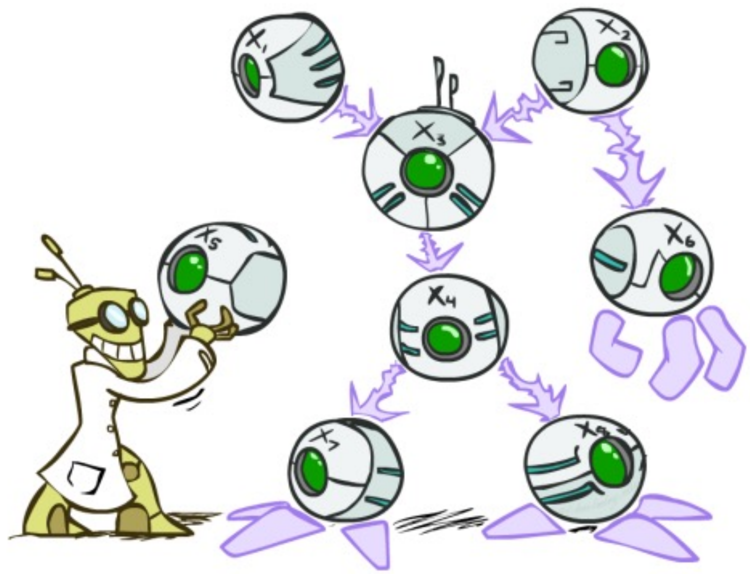
AI: Representation and Problem Solving

Bayes Nets Inference



Instructors: Tuomas Sandholm and Nihar Shah

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

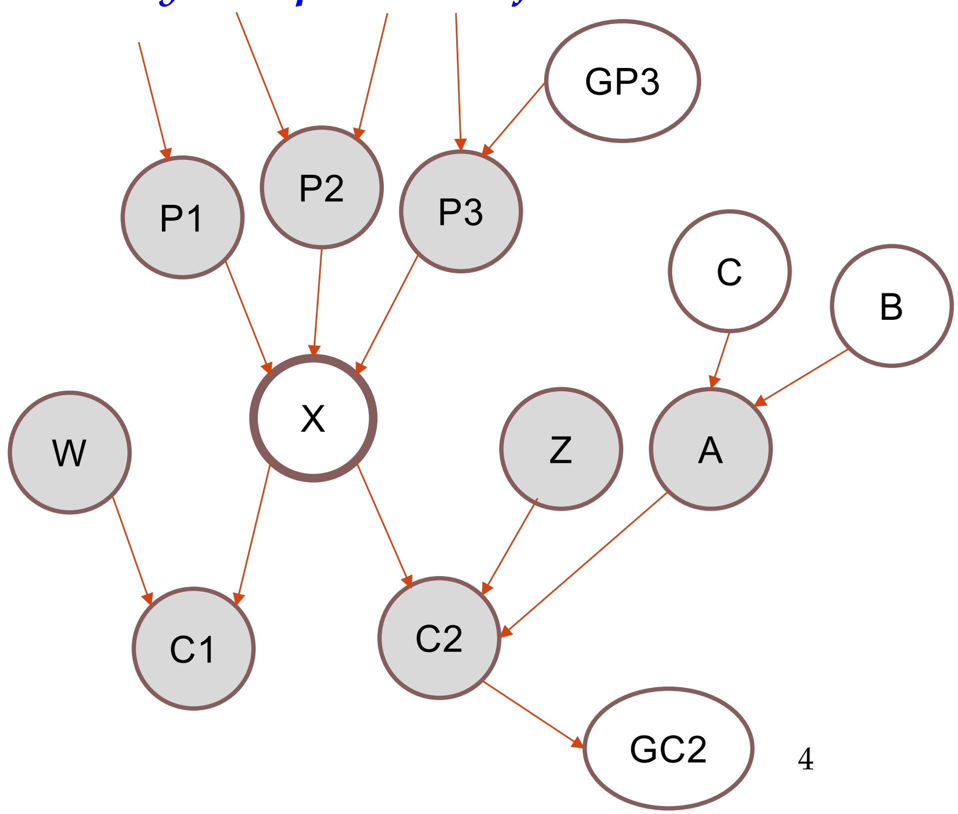
Part III: Approximate Inference

Markov blanket

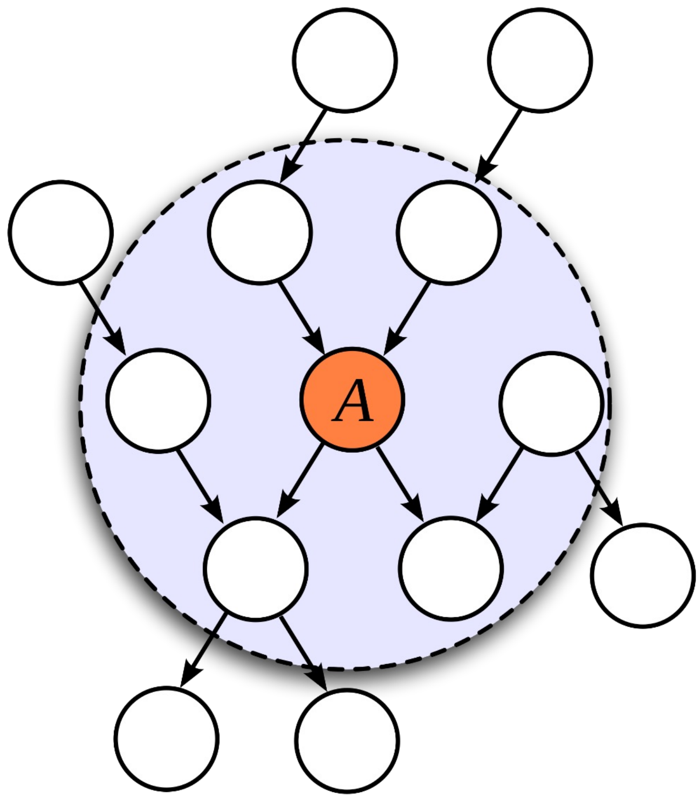
- Markov blanket of X - subset of variables such that all other variables are independent of X conditioned on the blanket

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



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)?

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

Normalization

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

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

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

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

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

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

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

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

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

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

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

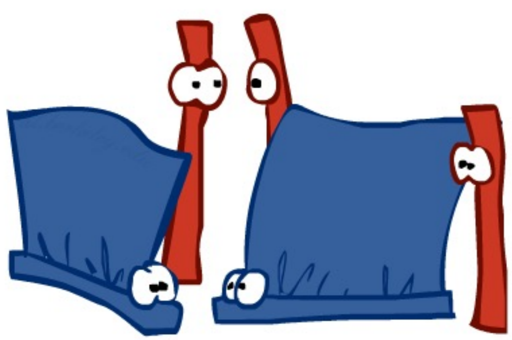
$$= \text{argmax}_{\boldsymbol{second}}\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid \boldsymbol{second})$$

