CSCE 970 Lecture 5: More Properties of Bayes Nets

Stephen D. Scott

Introduction

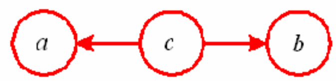
- So far, have introduced Bayes nets and discussed the Markov condition
- As mentioned previously, Markov condition entails conditional independencies among variables
- Does not imply any entailed dependencies
- Throughout lecture, unless otherwise stated, assume that (P, G) satisfies Markov condition

Outline

- Entailed conditional independencies
- Markov equivalence
- Entailing dependencies: faithfulness and embedded faithfulness
- Minimality
- Markov blankets and Markov boundaries

Entailed Conditional Independencies

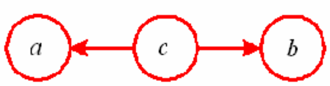
Tail-to-Tail Connections



Are a and b independent? Conditionally independent given c?

Entailed Conditional Independencies

Tail-to-Tail Connections (cont'd)



- Factorization via Theorem 1.4:

$$P(a, b, c) = P(a \mid c)P(b \mid c)P(c)$$

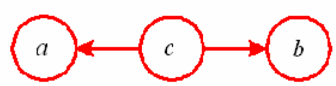
- When c unknown, get $P(a, b)$ by marginalizing:

$$P(a, b) = \sum_c P(a \mid c)P(b \mid c)P(c) \ ,$$

which generally does not equal $P(a)P(b)$

Entailed Conditional Independencies

Tail-to-Tail Connections (cont'd)



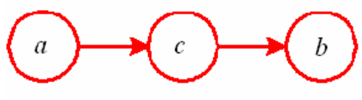
- But when conditioning on c, get:

$$P(a, b \mid c) = \frac{P(a, b, c)}{P(c)} = \frac{P(c)P(a \mid c)P(b \mid c)}{P(c)} = P(a \mid c)P(b \mid c)$$

- Thus a and b conditionally independent given c
- Say that connection between a and b is blocked by c when it is observed and unblocked when unobserved
- Always true for uncoupled tail-to-tail connections $a \leftarrow c \rightarrow b$ (where there's no edge between a and b)

Entailed Conditional Independencies

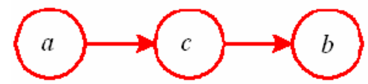
Head-to-Tail Connections



Are a and b independent? Conditionally independent given c?

Entailed Conditional Independencies

Head-to-Tail Connections (cont'd)



- Factorization via Theorem 1.4:

$$P(a,b,c) = P(a)P(c \mid a)P(b \mid c)$$

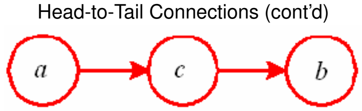
- When c unknown, get $P(a,b)$ by marginalizing:

$$P(a,b) = P(a)\sum_c P(c \mid a)P(b \mid c) = P(a)P(b \mid a) \;\; ,$$

which generally does not equal $P(a)P(b)$

Entailed Conditional Independencies

Head-to-Tail Connections (cont'd)



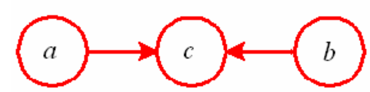
- But when conditioning on c, get:

$$P(a,b \mid c) = \frac{P(a,b,c)}{P(c)} = \frac{P(a)P(c \mid a)P(b \mid c)}{P(c)} = P(a \mid c)P(b \mid c)$$

- Thus a and b conditionally independent given c
- Say that connection between a and b is blocked by c when it is observed and unblocked when unobserved
- Always true for uncoupled head-to-tail connections $a \to c \to b$

Entailed Conditional Independencies

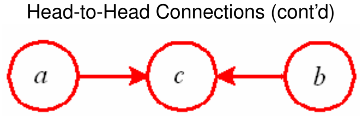
Head-to-Head Connections



Are a and b independent? Conditionally independent given c?

