FUDAN SUMMER SCHOOL INTRODUCTION TO AI

Probabilistic Reasoning and Decision Making

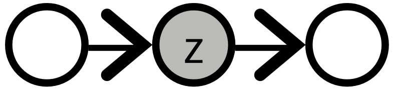
Day 5 – d-separation, variable elimination

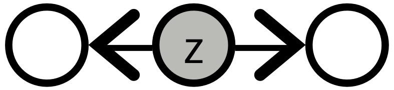
Conditional Independence

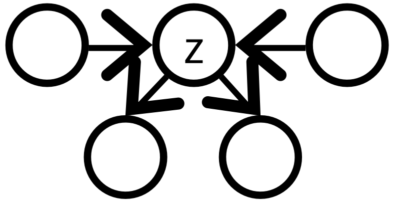
$P(J, M|A) = P(J|A)P(M|A)$
A. True
B. False

<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.

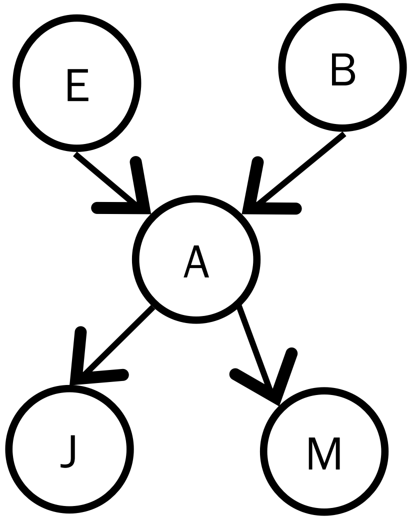
<u>Definition</u>: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

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

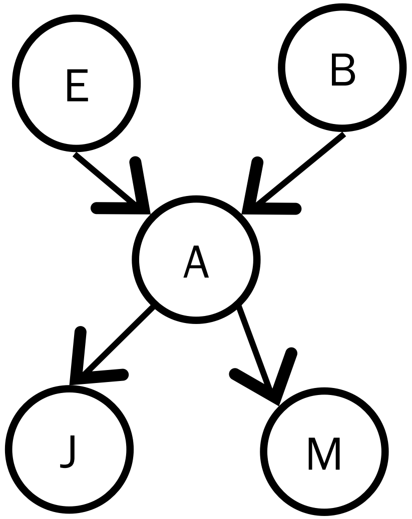
$P(B|E,J) = P(B|J)$

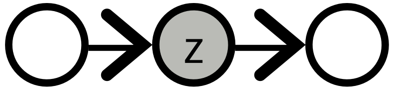
A. True

B. False

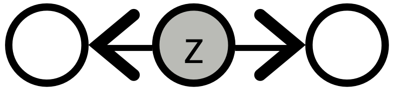
<u>Theorem</u>: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

<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:

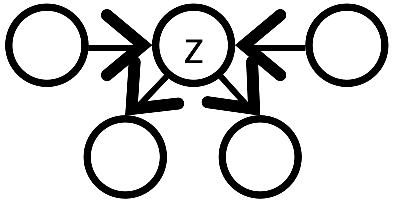


1. 

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

2. 

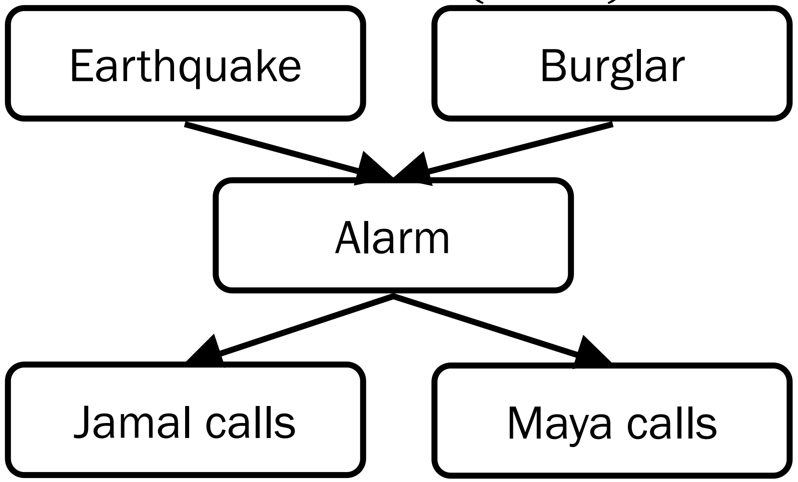
$z \in E$ (observed common explanation)

