FUDAN SUMMER SCHOOL INTRODUCTION TO AI

Probabilistic Reasoning and Decision Making

Day 3 – Bayes Nets and Inference on BN

Consider the following three events

- E: There was an earthquake
- B: A burglar broke into my apartment
- A: My alarm at my apartment went off

Assume that B and E are marginally independent

$P(E|B)=P(E)$

$P(B|E)=P(B)$

1. Multiple explanations for a single event

- Compare :
  - $P(B=1)$
  - $P(B=1|A=1)$
  - $P(B=1|A=1,E=1)$

$P(B|E)=P(B)$

$P(B|EA) \neq P(B|A)$

Which of the following represents the most intuitive relationship between these probabilities?

$A.\ P(B=1) < P(B=1|A=1)$ ✓
$B.\ P(B=1) < P(B=1|A=1,E=1)$ ✓
$C.\ P(B=1) > P(B=1|A=1)$
$D.\ P(B=1) > P(B=1|A=1,E=1)$
E. More than one of these (which ones?)

1. Multiple explanations for a single event

- Compare :
  - $P(B = 1)$
  - $P(B = 1|A = 1)$
  - $P(B = 1|A = 1, E = 1)$

Which of the following represents the most intuitive relationship between these probabilities?

$A.\ P(B = 1|A = 1) < P(B = 1|A = 1, E = 1)$

$B.\ P(B = 1|A = 1) > P(B = 1|A = 1, E = 1)$ explain away

$C.\ P(B = 1|A = 1) = P(B = 1|A = 1, E = 1)$

1. Multiple explanations for a single event

- Compare :
  - $P(B = 1)$
  - $P(B = 1|A = 1)$
  - $P(B = 1|A = 1, E = 1)$

Intuitively, we expect:
$P(B = 1) < P(B = 1|A = 1, E = 1) < P(B = 1|A = 1)$

I.e. "I generally believe there is a very low chance that a burglar breaks in. But if I get notified that my alarm is going off, then I think the alarm is probably caused by the burglar and my belief that a burglar broke in goes up. However, if I then hear about an earthquake, I no longer am as fearful that a burglar broke in, but I'm still more worried than I would have been if the alarm hadn't gone off at all"

1. Multiple explanations for a single event

$$\boxed{\begin{array}{l} P(B=1) \\ P(B=1|A=1) \\ P(B=1|A=1, E=1) \end{array}}$$

Domain knowledge:

Assumption: B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$

<u>Prior probabilities:</u>

$$P(B=1) = 0.001 \quad P(B=0) = 1 - 0.001 = 0.999$$
$$P(E=1) = 0.005 \quad P(E=0) = 1 - 0.005 = 0.995$$

<u>Conditional probabilities (i.e. properties of the alarm):</u>

B	E	$P(A=1\|B,E)$	$P(A=0\|B,E)$
0	0	0.002	1-.002 = .998
0	1		
1	0		
1	1		

CPT

(conditional probability table)

1. Multiple explanations for a single event

Marginalization
product
Bayes $P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$

$P(B = 1)$ $= .001$
$P(B = 1|A = 1)$ $= .2044$
$P(B = 1|A = 1, E = 1)$ $\simeq .002795$

Domain knowledge:

Assumption: B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$

Prior probabilities:

$P(B = 1) = 0.001$ $P(B = 0) = 1 - 0.001 = 0.999$

$P(E = 1) = 0.005$ $P(E = 0) = 1 - 0.005 = 0.995$

$P(B=1|A=1, \boxed{E=1})$
$= \frac{P(A=1|B=1,E=1)P(B=1|E=1)}{P(A=1|E=1)}$
$= \frac{P(A=1|B=1)P(B=1)}{P(A=1)}$

Conditional probabilities (i.e. properties of the alarm):

M $P(A=1|E=1)$
$= P(A=1,B=1|E=1) + P(A=1,B=0|E=1)$
$P = P(B=1)P(A=1|B=1,E=1) + P(B=0)P(A=1|B=0,E=1)$

B	E	$P(A = 1\|B, E)$	$P(A = 0\|B, E)$
0	0	0.002	0.998
0	1	.35	.65
1	0	.96	.04
1	1	.98	.02

$P(B=1 \mid A=1)$

$\overset{B}{=} \dfrac{P(A=1 \mid B=1)\, P(B=1)}{P(A=1)}$

$= \dfrac{.9601 \times .001}{.004696} \approx .2044$

$P(A=1 \mid B=1)$

$\overset{M}{=} P(A=1, E=1 \mid B=1) + P(A=1, E=0 \mid B=1)$

$\overset{P}{=} P(E=1 \mid \cancel{B=1})\, P(A=1 \mid E=1, B=1) + P(E=0 \mid \cancel{B=1})\, P(A=1 \mid E=0, B=1)$

independence

$= .005 \times .96 + .995 \times .96$

$= .9601$

A: Bayes

B: Marginal

C: Product

$P(A=1)$

$\overset{M}{=} P(B=1, E=1, A=1) + P(B=0, E=1, A=1) + P(B=1, E=0, A=1) + P(B=0, E=0, A=1)$

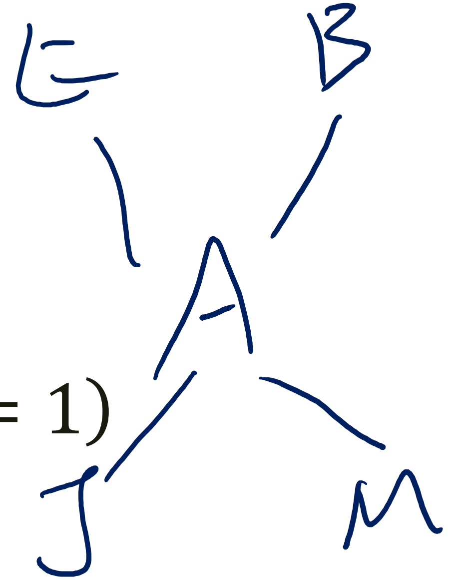
$\overset{P}{=} P(B=1)\, P(E=1 \mid \cancel{B=1})\, P(A=1 \mid B=1, E=1) + P(E=1)\, P(B=0)\, P(A=1 \mid B=0, E=1) + P(B=1)\, P(E=0)\, P(A=1 \mid B=1, E=0) + P(B=0)\, P(E=0)\, P(A=1 \mid B=0, E=0)$

$= .0046936$

2. Multiple events with a common explanation

Two new variables: J: Jamal calls, M: Maya calls

- Compare :
  - $P(A = 1)$
  - $P(A = 1|J = 1)$
  - $P(A = 1|M = 0, J = 1)$



M : maya
J : Jamal
E : earthquake
B : Burgalory
A : alarm

If we assume J, M, B and E are conditionally independent given A, which of the following statements is true?