Entailed Conditional Independencies

Head-to-Head Connections (cont'd)



- Factorization via Theorem 1.4:

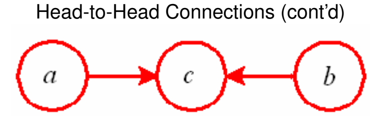
$$P(a,b,c) = P(a)P(b)P(c \mid a,b)$$

- When c unknown, get $P(a,b)$ by marginalizing:

$$P(a,b) = P(a)P(b)\sum_c P(c \mid a,b) = P(a)P(b)$$

Entailed Conditional Independencies

Head-to-Head Connections (cont'd)



- But when conditioning on c, get:

$$P(a,b \mid c) = \frac{P(a,b,c)}{P(c)} = \frac{P(a)P(b)P(c \mid a,b)}{P(c)} \;\; ,$$

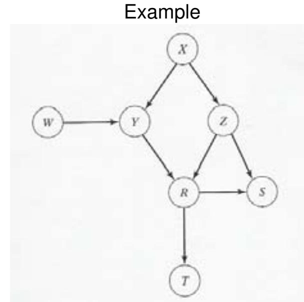
which generally does not equal $P(a \mid c)P(b \mid c)$

- Say that connection between a and b is blocked by c when it is unobserved and unblocked when observed (also unblocks if one of c's descendants is observed)
- Always true for uncoupled head-to-head connections $a \to c \leftarrow b$

D-Separation

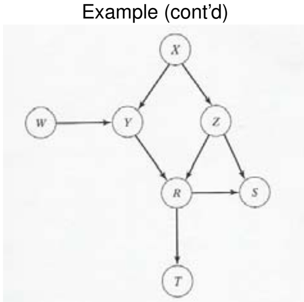
- Let a chain of nodes be a sequence of vertices in the DAG G that are pairwise adjacent, ignoring direction of the edges
  - E.g. on the next slide, $[W, Y, X, Z, S, R]$ is a chain
- Two nodes X and Y from G are d-separated by a set of nodes $\mathcal{A} \subset \mathcal{V}$ if every chain from X to Y is blocked by some node in $\mathcal{A}$
- This generalizes to sets of nodes $\mathcal{X}$ and $\mathcal{Y}$ if every pair of nodes (one from $\mathcal{X}$ and one from $\mathcal{Y}$) is d-separated by a node from $\mathcal{A}$
- Theorem 2.1: Based on the Markov condition, a DAG G entails all and only the conditional independencies that are identified by d-separation in G
  - I.e. if (P, G) satisfies the Markov condition, then if one finds a CI in P implied by G, this CI will also be found via d-separation in G
  - Won't necessarily find all CIs in P, since some CIs may not be captured in G

D-Separation
Example



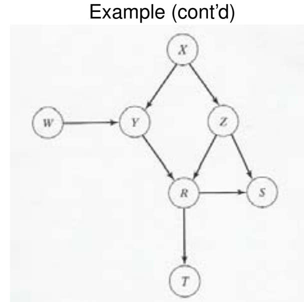
- W and T:
  - Chain $[W, Y, R, T]$ is blocked by Y or R
  - Chain $[W, Y, X, Z, R, T]$ is blocked by X or Z or R
  - Chain $[W, Y, X, Z, S, R, T]$ is blocked by X or Z or R but not by S since observing S unblocks the chain

D-Separation
Example (cont'd)



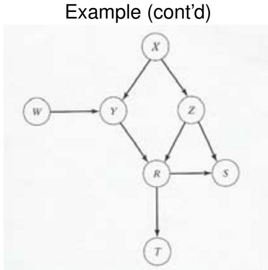
- Y and T:
  - Chain $[Y, R, T]$ is blocked by R
  - Chain $[Y, X, Z, R, T]$ is blocked by X or Z or R
  - Chain $[Y, X, Z, S, R, T]$ is blocked by X or Z or R

D-Separation
Example (cont'd)



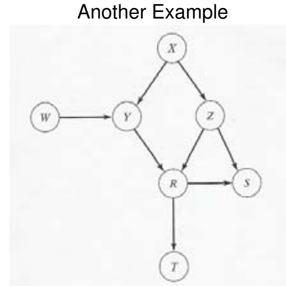
- W and S:
  - Chain $[W, Y, R, S]$ is blocked by Y or R
  - Chain $[W, Y, X, Z, R, S]$ is blocked by X or Z or R
  - Chain $[W, Y, X, Z, S]$ is blocked by X or Z
  - Chain $[W, Y, R, Z, S]$ is blocked by Y or Z

