Discussion 5

王欣奕，wangxy6@shanghaitech.edu.cn

Review

- Graphical models
- Bayes Nets
- D-separation
- Sample inference

Graphical models

- **Graphical models express sets of conditional independence assumptions via graph structure.**

- **Graphical models allow combining:**
  - Prior knowledge in form of dependencies/independencies
  - Prior knowledge in form of priors over parameters
  - Observed training data

- **Two types of graphical models:**
  - Directed graphs (aka Bayes Networks)
  - Undirected graphs (aka Markov Random)

Bayes Networks

A Bayes network represents the joint probability distribution over a collection of random variables

A Bayes network is a directed acyclic graph and a set of conditional probability distributions (CPD's)

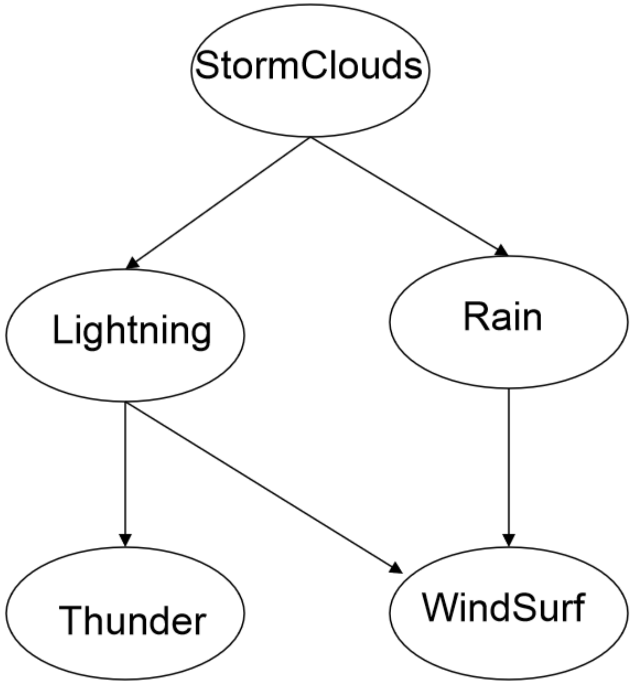
- Each node denotes a random variable
- Edges denote dependencies
- For each node X_i its CPD defines $P(X_i \mid Pa(X_i))$
- The joint distribution over all variables is defined to be

$$P(X_1 \ldots X_n) = \prod_i P(X_i | Pa(X_i))$$

Each node is conditionally independent of its non-descendents, given only its immediate parents.

Bayes Networks

- Number of parameters decreases



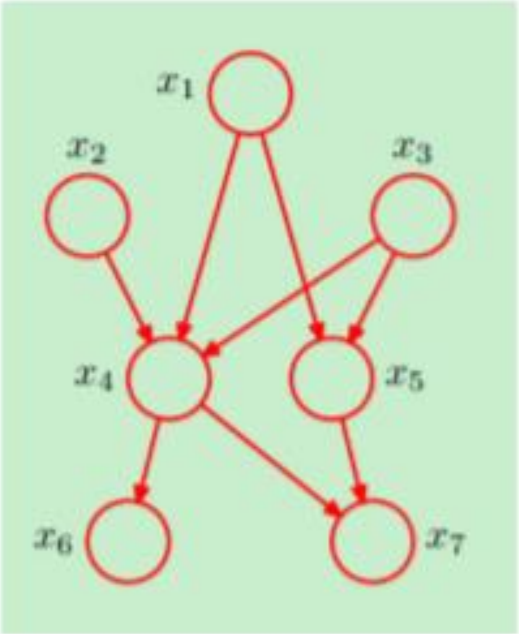
$$P(S, L, R, T, W) = P(S)P(L|S)P(R|S, L)P(T|S, L, R)P(W|S, L, R, T)$$

