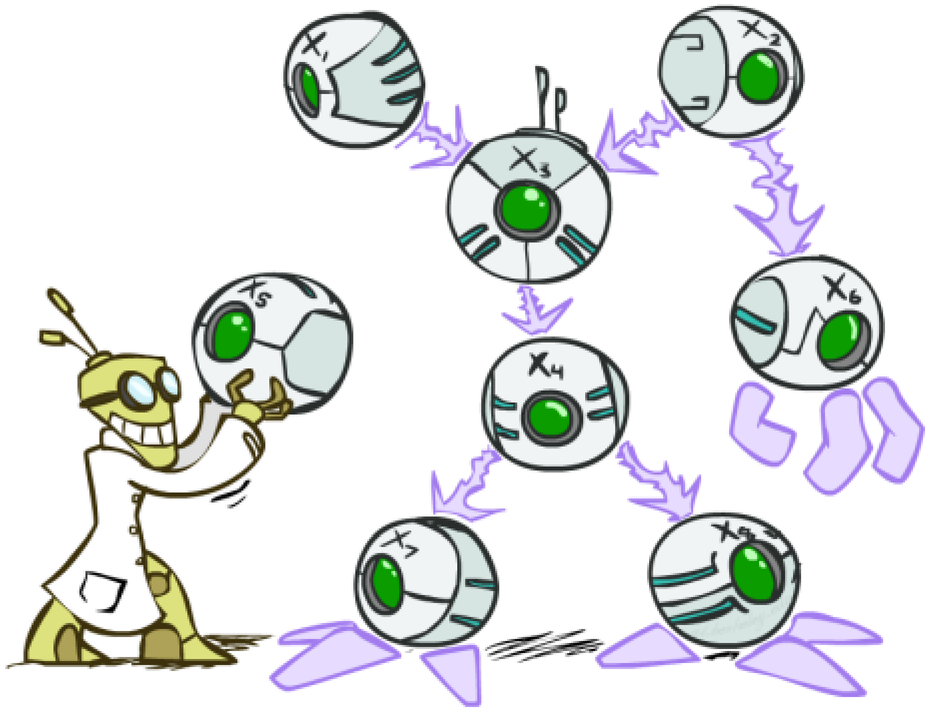
AI: Representation and Problem Solving

Bayes Nets Inference



Instructors: Tuomas Sandholm and Vincent Conitzer

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

Announcements

- HW8 (online) due tomorrow Thursday April 3

- HW9 (online, written) released tomorrow Thursday April 3, due next Thursday April 10

