recursive factorization* according to $\mathscr{G}$, if there exist nonnegative functions, henceforth referred to as *kernels*, $k^\nu(\cdot,\cdot)$, $\nu \in V$ defined on $\mathscr{X}_\nu \times \mathscr{X}_{\mathrm{pa}(\nu)}$ such that

$$\int k^\nu(y_\nu, x_{\mathrm{pa}(\nu)}) \mu_\nu(dy_\nu) = 1$$

and

$$P = f \cdot \mu, \quad \text{where} \quad f(x) = \prod_{\nu \in V} k^\nu(x_\nu, x_{\mathrm{pa}(\nu)}).$$

It is an easy induction argument to show that if P admits a recursive factorization as above, then the kernels $k^\nu(\cdot, x_{\mathrm{pa}(\nu)})$ are, in fact, densities for the conditional distribution of X_ν, given $X_{\mathrm{pa}(\nu)} = x_{\mathrm{pa}(\nu)}$. Also it is immediate, as was noted in [7], that if we form the, undirected, moral graph $\mathscr{G}^m$ (marrying parents and dropping directions) such as described in Section, 2, we obtain:

Lemma 1. *If P admits a recursive factorization according to the directed, acyclic graph $\mathcal{G}$, it factorizes according to the moral graph $\mathcal{G}^m$ and obeys therefore the global Markov property relative to $\mathcal{G}^m$.*

Proof. The factorization follows from the fact that, by construction, the sets $\{v\} \cup \mathrm{pa}(v)$ are complete in $\mathcal{G}^m$ and we can therefore let $\psi_{\{v\}\cup \mathrm{pa}(v)} = k^v$. The remaining part of the statement follows from Proposition 1. ■

Further, we have

Lemma 2. *If P admits a recursive factorization according to $\mathcal{G}$ and A is an ancestral set in $\mathcal{G}$, the marginal distribution P_A of X_A admits a recursive factorization according to the subgraph $\mathcal{G}_A$.*

Proof. The result follows directly from the recursive factorization, integrating out the variables that are not in A. ■

As a direct consequence:

Corollary 1. *Let P factorize recursively according to $\mathcal{G}$. Then*

$$A \perp\!\!\!\perp B \mid S$$

whenever A and B are separated by S in $(\mathcal{G}_{\mathrm{An}(A\cup B\cup S)})^m$, the moral graph of the smallest ancestral set containing $A \cup B \cup S$.

Proof. The result follows immediately from Lemma 2 and Lemma 1. ■

The property in Corollary 1 shall be referred to as the *directed global Markov property* (DG). Note that our terminology at this point differs from that used by Kiiveri et al. [6]. Further, we say that P obeys the *directed local Markov property* (DL) if any variable is conditionally independent of its nondescendants, given its parents:

$$v \perp\!\!\!\perp (\mathrm{nd}(v)\backslash \mathrm{pa}(v)) \mid \mathrm{pa}(v).$$

The property that P admits a recursive factorization will similarly be denoted by (DF). In contrast to the undirected case, we have that all the three properties (DF), (DL), and (DG) are equivalent just assuming existence of the density f, stated formally as

Theorem 1. *Let $\mathcal{G}$ be a directed, acyclic graph. For a probability measure P on $\mathcal{X}$ that is absolutely continuous w.r.t. a product measure μ, the following conditions are equivalent:*

(DF) *P admits a recursive factorization according to $\mathcal{G}$.*

(DG) *P obeys the global directed Markov property, relative to* $\mathcal{G}$.
(DL) *P obeys the local directed Markov property, relative to* $\mathcal{G}$.

Proof. That (DF) implies (DG) is Corollary 1. That (DG) implies (DL) follows by observing that $\{v\} \cup \text{nd}(v)$ is an ancestral set and that $\text{pa}(v)$ obviously separates $\{v\}$ from $\text{nd}(v)\backslash\text{pa}(v)$ in $(\mathcal{G}_{\{v\}\cup\text{nd}(v)})^m$. The final implication is shown by induction on the number of vertices $|V|$ of $\mathcal{G}$. Let v_0 be a terminal vertex of $\mathcal{G}$. Then, we can let k^{v_0} be the conditional density of X_{v_0}, given $X_{V\backslash\{v_0\}}$, which by (DL) can be chosen to depend on $x_{\text{pa}(v_0)}$ only. The marginal distribution of $X_{V\backslash\{v_0\}}$ trivially obeys the directed local Markov property and admits a factorization by the inductive assumption. Combining this factorization with k^{v_0} yields the factorization for P. This completes the proof. ∎

Since the three conditions in Theorem 1 are equivalent, it makes sense to speak of a *directed Markov field* as one where any of the conditions is satisfied.