D-Separation
Example (cont'd)



- Y and S:
  - Chain $[Y, R, S]$ is blocked by R
  - Chain $[Y, R, Z, S]$ is blocked by Z
  - Chain $[Y, X, Z, R, S]$ is blocked by X or Z or R
  - Chain $[Y, X, Z, S]$ is blocked by X or Z
- Thus we say that $\{W, Y\}$ and $\{S, T\}$ are conditionally independent given $\{R, Z\}$, i.e. $I_G(\{W, Y\}, \{S, T\} \mid \{R, Z\})$

D-Separation
Another Example



- W and X:
  - Chain $[W, Y, X]$ is blocked by Y when not observed
  - Chain $[W, Y, R, Z, X]$ is blocked by R when not observed
  - Chain $[W, Y, R, S, Z, X]$ is blocked by S when not observed
- Thus we say that W and X are independent, i.e. $I_G(\{W\}, \{X\} \mid \emptyset)$

Finding D-Separations

- Problem: Given a DAG $G = (\mathcal{V}, \mathcal{E})$, and disjoint subsets $\mathcal{A}, \mathcal{B} \subset \mathcal{V}$, find the set of nodes $\mathcal{D}$ that is d-separated from $\mathcal{B}$ by $\mathcal{A}$
  - I.e. find the set of nodes $\mathcal{D}$ that are blocked from those in $\mathcal{B}$ by $\mathcal{A}$
  - I.e. if there is an active path from a node $X \in \mathcal{B}$ to some node $Y \notin \mathcal{A} \cup \mathcal{B}$ (a path from X to Y not blocked by something in $\mathcal{A}$), then Y is NOT in $\mathcal{D}$
- Thus we'll find

  $\mathcal{R} = \{Y : Y \in \mathcal{B}$ or $\exists X \in \mathcal{B}$ that can reach Y with no block from $\mathcal{A}\}$

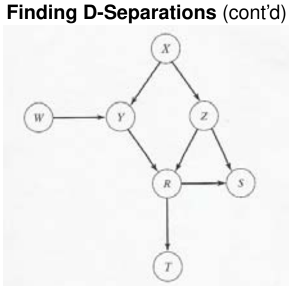
  (the set of reachable nodes) and set $\mathcal{D} = \mathcal{V} \setminus (\mathcal{A} \cup \mathcal{R})$

Finding D-Separations
(cont'd)

- How does node Z block a chain?
  1. By being in a head-to-tail or tail-to-tail arrangement in the chain and being in $\mathcal{A}$

     OR
  2. By being in a head-to-head arrangement in the chain not being in $\mathcal{A}$ and not having a descendent in $\mathcal{A}$
- Since we're initially seeking (sort of) the complement of $\mathcal{D}$, we'll turn the above two conditions on their heads and look for a set of nodes $\mathcal{R}$ that are reachable from $\mathcal{B}$ via active chains
- A chain is active iff each of its 3-node subchains $U - V - W$ satisfies one of
  1. $U - V - W$ is not head-to-head at V and $V \notin \mathcal{A}$
  2. $U - V - W$ is head-to-head at V and $V \in \mathcal{A}$ or a descendent of V is in $\mathcal{A}$

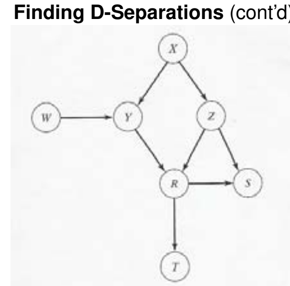
Finding D-Separations (cont'd)



- Let $\mathcal{B} = \{W, Y\}$ and $\mathcal{A} = \{X\}$
  - Then the active chains out of nodes in $\mathcal{B}$ are $[Y, R, T]$, $[Y, R, S]$, $[W, Y, R, T]$, $[W, Y, R, S]$, and $[W, Y, R]$

  $\Rightarrow$ D-separation from $\{Z\}$

Finding D-Separations (cont'd)



