The Annals of Statistics
1989, Vol. 17, No. 2, 654–672

INFLUENCE DIAGRAMS FOR STATISTICAL MODELLING

BY J. Q. SMITH

University of Warwick

A directed graph with identified nodes is defined to represent a set of conditional independence (c.i.) statements. It is shown how new c.i. statements can be read from the graph of an influence diagram and results of Howard and Matheson are rigorised and generalized. A new decomposition theorem, analogous to Kiiveri, Speed and Carlin and requiring no positivity condition, is proved. Connections between influence diagrams and Markov field networks are made explicit. Because all results depend on only three properties of c.i., the theorems proved here can be restated as theorems about other structures like second order processes.

1. Introduction. When eliciting the beliefs of a client about the basic structure of a problem, a statistician will all too often use convenient but unjustifiable modelling assumptions. These assumptions may concern the distribution of pertinent random variables or the existence between these variables of certain linear relationships. The assumptions will be made *before* the statistician has thoroughly discussed the problem with his client, in particular before he has ascertained which variables the client believes are related to each other and which are unrelated. When such premature assumptions are made we often find:

1. A model is imposed on a client that implicitly states that variables are unrelated when he believes that they are related (and vice versa). One consequence of this is that probabilities subsequently elicited from the client are contaminated and unreliable.
2. It is difficult for the statistician to distinguish which parts of his inference depend on the genuine information about the process as stated by the client and which depend on the convenient distributional assumptions he has imposed. Here even a sensitivity analysis may not help him fully differentiate these effects, since he can only perturb his model within a specified class of larger models still contaminated with a "convenient" structure.

Eliciting and then analysing how variables in a problem are related before imposing distributional assumptions is clearly desirable and allows the statistician or decision analyst not only to obtain direct information about the underlying process which is distribution-free but also to see how his problem might be decomposed into smaller component submodels which can be studied separately and more easily.

Influence diagrams are now a well used practical tool for graphically representing and manipulating the relationships between a set of random variables

Received December 1986; revised June 1988.

AMS 1980 *subject classifications*. Primary 62A99; secondary 62A15.

Key words and phrases. Causal networks, conditional independence, influence diagrams, Markov fields.

[see, e.g., Howard and Matheson (1981), Agogino (1985), Barlow and Zhang (1987), Barlow, Mensing and Smiriga (1986) and Rege and Agogino (1986)]. More recently, attempts have been made to formalise the approach [see, e.g., Olmsted (1984) and Shachter (1986a, b)]. The main difficulty with these authors' approach stems from the fact that they try to prove their results about conditional independence by manipulating the corresponding densities or mass functions. This approach is inelegant and not simple to adapt when it is not known beforehand that random variables are all discrete or all absolutely continuous. Technical problems can also arise when it is possible that some of the random variables in a model are functionally related. This will happen, for example, when some of the variables in your model represent your beliefs about a client's decisions [see, e.g., Smith (1988a, 1989)]. In this paper, by manipulating conditional independence statements *directly* we overcome these problems in a simple and straightforward way. In particular a key theorem, Theorem 6.2, can be proved whether or not functional relationships exist within the system.

The approach used here most clearly resembles that of Pearl (1986) on Bayes networks. However, the results given in the paper are more general than his because they do not require positivity conditions on random variables to hold. Furthermore since we use just three properties of conditional independence in our proofs we can obtain graphical representations of lack of influence not only appropriate for depicting and manipulating conventional conditional independence between random vectors but also for representing other types of lack of influence relating to best linear estimates. So the results given here are relevant not only to conventional Bayesian inference but other types of inference such as those that will be mentioned in Section 3.

Some exciting new developments, based on an axiomatic system analogous to the one developed here are reported in Pearl and Verma (1987). They assert that conditional independences can be read from a graph by examining whether a graphical property holds between all undirected paths between its nodes.

Another parallel area where graphs have been used to represent conditional independence between random variables stems from Markov field theory [Darroch, Lauritzen and Speed (1980), Wermuth (1980), Lauritzen, Speed and Vijayan (1984) and Lauritzen and Wermuth (1984)]. Within this discipline a directed graph can be defined which represents a set of conditional independence statements which are similar but not identical to those of the influence diagram [see Wermuth and Lauritzen (1983) and Kiiveri, Speed and Carlin (1984)]. Theorem 7.1 shows that a result by the last authors concerning these directed graphs extends to a theorem about influence diagrams where none of the awkward positivity conditions which they need to impose are necessary. Section 8 explores the relationship between influence diagrams and the type of graphs these authors use.

We begin our discussion by introducing the three properties that will be used as a basis of our graphical representation of conditional independence.

2. Three important properties of conditional independence. Let $A_m = \{\mathbf{X}_t: 1 \le t \le m\}$ be a finite set of random vectors. Dawid (1979) showed that the following three properties hold for the tertiary operator $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ where $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ reads "$\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$." Let $\mathbf{B}$ be the power set of A_m.

PROPERTY 1. For all $\mathbf{X}, \mathbf{Y}, \mathbf{W} \in \mathbf{B}$, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Y} \cup \mathbf{W}$.

This reads "Once $\mathbf{Y}$ is known (together with anything else, $\mathbf{W}$), $\mathbf{X}$ conveys no further information about $\mathbf{Y}$."

PROPERTY 2. For all $\mathbf{X}, \mathbf{Y}, \mathbf{W} \in \mathbf{B}$, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{W} \Leftrightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{W}$.

This reads "If $\mathbf{X}$ provides no further information about $\mathbf{Y}$ when $\mathbf{W}$ is known, then $\mathbf{Y}$ provides no information about $\mathbf{X}$ when $\mathbf{W}$ is known."

PROPERTY 3. For all $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \in \mathbf{B}$,

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{Z} | \mathbf{W} \Leftrightarrow \begin{cases} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \cup \mathbf{W}, \\ \mathbf{X} \perp\!\!\!\perp \mathbf{Z} | \mathbf{W}, \end{cases}$$

which reads "$\mathbf{X}$ is uninformative about both $\mathbf{Y}$ and $\mathbf{Z}$ given $\mathbf{W}$, is equivalent to saying that $\mathbf{X}$ is uninformative about $\mathbf{Y}$ given both $\mathbf{Z}$ and $\mathbf{W}$, together with the statement $\mathbf{X}$ is uninformative about $\mathbf{Z}$ given $\mathbf{W}$."

Henceforth we shall write $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ as shorthand for $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \varnothing$, where $\varnothing$ is the empty set, and abbreviate conditional independence to c.i. Sets which contain only one element are written $\mathbf{X}$ rather than $\{\mathbf{X}\}$.

In Section 4 it will be shown how sets of c.i. statements can be depicted and then manipulated on a directed graph. First let us discuss why c.i. statements are so important.

3. The manipulation of information assuming three c.i. properties. In a Bayesian statistical or decision analysis it is common to be told that, given certain information $\mathbf{W}$, a variable $\mathbf{X}$ will have no bearing on another $\mathbf{Y}$. It is often quite easy to ascertain this type of information from a client for various combinations of variables. Such information can be gathered before it is necessary to quantify subjective probabilities which, in contrast, are often very difficult to elicit with any degree of accuracy. By double checking a set of c.i. statements given to you by a client at the beginning of the modelling procedure you can ensure that you build a model which does not implicitly state that variables are conditionally independent when your client states they are linked (and vice versa). This checking procedure can be performed using the rules Properties 1, 2 and 3 which must hold if your client is being logically consistent. We will argue that this is most easily done using a directed graph. A case study outlining how such model checking is performed is given in Smith (1989). Now it is extremely important at this stage to notice that we need not demand that A_m be a set of random vectors nor that $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ be the conventional c.i. Provided that a tertiary operator satisfies Properties 1, 2 and 3 on a set A_m of "uncertain quantities," all the theory we develop about how "information" is transferred between those uncertain quantities will also hold on this new set A_m. So the results proved here are not just valid in Bayesian problems. Whether you use classical statistics, fuzzy probability, belief functions or linear inferences, a tertiary operator $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ can be usefully defined on your client's uncertain quantities.

