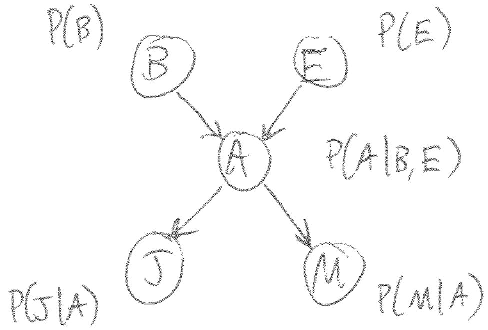
Review



- Belief network (BN) = DAG + CPTs

- Conditional Independence

$$P(X_1, X_2, \ldots, X_n) = \text{[illegible]}$$

$$P(X_1)P(X_2|X_1) \ldots P(X_n|X_1, X_2, \ldots, X_{n-1}) \quad \text{(product rule)}$$

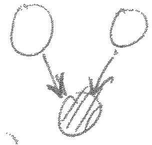
$$= \prod_i P(X_i|X_1, X_2, \ldots, X_{i-1})$$

$$= \prod_i P(X_i|pa(X_i))$$

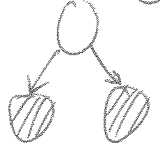
parents of X_i

- Types of reasoning

(1) Competing explanations of observed event



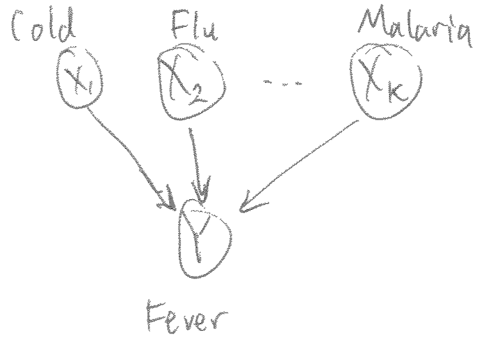
(2) Multiple events w/ common explanation



(shaded node = observed event)

(3) Intervening event

- Representing CPTs



Suppose $X_i \in \{0, 1\}$

$Y \in \{0, 1\}$

How to represent CPT $P(Y=1|X_1, X_2, \ldots, X_K)$?

i) Lookup table

$O(2^k)$ can store arbitrary CPT

X_1	X_2	$\cdots$	X_K	$P(Y=1 \mid X_1, X_2, \ldots, X_k)$
0	0		0	
1	0		0	
0	1		0	

ii) "Deterministic" node

"AND": $P(Y=1 \mid X_1, X_2, \ldots, X_K) = \prod_{i=1}^{k} X_i$

"OR": $P(Y=0 \mid X_1, X_2, \ldots, X_K) = \prod_{i=1}^{k} (1-X_i)$

iii) Noisy-OR node

Use k numbers $p_i \in [0,1]$ to parameterize $O(2^k)$ elements

$$P(Y=0 \mid X_1, X_2, \ldots, X_k) = \prod_{i=1}^{k} (1-p_i)^{X_i} \quad \text{for } X_i \in \{0,1\}$$

$$P(Y=1 \mid X_1, X_2, \ldots, X_k) = 1 - \prod_{i=1}^{k} (1-p_i)^{X_i}$$

Why called "noisy-OR"?

$$P(Y=1 \mid X_1=0, X_2=0, \ldots, X_k=0) = 1 - \prod_i \underbrace{(1-p_i)^0}_{1} = 0$$

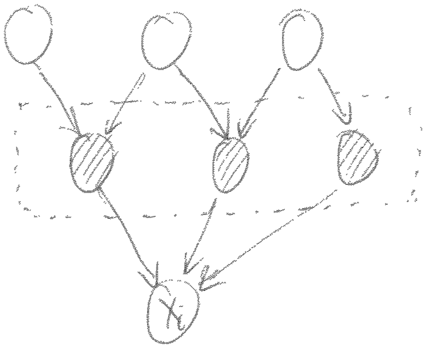
$$P(Y=1 \mid X_1=1, X_2=0, X_3=0, \ldots, X_K=0) = 1 - (1-p_1)^{(1)} \prod_{j>1}^{k} \underbrace{(1-p_j)^0}_{1}$$

$$= p_1$$

<u>Intuition</u>: $p_i \in [0,1]$ is the prob. that $X_i=1$ by itself triggers $Y=1$.

Conditional independence

A node is conditionally independent of its non-parent ancestors given its parent



$$P(X_i \mid pa(X_i)) = P(X_i \mid X_1, X_2, \ldots, X_{i-1})$$

More generally - Let X, Y, and E refer to sets of nodes. When is X conditionally independent of Y given evidence E?

When is $P(X \mid E, Y) = P(X \mid E)$?
$P(Y \mid E, X) = P(Y \mid E)$?
$P(X, Y \mid E) = P(X \mid E) P(Y \mid E)$?