- No TA office hours during carnival EXCEPT Saturday April 5 12-2pm

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

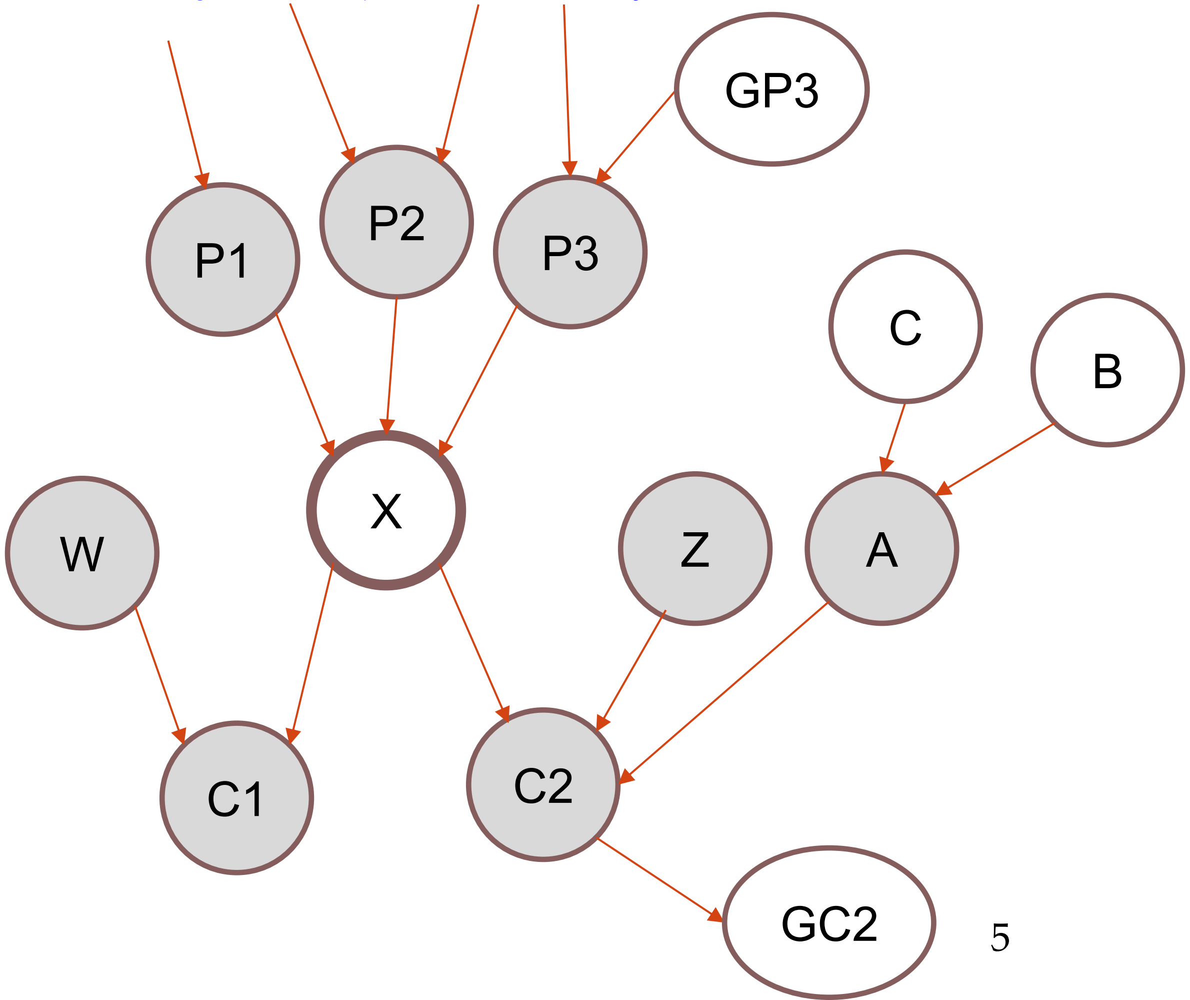
- Part III: Approximate Inference

Markov blanket

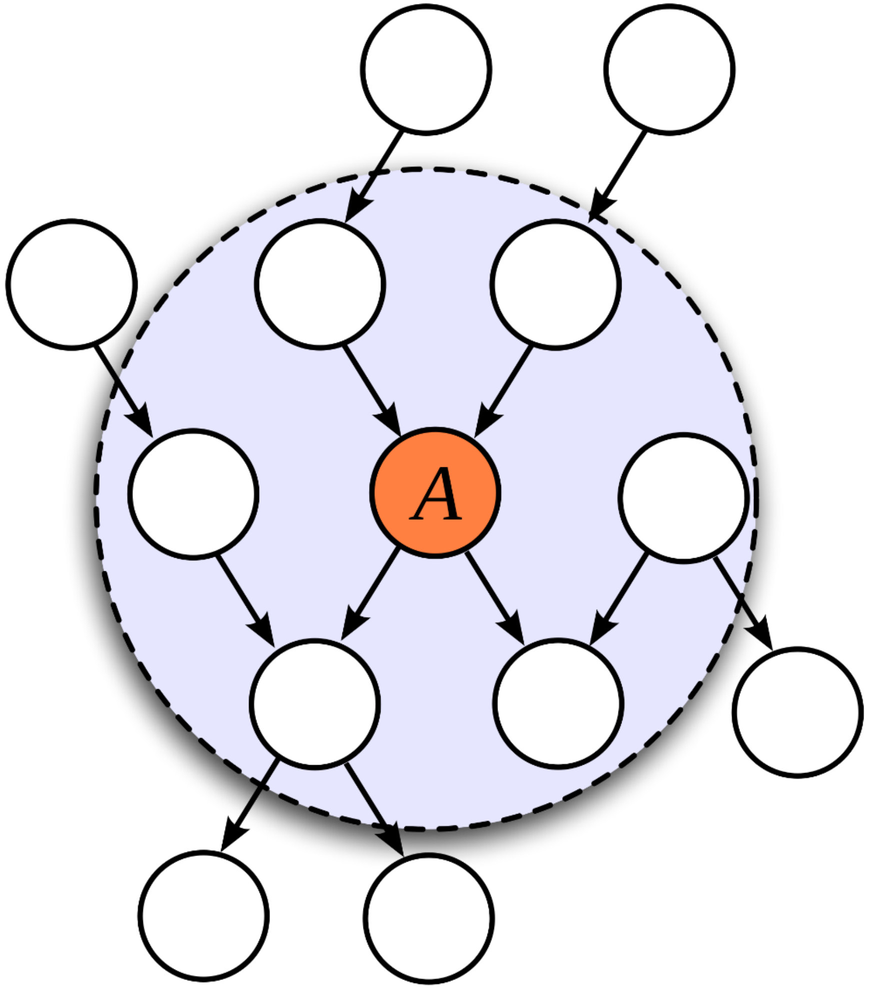
- Markov blanket of X - subset of variables such that all other variables are independent of X conditioned on the blanket

Markov blanket

- A variable's Markov blanket consists of parents, children, children's other parents
- ***Every variable is conditionally independent of all other variables given its Markov blanket***



Markov blanket



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

Queries

- What is the probability of *this* given what I know? $P(q \mid e)$

- What are the probabilities of all the possible outcomes (given what I know)? $P(Q \mid e)$

- Which outcome is the most likely outcome (given what I know)? $\text{argmax}_{q \in Q} P(q \mid e)$

Queries

- What is the probability of *this* given what I know?

$$P(\,q \mid e\,) = \frac{P(q,e)}{P(e)}$$

- What are the probabilities of all the possible outcomes (given what I know)?

$$P(\,Q \mid e\,) = \frac{P(Q,e)}{P(e)}$$

- Which outcome is the most likely outcome (given what I know)?

$$\text{argmax}_{q \in Q}\, P(\,q \mid e\,) = \text{argmax}_{q \in Q} \frac{P(q,e)}{P(e)}$$

Queries

- What is the probability of *this* given what I know?

$$P(\, q \mid e \,) = \frac{P(q, e)}{P(e)} = \frac{\sum_{h_1} \sum_{h_2} P(q, h_1, h_2, e)}{P(e)}$$

- What are the probabilities of all the possible outcomes (given what I know)?

$$P(\, Q \mid e \,) = \frac{P(Q, e)}{P(e)} = \frac{\sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)}{P(e)}$$

- Which outcome is the most likely outcome (given what I know)?

$$\operatorname{argmax}_{q \in Q} P(\, q \mid e \,) = \operatorname{argmax}_{q \in Q} \frac{P(q, e)}{P(e)} = \operatorname{argmax}_{q \in Q} \frac{\sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)}{P(e)}$$

Poll 1

- If we only have the joint table $P(Q, H_1, H_2, E)$, what is the **minimum number of times** we need to compute $P(e)$ to build $P(Q \mid e)$?

$$P(Q \mid e) = \frac{P(Q,\ e)}{P(e)} = \frac{\sum_{h_1} \sum_{h_2} P(Q,\ h_1,\ h_2,\ e)}{P(e)}$$

A) 0
B) 1
C) 10
D) 30
E) 200
F) 600

- Q can take on 10 different values
- H_1 can take on 4 different values
- H_2 can take on 5 different values
- E can take on 3 different values

Poll 2

- If we only have the joint table $P(Q, H_1, H_2, E)$, what is the **minimum** number of times we need to compute $P(e)$ to compute the following?

$$\text{argmax}_{q \in Q} P(q \mid e) = \frac{P(q,\ e)}{P(e)} = \frac{\sum_{h_1} \sum_{h_2} P(q,\ h_1,\ h_2,\ e)}{P(e)}$$

A) 0

B) 1

C) 10

D) 30

E) 200

F) 600

- Q can take on 10 different values
- H_1 can take on 4 different values
- H_2 can take on 5 different values
- E can take on 3 different values

Normalization

- Sometimes we don't care about exact probability and we skip $P(e)$

$$P(Q \mid e) = \frac{\sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)}{P(e)}$$

$$P(Q \mid e) = \frac{1}{Z} \sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)$$

$$P(Q \mid e) = \alpha \sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)$$

$$P(Q \mid e) \propto \sum_{h_1} \sum_{h_2} P(Q, h_1, h_2, e)$$

Bayes Nets in the Wild

Example: Speech Recognition

“artificial ……….”

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

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?

$P(\textbf{limb} \mid \text{artificial, [audio]})$

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

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

Break down problem

n-gram probability * audio probability

$P(\textbf{limb} \mid \text{artificial}) * P(\text{[audio]} \mid \textbf{limb})$

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

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

Bayes Nets in the Wild

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

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

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

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

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

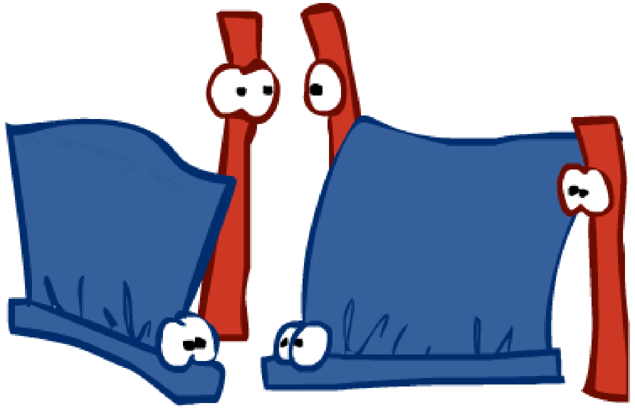
$$= \text{argmax}_{\boldsymbol{second}}\, P(\boldsymbol{second} \mid artificial)\, P([audio] \mid \boldsymbol{second})$$