A simple example of such an alternative c.i. structure is given below.

EXAMPLE 3.1. Let A_m be a set of random vectors and $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{W}$ read "a best linear estimate (under quadratic loss) of the components of $\mathbf{X}$ based on the components of $\mathbf{W}$ and $\mathbf{Y}$ need only include components of $\mathbf{W}$." This relation satisfies Properties 1, 2 and 3. To see this, first assume that $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{W}$ are jointly normally distributed. Then the statement above is equivalent to the statement that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{W}$. So, by the properties of conventional c.i., when A_m contains only (jointly) normal variates, then Properties 1, 2 and 3 hold. But the formulas for best linear estimates of arbitrarily distributed random variables always agree with those for normal variates with the same covariance structure. So Properties 1, 2 and 3 must be satisfied.

It can also be shown [see Smith (1988a, 1989)] that a c.i. operator satisfying Properties 1, 2 and 3 can be defined when the elements of A_m represent decisions and utilities as well as random vectors.

In Smith (1988b), I argue that *any* sensible definition of information transfer between variables should be expected to satisfy Properties 1, 2 and 3. In that paper several examples are given of how the theorems proved here can be applied to various non-Bayesian inference problems. Here, however, for the sake of clarity, we will henceforth assume that A_m is a set of random vectors and that $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ represents conventional c.i.

4. The preinfluence diagram and its properties. A single directed graph whose nodes label m random vectors $A_m = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m\}$ can be used to express $m - 1$ c.i. statements concerning the elements of A_m as follows.

Consider a pair (G, α) where $G := (A_m, E)$ is a directed acyclic graph and $\alpha: A_m \to \{1, \dots, m\}$ is a numbering of the vertices of G which is compatible with G, i.e., $\alpha(\mathbf{X}_i) < \alpha(\mathbf{X}_j)$ for a directed edge $(\mathbf{X}_i, \mathbf{X}_j) \in E$.

A *pre-influence diagram* (pre-I.D.) I on a set of random vectors $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ is a pair (G, α) together with the following $m - 1$ c.i. statements on A_m:

$$(4.1) \qquad \mathbf{X}_r \perp\!\!\!\perp \{\mathbf{X}_j \mid \alpha(\mathbf{X}_j) < \alpha(\mathbf{X}_r)\} \mid P(\mathbf{X}_r), \qquad \alpha(\mathbf{X}_r) = 2, 3, \dots, m,$$

where $P(\mathbf{X}_r) \subseteq \{\mathbf{X}_j \mid \alpha(\mathbf{X}_j) < \alpha(\mathbf{X}_r)\}$ is the set of nodes such that

$$\mathbf{X}_i \in P(\mathbf{X}_r) \Leftrightarrow (\mathbf{X}_i, \mathbf{X}_r) \in E.$$

So $P(X_r)$ is just the set of nodes attached by an edge directed into $\mathbf{X}_r$ in G. (G, α) is called a *numbered graph* of I.

Note that properties (4.1) hold for any set of random vectors when

$$P(\mathbf{X}_r) = \{\mathbf{X}_j \mid \alpha(\mathbf{X}_j) < \alpha(\mathbf{X}_r)\}, \qquad \alpha(\mathbf{X}_r) = 2, 3, \dots, m,$$

by Property 1. So, given any set A_m of random vectors, there always exists at least one pre-I.D. I (i.e., one whose graph is completely connected) whose c.i. statements (4.1) are true for A_m. The most informative pre-I.D.'s are those whose corresponding edge set E is small.

Before motivating the definition of a pre-I.D. it will be shown that, although α was needed to formulate the statements (4.1), any other compatible numbering β would produce a set of $m - 1$ c.i. statements (4.1) implying and implied by the original set of $m - 1$ c.i. statements under α. Hence by knowing only the directed graph G of a pre-I.D. I on A_m, without knowing the compatible numbering α, a statistical model is defined uniquely.

LEMMA 4.1. *Let I_1 be a pre-I.D. on $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ with numbered graph (G, α) and I_2 a pre-I.D. on the same set of random vectors with numbered graph (G, β), where $\beta := \sigma \circ \alpha$ and $\sigma: \{1, \dots, m\} \to \{1, \dots, m\}$ is the permutation with $\sigma(m-1) = m$, $\sigma(m) = m - 1$ and $\sigma(i) = i$ otherwise. Suppose the edge $(\mathbf{X}_s, \mathbf{X}_t)$ is not in the edge set E of G, where $\alpha(\mathbf{X}_s) = m - 1$ and $\alpha(\mathbf{X}_t) = m$. Then I_1 is a pre-I.D. on A_m iff I_2 is a pre-I.D. on A_m.*

PROOF. Note that the permutation σ only changes the c.i. statements (4.1) for $r = s$ or t. So by symmetry it is sufficient to prove that

$$\mathbf{X}_s \perp\!\!\!\perp A_m \setminus \{\mathbf{X}_s, \mathbf{X}_t\} | P(\mathbf{X}_s), \tag{4.2}$$

$$\mathbf{X}_t \perp\!\!\!\perp A_m \setminus \mathbf{X}_t | P(\mathbf{X}_t) \tag{4.3}$$

imply

$$\mathbf{X}_t \perp\!\!\!\perp A_m \setminus \{\mathbf{X}_s, \mathbf{X}_t\} | P(\mathbf{X}_t), \tag{4.4}$$

$$\mathbf{X}_s \perp\!\!\!\perp A_m \setminus \mathbf{X}_s | P(\mathbf{X}_s). \tag{4.5}$$

By (4.1) on I_1,

$$P(\mathbf{X}_s) \subseteq A_m \setminus \{\mathbf{X}_s, \mathbf{X}_t\}. \tag{4.6}$$

Also by (4.1) on I_2, since $(\mathbf{X}_5, \mathbf{X}_t) \notin E$,

$$P(\mathbf{X}_t) \subseteq A_m \setminus \{\mathbf{X}_s, \mathbf{X}_t\}. \tag{4.7}$$

By Property 3 (4.3) implies (4.4) and [with (4.7)]

$$\mathbf{X}_t \perp\!\!\!\perp \mathbf{X}_s | A_m \setminus \{\mathbf{X}_s, \mathbf{X}_t\}. \tag{4.8}$$

Statement (4.5) now follows from using Property 2 on (4.8) and combining it with (4.2) using Property 3 and (4.6). □

THEOREM 4.2. *Let I_1 and I_2 be two pre-I.D.'s on $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ with respective numbered graphs (G, α) and (G, β) where α and β are both compatible with G. Then I_1 is a pre-I.D. on A_m iff I_2 is a pre-I.D. on A_m.*

PROOF. Go by induction on m. The statement is clearly true for $m = 1$; suppose it is true for $m = n$. Let I_1 and I_2 both be pre-I.D.'s on $\{\mathbf{X}_1, \dots, \mathbf{X}_{n+1}\}$ with respective numbered graphs (G, α) and (G, β) where without loss we assume $(\mathbf{X}_{n+1}) = n + 1$. If $\beta(\mathbf{X}_{n+1}) = n + 1$, then the inductive hypothesis and the last statement in (4.1) prove our required result holds for $m = n + 1$. If $\beta(\mathbf{X}_{n+1}) \neq n + 1$, then without loss assume $\beta(\mathbf{X}_n) = n + 1$. Because no edge emanates from $\mathbf{X}_n$ or $\mathbf{X}_{n+1}$ in G there is a compatible numbering γ of G with $\gamma(\mathbf{X}_n) = n$ and $\gamma(\mathbf{X}_{n+1}) = n + 1$. Again, by the inductive hypothesis and the last statement of (4.1), I_1 is a pre-I.D. on A_{n+1} iff I_3 is a pre-I.D. on A_{n+1}, where I_3 has numbered graph (G, γ). Also by Lemma 4.1, I_3 is a pre-I.D. on A_{n+1} iff I_4 is a pre-I.D. on A_{n+1} where I_4 has numbered graph $(G, \sigma \circ \gamma)$ where σ is defined in Lemma 4.1. A third use of the inductive hypothesis and the last statement of (4.1) gives that I_4 is a pre-I.D. on A_{n+1} iff I_2 is a pre-I.D. on A_{n+1}. Hence I_1 is a pre-I.D. on A_m iff I_2 is a pre-I.D. on A_{n+1}. Hence the inductive step, and so the theorem is proved. □