The directed global Markov property (DG) gives a simple and efficient criterion for deciding when two groups of variables are conditionally independent given a third. The procedure is illustrated in the following.

Example 1. Consider a directed Markov field on the graph in Figure 1 and the problem of deciding, whether $a \perp\!\!\!\perp b \mid S$? The moral graph of the smallest ancestral set containing all the variables involved is shown in Figure 2. Obviously, S separates a from b in this graph, implying $a \perp\!\!\!\perp b|S$. ∎

5. RELATED SEPARATION RESULTS

The directed global Markov property as defined by Kiiveri et al. [6] involves the notion of sets (A,B,S) being in *configuration* [>] in a directed graph $\mathcal{G}$, if there is a chain in $\mathcal{G}$ from an element $a \in A$ to an element $b \in B$ including three vertices (i,j,k) with $i,j \notin S$, $k \in S$ and $i \nrightarrow j$. It also involves the subgraphs $\mathcal{G}_{\text{nd}(\alpha)\cup\{\alpha\}}$ for $\alpha \in V$ and their corresponding undirected counterparts $\mathcal{G}_{\widetilde{\text{nd}}(\alpha)\cup\{\alpha\}}$ (where just directions have been ignored). The main theorem of [6] states that if A, B, and S are subsets of V such that there exists an $\alpha \in V$ with all conditions below satisfied

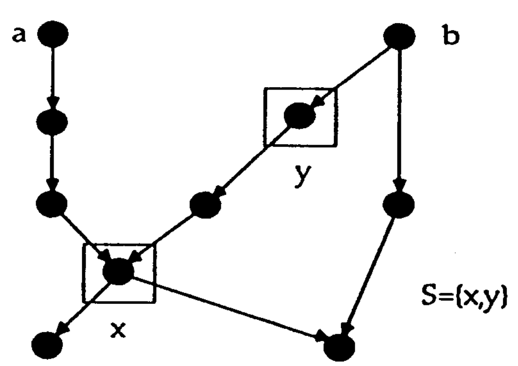

FIG. 1. The directed, global Markov property. Is $a \perp\!\!\!\perp b \mid S$?

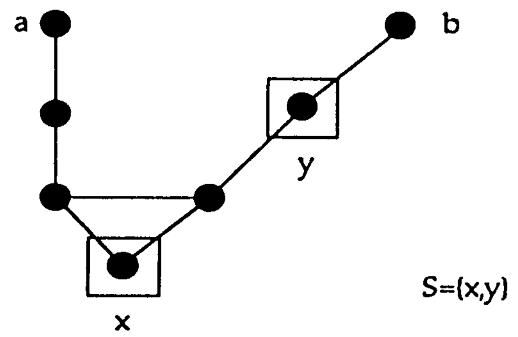

FIG. 2. The moral graph of the smallest ancestral set containing $\{a\} \cup \{b\} \cup S$.

- $A \cup B \cup S \subseteq \mathrm{nd}(\alpha) \cup \{\alpha\}$
- (A, B, S) are not in configuration $[>]$ in $\mathcal{G}_{\mathrm{nd}(\alpha)\cup\{\alpha\}}$
- A and B are separated by S in $\mathcal{G}^{\sim}_{\mathrm{nd}(\alpha)\cup\{\alpha\}}$,

then $A \perp\!\!\!\perp B | S$. The criterion leaves the question posed in Example 1 undecided.

