CS 6347

Lecture 3

More Bayesian Networks

Recap

- Last time:
  - Complexity challenges
    - Representing distributions
    - Computing probabilities/doing inference
  - Introduction to Bayesian networks

- Today:
  - D-separation, I-maps, limits of Bayesian networks

Bayesian Networks

- A **Bayesian network** is a directed graphical model that represents independence relationships of a given probability distribution
  - Directed acyclic graph (DAG), $G = (V, E)$
    - Edges are still pairs of vertices, but the edges $(1,2)$ and $(2,1)$ are now distinct in this model
  - One node for each random variable
  - One conditional probability distribution per node
  - Directed edge represents a direct statistical dependence

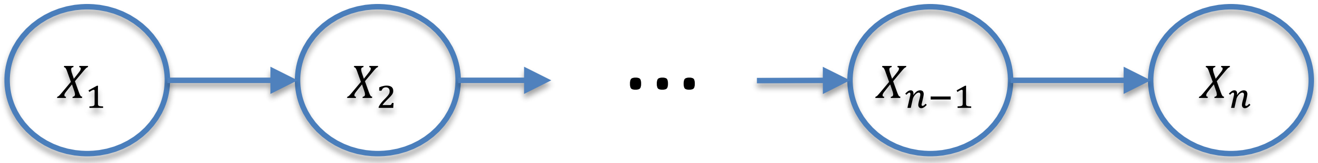
Bayesian Networks

- A **Bayesian network** is a directed graphical model that represents independence relationships of a given probability distribution
  - Encodes **local Markov** independence assumptions that each node is independent of its non-descendants given its parents
  - Corresponds to a **factorization** of the joint distribution

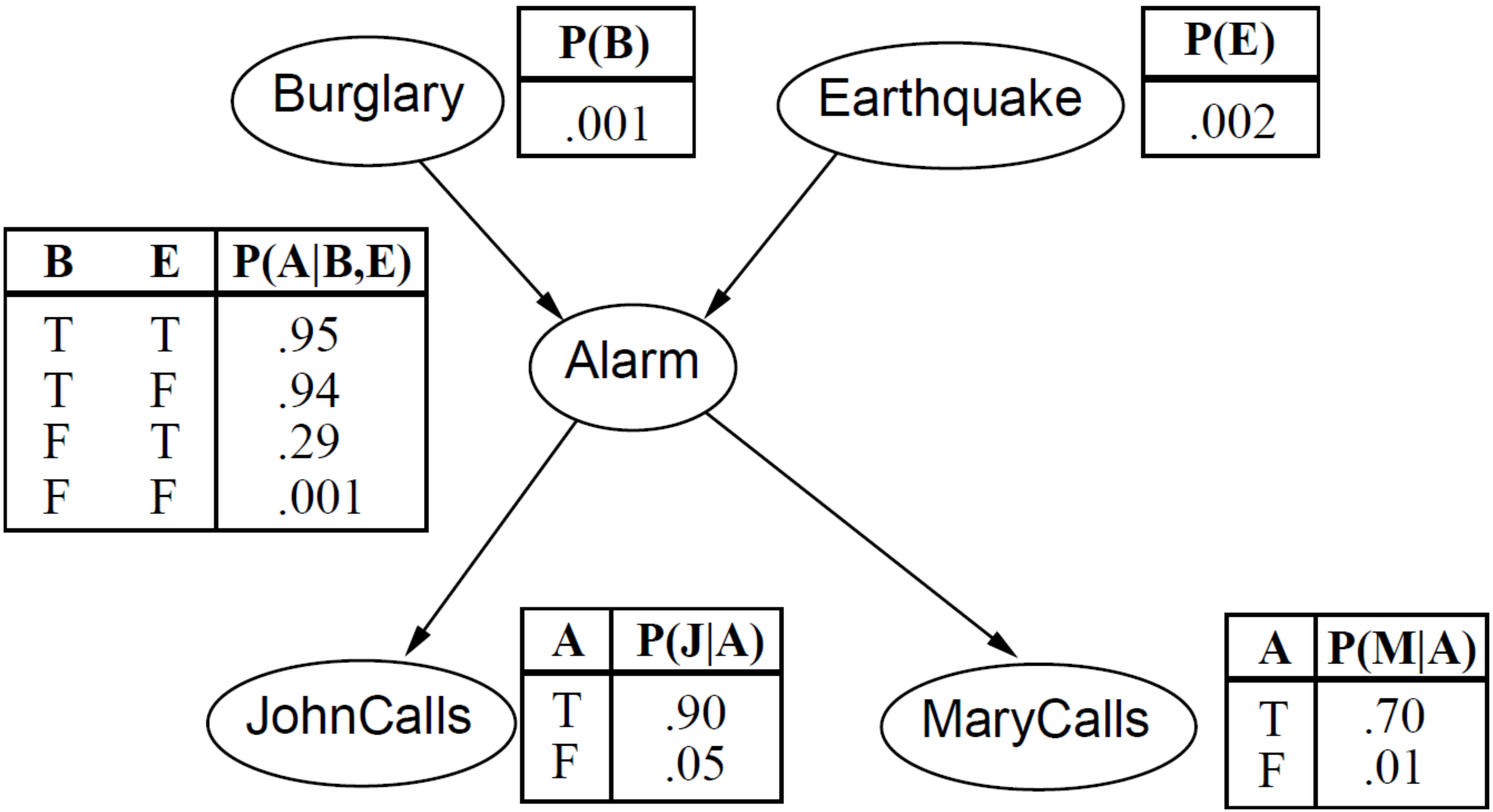
$$p(x_1, \ldots, x_n) = \prod_i p(x_i | x_{parents(i)})$$