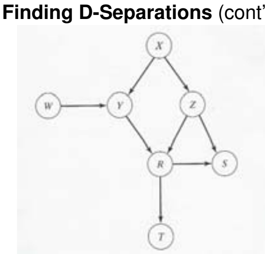
- Let $\mathcal{B} = \{W, Y\}$ and $\mathcal{A} = \{X, T\}$
  - Then the active chains out of nodes in $\mathcal{B}$ are $[Y, R, Z]$, $[Y, R, S]$, $[Y, R, Z, S]$, $[W, Y, R]$, $[W, Y, R, Z]$, $[W, Y, R, S]$, and $[W, Y, R, Z, S]$

  $\Rightarrow$ D-separation from $\emptyset$

Finding D-Separations
(cont'd)

- This problem is a node reachability problem with restrictions to legal pairs of edges
- Define a pair of edges $((U, V), (V, W))$ to be legal iff they satisfy one of the two active chain conditions described earlier
- Then $\mathcal{R}$ is the set of nodes reachable from a node in $\mathcal{B}$ via only legal pairs of edges

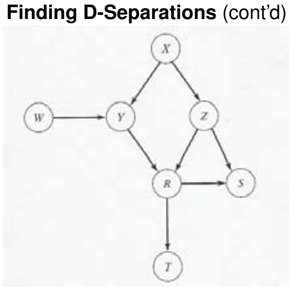
Finding D-Separations (cont'd)



- Let $\mathcal{B} = \{W, Y\}$ and $\mathcal{A} = \{X\}$
  - Then the set of legal pairs of edges is (excluding symmetries)

$$\mathcal{L} = \{((X, Z), (Z, R)), ((X, Z), (Z, S)), ((X, Y), (Y, R)),$$
$$((W, Y), (Y, R)), ((Y, R), (R, T)), ((Y, R), (R, S)),$$
$$((Z, R), (R, T)), ((Z, R), (R, S)), ((R, Z), (Z, S))\}$$

Finding D-Separations (cont'd)



- Let $\mathcal{B} = \{W, Y\}$ and $\mathcal{A} = \{X, T\}$
  - Then the set of legal pairs of edges is (excluding symmetries) the same as before, but add $((Y, R), (R, Z))$ and $((W, Y), (Y, X))$ (why?)

Finding D-Separations
The Algorithm

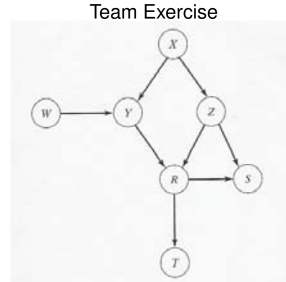
1. Given $G = (\mathcal{V}, \mathcal{E})$, $\mathcal{B}$, and $\mathcal{A}$, compute the set of legal edge pairs $\mathcal{L}$
2. Create $G' = (\mathcal{V}, \mathcal{E}')$, which is G with opposite edges added:
$$\mathcal{E}' = \mathcal{E} \cup \{(X, Y) : (Y, X) \in \mathcal{E}\}$$
   - Because the reachability algorithm respects edges' directions, but d-separation does not
3. Run as a subroutine an algorithm to return $\mathcal{R}$, the set of nodes in G' that are reachable from $\mathcal{B}$ via edge pairs from $\mathcal{L}$
4. The set of nodes that are d-separated from $\mathcal{B}$ by $\mathcal{A}$ is $\mathcal{D} = \mathcal{V} \setminus (\mathcal{A} \cup \mathcal{R})$

Finding D-Separations
Reachability Subroutine

- A breadth-first search of graph G', but over edges rather than nodes

1. Initialize $i = 1$ and
$$\mathcal{R} = \mathcal{B} \cup \{V : V \in \mathcal{V} \text{ and } (X, V) \in \mathcal{E}' \text{ for some } X \in \mathcal{B}\}$$
2. Label each such edge (X, V) with a 1
3. While new nodes added to $\mathcal{R}$
   (a) For each V such that edge (U, V) is labeled i
      i. For each unlabeled edge (V, W) s.t. $((U, V), (V, W)) \in \mathcal{L}$
         A. $\mathcal{R} = \mathcal{R} \cup \{W\}$
         B. Label (V, W) with $i + 1$
   (b) $i{+}{+}$