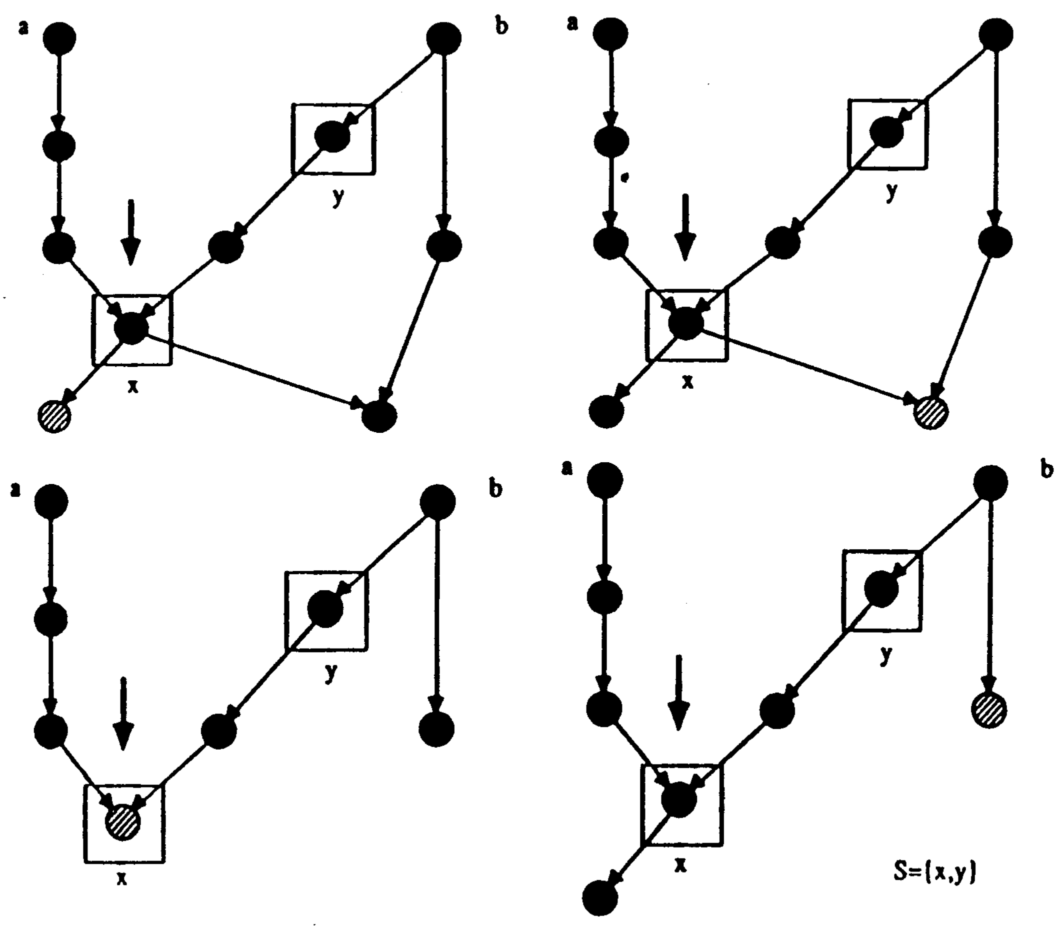

FIG. 3. All subgraphs of the form $\mathcal{G}_{\mathrm{nd}(\alpha)\cup\{\alpha\}}$ that contain $(\{a\},\{b\},S)$. The shaded vertex is α and the arrow indicates the critical configuration.

In this example, ($\{a\}$, $\{b\}$, S) are in configuration [>] in all the relevant subgraphs as illustrated in Figure 3. It seems plausible that it be possible to give a minor modification of this criterion, such as to enable it to disclose the conditional independence present in our example, but we do not think it worthwhile in the light of the simplicity of our criterion and the following argument that shows that, *when the above criterion is satisfied, so is ours*. Let $\mathcal{G}_0 = \mathcal{G}_{\mathrm{An}(A \cup B \cup S)}$. We show that both of the following:

(a) S separates A and B in $\mathcal{G}_0$
(b) S does not separate A and B in $\mathcal{G}_0^m$,

together imply that (A, B, S) are in configuration [>] in $\mathcal{G}_0$.

If (b) holds, we can find a chain in $\mathcal{G}_0^m$ linking A and B and avoiding S. By (a), this must contain at least one link formed from a marriage in $\mathcal{G}_0$, i.e., $i \sim j$ in $\mathcal{G}_0^m$, where in $\mathcal{G}_0$ we have $i \rightarrow k$, $j \rightarrow k$, and $i \not\sim j$. For each such link, substitute the double link $i \rightarrow k \leftarrow j$, whereby we obtain a chain linking A and B in $\mathcal{G}_0$. By (a) it must intersect S, but can only do so at one of the new vertices k introduced. This then yields (i,j,k) in configuration [>].

It should be mentioned that the graphs considered by Kiiveri et al. [6] are more general than ours in that they can comprise exogenous variables. However, as reported by Frydenberg [3], the method of constructing moralized ancestral sets extends to an even more general class of graphs, the so-called *chain graphs*, such graphs being the basis for the graphical chain models of Lauritzen and Wermuth [8].