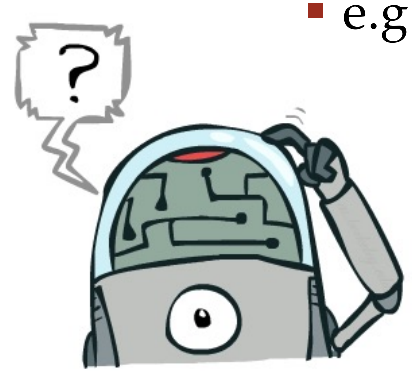
n-gram probability * audio probability

Inference

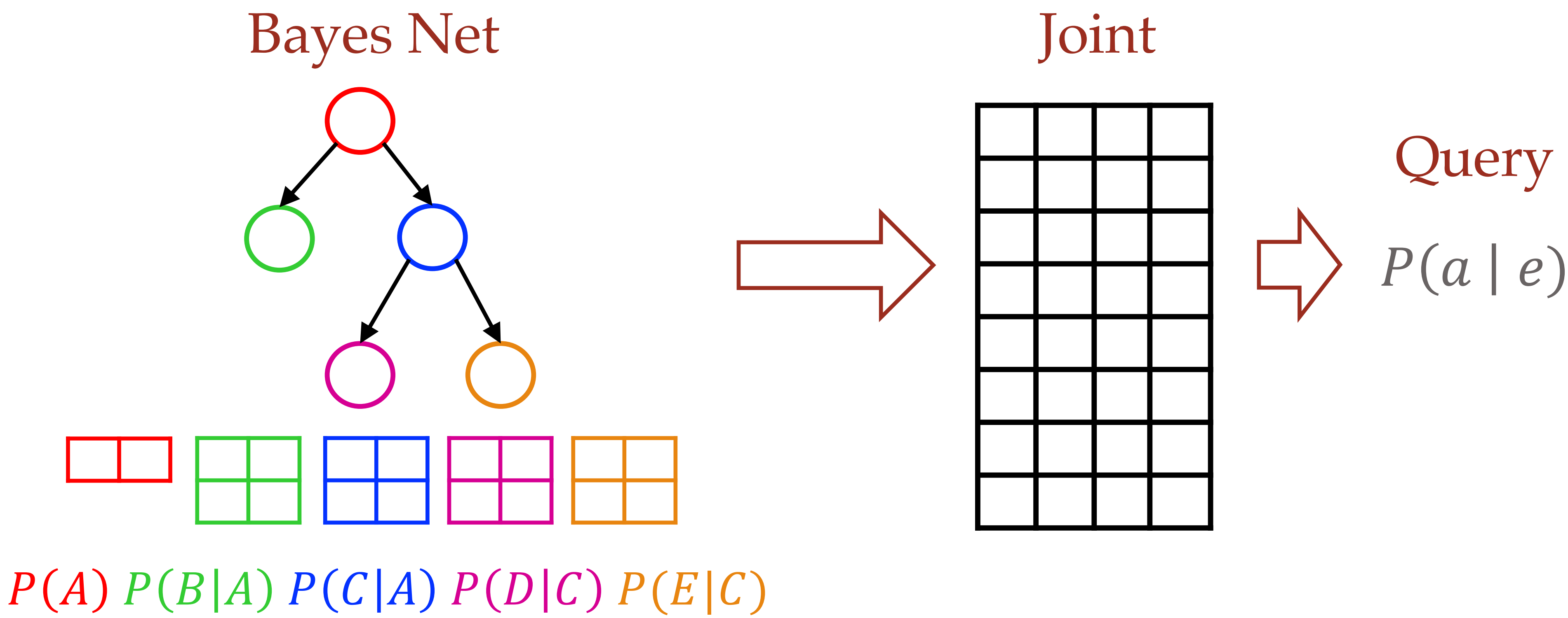
- **Inference:** calculating some useful quantity from a probability model (joint probability distribution)

- 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?

