Chapter 3

Independence Graphs

After a summary of the requisite concepts from graph theory, the independence graph of a k-dimensional vector of random variables is introduced and defined. Constructed from selected independences between pairs of variables conditioned on all the remaining variables in the vector, the graph is more correctly called the conditional independence graph. Following some examples, the most characteristic feature of the graph, the global Markov property, is established by proving the separation theorem. It is this property, above all others, that permits such a concise summary of the interaction pattern between the random variables.

The material has been heavily motivated by the the lecture notes of Speed (1978a) to the Institute of Mathematical Statistics in Copenhagen, some of which later appears in Speed and Kiiveri (1986); the seminal Annals paper of Darroch, Lauritzen and Speed (1980); and the elegant exposition by Lauritzen (1982) of the ideas for contingency tables.

The proof of the separation theorem given here has the merit that its basic inductive idea is a relatively straightforward exercise in the application of the factorisation criterion for conditional independence. It is restricted to random variables with a positive joint probability density function so as to apply the lemma of block independence, discussed in the previous chapter. In this sense the proof is distribution free.

There are more elegant proofs for particular distributions: the proof given for the multivariate Normal distribution by Speed and Kiiveri (1986); for Multinomial distributions appropriate for cross-classified categorical data, Lauritzen (1982), and for the family of conditional Gaussian distributions for mixed variables, Lauritzen and Wermuth (1984, 1989). These expositions start by proving that, with respect to a given graph on a finite number of vertices, the following properties are equivalent:

- the pairwise Markov property: that non-adjacent pairs of variables are independent conditional on the remaining variables;
- the local Markov property: that conditional only on the adjacent variables, any variable is independent of all the remaining variables; and
- the global Markov property: that any two subsets of variables separated by a third is independent conditionally only on variables in the third subset.

The separation theorem states that the pairwise property implies the global property.

In this text an independence graph is defined by the pairwise Markov property. There are two essential reasons for this choice: firstly, the pairwise definition of the independence relationship naturally corresponds to the pairwise graph-theoretic definition of the edge set in a graph. Secondly, from the applied point of view, the list of requirements to be verified in constructing the independence graph to model a given data set is the least stringent, while the interpretations that follow are the strongest.

These three properties are indeed equivalent if the density is positive. However, if some of the random variables are logically or mathematically dependent, for example if $X_1 = Y$, $X_2 = Z$ and $X_3 = Y + Z$, the joint density of (X_1, X_2, X_3) is zero whenever $X_1 + X_2 \neq X_3$. In this case, the local and global properties remain equivalent, but cannot be derived from the pairwise property, see the discussion at the end of Section 2.2. One is then obliged to adopt the local Markov property for the definition of an independence graph. However, as the densities in our applications are always positive, we resist the temptation to generalise and stay with the simpler pairwise definition.

When the variables are naturally ordered, for instance, according to time or to some *a priori* causal notion, the definition of the independence graph above is inappropriate. The conditioning set for a pairwise independence containing all remaining variables is too large, and in temporal studies will include the 'future'. A theory for directed independence graphs, based on the ideas of Wermuth and Lauritzen (1983), and Kiiveri, Speed and Carlin (1984), is developed by limiting the conditioning set to the 'past'. This material is very much related to the original path analysis ideas of Wright (1921, 1923). A recent proof of the Markov properties is given in Lauritzen *et al.* (1988).

Finally the extension is made to variables that only satisfy a weaker ordering, one in which the variables can be partitioned into disjoint sets, termed blocks, which are completely ordered and so form a chain. The conditioning sets for each pairwise independence statement consist of all variables in the 'past' blocks and all the remaining variables in the 'present' block. The resultant graph is known as a chain independence graph and provides the conditional independence framework necessary for discussion of multivariate regression and simultaneous equation models.

A different and more general axiomatic approach has been developed by Pearl (1986, 1988) and his co-workers. The idea is to write down a set of rules that encapsulate the relationship of irrelevance within a set of objects. For example, the rule, $A \perp\!\!\!\perp B | C \Rightarrow B \perp\!\!\!\perp A | C$, of symmetry can be interpreted as 'if A is irrelevant to B, given knowledge of C' then 'B is irrelevant to A, given knowledge of C', and is taken as an axiom of the system. Probability with its notion of conditional independence is just one of several application areas that satisfy these axioms of irrelevance. In this context, see Oliver and Smith (1990) and Smith (1989).

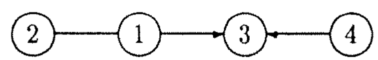
Finally we do no more than point out that the material has a deep connection to the probabilistic theory of Markov fields, see for instance Spitzer (1971), Kemeny *et al.* (1976), Speed (1979) and Isham (1981).

3.1 Graph Theory

There are several good books on graph theory, in particular we mention Harary (1969), Berge (1973), and Golumbic (1980). Fortunately its basic notions are easily understandable and a brief summary of the principal concepts and objects of graph theory will suffice for our purposes.

A *graph* G is a mathematical object that consists of two sets, a set of vertices, K, and a set of edges, E, consisting of pairs of elements taken from K. We usually take K to be the set of natural numbers $\{1, 2, \ldots, k\}$. There is a *directed edge* or *arrow* between vertices i and j in K if the set E contains the ordered pair (i, j); vertex i is a *parent* of vertex j, and vertex j is a *child* of vertex i. There is an *undirected edge* or *line* between these vertices if E contains both pairs, (i, j) and (j, i). The graph is *undirected* if all edges are undirected.

We only consider simple graphs, those without multiple edges or loops. The diagram of the graph is a picture, in which circles represent vertices, a *line* represents an undirected edge, and a *arrow* represents a directed edge. The graph with $K = \{1, 2, 3, 4\}$ and $E = \{(1, 2), (2, 1), (1, 3), (4, 3)\}$ has the diagram



Vertices i and j are *adjacent* if the undirected edge between i and j is in E, and a line connects them in the diagram of the graph. Thus 1 and 2 are adjacent, but neither pair 1 and 4, nor pair 1 and 3 are adjacent. The undirected graph G^u associated with G is the graph obtained by replacing all arrows in G by lines.