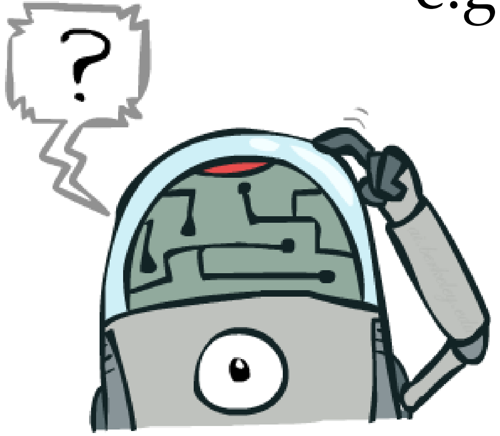
n-gram probability * [audio] probability

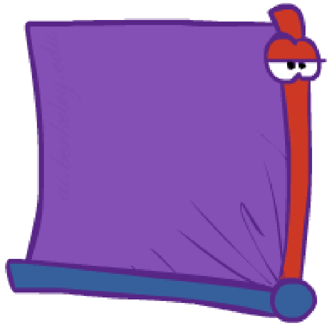
Inference

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

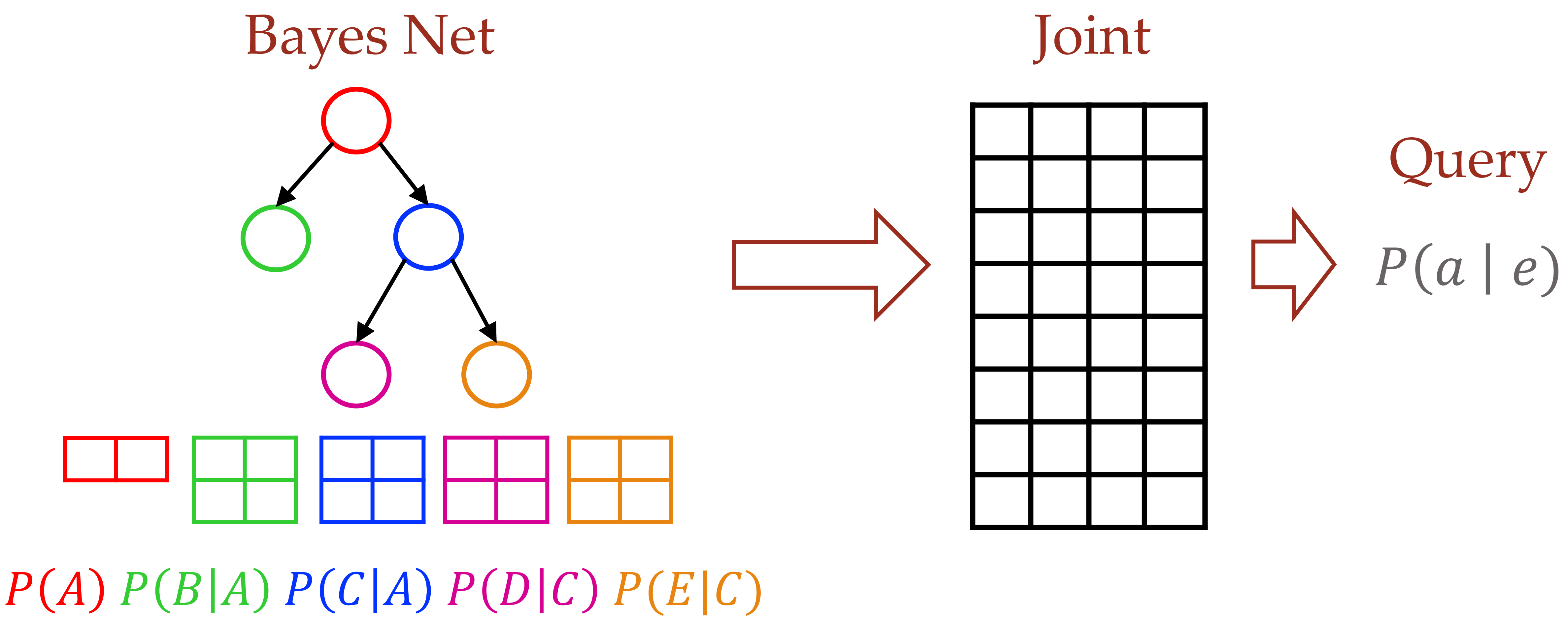
- Examples:
  - Posterior marginal probability
    - $P(Q \mid e_1,..,e_k)$
    - e.g., what disease might I have?
  - Most likely explanation:
    - $\text{argmax}_{q,r,s}\ P(Q=q,R=r,S=s \mid e_1,..,e_k)$
    - e.g., what was just said?