✓ A. $P(J|A, \cancel{M}, \cancel{B}, \cancel{E}) = P(J|A)$
B. $P(J|A, \cancel{M}, \cancel{B}, \cancel{E}) = P(J)$ ×
C. $P(J, M, B, E|A) = P(J|A)$ ×
D. More than one of these

$$P(J, M, B, E | A)$$
$$= P(J|A)\,P(M|\cancel{J}A)\,P(B|\cancel{J}A\cancel{M})\,P(E|\cancel{J}\cancel{M}\cancel{B}A)$$

2. Multiple events with a common explanation

$P(A=1) = 0.00469636$ smallest

$P(A=1|J=1)$ largest $= .042$

$P(A=1|M=0, J=1) \leftarrow = .0152$

$P(A=0) = 0.99530364$

Domain knowledge:

Assumptions:

- B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$
- J, M conditionally independent from each other and from B and E, given A, e.g. $P(J|A,M,B,E) = P(J|A)$

CPT

A	$P(J=1\|A)$
0	.1
1	.93

Prior probabilities:

$P(B=1) = 0.001 \quad P(B=0) = 1 - 0.001 = 0.999$

$P(E=1) = 0.005 \quad P(E=0) = 1 - 0.005 = 0.995$

Conditional probabilities (i.e. properties of the alarm, Jamal and Maya):

B	E	$P(A=1\|B,E)$	$P(A=0\|B,E)$
0	0	0.002	0.998
0	1	0.35	0.65
1	0	0.96	0.04
1	1	0.98	0.02

$P(J=1|A=1) = .93$ $\quad P(J=0|A=1) = .07$

$P(J=1|A=0) = .1$ $\quad P(J=0|A=0) = .9$

$P(M=1|A=1) = .65$ $\quad P(M=0|A=1) = .35$

$P(M=1|A=0) = .01$ $\quad P(M=0|A=0) = .99$

A	$P(M=1\|A)$
0	--
1	–

$$P(A=1 \mid J=1) \stackrel{B}{=} \frac{P(J=1 \mid A=1)\, P(A=1)}{P(J=1)}$$

$$P(J=1) \stackrel{M}{=}$$

$$P(A=1 \mid M=0, J=1) = \frac{P(J=1 \mid A=1, M=0)\, P(A=1 \mid M=0)}{P(J=1 \mid M=0)}$$

$$\overset{B}{=} \frac{P(M=0, J=1 \mid A=1)\, P(A=1)}{P(M=0, J=1)}$$

$$\overset{P,M}{=} \frac{P(M=0 \mid A=1)\, P(J=1 \mid \cancel{M=0}, A=1)\, P(A=1)}{P(M=0, J=1, A=1) + P(M=0, J=1, A=0)}$$

$$\overset{P}{=} \frac{}{P(A=1)\, P(M=0 \mid A=1)\, P(J=1 \mid \cancel{M=0}, A=1) + P(A=0)\, P(M=0 \mid A=0)\, P(J=1 \mid \cancel{M=0}, A=0)}$$

3. Chain of Linked Events

- Compare :
  - $P(A = 1) \approx .004686$
  - $P(A = 1|J = 1) \approx .042$
  - $P(A = 1|J = 1, B = 1)$? $= .99555$

A : ? largest

B : ? in middle

C : smallest

$$P(A=1|J=1,B=1)$$

$$\overset{B}{=} \frac{P(J=1|A=1,B=1)\,P(A=1|B=1)}{P(J=1|B=1)}$$

$$= \frac{P(J=1|A=1)\,P(A=1|B=1)}{??}$$

$$P(J=1|B=1)$$

$$\overset{m}{=} P(J=1, A=1|B=1) + P(J=1, A=0|B=1)$$

$$\overset{P}{=} P(A=1)\,P(J=1|A=1,\cancel{B=1}) + P(A=0)\,P(J=1|A=0,\cancel{B=1})$$

3. Chain of Linked Events

$P(A=1) = 0.00469636$
$P(A=1|J=1) \approx 0.0420375$
$P(A=1|J=1, B=1)$

$P(A=0) = 0.99530364$

Domain knowledge:

Assumptions:

- B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$
- J, M conditionally independent from each other and from B and E, given A, e.g. $P(J|A, M, B, E) = P(J|A)$

Prior probabilities:

$P(B=1) = 0.001 \quad P(B=0) = 1 - 0.001 = 0.999$
$P(E=1) = 0.005 \quad P(E=0) = 1 - 0.005 = 0.995$

Conditional probabilities (i.e. properties of the alarm, Jamal and Maya):

B	E	$P(A=1\|B,E)$	$P(A=0\|B,E)$
0	0	0.002	0.998
0	1	0.35	0.65
1	0	0.96	0.04
1	1	0.98	0.02

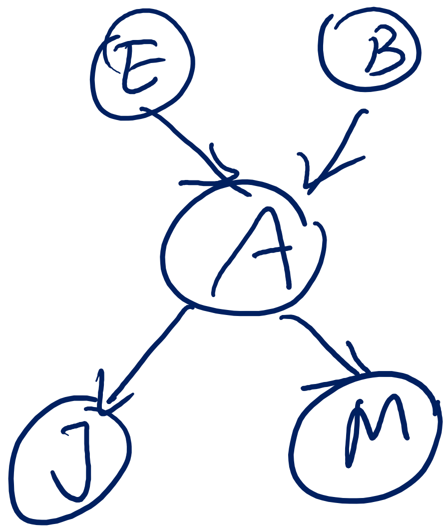
$P(J=1|A=1) = 0.93 \quad P(J=0|A=1) = 1 - 0.93 = 0.07$
$P(J=1|A=0) = 0.1 \quad P(J=0|A=0) = 1 - 0.1 = 0.9$

$P(M=1|A=1) = 0.65 \quad P(M=0|A=1) = 1 - 0.65 = 0.35$
$P(M=1|A=0) = 0.01 \quad P(M=0|A=0) = 1 - 0.01 = 0.99$

Reality check: Reasoning about multiple events

- Order the following probabilities (beliefs) from smallest to largest, using the domain knowledge from last class:

$$P(A = 1)$$
$$P(A = 1|J = 1)$$
$$P(A = 1|J = 1, B = 0)$$



A. $P(A = 1) < P(A = 1|J = 1) < P(A = 1|J = 1, B = 0)$
B. $P(A = 1) < P(A = 1|J = 1, B = 0) < P(A = 1|J = 1)$
C. $P(A = 1|J = 1, B = 0) < P(A = 1) < P(A = 1|J = 1)$
D. $P(A = 1|J = 1) < P(A = 1|J = 1, B = 0) < P(A = 1)$
E. None of these

Reality Check: Reasoning about multiple events

1. Multiple explanations (Burglary, Earthquake) for a single event (the Alarm):

$$P(B = 1) = 0.001$$
$$P(B = 1|A = 1) \approx 0.2044$$
$$P(B = 1|A = 1, E = 1) \approx 0.002795$$

2. Multiple events (Jamal calls, Maya calls) with a common explanation (the Alarm):

$$P(A = 1) = 0.00469636$$
$$P(A = 1|J = 1) \approx 0.0420375$$
$$P(A = 1|M = 0, J = 1) \approx 0.015277$$

3. Chain of linked events: Burglary causes Alarm causes Jamal to call

$$P(A = 1) = 0.00469636$$
$$P(A = 1|J = 1) \approx 0.0420375$$
$$P(A = 1|J = 1, B = 1) \approx 0.99555$$

Joint distribution

Assumptions about the world:

- B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$
- J, M conditionally independent from each other and from B and E, given A, e.g. $P(J|A, M, B, E) = P(J|A)$

$$P(B, E, A, J, M)$$

Which of the following also correctly represents the joint distribution?