A *path* is a sequence of distinct vertices, $i_1, i_2, \ldots, i_m$ for which (i_l, i_{l+1}) is in E for each $l = 1, 2, \ldots, m - 1$; there is an arrow between each successive pair. It is a *short path* if no subsequence of the sequence is also a path. The path is a *cycle* if the end points are allowed to be the same, $i_1 = i_m$. In an undirected graph each successive pair of vertices in a path are adjacent; and the cycle is *chordless* if no other than successive pairs of vertices in the cycle are adjacent. Two vertices, i and j, are *connected* if there is a path from i to j and a path from j to i, and a graph is *connected* if all pairs of vertices are connected.

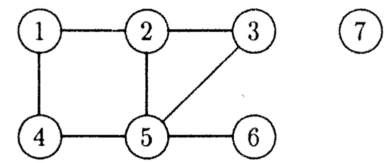
A subset of vertices *separates* two vertices, i and j, if every path joining the two vertices contains at least one vertex from the separating subset. A subset separates two subsets a and b of vertices in K if it separates every pair of vertices $i \in a$ and $j \in b$.

Let $a \subseteq K$ denote a subset of vertices of the graph. The *neighbours* of a are all those vertices in K, but not in a, that are adjacent to a vertex in a. The set of *parents* of a is $\text{pa}(a)$, the set of all those vertices in K, but not in a, that have a child in a. The *boundary* of a is $\text{bd}(a)$, the union of the neighbours and the parents of a. In an undirected graph the boundary and the set of neighbours are one and the same.

The induced *subgraph* of a, G_a, is the graph obtained by deleting all the vertices not in a from the graph on K, together with all edges that do not join two elements of a. A graph or subgraph is *complete* if all vertices are joined with either directed or undirected edges.

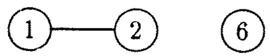
A *clique* is a subset of vertices which induce a complete subgraph but for which the addition of a further vertex renders the induced subgraph incomplete; that is, a clique is *maximally* complete.

EXAMPLE 3.1.1 The diagram of the undirected graph (K, E) with $K = \{1, 2, 3, 4, 5, 6, 7\}$ and edge set $E = \{(1, 2), (1, 4), (2, 3), (2, 5), (3, 5), (4, 5), (5, 6)\} \cup \{(2, 1), (4, 1), (3, 2), (5, 2), (5, 3), (5, 4), (6, 5)\}$ is



There are many paths from 1 to 6, and 1, 2, 5, 6 is one, but the graph is not connected as there is no path between vertex 7 and the other vertices. The boundary of vertex 1, $\text{bd}(1)$, is the set of neighbours $\{2, 4\}$, and of $\{1, 2\}$ is the set $\{3, 4, 5\}$. The cycle 1, 2, 5, 4, 1 is chordless. The cycle 1, 2, 3, 5, 4, 1 is not. The cliques of this graph are the subsets $\{1, 2\}, \{1, 4\}, \{4, 5\}, \{2, 3, 5\}, \{5, 6\}$

and {7}. The subgraphs induced by {1, 2, 3} and by {1, 2, 6} have diagrams

 and 

The first is connected, the second is not. □

3.2 Independence Graphs

Let $X = (X_1, X_2, \ldots, X_k)$ denote a vector of random variables, and $K = \{1, 2, \ldots, k\}$ the corresponding set of vertices. The graph is an independence graph, or more precisely a conditional independence graph, if there is no edge between two vertices whenever the pair of variables is independent given all the remaining variables. The vector of the remaining variables are referred to as the *rest*. We use the shorthand $1 \perp\!\!\!\perp 2 | \{3, 4\}$ for $X_1 \perp\!\!\!\perp X_2 | (X_3, X_4)$, so that the independence of X_i and X_j given the rest can be written as $i \perp\!\!\!\perp j | K \backslash \{i, j\}$; occasionally and rather loosely we write $X_i \perp\!\!\!\perp X_j |\text{rest}$. The resulting undirected graph gives a picture of the pattern of dependence or association between the variables.

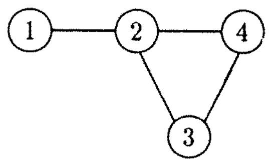
Definition. The *conditional independence graph* of X is the undirected graph $G = (K, E)$ where $K = \{1, 2, .., k\}$ and (i, j) is *not* in the edge set E if and only if $X_i \perp\!\!\!\perp X_j | X_{K \backslash \{i,j\}}$. □

Because of its Markov properties, a better name might be a *Markov* graph, but unfortunately this term is extensively used in the theory of random graphs. We remark on the importance of conditioning; there is no suitable theory for graphs constructed from pairwise marginal independences. It is often easy to construct the independence graph if we are given the joint density function of X by repeated application of the factorisation criterion for conditional independence.

EXAMPLE 3.2.1 Take $k = 4$ and consider the density function of $X = (X_1, X_2, X_3, X_4)$, $f_X(x) = \exp(u + x_1 + x_1 x_2 + x_2 x_3 x_4)$, on the four dimensional cube, $\{x; x = (x_1, x_2, x_3, x_4),\ 0 < x_i < 1,\ i = 1, 2, 3, 4\}$, where the constant u ensures the density integrates to 1. Direct application of the factorisation criterion implies

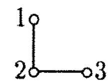
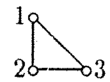
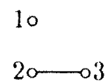
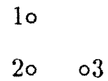
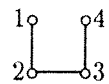
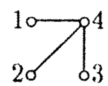
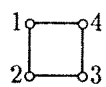
$$X_1 \perp\!\!\!\perp X_4 | (X_2, X_3) \quad \text{and} \quad X_1 \perp\!\!\!\perp X_3 | (X_2, X_4),$$

and consequently the independence graph is



The graph uses the fact that vertex 1 is not adjacent to either 3 or 4 for construction but highlights the fact that the cliques of the graph are {1, 2} and {2, 3, 4}. □