5. Influence diagrams and graphical representations of conditional independence. By definition an *influence diagram* (I.D.) I on a set of random vectors $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ is an acyclic directed graph $G := (A_m, E)$ together with the set of $m - 1$ c.i. statements (4.1) for some (and hence all) numberings α compatible with G. If $\mathbf{X} \in P(\mathbf{X}_r)$, where $P(\mathbf{X}_r)$ is defined in statements (4.1), then $\mathbf{X}$ is called a *direct predecessor* (d.p.) of $\mathbf{X}_r$ in I and $\mathbf{X}_r$ is called a *direct successor* (d.s.) of $\mathbf{X}$ in I. The set of all direct successors of $\mathbf{X}$ in I is denoted by $S(\mathbf{X})$. Henceforth G will be called the *graph of I*.

As for pre-I.D.'s, given a set of c.i. statements (4.1) there is at least one, and usually many, I.D.'s I which are consistent with the statements (4.1) on a given joint distribution on A_m. Again, those I.D.'s whose graphs have the least number of edges tend to be the most useful. This definition of an I.D. agrees with one originally given by Howard and Matheson (1981) and later by Olmsted (1984) and Shachter (1986a, b) where $\{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ are random variables with a joint density/joint probability mass function which is strictly positive. By their definition, for example, if $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ have a joint mass function $p(x_1, x_2, x_3, x_4)$ which can be written in the form

$$p(x_1, x_2, x_3, x_4) = p_1(x_1)p_2(x_2)p_3(x_3)p_4(x_4)(x_4|x_2, x_3), \tag{5.1}$$

then this can be represented by them by an influence diagram whose graph G is given in Figure 1.

One compatible numbering here is $\alpha(\mathbf{X}_i) = i$ which, with G, represents the three c.i. statements on $A_4 = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4\}$

$$\mathbf{X}_2 \perp\!\!\!\perp \mathbf{X}_1,$$

$$\mathbf{X}_3 \perp\!\!\!\perp \{\mathbf{X}_1, \mathbf{X}_2\},$$

$$\mathbf{X}_4 \perp\!\!\!\perp \mathbf{X}_1 | \{\mathbf{X}_2, \mathbf{X}_3\}.$$

These conditions imply and are implied by the mass function breakdown (5.1). Another compatible numbering is $\alpha(X_1) = 4$, $\alpha(X_2) = 2$, $\alpha(X_3) = 1$ and $\alpha(X_4) = 3$. It is easy to check that the three c.i. statements corresponding with this new compatible numbering also agree with (5.1).

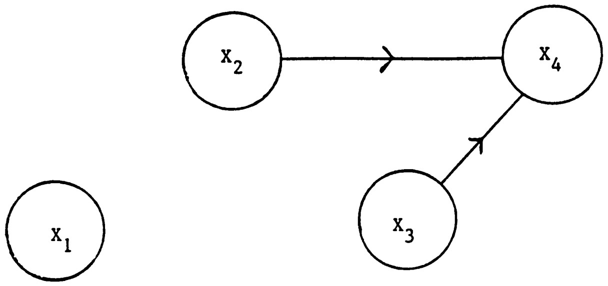

FIG. 1. *A simple influence diagram on four variables.*

Why then is it necessary to create all this abstraction using the c.i. notation? There are two answers to this:

1. It is much easier to prove theorems rigorously using the definition above. If the density/mass functions definition is used, then it is necessary to distinguish between continuous discrete and mixed variables and also whether the corresponding conditional densities are degenerate. By using our c.i. properties directly all these technical difficulties are overcome.
2. Because we have used just Properties 1, 2 and 3 we know that our theorems carry over to more general types of c.i. structure (see Section 3).

It is only fair to note that Howard and Matheson (1981) and Shachter (1986a, b) use influence diagrams to represent full decision problems rather than just relationships between random variables. However, Smith (1988a, 1989) shows how the c.i. structure above can be used to represent full decision problems as well but with more rigor and in much more generality.

We now restate Theorem 4.2 in terms of I.D.'s.

THEOREM 5.1. *Let $Q(\mathbf{X}_i)$ be the set of nodes $\mathbf{X}_j$ such that there exists no directed path from $\mathbf{X}_i$ to $\mathbf{X}_j$ in the graph of the I.D. I on A_m, $1 \le i \le m$. Then the $m - 1$ c.i. statements in I on A_m as stated in* (4.1) *imply* (*and are implied by*) *the m c.i. statements*

$$\mathbf{X}_i \perp\!\!\!\perp Q(\mathbf{X}_i) | P(\mathbf{X}_i), \qquad 1 \le i \le m, \tag{5.2}$$

where $P(\mathbf{X}_i)$ are the d.p.'s of X_i in I.

PROOF. For each i we can find a compatible numbering of the vertices in the graph of I which numbers each element of $Q(\mathbf{X}_i)$ before $\mathbf{X}_i$. The result now follows from Theorem 4.2. □

One advantage of using I.D.'s I rather than sets of c.i. statements (5.2) is that it is very easy to identify from the graph of I, the sets $Q(\mathbf{X}_i)$ given above. Hence with no algebra it is possible to invoke Theorem 5.1 to pick out useful implied c.i. statements from the c.i. statements originally given in (4.1). Ways of deducing c.i.

statements which involve the reversing of conditional statements are given in later sections.

Another directed graph with numbered nodes which have been used to represent a set of c.i. relationships between random variables is called the recursive causal graph on endogenous variables [Wermuth (1980), Wermuth and Lauritzen (1983) and Kiiveri, Speed and Carlin (1984)]. Within the terminology developed above we can form the following definitions. Consider the pair (G, α) where $G := (A_m, E)$ is a directed acyclic graph and $\alpha: A_m \to \{1, \dots, m\}$ is a numbering of the vertices of E which is compatible with G. A *pairwise influence diagram* (P.I.D.) K *on a set of random vectors* $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ is a pair (G, α) given above together with the following set of c.i. statements:

$$\text{When} \quad \alpha(\mathbf{X}_i) < \alpha(\mathbf{X}_j),\ (\mathbf{X}_i, \mathbf{X}_j) \notin E \text{ and } 1 \le i \ne j \le m$$

$$\text{then} \quad \mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | \{\mathbf{X}_k | \alpha(\mathbf{X}_k) < \alpha(\mathbf{X}_j), \mathbf{X}_k \ne \mathbf{X}_i\}.$$

G is then called a *graph* of the P.I.D. K.

A disadvantage the P.I.D. has over the I.D. is that different compatible numberings imply different sets of c.i. statements. So the index α on a P.I.D. needs to be retained.

For example suppose $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3$. Then, under the compatible numbering $\alpha(\mathbf{X}_i) = i$, the graph of the P.I.D. is given in Figure 2.

However if an alternative compatible numbering $\alpha(\mathbf{X}_1) = 1$, $\alpha(\mathbf{X}_2) = 3$ and $\alpha(\mathbf{X}_3) = 2$ was used, then Figure 2 would imply $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_3$ which is clearly false if $\mathbf{X}_1 = \mathbf{X}_3$ unless $\mathbf{X}_1$ and $\mathbf{X}_3$ both have a degenerate distribution.