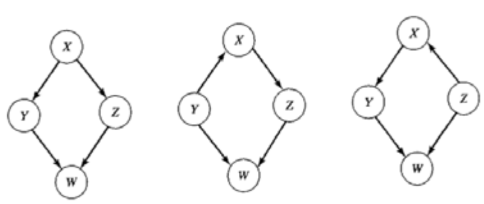
Finding D-Separations
Team Exercise



- Let $\mathcal{B} = \{W, Y\}$ and $\mathcal{A} = \{X\}$
- Everybody join one of four teams (even if you're just sitting in), draw this graph, and simulate the algorithm, including labeling edges

Markov Equivalence

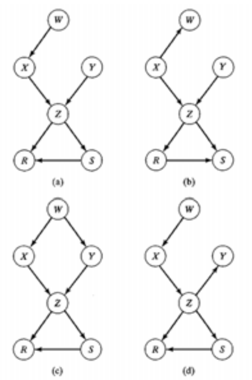
- Many DAGs with the same set of vertices have the same d-separations
- DAGs $G_1 = (\mathcal{V}, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}, \mathcal{E}_2)$ are Markov equivalent if for every three mutually disjoint subsets $\mathcal{A}, \mathcal{B}, \mathcal{C} \subseteq \mathcal{V}$, $\mathcal{A}$ and $\mathcal{B}$ are d-separated by $\mathcal{C}$ in G_1 iff $\mathcal{A}$ and $\mathcal{B}$ are d-separated by $\mathcal{C}$ in G_2
  - I.e. $I_{G_1}(\mathcal{A}, \mathcal{B} \mid \mathcal{C}) \Leftrightarrow I_{G_2}(\mathcal{A}, \mathcal{B} \mid \mathcal{C})$



Markov Equivalence
(cont'd)

Theorem 2.4: DAGs G_1 and G_2 are Markov equivalent iff they have the same links (ignoring edge direction) and the same set of uncoupled head-to-head matchings

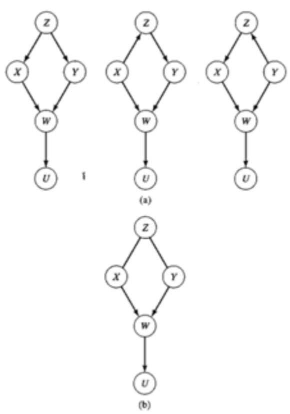
Markov Equivalence
(cont'd)



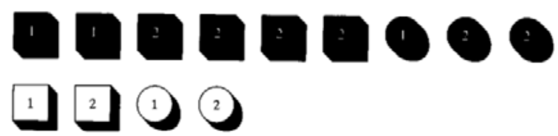
DAG Patterns

- Can represent a set of Markov equivalent DAGs in a single graph
- If an edge can be directed either way and still yield a Markov equivalent DAG, then the edge in the DAG pattern is undirected
- If the edge must be oriented only one way, then the edge in the DAG pattern remains directed

DAG Patterns
(cont'd)



Entailing Dependencies

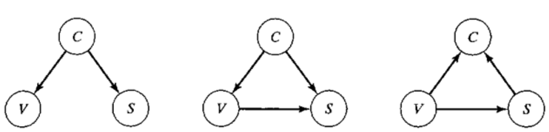


P is uniform

Var	Values	Outcomes
V	$\{v1, v2\}$	obj with "1"/"2"
S	$\{s1, s2\}$	square/round
C	$\{c1, c2\}$	black/white

Entailing Dependencies (cont'd)

We earlier showed that $I_P(\{V\},\{S\} \mid \{C\})$. All of the following three graphs have the Markov property with P.



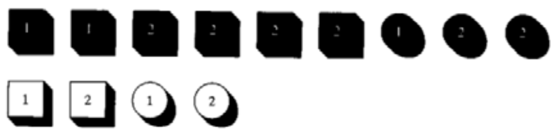
(a) (b) (c)

Graphs (b) and (c) have no independencies, so they satisfy the Markov condition with any distribution P

Entailing Dependencies
Faithfulness

- Given a DAG G and a distribution P, (G, P) satisfies the faithfulness condition if both of these conditions hold
  1. (G, P) satisfies the Markov condition
  2. All conditional independencies in P are entailed by G, based on the Markov condition