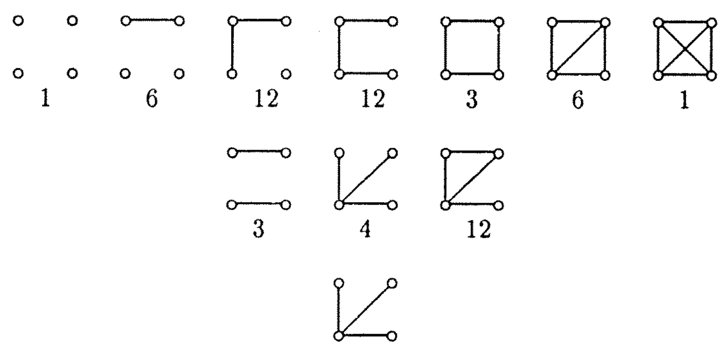
EXAMPLE 3.2.2 Some independence graphs in three and four dimensions. The graphs are undirected but, for simplicity, only one ordering for an edge pair is included in the edge set.

independences	edge set E	diagram	comments
Three dimensions			
$1 \perp\!\!\!\perp 3 \| 2$	{(1, 2), (2, 3)}		one independence
none	{(1, 2), (1, 3), (2, 3)}		complete inter-dependence
$1 \perp\!\!\!\perp 2 \| 3$ $1 \perp\!\!\!\perp 3 \| 2$	{(2, 3)}		independent subsets
$1 \perp\!\!\!\perp 2 \| 3$ $1 \perp\!\!\!\perp 3 \| 2$ $2 \perp\!\!\!\perp 3 \| 1$	{}		mutual independence
Four dimensions			
$1 \perp\!\!\!\perp 3 \| \{2, 4\}$ $1 \perp\!\!\!\perp 4 \| \{2, 3\}$ $2 \perp\!\!\!\perp 4 \| \{1, 3\}$	{(1, 2), (2, 3), (3, 4)}		a Markov chain
$1 \perp\!\!\!\perp 2 \| \{3, 4\}$ $1 \perp\!\!\!\perp 3 \| \{2, 4\}$ $2 \perp\!\!\!\perp 3 \| \{1, 4\}$	{(1, 4), (2, 4), (3, 4)}		possible latent structure
$1 \perp\!\!\!\perp 3 \| \{2, 4\}$ $2 \perp\!\!\!\perp 4 \| \{1, 3\}$	{(1, 2), (1, 4), (2, 3), (3, 4)}		a chordless 4-cycle

The assumption generally made in latent structure models is that the observed variables are independent conditionally upon the unobserved latent variable: so the example requires that variable 4 is not observed. □

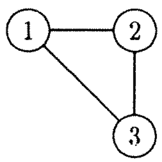
All possible graphs

In the exploratory analysis of data it may well be necessary to examine all possible graphs in order to choose one that may fit the data. In k dimensions there are $2^{\binom{k}{2}}$ different graphs, so that with $k = 4$ there are 64 such graphs. They can be classified in groups according to the number of edges together with the number of permutations:

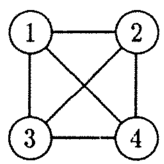


Edges	0	1	2	3	4	5	6
	1	6	12	12	3	6	1
			3	4	12		
				4			

All the graphs above have exactly 4 vertices. If the enumeration is extended to include subgraphs, for example, by including graphs based on



as well as those based on



then this figure climbs to

$$\sum_{i=0}^{k} \binom{k}{i} 2^{\binom{i}{2}}$$

When $k = 4$ this is 113.

3.3 Separation

An independence graph highlights certain aspects of information about the interdependence of the variables, in particular, whether or not two variables are adjacent, and if not, how they are separated. It is tempting to conclude that non-adjacent variables are independent given the separating set alone. The separation theorem provides the theoretical justification for this interpretation.

EXAMPLE 3.3.1 Consider the independence statements that can be made from the graph

The manner in which an independence graph is defined means that

$$X_2 \perp\!\!\!\perp X_3 | (X_1, X_4)$$

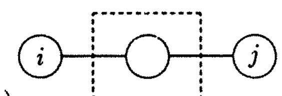
but we would like to conclude that

$$X_2 \perp\!\!\!\perp X_3 | X_1,$$

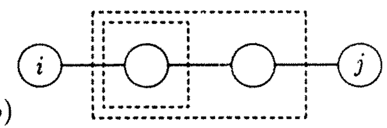
where X_4 has been dropped from the conditioning set because the graph suggests that it is redundant in explaining the dependence between X_2 and X_3. This is a statement about the marginal distribution of (X_1, X_2, X_3) and its truth is a consequence of the separation theorem. □

Before proceeding to a proof of the theorem it is worth pondering over what the term separation implies. Recall that two vertices, i and j, are separated by a subset a if and only if all paths connecting the two pass through at least one member of the subset. Consider the following example.

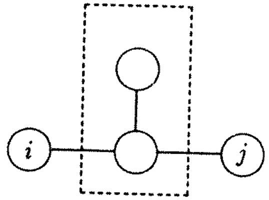
EXAMPLE 3.3.2 Separating subsets. The vertices i and j are distinct and are not members of the separating subset. In each of the following graphs the ringed subsets are separating subsets for i and j. Not all possible separating subsets are explicitly indicated.



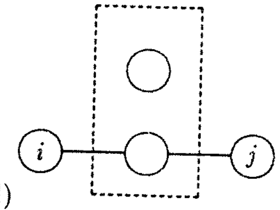
(a)



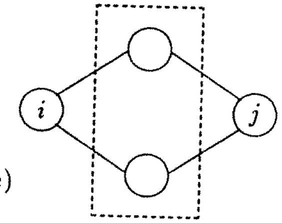
(b)



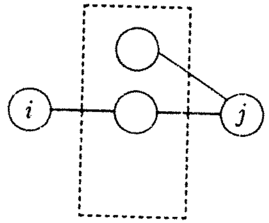
(c)



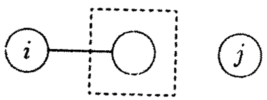
(d)



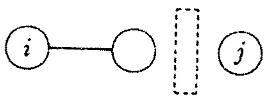
(e)



(f)



(g)



(h)

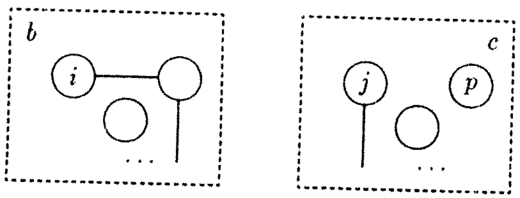
When the vertices are not connected then any subset not containing i and j is separating, including the empty set. □

The argument used here to establish the separation theorem depends upon successive application of block independence and reduction. To reveal this structure explicitly we first establish a version of the separation theorem for a graph that is composed of two disconnected subgraphs.