Directed Chain

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_2) \ldots p(x_n|x_{n-1})$$



An Example



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

from Artificial Intelligence: A Modern Approach

D-separation

- Independence relationships can be figured out by looking at the graph structure!
  - Easier than looking at the tables and plugging into the definition
- We look at all of the paths from X to Y in the graph and determine whether or not they are **blocked**
  - $X \subset V$ is d-separated from $Y \subset V$ given $Z \subset V$ iff every path from X to Y in the graph is blocked by Z

D-separation

- Three types of situations can occur along any given path

(1) Sequential

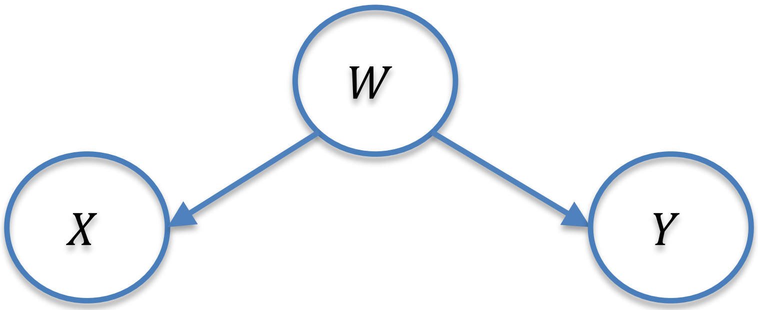
The path from X to Y is blocked if we condition on W

Intuitively, if we condition on W, then information about X does not affect Y and vice versa

D-separation

- Three types of situations can occur along any given path

(2) Divergent



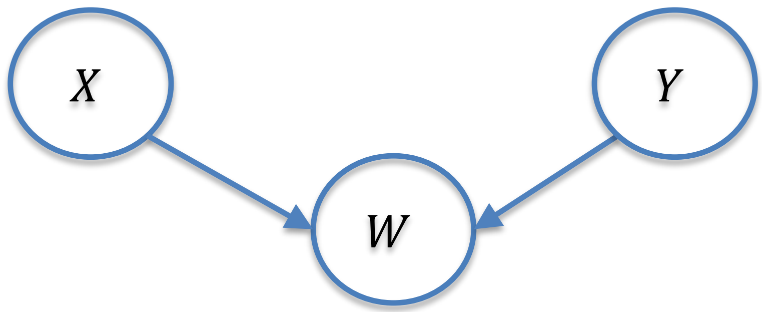
The path from X to Y is blocked if we condition on W

If we don't condition on W, then information about W could affect the probability of observing either X or Y