Entailing Dependencies
Faithfulness Example



P is uniform

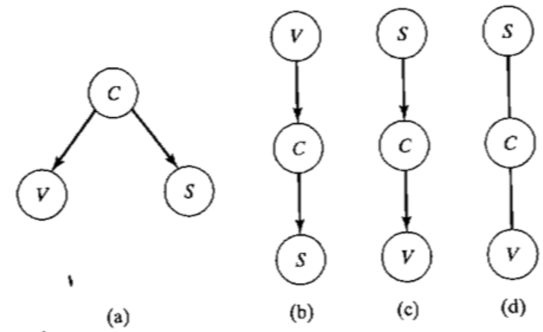
Var	Values	Outcomes
V	$\{v1, v2\}$	obj with "1"/"2"
S	$\{s1, s2\}$	square/round
C	$\{c1, c2\}$	black/white

c	s	v	$P(v)$	$P(s)$	$P(v,s)$
$c1$	$s1$	$v1$	$5/13$	$8/13$	$3/13$
$c1$	$s1$	$v2$	$8/13$	$5/13$	$5/13$
$c1$	$s2$	$v1$	$5/13$	$8/13$	$2/13$
$c1$	$s2$	$v2$	$8/13$	$5/13$	$3/13$
$c2$	$s1$	$v1$	$5/13$	$8/13$	$3/13$
$c2$	$s1$	$v2$	$8/13$	$5/13$	$5/13$
$c2$	$s2$	$v1$	$5/13$	$8/13$	$2/13$
$c2$	$s2$	$v2$	$8/13$	$5/13$	$3/13$

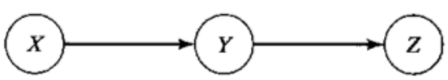
$\Rightarrow \neg I_P(\{V\},\{S\})$. Can show P's only CI is $I_P(\{V\},\{S\} \mid \{C\})$

Entailing Dependencies
Faithfulness Example (cont'd)

These are all faithful to P



Entailing Dependencies
Another Faithfulness Example



$P(x1) = a$
$P(x2) = 1 - a$

$P(y1|x1) = 1 - (b + c)$
$P(y2|x1) = c$
$P(y3|x1) = b$

$P(y1|x2) = 1 - (b + d)$
$P(y2|x2) = d$
$P(y3|x2) = b$

$P(z1|y1) = e$
$P(z2|y1) = 1 - e$

$P(z1|y2) = e$
$P(z2|y2) = 1 - e$

$P(z1|y3) = f$
$P(z2|y3) = 1 - f$

G does not entail unconditional independence of X and Z, but P does
$\Rightarrow$ Markov property holds, but P not faithful to G

Entailing Dependencies
Another Faithfulness Example (cont'd)

Turns out that $P(X,Z) = P(X)P(Z)$. E.g.

$$\begin{aligned}
P(y3) &= \sum_x P(y3 \mid x)P(x) = ba + b(1-a) = b \\
P(y2) &= \sum_x P(y1 \mid x)P(x) = ca + d(1-a) = ca + d - da \\
P(y1) &= (1-(b+c))a + (1-(b+d))(1-a) = 1 - ac + ad - b - d
\end{aligned}$$

$$\begin{aligned}
P(z1) &= e(1 - ac + ad - b - d) + e(ca + d - da) + fb \\
&= e - eb + fb
\end{aligned}$$

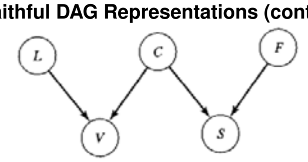
$$\Rightarrow P(x1)P(z1) = a(e - eb + fb)$$

$$\begin{aligned}
P(z1, x1) &= P(z1 \mid x1)P(x1) = P(x1)\sum_y P(z1 \mid y)P(y \mid x1) \\
&= a[e(1-(b+c)) + ec + fb] = a(e - eb + fb)
\end{aligned}$$

Faithful DAG Representations

- Theorem 2.6: If (G, P) satisfies the faithfulness condition, then P satisfies this with all and only those DAGs that are Markov equivalent with G
- The graph pattern representing the class of Markov equivalent DAGs that P is faithful to is called a perfect map of P
- P admits a faithful DAG representation if it is faithful to some DAG
  - Not all distributions admit a faithful DAG representation

Faithful DAG Representations (cont'd)