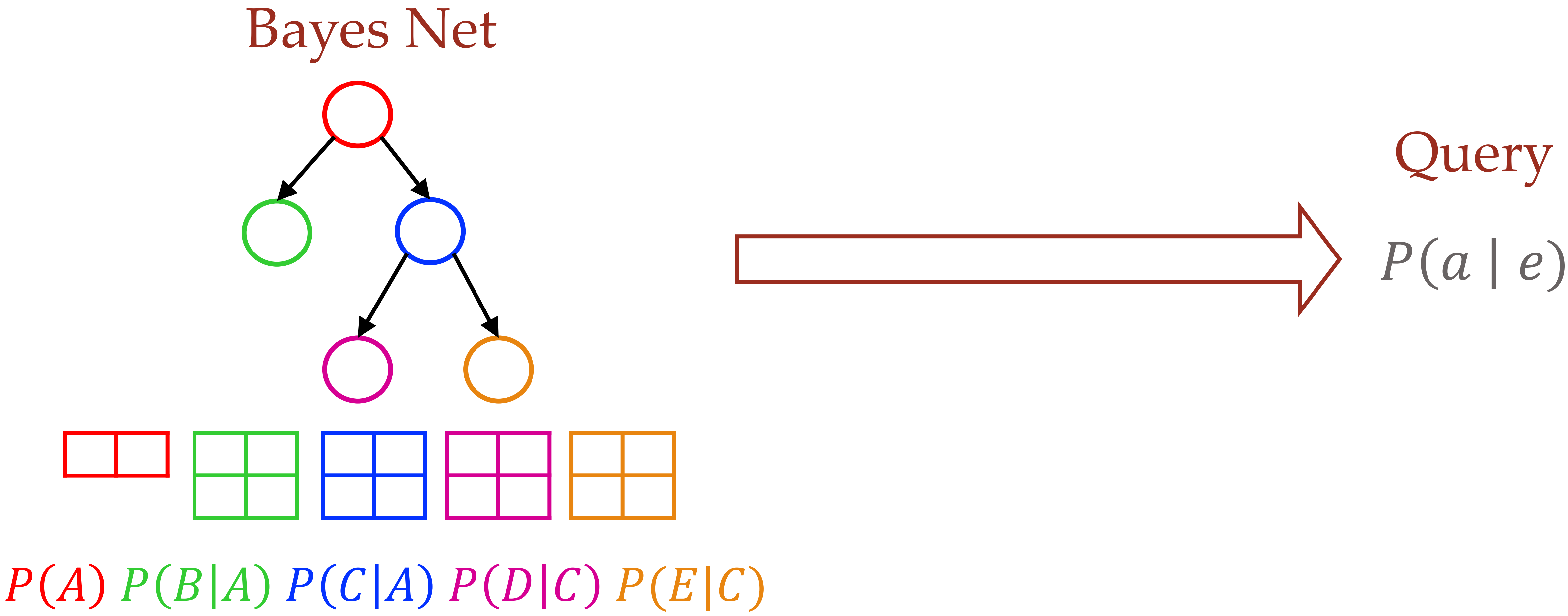




Answer Any Query from Bayes Net

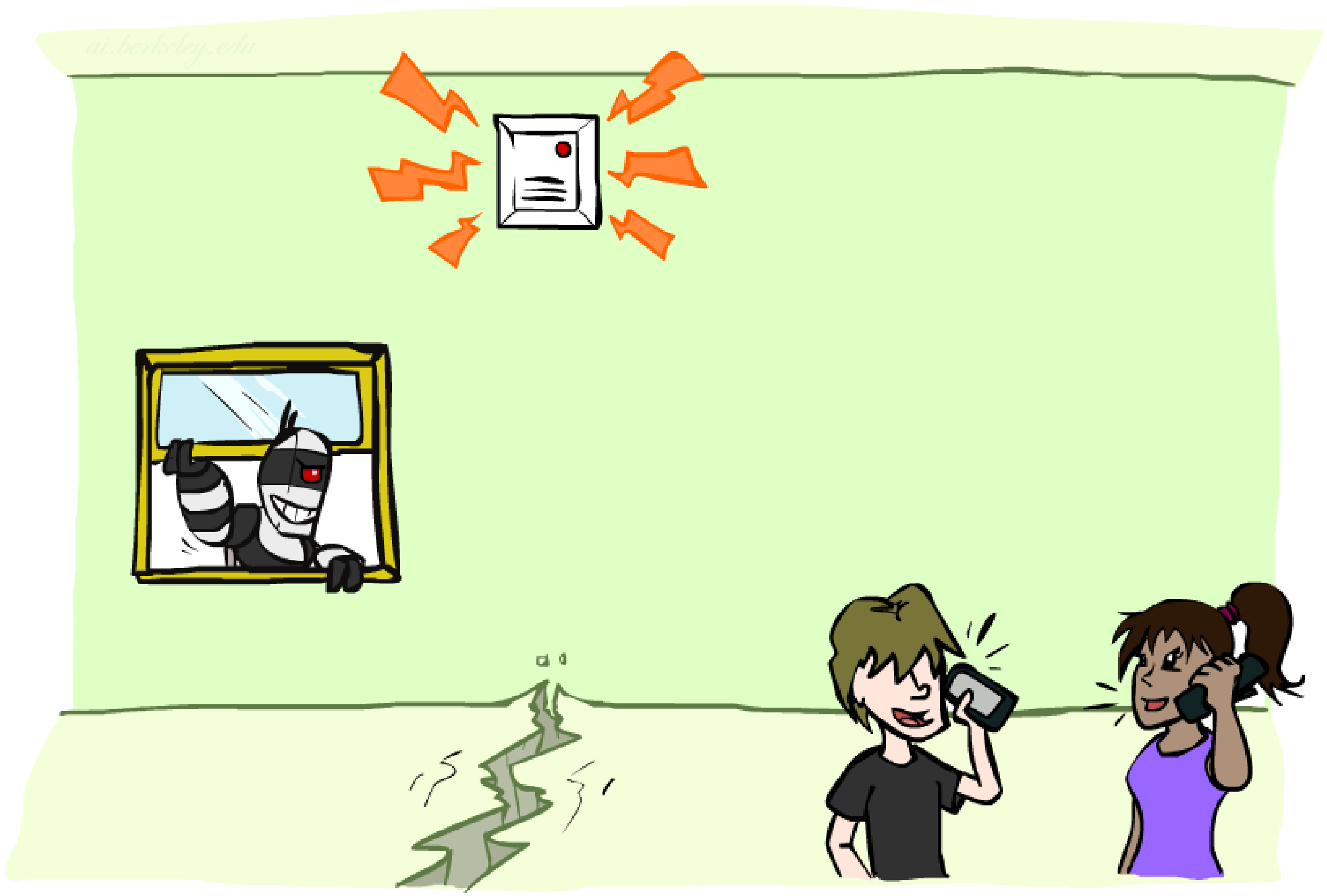


Next: Answer Any Query from Bayes Net



Example: Alarm Network

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



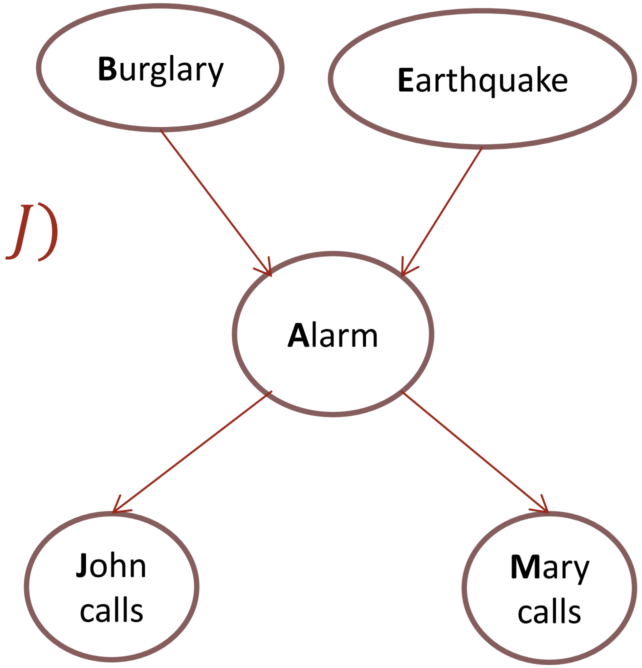
Example: Alarm Network

- Joint distribution factorization example

- Generic chain rule
  - $P(X_1 \ldots X_n) = \prod_i P(X_i | X_1 \ldots X_{i-1})$