Lemma 3.3.1 *Suppose that the vertex set $K = \{1, 2, \ldots, k\}$ can be partitioned into two sets b and c where in the independence graph of K there is no path between any vertex in c with any vertex in b. Then*

$$i \perp\!\!\!\perp j \quad \textit{for all } i \in b \textit{ and } j \in c.$$

Proof: Bear this picture in mind:



Fix vertices i and j, select an arbitrary vertex, say p, in c and consider integrating out X_p. By construction we know that

$$i \perp\!\!\!\perp j | K \backslash \{i,j\} \quad \text{and} \quad i \perp\!\!\!\perp p | K \backslash \{i,p\}.$$

By the block independence lemma it follows that

$$i \perp\!\!\!\perp \{j,p\} | K \backslash \{i,j,p\}$$

and by the reduction lemma

$$i \perp\!\!\!\perp j | K \backslash \{i,j,p\}.$$

The vertex p has been eliminated from the conditioning set. As this independence statement $i \perp\!\!\!\perp j | K \backslash \{i,j,p\}$ holds for all i in b and for all $j \neq p$ in c, we can conclude that in the independence graph for $K \backslash \{p\}$ with k-1 vertices, there is no edge between any vertex in b with any vertex in c. This graph again has two disconnected subgraphs.

Repetition of this argument eliminates further vertices from c until c is empty. The same process is applied to b, and the process terminates as K is finite, leaving $i \perp\!\!\!\perp j$, as required. □

We remarked above that any subset is a separating set for unconnected vertices so we need the somewhat more general:

Lemma 3.3.2 *Under the same conditions as in the preceding lemma*

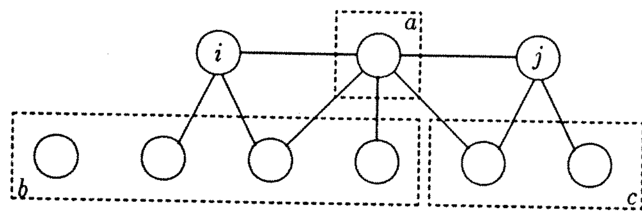
$$i \perp\!\!\!\perp j | a \quad \textit{for all } i \in b \textit{ and } j \in c$$

and any subset a of K not containing i or j.

Proof: Modify the proof of the lemma above by choosing to integrate out those variables, X_p, only for p not in a. Then generalise the argument to an arbitrary finite graph. □

Lemma 3.3.3 *If a is any subset of vertices of K that separates two vertices i and j then $i \perp\!\!\!\perp j | a$, or more precisely, $X_i \perp\!\!\!\perp X_j | X_a$.*

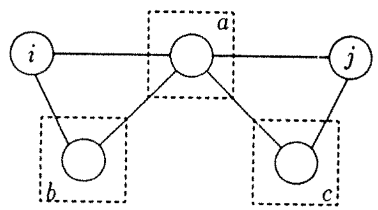
Proof: We suppose that i and j are connected, for if not the preceding lemma applies. The first part of the proof blocks certain vertices together; to help understand the construction hold the following diagram in mind:



By hypothesis vertices i and j are separated by a. The remaining vertices can be partitioned in the following manner: any other vertex is either connected or is not connected to $a \cup \{i,j\}$. If connected, it is either separated from i by a or separated from j by a or both. So partition these remaining vertices into subsets b and c where

$$\begin{aligned} b &= \{l;\ l \text{ is not connected to } a \cup \{i,j\} \textit{ or } l \text{ is separated from } j \text{ by } a\}, \\ c &= \{l;\ l \text{ is not in } b \textit{ and } l \text{ is separated from } i \text{ by } a\ \}. \end{aligned}$$

Now map the independence graph of X into the blocked graph with 5 vertices $\{i,j,a,b,c\}$, and draw an edge between any pair if an edge can exist between the elements of that pair in the original independence graph. The blocked graph must have the following 'butterfly' structure:



because, by hypothesis, there can be no edge between i and j, and by construction, no edges between i and c, j and b, and finally, between b and c.

The second part of the proof is to repeat the block independence and reduction argument. First consider two vertices p and q in c, and integrate out X_p. The construction of an independence graph implies that

$$i \perp\!\!\!\perp j | K \backslash \{i,j\} \quad \text{and} \quad i \perp\!\!\!\perp p | K \backslash \{i,p\}$$

so that application of block independence and reduction implies

$$i \perp\!\!\!\perp j | K \backslash \{i,j,p\}.$$

Thus, in the independence graph on $K \backslash \{p\}$, there is no edge between i and j. Furthermore because q is in c and all vertices in c are separated from i by a, a similar argument reveals

$$i \perp\!\!\!\perp q | K \backslash \{i,q,p\} \quad \text{for all } q \text{ in } c.$$

Thus integrating out X_p has not induced an edge between i and j, nor between i and any element in $c \backslash \{p\}$.

A similar argument establishes that there can be no induced edge between j and b nor between b and $c \backslash p$. Thus the blocked graph on the 5 vertices $\{i,j,a,b,c \backslash p\}$ has the identical 'butterfly' structure to the blocked graph on $\{i,j,a,b,c\}$. This argument can be repeated until c is empty and then reapplied and repeated on b in the same manner. The process must terminate because K is finite, leaving $i \perp\!\!\!\perp j | a$. □

The construction used for this lemma immediately generalises to random vectors and the same argument goes through with X_b and X_c replacing X_i and X_j. Under the condition that the density function be positive and with the pairwise definition of an independence graph, we have proved

Theorem 3.3.4 The separation theorem. *If X_a, X_b and X_c are vectors containing disjoint subsets of variables from X, and if, in the independence graph of X, each vertex in b is separated from each vertex in c by the subset a, then*

$$X_b \perp\!\!\!\perp X_c | X_a.$$

The converse of the separation theorem is immediate: if it is true that whenever a separates i and j that $i \perp\!\!\!\perp j | a$ then it is true that, whenever i and j are not adjacent, $i \perp\!\!\!\perp j |$rest. However it is trivial because the 'rest' is always a separating set for non-adjacent vertices and hence the conclusion forms part of the premiss to the assertion.