3. 

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

Conditional Independence

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



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

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

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

$P(B \mid E, J, M) = P(B \mid J, M)$?

A. True

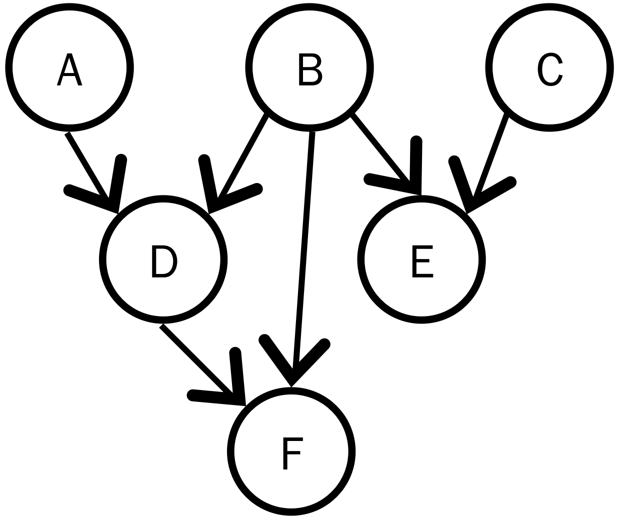
B. False

Conditional Independence

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

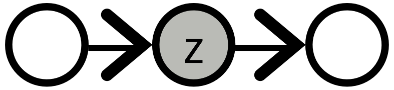
A. True

B. False

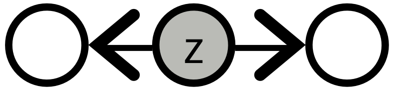


Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

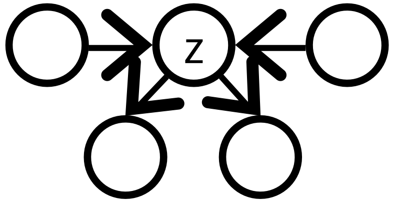
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

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

2. 

$z \in E$ (observed common explanation)

3. 

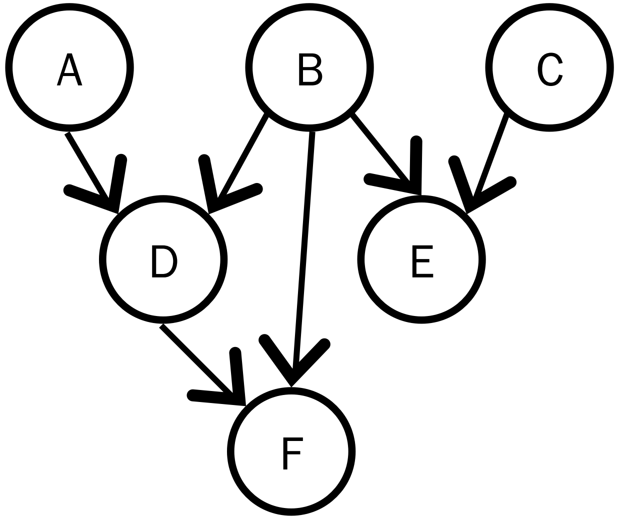
$z \notin E, descendants(z) \notin E$ (no observed common effect)

