Graphical Models and Bayesian Networks

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Usefulness of Graphical Models

- Reasons for probabilistic graphical models
  - Visualizations
  - Make model properties (e.g., conditional independence) more obvious
  - Decompose complex computations and inference
  - Shows how to improve computational complexity
- In a probabilistic graph, the nodes are random variables and the edges show relationships between variables (i.e., conditional dependence)
- The graph shows the decompositions of the joint distribution into a product of factors depending on only a subset of the variables

Types of Probabilistic Graphs

- An edge is a pair of nodes or a set of two nodes
- If the set is ordered then this is a *directed graph* - Bayesian network
  - Generative model show causal relationships between random variables
- If the set is unordered then this is an *undirected graph* - Markov random field or Markov network
  - Show soft constraints between random variables
- Factor graphs use both directed and undirected graphs for inference problems

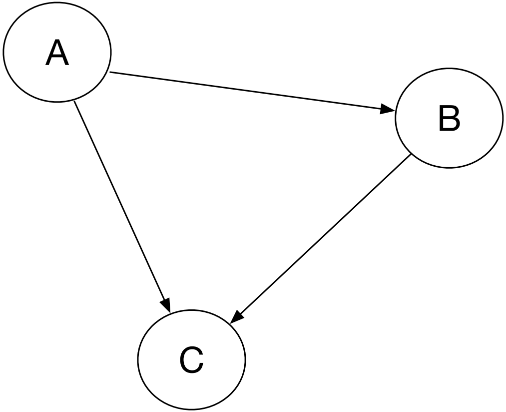
Directed Graphical Models

- Let $p(A, B, C)$ be the joint distribution for three random variables, A, B, C
- By the product rule:

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

- Let each RV be a node
- Let each conditional distribution have a directed arc
- So, $p(C|B, A)$ has two arcs
- Call A the *parent* of B and C, and B the parent of C
- Call B and C the *children* of A, and C the child of B

Example Joint Distribution



Comment on Ordering

- The variables in the joint distribution, $p(A, B, C)$ are not ordered
- The variables in the chosen conditional distributions, $p(C|A, B); p(B|A)$ are ordered
- A different choice for the decomposition would have produced a different ordering and, hence, a different graph
- In this case, we chose the ordering we wanted and in general, we choose based on the problem

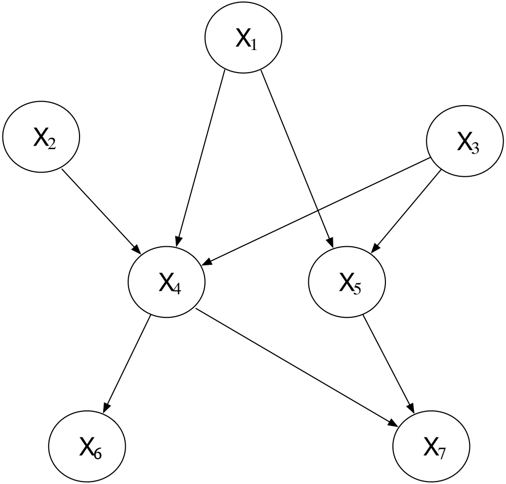
More General Distribution

- Suppose K random variables, $X_1, \ldots, X_K$
- Joint probability decomposition

$$p(X_1, \ldots, X_K) = p(X_K|X_1, \ldots, X_{K-1}) \cdots p(X_2|X_1)p(X_1)$$

- For some K, we can represent this joint distribution as a directed graph with K nodes
- There are arcs between each lower indexed node and every higher indexed node
- Called *fully connected*
- Example: draw a fully connected, lexicographically ordered graph with the nodes $\{A, B, C, D\}$
- What is the equation for the joint distribution in terms of conditional distributions given by the following graph ?

Example Restricted Joint Distribution



Absence of Arcs

- The presence of an arc shows a conditional distribution or relationship
- So, in the previous graph, X_5 is conditioned on only X_1 and X_3
- The joint distribution is

$$p(X_1, \ldots, X_7) =$$

$$p(X_1)p(X_2)p(X_3)p(X_4|X_1,X_2,X_3)p(X_5|X_1,X_3)p(X_6|X_4)p(X_7|X_4,X_5)$$

General Graphs from Joint Distributions

- The joint distribution defined by a graph is shown by the product over all nodes of the graph where the elements of the product are the distributions of the node conditioned on its parent(s) or the marginal distribution if there are no parents for the node
- So, for K nodes, the joint distribution of $\mathbf{X} = (X_1, \ldots, X_K)$

$$p(\mathbf{X}) = \prod_{k=1}^{K} p(X_k | \mathsf{pa}_k)$$

where pa_k are the parents of node k

- Nodes can also be sets of variables or vectors of variables

Directed Acyclic Graphs (DAGs)

- A condition on this factorization and Bayesian networks is that there are no directed cycles
- So, we cannot start at a node and return to that node through a set of directed arcs
- This means there exists an ordering of variables or nodes, such that there are no arcs that go from a higher indexed node to a lower indexed node