Minimal separating subsets

One consequence of the separation theorem is that some of the conditioning variables in an independence relation may be redundant. A conditional independence between a pair of variables is said to be *minimal* if it is not possible to apply the separation theorem to eliminate any variable from the conditioning set.

EXAMPLE 3.3.3 The conditional independences embodied in the graph

are $1 \perp\!\!\!\perp 3 | \{2,4\}$, $1 \perp\!\!\!\perp 4 | \{2,3\}$ and $2 \perp\!\!\!\perp 4 | \{1,3\}$. The minimal independences are $1 \perp\!\!\!\perp 3 | 2$, $1 \perp\!\!\!\perp 4 | 2$, $1 \perp\!\!\!\perp 4 | 3$ and $2 \perp\!\!\!\perp 4 | 3$. In this example the set of minimal independences is larger than the original set of pairwise conditional independences. □

One might conjecture that a strong converse to the separation theorem holds: if, for all non-adjacent vertices i and j and for all minimal separators a, it is true that $X_i \perp\!\!\!\perp X_j | X_a$, then it is true that $X_i \perp\!\!\!\perp X_j |$rest. However the following is a counter-example.

EXAMPLE 3.3.4 Consider the graph on three vertices

The set of all pairwise conditional independences given the rest is

$$I_{pair} = \{\, 1 \perp\!\!\!\perp 2 | 3,\ 1 \perp\!\!\!\perp 3 | 2 \,\},$$

while the set of minimal conditional independences is

$$I_{min} = \{\, 1 \perp\!\!\!\perp 2,\ 1 \perp\!\!\!\perp 3 \,\}.$$

Now block independence establishes that I_{pair} implies $1 \perp\!\!\!\perp \{2,3\}$ and hence that I_{pair} implies I_{min}, but the converse is false, as marginal independence does not imply joint independence. For instance, see the counterexample in Example 2.1.3. □

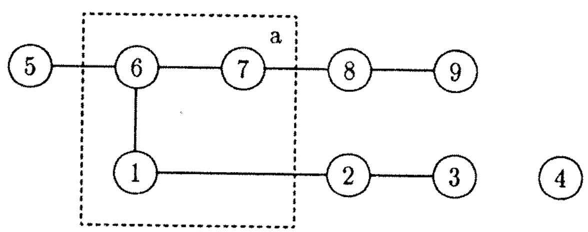
3.4 Markov Properties

In this section we take up the issue of equivalent Markov properties.

Separation and prediction

The following may be obvious but its importance makes it worth stating explicitly. Consider predicting the value of one random vector, say Y, from information on another, say Z. We take it to be axiomatic that the structural information relating Z to Y is entirely contained in the conditional density function of Y given Z, $f_{Y|Z}$, which thereby determines the form of the optimal predictor. Now suppose we are given the independence graph of a random vector X, that includes the subset, X_a $(= Y)$, which we wish to predict. What information does this graph have about the form of an optimal predictor? Let X_b denote the vector containing exactly those variables that are adjacent to at least one element of X_a, that is the boundary of a, and let X_c denote the remaining variables in X. By construction the vectors X_a and X_c are separated by X_b and so by the corollary to the separation theorem $X_a \perp\!\!\!\perp X_c | X_b$. But this independence statement is equivalent to $f_{a|b\cup c} = f_{a|b}$. Hence, given X_b, no more information for predicting X_a can be extracted from X_c.

EXAMPLE 3.4.1 Suppose X has the graph



and X_a corresponds to the set $\{1,6,7\}$, then the boundary of a is $b = \{2,5,8\}$, and the remaining vertices are $c = \{3,4,9\}$. The information on (X_3, X_4, X_9) is irrelevant for predicting (X_1, X_6, X_7). □

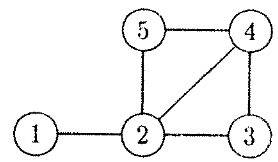
Local Markov property

The local Markov property is closely related to prediction, and expresses the independence statements in the graph in terms of a specified vertex and its nearest neighbours in the boundary set. It turns out, in treatments of graphs with infinite numbers of vertices, that this is the easiest property to generalise. A random vector with graph G has the *local Markov property*, if, for every vertex i, with boundary $a = \mathrm{bd}(i)$, and b the set of remaining vertices, then

$$X_i \perp\!\!\!\perp X_b | X_a, \quad \text{where } b = K \backslash (\{i\} \cup a).$$

A mnemonic formulation is that $i \perp\!\!\!\perp$rest|boundary. Each of the k independence statements are statements about the joint distribution of all k variables, because $\{i\} \cup a \cup b = K$. Note that the list of independences may contain some redundancies.

EXAMPLE 3.4.2 For the graph with five vertices



the full list of independences specified by the local Markov property is

$$1 \perp\!\!\!\perp \{3,4,5\}|2, \quad 3 \perp\!\!\!\perp \{1,5\}|\{2,4\}, \quad 4 \perp\!\!\!\perp 1|\{2,3,5\}, \quad 5 \perp\!\!\!\perp \{1,3\}|\{2,4\}.$$

There is redundancy here as the independence between X_3 and X_5 given (X_2, X_4) occurs twice in the list; and the first independence implies $X_1 \perp\!\!\!\perp X_5 | X_2$ so that part of the final independence, namely $X_1 \perp\!\!\!\perp X_5 | (X_2, X_4)$, is made redundant. □

To check that a set of variables is locally Markov with respect to a given graph means checking that the distribution satisfies, for each vertex i, the condition $i \perp\!\!\!\perp$rest|boundary. The converse problem, of constructing the graph given the information that the distribution of the set is locally Markov, is rather more difficult than the equivalent exercise based on the pairwise conditional independence. This is because, for any given vertex, there are 2^{k-1} ways of partitioning the remaining variables into the two sets, the boundary and the rest. Solving this problem of determining the neighbours of a given vertex, by checking for conditional independence with each of its k-1 potential neighbours, presupposes the equivalence of the pairwise independence and local Markov properties.