- Consider a joint distribution $P(v, s, c, \ell, f)$ faithful to the above DAG G
- Only independencies (excluding those with C) are $I_P(\{L\},\{F,S\})$, $I_P(\{L\},\{S\})$, $I_P(\{L\},\{F\})$, $I_P(\{F\},\{L,V\})$, $I_P(\{F\},\{V\})$
- Now consider marginal distribution $P(v, s, \ell, f)$. If the marginal is faithful to a DAG G', then the above independencies imply G''s only d-separations

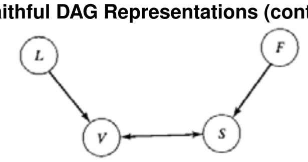
Faithful DAG Representations (cont'd)

- If two nodes cannot be d-separated, then they must be adjacent (Lemma 2.4), so G' has links $L - V$, $V - S$, and $S - F$

- Since $I_{G'}(\{L\}, \{S\})$, the uncoupled meeting $L - V - S$ must be head-to-head

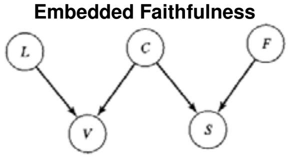
- Also, since $I_{G'}(\{V\}, \{F\})$, the uncoupled meeting $V - S - F$ must be head-to-head

Faithful DAG Representations (cont'd)



Thus G' doesn't exist as a DAG, and the marginal $P(v, s, \ell, f)$ does not admit a faithful DAG representation

Embedded Faithfulness



- So $P(v, s, \ell, f)$ does not admit a faithful DAG representation, but if we allow node C to exist as well, then everything works
- Let P be a distribution over $\mathcal{V} \subseteq \mathcal{W}$ and let $G = (\mathcal{V}, \mathcal{E})$ be a DAG. (G, P) satisfies the embedded faithfulness condition if

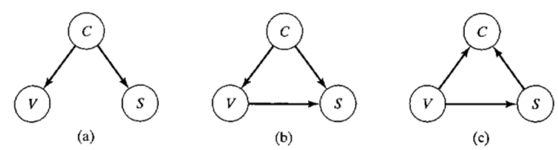
  1. The CIs entailed by G (when restricting to nodes in V) all exist in P

  2. All CIs in P are entailed by G

- P also embedded faithfully in DAG G' that is Markov equivalent to G (and possibly others)

Minimality

Here's that distribution again:  , etc.
The only CI is $I_P(\{V\}, \{S\} \mid C)$, so these have the Markov property:



- If we remove edge (V, S) from (b), it still has the Markov property
- Can we remove any edge from (a) or (c) and still satisfy Markov?
- Given distribution P and DAG $G = (\mathcal{V}, \mathcal{E})$, (G, P) satisfies the minimality condition if (1) (G, P) satisfies the Markov condition and (2) removing any edge from G results in a graph that does not
- Faithfulness $\Rightarrow$ Minimality, but Minimality $\not\Rightarrow$ Faithfulness

Markov Blankets and Boundaries

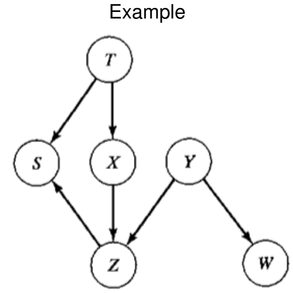
- Let $\mathcal{V}$ be a set of RVs, P their joint distribution, and $X \in \mathcal{V}$. A Markov blanket $\mathcal{M}_X$ of X is any set of variables such that X is CI of all other variables given $\mathcal{M}_X$:

$$I_P(\{X\}, \mathcal{V} \setminus (\mathcal{M}_X \cup \{X\}) \mid \mathcal{M}_X)$$

- If no proper subset of $\mathcal{M}_X$ is a Markov blanket, then $\mathcal{M}_X$ is a Markov boundary
- Theorem 2.13: If (G, P) satisfies the Markov condition, then the set of X's parents, children, and co-parents (other parents of X's children) form a Markov blanket of X
  - "Parent" respects edge direction
- Theorem 2.14: If (G, P) satisfies the faithfulness condition, then the set of X's parents, children, and co-parents form the unique Markov boundary of X

Markov Blankets and Boundaries

Example



- If the faithfulness condition is satisfied, then what is X's Markov boundary?

- What if the edge (T, X) is deleted?