

## 5 Minimal I-Maps, Chordal Graphs, Trees, and Markov Chains

Recall that some kinds of structure in distributions cannot be efficiently captured with either directed or undirected graphical models. An example of such a distribution is

$$p_{xyz}(x, y, z) = f_1(x, y) f_2(y, z) f_3(x, z). \quad (1)$$

This is captured by (cf. Hammersley-Clifford) an undirected complete graph on 3 nodes. And since such a complete graph captures any distributions, it fails to capture the special *factorization* structure in the distribution. By contrast, this structure *can* be well captured by a factor graph in which there are three factors, one each for  $f_1$ ,  $f_2$  and  $f_3$ . In this respect, we say that a factor graph can provide a finer grained representation. The “complexity” of a factor graph representation is  $|F||\mathcal{X}|^D$  where  $D$  is the size of the largest factor (i.e. number of variables participating in the factor).

Now all the graphical models are universal, in the sense, that each of them can be used to represent any distribution. However, the goal is not merely representation but one of trying to find the “most efficient” representation in terms of its ability to capture the structure of the underlying distribution. To begin to quantify such a notion of efficiency, we introduced the concepts of I-map, D-map and P-map. The basic idea was that P-map captures the conditional independence structure of the distribution precisely. An I-map for a given distribution implies no more independencies than the distribution satisfies.

Typically, we are interested in I-maps that represent a family of distributions that is as small as possible (subject to the constraint that it contains our distribution of interest.) For this purpose, we have the following definition:

**Definition 1** (Minimal I-Map). *A minimal I-map is an I-map with the property that removing any single edge would cause the graph to no longer be an I-map.*

In a similar manner, one can define a *maximal D-map*.

### 5.1 Generating Minimal I-Maps

Consider a distribution  $D$  for which we wish to generate a minimal I-map.

We first consider the construction of a minimal directed I-map. Say there are  $N$  variables  $x_1, \dots, x_N$  with chain rule representation being

$$p_{\mathbf{x}}(\mathbf{x}) = p_{x_1}(x_1)p_{x_2|x_1}(x_2|x_1) \cdots p_{x_N|x_{N-1}\dots x_1}(x_N|x_{N-1} \dots x_1). \quad (2)$$

Now for each component (conditional probability), reduce it if conditional independence exists, i.e. remove a variable in the conditioning if the distribution’s conditional

independencies allows for it. Inherently, the minimal I-map satisfies “local minimality” and therefore such a “greedy” procedure leads to one such minimal I-map.

Now consider the construction of a minimal undirected I-map. Let us start with a valid representation for the given distribution. Iteratively, check, for all edges  $(i, j)$  if  $x_i \perp\!\!\!\perp x_j | x_{rest}$ , and if edge  $(i, j)$  satisfies such conditional independence, then remove it. Continue doing this iteratively until one cannot remove an edge. The algorithm can take at most  $O(N^3)$  time.

**Theorem 1.** *If  $p > 0$ , the resulting undirected graphical model is the unique undirected minimal I-map for  $p$ .*

For some distributions, the resulting minimal I-map obtained in this way will actually be a P-map. In this case, the associated directed or undirected graph is a perfect fit for the distribution of interest. Moreover, a further subset of distributions will have *both* directed *and* undirected P-maps. Such distributions (and the associated graphical models) are particularly special, and allow for highly efficient inference, as we will see. In the meantime, let’s develop the characteristics of such distributions (and graphs).

## 5.2 Moralization and Triangulation

We begin by determining conditions such that converting directed graphical model to undirected graphical model does not lead to the introduction of “inefficiency.” First recall the conversion: 1) retain all edges from the directed model in the undirected one; and 2) connect ever pair of parents of a node with an edge (a process quaintly called “moralization.” The resulting of this procedure is called the *moralized graph*.