Equivalent Markov properties

We have used the independence of a pair of variables given the rest to define the conditional independence graph, G, for the random vector X. However there are alternative formulations:

- Firstly, the *pairwise Markov* property, that for all non-adjacent vertices i and j,

$$X_i \perp\!\!\!\perp X_j | X_a, \text{ where } a = K \backslash \{i,j\}.$$

- Secondly, the *global Markov* property, that, for all disjoint subsets a, b and c, of K, whenever b and c are separated by a in the graph, then X_b and X_c are independent given X_a alone:

$$X_b \perp\!\!\!\perp X_c | X_a.$$

It is global in the sense that the subsets, and in particular the separating set, are potentially arbitrary subsets of vertices.

- Thirdly, the *local Markov* property, that, for every vertex i, if $a = bd(i)$ is its boundary set, and b the set of remaining vertices, then

$$X_i \perp\!\!\!\perp X_b | X_a, \text{ where } b = K \backslash (\{i\} \cup a).$$

It is a remarkable fact that these are equivalent.

Theorem 3.4.1 *The three Markov properties: pairwise Markov, local Markov and global Markov, are equivalent.*

Proof: To summarise the proof of these equivalences we argue

1. The global Markov property implies the local Markov property, because the boundary set is always a separating subset.

2. The local Markov property implies the pairwise independence property, because, if the graph with vertices $K = \{1, 2, \ldots, k\}$ satisfies the local Markov property then for every vertex i, with boundary set $a = \text{bd}(i)$,

$$X_i \perp\!\!\!\perp X_b | X_a,$$

where b denotes the remaining vertices, i.e. $b = K \backslash (\{i\} \cup a)$. Select any vertex j not adjacent to i, then j is in b; put $c = b \backslash \{j\} = K \backslash (\{i,j\} \cup a)$ and rewrite the independence as

$$X_i \perp\!\!\!\perp (X_j, X_c) | X_a.$$

By the block independence lemma this is equivalent to

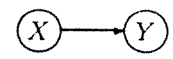
$$X_i \perp\!\!\!\perp X_j | (X_a, X_c) \quad \text{and} \quad X_i \perp\!\!\!\perp X_c | (X_a, X_j).$$

But $a \cup c = K \backslash \{i,j\}$ is the 'rest', and the first independence is exactly the pairwise independence property.

3. The separation theorem asserts that the pairwise conditional independence property implies the global Markov property. □

3.5 Directed Acyclic Independence Graphs

In many, if not most, studies of several interacting variables there is a striking lack of symmetry in the roles played by the variables that corresponds to a notion of causality and the premiss that if X causes Y then Y cannot cause X. We will not digress into the philosophy of causation, with its notions of time and time irreversibility, but we certainly wish to incorporate methods for dealing with such assymmetries. Now a neat way to portray the relationship that 'X effects Y' is by means of a directed graph and its diagram



together with the conditional probability density function, $f_{Y|X}$, a natural object of study for the probability modeller.

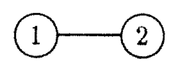
For example, suppose that in an educational study, measures are made on on the social class, X_1, and income X_2, of the head of a family, and also on the educational achievement of the eldest child, Y. The variables are not symmetric but satisfy a partial ordering. Our approach is firstly, to inquire if Y depends on both of X_1 and X_2, by assessing the independence statements $Y \perp\!\!\!\perp X_1 | X_2$ and $Y \perp\!\!\!\perp X_2 | X_1$ in the conditional distribution of Y given X; and secondly to assess the interaction between X_1 and X_2 in the distribution of (X_1, X_2) alone, without reference to Y. The conditional independence $X_1 \perp\!\!\!\perp X_2 | Y$ is of no interest.

We investigate how to represent such structures using directed independence graphs. In a directed graph a subgraph with edge (1,2) and a subgraph with the edge (2,1) have the respective diagrams

and

We start by requiring all edges to be directed. When later we do allow a mix of directed and undirected edges then we suppose

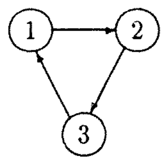
and



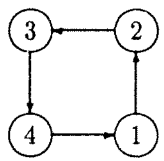
to be equivalent.

Directed cycles and orderings

Extending independence graphs to include directed edges immediately faces the problem of directed cycles such as



and



At first sight allowing directed cycles appears to have good possibilities of modelling 'feed-back', so that X_1 effects X_2 which effects X_3 which in turn effects X_1. But unfortunately, there is no suitable joint probability to model this situation; for instance, in the directed 3-cycle pictured here, we would like to express the joint density function as $f_{3|2}f_{2|1}f_{1|3}$, but apart from very special instances, this is not a well defined probability density function. Hence, by decree, we shall not allow a graph that contains a directed cycle to represent a directed version of an independence graph.

It turns out that an equivalent assumption to excluding directed cycles is to suppose that the vertices are completely ordered, that is, there exists a relation $\prec$ on the elements of $K = \{1, 2, \ldots, k\}$ such that: for all i and j in the set, (i) either $i \prec j$ or $j \prec i$, (ii) $\prec$ is irreflexive, and (iii) $\prec$ is transitive, so that if $i \prec j$ and $j \prec l$ then $i \prec l$. In this case we can write $1 \prec 2 \prec \ldots \prec k$, and think that each variable has a well defined past and future. Applied to the directed graph, an ordering means that any edge in the graph can have only one possible direction. By restricting attention to acyclic graphs, we may maintain the notion of parenthood, in which the direct antecedents of vertex i are known as the parents of i, and represented as $\text{pa}(i)$.

