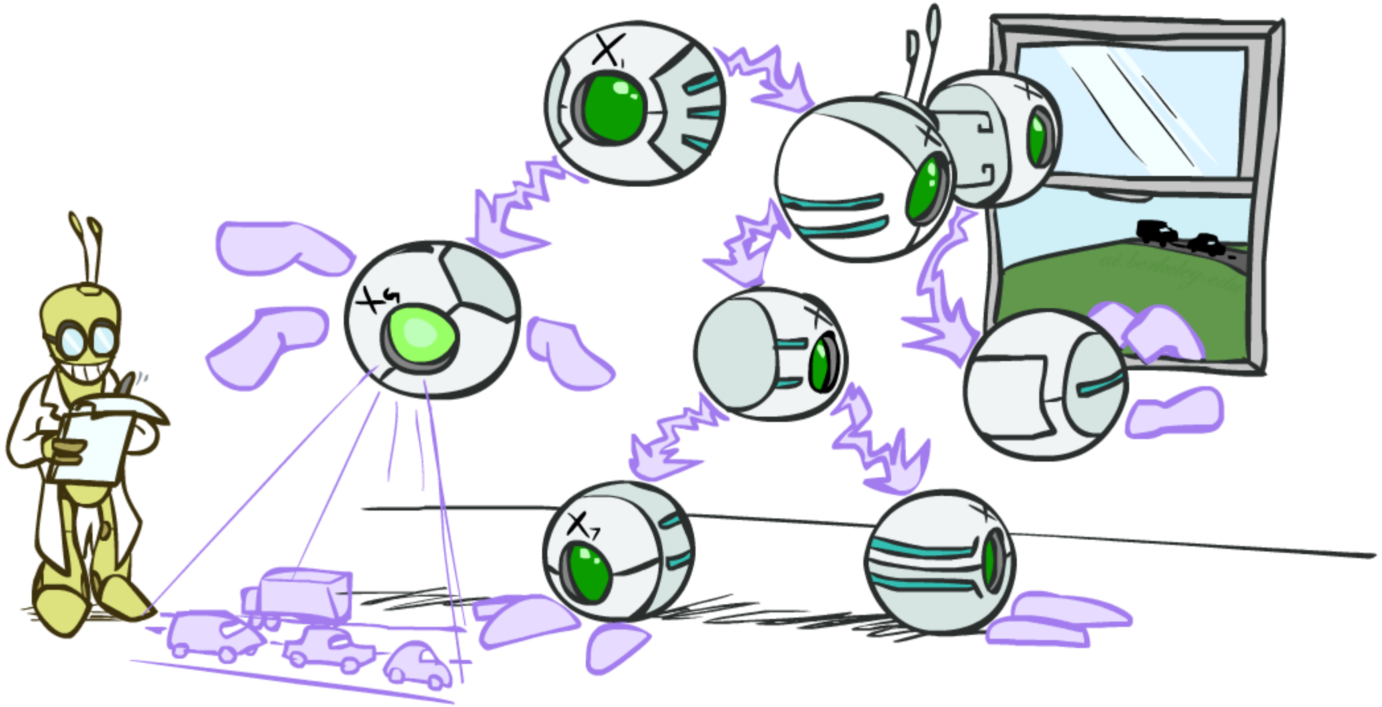
AI: Representation and Problem Solving

Bayes Nets Inference



Instructor: Pat Virtue

Slide credits: CMU AI and http://ai.berkeley.edu

Bayes Nets

✓ Part I: Representation and Independence

Part II: Exact inference

- Enumeration (always exponential complexity)
- Variable elimination (worst-case exponential complexity, often better)
- Inference is NP-hard in general

Part III: Approximate Inference

Bayes Nets in the Wild

[Wheel of Fortune]

Bayes Nets in the Wild

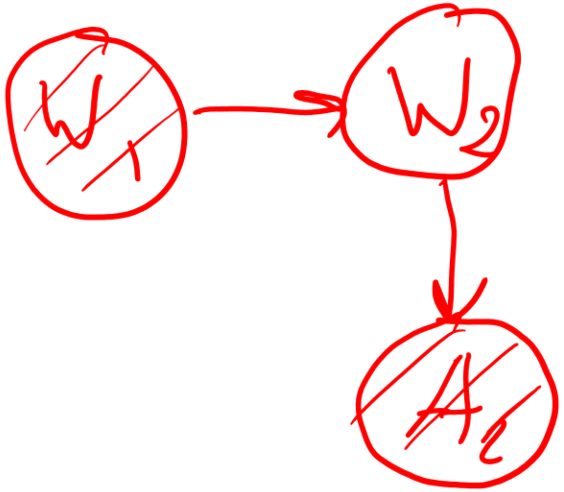
Example: Speech Recognition

“artificial ……….”

Find most probable next word given “artificial” and the audio for second word.

$$\hat{w}_2 = \underset{w_2 \in Vocab}{\arg\max} \; P(w_2 \mid w_1, a_2)$$

Bayes Nets in the Wild



Example: Speech Recognition

"artificial ………."

Find most probable next word given "artificial" and the audio for second word.

Which second word gives the highest probability?

Break down problem

n-gram probability * audio probability

LM

$P(\textbf{limb} \mid \text{artificial}, \text{audio})$ → $P(\textbf{limb} \mid \text{artificial})$ * $P(\text{audio} \mid \textbf{limb})$

$P(\textbf{intelligence} \mid \text{artificial}, \text{audio})$ $P(\textbf{intelligence} \mid \text{artificial})$ * $P(\text{audio} \mid \textbf{intelligence})$

$P(\textbf{flavoring} \mid \text{artificial}, \text{audio})$ $P(\textbf{flavoring} \mid \text{artificial})$ * $P(\text{audio} \mid \textbf{flavoring})$

Bayes Nets in the Wild

$$second^* = \text{argmax}_{second}\, P(\boldsymbol{second} \mid artificial, audio)$$

def cond prob

$$= \text{argmax}_{second} \frac{P(\boldsymbol{second}, artificial, audio)}{P(artificial, audio)}$$

$$= \text{argmax}_{second}\, P(\boldsymbol{second}, artificial, audio)$$

chain

$$= \text{argmax}_{second}\, P(artificial)\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid artificial, \boldsymbol{second})$$

$$= \text{argmax}_{second}\, P(artificial)\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid \boldsymbol{second})$$

Bayes

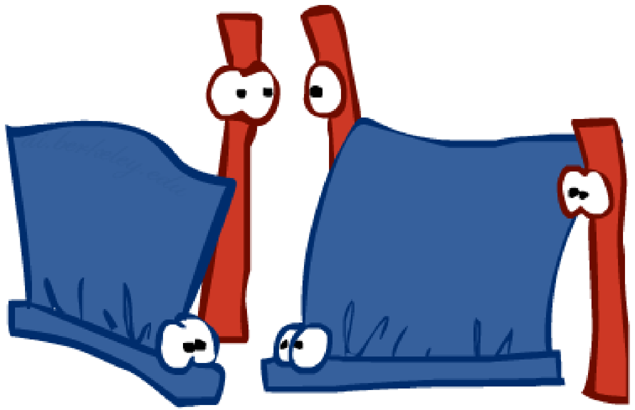
$$= \text{argmax}_{second}\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid \boldsymbol{second})$$

