

Directed Graphs

Mark Alan Peot

1.0 Notation

For a directed acyclic graph (or *DAG*), we will use the letter G to denote the graph instead of D used in the book. It will be understood that I_G is the appropriate separation operator for the graph that it is applied to.

The *family* for a node x in a directed graph is $\{x\} \cup \text{pa}(x)$.

A set of nodes S is an *ancestral set* if it contains the ancestors of all of its nodes, that is, $\text{an}(s) \in S$ for all $s \in S$. Define $\text{An}(B) = B \cup \text{an}(B)$ to be the *ancestral set* induced by B .

Let $\text{De}(A) = A \cup \text{de}(A)$.

The *nondescendants*, $\text{nd}(a)$, of a are those nodes that are not descendants of a , excluding the node a itself. $\text{nd}(a) = X \setminus \text{De}(\{a\})$.

The *moral graph* G^m for directed acyclic graph G contains an edge $a - b$ iff

$a \rightarrow b$ or $b \rightarrow a$ is contained in G , or
there exists a node c in G , such that $a \rightarrow c$ and $b \rightarrow c$.

Warning: [CGH] like to use the notation Π_i for the parents of node i .

2.0 Recursive factorization and D-separation

(RF) will denote the *recursive factorization property*: $P(X)$ admits a *recursive factorization* according to G , if for each $a \in X$, there exists non-negative functions $f_a(x_a, x_{\text{pa}(a)})$

(called *kernels* in [Lauritzen] or *canonical CPD's* in [CGH]), s.t. $\int_{x_a \in \mathcal{X}_a} f_a(x_a, x_{\text{pa}(a)}) = 1$

and $P(x) = \prod_{a \in X} f_a(x_a, x_{\text{pa}(a)})$.

Thm 1: If a directed acyclic graph G and distribution P obey (RF), then the probability distribution factors according to the moral graph, G^m .

Thm 2: Suppose P and DAG G satisfy (RF), then $I(A, B|C)$ when C separates A and B in $(G_{\text{An}(A \cup B \cup C)})^m$, the moral graph of the ancestral set induced by $A \cup B \cup C$.

This last theorem is equivalent to *d-separation*: (due to Verma and Pearl):

A *undirected path* (or *chain*) from a to b is an ordered sequence of two or more distinct nodes (x_0, \dots, x_n) such that $a = x_0$, $b = x_n$ and either $(x_{i-1}, x_i) \in L$ or $(x_i, x_{i-1}) \in L$ for $i = \{1, \dots, n\}$.

Node x_i of undirected path π is said to be *head-to-head* in undirected path

$\pi = (x_0, \dots, x_n)$ if $(x_{i-1} \rightarrow x_i) \in L$ and $(x_{i+1} \rightarrow x_i) \in L$.

An undirected path π is *blocked* by C if π contains a node x_i such that either

$x_i \in C$ and x_i is not head-to-head in π , or

$\text{De}(x_i) \cap C = \emptyset$, and x_i is head-to-head in π .

If undirected path π is not blocked by C , then it is *active*.

C *d-separates* A from B if all undirected paths from A to B are blocked by C .

Thm 3: (d-separation): Let A , B , and C be disjoint subsets of a DAG G . Then C d-separates A from B iff C separates A and B in $(G_{\text{An}(A \cup B \cup C)})^m$.

3.0 Markov Properties of Directed Graphs

This section is drawn from Lauritzen [97]. The separation operation $I(A, B|C)_G$ on DAG G is d-separation.

(**DP**) will denote *the directed pairwise Markov property*: (DP) holds relative to a DAG $G = (X, L)$ and joint probability distribution $P\{X\}$, if for all non-adjacent $a, b \in X$ such that $b \in \text{nd}(a)$ then,

$$I(a, b | \text{nd}(a) \setminus \{b\}).$$

(DL) will denote *the directed local Markov property*: (DL) holds if any variable is conditionally independent of its non-descendants given its parents:

$$I(a, \text{nd}(a) | \text{pa}(a))$$

(DG) will denote *the directed global Markov property*: (DG) holds relative to DAG G and JPD P if d-separation in the graph implies independence, that is:

$$I(A, B | C)_G \Rightarrow I(A, B | C).$$

In the language of (Pearl and Paz), G is a directed I-map for P .

Thm 4: (Chain rule) Every distribution has a recursive factorization according to some directed acyclic graph G .

Thm 5: (DL) \Rightarrow (DP)

Thm 6: (DF) \Leftrightarrow (DG) \Leftrightarrow (DL)

In chess, we might write (!!!) after a move that is comparable to this theorem. This is really neat. It says that if we can show the directed local or directed global properties in our belief net, we know that it **MUST** factor recursively according to the structure of the graph -- this is not true for undirected graphs unless the distributions are strictly positive. Thus, the tie between directed graphs and JPDs is ‘tighter’ than the link between undirected graphs and JPDs.

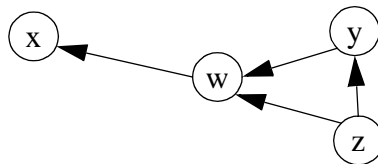
But don’t feel that undirected graphs are unimportant. We will use undirected graphs for a number of reasons, mainly as a tool for computation. The **STRONG UNION** property of undirected graphs, for example, allows us to use a single factor tree for the calculation of many different probability queries.

Thm 7: If P is positive, then (DP) \Rightarrow (DL).

Discussion:

(DP) does not imply (DL), in general.

For example, consider binary variables w, x, y, z with $x = y = z$ and w independent with distributions $P\{x = 1\} = P\{w = 1\} = 1/2$.



(DP) holds for this graph, but (DL) does not because $\neg I(x, \{y, z\} | w)$.