**Theorem 2.** *The moralized undirected graphical model of a given directed graphical model is a minimal I-map for that family of distributions represented by the directed graphical model. Further, if the moralization does not add any edges, the moralized graph is a P-map.*

To understand the later part of the above stated result, we need to understand the conversion from undirected graphical model to directed graphical model. This requires addition of “chords.” Specifically, for a loop (cycle) in a graph, a “chord” in the loop is an edge connecting two non-consecutive nodes (of the loop). A graph is called *chordal* if any cycle of the graph of size  $\geq 4$  has a chord.

**Theorem 3.** *If a directed acyclic graph is a minimal I-map for an undirected graphical model, then the directed graph must be chordal.*

This suggests that to obtain a directed graphical model out of an undirected graphical model, one needs to add “chords.” The detailed procedure is referred to as

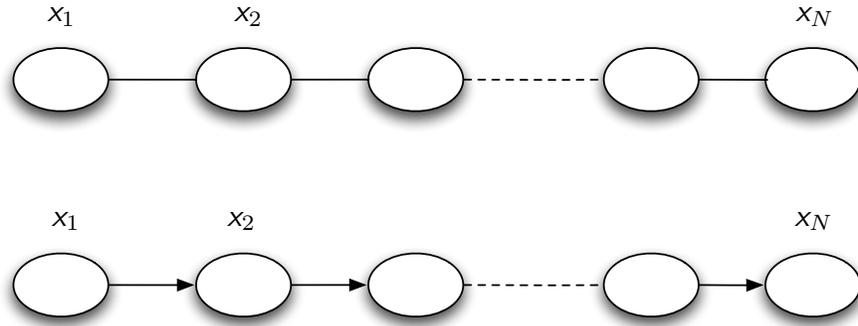


Figure 1: Markov chain in undirected and directed representations.

“triangulation” of the graphical.<sup>1</sup> We shall study this procedure in detail soon, when we begin investigating efficient inference.

From the above development, we can now deduce when a set of conditional independencies of distribution have both directed and undirected P-maps. In particular, an undirected graph  $\mathcal{G}$  has a directed P-map if and only if  $\mathcal{G}$  is chordal. Alternatively, a directed graph  $\mathcal{H}$  has an undirected P-map if and only if moralization of  $\mathcal{H}$  does not add any edges.

## 5.3 Simple Examples of Chordal Graphs

### 5.3.1 Markov Chains

Perhaps the simplest example of a chordal graph corresponds to a Markov chain, as shown in Figure 1. In the figure, the directed graphical model and undirected graphical model have exactly the same “structure” (ignoring directionality of edges). And it is chordal, because there are no loops or cycles in the graph. Another such example is that of hidden Markov model that we have seen before (also shown in Figure 2). The random variables of interest are  $x_1, \dots, x_N$  and the observations are  $y_1, \dots, y_N$  where  $y_i$  is dependent on  $x_i$ ;  $x_i$  immediately depends on  $x_{i-1}$ . The undirected and directed graphical representation for both of these distributions is the same and both of these graphical models are chordal (again, because it does not have any loop).

### 5.3.2 Trees

A very slightly richer class of chordal graphs are tree models. Consider an undirected graph as shown in Figure 3. To make its directed graph counterpart, start with any

<sup>1</sup>It is important to emphasize that the term “triangulation” is a bit generic in that there are many ways one might imagine creating triangles of nodes in a graph, not all of which result in a chordal graph. We use the term to refer to a very particular procedure of creating a chordal graph.