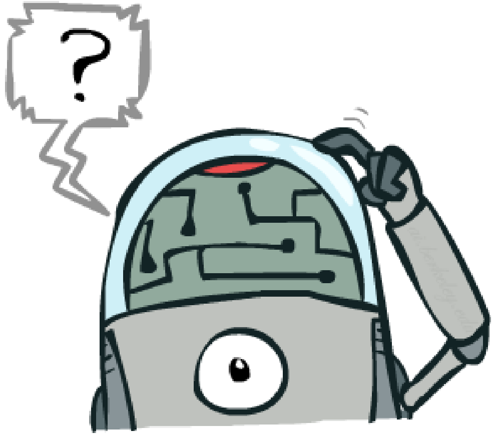
n-gram probability * audio probability

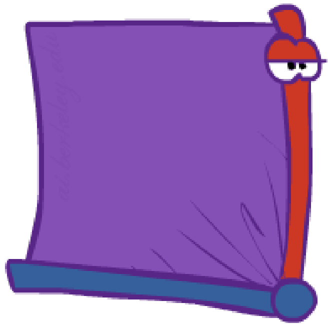
Inference

Inference: calculating some useful quantity from a probability model (joint probability distribution)

- Examples:
  - Posterior marginal probability
    - $P(Q|e_1,..,e_k)$
    - e.g., what disease might I have?
  - Most likely explanation:
    - $\text{argmax}_{q,r,s}\, P(Q=q,R=r,S=s|e_1,..,e_k)$
    - e.g., what was just said?







Inference Overview

Given random variables Q, H, E (query, hidden, evidence)

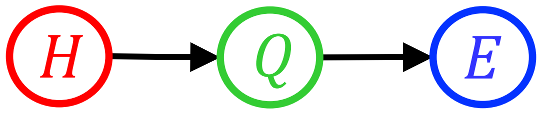
Enumeration

We know how to do inference on a joint distribution

$$P(q|e) = \alpha\, P(q,e)$$
$$= \alpha \textstyle\sum_{h \in \{h_1,h_2\}} P(q,h,e)$$

$$\alpha = \frac{1}{P(e)} = \frac{1}{\sum_{hq} P(hqe)}$$

We know Bayes nets can break down joint in to CPT factors



$$P(q|e) = \alpha \textstyle\sum_{h \in \{h_1,h_2\}} P(h)\, P(q|h)\, P(e|q)$$
$$= \alpha\, [P(h_1)\, P(q|h_1)\, P(e|q) + P(h_2)\, P(q|h_2)\, P(e|q)]$$