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

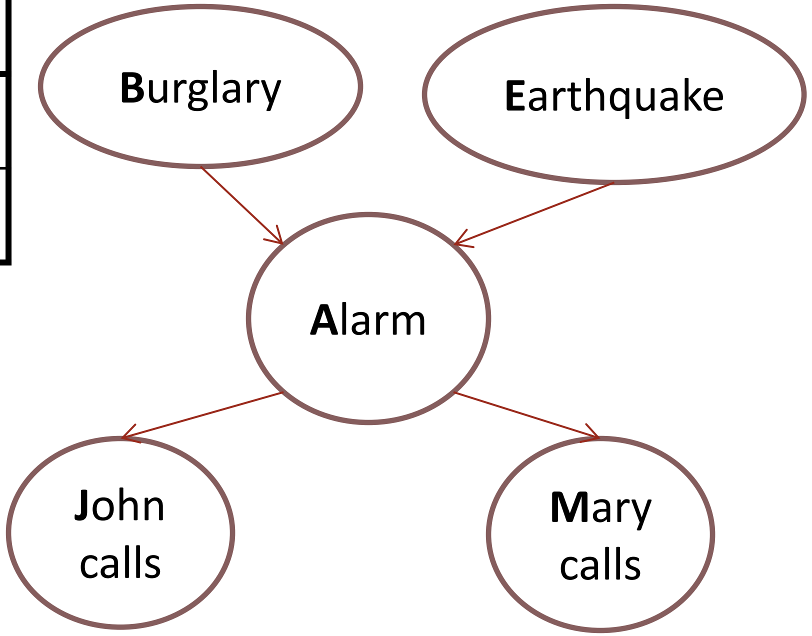
$$P(B,E,A,J,M) = P(B)\, P(E) \quad P(A|B,E)\, P(J|A) \qquad P(M|A)$$



- Bayes nets
  - $P(X_1 \ldots X_n) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



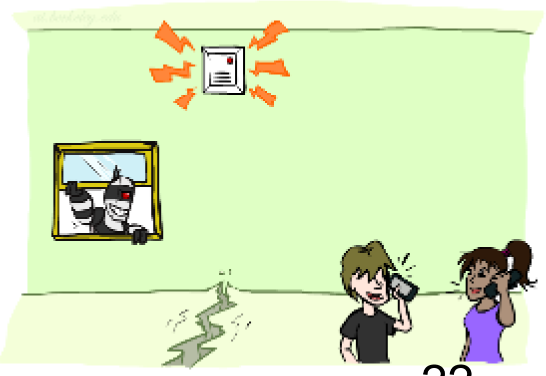
E	P(E)
+e	0.002
-e	0.998

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

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: Alarm Network

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

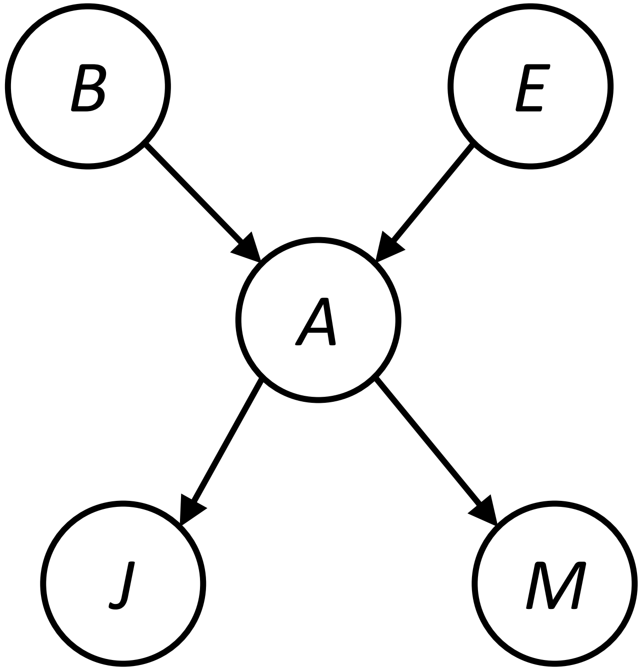
$$P(+b,-e,-a,-j,-m) = P(-e) * P(-a|+b,-e) * P(+b) * P(-j|-a) * P(-m|-a)$$
$$= 0.998 * 0.06 * 0.001 * 0.95 * 0.99$$

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

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

Inference by Enumeration in Bayes Net

- 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



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

- So inference in Bayes nets means computing sums of products of numbers: sounds easy!!
- Problem: sums of ***exponentially many*** products!

Can we do better?

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

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

  - **Lots of repeated subexpressions!**

Can we do better?

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

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

  - **Lots of repeated subexpressions!**

Inference Overview

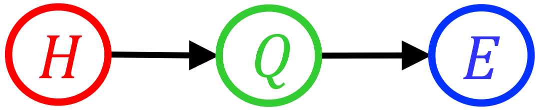
- Given random variables Q, H, E (query, hidden, evidence)

Enumeration

- We know how to do inference on a joint distribution

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

- Bayes nets break down joint distribution into CPT factors



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

Variable Elimination

- But we can be more efficient

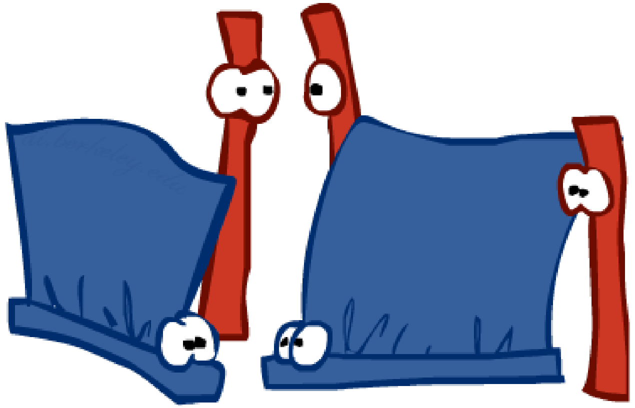
$$P(q|e) = \alpha\, P(e|q) \sum_{h \in \{h_1,h_2\}} P(h) P(q|h)$$
$$= \alpha\, P(e|q)\; [P(h_1)P(q|h_1) + P(h_2)P(q|h_2)]$$
$$= \alpha\, P(e|q)\; P(q)$$

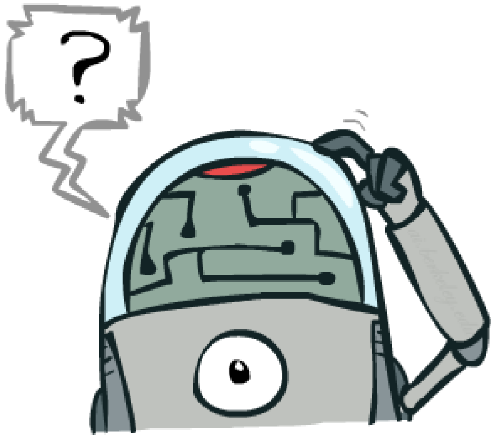
- Now just extend to larger Bayes nets and a variety of queries

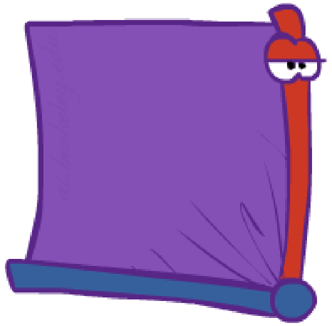
Factor Tables

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

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







Variable elimination: The basic ideas

- Move summations inwards as far as possible

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

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

Variable elimination: The basic ideas

- Move summations inwards as far as possible, inner sum is easier to compute

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

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

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

Variable Elimination

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

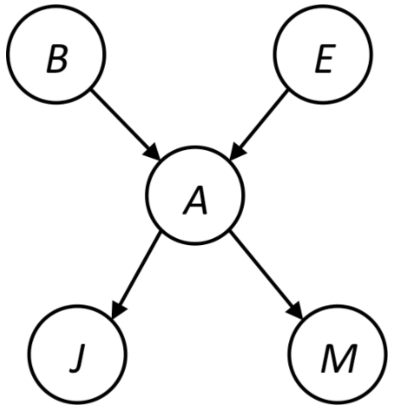
- Start with initial factors:
  - Local CPTs (but instantiated by evidence)