Conditional Independence

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

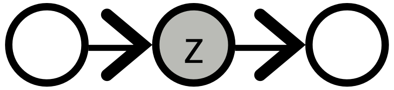
A. True

B. False

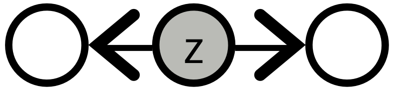


Theorem: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

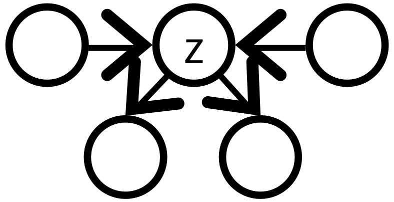
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

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)

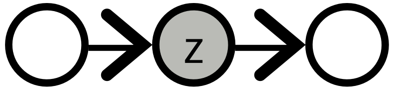
Conditional Independence

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

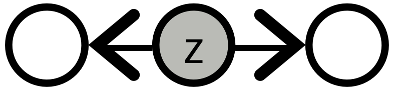
A. X={F}, Y={H}, Ev={A, E}
B. X={F, H}, Y={E}, Ev={A}
C. X={F}, Y={H, E}, Ev={A}
D. X={F, H, E}, Y={A}, Ev={}

<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.

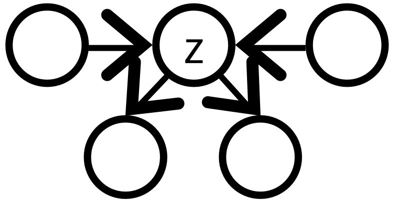
<u>Definition</u>: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

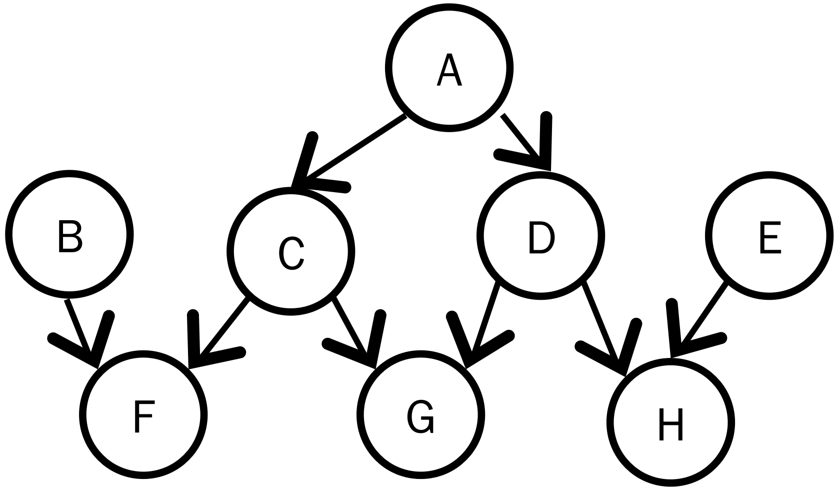
$z \in E$ (observed event in causal chain)

2. 

$z \in E$ (observed common explanation)

3. 

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



Conditional Independence

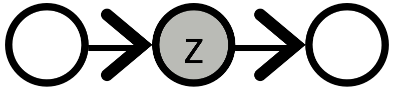
$\boxed{P(F,H|A,E) = P(F|A)P(H|A,E)?}$

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

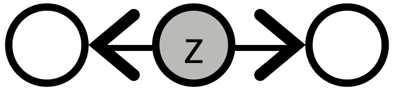
A. True
B. False

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.

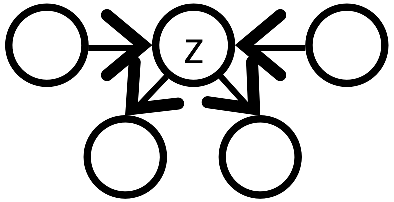
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:

1. 