Lemma 3.5.1 Ordering an acyclic directed graph. *In a directed graph, the conditions that: (i) there is no directed cycle, and (ii) there exists a complete ordering of the vertices that is respected in the graph, are equivalent.*

Proof: To go one way: if there is a directed cycle in the graph then for any element, i, in the cycle, we may derive $i \prec i$, violating the irreflexive nature of the ordering. The proof of the converse is left as an exercise. □

We have just seen that such an assumption is equivalent to supposing the existence of a complete ordering for the vertices. Note that it is not usually possible to deduce the underlying complete ordering of the vertices from a particular graph, for example, the graph may have no edges at all. The order is assumed *a priori*; at no point do we suggest it is possible for the probability modeller or data analyst to determine the direction of an edge, or the implicit order underlying the graph, other than by presupposition.

Directed independence graphs

The *a priori* ordering of the vertices endows each variable with a past, present and future. We can now define a directed independence graph, termed a *recursive* graph, by Wermuth and Lauritzen (1983): the natural conditioning set for each pairwise independence statement is the past, so define $K(i) = \{1, 2, \ldots, i\}$, to be the set which comprises the past and present for the i-th variable.

Definition. The *directed independence graph* of X is the directed graph $G^{\prec} = (K, E^{\prec})$, where $K = \{1, 2, \ldots, k\}$, $K(j) = \{1, 2, \ldots, j\}$ and the edge (i, j), with $i \prec j$, is *not* in the edge set $E^{\prec}$ if and only if $j \perp\!\!\!\perp i | K(j) \setminus \{i, j\}$. □

This is the same definition used for the undirected independence graph with the conditioning set modified, from the 'rest' comprising all past and future variables, to just the past. This one crucial difference between directed and undirected independence graphs means that for an undirected graph the independence statements are statements about a single joint distribution while for a directed graph they are statements about a sequence of marginal distributions. Though, we must remark that this sequence has the property that enough information is given to define the joint distribution by virtue of the *recursive factorisation identity*:

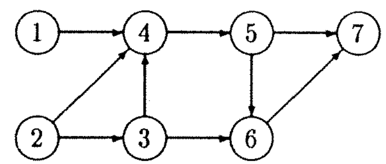
$$f_{12\ldots k} = f_{k|K(k)\setminus\{k\}} f_{k-1|K(k-1)\setminus\{k-1\}} \cdots f_{2|1} f_1.$$

Because we have an implicit *a priori* ordering of the vertices, application of the independences to evaluate the joint density function is entirely straightforward.

EXAMPLE 3.5.1 Given a seven dimensional vector X and the following pairwise conditional independences

$2 \perp\!\!\!\perp 1$	$5 \perp\!\!\!\perp 3 \| \{1,2,4\}$	$6 \perp\!\!\!\perp 4 \| \{1,2,3,5\}$	$7 \perp\!\!\!\perp 4 \| \{1,2,3,5,6\}$
$3 \perp\!\!\!\perp 1 \| \{2\}$	$5 \perp\!\!\!\perp 1 \| \{2,3,4\}$	$6 \perp\!\!\!\perp 2 \| \{1,3,4,5\}$	$7 \perp\!\!\!\perp 3 \| \{1,2,4,5,6\}$
	$5 \perp\!\!\!\perp 2 \| \{1,3,4\}$	$6 \perp\!\!\!\perp 1 \| \{2,3,4,5\}$	$7 \perp\!\!\!\perp 2 \| \{1,3,4,5,6\}$
			$7 \perp\!\!\!\perp 1 \| \{2,3,4,5,6\}$,

the independence graph has the diagram



Each pairwise independence can be immediately applied to the recursive factorisation identity to derive the form of the joint density as

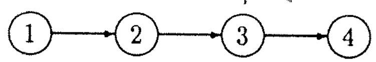
$$f_{12\ldots 7} = f_{7|56} f_{6|53} f_{5|4} f_{4|123} f_{3|2} f_2 f_1.$$

It is the ordering of the vertices, here specified numerically, that determines the conditioning set. □

Wermuth condition

We wish to elucidate the Markov properties of a directed independence graph, but hope to avoid much further work in proving such properties by exploiting the relationship of the directed graph to independence statements elicited from its associated undirected graph. If $G^{\prec} = (K, E^{\prec})$ then the *associated undirected graph* is defined as $G^u = (K, E^u)$ with the same vertex set and an undirected edge replacing each directed edge.

EXAMPLE 3.5.2 Consider the simple Markov chain defined from the independences $4 \perp\!\!\!\perp 2|\{1,3\}$, $4 \perp\!\!\!\perp 1|\{2,3\}$, and $3 \perp\!\!\!\perp 1|\{2\}$. Its diagram is



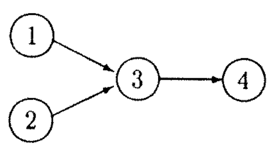
Our intuition suggests that as the path from 1 to 4 is cut by vertex 2, then $4 \perp\!\!\!\perp 1|2$, and variable 2 alone is sufficient in the conditioning set for this independence to hold. But how can we prove this? It is not directly obvious: for instance, applying block independence to the two statements $4 \perp\!\!\!\perp 2|\{1,3\}$ and $4 \perp\!\!\!\perp 1|\{2,3\}$ and deriving $4 \perp\!\!\!\perp \{1,2\}|3$ does not help, because we cannot integrate out 3 from the conditioning set.

Now the associated undirected graph, G^u, has the diagram

If we interpret this as an independence graph then firstly, three independence statements hold: $4 \perp\!\!\!\perp 2|\{1,3\}$, $4 \perp\!\!\!\perp 1|\{1,3\}$ and $3 \perp\!\!\!\perp 1|\{2,4\}$; and secondly, these three statements imply the three statements in the list defining the directed independence graph. Will this insight do the trick? □

Before launching forward, consider another example, which indicates we need to be careful.

