Undirected Graphical Models

Mark Alan Peot

Read: Chapters 5 and 6 of [CGH].

1.0 Goal

The first model for a joint probability distribution that we will consider is the *undirected graph*.

We will undirected graphs using two different methods. First, we will show the relationship between undirected graphs and joint probability distributions. Second, we will show the relationship between undirected graphs, and an abstract independence model (called an *input list*).

2.0 Independence, Graphoids and Semi-Graphoids

Let $I(A, B|C)$ denote probabilistic independence: A is independent of B given C in the joint probability distribution.

If $I(A, B|C)$ and $f(C = c) > 0$, then

$$f(A = a, B = b|C = c) = f(A = a|C = c)f(B = b|C = c).$$

In order to make this definition work for all values of C, we will rewrite this condition as

$$f(A = a, B = b, C = c)f(C = c) = f(A = a, C = c)f(B = b, C = c)$$

The relationship I has four properties that hold for any distribution.

P1. **Symmetry**: $I(A, B|C) \Leftrightarrow I(B, A|C)$, where $\Leftrightarrow$ is understood to mean "iff" in this context.

P2. **Decomposition**: $I(X, W \cup Y|Z) \Rightarrow I(X, Y|Z)$
(and this also implies $I(X, W|Z)$)

P3. **Weak Union**: $I(X, Y \cup W|Z) \Rightarrow I(X, W|Z \cup Y)$
(and this also implies $I(X, W|Z \cup W)$)

P4. **Contraction**: $I(X, W|Z \cup Y) \wedge I(X, Y|Z) \Rightarrow I(X, Y \cup W|Z)$
(with decomposition, this implies $I(X, W|Z)$)

If the underlying probability density function is known to be positive then an additional relationship holds:

P5. **Intersection**: $I(X, W|Z \cup Y) \wedge I(X, Y|Z \cup W) \Rightarrow I(X, W \cup Y|Z)$

To see why (P5) does not hold, in general, consider the distribution over binary variables w, x, y, z given by $P\{x = \text{true}\} = 0.5, P\{z = \text{true}\} = 0.5$ and $w = x = y$. Clearly, x is independent of w given y and z, since y determines both x and w. x is also independent of y given z and w for similar reasons. x is not independent, however, of both w and y given z.

P1-P5 are *properties* of a joint probability distribution, but we can also formulate an *abstract notion of independence* that uses these properties of independence as axioms (See Section 4.0 on page 4).

3.0 Markov Properties of Undirected Graphs

Let $I(A, B|C)_G$ denote graph separation: C separates A from B.

The following development follows Lauritzen, 97 [recommended reading]:

(P) will denote *the pairwise markov property*: (P) holds relative to a graph $G = (X, L)$ and joint probability distribution $P\{X\}$, if

$$I(a, b|X \setminus \{a, b\})_G \Rightarrow I(a, b|X \setminus \{a, b\}).$$

In words, this says: “(P) is true if nonadjacent nodes represent variables that are independent given all of the other nodes in the graph.

(L) will denote *the local markov property*: (L) holds relative to a graph G and joint probability distribution P if

$$I(a, X \setminus \text{cl}(a)|\text{bd}(a))_G \Rightarrow I(a, X \setminus \text{cl}(a)|\text{bd}(a)), \text{ where } \text{cl}(a) = \{a\} \cup \text{bd}(a).$$

In words, (L) is true when a variable is independent of all other variables in the graph given the variables on the boundary of the corresponding undirected graph.

(G) will denote *the global markov property*: (G) holds relative to graph G and JPD P if

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

Pearl, et. al, reads this implication as "Graph G is an *I-map* for joint probability distribution P: every independence relationship implied by G is true in P. The global markov property gives our definition for *U-separation*. C u-separates A and B if C separates A and B.

Finally, (F) will denote *the graph factorization property*. (F) holds relative to graph G and JPD P if P *factorizes* according to the cliques (not necessarily maximal) in G. That is,

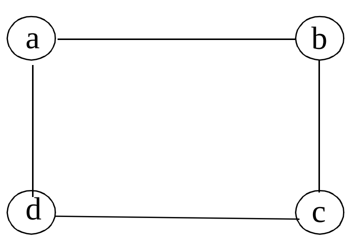
$$P\{X\} = \prod_{A \in C'} \phi(x_A), \text{ where } C \text{ is a subset of the cliques in } G.$$

OK, First of all, the first four properties of JPDs (P1-4 above) can be used to show that (G) implies (L), which implies (P).

Thm 1: (F) => (G) => (L) => (P)

Theorem 1 says that if we construct an undirected graph G representing the factors of a JPD, then u-separation in G implies independence. If a distribution factorizes according to G, G is an I-map for that distribution.