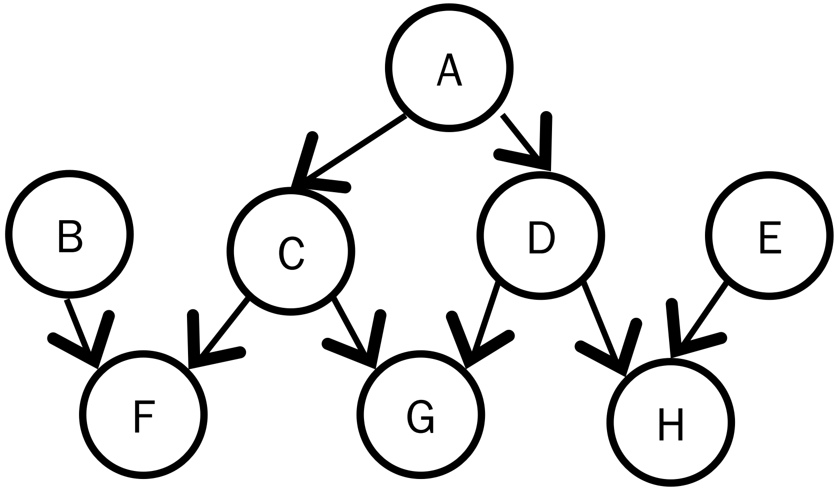
$z \in E$ (observed event in causal chain)

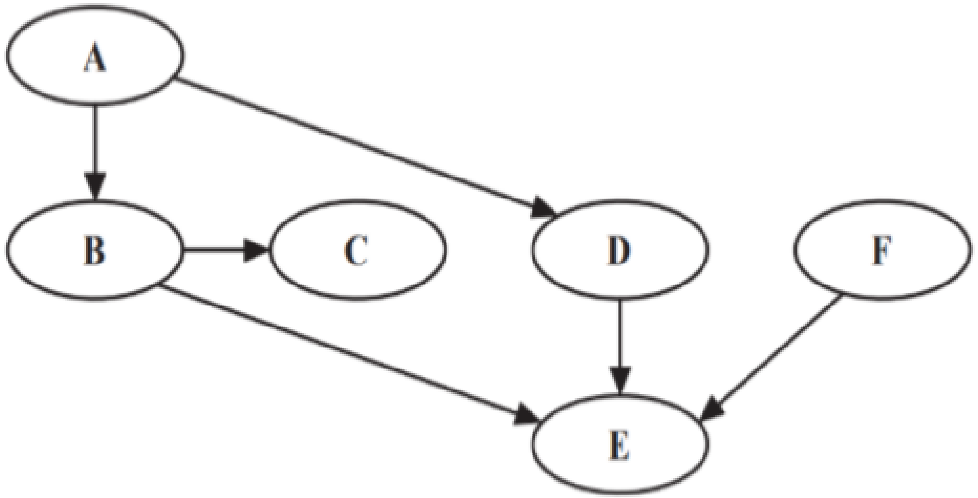
2. 

$z \in E$ (observed common explanation)

3. 

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





(a) Does D d-separate C and F?

(b) Do D and E d-separate C and F?