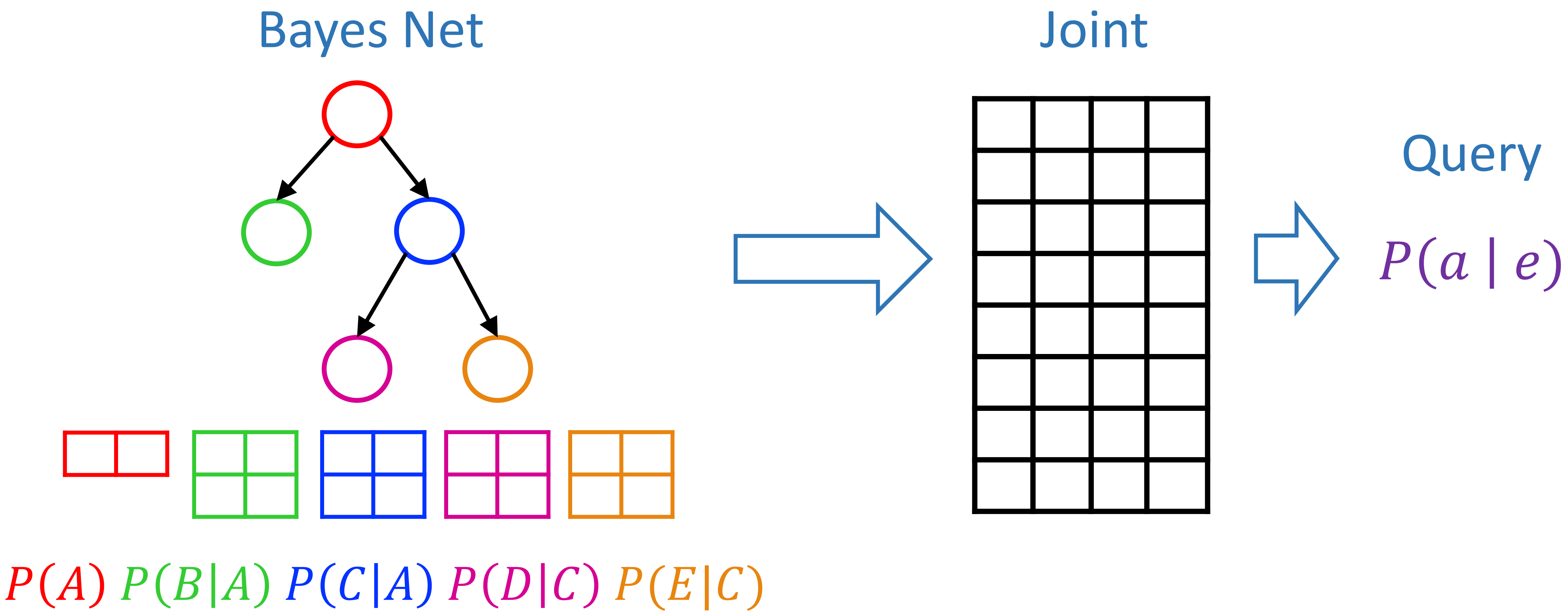
Variable Elimination

But we can be more efficient

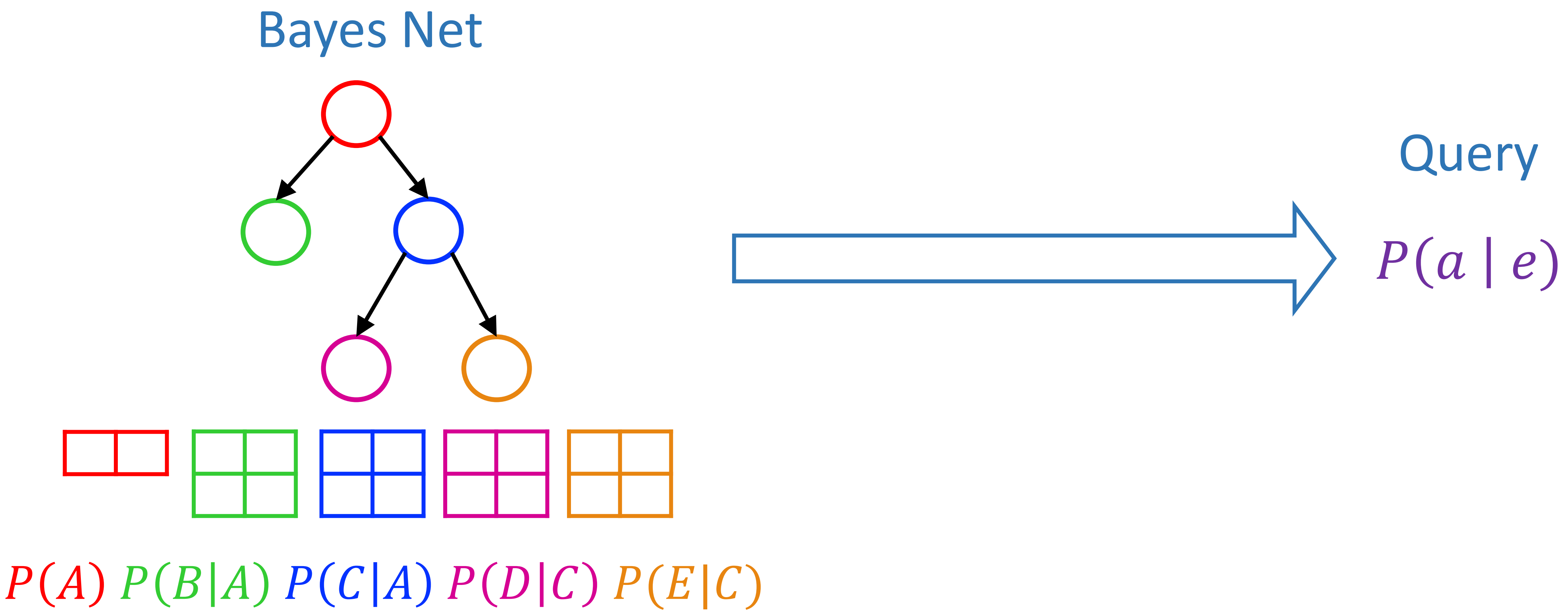
$$P(q|e) = \alpha\, P(e|q) \textstyle\sum_{h \in \{h_1,h_2\}} P(h) P(q|h)$$
$$= \alpha\, P(e|q)\, [P(h_1)P(q|h_1) + P(h_2)P(q|h_2)]$$
$$= \alpha\, P(e|q)\, P(q)$$

Now just extend to larger Bayes nets and a variety of queries

Answer Any Query from Bayes Net



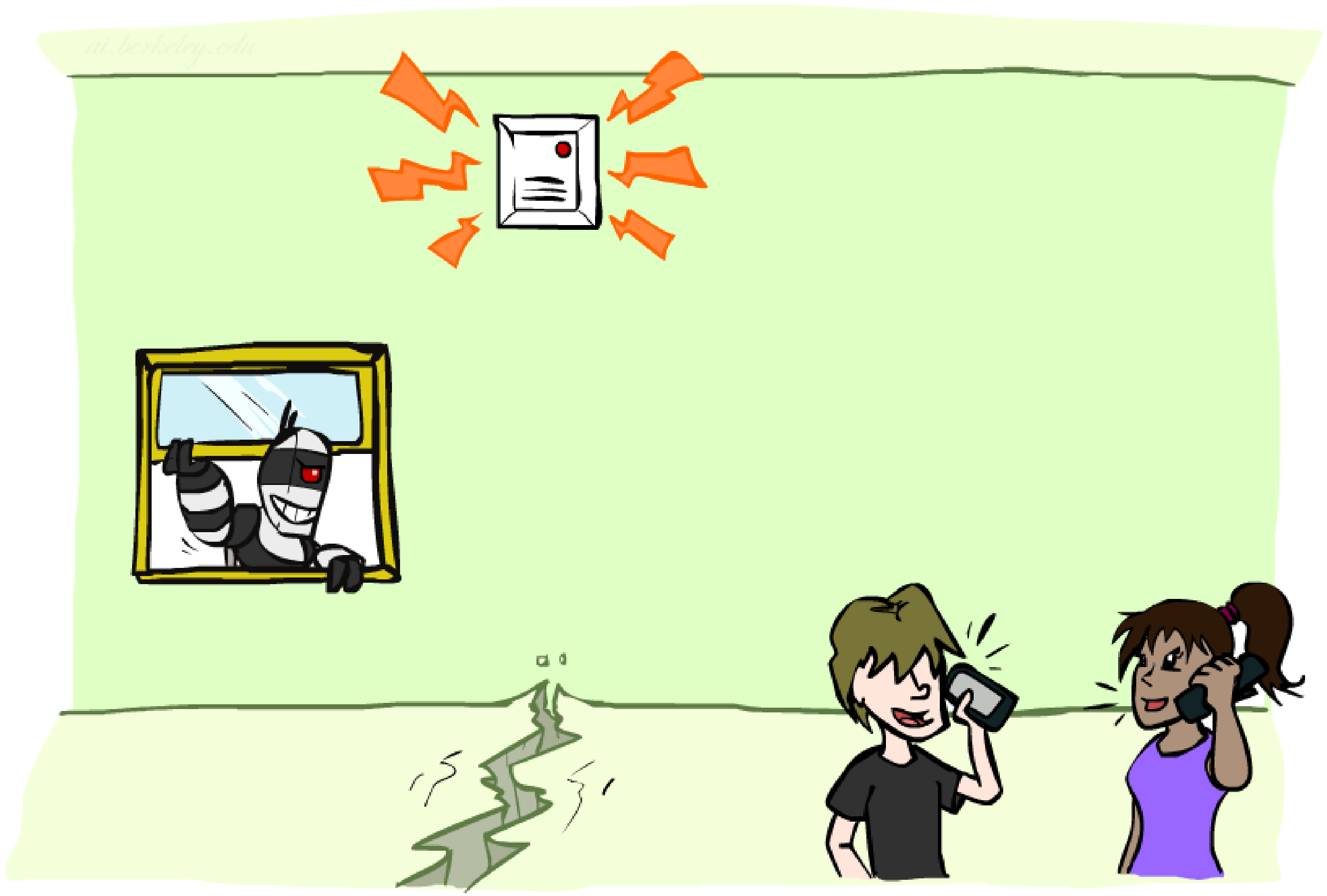
Next: Answer Any Query from Bayes Net



Example: Alarm Network

Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



Example: Alarm Network

Joint distribution factorization example

Generic chain rule

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

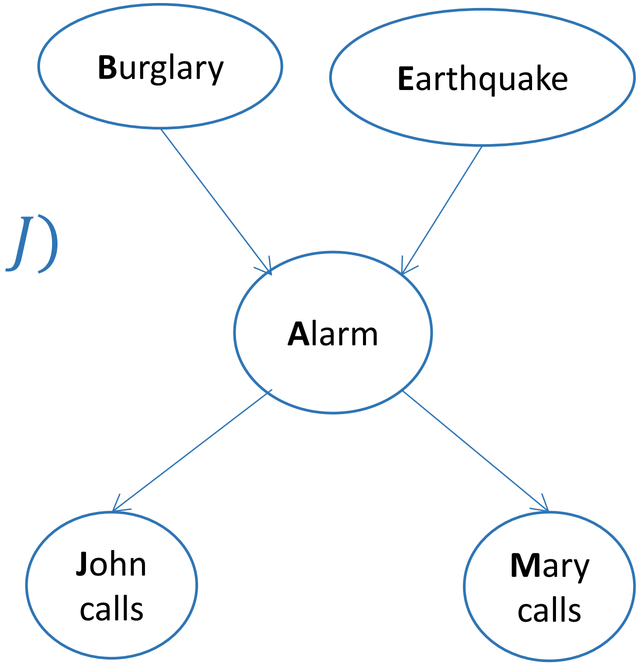
$P(B,E,A,J,M) = P(B)\, P(E|B)\, P(A|B,E)\, P(J|B,E,A)\, P(M|B,E,A,J)$