To obtain interesting results about P.I.D.'s it has been necessary to assume that no probabilities in the joint mass function of $\{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ are zero. This, however, is a very strong assumption to make in practice. For example it is often the case that variables are functionally linked [Spiegelhalter (1987)]. Problems are amplified if we require decisions as well as variables to be represented in a diagram [see Smith (1988a, 1989)] since by definition Bayes decisions will nearly always be deterministic functions of random variables.

We will show in Sections 7 and 8 how theorems about P.I.D.'s can be reformulated and reproved as theorems about I.D.'s with none of these uncomfortable side conditions. We will also see how I.D.'s and P.I.D.'s relate to one

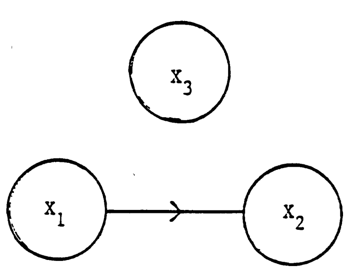

FIG. 2.

another. First we need to introduce some notation in order to prove some basic results about I.D.'s.

An I.D. I_1 on random vectors A_m is said to be implied by an I.D. I_2 on A_m (written $I_2 \Rightarrow I_1$) if all conditional independence statements contained in I_1 can be deduced from the conditional independence statements in I_2. Two I.D.'s I_1 and I_2 on A_m are said to be equivalent (written $I_1 \equiv I_2$) if $I_2 \Rightarrow I_1$ and $I_1 \Rightarrow I_2$. Two I.D.'s I_1 and I_2 on A_m having respective graphs $G_1 := (A_m, E_1)$ and $G_2 := (A_m, E_2)$ are called *similar* (written $I_1 \sim I_2$) when

$$(\mathbf{X}_j, \mathbf{X}_i) \text{ or } (\mathbf{X}_i, \mathbf{X}_j) \in E_1 \Leftrightarrow (\mathbf{X}_i, \mathbf{X}_j) \text{ or } (\mathbf{X}_j, \mathbf{X}_i) \in E_2.$$

Note that both equivalence and similarity defined above define equivalence relations of the class of all I.D.'s on A_m.

The directed graph $\hat{G} := (A_m \setminus \mathbf{X}_i, \hat{E})$ obtained from an I.D. I with graph $G := (A_m, E)$ by deleting the node $\mathbf{X}_i$ together with all edges into it is a graph of an I.D. on A_{m-1} expressing the $m - 2$ c.i. statements

$$\mathbf{X}_r \perp\!\!\!\perp \{\mathbf{X}_j | \alpha(\mathbf{X}_j) < \alpha(\mathbf{X}_r)\} | P(\mathbf{X}_r), \qquad \alpha(\mathbf{X}_r) = 2, \dots, m, \ \alpha(\mathbf{X}_r) \neq \alpha(\mathbf{X}_i),$$

where $P(\mathbf{X}_r)$ is defined in (4.1); α is a numbering of the vertices of E which is compatible with G and $\alpha(\mathbf{X}_i) = m$. Let this I.D. on $A_m \setminus \mathbf{X}_i$ be denoted by $I - \mathbf{X}_i$. Note that $I - \mathbf{X}_i$ is defined iff $\mathbf{X}_i$ has no d.s.'s in I.

6. Bayes rule for influence diagrams. Very often our interest lies in the statements we can make about a subset C of uncertain quantities $\{\mathbf{X}_1, \dots, \mathbf{X}_m\} = A_m$. Before we try to make precise probabilistic statements of one sort or another it will be important which quantities in $\bar{C} = A_m \setminus C$ might influence statements we would like to make about C. This requires us to introduce variables into an influence diagram in such a way that quantities in $\bar{C}$ precede quantities in C. Unfortunately we often find that the most natural representation of a problem introduces the quantities of interest first. For example in most Bayesian models, prior distributions on the quantities affecting predictive distribution (parameters) are introduced first and the distribution of observations given these parameters second. This happens because it is often easier to *specify* relationships between random variables in an order which follows causality, but we are usually *interested* in an unobserved "cause" given some observed "effects."

In more classical inference about parameters we have the same problems. A set of parameter θ of interest is defined in terms of other hyperparameters ψ. Ideally we want to reverse this conditioning in order to distill information about θ in our data without reference to ψ.

For probabilistic c.i. on discrete variables, Howard and Matheson (1981) and more formally Shachter (1986a) took the first steps by showing how an I.D. needs to be adjusted when an edge in its graph needs to be reversed. Here is the formal proof of their theorem generalised to all I.D.'s on a set of quantities with a c.i. structure defined by Properties 1, 2 and 3.

LEMMA 6.1. *Let* $G := (A_m, E)$ *be the graph of an I.D.* I *on* $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ *and* $(\mathbf{X}_i, \mathbf{X}_j) \in E$ *fixed with* (6.2). *Let* $G_1 := (A_m, E_1)$ *be another* *directed graph obtained by defining E_1 as*

$$E_1 = \left[E \cup \{(\mathbf{X}_j, \mathbf{X}_i)\} \cup \hat{E}\right] \setminus \{(\mathbf{X}_i, \mathbf{X}_j)\}, \tag{6.1}$$

where $\hat{E} = \{(\mathbf{X}_k, \mathbf{X}_i), (\mathbf{X}_k, \mathbf{X}_j): k \in K\}$ where $K = \{1 \le k \le m | (\mathbf{X}_k, \mathbf{X}_i) \in E$ or $(\mathbf{X}_k, \mathbf{X}_j) \in E\}$. Suppose that under some ordering α compatible with I,

$$\begin{aligned} \alpha(\mathbf{X}_i) &= m - 1, \\ \alpha(\mathbf{X}_j) &= m. \end{aligned} \tag{6.2}$$

Then $I \Rightarrow I_1$ where I_1 is the I.D. on A_m whose graph is G_1.

PROOF. First note that, by (6.2), the graph G_1 defined above is acyclic. Also by our construction

$$I - \mathbf{X}_j - \mathbf{X}_i \equiv I_1 - \mathbf{X}_i - \mathbf{X}_j.$$

So it is sufficient to prove that the remaining two c.i. statements in I,

$$\mathbf{X}_i \perp\!\!\!\perp A | P(\mathbf{X}_i), \tag{6.3}$$

$$\mathbf{X}_j \perp\!\!\!\perp A \cup \mathbf{X}_i | \hat{P}(\mathbf{X}_j) \cup \mathbf{X}_i, \tag{6.4}$$

where $A = A_m \setminus \{\mathbf{X}_i, \mathbf{X}_j\}$, $\hat{P}(\mathbf{X}_j)$ is the set of d.p.'s of $\mathbf{X}_j$ other than $\mathbf{X}_i$ and $P(\mathbf{X}_i)$ are the d.p.'s of $\mathbf{X}_i$, imply the remaining two c.i. statements are I_1, which from (6.1) are

$$\mathbf{X}_j \perp\!\!\!\perp A | \hat{P}(\mathbf{X}_j) \cup P(\mathbf{X}_i), \tag{6.5}$$

$$\mathbf{X}_i \perp\!\!\!\perp A \cup \mathbf{X}_j | \mathbf{X}_j \cup \hat{P}(\mathbf{X}_j) \cup P(\mathbf{X}_i). \tag{6.6}$$

Now if $B := \hat{P}(\mathbf{X}_i) \cup P(\mathbf{X}_j) \subseteq A$, (6.3) and (6.4) imply by Properties 2 and 3,

$$A \perp\!\!\!\perp \mathbf{X}_i | B, \tag{6.7}$$

$$A \perp\!\!\!\perp \mathbf{X}_j | B \cup \mathbf{X}_i, \tag{6.8}$$

which by Property 3 is equivalent to stating that

$$A \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{X}_j\} | B$$

using Property 3. Then Property 2 gives us the two equivalent statements

$$\mathbf{X}_j \perp\!\!\!\perp A | B \tag{6.9}$$

and

$$\mathbf{X}_i \perp\!\!\!\perp A | \mathbf{X}_j \cup B, \tag{6.10}$$

which, on using Property 1 on (6.10) are equivalent to (6.5) and (6.6) as required. □

NOTE. We may lose c.i. information by this manipulation because (6.3) and (6.4) imply but are not necessarily implied by (6.7) and (6.8).