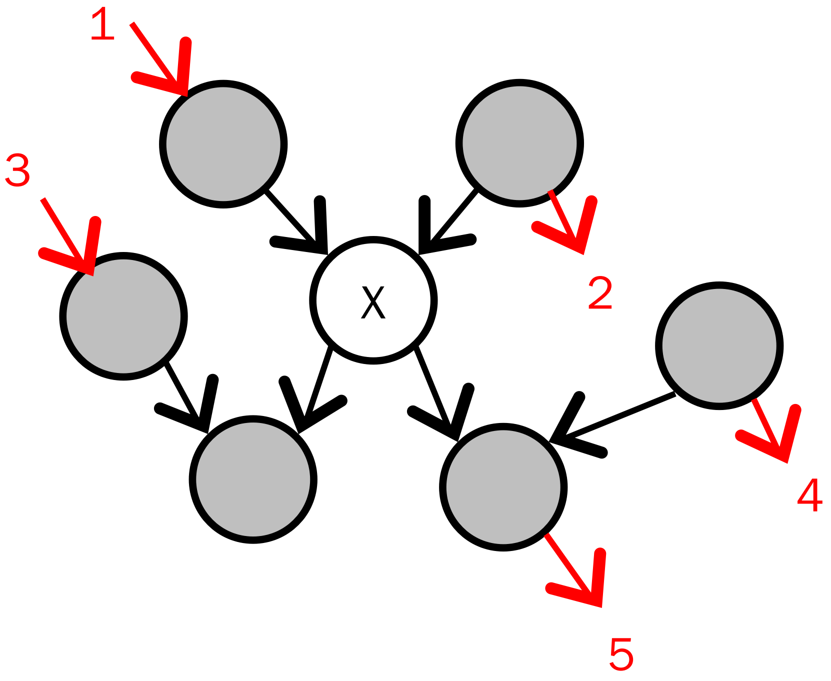
(c) Write down all pairs of nodes which are independent of each other.

(d) Which pairs of nodes are independent of each other given B?

(e) $P(A, F|E) = P(A|E)P(F|E)$?

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.

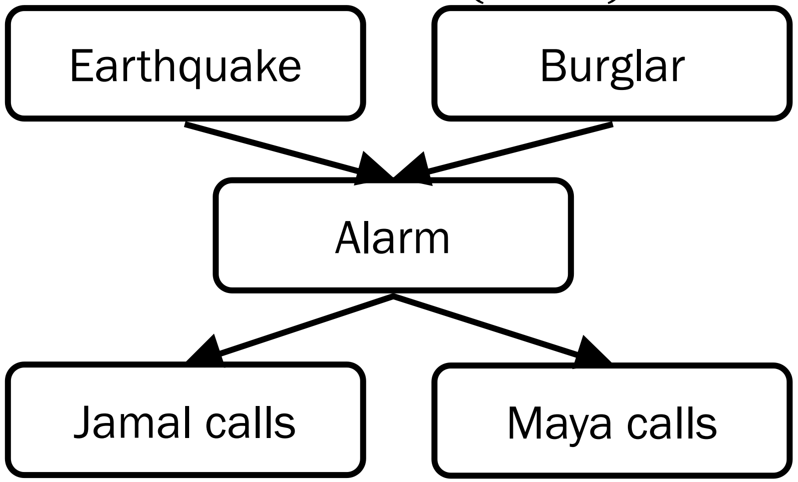


Theorem: $P(X|B_x, Y) = P(X|B_x)$
$if\ Y \notin \{X, B_x\}$

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

Inference in Bayes Nets: General Approach

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



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

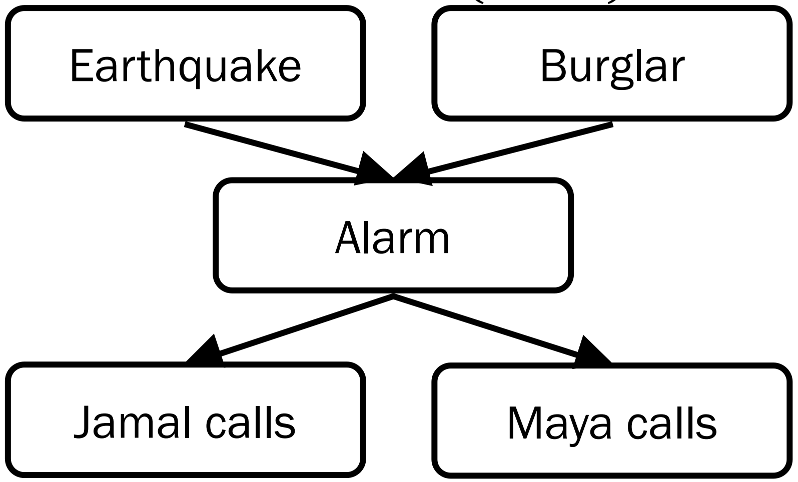
A	$P(J=1\|A)$
0	0.1
1	0.93

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

Compute $P(B\|\,J=1, M=1)$

Inference in Bayes Nets: Enumeration

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



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

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

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

Compute $P(B \mid J = 1, M = 1)$

Using a general approach:

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

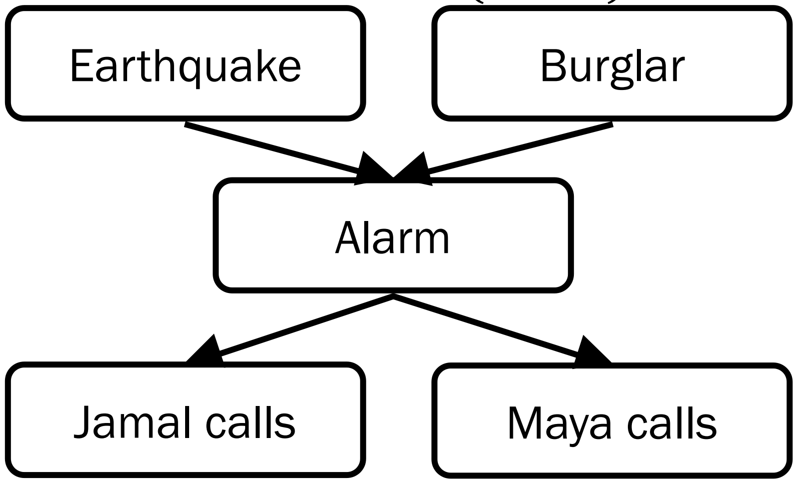
$$= \frac{\sum_a \sum_e P(B, J = 1, M = 1, A = a, E = e)}{\sum_b P(B = b, J = 1, M = 1)}$$

How many operations (additions, multiplications) does it take to compute $P(B = 1, J = 1, M = 1)$?

Inference in Bayes Nets: Enumeration

$P(E = 1) = 0.005$

$P(B = 1) = 0.001$



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

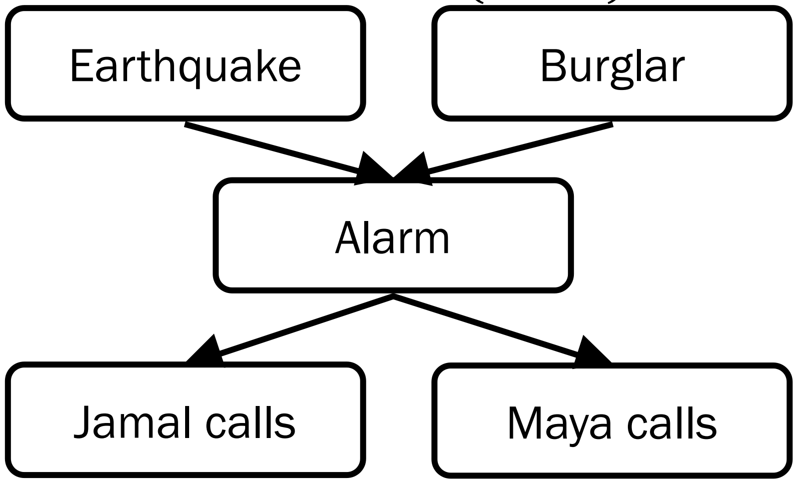
A	$P(J = 1 \mid A)$
0	0.1
1	0.93

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

Compute $P(B = 1, J = 1, M = 1)$ by expanding out:

Inference in Bayes Nets: Enumeration

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



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

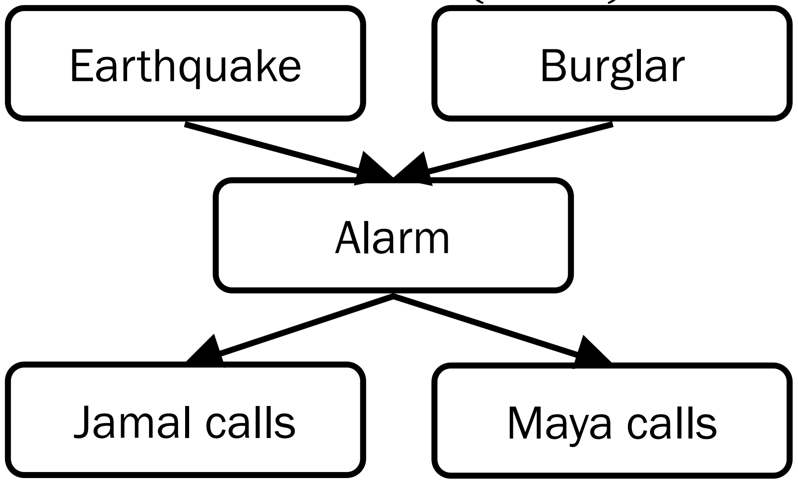
A	$P(M = 1 \mid A)$
0	0.01
1	0.65

How many multiplications and additions does it take to compute $P(B = 1 \mid J = 1, M = 1)$?

A. 16 multiplications, 3 additions
B. 18 multiplications, 4 additions
C. 32 multiplications, 6 additions
D. 32 multiplications, 7 additions
E. I have no idea

Inference in Bayes Nets: Enumeration

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



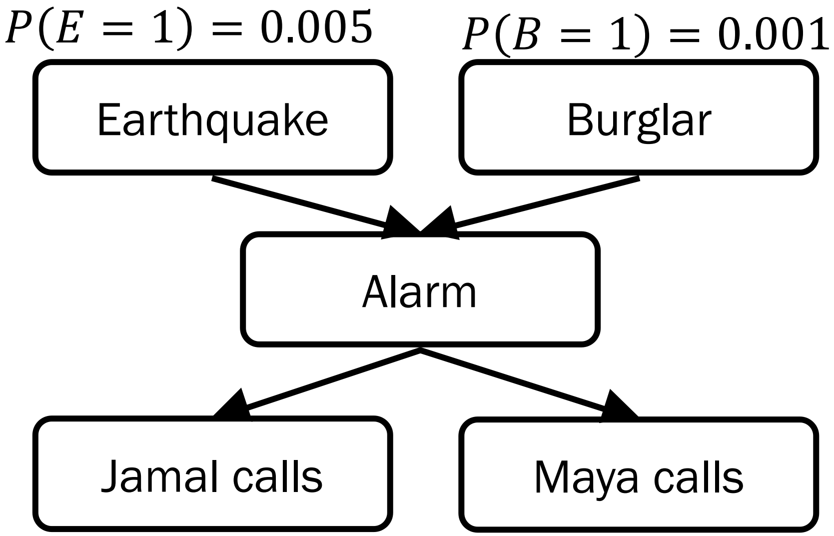
B	E	$P(A=1\|B,E)$
0	0	0.002
0	1	0.35
1	0	0.96
1	1	0.98

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

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

$$
\begin{aligned}
&P(B=1,J=1,M=1)\\
&\quad = P(B=1)P(E=1)P(A=1|E=1,B=1)P(J=1|A=1)P(M=1|A=1)\\
&\quad + P(B=1)P(E=0)P(A=1|E=0,B=1)P(J=1|A=1)P(M=1|A=1)\\
&\quad + P(B=1)P(E=1)P(A=0|E=1,B=1)P(J=1|A=0)P(M=1|A=0)\\
&\quad + P(B=1)P(E=0)P(A=0|E=0,B=1)P(J=1|A=0)P(M=1|A=0)
\end{aligned}
$$