$f_1(B)\; f_2(E)\; f_3(A,B,E)$

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

Bayes nets

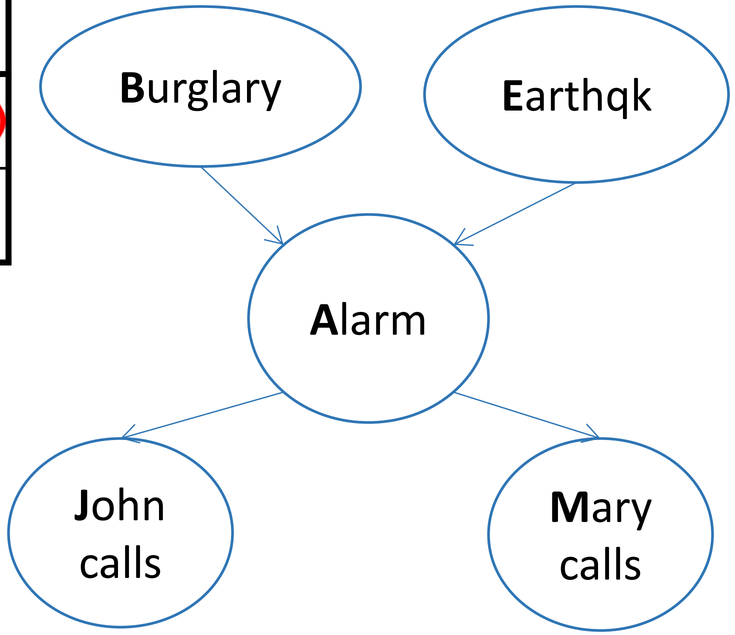
- $P(X_1 \ldots X_2) = \prod_i P(X_i | Parents(X_i))$



Example: Alarm Network

$P(+b, -e, -a, -j, -m) =$

B	P(B)
+b	0.001
-b	0.999

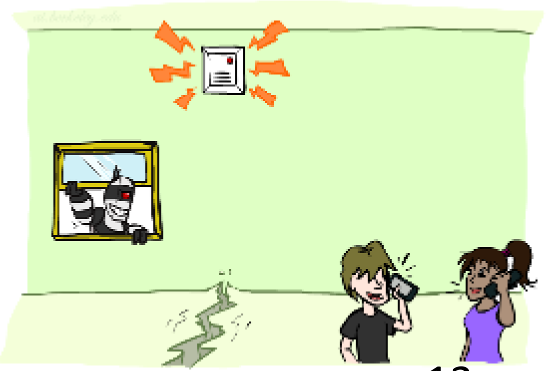


E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999



Example: Alarm Network

$$P(+b,-e,-a,-j,-m) = P(+b) * P(-e) * P(-a|+b,-e) * P(-j|-a) * P(-m|-a)$$
$$= 0.001 * 0.998 * 0.06 * 0.95 * 0.99$$

$$P(+b,-e,-a,-j,-m) = P(-e) * P(-a|+b,-e) * P(+b) * P(-j|-a) * P(-m|-a)$$
$$= 0.998 * 0.06 * 0.001 * 0.95 * 0.99$$

$$P(+b,-e,-a,-j,-m) = P(-e) * P(-a|+b,-e) * P(+b)\, P(-j|-a) * P(-m|-a)$$
$$= 0.998 * 0.06 * 0.0095 * 0.99$$

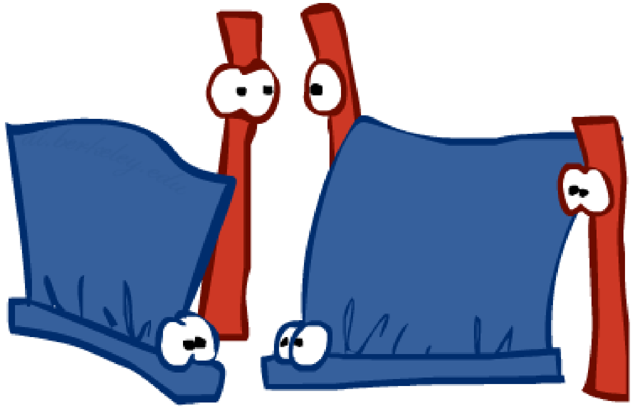
$f_6(+b,-j,-a)$

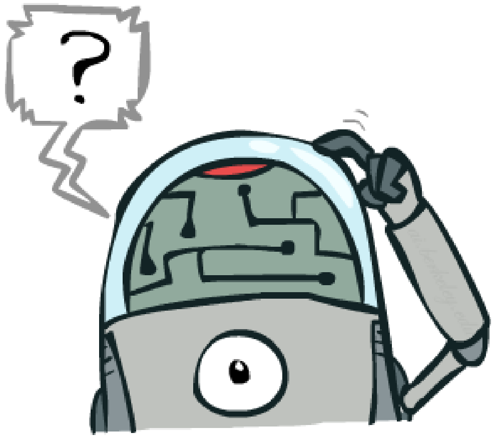
- Multiplication order can change (commutativity)
- Multiplication pairs don't have to make sense (associativity)

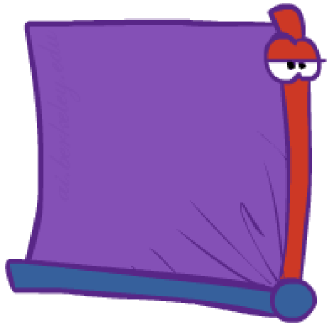
Factor Tables

$$P(+b,-e,-a,-j,-m) = P(+b) * P(-e) * P(-a|+b,-e) * P(-j|-a) * P(-m|-a)$$
$$= 0.001 * 0.998 * 0.06 * 0.95 * 0.99$$

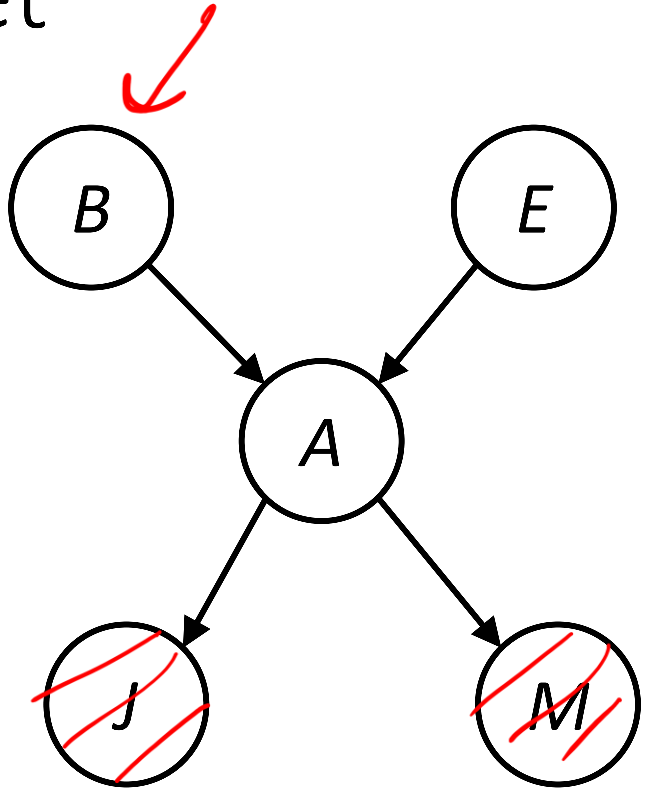
$$P(B,E,A,J,M) = P(B) * P(E) * P(A|B,E) * P(J|A) * P(M|A)$$