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



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

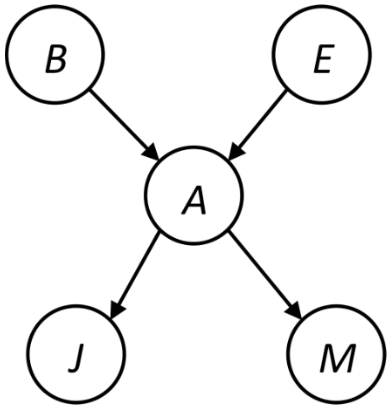
Push summations inwards such that products that do not depend on the variable are pulled out of the sum.

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

Example

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

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



Choose A (innermost sum)

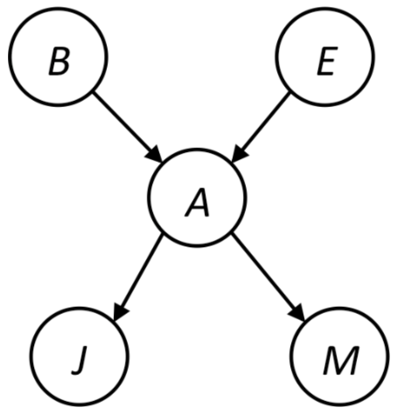
Create a table $t_1 = P(A \mid B,E)P(j \mid A)P(m \mid A)$
How many entries does this table have?

$$= \alpha P(B) \sum_{e} P(e) \sum_{a} t_1(a, B, e, j, m)$$

Example

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

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



Choose A (inner most sum)

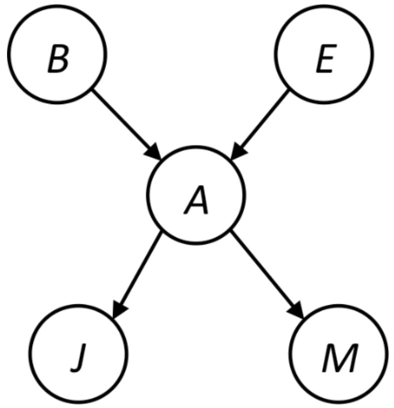
Sum over A in the table to create a factor $f_1 = \sum_a t(a, B, e, j, m)$

How many entries does this new factor table have?

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

Example

$$= \alpha P(B) \sum_{e} P(e) f_1(B, e, j, m)$$



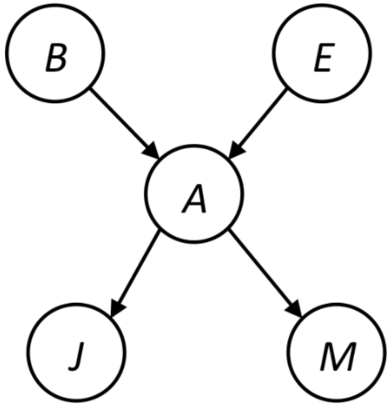
Choose E (innermost sum)

Create a table $t_2 = P(E)\, f_1(B,E,j,m)$
How many entries does this table have?

$$= \alpha P(B) \sum_{e} t_2(B, e, j, m)$$

Example

$$= \alpha P(B) \sum_{e} t_2(B, e, j, m)$$



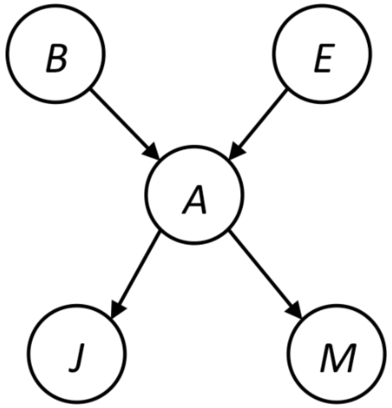
Choose E (innermost sum)

Sum over E in the table to create a factor $f_2 = \sum_e t_2(B, e, j, m)$
How many entries does this new factor table have?

$$= \alpha P(B) f_2(B, j, m)$$

Example

$$= \alpha P(B) f_2(B, j, m)$$



Multiply remaining probability to create joint probability *P(B,j,m)*

How many entries does this probability table have?

Don't forget the normalization to compute the conditional probability!

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

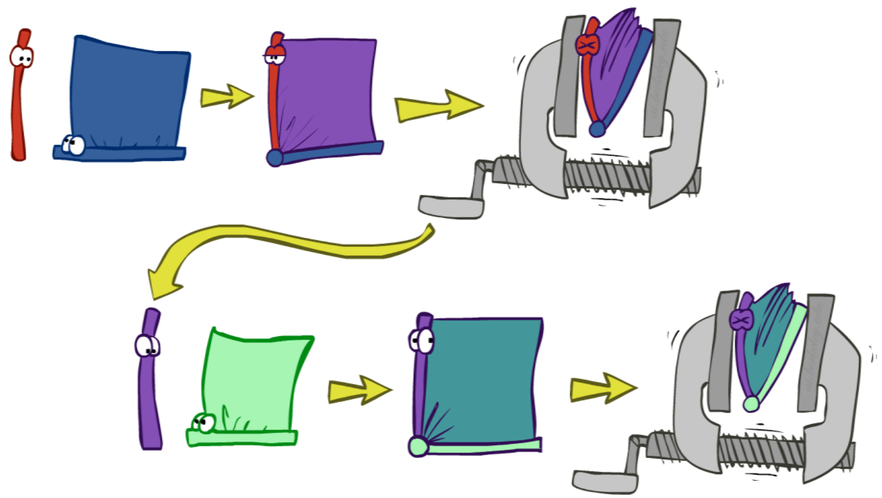
$$P(B|j, m) = \alpha P(B, j, m)$$

Example summary

- Query $P(B \mid j,m)\ \alpha \sum_e \sum_a P(j|a)\ P(e)\ P(m|a)\ P(a|B,e)\ P(B)$
- Join A to get table $P(A \mid B,E)P(j \mid A)P(m \mid A)$
- Eliminate A to get factor $f_1(B,E,j,m)$
- Join E to get table $P(E)f_1(B,E,j\ m)$
- Eliminate E to get table $P(B)f_2(B,j,m)$
- Join B to get $P(B,j,m)$ and then normalize