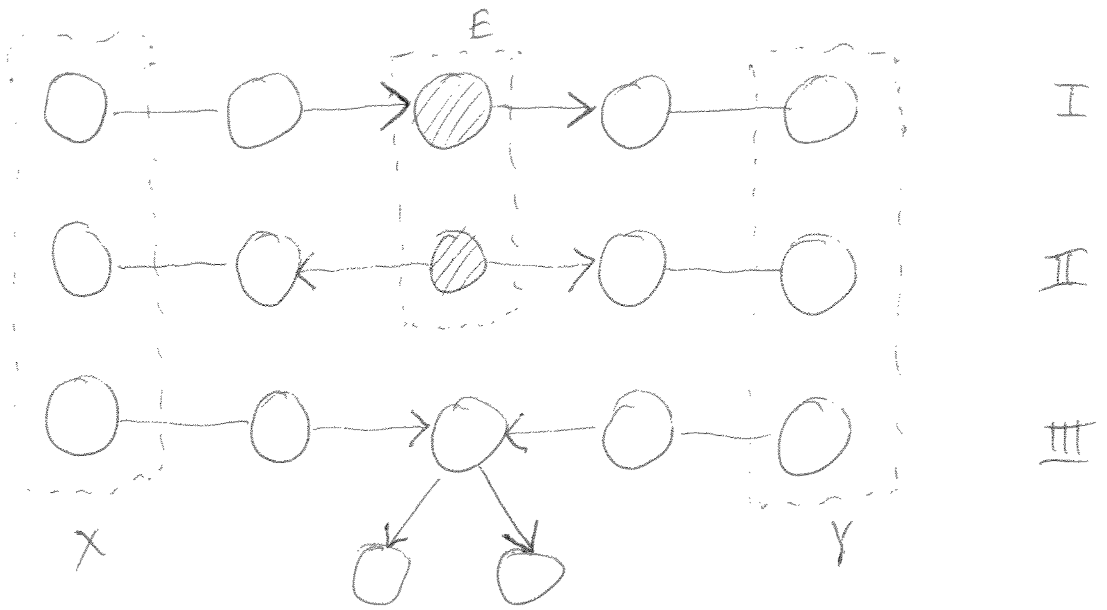
e.g. $X = \{X_i\}$ (special case above)
$E = \{pa(X_i)\}$
$Y = \{X_1, X_2, \ldots, X_{i-1}\} - pa(X_i)$

- d-separation

↳ "direction-dependent" separation

Relates conditional independence to graph-theoretic properties

$P(X, Y \mid E) = P(X \mid E) P(Y \mid E)$ if and only if <u>every</u> <u>undirected</u> path (ignoring edge direction) from any node in X to any node in Y is "d-separated" by E.

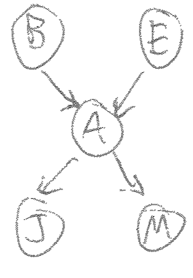


<u>Definition</u>: a path π is d-separated if there exists a node $z \in \pi$ for which one of 3 conditions hold:

1) $z \in E$ with →▨→ an intervening event
2) $z \in E$ with ←▨→ a common explanation
3) $z \notin E$, $desc(z) \notin E$, →Ⓩ← no observed common effects

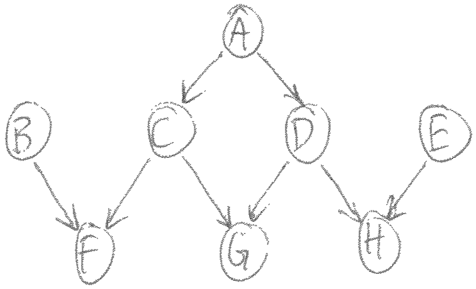
- Proof that d-separation ⟺ conditional independence is difficult, beyond cours
- Efficient algorithm exists for tests of d-separation

- Alarm example



1) $P(B|A,M) \overset{?}{=} P(B|A)$ $X = \{B\}$, $E = \{A\}$, $Y = \{M\}$
true: alarm is an intervening event (I)
2) $P(J,M|A) \overset{?}{=} P(J|A)P(M|A)$ $X = \{J\}$, $E = \{A\}$, $Y = \{M\}$
true: alarm is common explanation of J, M (II)
3) $P(B|E) \overset{?}{=} P(B)$ $X = \{B\}$, $E = \emptyset$, $Y = \{E\}$
true (III)
4) $P(B,E|J) \overset{?}{=} P(B|J)P(E|J)$ $X = \{B\}$, $E = \{J\}$, $Y = \{E\}$
false, explaining away; III not true: $desc(A) \in E$

- Loopy BN example



Statement	T/F
$P(D \mid H) \overset{?}{=} P(D \mid E, H)$	false
$P(F, H \mid A) \overset{?}{=} P(F \mid A)\, P(H \mid A)$	true
$P(F, G, H \mid A) \overset{?}{=} P(F \mid A)\, P(G \mid A)\, P(H \mid A)$	false

2 paths to check

$F \leftarrow C \leftarrow A \rightarrow D \rightarrow H$ I

$F \leftarrow C \rightarrow G \leftarrow D \rightarrow H$ II

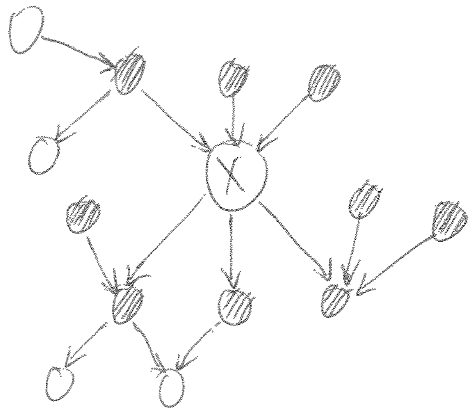
$\downarrow$

$$P(F, G, H \mid A) = P(F \mid A)\, \underset{\overset{\shortparallel\, ?}{P(G \mid A)}}{P(G \mid A, F)}\, \underset{\overset{\shortparallel\, ?}{P(H \mid A)}}{P(H \mid A, F, G)} \qquad \text{(product rule)}$$

$$\overset{?}{=} P(F \mid A)\; P(G \mid A)\; P(H \mid A)$$

No! (G-C-F not d-separated)

Def: Markov blanket B_X of individual node X consists of parents of X, children of X, and parents of children ("spouses") of X (not including X).



Theorem: A node X is conditionally independent of all nodes outside B_X given nodes inside B_X

$$P(X \mid B_X, Y) = P(X \mid B_X) \quad \text{where} \quad Y \notin \{B_X, X\}.$$