Inference by Enumeration in Bayes Net



Reminder of inference by enumeration:

- Any probability of interest can be computed by summing entries from the joint distribution
- Entries from the joint distribution can be obtained from a BN by multiplying the corresponding conditional probabilities

$$P(B \mid j, m) = \alpha P(B, j, m)$$
$$= \alpha \sum_{e,a} P(B, e, a, j, m)$$
$$= \alpha \sum_{e,a} P(B)\, P(e)\, P(a \mid B,e)\, P(j \mid a)\, P(m \mid a)$$

So inference in Bayes nets means computing sums of products of numbers: sounds easy!!

Problem: sums of ***exponentially many*** products!

Can we do better?

Consider

- $x_1y_1z_1 + x_1y_1z_2 + x_1y_2z_1 + x_1y_2z_2 + x_2y_1z_1 + x_2y_1z_2 + x_2y_2z_1 + x_2y_2z_2$
- 16 multiplies, 7 adds
- Lots of repeated subexpressions!

Rewrite as

- $(x_1 + x_2)(y_1 + y_2)(z_1 + z_2)$
- 2 multiplies, 3 adds

$$\sum_e \sum_a P(B)\, P(e)\, P(a \mid B, e)\, P(j \mid a)\, P(m \mid a)$$

$$\begin{aligned} = \; & P(B)\, P(+e)\, P(+a \mid B, +e)\, P(j \mid +a)\, P(m \mid +a) \\ + \; & P(B)\, P(-e)\, P(+a \mid B, -e)\, P(j \mid +a)\, P(m \mid +a) \\ + \; & P(B)\, P(+e)\, P(-a \mid B, +e)\, P(j \mid -a)\, P(m \mid -a) \\ + \; & P(B)\, P(-e)\, P(-a \mid B, -e)\, P(j \mid -a)\, P(m \mid -a) \end{aligned}$$

- Lots of repeated subexpressions!

Variable elimination: The basic ideas

Move summations inwards as far as possible

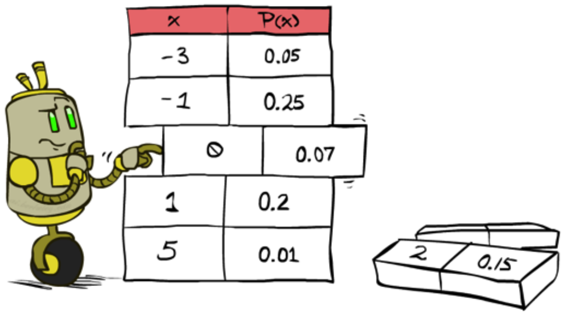
$$P(B \mid j, m) = \alpha \sum_e \sum_a P(B, e, a, j, m)$$

$$= \alpha \sum_e \sum_a P(j|a)\, P(e)\, P(m|a)\, P(a|B, e)\, P(B)$$

$$= \alpha\, P(B) \sum_e P(e) \sum_a P(j|a)\;\; P(m|a)\, P(a|B, e)$$

Variable Elimination

- Query: $P(Q_1 ,\ldots, Q_m \mid E_1=e_1,\ldots, E_k=e_k)$

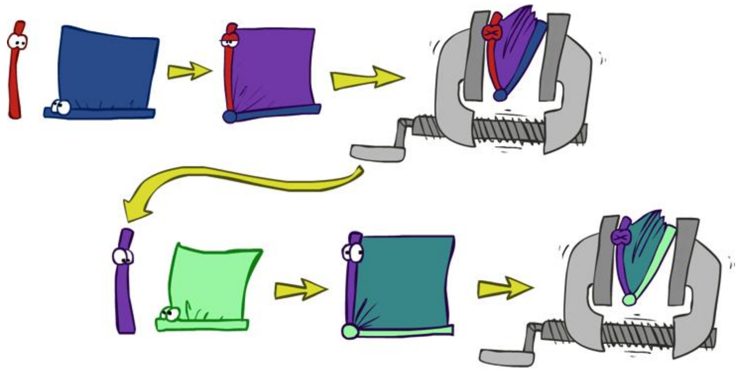


Start with initial factors:

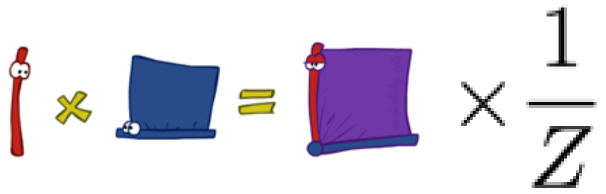
- Local CPTs (but instantiated by evidence)

While there are still hidden variables (not Q_i or evidence):

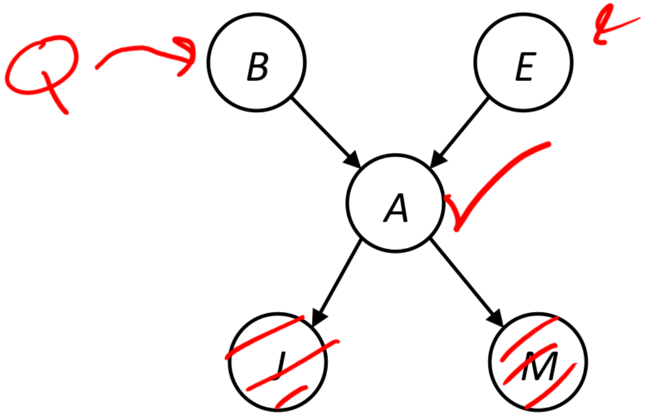
- Pick a hidden variable H
- Join all factors mentioning H
- Eliminate (sum out) H



Join all remaining factors and normalize



Example



Query $P(B \mid j,m)$

$P(B)$	$P(E)$	$P(A \mid B,E)$	$P(j \mid A)$	$P(m \mid A)$

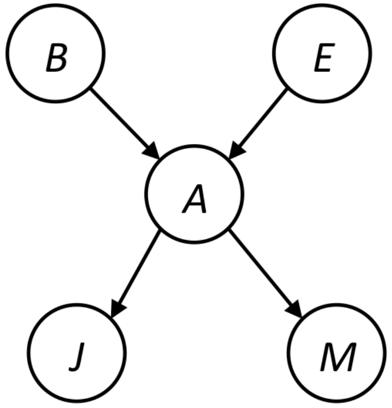
Choose A

$P(A \mid B,E)$
$P(j \mid A)$
$P(m \mid A)$

$\times \Rightarrow P(A,j,m \mid B,E) \quad \Sigma \Rightarrow P(j,m \mid B,E)$

$P(B)$	$P(E)$	$P(j,m \mid B,E)$

Example



P(B)	~~P(E)~~	~~P(j,m|B,E)~~

Choose E

P(E)
P(j,m|B,E)