D-separation

- Three types of situations can occur along any given path

(3) Convergent



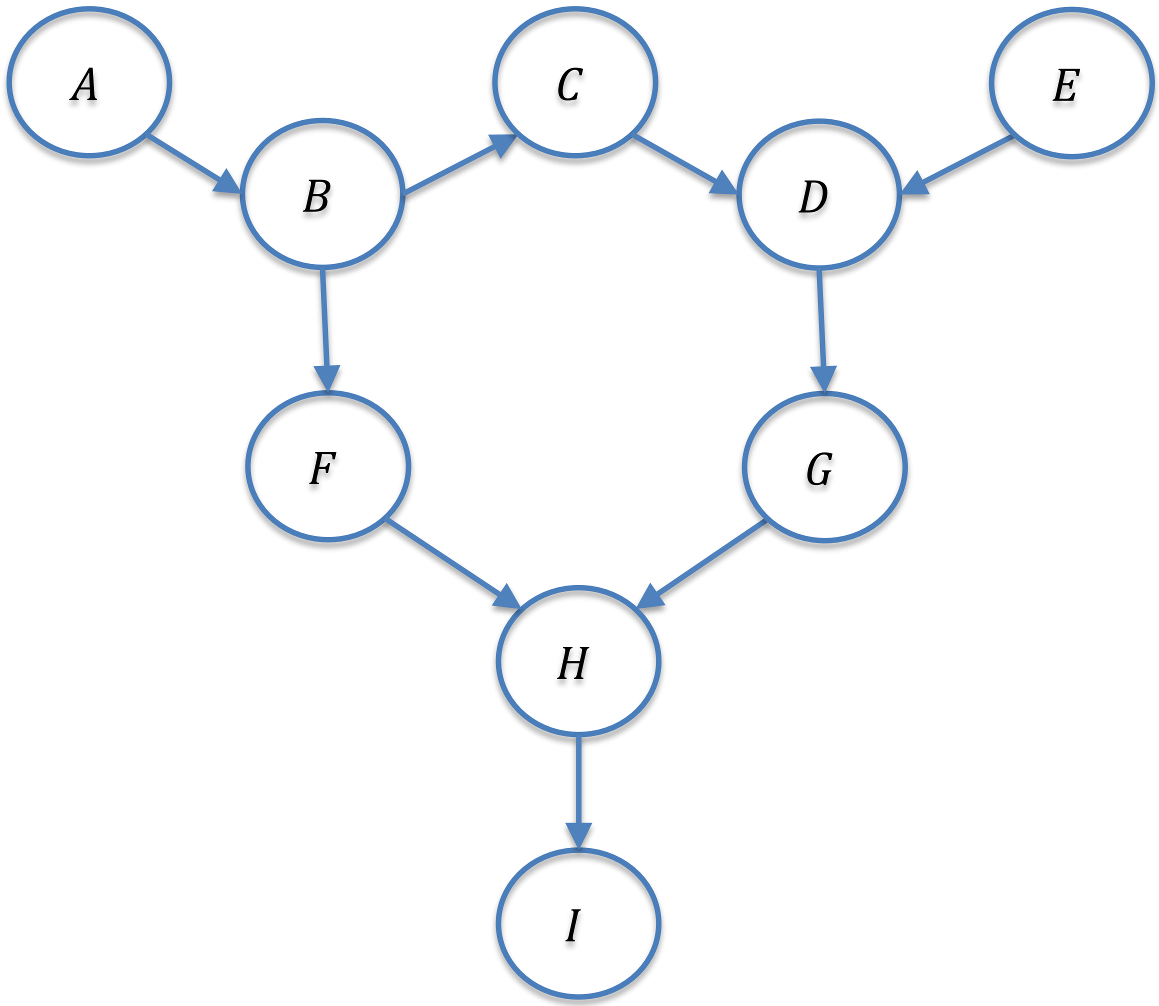
The path from X to Y is blocked if we **do not** condition on W or any of its descendants

Conditioning on W couples the variables X and Y: knowing whether or not X occurs impacts the probability that Y occurs

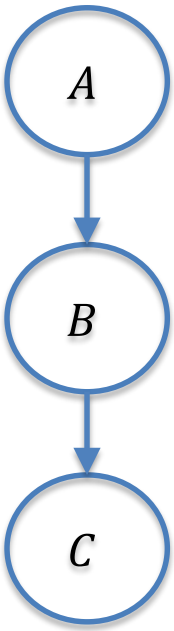
D-separation

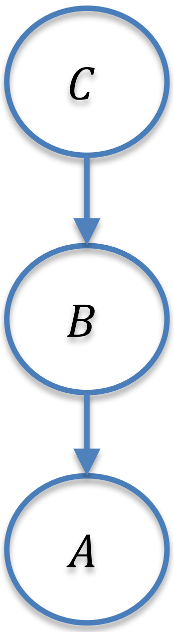
- If the joint probability distribution factorizes with respect to the DAG $G = (V, E)$, then X is d-separated from Y given Z implies $X \perp Y \mid Z$
  - We can use this to quickly check independence assertions by using the graph
  - In general, these are only a subset of all independence relationships that are actually present in the joint distribution
  - If X and Y are not d-separated in G given Z, then there is some distribution that factorizes over G in which X and Y dependent