✓ A. $P(B)P(E|B)P(A|E, B)P(J|A, E, B)P(M|J, A, E, B)$ ✓ product rule

✓ B. $P(B)P(E)P(A|E, B)P(J|A)P(M|A)$ ✓

C. $P(B)P(E)P(A)P(J)P(M)$ ✗

D. $P(B|A)P(E|A)P(A)P(J|A)P(M|A)$ ✗

E. More than one of these (which?) ✗

$P(A)P(B|A)P(E|AB)$

Joint distribution and Belief/Bayes Nets

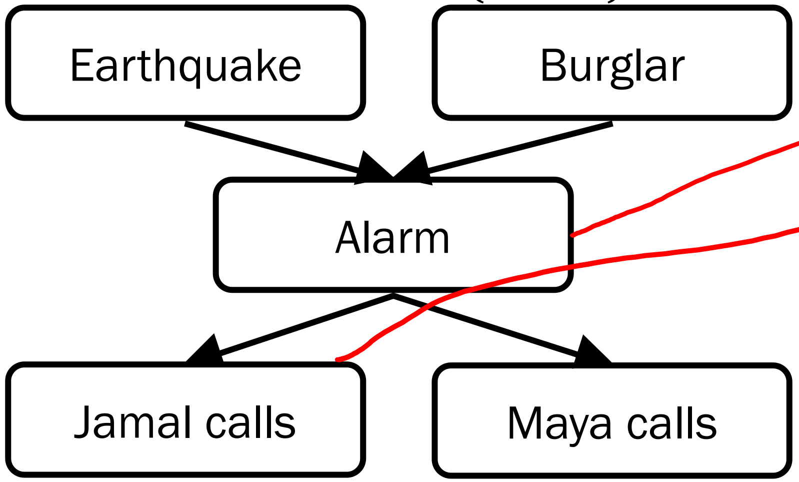
DAG: directed acyclic graph

<u>*Assumptions about the world*</u>:

- B and E are marginally independent: $P(B|E) = P(B)$, $P(E|B) = P(E)$
- J, M conditionally independent from each other and from B and E, given A, e.g. $P(J|A, M, B, E) = P(J|A)$

$$P(B,E,A,J,M) = P(B)P(E|B)P(A|E,B)P(J|A,E,B)P(M|J,A,E,B) = P(B)P(E)P(A|E,B)P(J|A)P(M|A)$$

$P(E=1) = 0.001$ $\quad$ $P(B=1) = 0.005$



CPT

B	E	$P(A=1\|B,E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

A	$P(J=1\|A)$
0	0.1
1	0.93

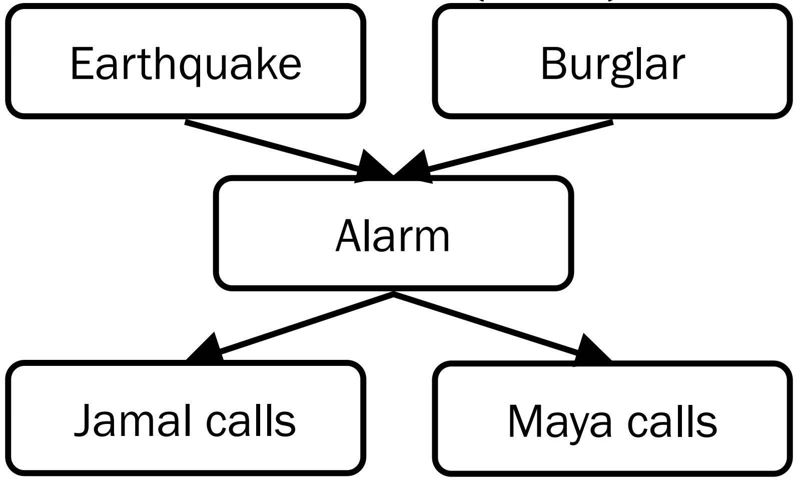
A	$P(M=1\|A)$
0	0.01
1	0.65

A BN (Belief Network or Bayes Net) is a DAG in which:

1. node in BN represents r.v.
2. edge in BN represents conditional dependence
3. Conditional probability describes how a node depends on its parents.

Reasoning using Bayes Nets

$P(E = 1) = 0.001$ $\quad$ $P(B = 1) = 0.005$



B	E	$P(A = 1 \mid B, E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

A	$P(J = 1 \mid A)$
0	0.1
1	0.93

A	$P(M = 1 \mid A)$
0	0.01
1	0.65

$P(A=1, J=0, M=0)$

$= P(J=0 \mid \cancel{A=0}, A=1)\, P(M=0 \mid A=1)\, P(A=1)$

Bayes Net represents the joint prob over its vars.

Which of the following probabilities is most straightforward to computing using the information in the Bayes net?

A. $P(A = 1, J = 0, M = 0)$

B. $P(A = 1, J = 0, M = 0 \boxed{|} B = 1, E = 0)$

C. $P(A = 1, J = 0, M = 0, B = 1, E = 0)$

D. $P(A = 1)$

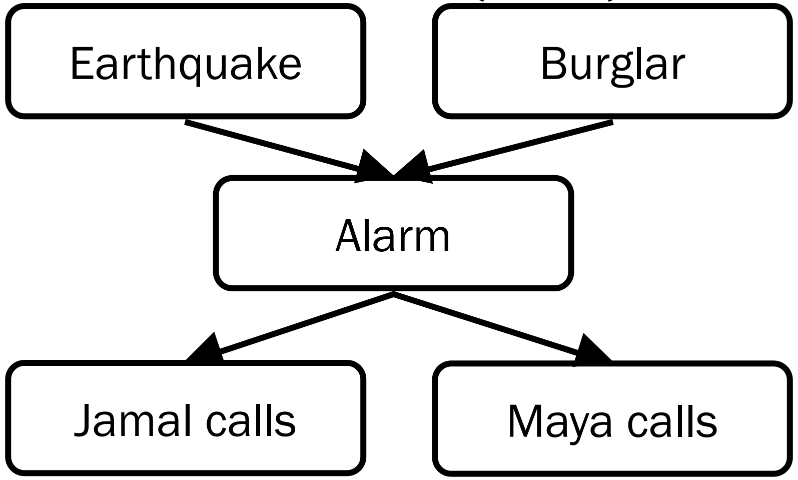
C

$\Rightarrow P(A=1, J=0, M=0, B=1, E=0) = P(E=0)\, P(B=1)\, P(A=1 \mid B=1, E=0)\, P(J=0 \mid A=1)\, P(M=0 \mid A=1)$

.999 $\quad$ .005 $\quad$ .96 $\quad$ .07 $\quad$ .35

Reasoning using Bayes Nets

$P(E=1) = 0.001$ $P(B=1) = 0.005$



A	$P(J=1\|A)$
0	0.1
1	0.93

A	$P(M=1\|A)$
0	0.01
1	0.65

B	E	$P(A=1\|B,E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

$P(A=1 \mid B=1)$

$\overset{M}{=} P(A=1, E=1 \mid B=1) + P(A=1, E=0 \mid B=1)$

$= P(E=1)\,P(A=1 \mid B=1, E=1) + P(E=0)\,P(A=1 \mid E=0, B=1)$

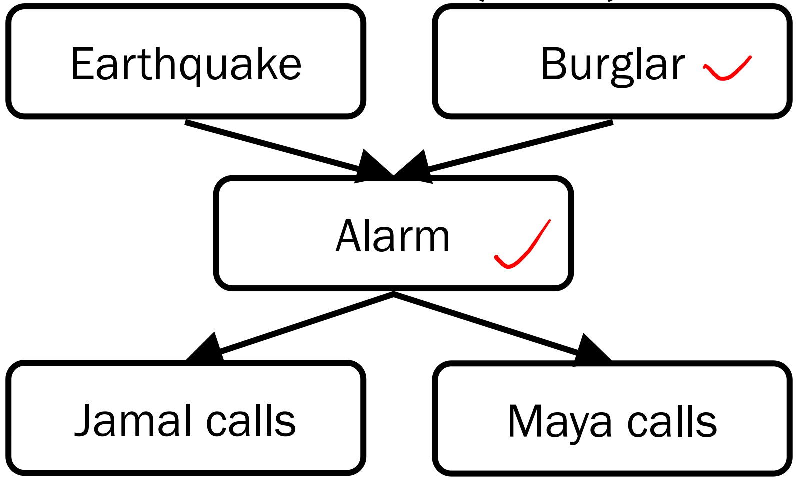
$= .9601$

Bayes' Nets also give a straightforward way to calculate any probability:

Reasoning using Bayes Nets

w/ BN

$P(E=1) = 0.001$ $P(B=1) = 0.005$



A	$P(J=1\|A)$
0	0.1
1	0.93

A	$P(M=1\|A)$
0	0.01
1	0.65

B	E	$P(A=1\|B,E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

$$P(A=1 \mid B=1)$$

$$= \frac{P(A=1, B=1)}{P(B=1)}$$

$$= \frac{\sum_{i,j,k} P(A=1, B=1, E=e_i, J=j_j, M=m_k)}{\sum_{i,j,k,p} P(B=1, A=a_i, E=e_j, J=j_k, M=m_p)}$$

Bayes' Nets also give a straightforward way to calculate any probability:

$$= .968$$

Reasoning using Bayes Nets

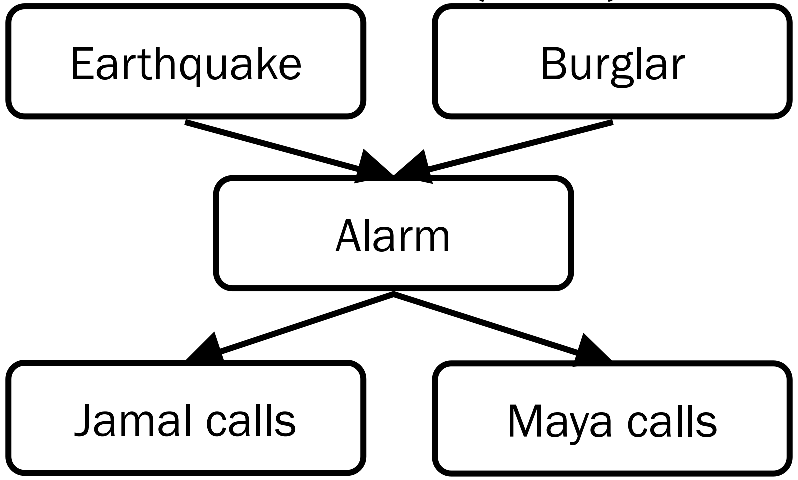
10 entries

1

$P(E = 1) = 0.001$

1

$P(B = 1) = 0.005$



B	E	$P(A = 1 \mid B, E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

4

A	$P(J = 1 \mid A)$
0	0.1
1	0.93

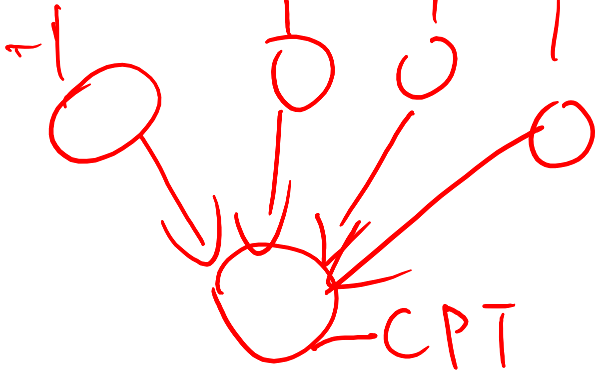
2

A	$P(M = 1 \mid A)$
0	0.01
1	0.65

2

Bayes' Nets also give a straightforward way to calculate any probability:

20



$2^4 = 16$ values

Bayes' nets give you a compact way to represent joint probabilities

- Let $\{X_1, X_2, X_3, \ldots, X_n\}$ be n distinct binary random variables. (Approximately) how many numbers do you need to specify the full joint distribution over all X's?

  A. n
  B. 2n
  C. 2^n
  D. n^n

$P(x_1, x_2 \ldots x_n)$

0 1 0 1

2 2 . . .

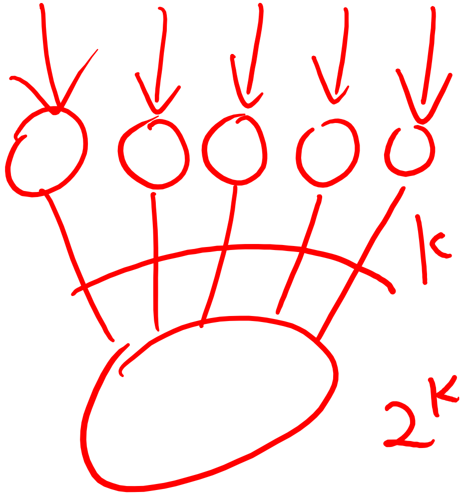
2^n

Bayes' nets give you a compact way to represent joint probabilities

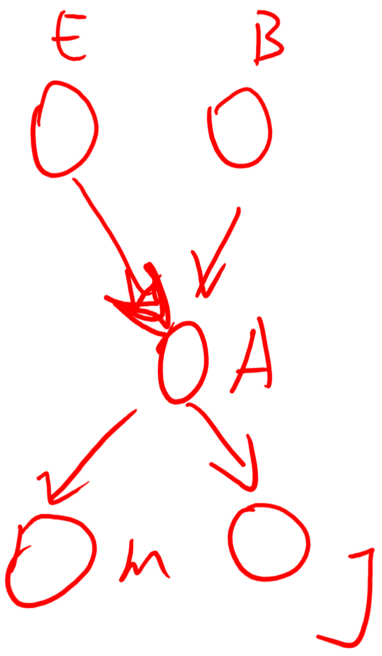
- Let $\{X_1, X_2, X_3, \ldots, X_n\}$ be n distinct binary random variables. (Approximately) how many values do you need to specify the full joint distribution over all X's? 2^n
- What does the number of values you need to specify in a Bayes' net depend on?
  - A. The average number of parents a node has
  - B. The maximum number of parents any node has
  - C. The average number of children a node has
  - D. The maximum number of children any node has

k : # of max parents in BN

$CPT \in O(n \cdot \underbrace{2^k}_{})$



Conditional independence revisited



$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_2, \ldots, X_{n-1})$$

$$= \prod_{i=1}^{n} P(X_i|\underbrace{X_1, X_2, \ldots, X_{i-1}})$$

Suppose that in any given domain:

$$P(x_1, x_2 \cdots x_n) = \prod_{i=1}^{n} P(x_i | \underline{\text{parents}(x_i)})$$

$$\text{parents}(x_i) \subseteq \{x_1 \cdots x_{i-1}\}$$

How to build a Bayes Net?

1. Choose random variables
2. Choose an ordering
3. While there are variables left:
    1. *Add node* X_i *to the BN*
    2. *Set parents of* X_i *to the minimal subset such that* X_i *is conditionally independent from non-parent nodes preceding i, given its parents.*
    3. *Define* $P(X_i|parent(X_i))$ *CPT*

How to build a Bayes Net?

