
Graphical Models

Contents

1	Introduction	2
2	Directed graphical models	2
2.1	Examples	3
3	Undirected graphical models	4
4	DGMs versus UGMs	5
5	Conditional independence criteria	6
5.1	Undirected graphical models	6
5.2	Directed graphical models	6

Graphical models are a way of visually representing conditional independence properties among random variables. A graphical model is useful for communicating and understanding the relationships between variables in a model, for deriving conditional independence properties of a model, and for developing inference algorithms. (Note: Ideally, these notes would be accompanied by diagrams of graphical models, however, for now they will only provide the mathematical formulas. You will need to refer to Bishop chapter 8 or Murphy chapter 19 for examples and visualizations of graphical models themselves.)

1 Introduction

- We will consider two types of graphical models: directed graphical models (DGMs) and undirected graphical models (UGMs).
- Directed graphical models are sometimes also called Bayesian networks (although sometimes this has a slightly different meaning).
- Undirected graphical models are sometimes also called Markov random fields (MRFs).
- I'm not a big fan of any of these names, since they are all misleading in some way. First of all, it is important to note that a DGM or UGM does not completely specify a model, but rather, a set of conditional independence properties. There is a whole class of potential models corresponding to any given “graphical model”. Also, there is nothing necessarily Bayesian about graphical models.
- Better names would be “directed factorization graphs” and “undirected factorization graphs” (but these are not standard terms, so if you use them, you will need to clarify what you're talking about).
- It is also worth noting that a graphical model is purely a conceptual device—mathematically speaking, the graphical model doesn't give you anything more than what you could derive by just applying the rules of probability using the factorization implied by the graphical model.

2 Directed graphical models

- Directed graphical models (DGMs) arise naturally in the context of Bayesian models, particularly hierarchical models and time series.
- A *directed graph* is a graph in which all the edges are directed (i.e., have “arrow heads”). A directed graph with n vertices can be represented by an $n \times n$ square matrix in which entry (i, j) is 1 if there is an edge from i to j , and 0 otherwise.
- A *directed acyclic graph (DAG)* is a directed graph in which there are no “cycles”, i.e., there is no way to start at one vertex, follow a sequence of edges in the directions that they point, and end up back at the same vertex.
- The set of *parents of vertex i* , denoted $\text{pa}(i)$, is the set of vertices j such that there is a directed edge from j to i .
- Suppose $p(x_1, \dots, x_n)$ is a probability distribution, and G is a DAG with vertices $1, \dots, n$. We say that p *respects* G if

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\text{pa}(i)}).$$

2.1 Examples

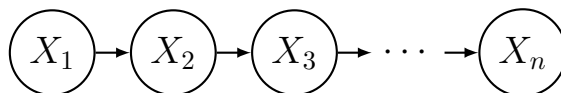
- *Markov chain.* A Markov chain is a distribution on x_1, \dots, x_n that respects the following directed graph:



In other words, it factors as

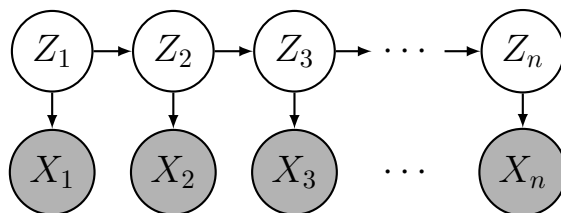
$$p(x_1, \dots, x_n) = p(x_1) \prod_{i=2}^n p(x_i | x_{i-1}).$$

Often, it is more convenient to label the vertices of the graph by the names of random variables themselves, rather than their indices, for example:



We will use both conventions interchangeably.

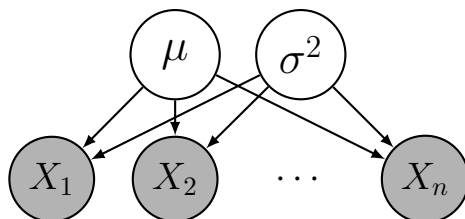
- *Hidden Markov model.* A hidden Markov model is a distribution on $(x_{1:n}, z_{1:n})$ that respects the following directed graph:



- Often, some of the random variables in a model will be part of the observed data, and some of them will be unobserved (hidden variables). A common practice is to indicate which variables are observed by shading them, as above.
- *Normal model.* Consider the following model:

$$\begin{aligned} X_1, \dots, X_n | \mu, \sigma^2 &\text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma^2) \\ \mu &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\ \sigma^2 &\sim \text{InverseGamma}(a, b). \end{aligned}$$

This respects the following directed graph:



3 Undirected graphical models

- Undirected graphical models (UGMs) arise naturally in physics, image analysis, spatial statistics, and biochemistry, for example. UGMs are often useful when there are a number of interacting variables that do not necessarily have a natural ordering in their dependencies.
- An *undirected graph* is a graph in which all of the edges are undirected (i.e., have no “arrow heads”).
- A *clique* is a subset of vertices that is fully connected, i.e., there is an edge between every pair of vertices in the subset. Denote the set of all cliques by \mathcal{C} .
- A *maximal clique* is a clique that is not contained in any other clique (i.e., adding any vertex would make it no longer clique). Denote the set of maximal cliques by \mathcal{C}_m .
- Suppose $p(x_1, \dots, x_n)$ is a probability distribution, and G is an undirected graph with vertices $1, \dots, n$. We say that p *respects* G if

$$p(x_1, \dots, x_n) = \prod_{C \in \mathcal{C}_m} \varphi_C(x_C)$$

for some functions $\varphi_C(x_C) \geq 0$. Here, as usual, $x_C = (x_i : i \in C)$. Note that we use a subscript C on each function φ_C to indicate that a different function may be used for each clique.