Another criterion was given by Pearl [9]. A chain π from a to b in a directed, acyclic graph $\mathcal{G}$ is said to be *blocked* by S, if it contains a vertex $\gamma \in \pi$ such that either

- $\gamma \in S$ and arrows of π do not meet head to head at γ, or
- $\gamma \notin S$, nor has γ any descendants in S, and arrows of π do meet head to head at γ.

A chain that is not blocked by S is said to be *active*. Two subsets A and B are now said to be *d-separated* by S if all chains from A to B are blocked by S. We then have

Proposition 2. *If A and B are d-separated by S, $A \perp\!\!\!\perp B \mid S$.*

Proof. The result was stated without proof in [9] and later proved by Verma [14]; see also [11]. It is an immediate consequence of Proposition 3 to be given below. ■

We illustrate the notion by applying it to the query of our example. As Figure 4 indicates, all chains between a and b are blocked by S, whereby Proposition 2 gives that $a \mathcal{G} b \mid S$.

It turns out that the criterion used in connection with our formulation of the

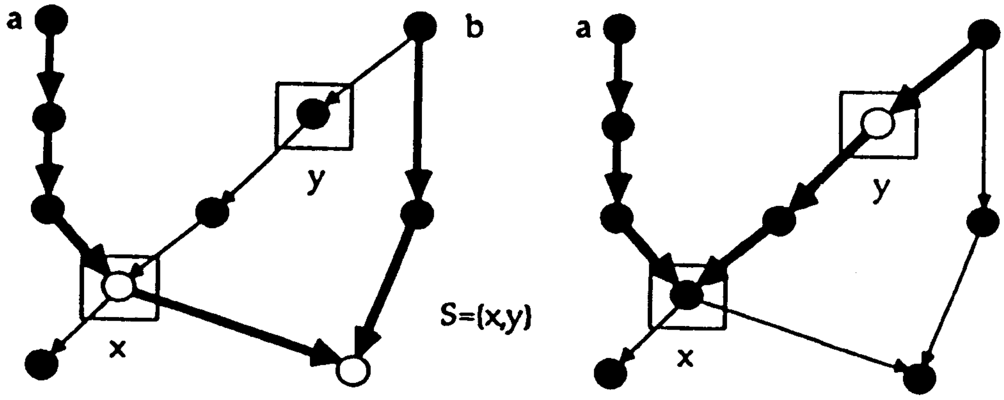

FIG. 4. Illustration of Pearl's separation criterion. There are two chains from a to b, drawn with fat lines. Both are blocked, but different vertices γ, indicated with open circles, play the role of blocking vertices.

directed, global Markov property is equivalent to that of Pearl. More precisely, we have

Proposition 3. *Let A, B, and S be disjoint subsets of a directed, acyclic graph $\mathcal{G}$. Then, S d-separates A from B if and only if S separates A from B in $(\mathcal{G}_{\mathrm{An}(A \cup B \cup S)})^m$.*

Proof. Suppose S does not d-separate A from B. Then, there is an active chain from A to B such as, for example, indicated in Figure 5. All vertices in this chain must lie within $\mathrm{An}(A \cup B \cup S)$. Because if the arrows meet head to head at some vertex γ, either $\gamma \in S$ or γ has descendants in S. If not, either of the subpaths away from γ either meets another arrow, in which case γ has descendants in S, or leads all the way to A or B. Each of these head-to-head meetings will give rise to a marriage in the moral graph such as illustrated in Figure 6, thereby creating a chain from A to B in $(\mathcal{G}_{\mathrm{An}(A \cup B \cup S)})^m$, circumventing S.

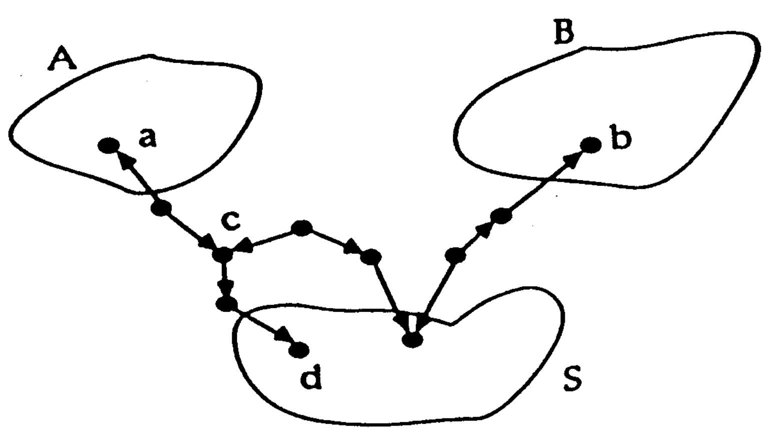

FIG. 5. Example of an active chain from A to B. The path from c to d is not part of the chain, but indicates that c must have descendants in S.