THEOREM 6.2 (The Howard–Matheson theorem). *Let $G := (A_m, E)$ be the graph of an I.D. I on A_m and let $G_1 := (A_m, E_1)$ where E_1 is as defined in* (6.1).

Then, provided G_1 *is acyclic,* $I \Rightarrow I_1$ *where* I_1 *is the I.D. on* A_m *whose graph is* G_1.

PROOF. Go by induction. The theorem is trivially true for $m = 2$. Assume that it is true for all I when $m = n$. Let I_1 be an I.D. on $n + 1$ variables. If there exists a compatible numbering α of I such that

$$\alpha(\mathbf{X}_r) = n + 1, \qquad r \neq j, \tag{6.11}$$

then because from G $\alpha(\mathbf{X}_i) < \alpha(\mathbf{X}_j)$ the inductive hypothesis gives us that $I - \mathbf{X}_r \Rightarrow I_1 - \mathbf{X}_r$ whence $I \Rightarrow I_1$, as required.

On the other hand if no numbering with property (6.11) exists for I, then $\mathbf{X}_j$ must be the *only* variable in I with no d.s. It follows that there exists a numbering compatible with I which sets $\alpha(\mathbf{X}_i) = n$. Otherwise G_1 would contain a directed cycle consisting of a directed path from $\mathbf{X}_i$ to $\mathbf{X}_j$ of at least two edges inherited from G and the edge $(\mathbf{X}_j, \mathbf{X}_i)$. Therefore $\alpha(\mathbf{X}_j) = n + 1$ and $\alpha(\mathbf{X}_i) = n$. The inductive step now follows from Lemma 6.1. So the theorem is proved. □

Shachter (1986b) shows that for probabilistic c.i. it is possible to use this algorithm to reverse all edges from one node until it has no d.s.'s. However, his method can be poor if used directly in the sense that many c.i. statements originally made in I can no longer be read from I_1. Theorem 6.2 is very useful, however, as a stepping stone to more powerful methods of using I.D.'s to rerepresent sets of c.i. statements.

The following corollary is useful in the proof of a later result. It gives conditions when the reversal of an edge in the graphs to an influence diagram need introduce no additional edge in the graph of an implied influence diagram.

COROLLARY 6.3. *Let* I *have the graph* $G := (A_m, E)$. *Suppose two variables* $\mathbf{X}_i, \mathbf{X}_j \in I$, *have the properties*:

(i) $\mathbf{X}_i \in P(\mathbf{X}_j)$ *where* $P(\mathbf{X}_j)$ *are the d.p.'s of* $\mathbf{X}_j$ *in* I.
(ii) $P(\mathbf{X}_i) = P(\mathbf{X}_j) \setminus \mathbf{X}_i$ *where* $P(\mathbf{X}_i)$ *are the d.p.'s of* $\mathbf{X}_i$ *in* I.
(iii) *Every directed path in* G *from* $\mathbf{X}_i$ *to* $\mathbf{X}_j$ *contains the edge* $(\mathbf{X}_i, \mathbf{X}_j)$.

Then $I \equiv I_1$ *where* I_1 *has the graph* $G_1 := (A_m, E_1)$ *where*

$$E_1 = \left(E \setminus \{(\mathbf{X}_i, \mathbf{X}_j)\}\right) \cup \{(\mathbf{X}_j, \mathbf{X}_i)\}.$$

PROOF. Property (iii) guarantees that the graph G_1, derived from G by reversing the edge (i, j), is again acyclic. The corollary now follows directly from the Howard–Matheson theorem. □

7. A decomposition theorem for influence diagrams. A key theorem was proved by Kiiveri, Speed and Carlin (1984) which made a vital link between P.I.D.'s and undirected graphs of Markov fields as developed by Darroch, Lauritzen and Speed (1980). A consequence of this theorem was (using the terminology developed here for I.D.'s) that if all the d.p.'s of each node **X** in a P.I.D. K_1 were connected and all the d.p.'s of each node **X** on a P.I.D. K_2 were also connected, then, when $K_1 \sim K_2$, $K_1 \equiv K_2$. This corollary is very useful. For example, it indicates how information can be efficiently propagated through a probabilistic system [see Spiegelhalter (1986)]. The undirected graphs formed by deleting the arrows on edges on P.I.D.'s with this property were called "decomposable" by Darroch, Lauritzen and Speed (1980). This motivates the following terminology.

DEFINITION. An I.D. I is said to be *decomposable* if the d.p.'s $P(\mathbf{X})$ of any node $\mathbf{X}$ in I have the property that $\mathbf{X}_1, \mathbf{X}_2 \in P(\mathbf{X}) \Rightarrow \mathbf{X}_1$ and $\mathbf{X}_2$ are joined by an edge in the graph of I.

Figure 3 illustrates four I.D.'s, two of which are decomposable and two which are not. The following assertion is immediate from the definition above.

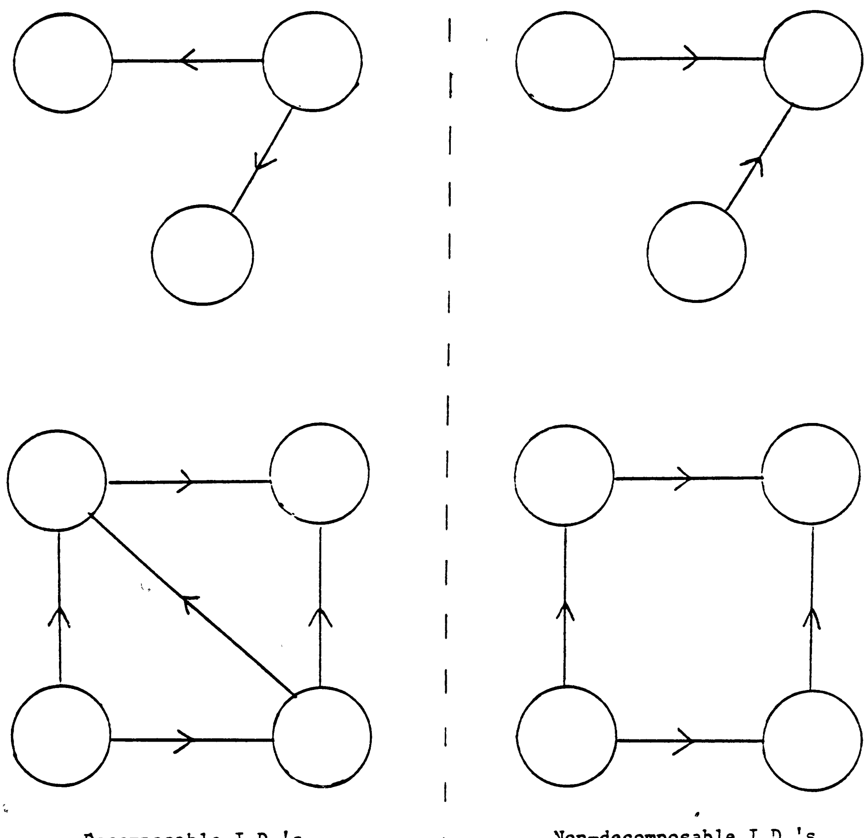

FIG. 3. *Two decomposable and two nondecomposable I.D.'s.*

ASSERTION. When I is decomposable then $I - \mathbf{X}$ is decomposable where $\mathbf{X}$ is any node in I with no d.s.

In this section we prove that the above mentioned corollary of the theorem of Kiiveri, Speed and Carlin (1984) carries over to I.D.'s but without needing any restrictive side conditions concerning strict positivity of the variable's joint distribution. Again we only use Properties 1, 2 and 3 and so this theorem is again true whenever a c.i. structure can be defined.

