Directed Graphical Models

Graphical models provide a visual representation of the underlying structure of a joint probablity distribution. As we will see below, the structure encodes information about the conditional independence relationships among the random variables. Recall that such independence relationships are important for understanding the computational costs associated with representation and inference for a given joint probability distribution. Our first goal will be to use the model to answer simple questions such as, "is the random variable X_A independent of the random variable X_B conditioned on the random variable X_C?" or, "are the random variables X_A and X_B independent?" Although these questions seem simple, our only way to answer these questions, thus far, would be to apply Bayes rule. Instead, graphical models will allow us to reduce these types of questions to simple queries over directed or undirected graphs.

Bayesian Networks

A Bayesian network (BN) is a directed graphical model that captures a subset of the independence relationships of a given joint probability distribution. Each BN is represented as a directed acyclic graph (DAG), $G = (V, D)$, together with a collection of conditional

is no directed cycle (i.e., a series of directed edges starting at a vertex $v \in V$ such that if the edges are traversed in the direction of the arrows, you will eventually return to the starting vertex). In a BN, each node in the directed graph corresponds to a random variable and each directed edge represents a statistical dependence. In addition, each node is associated with a conditional probability distribution of the corresponding random variables given its parents in the DAG. Node $j \in G$ is a parent of a node $i \in G$ if there is a directed edge from j to i in the graph G. We say that a joint probability distribution **factorizes** with respect to the directed graph G if,

$$p(x_1, \ldots, x_n) = \prod_{i \in V} p(x_i | x_{\text{parents}(i)}).$$
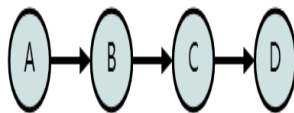
Figure 1. A directed graph G with four vertices $A, B, C,$ and D. If $p(x_A, x_B, x_C, x_D)$ factorizes with respect to G, then we must have

$$p(x_A, x_B, x_C, x_D) = p(x_A)p(x_B|x_A)p(x_C|x_B)p(x_D|x_C)$$

A given joint distribution may factorize over many different DAGs. For example, every joint distribution on n variables factorizes over a directed graph with n nodes such that

example, one could number the vertices from 1 to n and place a directed edge from each vertex to all of the vertices in the graph that received a higher number).
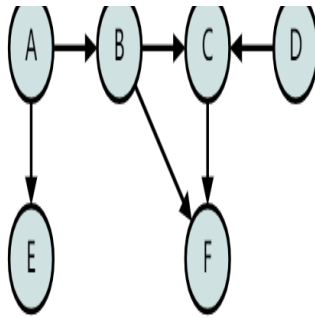
Local Markov Independence Assumptions

Every Bayesian network demands a particular factorization of a joint probability distribution. This factorizaiton, in turn, implies certain independence assumptions about the underlying model. As a simple example, recall that if X_A and X_B are independent random variables, then $p(X_A, X_B) = p(X_A)p(X_B)$. This probability distribution factorizes over the directed graph with two nodes and no edges depicted in Figure 2 below.
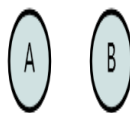
Figure 2. A directed graph G with two vertices A and B and no edges representing two independent random variables.

More generally, BNs encode a collection of local Markov independence assumptions: if a joint distribution factorizes with respect to a directed graph G, then each variable in the joint distribution (which corresponds to a node in the graph) is independent of its non-descendants in G given its parents in G. Consider the following directed graph.

Figure 3. A directed graph G with six vertices $A, B, C, D, E,$ and F.

What local Markov independence assumptions are implied by the directed graph in Figure 3?

Show that if a joint probability distribution factorizes over the DAG in Figure 3, then $X_A \perp X_D$.

Show that if a joint probability distribution factorizes over the DAG in Figure 3, then $X_B \perp X_E | X_A$.

The local Markov independence assumptions are only a subset of the conditional independence relations that can be inferred from the factorized form.

Show that if a joint probability distribution factorizes over the DAG in Figure 3, then $X_A \perp X_C | X_B$. Note that this is not implied by any single local Markov indepenence assumption.

Having to produce an argument of this type every time that we want to check a conditional independence

is a close connection between the structure of the graph and the independence relationships that can be inferred from the factorization.

D-separation

All of the independence relationships implied by the factorization can be found using only the DAG. The crux of this argument is based on the notion of d-separation: a graph theoretic concept that states that a subset $S_1 \subseteq V$ is d-separated from a subset $S_2 \subseteq V$ of vertices given a third set $S_3 \subseteq V$ whenever all paths from vertices in S_1 to vertices in S_2 are blocked given the vertices in S_3. For purposes of this section, we will consider all paths ignoring the direction of the edges. As depicted in Figure 4, there are three different types of segments that can be encountered on a path from a vertex in S_1 to a vertex in S_2. Each of these segments is considered blocked depending on whether or not the second node in the segment is in the separating set, S_3.