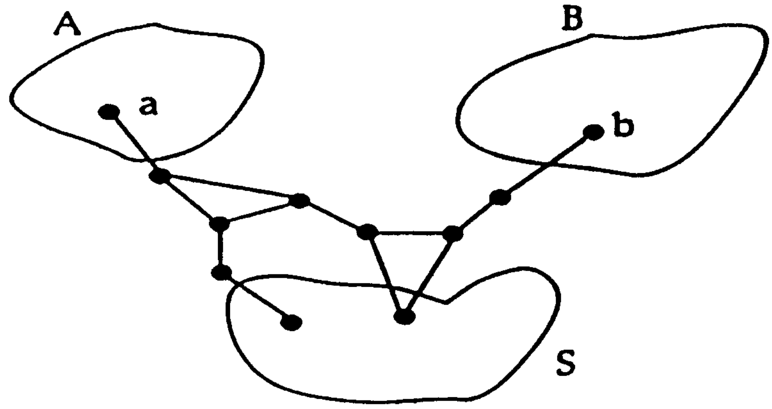

FIG. 6. The moral graph corresponding to the active chain in $\mathcal{G}$.

Suppose conversely that A is not separated from B in $(\mathcal{G}_{\mathrm{An}(A \cup B \cup S)})^m$. Then, there is a chain in this graph that circumvents S. The chain has pieces that correspond to edges in the original graph and pieces that correspond to marriages. Each marriage is a consequence of a meeting of arrows head to head at some vertex γ. If γ is in S or it has descendants in S, the meeting does not block the chain. If not, γ must have descendants in A or B, since the ancestral set was smallest. In the latter case, a new chain can be created with one head-to-head meeting less, using the line of descent, such as illustrated in Figure 7. Continuing this substitution process eventually leads to an active chain from A to B, and the proof is complete. ■

Geiger and Pearl [5] show in their Theorem 4 that the criterion of d-separation cannot be improved in the sense that for any given directed acyclic graph one

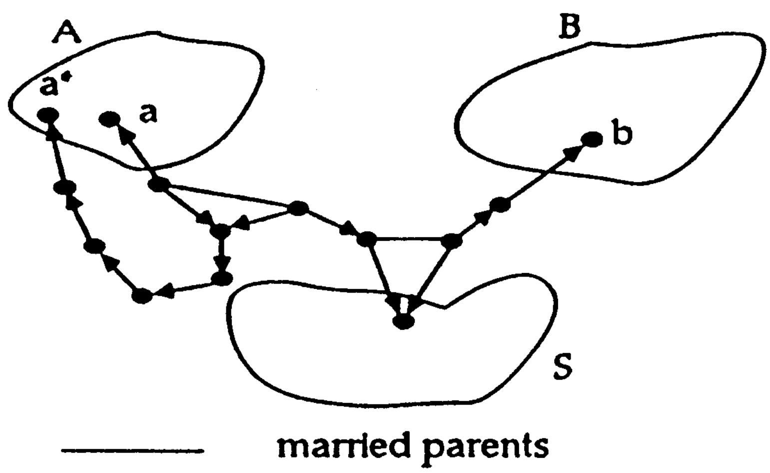

FIG. 7. The chain in the graph $(\mathcal{G}_{\mathrm{An}(A \cup B \cup S)})^m$ makes it possible to construct an active chain in $\mathcal{G}$ from A to B.

can find state spaces $\mathscr{X}_v$, $v \in V$ and a probability P such that

$$A \perp\!\!\!\perp B \mid S \Leftrightarrow S \text{ } d\text{-separates } A \text{ from } B\,. \tag{3}$$

The same holds for our global, directed Markov property by Proposition 3. But it is of interest to consider such statements for given sample spaces. Geiger and Pearl ([5], Theorem 3) show that in the case of real sample spaces a Gaussian distribution satisfying (3) exists and conjecture this to be true also for the case, where the state spaces all have two points, i.e., $\mathscr{X}_v = \{1, -1\}$.