THEOREM 7.1 (Influence diagram decomposition). *If I_1 and I_2 are two decomposable I.D.'s on the same set of m nodes and $I_1 \sim I_2$, then $I_1 \equiv I_2$.*

PROOF. Go by induction. Clearly the statement is true for $m = 1$. Suppose that the statement is true for all I.D.'s on $m - 1$ nodes. Let $\mathbf{X}$ be any node with no d.s. in I_1. Let $\{\mathbf{X}_1, \dots, \mathbf{X}_k\}$ be the d.p.'s of $\mathbf{X}$ in I_1.

If $\mathbf{X}$ has no d.s. in I_2, then by the assertion above, both $I_1 - \mathbf{X}$ and $I_2 - \mathbf{X}$ are decomposable and clearly $I_1 - \mathbf{X} \sim I_2 - \mathbf{X}$. So by the inductive hypothesis $I_1 - \mathbf{X} \equiv I_2 - \mathbf{X}$ whence $I_1 \equiv I_2$. So if $I_1 \not\equiv I_2$, $\mathbf{X}$ must have at least one d.s. in I_2. Label $\{\mathbf{X}_1, \dots, \mathbf{X}_k\}$ so that $\{\mathbf{X}_1, \dots, \mathbf{X}_p\}$, $1 \le p \le k$, are the d.s.'s of $\mathbf{X}$ in I_2. Assume I_2 has the least number p of d.s.'s of $\mathbf{X}$ such that $I_1 \not\equiv I_2$. By the above $p \geq 1$.

Any d.p. $\mathbf{Y}$ of $\mathbf{X}_i \in \{\mathbf{X}_1, \dots, \mathbf{X}_p\}$ with $\mathbf{X} \neq \mathbf{Y}$ in I_2 has the property

$$\mathbf{Y} \in \{\mathbf{X}_1, \dots, \mathbf{X}_k\}, \tag{7.1}$$

for otherwise $\mathbf{X}$ and $\mathbf{Y}$ would be disconnected d.p.'s of $\mathbf{X}_i$ contradicting the decomposability of I_2. In the graph of I_2 an edge connects each node $\{\mathbf{X}_{p+1}, \dots, \mathbf{X}_k\}$ to each node in

$$\{\mathbf{X}_1, \dots, \mathbf{X}_p\}. \tag{7.2}$$

This is because all nodes $\{\mathbf{X}_1, \dots, \mathbf{X}_k\}$ are connected to each other by the decomposability of I_1 and if node $\mathbf{X}_i \in \{\mathbf{X}_1, \dots, \mathbf{X}_p\}$ were connected to a node $\mathbf{X}_j \in \{\mathbf{X}_{p+1}, \dots, \mathbf{X}_k\}$, then the graph of I_2 would exhibit a cycle $(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}, \mathbf{X}_i)$.

Without loss let $\mathbf{X}_1$ denote the first element from $\{\mathbf{X}_1, \dots, \mathbf{X}_p\}$ introduced into I_2 with a compatible numbering of its nodes. Then by (7.1) and (7.2) the d.p.'s of $\mathbf{X}_1$ in I_2,

$$P_2(\mathbf{X}_1) = \{\mathbf{X}_{p+1}, \dots, \mathbf{X}_k, \mathbf{X}\}.$$

Since $\mathbf{X}$ and $\mathbf{X}_1$ satisfy conditions (i), (ii) and (iii) of Corollary 6.3 in I_2 (relabelling $\mathbf{X}$ as $\mathbf{X}_i$ and $\mathbf{X}_1$ as $\mathbf{X}_j$) we can conclude

$$I_2 \equiv I_2(1),$$

where $I_2(1)$ is the I.D. whose graph is obtained by reversing the edge from $\mathbf{X}$ to $\mathbf{X}_1$ in the graph of I_2. So by our hypothesis $I_2(1) \not\equiv I_1$. Clearly $I_2(1)$ is decomposable and has one less d.s. of $\mathbf{X}$ than I_2. So, contrary to our hypothesis, I_2 does not have the least number of d.s.'s of $\mathbf{X}$ with the property $I_2 \equiv I_1$. Our theorem is thus proved by contradiction. □

There are two problems with using the Howard–Matheson theorem directly to reverse the edges of the graph of an I.D. First, it can be computationally

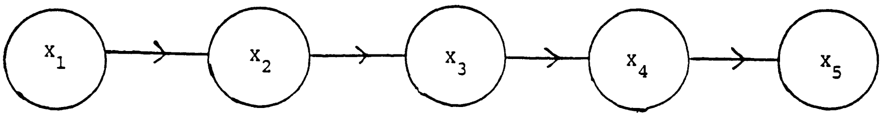

FIG. 4. *A Markov chain I.*

costly to discover whether the edge reversal introduces a cycle. Second, unless used with guile, it loses information in the system very quickly since more edges may need to be introduced in the graph of the implied I.D. at each reversal. Consider the I.D. I whose graph is given in Figure 4. You expect to observe $\mathbf{X}_5$ and want to make inferences about the other random vectors in your model. So you want to find an I.D. I_1 implied by I with $\mathbf{X}_5$ having no d.p.'s. Using the Howard–Matheson theorem directly you would reverse the $(\mathbf{X}_4, \mathbf{X}_5)$ edge introducing $(\mathbf{X}_3, \mathbf{X}_5)$ edge. You would then need to reverse the $(\mathbf{X}_3, \mathbf{X}_5)$ edge introducing the $(\mathbf{X}_2, \mathbf{X}_5)$ edge and so on. Each time you have introduced a new edge you have lost information. On the other hand it is easy to check that I is decomposable and is similar to the decomposable I.D. I_1 whose graph is obtained from the graph of I by reversing the direction of all its edges. You can conclude from the decomposition theorem that $I \equiv I_1$ and no information has been lost. (Note, incidently, that this shows that the decomposition theorem is a generalisation of the result which says that the Markov property is preserved under time index reversal.)

Now of course the decomposition theorem was proven using only the Howard–Matheson theorem. In fact it reverses edges in the order $(\mathbf{X}_1, \mathbf{X}_2)$, $(\mathbf{X}_2, \mathbf{X}_3), (\mathbf{X}_3, \mathbf{X}_4), (\mathbf{X}_4, \mathbf{X}_5)$. What the decomposition theorem does do is make it much easier to identify in which order edges need to be reversed so as to lose no information when "conditioning out" random vectors.

To my knowledge it is an open question how to identify an order of edge reversal which minimises the number of additional edges in the graph of an I.D. I_1 implied by an I.D. I when I_1 is constrained to have a graph with a set of edges between nodes of a given direction. The following algorithm often appears to achieve this minimum.

Suppose we need to find an I.D. I_1 such that $I \Rightarrow I_1$ and a subset of edges $\hat{E}$ of the graph G of I needs to be reversed in the graph G_1 of I_1 and certain others kept in a way that does not create cycles. Let C denote the set of nodes attached by edges in $\hat{E}$ in $G := (A_m, E)$. Let $D \subseteq A_m$ denote the set of nodes $\mathbf{Y}$ for which there exists a compatible numbering of G with $\alpha(\mathbf{Y}) > \alpha(\mathbf{X})$ for all $\mathbf{X} \in C$.

Define $\tilde{G} := (A_m \setminus D, \tilde{E})$ where $\tilde{E} := \{(\mathbf{X}_i, \mathbf{X}_j) \in E | \mathbf{X}_j \notin D\}$ and $A_m = \{\mathbf{X}_1, \dots, \mathbf{X}_m\}$. Let $\overline{G} := (A_m \setminus D, \overline{E})$ denote the graph of a decomposable I.D. $\overline{I}$ on $A_m \setminus D$ satisfying:

1. $\tilde{E} \subseteq \overline{E}$.
2. There exists a graph $\overline{G}_1 := (A_m \setminus D, \overline{E}_1)$ of a decomposable I.D. $\overline{I}_1 \sim \overline{I}$ (so that $\overline{I} \equiv \overline{I}_1$ by Theorem 7.1) where $\hat{E}_2 \subseteq \overline{E}_1$ and where $\hat{E}_2 := \{(\mathbf{X}_i, \mathbf{X}_j) | (\mathbf{X}_j, \mathbf{X}_i) \in \hat{E}\}$.
3. $\overline{G}$ is a graph of an I.D. with the least number of edges satisfying (i) and (ii).