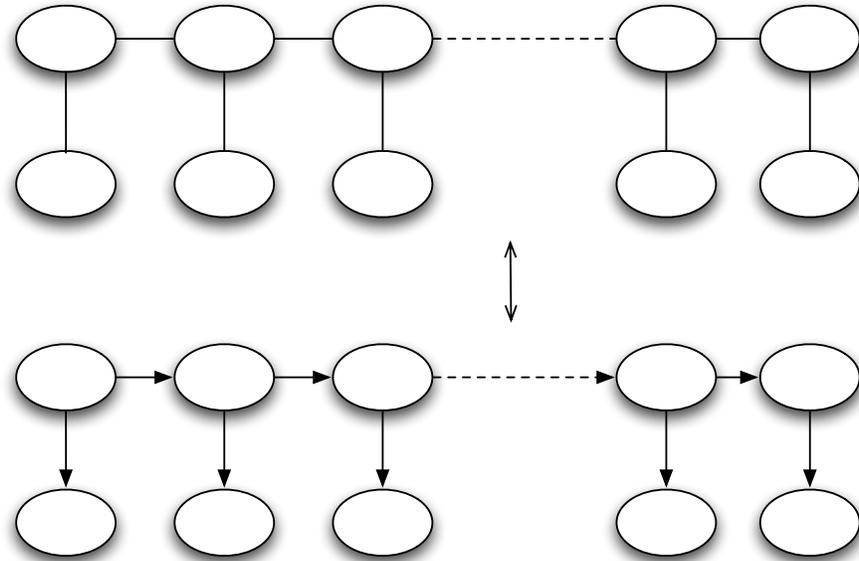


Figure 2: Hidden Markov chain in undirected and directed representations.

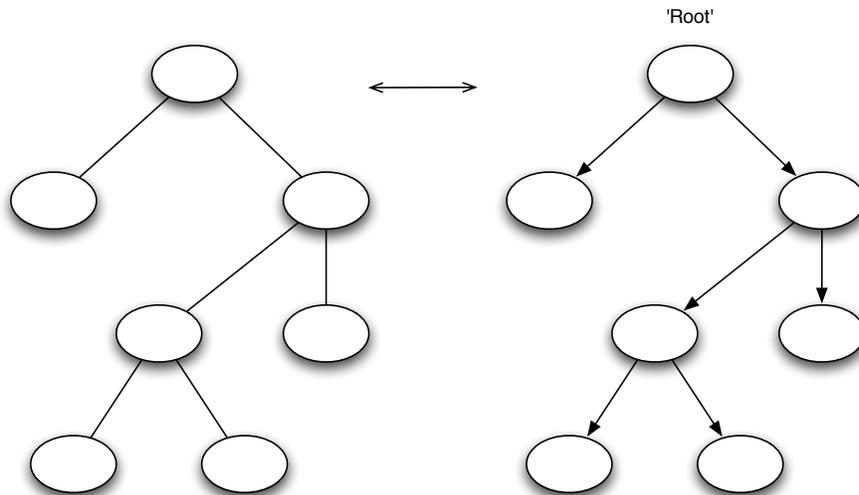


Figure 3: Converting from undirected to a directed tree.

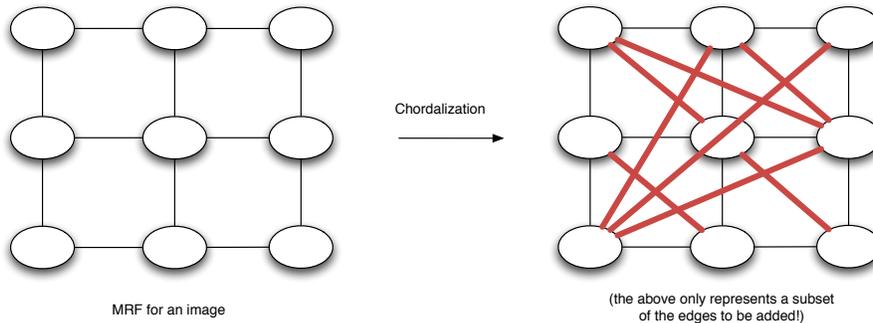


Figure 4: Chords in the MRF representation of an image.

node as a “root” and start “directing” edges away from the root along all directions in the tree to obtain the associated directed graphical model, as shown.

As we shall see, inference is especially easy on trees.

## 5.4 Discussion

As we shall see, chordal graphical models are good representation and allow for efficient inference (with respect to the size of the graphical structure, which could be exponentially large in number of variables). Figure 4 represents an example of MRF that we have seen earlier for representing image. This graphical model (two dimensional grid graph) seems nice. But to make it chordal, we have to add *lots* of edges and thus making it lot more complicated.

Figure 5 shows a useful visual classification of distribution in terms of perfect maps, chordal graphical models, trees and Markov chains.