Number of parameters: 1+2+4+8+16 = 31

$$P(S,L,R,T,W) = P(S)\ P(L|S)\ P(R|S)\ P(T|L)\ P(W|L,R)$$

Number of parameters: 1+2+2+2+4 = 11

Bayes Networks

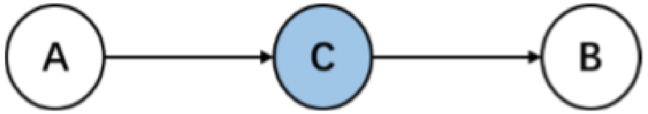


The corresponding decomposition of the joint distribution is given by

$$p(x_1, \cdots, x_7) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

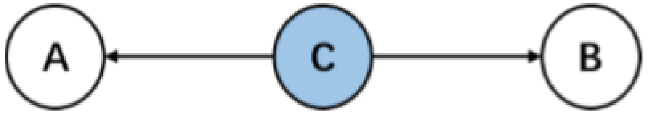
D-separation

Head-to-Tail



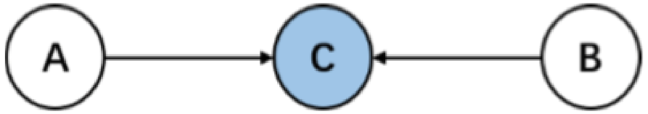
- None of the variables are observed: A is not cond indep of B
- Given C: A is cond indep of B

Tail-to-Tail



- None of the variables are observed: A is not cond indep of B
- Given C: A is cond indep of B

Head-to-Head



- None of the variables are observed: A is cond indep of B
- Given C: A is not cond indep of B

D-separation

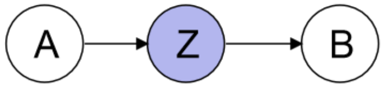
X and Y are conditionally independent given Z,
if and only if X and Y are D-separated by Z.

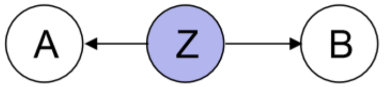
[Bishop, 8.2.2]

Suppose we have three sets of random variables: X, Y and Z

X and Y are **D-separated** by Z (and therefore conditionally indep, given Z) iff every path from every variable in X to every variable in Y is **blocked**

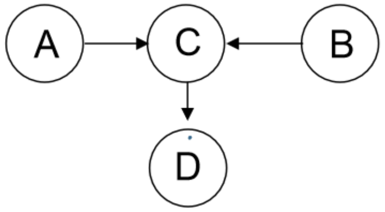
A path from variable X to variable Y is **blocked** if it includes a node in Z such that either