Inference in Bayes Nets: Variable Elimination



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

Eliminate redundant calculations by storing intermediate results in "factors"

A factor is:

Operations on Factors: Multiplication

$f_0(E)$

E	val
0	0.995
1	0.005

$f_1(B)$

B	val
0	0.999
1	0.001

$f_2(A, B, E)$

A	B	E	val
0	0	0	0.998
0	0	1	0.65
0	1	0	0.04
0	1	1	0.02
1	0	0	0.002
1	0	1	0.35
1	1	0	0.96
1	1	1	0.98

$f_3 * f_4$

$f_3(J, A)$

J	A	val
0	0	0.9
0	1	0.07
1	0	0.1
1	1	0.93

$f_4(M, A)$

M	A	val
0	0	0.99
0	1	0.35
1	0	0.01
1	1	0.65

Operations on Factors: Multiplication

$f_0(E)$

E	val
0	0.995
1	0.005

$f_1(B)$

B	val
0	0.999
1	0.001

$f_2(A, B, E)$

A	B	E	val
0	0	0	0.998
0	0	1	0.65
0	1	0	0.04
0	1	1	0.02
1	0	0	0.002
1	0	1	0.35
1	1	0	0.96
1	1	1	0.98

$f_3(J, A)$

J	A	val
0	0	0.9
0	1	0.07
1	0	0.1
1	1	0.93

$f_4(M, A)$

M	A	val
0	0	0.99
0	1	0.35
1	0	0.01
1	1	0.65

How many values (rows) are in the table for the factor $f_2 * f_3$

A. 4
B. 8
C. 16
D. 32

Operations on Factors: Multiplication

$f_0(E)$

E	val
0	0.995
1	0.005

$f_1(B)$

B	val
0	0.999
1	0.001

$f_2(A, B, E)$

A	B	E	val
0	0	0	0.998
0	0	1	0.65
0	1	0	0.04
0	1	1	0.02
1	0	0	0.002
1	0	1	0.35
1	1	0	0.96
1	1	1	0.98

$f_3(J, A)$

J	A	val
0	0	0.9
0	1	0.07
1	0	0.1
1	1	0.93

$f_4(M, A)$

M	A	val
0	0	0.99
0	1	0.35
1	0	0.01
1	1	0.65

How many total multiplications are needed to produce the table $f_2 * f_3$

A. 0
B. 16
C. 32
D. 64

Operations on Factors: Summing out

$f_0(E)$

E	val
0	0.995
1	0.005

$f_1(B)$

B	val
0	0.999
1	0.001

$f_2(A, B, E)$

A	B	E	val
0	0	0	0.998
0	0	1	0.65
0	1	0	0.04
0	1	1	0.02
1	0	0	0.002
1	0	1	0.35
1	1	0	0.96
1	1	1	0.98

$$\sum_e f_2$$

$f_3(J, A)$

J	A	val
0	0	0.9
0	1	0.07
1	0	0.1
1	1	0.93

$f_4(M, A)$

M	A	val
0	0	0.99
0	1	0.35
1	0	0.01
1	1	0.65

Operations on Factors: Conditioning

$f_0(E)$

E	val
0	0.995
1	0.005

$f_1(B)$

B	val
0	0.999
1	0.001

$f_2(A, B, E)$

A	B	E	val
0	0	0	0.998
0	0	1	0.65
0	1	0	0.04
0	1	1	0.02
1	0	0	0.002
1	0	1	0.35
1	1	0	0.96
1	1	1	0.98

$f_2(A, B = 1, E)$

$f_3(J, A)$

J	A	val
0	0	0.9
0	1	0.07
1	0	0.1
1	1	0.93

$f_4(M, A)$

M	A	val
0	0	0.99
0	1	0.35
1	0	0.01
1	1	0.65