The opposite is not true, in general. Moussouris (74) gives the following example:

> Say that the following elements of the JPD for $\{a, b, c, d\}$ all have probability $1/8$:
>
> (0,0,0,0), (1,0,0,0), (1,1,0,0), (1,1,1,0)
> (1,1,1,1), (0,1,1,1), (0,0,1,1), (0,0,0,1)
>
> This distribution satisfies (G) with respect to a 4-cycle:
>
> 
>
>
> For each pair of values for (a, c), one of either b or d is determined. Thus, $I(b, d|\{a, c\})$. For each pair of values for (b, d), one of $\{a, c\}$ is determined. Thus $I(a, c|\{b, d\})$ and (G) holds for the JPD with respect to the 4-cycle. This probability distribution, however does not factorize according to the graph. For example,

$P\{0, 0, 0, 0\} = P\{a = 0, b = 0\}P\{b = 0, c = 0\}P\{c = 0, d = 0\}P\{d = 0, a = 0\} = \frac{1}{8}$ and

$P\{0, 0, 1, 0\} = P\{a = 0, b = 0\}P\{b = 0, c = 1\}P\{c = 1, d = 0\}P\{d = 0, a = 0\} = 0$ implying that

$P\{b = 0, c = 1\}P\{c = 1, d = 0\} = 0$.

Since $P\{0, 0, 1, 1\} = P\{a = 0, b = 0\}P\{b = 0, c = 1\}P\{c = 1, d = 1\}P\{d = 1, a = 0\} = \frac{1}{8}$, then $P\{b = 0, c = 1\} \neq 0$, implying that $P\{c = 1, d = 0\} = 0$. This contradicts the fact that $P\{1, 1, 1, 0\} = 1/8$. [Lauritzen, pg. 38]

Thm 2: (Pearl and Paz) If the intersection property (P5) holds for disjoint sets of variables A, B, C, D in the JPD P, then (P) => (L) => (G).

Theorem 2 says that when the intersection property holds for the JPD, the pairwise markov properties implies the local markov property and both of these properties imply the global property.

Thm 3: (Hammersley and Clifford) A probability distribution P with positive and continuous density f with respect to a product measure μ satisfies (P) iff (F) is true.

Theorem 3 says that when distributions are positive, that the pairwise markov property can be used to construct a graph that satisfies the factorization property.

4.0 Abstract Independence

Now we will use the properties (P1-P5) to form two algebraic-structures for representing independence distributions. We will refer to properties (P1-P5) as axioms for these algebras. Axioms P1-P4 will be referred to as the *semi-graphoid axioms*. Axioms P1-P5 will be referred to as the *graphoid axioms*.

A set of abstract independence relations M will be called an *input list*. Each independence relation has the structure $I(A, B|C)_M$, for disjoint sets A, B and C.

A *semi-graphoid* is a set of independence relations that are closed under the semi-graphoid axioms (symmetry, decomposition, weak union and contraction).

A *graphoid* is a set of independence relations that is closed under the graphoid axioms (symmetry, decomposition, weak union, contraction, and intersection).

Recall that all JPDs satisfy the semi-graphoid properties, thus the conditional independence relationships defined on the JPD form a semi-graphoid. Similarly, positive JPDs satisfy the graphoid properties, thus the conditional independence relations defined on a JPD form a graphoid.

A graph is a *perfect map* of a model M, if $I(A, B|C)_G \Leftrightarrow I(A, B|C)_M$.

Thm 4: A necessary and sufficient condition for a dependency model M to have an undirected perfect map is that M must satisfy:

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

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

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

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

Strong Transitivity: $I(X, Y|Z)_M \Rightarrow I(X, a|Z)_M \vee I(Y, a|Z)_M$, where a is a single node that is not in X, Y or Z.

Discussion:

Strong Union is violated by all recursive models with more than one parent. Say that a and b are marginally independent and one of the factors of a JPD is $P(c|a, b)$, then a is dependent on b, given c.

Strong union implies weak union.

Strong union and intersection imply contraction.

Thus, these conditions are stronger than the graphoid properties.

The main point: It is a pretty odd-looking distribution that had independence properties that can be represented exactly using an undirected graph.

5.0 The Undirected Graphical Models

A undirected graphical model, or *Markov Net*, is a pair (Φ, G), where G is an undirected graph and Φ is a set of factors corresponding to the cliques in G. Since we construct the graph to reflect the structure of the factors in Φ, graph separation implies independence and vice versa.

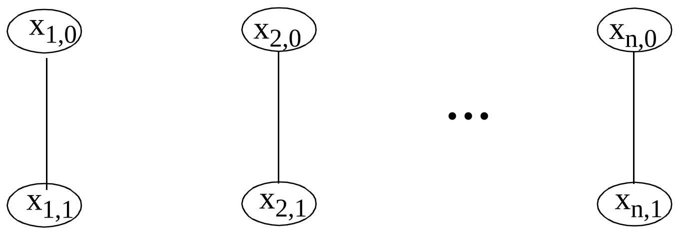
We will see that the undirected graph will be our workhorse for the calculations of marginals. We can assess an undirected graph directly, but will usually find that it is more convenient to assess a directed graphical model and convert it into its undirected (moralized) form.