× → f_8: P(E,j,m|B) → Σ → f_9: P(j,m|B)

P(B)	P(j,m|B)

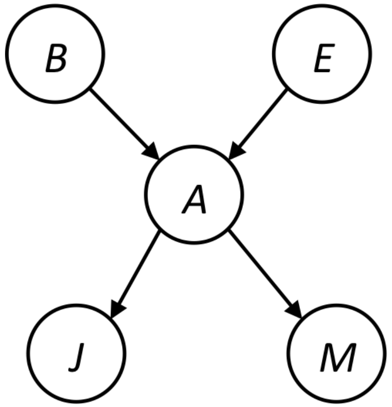
Finish with B

P(B)
P(j,m|B)

× → f_{10}: P(j,m,B) → Normalize → P(B | j,m)

$P(B \mid j, m)$

See video: Math for Variable Elimination



$$P(B \mid j, m) = \alpha P(B, j, m)$$

$$= \alpha \sum_a \sum_e P(B, e, a, j, m)$$

$$= \alpha \sum_a \sum_e f_1(B) f_2(e) f_3(a, B, e) f_4(j, a) f_5(m, a)$$

$$= \alpha \sum_e f_1(B) f_2(e) \sum_a f_3(a, B, e) f_4(j, a) f_5(m, a)$$

$$= \alpha \sum_e f_1(B) f_2(e) \sum_a f_6(B, e, a, j, m)$$

$$= \alpha \sum_e f_1(B) f_2(e) f_7(B, e, j, m)$$

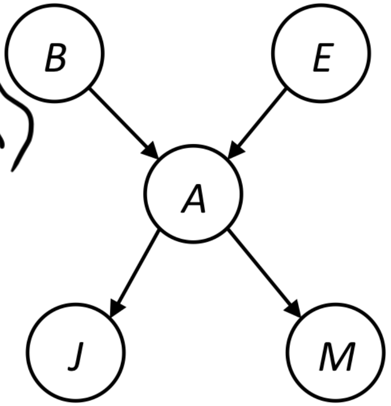
$$= \alpha f_1(B) \sum_e f_8(B, e, j, m)$$

$$= \alpha f_1(B) f_9(B, j, m)$$

$$= \alpha f_{10}(B, j, m)$$

$$= P(B \mid j, m)$$

$$\alpha = \frac{1}{P(j, m)} = \frac{1}{\sum_b P(b, j, m)}$$

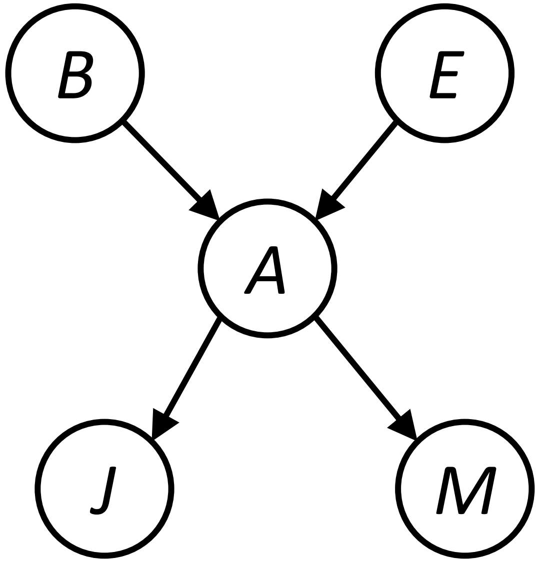


$$\alpha = \frac{1}{P(j, m)} = \frac{1}{\sum_b f_{10}(b, j, m)}$$

Example (with numbers!)

Start with a Bayes net, the associated CPTs, and a query, $P(B \mid +j, +m)$

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

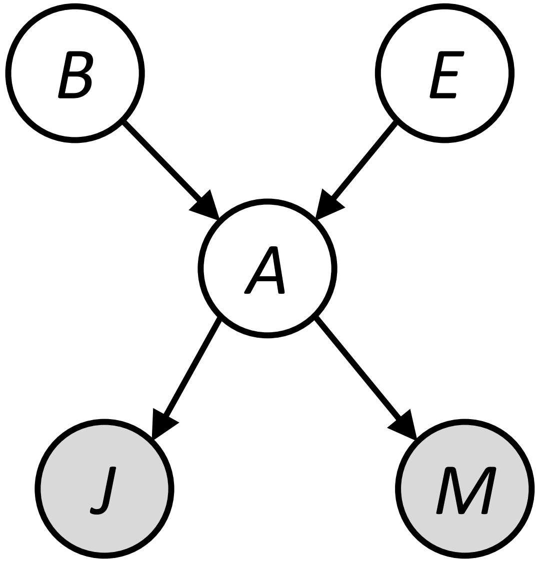
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 1: Remove any entries in tables that don't match the evidence, $+j, +m$

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
-a	+j	0.05

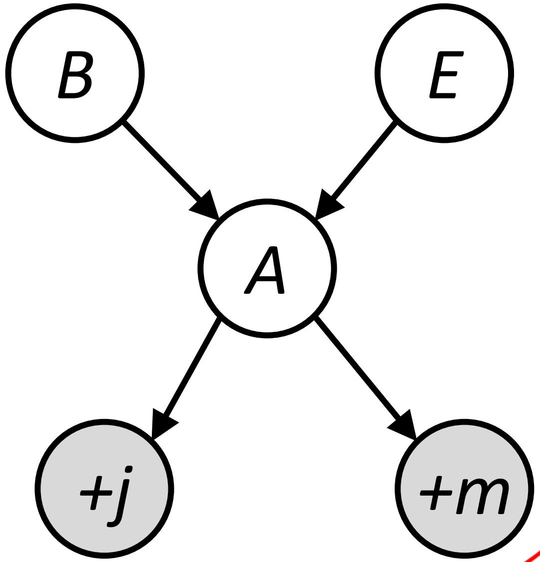
A	M	P(M|A)
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 1: Remove any entries in tables that don't match the evidence, $+j, +m$

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	+j	P(+j|A)
+a	+j	0.9
-a	+j	0.05

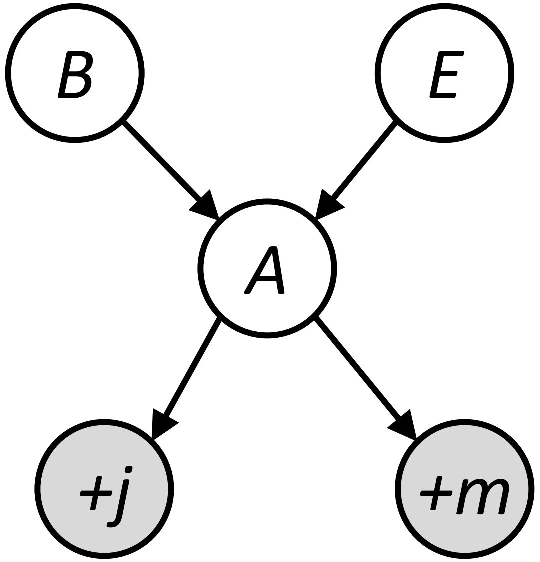
A	+m	P(+m|A)
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 2: Think each table as just a factor table rather than probabilities