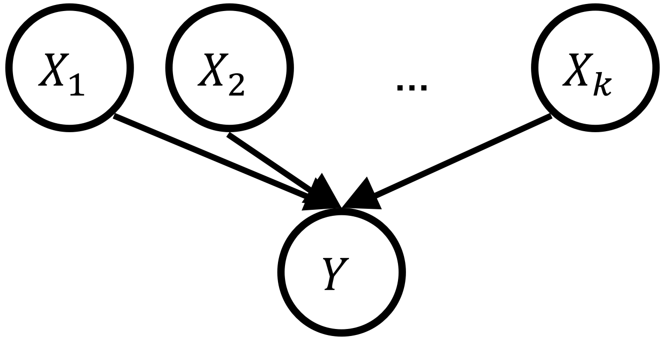
1. Choose random variables
2. Choose an ordering
3. While there are variables left:
   1. *Add node X_i to the BN*
   2. *Set parents of X_i to the minimal subset such that X_i is conditionally independent from non-parent nodes preceding i, given its parents.*
   3. *Define $P(X_i|parent(X_i))$ CPT*

Which is the better ordering to choose?

A. Root events first (i.e. causes), e.g. B, E, A, J, M
B. Child events first (i.e. consequences), e.g. M, J, A, E, B
C. It doesn't matter

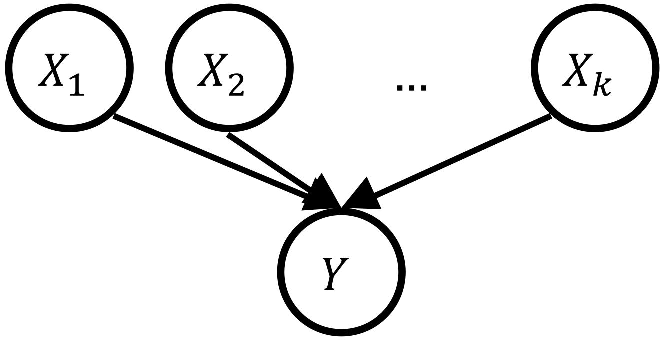
Representing CPTs

How do we represent the CPT?



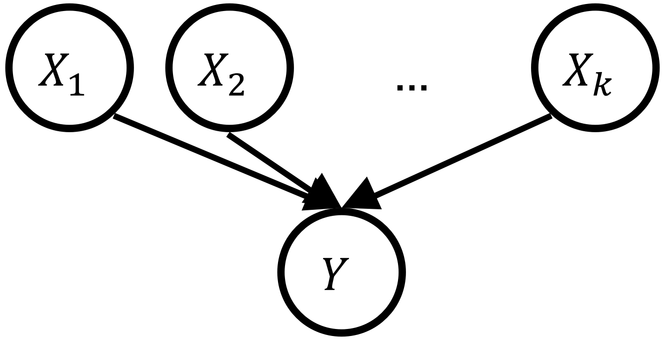
Representing CPTs

How do we represent the CPT?

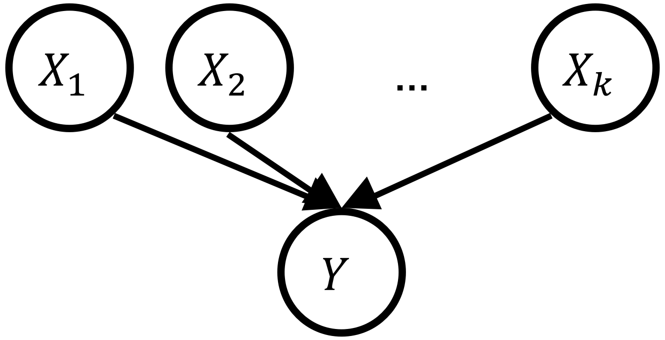


Representing CPTs

How do we represent the CPT?



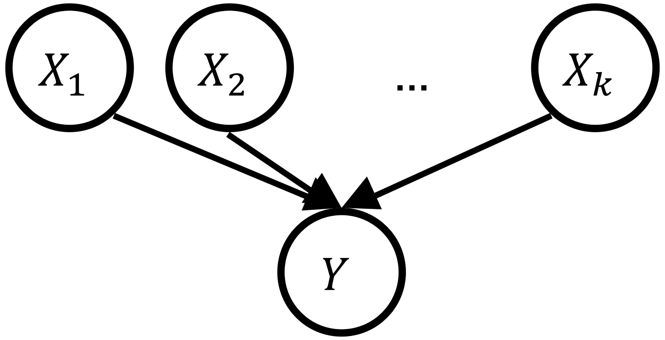
Noisy-OR



What is the value of $P(Y = 1 | X_1 = 0, X_2 = 0, \ldots, X_k = 0)$?

A. 0
B. p_i
C. 1
D. None of these
E. Not enough information given

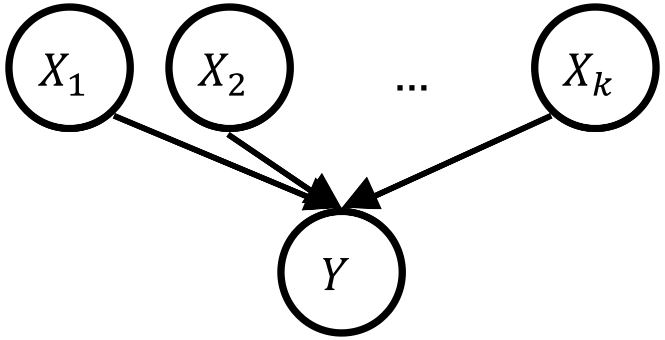
Noisy-OR



What is the value of $P(Y = 1|X_1 = 0, X_2 = 0, \ldots, X_i = 1, \ldots, X_k = 0)$?

A. 0
B. p_i
C. 1
D. None of these
E. Not enough information given

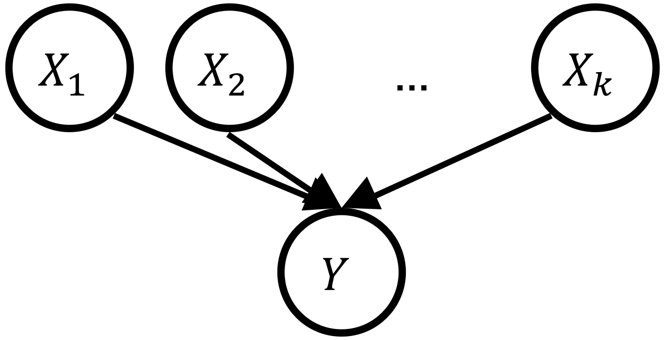
Noisy-OR



Let $P(Y = 1 | X_1 = 0, \dots, X_i = 1, X_j = 1, \dots, X_k = 0) = \frac{1}{3}$
and $p_i = \frac{1}{5}$
What is p_j?

A. $^1/_6$
B. $^2/_5$
C. $^1/_{15}$
D. $^2/_3$
E. I have no idea

Noisy-OR



Let $P\left(Y = 1 | X_1 = 0, \ldots, X_i = 1, X_j = 1, \ldots, X_k = 0\right) = \frac{1}{3}$

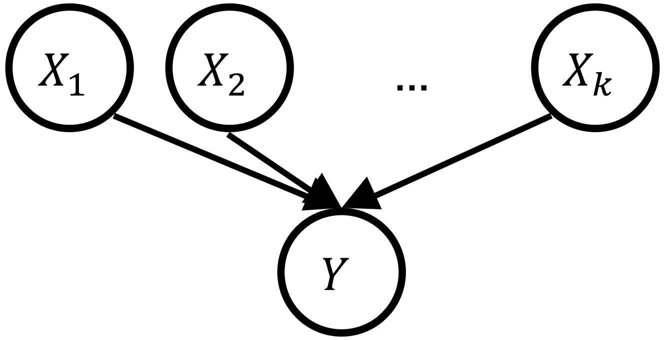
and $p_i = \frac{1}{5}$

What is p_j?