6.0 Interesting Results

Noam Shazeer and Michael Littman point out that there can be a lot of cliques in an undirected graph. In fact, the problem of finding the largest clique in a graph is known to be NP-complete.

Thm 5: (Thanks to Shazeer and Littman) There may be $2^{n/2}$ maximal cliques in an undirected graph with n nodes.

We can show this via constructing the following bipartite graph:



There are an exponential number of minimal covers for this graph: a cover will have one node from every vertical "stick" in the above graph and there are 2^n ways to select this cover.

Now consider the complement to this graph. Since each minimal cover corresponds to a maximal clique, there are an exponential number (2^n)of maximal cliques.

Now this bothered me a lot. It doesn't seem that factoring a JPD into an exponential number of factors can possibly result in a compact representation for the joint probability distribution. Fortunately, this intuition is wrong:

Thm 6: (Shazeer and Peot, but reportedly an older result...) Say that we factor a distribution over *n* variables according to the maximal cliques of an undirected graph. The number of independent parameters in the factored distribution is always smaller than the number of independent parameters in the full joint distribution.

We most often encounter UGs in connection with directed acyclic graphs. Fortunately, it is easy to count the maximal cliques in these distributions.

Conjecture 7: The number of maximal cliques in a moralized directed acyclic graph is linear in the number of variables.

It is also easy to count the number of maximal cliques in a decomposable graph.

Thm 8: The number of maximal cliques in a decomposable graph is linear in the number of variables.

7.0 Minimality

7.1 I-Maps

A graph G is an *I-map* of a dependency model M if $I(A, B|C)_G \Rightarrow I(A, B|C)_M$.

> A complete graph is a trivial I-map.

Define $D(A, B|C)$ to mean that $\neg I(A, B|C)$. A graph G is a *D-map* of a dependency model M if $D(A, B|C)_G \Rightarrow D(A, B|C)_M$.

A graph G is a *minimal I-map* for M if it is no longer an I-map if any edge is deleted.

A graph G is a *perfect map* of M if $I(A, B|C)_G \Leftrightarrow I(A, B|C)_M$.

Easy theorem: A perfect map is both an I-map and a D-map.

7.2 Minimal I-maps

Thm 9: If M is a graphoid, the unique minimal undirected I-map for M is the $G = (X, L)$, where $(a, b) \notin L$ iff $I(a, b|X \setminus \{a, b\})_M$.

> *Discussion*: Since (P) does not imply (G) unless the intersection property holds, M must be a graphoid.

8.0 Decomposition and Undirected Graphs

Running Intersection Property. Say that C is a set of cliques in an undirected graph. The sequence $(C_1, \ldots, C_n)$ obeys the *running intersection property* if the set $C_k \cap (C_1 \cup \ldots \cup C_{k-1})$ is contained in at least one of the previous cliques $C_1, \ldots, C_{k-1}$.

Thm 10: There is a sequence of cliques obeying the running intersection property iff the graph is decomposable.

A *chain of cliques* obeying the running intersection property can be derived from a perfect numbering of the nodes in a triangulated graph. Assign a number to the clique equal to the largest “perfect number” of any of its nodes. Order the cliques in ascending order of this assigned number, breaking ties arbitrarily.

OK, now say that we have a JPD f over a set of variables and we know that the density is positive everywhere. We can derive an undirected graph by using the pairwise Markov property (P) to determine independence:

$$I(a, b | X \setminus \{a, b\}) \Leftrightarrow f(X \setminus \{a\}) f(X \setminus \{b\}) = f(X) f(X \setminus \{a, b\}) \qquad \text{(EQ 1)}$$

Well, we can do this by computing the marginals given in Equation 1.

$$f(X \setminus \{a\}) = \sum_a f(X) \qquad \text{(EQ 2)}$$

$$f(X \setminus \{b\}) = \sum_b f(X) \qquad \text{(EQ 3)}$$

$$f(X \setminus \{a, b\}) = \sum_{a, b} f(X) \qquad \text{(EQ 4)}$$

After determining these marginals, we can use Equation 1 to test for pairwise independence.

OK, given this graph, we know that we can factor f into potentials on each clique, but these potentials do not have a clear meaning. One of the nice properties of decomposable graphs is that we can give these potentials a probabilistic interpretation:

Thm 11: Say that undirected graph has a sequence $(C_1, \ldots, C_n)$ of cliques obeying the running intersection property. Define $S_k = C_k \cap (C_1 \cup \ldots \cup C_{k-1})$ and $R_k = C_k \setminus S_k$, with $S_1 = \varnothing$.

Call $S = \{S_1, \ldots, S_n\}$ the set of *separators*. Call $R = \{R_1, \ldots, R_n\}$ the set of *residues* Then the JPD factors into:

$$f(X) = \frac{\prod_{k=1,n} f(C_k)}{\prod_{k=1,n} f(S_k)}$$

or equivalently,

$$f(X) = \prod_{k=1,n} f(R_k | S_k).$$

This is cool! If a graph is decomposable, the factors have a probabilistic interpretation.