6. EXTENSIONS AND DISCUSSION

It is of interest to observe that the equivalence of the local and global directed Markov properties holds even if the assumption of existence of densities is dropped. The only disadvantage of dropping this assumption is that the proof becomes slightly more involved.

Proposition 4. *Let P be a probability measure on $\mathscr{X}$ and $\mathscr{G}$ a directed acyclic graph as above. Then P obeys the local directed Markov property* (DL) *if and only if it obeys the global directed Markov property* (DG).

Proof. That (DG) $\Rightarrow$ (DL) is proved exactly as in the proof of Theorem 1. The direction of the proof that is different is (DL) $\Rightarrow$ (DG). The proof is by induction on the number of vertices of $\mathscr{G}$. For less than or equal to two vertices, there is nothing to show. Assume that the proposition is true for graphs with less than or equal to n vertices and assume that $|V| = n + 1$. To show that P satisfies (DG), we only need consider the case $\mathrm{An}(A \cup B \cup S) = V$, since cases with smaller ancestral sets will follow from the inductive assumption combined with observing that the directed local Markov property is inherited for marginal distributions to ancestral sets. We can then extend A and B if necessary such that $A \cup B \cup S = V$.

Thus, assume that $A \cup B \cup S = V$ and that A is separated from B by S in $\mathscr{G}^m$. Let λ be a terminal vertex in $\mathscr{G}$, and note that the separation implies that either $\mathrm{pa}(\lambda) \subseteq A \cup S$ or $\mathrm{pa}(\lambda) \subseteq B \cup S$, since parents in A and B would be married in $\mathscr{G}^m$.

Consider first the case $\lambda \in A$. Then, separation implies that $\mathrm{pa}(\lambda) \subseteq (A \setminus \{\lambda\}) \cup S$, so that, by (DL), we have that $\lambda \perp\!\!\!\perp B \mid (A\setminus\{\lambda\}) \cup S$. Also, S separates $A\setminus\{\lambda\}$ from B in $(\mathscr{G}^m)_{V\setminus\{\lambda\}}$ and, hence, also in the moral graph $(\mathscr{G}_{V\setminus\{\lambda\}})^m$ of $\mathscr{G}_{V\setminus\{\lambda\}}$, since this contains no more edges. By the inductive hypotheses $(A\setminus\{\lambda\}) \perp\!\!\!\perp B \mid S$. From these two conditional independencies, we deduce from basic properties of conditional independence ([1]) that $A \perp\!\!\!\perp B \mid S$.

The case $\lambda \in B$ is analogous.

To deal with the case $\lambda \in S$, we observe that if S separates A from B in $\mathscr{G}^m$, then $S\setminus\{\lambda\}$ separates A from B in $(\mathscr{G}^m)_{V\setminus\{\lambda\}}$ and, hence, also in $(\mathscr{G}_{V\setminus\{\lambda\}})^m$. By the inductive hypothesis, we deduce $A \perp\!\!\!\perp B \mid (S\setminus\{\lambda\})$.

If $\mathrm{pa}(\lambda) \subseteq A \cup S$, the directed local Markov property implies that $\lambda \perp\!\!\!\perp$

$B \mid (A \cup (S \setminus \{\lambda\}))$, whence — by properties of conditional independence — $B \perp\!\!\!\perp (A \cup \{\lambda\}) \mid S \setminus \{\lambda\}$ and, therefore, $B \perp\!\!\!\perp A \mid S$. Similarly in the case with $\text{pa}(\lambda) \subseteq B \cup S$. ■

It is not obvious to see how one should construct a probability measure with the directed Markov property other than via a recursive factorization of densities. However, suppose that we number the vertices of $\mathcal{G}$ in such a way that

$$\mu < \lambda \Rightarrow \mu \in \text{nd}(\lambda) ,$$

i.e., in such a way that arrows always point from low numbers to high numbers. We term such a numbering a a *well-numbering* of the vertices. Then, we can consider the following version of the local Markov property

$$\text{(WL)} \qquad \lambda \perp\!\!\!\perp \{\mu \mid \mu < \lambda\} \mid \text{pa}(\lambda) .$$

It is clear how to construct a probability measure recursively that satisfies (WL), but we do, in fact, have

Proposition 5. *Let P be a probability measure on $\mathcal{X}$ and $\mathcal{G}$ a directed acyclic graph as above. Then, P obeys the local directed Markov property* (DL) *if and only if it obeys the local well-numbering Markov property* (WL).