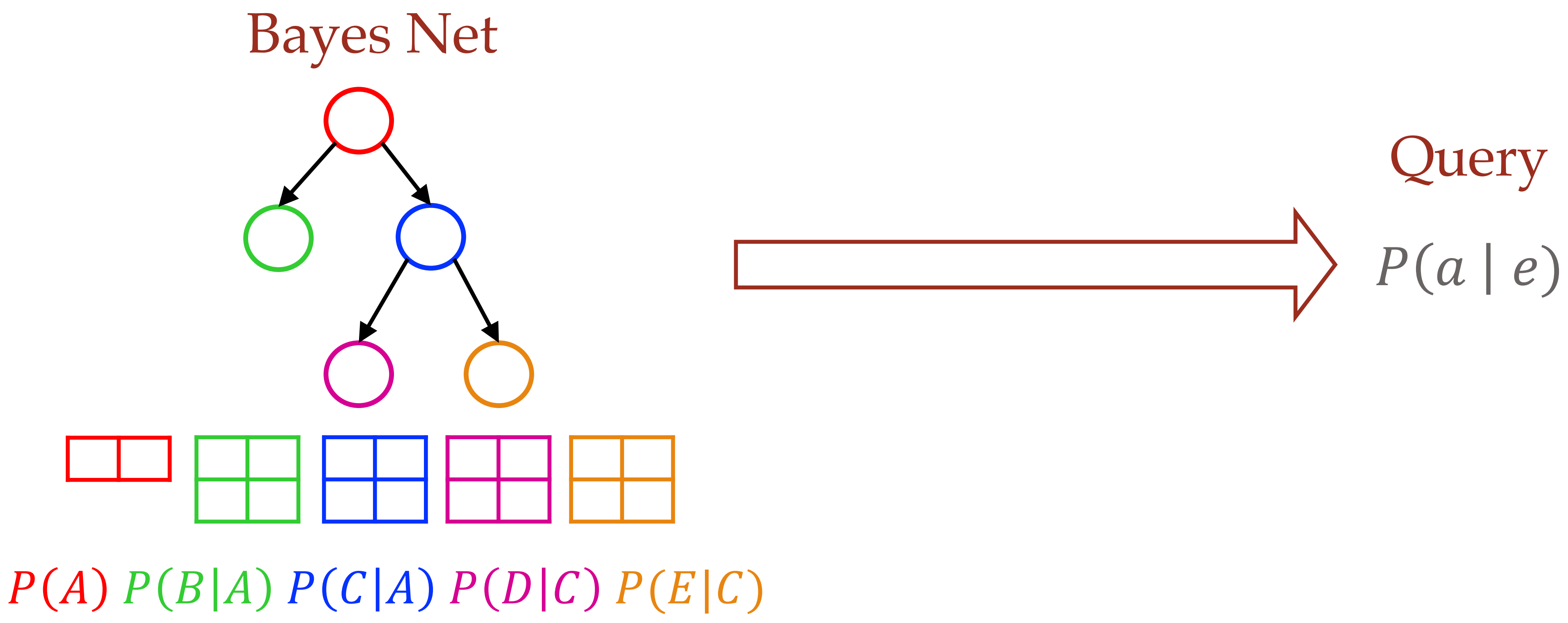




Answer Any Query from Bayes Net

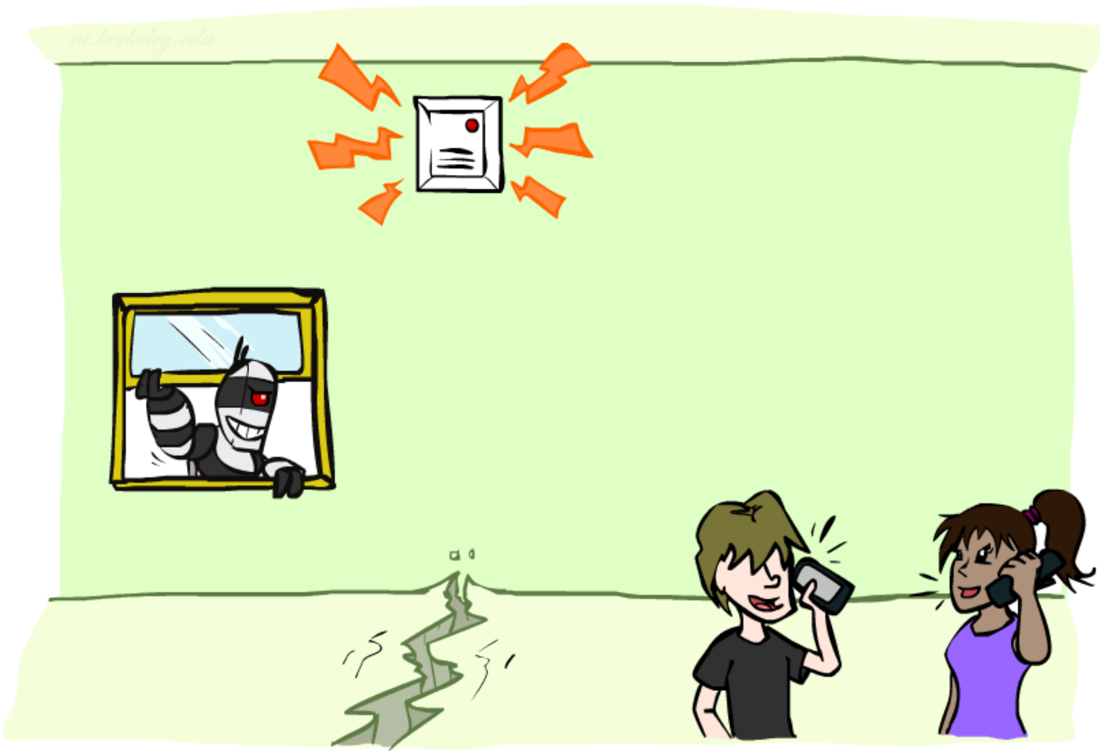


Next: Answer Any Query from Bayes Net



Example: Alarm Network

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



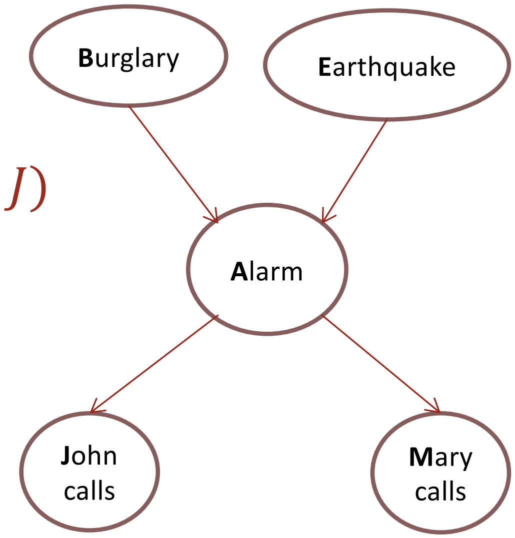
Example: Alarm Network

- Joint distribution factorization example

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

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

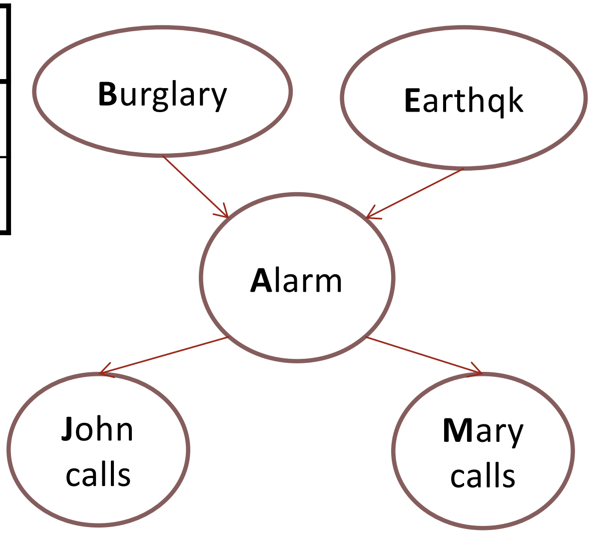
$$P(B,E,A,J,M) = P(B)\, P(E) \quad P(A|B,E)\, P(J|A) \qquad P(M|A)$$



- 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



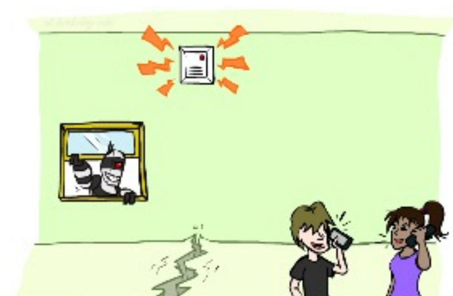
E	P(E)
+e	0.002
-e	0.998

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