Such a $\overline{G}$ and $\overline{G}_1$ must exist since any completely connected acyclic directed graph is decomposable. Finally let $G_1 := (A_m, E_1)$ where $E_1 = \overline{E}_1 \cup \hat{E}_1$ and where

$$\hat{E}_1 = \{(\mathbf{X}_i, \mathbf{X}_j) \in E | \mathbf{X}_j \in D\}.$$

It is easily checked that if I_1 is the I.D. with graph G_1, then $I \Rightarrow I_1$. Graph G_1 also has all edges in $\hat{E}$ reversed.

The construction of G_1 from G is illustrated in Figure 5 where $\alpha(\mathbf{X}_i) = i$, $1 \leq i \leq 6$ in I, and $\hat{E} = \{(\mathbf{X}_3, \mathbf{X}_4)\}$.

8. Some links between I.D.'s and P.I.D.'s. Here we prove some new results on the relationships between I.D.'s and P.I.D.'s.

THEOREM 8.1. *The directed graph G of an I.D. I, implies the c.i. statements of a P.I.D. with the same directed graph G.*

PROOF. Without loss label the nodes of G such that $\alpha(\mathbf{X}_i) = i$ where α is a numbering compatible with I. From the definition of a P.I.D. (see Section 5) $\mathbf{X}_i, \mathbf{X}_j$ can only be left unconnected when

$$\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | A_{j-1} \setminus \mathbf{X}_i, \qquad i < j \text{ where } A_{j-1} = \{\mathbf{X}_1, \dots, \mathbf{X}_{j-1}\}.$$

Well, under this labelling, $\mathbf{X}_i, \mathbf{X}_j$ are not connected in the graph of I, $i < j$, implies

$$\mathbf{X}_j \perp\!\!\!\perp A_{j-1} | P(\mathbf{X}_j),$$

where $P(\mathbf{X}_j)$ are the d.p.'s of $\mathbf{X}_j$ and $\mathbf{X}_i \in A_{j-1} \setminus P(\mathbf{X}_j)$,

$$\overset{\text{P3}}{\Rightarrow} \mathbf{X}_j \perp\!\!\!\perp \mathbf{X}_i | P(\mathbf{X}_j) \cup A_{j-1} \setminus \mathbf{X}_i,$$

$$\overset{\text{P2}}{\Rightarrow} \mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | A_{j-1} \setminus \mathbf{X}_i$$

as required. □

Note that the converse of this theorem is *not* true as is illustrated in the example of Figure 2.

The next theorem gives a c.i. relationship between a particular node $\mathbf{X}_i$ in an I.D. I and all the other nodes in the diagram. Theorem 8.1 shows that the c.i. statements in an I.D. I imply those in the P.I.D. with the same directed graph as I. Wermuth and Lauritzen (1983) and Kiiveri, Speed and Carlin (1984) prove the result that if a P.I.D. K is decomposable (in the sense defined above for I.D.'s), then we can conclude that for every node $\mathbf{X}_i$ and $\mathbf{X}_j$ unconnected in K,

$$\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | A_m \setminus \{\mathbf{X}_i, \mathbf{X}_j\} \tag{8.1}$$

provided that the joint density (mass function) over A_m is strictly positive. Theorem 8.2 gives an analogous result for I.D.'s when no positivity condition is imposed. Its corollary may be particularly useful to those who work with P.I.D.'s.

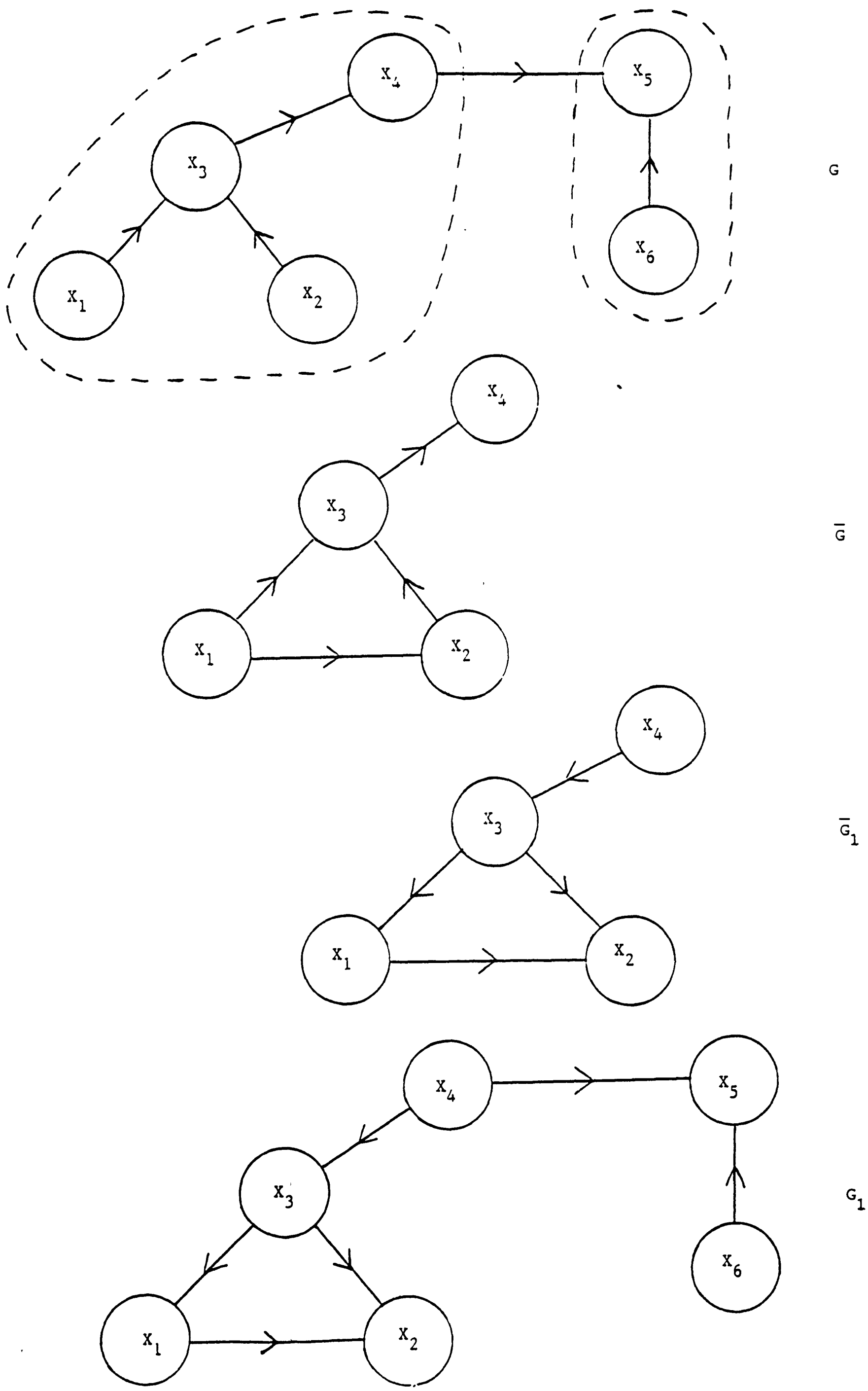

FIG. 5.