4.0 Abstract Independence

4.1 Properties of Models with Perfect Maps

Thm 8: (Necessary conditions on M for the existence of a perfect directed map) In order for dependency model M to have a directed perfect map G , M must satisfy the following conditions. These conditions are necessary, but not sufficient.

Symmetry: $I(X, Y|Z)_M \Leftrightarrow I(Y, X|Z)_M$

Decomposition/Composition: $I(X, W \cup Y|Z)_M \Leftrightarrow I(X, W|Z)_M \wedge I(X, Y|Z)_M$

Intersection: $I(X, W|Z \cup Y)_M \wedge I(X, Y|Z \cup W)_M \Rightarrow I(X, W \cup Y|Z)_M$

Weak Union: $I(X, Y \cup Z|W)_M \Rightarrow I(X, Y|W \cup Z)_M$

Weak Transitivity: $I(X, Y|Z)_M \wedge I(X, Y|Z \cup \{a\}) \Rightarrow I(X, a|Z)_M \vee I(Y, a|Z)_M$,

where a is a single node that is not in X , Y or Z .

Contraction: $I(X, W|Z \cup Y)_M \wedge I(X, Y|Z)_M \Rightarrow I(X, Y \cup W|Z)_M$

Chordality: $I(a, b|\{c, d\})_M \wedge I(c, d|\{a, b\})_M \Rightarrow I(a, b|c)_M \vee I(a, b|d)_M$,

where a , b , c , and d are single nodes.

Discussion:

These are not sufficient conditions. The book gives a the example of $M = \{I(x, y|z), I(y, x|z), I(x, y|w), I(y, x|w)\}$ which satisfies the seven properties of the theorem, but does not have a directed perfect map.

Weak transitivity, composition, and chordality are violated by JPDs.

4.2 Constructing DAGs from JPDs

Thm 9: (Minimal directed I-map of a dependency model) Every semigraphoid has a minimum directed I-map corresponding to an arbitrary ordering α of the variables. Let $A_i = \{\alpha(1), \dots, \alpha(i)\}$. Construct the I-map by designating as parents of each node $\alpha(i)$ any minimal set of predecessors $\pi_i \subseteq A_{i-1}$ satisfying:

$$I(\alpha(i), A_{i-1} \setminus \pi_i | \pi_i).$$

Thm 10: (Unique I-map) If the JPD is positive, the minimal sets of predecessors π_1, \dots, π_n in Theorem 9 are unique.

4.3 Causal Input Lists

Say that x_1, \dots, x_n is a set of variables. Given an ordering α of these variables, define a *causal input list* to be a set of n conditional independence statements, one for each variable, of the form $I(\alpha(i), A_{i-1} \setminus \pi_i | \pi_i)$. The DAG *generated* from this causal input list is a directed acyclic graph containing nodes x_1, \dots, x_n where the parents of each node $\alpha(i)$ in the DAG is the set π_i .

Let $\text{CIS}(G)_D$ be the set of all conditional independence statements that can be generated from a directed acyclic graph using d-separation. Let $\text{CIS}(M)_S$ be all of the conditional independence statements that can be generated from M using the semi-graphoid axioms.

Thm 11: $\text{CIS}(G)_D = \text{CIS}(M)_S$.

Thus, a DAG is a perfect map for a causal input list.

5.0 Independence Equivalence.

Two graphs G_1 and G_2 are *independence equivalent* if $I(A, B|C)_{G_1} \Leftrightarrow I(A, B|C)_{G_2}$.

For undirected graphs, the answer is easy. For directed graphs, the answer is much harder.

Thm 12: Two undirected graphs G_1 and G_2 are independence equivalent only if they are identical.

A *V-structure* in a DAG G is a triplet of nodes (x, y, z) from G , where $x \rightarrow y$, $z \rightarrow y$, but there is no edge between x and z .

Thm 13: Two DAGs G_1 and G_2 are independence equivalent if

the undirected versions G_1^- and G_2^- of these graphs are identical, and
the V-structures of both G_1 and G_2 are identical.

Thm 14: Intersection of directed and undirected graphical models. An undirected graph G_u and a directed graph G_d represent the same conditional independence state-

ments if (a) the undirected version of G_d is identical to the moral graph of G_d , (b) the undirected version of G_d is identical to G_u , and (c) G_u is decomposable.

Thm 15: All of the conditional independence relationships in an undirected graph can be represented by an undirected graph with auxiliary nodes.



$$\{I(A, D|\{B, C\})_G, I(B, C|\{A, D\})_G\} \quad \{I(A, D|\{B, C, E\})_G, I(B, C|\{A, D, E\})_G\}$$

For example, in the graph above, if we condition the independence statements represented by the DAG on an arbitrary value for E , we can obtain a list of conditional independence statements that are otherwise identical to the undirected graph on the left. This *trick* is often used to construct an undirected graphical model using a commercial belief network tool.

6.0 Bayes Nets

A *Bayesian network* (aka *Bayes net* aka *causal probabilistic model*¹ aka *belief network*²) is a pair (G, P) where G is a directed acyclic graph and P is a set of conditional probability distributions, one for each node in G , of the form $P\{x_i, \text{pa}(x_i)\}$; where $\text{pa}(x_i)$ is defined on G .

7.0 Multi-graphs

In general, a single graph, directed or undirected, cannot represent all of the independence relationships that can be derived from a JPD or other dependence model. Chapter 7 of [CHG] discusses the use of multiple graphs to represent a single model. The independence relationships represented by the multigraph are the union of the independence relationships implied by the individual graphs that comprise that multigraph. Your instructor doesn't think that multigraphs are very useful for modeling, though a distribution over graphical structures might represent the appropriate result of a model selection process.

1. Used by the Danish (Hugin creators).
 2. Used by Judea Pearl and students.