Proof. Clearly (DL) $\Rightarrow$ (WL). The converse is shown by induction. Assume that P satisfies (WL), and let λ^* be the highest numbered vertex of $\mathcal{G}$. Then, the marginal distribution of the variables $V \setminus \{\lambda^*\}$ satisfies (WL) and by the inductive hypothesis, we have for all $\alpha \neq \lambda^*$ that

$$\alpha \perp\!\!\!\perp (\text{nd}(\alpha) \setminus \{\lambda^*\}) \mid \text{pa}(\alpha) .$$

If $\lambda^* \in \text{de}(\alpha)$, it then follows that $\alpha \perp\!\!\!\perp \text{nd}(\alpha) \mid \text{pa}(\alpha)$. Otherwise, we must have $\text{pa}(\lambda^*) \subseteq \text{nd}(\alpha)$ and $\lambda^* \perp\!\!\!\perp (V \setminus \{\lambda^*\}) \text{pa}(\lambda^*)$, whence $\lambda^* \perp\!\!\!\perp \alpha \mid (\text{nd}(\alpha) \cup \text{pa}(\alpha))$ and we again deduce that $\alpha \perp\!\!\!\perp \text{nd}(\alpha) \mid \text{pa}(\alpha)$. Finally, the case $\alpha = \lambda^*$ follows trivially. ■

It follows by combining Propositions 4 and 5 that

Corollary 2.

$$\text{(WL)} \Leftrightarrow \text{(DL)} \Leftrightarrow \text{(DG)} .$$

The interesting aspect of these more general results is that no other features of the probability measure than the axioms of conditional independence are used in the proofs. Therefore, such results hold for other structures satisfying the same axioms; see, for example, the recent paper of Smith [12]. The results are really about an *algebra of relevance*, much more general than one based on probability.

Finally, we mention that the extension from statements about pairwise conditional independencies easily extend to statements about multiple, joint conditional independencies such as $\perp\!\!\!\perp_{i=1}^{n} A_i \mid S$ using Pearl's criterion of d-separation, since the fact that there is no active chain between any pair of sets A_i, clearly implies that there is no active chain between any one of them and the union of any of the others.

ACKNOWLEDGEMENTS

This paper was largely written while the first author was visiting the Statistical Laboratory, University of Cambridge. The author is indebted to the Laboratory for showing hospitality and to the Danish Natural Science Research Council for financial support. We are all grateful for having received useful comments from Morten Frydenberg and Nanny Wermuth.

References

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

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

[3] M. Frydenberg, The chain graph Markov property. Research Report No. 186, Dept. Theor. Stat., Århus University, Århus, Denmark (1989).

[4] M. Frydenberg, Marginalization and collapsibility in graphical association models. *Ann. Statist.* (to appear).

[5] D. Geiger and J. Pearl, On the logic of influence diagrams. *Proc. 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN (1988) 136–147.

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

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

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

[9] J. Pearl, A constraint propagation approach to probabilistic reasoning. L. M. Kanal and J. Lemmer, Eds., *Uncertainty in Artificial Intelligence* North-Holland, Amsterdam, (1986) 357–370.

[10] J. Pearl and A. Paz, Graphoids: A graph-based logic for reasoning about relevancy relations. *Proceedings of the European Conference on Artificial Intelligence*, Brighton, UK (1986).

[11] J. Pearl and T. Verma, The logic of representing dependencies by directed graphs. *Proceedings of the American Association of Artificial Intelligence Conference*, Seattle, WA (1987).

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

[13] T. P. Speed, A note on nearest-neighbour Gibbs and Markov probabilities. *Sankhya A* **41** (1979) 184–197.

[14] T. Verma, Causal networks: Semantics and expressiveness. *Proc. 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN (1988) 352–359.

[15] N. Wermuth, Linear recursive equations, covariance selection and path analysis. *J. Am. Statist. Assoc.* **75** (1980) 963–972.
[16] N. Wermuth and S. L. Lauritzen, Graphical and recursive models for contingency tables. *Biometrika* **70** (1983) 537–552.
[17] S. Wright, The method of path coefficients. *Ann. Math. Statist.* **5** (1934) 161–215.

Received October 1988
Accepted March 1990