D-separation Example



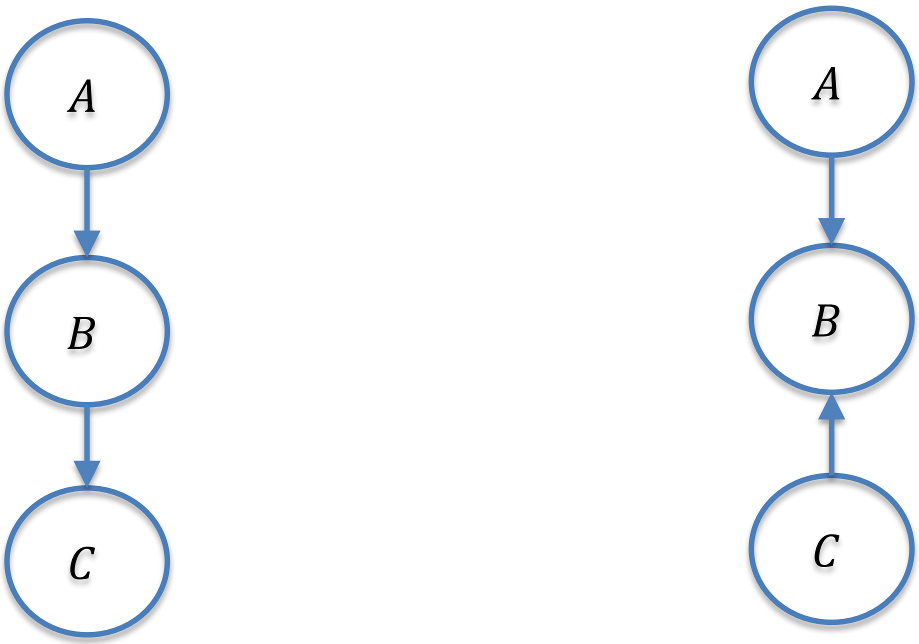
Equivalent Models?





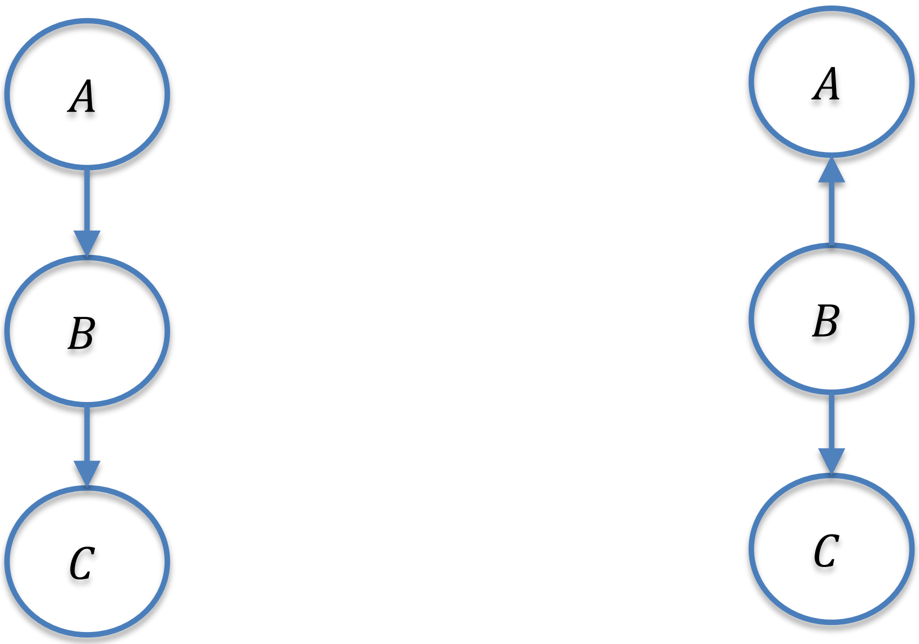
Do these models represent the same independence relations?

Equivalent Models?



Do these models represent the same independence relations?