B	$f_1(B)$
+b	0.001
-b	0.999



E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A, B, E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 2: Think each table as just a factor table rather than probabilities

B	$f_1(B)$
+b	0.001
-b	0.999

E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A,B,E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	+j	$f_4(+j,A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m,A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E (we chose E first this time)

i: Collect all factors tables associated with E

B	$f_1(B)$
+b	0.001
-b	0.999

E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A,B,E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	+j	$f_4(+j,A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m,A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A,B,E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

B	E	A	$f_4(A,B,E)$
+b	+e	+a	
+b	+e	-a	
+b	-e	+a	
+b	-e	-a	
-b	+e	+a	
-b	+e	-a	
-b	-e	+a	
-b	-e	-a	

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A,B,E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

B	E	A	$f_4(A,B,E)$
+b	+e	+a	$0.002 \cdot 0.95$
+b	+e	-a	$0.002 \cdot 0.05$
+b	-e	+a	$0.998 \cdot 0.94$
+b	-e	-a	$0.998 \cdot 0.06$
-b	+e	+a	$0.002 \cdot 0.29$
-b	+e	-a	$0.002 \cdot 0.71$
-b	-e	+a	$0.998 \cdot 0.001$
-b	-e	-a	$0.998 \cdot 0.999$

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

E	$f_2(E)$
+e	0.002
-e	0.998

B	E	A	$f_3(A, B, E)$
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

B	E	A	$f_6(A, B, E)$
+b	+e	+a	0.001900
+b	+e	-a	0.000100
+b	-e	+a	0.938120
+b	-e	-a	0.059880
-b	+e	+a	0.000580
-b	+e	-a	0.001420
-b	-e	+a	0.000998
-b	-e	-a	0.997002

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

iii: Create another new factor by summing over E

B	A	$f_7(A, B)$
+b	+a	
+b	-a	
-b	+a	
-b	-a	

B	E	A	$f_6(A, B, E)$
+b	+e	+a	0.001900
+b	+e	-a	0.000100
+b	-e	+a	0.938120
+b	-e	-a	0.059880
-b	+e	+a	0.000580
-b	+e	-a	0.001420
-b	-e	+a	0.000998
-b	-e	-a	0.997002

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

iii: Create another new factor by summing over E

B	A	$f_7(A, B)$
+b	+a	0.001900 + 0.938120
+b	-a	0.000100 + 0.059880
-b	+a	0.000580 + 0.000998
-b	-a	0.001420 + 0.997002

B	E	A	$f_6(A, B, E)$
+b	+e	+a	0.001900
+b	+e	-a	0.000100
+b	-e	+a	0.938120
+b	-e	-a	0.059880
-b	+e	+a	0.000580
-b	+e	-a	0.001420
-b	-e	+a	0.000998
-b	-e	-a	0.997002

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E

i: Collect all factors tables associated with E

ii: Create new factor table by multiplying them together

iii: Create another new factor by summing over E

B	A	$f_7(A,B)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

B	E	A	$f_6(A,B,E)$
+b	+e	+a	0.001900
+b	+e	-a	0.000100
+b	-e	+a	0.938120
+b	-e	-a	0.059880
-b	+e	+a	0.000580
-b	+e	-a	0.001420
-b	-e	+a	0.000998
-b	-e	-a	0.997002

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(E): Eliminate hidden variable E (and return to list of tables)

B	$f_1(B)$
+b	0.001
-b	0.999

B	A	$f_7(A,B)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j,A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m,A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

B	$f_1(B)$
+b	0.001
-b	0.999

B	A	$f_7(A, B)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

B	$f_1(B)$
+b	0.001
-b	0.999

B	A	$f_7(A, B)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

ii: Create new factor table by multiplying them together

B	A	$f_7(B, A)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

B	A	+j	+m	$f_8(B, A, +j, +m)$
+b	+a	+j	+m	
+b	-a	+j	+m	
-b	+a	+j	+m	
-b	-a	+j	+m	

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

ii: Create new factor table by multiplying them together

B	A	$f_7(B, A)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

B	A	+j	+m	$f_8(B, A, +j, +m)$
+b	+a	+j	+m	$0.940020 \cdot 0.9 \cdot 0.7$
+b	-a	+j	+m	$0.059980 \cdot 0.05 \cdot 0.01$
-b	+a	+j	+m	$0.001578 \cdot 0.9 \cdot 0.7$
-b	-a	+j	+m	$0.998422 \cdot 0.05 \cdot 0.01$

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

ii: Create new factor table by multiplying them together

B	A	$f_7(B, A)$
+b	+a	0.940020
+b	-a	0.059980
-b	+a	0.001578
-b	-a	0.998422

A	+j	$f_4(+j, A)$
+a	+j	0.9
-a	+j	0.05

A	+m	$f_5(+m, A)$
+a	+m	0.7
-a	+m	0.01

B	A	+j	+m	$f_8(B, A, +j, +m)$
+b	+a	+j	+m	0.592213
+b	-a	+j	+m	0.000030
-b	+a	+j	+m	0.000994
-b	-a	+j	+m	0.000499

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

ii: Create new factor table by multiplying them together

iii: Create another new factor by summing over A

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	
-b	+j	+m	

B	A	+j	+m	$f_8(B, A, +j, +m)$
+b	+a	+j	+m	0.592213
+b	-a	+j	+m	0.000030
-b	+a	+j	+m	0.000994
-b	-a	+j	+m	0.000499

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A

i: Collect all factors tables associated with A

ii: Create new factor table by multiplying them together

iii: Create another new factor by summing over A

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	0.592243
-b	+j	+m	0.001493

B	A	+j	+m	$f_8(B, A, +j, +m)$
+b	+a	+j	+m	0.592213
+b	-a	+j	+m	0.000030
-b	+a	+j	+m	0.000994
-b	-a	+j	+m	0.000499

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 3(A): Eliminate hidden variable A (and return to list of tables)

B	$f_1(B)$
+b	0.001
-b	0.999

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	0.592243
-b	+j	+m	0.001493

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 4: Multiply remaining factor tables

B	$f_1(B)$
+b	0.001
-b	0.999

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	0.592243
-b	+j	+m	0.001493

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	
-b	+j	+m	

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 4: Multiply remaining factor tables

B	$f_1(B)$
+b	0.001
-b	0.999

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	0.592243
-b	+j	+m	0.001493

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	$0.001 \cdot 0.592243$
-b	+j	+m	$0.999 \cdot 0.001493$

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 4: Multiply remaining factor tables

B	$f_1(B)$
+b	0.001
-b	0.999

B	+j	+m	$f_8(B, +j, +m)$
+b	+j	+m	0.592243
-b	+j	+m	0.001493

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	0.000592
-b	+j	+m	0.001492

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 5: Normalize

B	+j	+m	$P(B \mid +j, +m)$
+b	+j	+m	
-b	+j	+m	

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	0.000592
-b	+j	+m	0.001492

+j	+m	$Z = f_{10}(+j, +m)$
+j	+m	0.000592+0.001492

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 5: Normalize

$\alpha = \frac{1}{Z}$

B	+j	+m	$P(B \mid +j, +m)$
+b	+j	+m	
-b	+j	+m	

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	0.000592
-b	+j	+m	0.001492

+j	+m	$Z = f_{10}(+j, +m)$
+j	+m	0.002084

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 5: Normalize

B	+j	+m	$P(B \mid +j, +m)$
+b	+j	+m	0.000592/0.002084
-b	+j	+m	0.001492/0.002084

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	0.000592
-b	+j	+m	0.001492

+j	+m	$Z = f_{10}(+j, +m)$
+j	+m	0.002084

Example (with numbers!)