As a final comment, Markov chains also have other associated graphical representations with which you may already be familiar: state transition diagrams and trellis diagrams. While these representations continue to be useful, it is important not confuse these with our new graphical model representations of Markov chains: the nodes and edges mean very different things in these different representations!

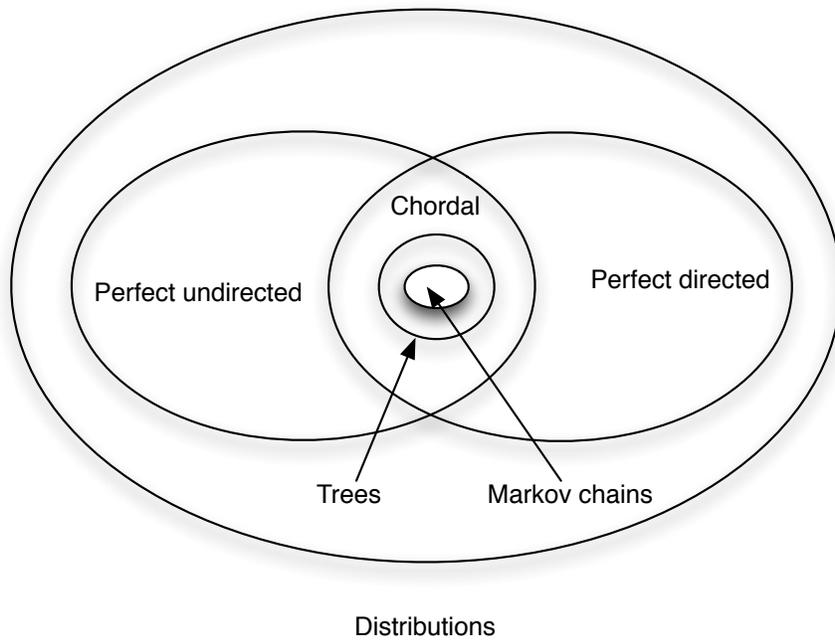


Figure 5: Venn diagram of graphical models.

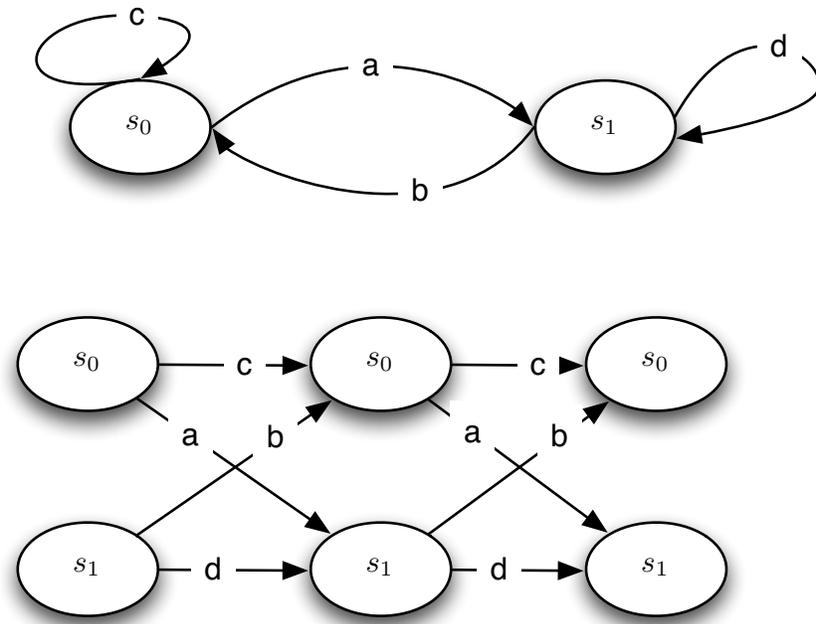


Figure 6: State Transition diagram and Trellis diagram for a Markov chain.

MIT OpenCourseWare  
<http://ocw.mit.edu>

6.438 Algorithms for Inference  
Fall 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.