Equivalent Models?



Do these models represent the same independence relations?

D-separation

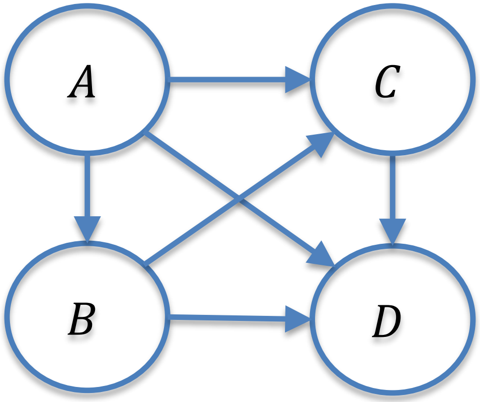
- Let $I(p)$ be the set of all independence relationships in the joint distribution p and $I(G)$ be the set of all independence relationships implied by the graph G

- We say that G is an I-map for $I(p)$ if $I(G) \subseteq I(p)$

- Theorem: the joint probability distribution, p, factorizes with respect to the DAG $G = (V, E)$ iff G is an I-map for $I(p)$

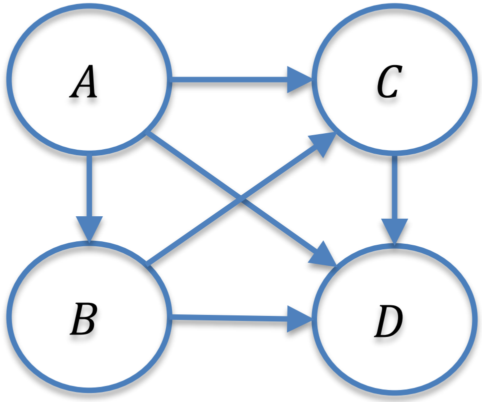
- An I-map is perfect if $I(G) = I(p)$
  - Not always possible to perfectly represent all of the independence relations with a graph

I-Maps



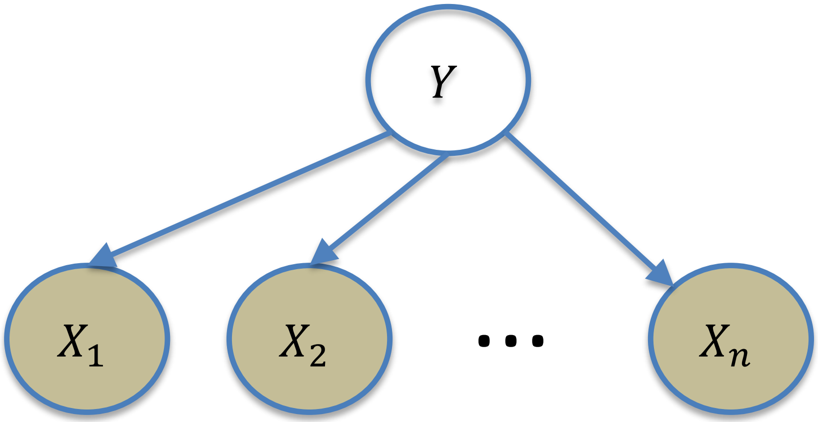
What independence relations does this model imply?

I-Maps



$I(G) = \emptyset$, this is an I-map for any joint distribution on four variables!

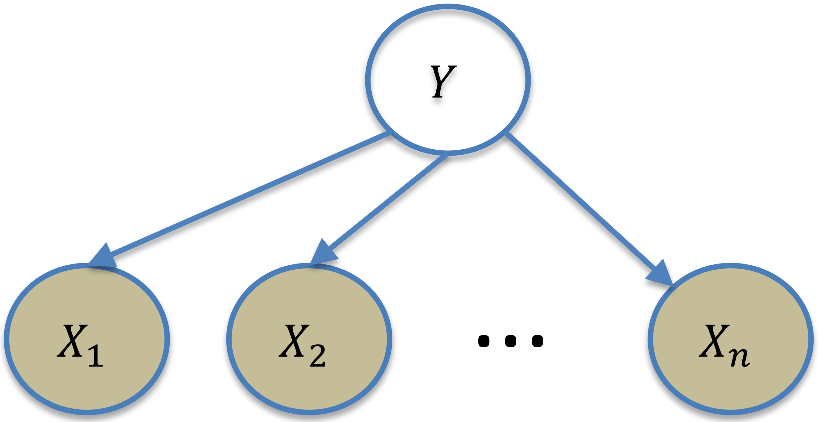
Naïve Bayes



$$p(y, x_1, \dots, x_n) = p(y)p(x_1|y) \dots p(x_n|y)$$

- In practice, we often have variables that we observe directly and those that can only be observed indirectly