A. ${}^{1}/_{6}$
B. ${}^{2}/_{5}$
C. ${}^{1}/_{15}$
D. ${}^{2}/_{3}$
E. I have no idea

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

Noisy-OR

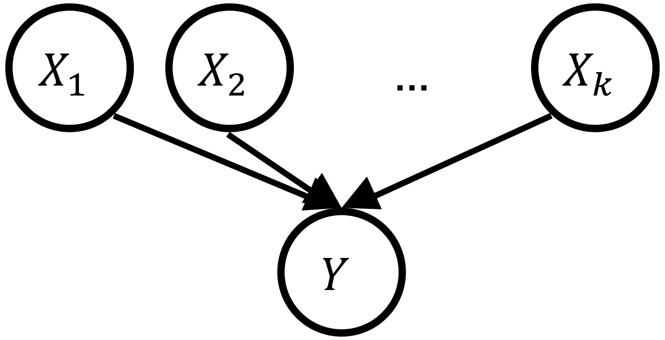


Is it true that P(Xi=1) = pi

A. Yes
B. NO

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

Noisy-OR: Intuition

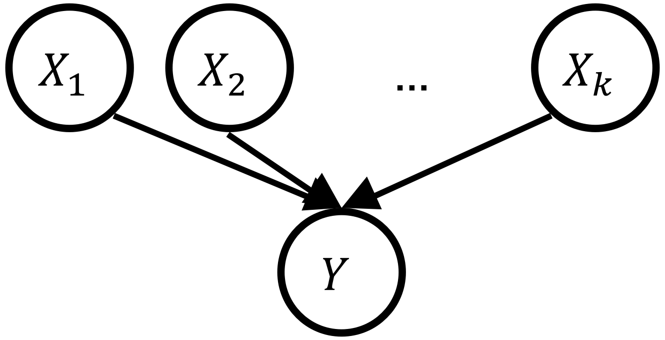


$$P(Y = 1 | X_1, \dots, X_k) = 1 - \prod_{i=1}^{k} (1 - p_i)^{X_i}$$

Let $P(Y = 1 | X_1 = 1, X_2 = 1, \dots, X_{i-1} = 1, X_i = 0, \dots, X_k = 0) = a$
and $P(Y = 1 | X_1 = 1, X_2 = 1, \dots, X_{i-1} = 1, X_i = 1, X_{i+1} = 0, \dots, X_k = 0) = b$
What is the relationship between a and b?

A. $a \leq b$
B. $a = b$
C. $a \geq b$
D. There is not enough information to determine the relationship

Noisy-OR: Intuition



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

- When all parents are off, child is off with certainty (just like OR).
- Each parent that turns on increases (or at least does not decrease) the probability that the child will turn on.
- When all the parents are on, it still might not be certain that the child is on, but it is probably very likely.

Bayesian Networks: Network structure and conditional independence

- By our construction of a Bayes' net, we know that a node is conditionally independent of its non-parent ancestors given its parents:

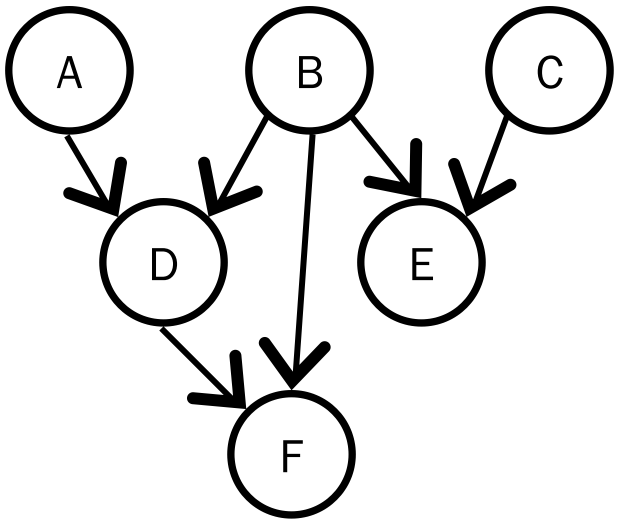
$$
\begin{aligned}
P(X_1, X_2, \ldots, X_n) &= P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_2, \ldots, X_{n-1}) \\
&= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \\
&= \prod_{i=1}^{n} P\big(X_i|parents(X_i)\big)
\end{aligned}
$$

But what about more general assertions of independence?

Bayesian Networks: Network structure and conditional independence

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

Write three equalities using $P(X|Y,E)$, $P(Y|X,E)$, $P(X,Y|E)$, etc, that indicate this conditional independence.

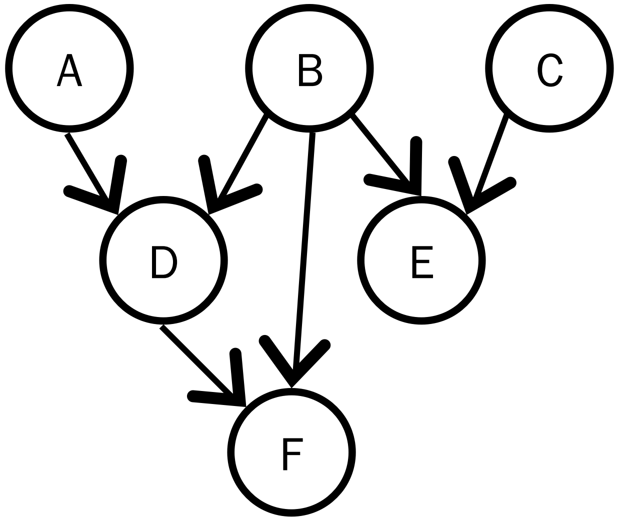


Bayesian Networks: Network structure and conditional independence

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

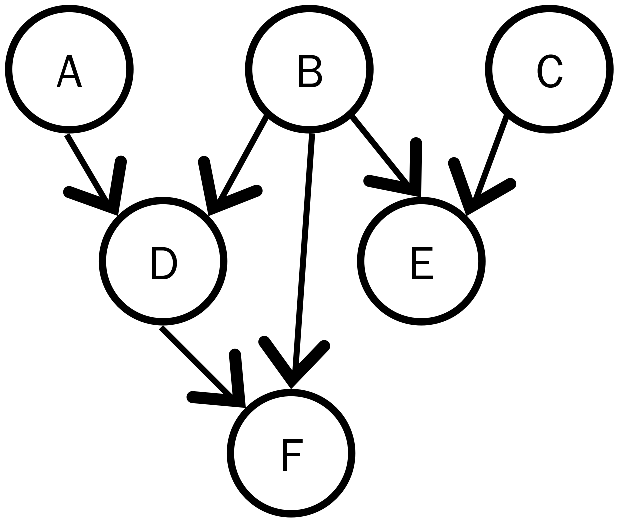
<u>Theorem</u>: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if every UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

(we will not prove this)



d-separation ("dependence" separation)

Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which one of the following three conditions hold:



Conditional Independence

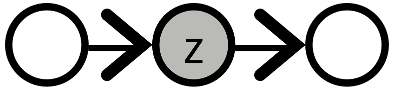
$P(B|A,M) = P(B|A)$

A. True

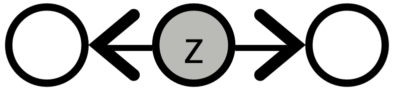
B. False

<u>Theorem</u>: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

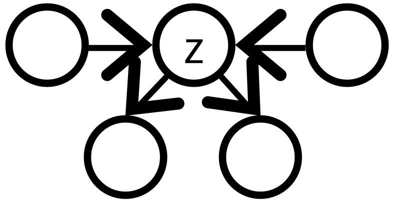
<u>Definition</u>: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

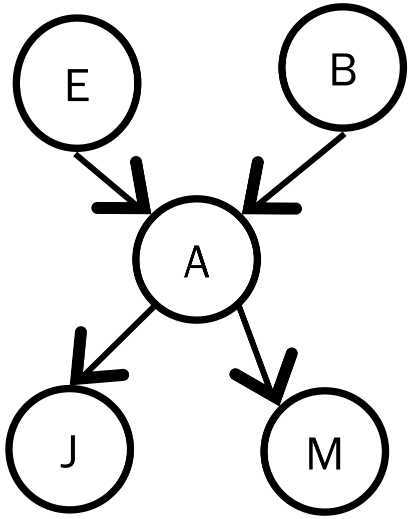
$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)

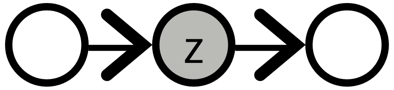


Conditional Independence

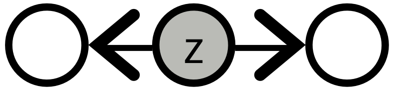
$P(J, M|A) = P(J|A)P(M|A)$
A. True
B. False

Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

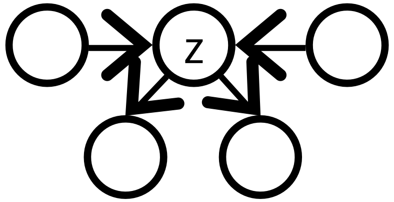
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

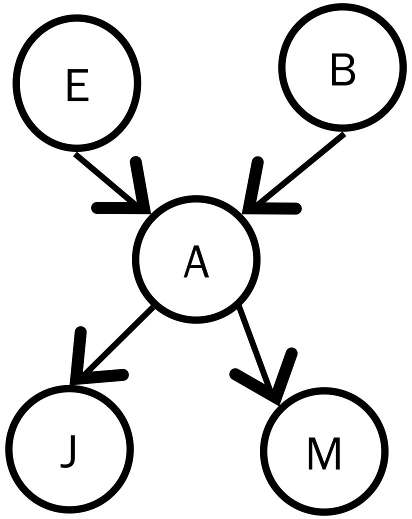
$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)

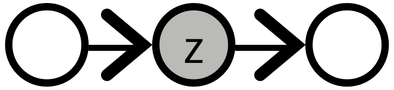


Conditional Independence

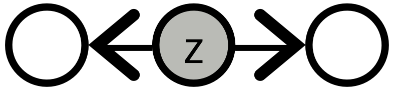
$P(B|E,J)=P(B|J)$
A. True
B. False

<u>Theorem</u>: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

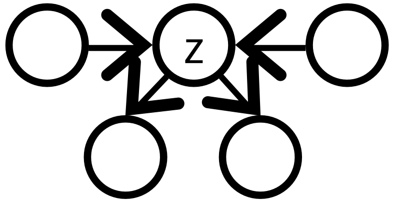
<u>Definition</u>: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

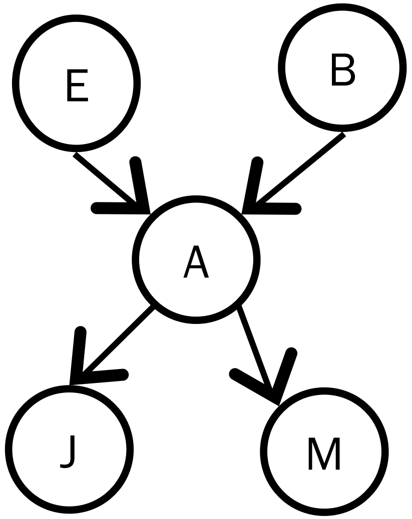
$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)

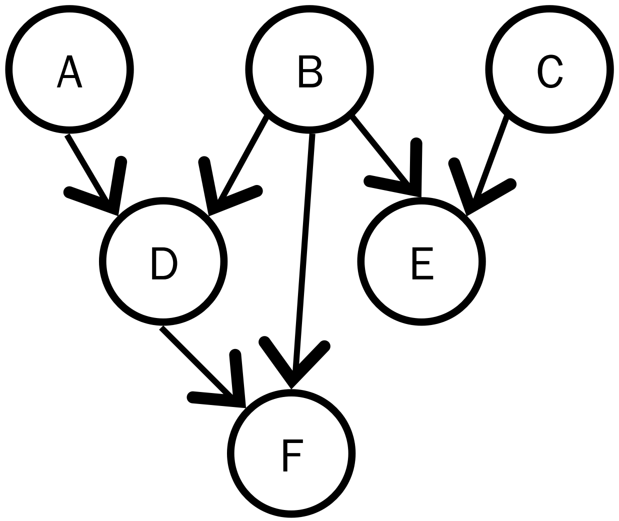


Conditional Independence

$P(A, B|D) = P(A|D)P(B|D)$?

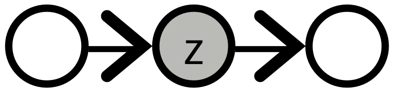
A. True

B. False

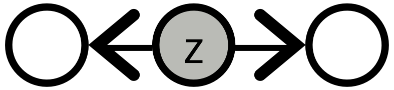


Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

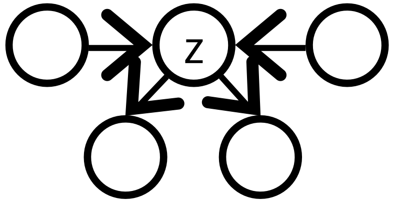
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

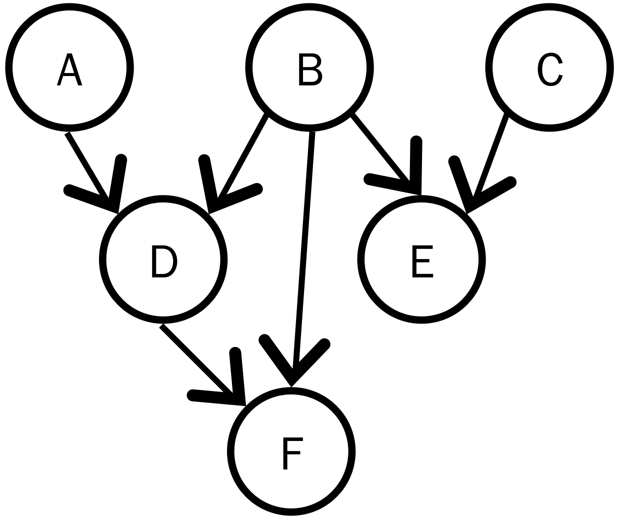
3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)

Conditional Independence

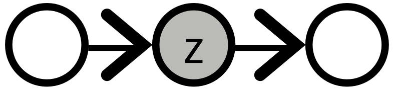
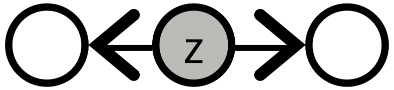
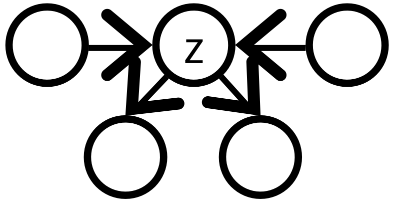
$P(D, E|B) = P(D|B)P(E|B)$?

