Review

For sets of nodes X, Y, E:

$$P(X, Y \mid E) = P(X \mid E)\, P(Y \mid E) \iff \text{all paths from } X \text{ to } Y \text{ are } \underline{\text{d-separated}}$$

d-separation cases:

(I) $z \in E$ with $\to z \to$ an intervening event

(II) $z \in E$ with $\leftarrow z \to$ a common cause

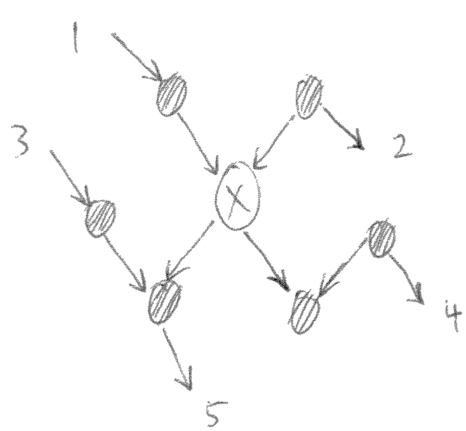
(III) $z \notin E$, $\operatorname{desc}(z) \notin E$ with $\to z \leftarrow$ no observed common effects

<u>Markov blanket</u>: B_X of node X consists of parents, children, and "spouses" of X.

<u>Theorem</u>: $P(X \mid B_X, Y) = P(X \mid B_X)$ where $Y \notin \{B_X, X\}$.

<u>Proof</u>: For any node $Y \notin \{B_X, X\}$, the undirected path from X to Y must pass through B_X.

There are five cases to consider:



(1) from parent of parent of X (I)
(2) from child of parent of X (II)
(3) from parent of spouse of X (I)
(4) from child of spouse of X (II)
(5) from child of child of X (I)

All X-Y paths are d-separated; hence, $P(X \mid B_X, Y) = P(X \mid B_X)$. ∎

Inference

- Problem

  E = set of evidence nodes

  Q = set of query nodes

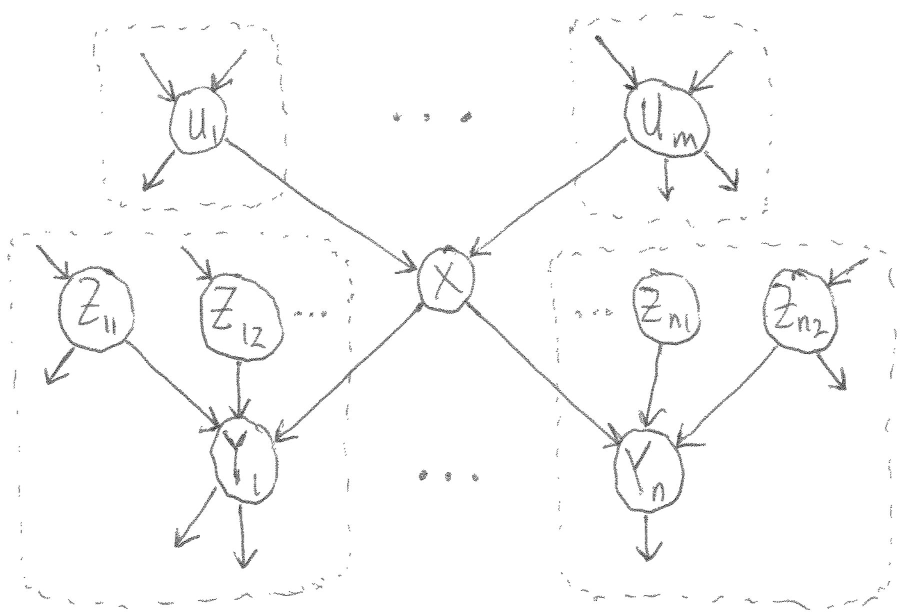
  How to compute posterior probabilities $P(Q|E)$?

- Question: when can we perform inference efficiently?
  (polynomial time in size of DAG and CPTs)

  Answer: polytrees — singly connected network; at most one undirected path between any two nodes; no loops

- Goal: compute $P(X|E)$

  Node X
  Evidence E
  Parents U_i
  Children Y_j
  Spouses Z_{jk}
  (k^{th} parent of j^{th} child)



Boxes don't overlap: no loops in polytrees

Types of evidence:

E_X^+ = upstream evidence "above" X, connected to X thru parents

E_X^- = downstream evidence "below" X, connected to X thru children

$E = E_X^+ \cup E_X^-$ (assume $X \notin E$, otherwise trivial)

Inference in polytree

$$P(X|E) = P(X|E_X^-, E_X^+)$$

$$= \frac{P(X, E_X^- | E_X^+)}{P(E_X^- | E_X^+)} \qquad \text{conditionalized product rule}$$

$$= \frac{P(X, E_X^- | E_X^+)}{\sum_x P(X=x, E_X^- | E_X^+)} \qquad \text{conditional marginalization}$$

Denominator is same computation as numerator, but summed over different values of X.