Variable elimination: The basic ideas

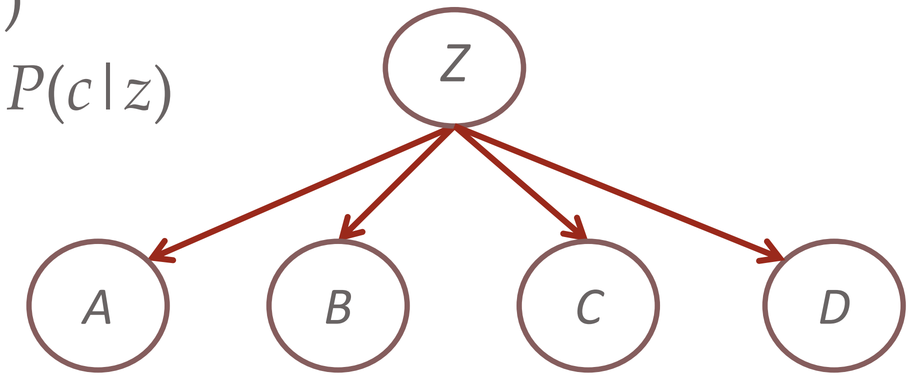
- Move summations inwards as far as possible
  - $P(B \mid j, m) = \alpha \sum_e \sum_a P(B)\, P(e)\, P(a \mid B,e)\, P(j \mid a)\, P(m \mid a)$
    $= \alpha\, P(B) \sum_e P(e) \sum_a P(a \mid B,e)\, P(j \mid a)\, P(m \mid a)$



- Do the calculation from the inside out
  - I.e., sum over *a* first, then sum over *e*
  - Problem: $P(a \mid B,e)$ isn't a single number, it's a bunch of different numbers depending on the values of *B* and *e*
  - Solution: use arrays of numbers (of various dimensions) with appropriate operations on them; these are called factors

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

VE: Computational and Space Complexity

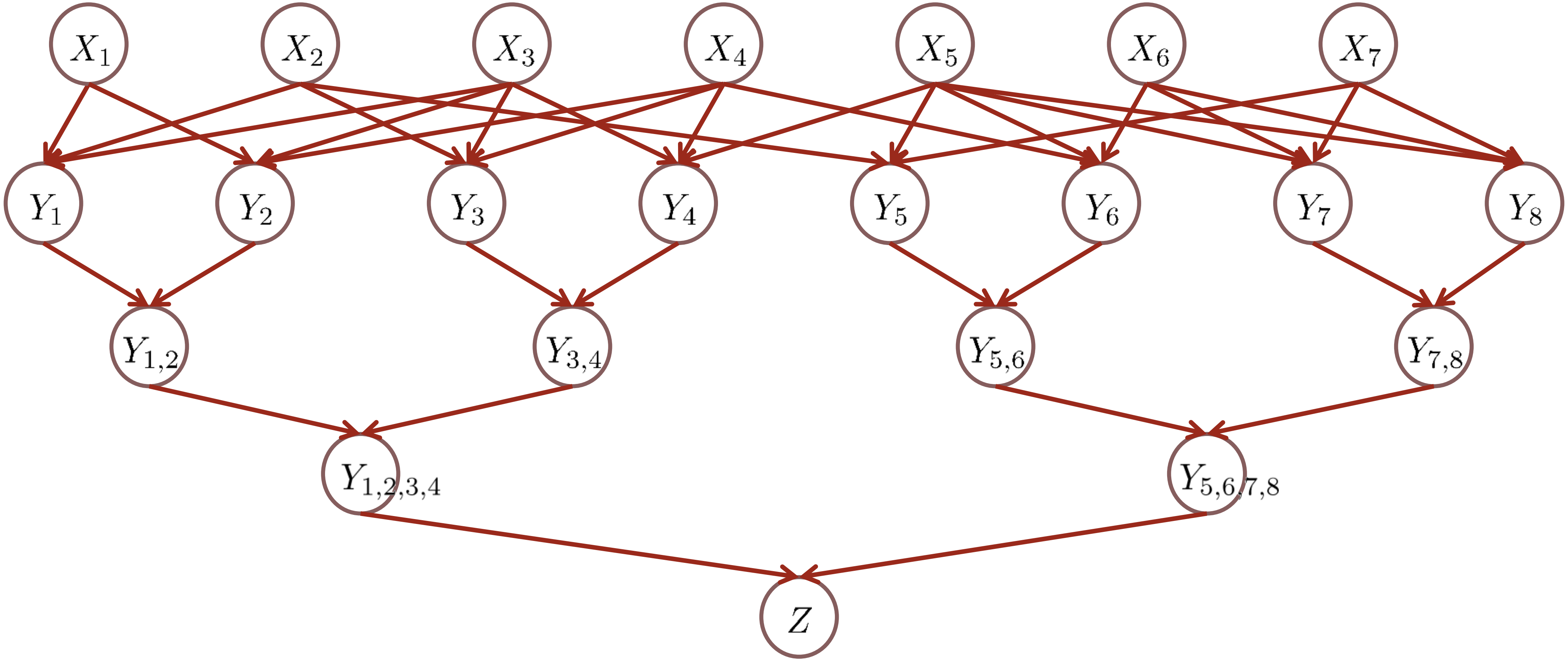
- The computational and space complexity of variable elimination is determined by the largest factor (and it's space that kills you)

- The elimination ordering can greatly affect the size of the largest factor.
  - E.g., previous slide's example 2^n vs. 2

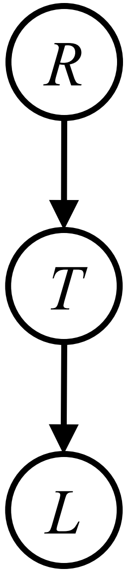
- Does there always exist an ordering that only results in small factors?
  - No!

VE: Computational and Space Complexity

- Inference in Bayes' nets is NP-hard.
- No known efficient probabilistic inference in general.



Another example



$$P(L) = ?$$

- Inference by Enumeration

$$= \sum_t \sum_r P(L|t)P(r)P(t|r)$$

Join on r

Join on t

Eliminate r

Eliminate t

- Variable Elimination

$$= \sum_t P(L|t) \sum_r P(r)P(t|r)$$

Join on r

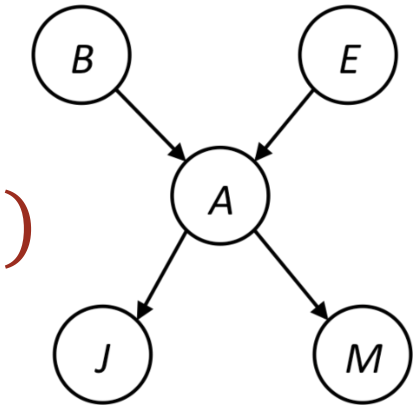
Eliminate r

Join on t

Eliminate t

New Example

Query $P(E \mid m)$



$$= \alpha \sum_b \sum_a \sum_j P(j|a)\, P(E)\, P(m|a)\, P(a|b,E)\, P(b)$$

Push summations inwards such that products that do not depend on the variable are pulled out of the sum.