A. True
B. False



Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

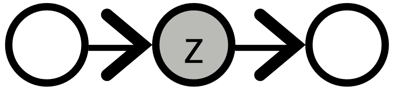
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1.  $z \in E$ (observed event in causal chain)
2.  $z \in E$ (observed common explanation)
3.  $z \notin E, descendants(z) \notin E$ (no observed common effect)

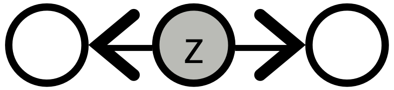
Conditional Independence

Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

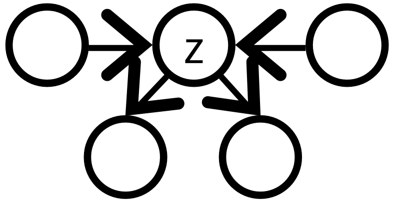
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

$z \in E$ (observed event in causal chain)

2. 

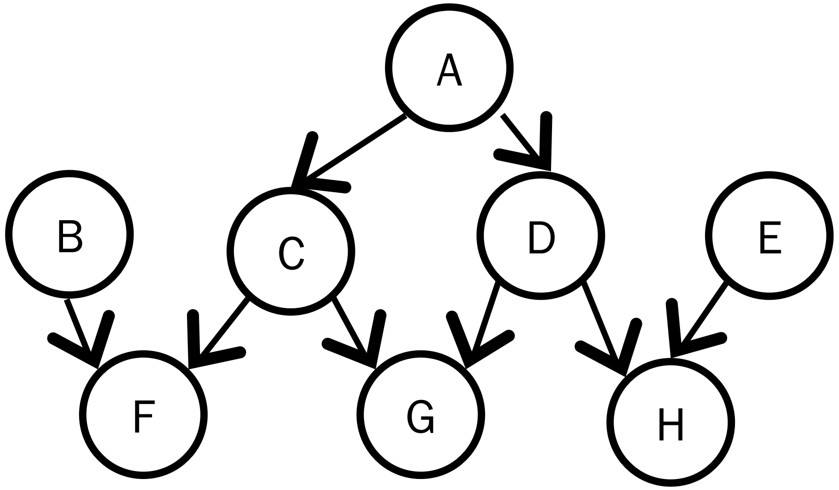
$z \in E$ (observed common explanation)

3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)

$\boxed{P(F, H|A, E) = P(F|A)P(H|A, E)?}$

A. X={F}, Y={H}, Ev={A, E}
B. X={F, H}, Y={E}, Ev={A}
C. X={F}, Y={H, E}, Ev={A}
D. X={F, H, E}, Y={A}, Ev={}



Conditional Independence

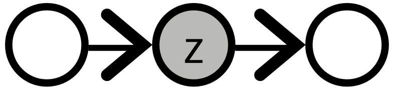
$\boxed{P(F, H|A, E) = P(F|A)P(H|A, E)?}$

X={F}, Y={H, E}, Ev={A}

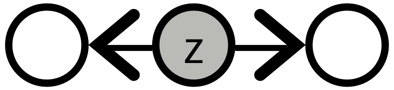
A. True
B. False

Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

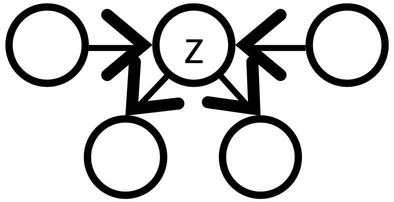
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

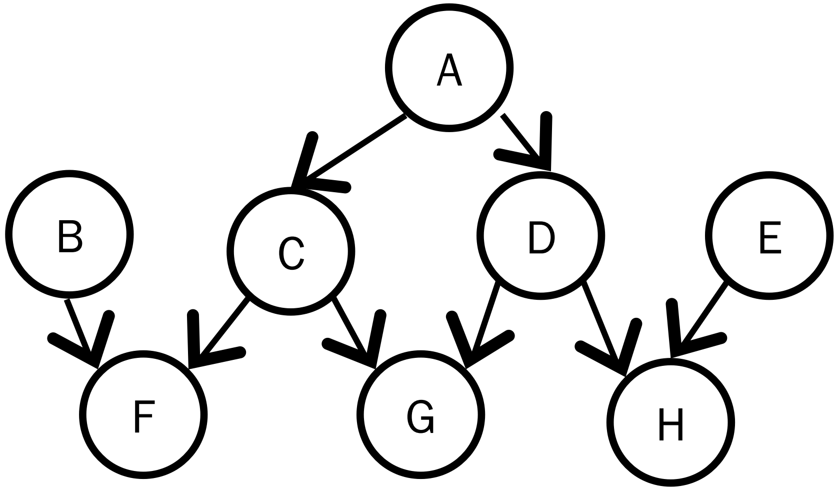
$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

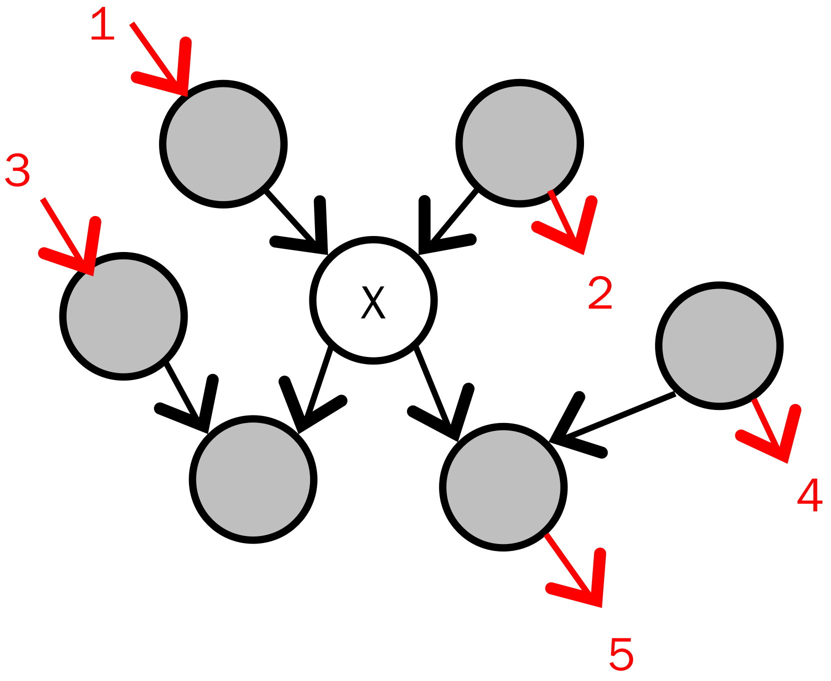
3. 

$z \notin E, descendants(z) \notin E$ (no observed common effect)



Markov Blanket

- A Markov Blanket B_x of node X consists of parents of X, children of X and "spouses" (other parents of children of X, but not X) of X.



<u>Theorem:</u> $P(X|B_x, Y) = P(X|B_x)$
$if\ Y \notin \{X, B_x\}$

Prove this theorem by showing that B_x d-separates each path to X, indicated by the 5 red arrows.