- There are a couple of equivalent representations that are sometimes useful:
 1. p respects G if and only if

$$p(x_1, \dots, x_n) \propto \prod_{C \in \mathcal{C}_m} \tilde{\varphi}_C(x_C)$$

for some functions $\tilde{\varphi}_C(x_C) \geq 0$.

2. p respects G if and only if

$$p(x_1, \dots, x_n) = \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

for some functions $\psi_C(x_C) \geq 0$. (Here, the product is over all cliques, not just maximal cliques.)

- Please see the textbook for examples.

4 DGMs versus UGMs

- It is natural to wonder whether there would be any disadvantage to always using directed graphical models rather than undirected, or vice versa. The answer is that sometimes DGMs are preferable, sometimes UGMs are preferable, and sometimes they are equivalent.
- It's also important to note that neither DGMs nor UGMs are “complete”, in the sense that there are some distributions whose conditional independence properties cannot be fully captured by either class.
- Sometimes a DGM and a UGM represent exactly the same set of conditional independence properties. For example, this is the case in a hidden Markov model:

$$p(z, x) = p(z_1)p(x_1|z_1) \prod_{i=2}^n p(z_i|z_{i-1})p(x_i|z_i)$$

if we consider the natural DGM and UGM corresponding to this factorization.

- Sometimes a DGM more fully captures the conditional independence properties of a distribution. For example, consider a typical normal model with independent priors on the mean and variance:

$$\begin{aligned}\mu &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\ \sigma^2 &\sim \text{InverseGamma}(a, b) \\ X|\mu, \sigma^2 &\sim \mathcal{N}(\mu, \sigma^2)\end{aligned}$$

where $\mu_0 \in \mathbb{R}$, $a, b, \sigma_0^2 > 0$. The natural DGM corresponding to this distribution is $\mu \rightarrow X \leftarrow \sigma^2$, and it turns out that this DGM captures all of the conditional independence properties of this distribution. On the other hand, there is no UGM that captures all of the conditional independence properties of this distribution.

- Sometimes a UGM more fully captures the conditional independence properties of a distribution. For example, consider a one-dimensional Ising model with periodic boundary:

$$p(x_1, \dots, x_n) \propto \exp(\beta(x_1x_2 + x_2x_3 + \dots + x_{n-1}x_n + x_nx_1)) \mathbb{1}(x_1, \dots, x_n \in \{-1, 1\}),$$

where $\beta > 0$. The natural UGM corresponding to this distribution has edges between i and $i + 1$ for each $i = 1, \dots, n - 1$, and an edge between n and 1. It turns out that this UGM captures all of the conditional independence properties of this model. On the other hand, there is no DGM that captures all of the conditional independence properties of this model.

- For more on this, see Murphy chapter 19.

5 Conditional independence criteria

5.1 Undirected graphical models

- Separation criterion for undirected graphical models: Suppose $p(x_1, \dots, x_n)$ respects an undirected graph G . Let A, B, C be disjoint subsets of vertices of G . If all paths from A to B are blocked by C (i.e., if every path from a vertex of A to a vertex of B passes through a vertex of C), then $p(x_A, x_B | x_C) = p(x_A | x_C)p(x_B | x_C)$ (i.e., $X_A \perp X_B | X_C$).
- Meanwhile, if there is a path from A to B that is not blocked by C , then we cannot determine, from G alone, whether or not $X_A \perp X_B | X_C$.

5.2 Directed graphical models

- Suppose G is a DAG. The *moralization of G* (a.k.a. the moral graph, or the moralized graph) is obtained by connecting all pairs of parents of each vertex, and then converting all directed edges to undirected edges.
- Suppose G is a DAG. The *set of ancestors* of a subset of vertices S is the set of all vertices i for which there is a directed path from i to a vertex of S . By convention, the vertices of S are included in the set of ancestors of S .
- The *subgraph of ancestors of S* is the DAG obtained by removing all non-ancestors and their edges.
- Moral ancestral separation criterion for directed graphical models: Suppose $p(x_1, \dots, x_n)$ respects a DAG G . Let A, B, C be disjoint subsets of vertices of G . Let G_{MA} be the moralization of the subgraph of ancestors of $S = A \cup B \cup C$. If all paths in G_{MA} from A to B are blocked by C , then $X_A \perp X_B | X_C$.
- Meanwhile, if there is a path in G_{MA} from A to B that is not blocked by C , then we cannot determine, from G alone, whether or not $X_A \perp X_B | X_C$.