Naïve Bayes



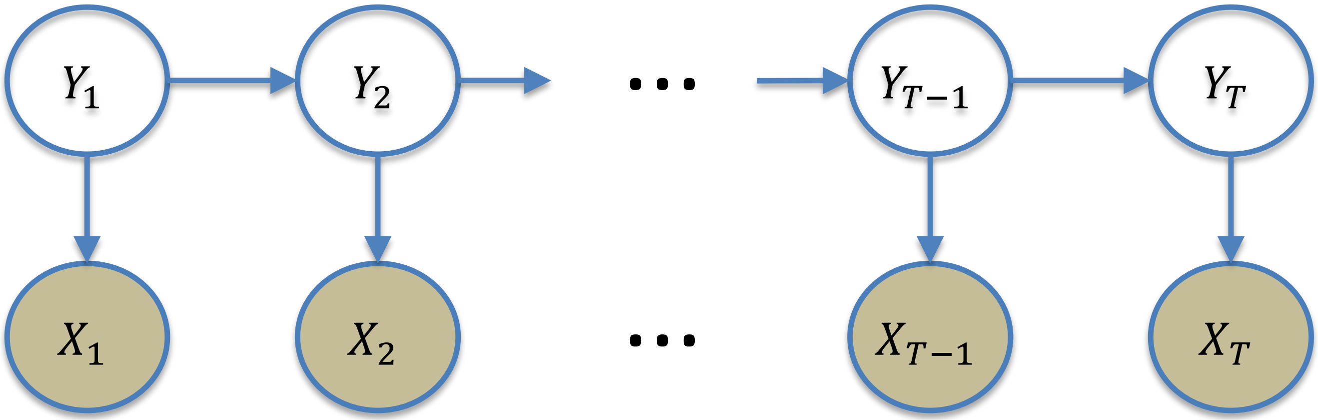
$$p(y, x_1, \dots, x_n) = p(y)p(x_1|y) \dots p(x_n|y)$$

- This model assumes that $X_1, \dots, X_n$ are independent given Y, sometimes called naïve Bayes

Example: Naïve Bayes

- Let Y be a binary random variable indicating whether or not an email is a piece of spam
- For each word in the dictionary, create a binary random variable X_i indicating whether or not word i appears in the email
- For simplicity, we will assume that $X_1, \ldots, X_n$ are independent given Y
- How do we compute the probability that an email is spam?

Hidden Markov Models

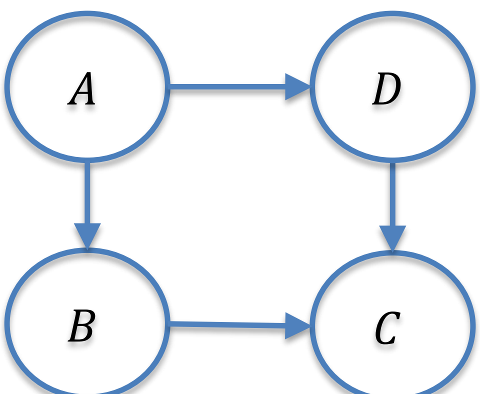


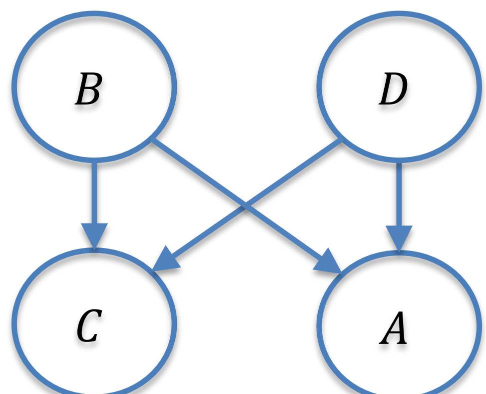
$$p(x_1, \ldots, x_T, y_1, \ldots, y_T) = p(y_1)p(x_1|y_1) \prod_{t=2} p(y_t|y_{t-1})p(x_t|y_t)$$

- Used in coding, speech recognition, etc.
- Independence assertions?