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

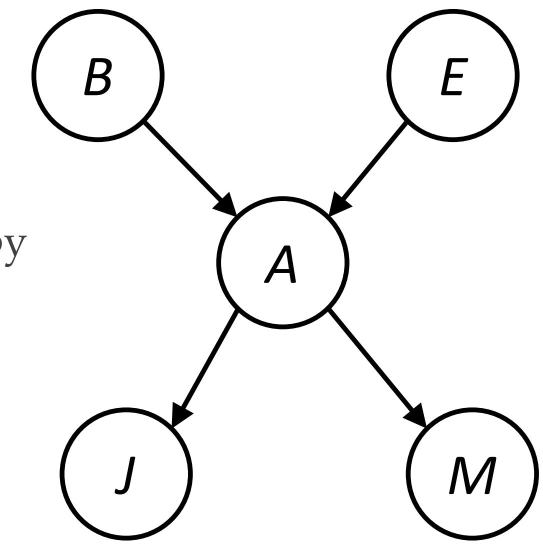
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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



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



$$
\begin{aligned}
P(B \mid j, m) &= \alpha\, P(B, j, m) \\
&= \alpha \textstyle\sum_{e,a} P(B, e, a, j, m) \\
&= \alpha \textstyle\sum_{e,a} P(B)\, P(e)\, P(\mathrm{a} \mid B, e)\, P(j \mid a)\, P(m \mid a)
\end{aligned}
$$

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

Can we do better?

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

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

  - **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!**

Can we do better?