So focus on numerator:

$$P(X, E_X^- | E_X^+) = P(X|E_X^+) P(E_X^- | X, E_X^+) \qquad \text{cond. product rule}$$

$$= \underbrace{P(X|E_X^+)}_{\text{upstream recursion}} \underbrace{P(E_X^- | X)}_{\text{downstream recursion}} \qquad \text{d-sep I}$$

Goal:

"Upstream" recursion

$$P(X|E_X^+) = \sum_{\vec{u}} P(X, \vec{U} = \vec{u} | E_X^+) \qquad \text{marginalization over parents}$$

$$= \sum_{\vec{u}} P(\vec{U} = \vec{u} | E_X^+) P(X | \vec{U} = \vec{u}, E_X^+) \qquad \text{cond. product rule}$$

$$= \sum_{\vec{u}} P(\vec{U} = \vec{u} | E_X^+) P(X | \vec{U} = \vec{u}) \qquad \text{d-sep. I or II}$$

$$= \sum_{\vec{u}} P(X | \vec{U} = \vec{u}) \prod_{i=1}^{m} P(U_i = u_i | E_X^+) \qquad \text{d-sep III (X is unobserved common effect)}$$

Let $E_{U_i \setminus X}$ denote evidence inside i^{th} parent's box: evidence connected to U_i except via path through X.

$$P(X|E_X^+) = \sum_{\vec{u}} P(X|\vec{U}=\vec{u}) \prod_{i=1}^{m} P(U_i=u_i|E_X^+)$$

$$= \underbrace{\sum_{\vec{u}} P(X|\vec{U}=\vec{u})}_{\text{CPT at node } X} \prod_{i=1}^{m} \underbrace{P(U_i=u_i|E_{U_i \setminus X})}_{\text{recursive instance of original problem}}$$

d-sep III (X is unobserved common effect)

Downstream recursion

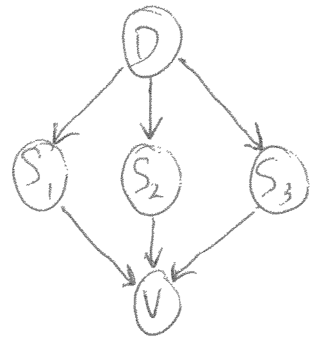
How to compute $P(E_X^-|X)$?

Possible but slightly more complicated

- Termination conditions
  - root node (no parents)
  - leaf node (no children)
  - evidence node (trivial)
- Running time
  - linear in # of nodes
  - linear in size of CPTs (must sum over parents $\sum_{\vec{u}} P(X|\vec{U}=\vec{u})$)

Loopy BNs - how to perform inference?

Medical example:



disease (all binary RVs)

symptoms

visit doctor?

How to do exact inference?

Turn loopy BN into polytree

One approach: node clustering

Merge nodes to form polytree

Merge S_1, S_2, S_3 into mega-node S

Merge CPTs $P(S_1|D)$, $P(S_2|D)$, $P(S_3|D)$ into mega-CPT $P(S|D)$

Apply polytree algorithm

- size of mega-node: 2^3
- size of mega-CPT: 2^4

$$P(S|D) = P(S_1, S_2, S_3 | D) = P(S_1|D)P(S_2|D)P(S_3|D)$$

product rule + conditional independence (d-sep II)

S_1	S_2	S_3	S	$P(S\|D)$
0	0	0	0	$P(S_1=0\|D)P(S_2=0\|D)P(S_3=0\|D)$
0	0	1	1	$P(S_1=0\|D)P(S_2=0\|D)P(S_3=1\|D)$
⋮	⋮	⋮	⋮	
1	1	1	7	$P(S_1=1\|D)P(S_2=1\|D)P(S_3=1\|D)$

- Polytree algorithm linear in size of CPTs, but CPTs grow exponentially with clustering
- How to choose optimal clustering?
  Computationally hard problem