Undirected Graphical Models - Representation

What are undirected graphical models?

Consider a random vector $X = (X_1, \ldots, X_p)$ with a multivariate distribution P_X. An undirected graphical model is a multivariate distribution together with an undirected graph that encodes (a subset of) conditional independence[1] relations implied by this distribution. The undirected graph $G = (V, E)$ associated with the distribution consists of $|V| = p$ nodes (each node is associated with a random variable), and the edges E in the graph encode the conditional independence relations as described next.

Undirected graphical models are also referred to as Markov random fields or Markov networks, due to the markov properties discussed below [2].

What conditional independence relations are encoded by the graph?

Pairwise Markov Property - An edge between two nodes X_i and X_j is absent in the graph if and only if X_i and X_j are conditionally independent given the other variables, i.e. $X_i \perp X_j | X_{\setminus i \setminus j}$.

Notice that the complete graph encodes no conditional independence assumptions. It is the *absence of edges* that makes a graphical representation useful for describing the distribution.

Are pairwise Markov relations the only conditional independence relations encoded by the graph? No, there are more. We can read off all the conditional independence relations encoded by the graph as follows.

Global Markov Property - For any disjoint vertex subsets A, B, and C in the graph G such that C separates A and B (i.e. every path between a node in A and a node in B passes through a node in C), the random variables X_A are conditionally independent of X_B given X_C, i.e. $X_A \perp X_B | X_C$, where $X_A = \{X_v\}_{v \in A}$.

We can also define a set of local Markov independencies as follows.
Local Markov Property - The random variable associated with a node X_i is independent of the rest of the nodes in the graph given its immediate neighbors $\mathcal{N}_i$, i.e. $X_i \perp X_{\setminus i \setminus \mathcal{N}_i} | X_{\mathcal{N}_i}$.

[1] When we refer to *conditional* independence relations, we also include any *independence* relations (conditioned on the empty set).

[2] Sometimes people refer to a collection of random variables as a Markov random field, which basically implies that they are distributed according to a distribution which admits an undirected graphical model representation.

Notice that pairwise Markov property is weaker than local Markov property which is weaker than the global Markov property. However, for positive distributions $P_X > 0$ (i.e. it does not assign zero probability to any assignment of the variables),the three properties are equivalent.

Is there an easy characterization of distributions that satisfy the conditional independence relations encoded by a graph?

Let I_G denote the set of all conditional independence relations encoded by the graph G, i.e. relations you can read off using the Global Markov property. Let I_P denote the set of all conditional independence relations implied by the distribution P_X. A graph G is called an **I-map** for a probability distribution P_X if all conditional independence relations implied by G hold true for P_X, i.e. $I_G \subseteq I_P$.

Definition. A probability distribution factors with respect to a graph in case it can be written as a product of factors, one for each of the cliques $\mathcal{C}$ in the graph:

$$P(X_1, ..., X_p) = \Pi_{\mathcal{C}} \phi_{\mathcal{C}}(X_{\mathcal{C}}).$$

Theorem. For any undirected graph G, a distribution P_X that factors with respect to the graph also satisfies the global Markov property on the graph. Equivalently, G is an I-map for P_X.

The other direction is known as Hammersley-Clifford Theorem and only holds for positive distributions.
Theorem (Hammersley-Clifford-Besag): If $P_X > 0$ is a positive distribution (does not assign zero probability to any assignment of the variables), then if G is an I-map for P_X, then P_X factorizes over G.

Can the graph encode all conditional independence relations implies by the distribution?

If P_X factorizes over G, then G is an I-map for P_X, i.e. $I_G \subseteq I_P$. The converse is not necessarily true, i.e. there are conditional independence relations implied by the distribution P_X that may not be encoded by the graph (its I-map). For example, a trivial example is the complete graph which is an I-map for every distribution (since the set of conditional independence relations implied by G, $I_G = \emptyset$), but it encodes no conditional independence relations implied by P_X. Therefore, we are usually interested in the **minimal I-map**, i.e. an I-map from which none of the edges could be removed without destroying its I-mapness. It is reasonable to require this as it leads to the most compact graph representation, which involves as few dependencies as possible. Every distribution has a unique minimal I-map which can be found as follows:

> Let $P > 0$. Define G to be the graph obtained by introducing edges between all pairs of variables X_i, X_j such that $X_i \perp X_j | X_{\setminus i \setminus j}$. Then G is the unique minimal I-map.

We may now ask if the minimal I-map encodes all of the independencies implied by P_X. If G implies no dependencies that are not indicated by P_X, then G represents precisely the set of conditional independence relations indicated by P_X, i.e. $I_G = I_P$. In this case, G is called a **perfect map** of P_X, or alternatively P_X is said to be **faithful** to G. The following theorem due to Meek says that almost all multinomial and multivariate normal distributions are faithful to their minimal I-map. This result probably extends to all parametric distributions, but I don't know of such a proof.

Theorem (Meek): The set of multinomial and multivariate normal distributions which are unfaithful to their I-map (graph G) has measure zero, i.e. the set of such parameters has Lebesgue measure zero.

However, there exist distributions for which the perfect map may not exist. For example, consider the following distribution:

$$P(A, B, C) = P(A)P(B)P(C|A, B)$$

In the undirected graph corresponding to this distribution, we must have an edge between A and C, and between B and C. Can we omit the edge between A and B? No, because A and B are dependent given C. Therefore, the only minimal I-map for this distribution is the fully connected graph, which does not capture the marginal independence ($A \perp B$) that holds in P.

Note: Meek's result should be taken with a grain of salt - it only says that if we consider a fixed graph and a set of parameters from the multinomial and multivariate normal distributions. Now consider any smooth distribution (dominated by the Lebesgue measure) over the possible parameter values. Then the probability of drawing an unfaithful distribution is zero. While this result has clear implications for the existence of faithful distributions and strong completeness of global markov property, it does not imply that a graphical model can be learnt easily from data. In fact the issue of reliably inferring a graphical model from data has to do with near violations of faithfulness.

Can all distributions be represented as a graphical model?

Yes, notice that any distribution can be trivially represented by the complete graph, as discussed above. But this representation is useless as the complete graph does not encode any conditional independence assumptions that might be implied by the distribution.

In literature, you might read statements like *"there are distributions that cannot be represented by a (directed/undirected) graphical model"*. What this means is that these distributions do not have a perfect-map.

Can all graphs correspond to the graphical representation of a distribution?

Yes. Given a graph, we can define any clique potentials along with a corresponding normalization factor which specifies a distribution on the nodes of the graph. Since the distribution factors over the graph, by the Hammersley-Clifford theorem, such a distribution satisfies all the conditional independence relations encoded by the graph, and hence the graph is an I-map for the distribution. It may however not be a perfect map.

Does the graphical model capture all properties of the distribution?

No. There are many properties that P_X can have that are not represented in the graph (i.e. many properties besides independence relations such as interactions etc.). As an example, a distribution can be written over cliques or maximal cliques and therefore, does not necessarily capture all higher-order interactions.