THEOREM 8.2. *In a decomposable I.D. I with nodes ordered* $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$,

$$\mathbf{X}_i \perp\!\!\!\perp T(\mathbf{X}_i) | P(\mathbf{X}_i) \cup S(\mathbf{X}_i), \tag{8.2}$$

where $P(\mathbf{X}_i)$ *and* $S(\mathbf{X}_i)$ *are, respectively, the d.p.'s and d.s.'s of* $\mathbf{X}_i$ *and* $T(\mathbf{X}_i)$ *is their complement in* $A_m \setminus \{\mathbf{X}_i\}$.

PROOF. Let I.D. I_1 have graph G_1 formed from the graph G of I by adding (where necessary) edges from all the d.p.'s and d.s.'s of $\mathbf{X}_i$ to each d.s. $\mathbf{Y} \in S(\mathbf{X}_i)$ consistent with some compatible numbering of I. Clearly G_1 is acyclic and $I \Rightarrow I_1$. The set of nodes $P(\mathbf{X}_i) \cup S(\mathbf{X}_i)$ is completely connected in G_1 since the decomposability of I at $\mathbf{Y}$ ensures that all nodes in $P(\mathbf{X}_i)$ are connected to one another. □

CLAIM. I_1 is decomposable.

It is sufficient to check that if $\mathbf{Y} \in S(\mathbf{X}_i)$ then any pair $\mathbf{Z}_1, \mathbf{Z}_2$ of d.p.'s of $\mathbf{Y}$ is connected by an edge in G_1. Well, by the decomposability of I at $\mathbf{Y}_i$, $\mathbf{X}_i$ is connected to $\mathbf{Z}_1$ and to $\mathbf{Z}_2$ so

$$\{\mathbf{Z}_1, \mathbf{Z}_2\} \subseteq P(\mathbf{X}_i) \cup S(\mathbf{X}_i).$$

Since $P(\mathbf{X}_i) \cup S(\mathbf{X}_i)$ is completely connected in G_1, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are connected in G_1 as required.

Now define G_2 to be the directed graph formed by reversing all edges in I_1 from $\mathbf{X}_i$ to nodes in $S(\mathbf{X})$. Clearly G_2 is acyclic. Let I_2 be an I.D. whose graph is G_2. Then $I_2 \sim I_1$. In I_2 the d.p.'s of $\mathbf{X}_i$, $P_2(\mathbf{X}_i) := P(\mathbf{X}_i) \cup S(\mathbf{X}_i)$ where $P(\mathbf{X}_i)$ and $S(\mathbf{X}_i)$ are, respectively, the d.p.'s and d.s.'s of $\mathbf{X}_i$ in I_1. By our construction of I_1 all nodes in $P_2(\mathbf{X}_i)$ are connected in G_1. It follows that I_2 is itself decomposable. By Theorem 7.1 we therefore have $I_1 \equiv I_2$ so in particular $I \Rightarrow I_2$. From I_2 we can read the c.i. statement

$$\begin{aligned}
&\mathbf{X}_i \perp\!\!\!\perp A_m \setminus \mathbf{X}_i | P(\mathbf{X}_i) \cup S(\mathbf{X}_i) \\
&\quad \Leftrightarrow \mathbf{X}_i \perp\!\!\!\perp T(\mathbf{X}_i) \cup P(\mathbf{X}_i) \cup S(\mathbf{X}_i) | P(\mathbf{X}_i) \cup S(\mathbf{X}_i) \\
&\quad \overset{\mathrm{P3,P1}}{\Leftrightarrow} \mathbf{X}_i \perp\!\!\!\perp T(\mathbf{X}_i) | P(\mathbf{X}_i) \cup S(\mathbf{X}_i)
\end{aligned}$$

as required.

COROLLARY 8.3. *In the notation of the theorem above, when* $\mathbf{X}_i$ *is unconnected to* $\mathbf{X}_j$ *in the graph of a decomposable I.D. I, then*

$$\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | A_m \setminus \{\mathbf{X}_i, \mathbf{X}_j\}.$$

PROOF. This is immediate from Theorem 8.2 on conditioning out $T(\mathbf{X}_i) \perp\!\!\!\perp \mathbf{X}_j$ using Property 3. □

What we have proved here is that provided at least one of the "decomposable" P.I.D.'s K_2, which is similar to K_1, can be interpreted as an I.D. as well, then no positivity conditions are necessary for Kiiveri's corollary, given in the last section, to hold. This is particularly useful since it is often the practice to elicit the I.D. rather than the P.I.D. from the client. In addition we have shown that their results emanated from Properties 1, 2 and 3 defining c.i. and so essentially holds for other structures such as those mentioned in Section 3.

Acknowledgments. I am indebted to the referees for their helpful and constructive comments both on the substance and the presentation of this paper.

REFERENCES

AGOGINO, A. M. (1985). Use of probabilistic inference in diagnostic expert systems. *Proc. of the 1985 ASME International Computers in Engineering Conference, Boston, Aug. 4–8* **2** 305–310.

BARLOW, R. E. and ZHANG, X. (1987). Bayesian analysis of acceptance sampling procedures discussed by Deming. *J. Statist. Plann. Inference* **16** 285–296.

BARLOW, R. E., MENSING, R. W. and SMIRIGA, N. G. (1986). Using influence diagrams to solve a calibration problem. In *Probability and Bayesian Statistics* (R. Vierl, ed.) 17–30. Plenum, New York.

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

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

GOLDSTEIN, M. (1981). Revising previsions: A geometric approach (with discussion). *J. Roy. Statist. Soc. Ser. B* **43** 105–130.

HOWARD, R. A. and MATHESON, J. E. (1981). Influence diagram. In *The Principles and Applications of Decision Analysis* **2**. Strategic Decisions Group, Menlo Park, Calif.

KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–51.

LAURITZEN, S. L. and WERMUTH, N. (1984). Mixed interaction models. Research Report R84-8, Institute for Elektroniske Systemer, Aalborg Universitetscenter, Denmark.

LAURITZEN, S. L., SPEED, T. P. and VIJAYAN, K. (1984). Decomposable graphs and hypergraphs. *J. Austral. Math. Soc. Ser. A* **36** 12–29.

OLMSTED, S. M. (1984). On representing and solving decision problems. Ph.D. dissertation, Engineering–Economic Systems Dept., Stanford Univ.

PEARL, J. (1986). Markov and Bayes networks: A comparison of two graphical representations of probabilistic knowledge. Technical Report CSD-860024, Cognitive Systems Lab., Comp. Sci. Dept., Univ. California, Los Angeles.

PEARL, J. and VERMA, T. (1987). On the logic of representing dependencies by directed graphs. In *AAAI-1987: Sixth National Conference on Artificial Intelligence* 374–379.

REGE, A. and AGOGINO, A. M. (1986). Sensor integrated expert system for manufacturing the process diagnostics. Report 86-15, Operations Research Center, Univ. California, Berkeley.

SHACHTER, R. D. (1986a). Evaluating influence diagrams. In *Reliability and Quality Control* (A. P. Basu, ed.) 321–344. North-Holland, Amsterdam.

SHACHTER, R. D. (1986b). Intelligent probabilistic inference. In *Uncertainty in Artificial Intelligence* (L. N. Kanal and J. Lemmer, eds.) 371–382. North-Holland, Amsterdam.

SMITH, J. Q. (1988a). Models, optimal decisions and influence diagrams. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 765–776. Oxford Univ. Press, New York.

SMITH, J. Q. (1988b). Statistical principles on graphs. Research Report 138, Dept. Statistics, Univ. Warwick.

SMITH, J. Q. (1989). Influence diagrams for Bayesian decision analysis. *European J. Opér. Res.* To appear.

SPIEGELHALTER, D. J. (1986). Probabilistic reasoning in predictive expert systems. In *Uncertainty in Artificial Intelligence* (L. N. Kanal and J. Lemmer, eds.). North-Holland, Amsterdam.

SPIEGELHALTER, D. J. (1987). Coherent evidence propagation in expert systems. *The Statistician* **36** 201–210.

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

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
ENGLAND