Query: $P(B \mid +j, +m)$

Step 5: Normalize

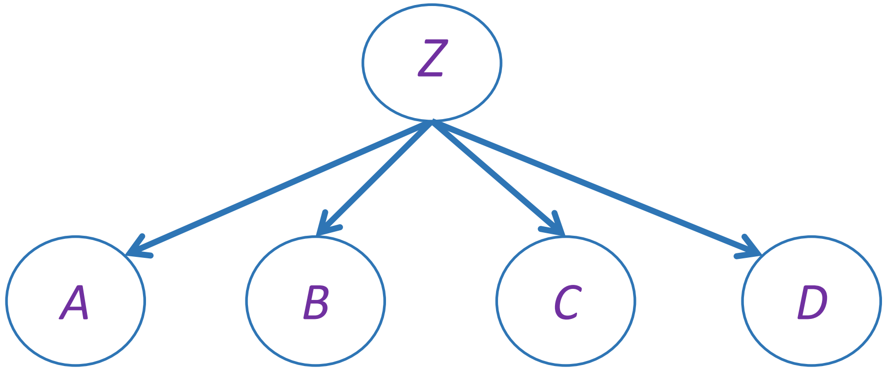
B	+j	+m	$P(B \mid +j, +m)$
+b	+j	+m	0.284172
-b	+j	+m	0.715828

B	+j	+m	$f_9(B, +j, +m)$
+b	+j	+m	0.000592
-b	+j	+m	0.001492

+j	+m	$Z = f_{10}(+j, +m)$
+j	+m	0.002084

Order matters

- Elimination Order: C, B, A, Z
  - $P(D) = \alpha \sum_{z,a,b,c} P(D|z)\, P(z)\, P(a|z)\, P(b|z)\, P(c|z)$
  - $\quad = \alpha \sum_{z} P(D|z)\, P(z) \sum_{a} P(a|z) \sum_{b} P(b|z) \sum_{c} P(c|z)$
  - Largest factor has 2 variables (D,Z)



- Elimination Order: Z, C, B, A
  - $P(D) = \alpha \sum_{a,b,c,z} P(a|z)\, P(b|z)\, P(c|z)\, P(D|z)\, P(z)$
  - $\quad = \alpha \sum_{a} \sum_{b} \sum_{c} \sum_{z} P(a|z)\, P(b|z)\, P(c|z)\, P(D|z)\, P(z)$
  - Largest factor has 4 variables (A,B,C,D) (or 5 if you count pre-summation over Z)

- In general, with *n* leaves, factor of size 2^n

VE: Computational and Space Complexity

The computational and space complexity of variable elimination is determined by the largest factor (and it's space that kills you)

The elimination ordering can greatly affect the size of the largest factor.

- E.g., previous slide's example 2^n vs. 2

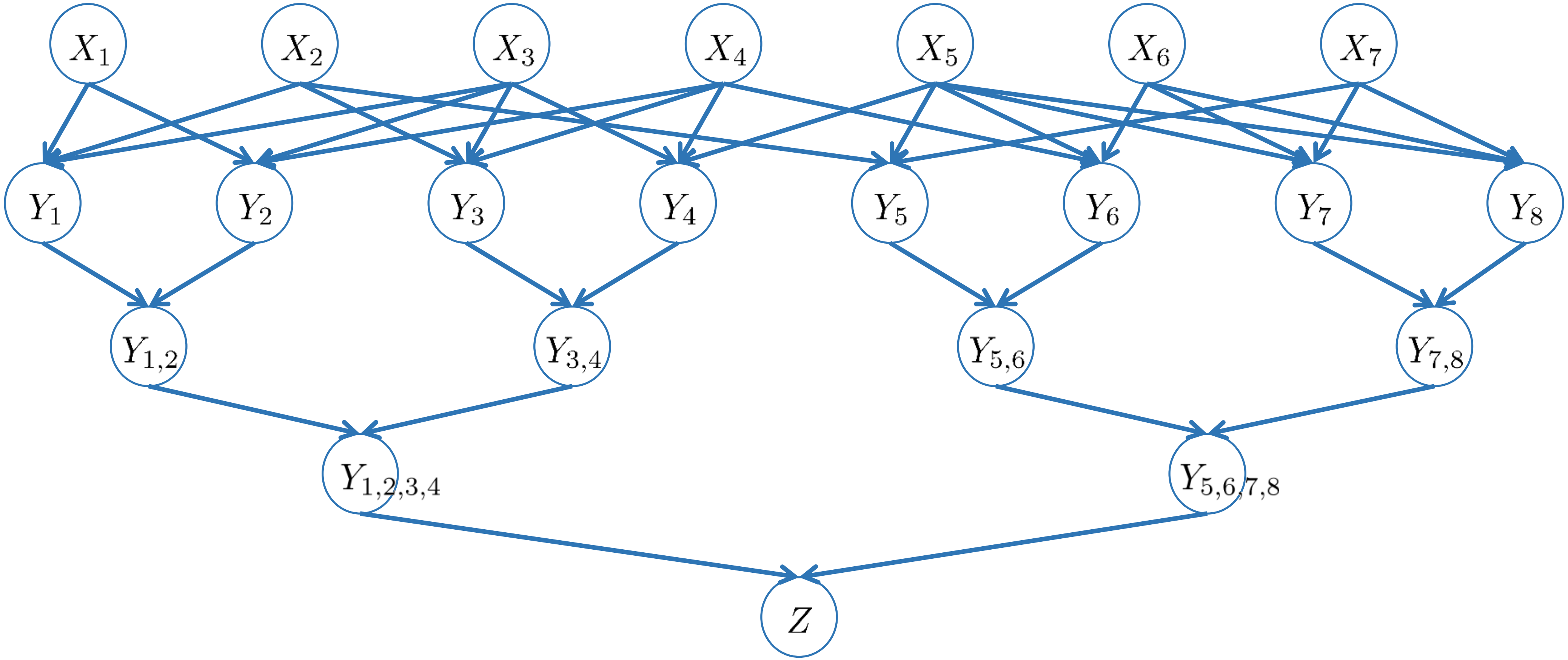
Does there always exist an ordering that only results in small factors?

- No!

VE: Computational and Space Complexity

Inference in Bayes' nets is NP-hard.

No known efficient probabilistic inference in general.



Bayes Nets

✓ Part I: Representation and Independence

✓ Part II: Exact inference

✓ ▪ Enumeration (always exponential complexity)

✓ ▪ Variable elimination (worst-case exponential complexity, often better)

✓ ▪ Inference is NP-hard in general

Part III: Approximate Inference