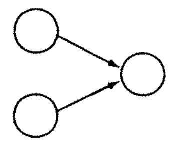
EXAMPLE 3.5.3 The independences $4 \perp\!\!\!\perp 2|\{1,3\}$, $4 \perp\!\!\!\perp 1|\{2,3\}$ and $1 \perp\!\!\!\perp 2$ define the recursive graph



If we interpret the associated undirected graph as an independence graph we could conclude that $1 \perp\!\!\!\perp 2|3$; but this independence statement cannot hold in the directed graph, because as we saw in Section 2.1, marginal independence does not imply conditional independence. □

Thus we have two examples, one where the directed and the undirected graph might well have the same independence interpretations, and another where the interpretations are certainly different. It transpires that there is an easy way to classify this behaviour:

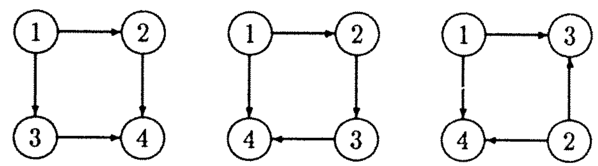
Definition. A directed graph satisfies the *Wermuth condition* if no subgraph has the configuration



□

Thus the graph of Example 3.5.2 is Wermuth, while the graph of Example 3.5.3 is not.

EXAMPLE 3.5.4 Any ordering of the (acyclic) chordless four cycle fails the Wermuth condition. There are only three distinct labellings:



and each contains a forbidden configuration. □

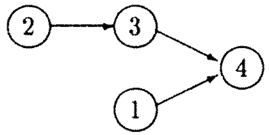
Moral graphs and Markov properties

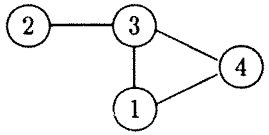
We return to the question of when the directed and the associated undirected graph have the same independence intepretations.

Definition. The *moral graph* associated with the directed graph $G^{\prec} = (K, E^{\prec})$ is the undirected graph $G^m = (K, E^m)$ on the same vertex set and with an edge set obtained by including all edges in $E^{\prec}$ together with all edges necessary to eliminate forbidden Wermuth configurations from $G^{\prec}$. □

It is termed a moral graph because it 'marries parents'; the adjective is due to Lauritzen and Spiegelhalter (1988).

EXAMPLE 3.5.5 A directed acyclic graph and its associated moral graph:





The Wermuth condition fails because variables 1 and 3, the parents of 4, are unmarried. They are married with an undirected edge and all other directions are dropped, giving the associated moral graph. This marriage does not make variable 1 a parent of 3, and so does not introduce another forbidden configuration on the subgraph $\{1, 2, 3\}$. □

Theorem 3.5.2 A Markov theorem for directed independence graphs. *The directed independence graph $G^{\prec}$ possesses the Markov properties of its associated moral graph, G^m.*

Proof: The recursive factorisation identity is simplified by application of each pairwise independence statement in the directed graph:

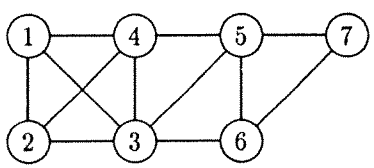
$$\begin{aligned} f_K &= \textstyle\prod_{j=2}^{k} f_{j|K(j)\setminus\{j\}} f_1 && \text{is an identity,} \\ &= \textstyle\prod_{j=2}^{k} f_{j|\mathrm{pa}\,(j)} f_1 && \text{application of independences,} \\ &= \textstyle\prod_{j=2}^{k} g_{j\cup \mathrm{pa}\,(j)} g_1 \end{aligned}$$

by appropriately defining the functions g. We thus have an expansion for the joint density function in terms of functions, g_a, which are functions of x_a for $a = \{1\}, \{2\} \cup \mathrm{pa}\,(2), \ldots, \{k\} \cup \mathrm{pa}\,(k)$. By application of the factorisation criterion to this expansion we may deduce all pairwise conditional independence statements of the form $i \perp\!\!\!\perp j|\text{rest}$. The edges of the undirected independence graph for f_K are characterised as edges between j and each of its parents, and edges between the members of each pair of parents of j. That is, the edge set of the moral graph, G^m. □

EXAMPLE 3.5.6 Example 3.5.1 continued: the seven dimensional vector X with directed independence graph given in the example above, factorised down to

$$f_{12\ldots7} = f_{7|56} f_{6|53} f_{5|4} f_{4|123} f_{3|2} f_2 f_1.$$

Its independences can be read off the moral graph, G^m, with diagram



For example $7 \perp\!\!\!\perp \{1, 2\}|\{3, 4\}$, a fact not easy to deduce from the original defining independences. □

The full moral graph can obscure certain independences. In this example, $G^{\prec}$ implies $\{3, 2\}| \perp\!\!\!\perp 1$, but the subgraph $\{1, 2, 3\}$ is complete in G^m. The fact may be deduced, though, by strengthening the assertion of the equivalence of directed and moral graphs to refer to the graph on any initial segment $K(i)$ of the vertex set, rather than just the one on the full vertex set K itself.

The following is a reformulation of a result from Wermuth (1980).

Corollary 3.5.3 *If $G^m = G^u$ then the Markov properties of the directed graph, $G^{\prec}$, are exactly identical to those of G^m.*

When the moral graph is identical to the undirected graph, so that no 'marrying' is required, then the Markov properties of the directed graph are exactly identical to those of the moral graph. We shall leave the proof as an exercise to the reader. Furthermore they will be the same as those of any directed graph whose undirected graph is identical to this moral graph; an application of this is the reversal of direction in a simple Markov chain.

EXAMPLE 3.5.7 The Markov properties of

and

are identical. □

3.6 Chain Independence Graphs

We want to extend the theory to cover independence graphs with a mixture of directed and undirected edges. Though this is a generalisation, for it must include the theory of both these special cases, its motivation is entirely practical. For example, consider an experiment on crop yield, in which the response vector records measures of plant size during the growing period, final yield and fruiting quality, all made under a variety of treatment combinations. The graph must allow the possibility of some directed edges, because it is unrealistic to imagine that final fruiting quality, say, effects the mid-term growth rate, while the converse effect is eminently reasonable. It must also allow the possibility of undirected edges, because any interaction between final fruiting quality and yield, say, may be a two-way affair.

Mathematically, the way ahead is to assume that the vertex set satisfies a particular type of partial ordering, $\preceq$, instead of the complete ordering imposed on directed graphs. The partial ordering is derived by supposing the rather strong condition that the vertex set K can be partitioned into subsets