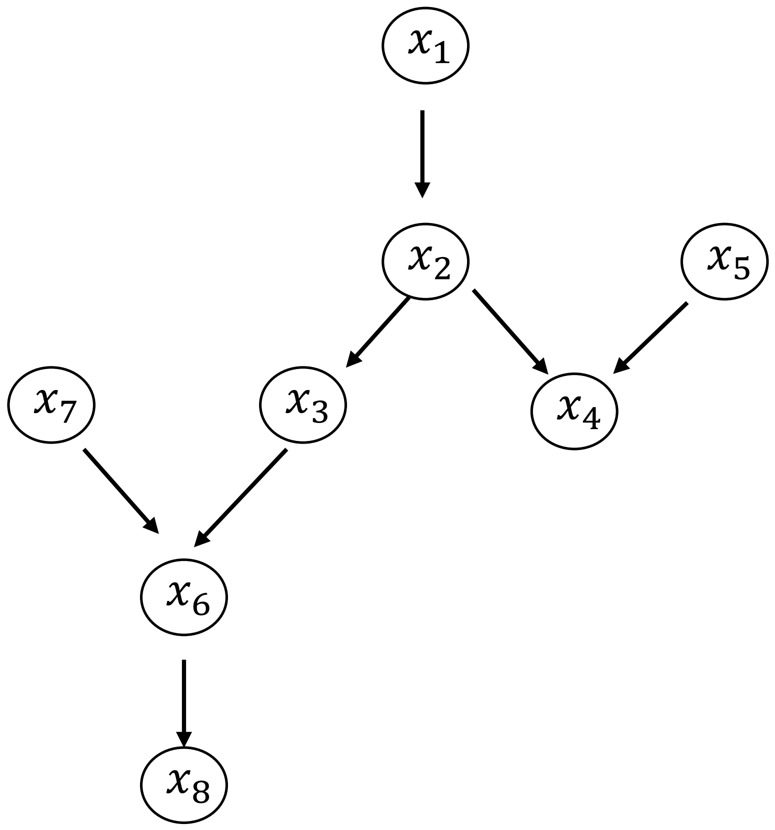


1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z

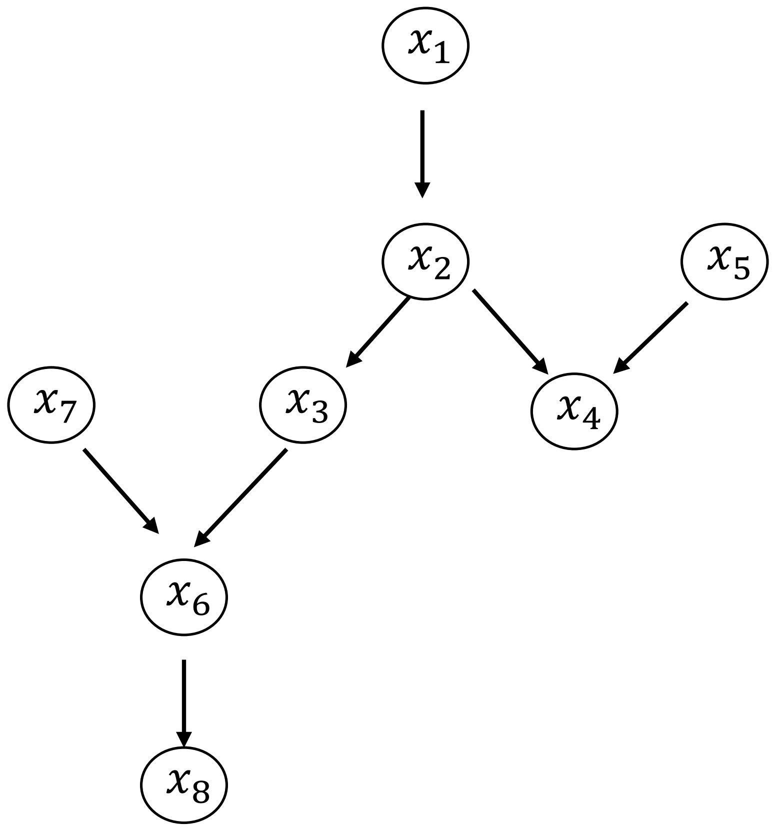


D-separation



$X5 \perp\!\!\!\perp X7 \mid \{X4, X8\}$?

D-separation



X1 ⫫ X7 | {X2, X6} ?

Inference

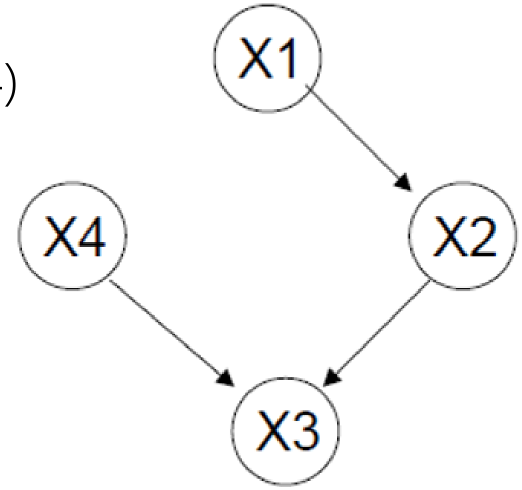
- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Belief propagation
- Sometimes use Monte Carlo methods
  - Generate many samples according to the Bayes Net distribution, then count up the results

Inference

$P(X1,X2,X3,X4) = \ ?$ $\qquad P(X1)P(X2 \mid X1)P(X3 \mid X2,X4)P(X4)$

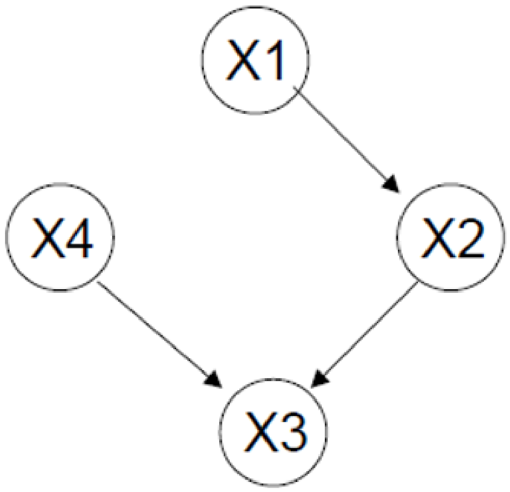
$P(X1 \mid X2,X3,X4) = \ ?$ $\qquad \dfrac{P(X1,X2,X3,X4)}{P(X2,X3,X4)}$

$P(X1) = \ ?$



Inference(Generating a sample from joint distribution: easy)

$P(X1,X2,X3,X4) = ?$



$P(X1 = 1) = \theta$,
draw a value of r uniformly from [0,1].
If $r < \theta$, then output $X1 = 1$, else $X1 = 0$.

The same process for X4, for X2 | X1, for X3 | X2, X4

General methods for any probability term but weak!