Limits of Bayesian Networks

- Not all sets of independence relations can be captured by a Bayesian network
  - $A \perp C \mid B, D$
  - $B \perp D \mid A, C$
- Possible DAGs that represent these independence relationships?





Markov Random Fields (MRFs)

- A **Markov random field** is an undirected graphical model
  - Undirected graph $G = (V, E)$
  - One node for each random variable
  - Potential function or "factor" associated with cliques, C, of the graph
  - Nonnegative potential functions represent interactions and need not correspond to conditional probabilities (may not even sum to one)

Markov Random Fields (MRFs)

- A **Markov random field** is an undirected graphical model
  - Corresponds to a **factorization** of the joint distribution

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)$$

$$Z = \sum_{x'_1, \dots, x'_n} \prod_{c \in C} \psi_c(x'_c)$$

Markov Random Fields (MRFs)

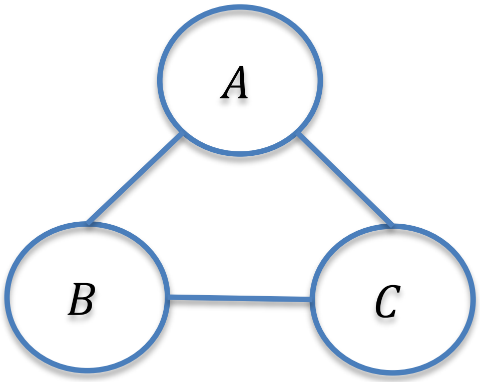
- A **Markov random field** is an undirected graphical model
  - Corresponds to a **factorization** of the joint distribution

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)$$

$$Z = \sum_{x_1', \dots, x_n'} \prod_{c \in C} \psi_c(x_c')$$

Normalizing constant, Z, often called the partition function

An Example



- $p(x_A, x_B, x_C) = \frac{1}{Z}\psi_{AB}(x_A, x_B)\psi_{BC}(x_B, x_C)\,\psi_{AC}(x_A, x_C)$

- **Each potential function can be specified as a table as before**

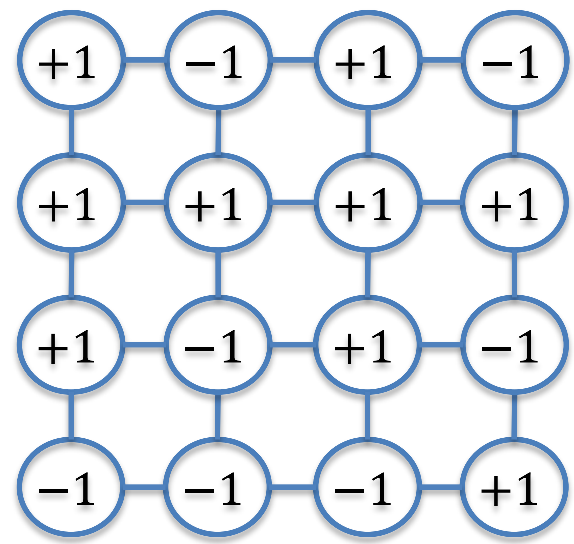
$\psi_{AB}(x_A, x_B) =$

	$x_A = 0$	$x_A = 1$
$x_B = 0$	1	1
$x_B = 1$	1	0

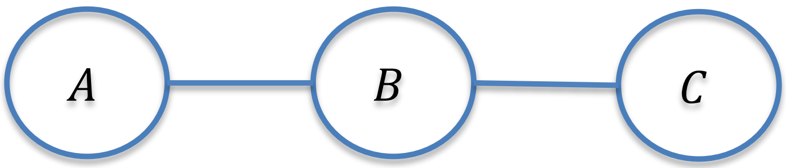
The Ising Model

- Mathematical model of ferromagnets
- Each atom has an associated spin that is biased by both its neighbors in the material and an external magnetic field
  - Spins can be either +1 or -1
  - Edge potentials capture the local interactions
  - Singleton potentials capture the external field

$$p(x_V) = \frac{1}{Z} \exp\left(\sum_{i \in V} h_i x_i + \sum_{(i,j) \in E} J_{ij} x_i x_j \right)$$



Independence Assertions



$$p(x_A, x_B, x_C) = \frac{1}{Z}\psi_{AB}(x_A, x_B)\psi_{BC}(x_B, x_C)$$

- **How does separation imply independence?**
- **Show that** $A \perp C \mid B$