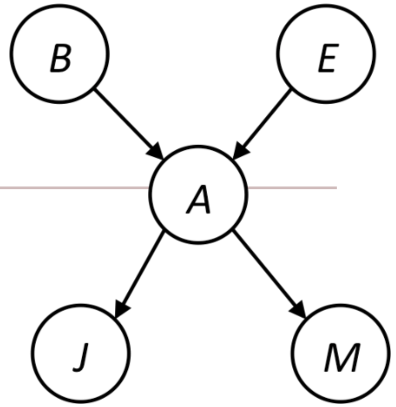
$$= \alpha P(E) \sum_b P(B) \sum_a P(m|a)\, P(a|b,E) \sum_j P(j|a)$$

Poll 3

- How many entries does the factor that results after summing out the second variable (a) on the previous slide have?

A) 0
B) 1
C) 2
D) 4
E) 6
F) 8

$P(B \mid j, m)$



See video: Math for Variable Elimination

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

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

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

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

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

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

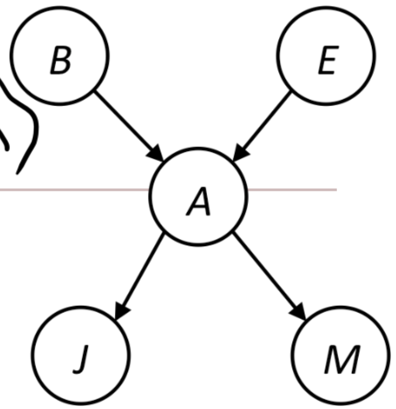
$$= \alpha f_1(B) \sum_e f_8(B, e, j, m)$$

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

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

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

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



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

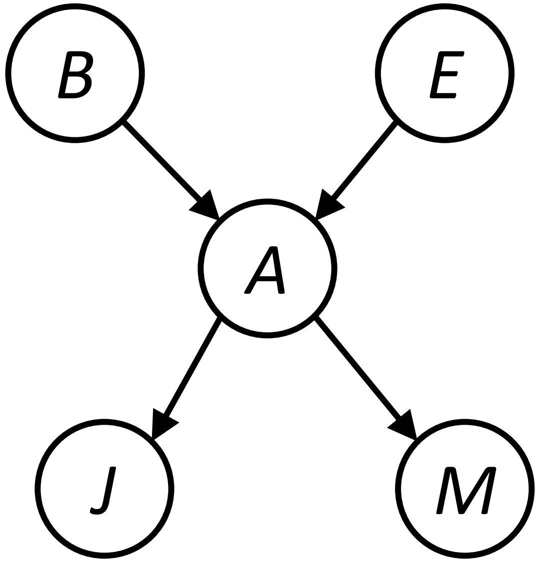
Bayes Nets

- ✓ 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

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

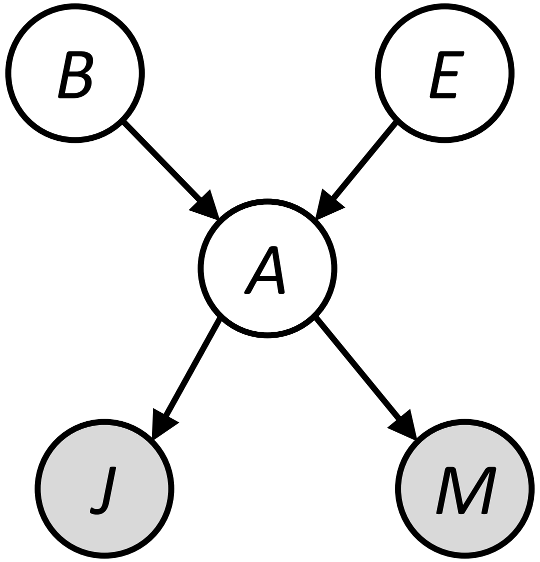
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

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

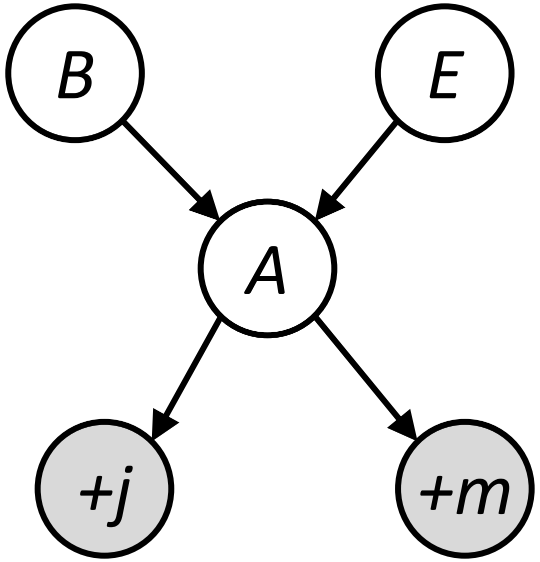
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Example (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

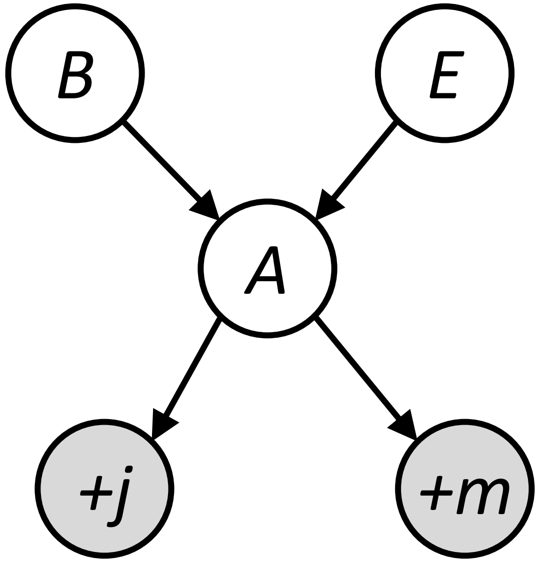
A	+m	P(+m|A)
+a	+m	0.7
-a	+m	0.01

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

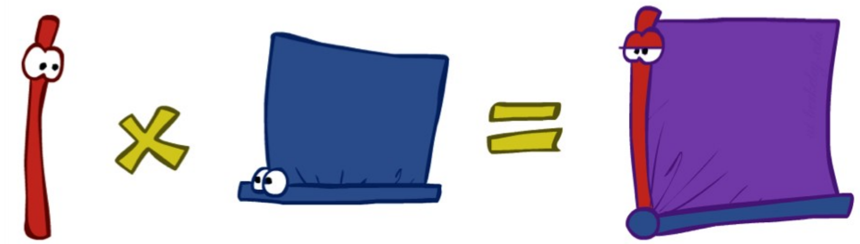
B	$f_1(B)$
+b	0.001
-b	0.999

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

Operation 1: Pointwise product

- First basic operation: pointwise product of factors (similar to a database join, ***not*** matrix multiply!)
  - New factor has ***union*** of variables of the two original factors
  - Each entry is the product of the corresponding entries from the original factors



- Example: $P(A) \times P(J \mid A) = P(A,J)$

P(A)

true	0.1
false	0.9

x

P(J|A)

A \ J	true	false
true	0.9	0.1
false	0.05	0.95

=

P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855