(a) Sequential Path   (b) Convergent Path   (c) Divergent Path
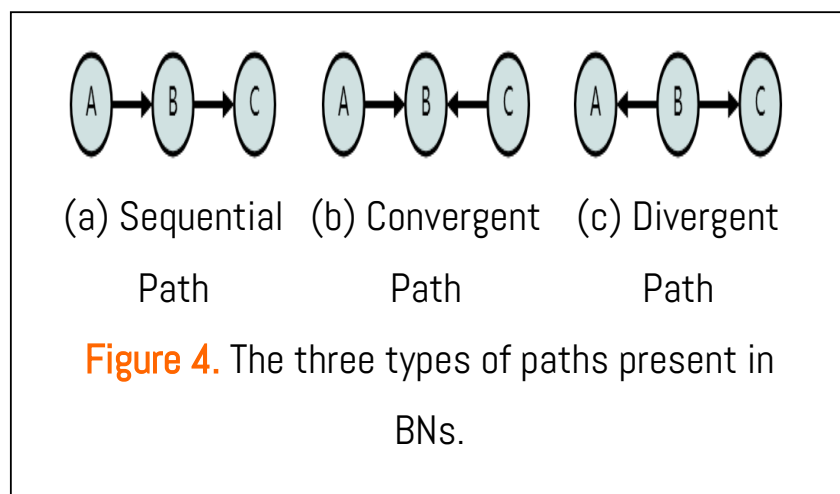
**Figure 4.** The three types of paths present in BNs.

Figure 4 (a), is blocked whenever vertex B is in the separating set.

- **Convergent path:** The convergent path, depicted in Figure 4 (b), is blocked when neither the vertex B nor any of its descendants (i.e., nodes that can be reached by following directed arrows starting from B) are in the separating set.

- **Divergent path:** The divergent path, depicted in Figure 4 (c), is blocked whenever vertex B is in the separating set.

A path between two nodes is blocked if any segment along the path is blocked. There are two special cases. First, if a node in S_1 is adjacent to a node in S_2, we say that the path is not blocked. Second, if there is no path bewteen two nodes, then we say that all of the paths are blocked. For disjoint sets $S_1, S_2, S_3 \subseteq V$, if every path, chosen ignoring the direction of the arrows, from a node $A \in S_1 \subseteq V$ to a node $B \in S_2 \subseteq V$ is blocked given the set S_3 we say that S_1 is **d-separated** from S_2 given S_3. Consider again the directed graph from Figure 3.
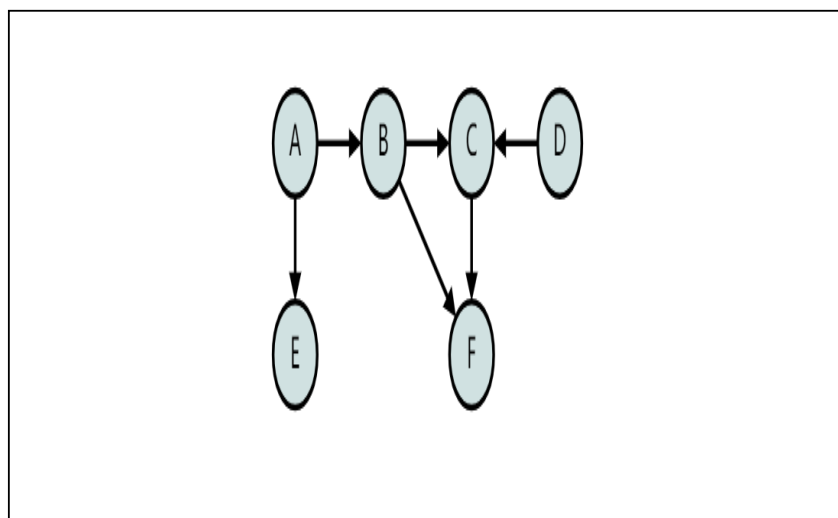
$$A, B, C, D, E, \text{and } F.$$

> Show that $\{A\}$ is d-separated from $\{F\}$ given $\{B\}$ for the directed graph in Figure 5.

> In the graph in Figure 5, is $\{B\}$ d-separated from $\{D\}$ given $\{F\}$?

> In the graph in Figure 5, is $\{B\}$ d-separated from $\{D\}$ given $\{C\}$?

For joint distributions that factorize over a DAG, d-separation can be directly converted into a conditional independence relation.

> **Theorem 1:** If S_1 is d-separated from S_2 given S_3 in a directed graph G, then for any probability distribution that factorizes over G, $X_{S_1} \perp X_{S_2} | X_{S_3}$.

That is, if a propability distribution p factorizes over a directed graph G, then d-separation implies a certain set of independence relationships among the random variables in p. Note that not necessarily every independence relationship in p can be deduced from d-separation. As a simple example, if p is the joint probability distribution over n mutually independent random variables, then p factorizes over any n-vertex DAG. However, if the DAG has even one edge, say between nodes A and B, then at

cannot be inferred from the graph. A formal proof of this theorm will be presented in the next chapter.

I-maps and Equivalent Models

Theorem 1 implies that if joint distribution p factorizes over G, then p must satisfy all of the conditional independence relationships implied by the graph. This turns out to be both a necessary and sufficient condition for strictly positive probability distributions to factorize over a DAG. Let $I(G)$ denote the set of conditional independence relations that can be obtained from d-separation on a DAG, G, and let $I(p)$ denote the set of conditional independence relations that hold in the distribution p. We say that G is an I-map for p if $I(G) \subseteq I(p)$.

> **Theorem:** A strictly positive probability distribution p factorizes over a directed graph G if and only if $I(G) \subseteq I(p)$ (i.e., G is an I-map for p).

An I-map is said to be **perfect** if $I(G) = I(p)$. Given a distribution p, it is not always possible to find a DAG G such that $I(G) = I(p)$. Consider a joint distribution over four random variables such that $X_A \perp X_C | X_B, X_D$ and $X_B \perp X_D | X_A, X_C$ are the only conditional independence relations. One can verify that there is no four vertex DAG that implies only these conditional independence assumptions.

conditional independence relations. As an example, consider the sequential path and segement and the divergent path segment from Figure 4. As DAGs, both of these segments imply the same set of conditional independence relations.

◄   ►