- Consider
  - $x_1 y_1 z_1 + x_1 y_1 z_2 + x_1 y_2 z_1 + x_1 y_2 z_2 + x_2 y_1 z_1 + x_2 y_1 z_2 + x_2 y_2 z_1 + x_2 y_2 z_2$
  - 16 multiplies, 7 adds
  - Lots of repeated sub expressions!
- Rewrite as
  - $(x_1 + x_2)(y_1 + y_2)(z_1 + z_2)$

Inference Overview

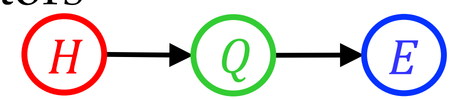
- Given random variables Q, H, E (query, hidden, evidence)

Enumeration

- We know how to do inference on a joint distribution

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

- We know Bayes nets can break down joint 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

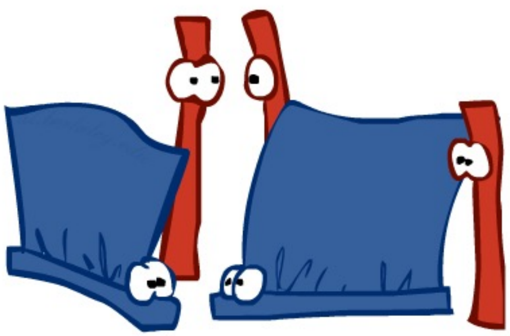
$$P(q|e) = \alpha\, P(e|q) \sum_{h \in \{h_1,h_2\}} P(h) P(q|h)$$
$$= \alpha\, P(e|q)\, [P(h_1)P(q|h_1) + P(h_2)P(q|h_2)]$$
$$= \alpha\, P(e|q)\, P(q)$$

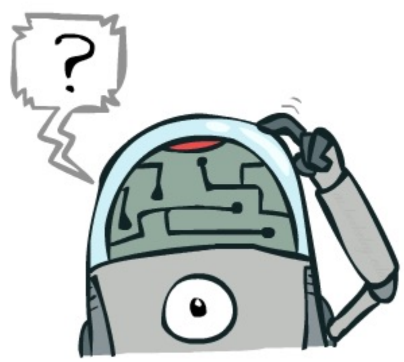
- Now just extend to larger Bayes nets and a variety of queries

Factor Tables

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

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





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)

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)

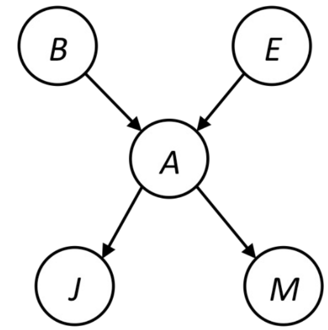
- While there are still hidden variables (not Q_i or evidence):
  - Pick a hidden variable H
  - **Join** all factors mentioning H
  - **Eliminate** (sum out) H

- Join all remaining factors and normalize

Example

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

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



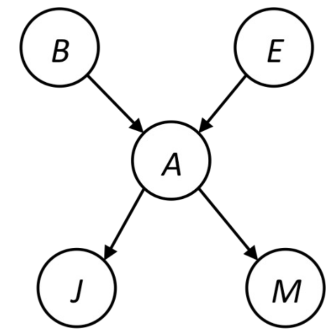
Push summations inwards such that products that do not depend on the variable are pulled out of the sum

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

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 (inner most sum)

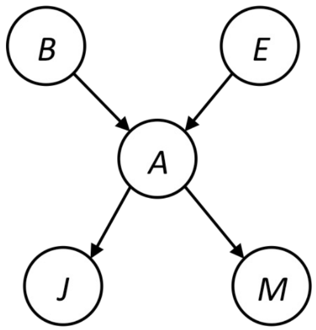
Create a table $t_1 = P(A \mid B,E)P(j \mid A)P(m \mid A)$
How many entries does this table have?

$$= \alpha P(B) \sum_e P(e) \sum_a t_1(a, B, e, j, m)$$

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)

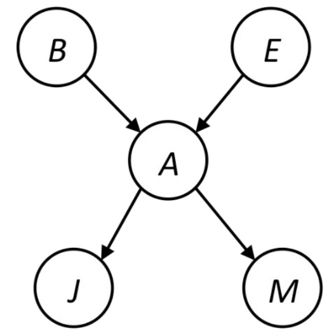
Sum over A in the table to create a factor $f_1 = \sum_a t(a, B, e, j, m)$

How many entries does this new factor table have?

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

Example

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



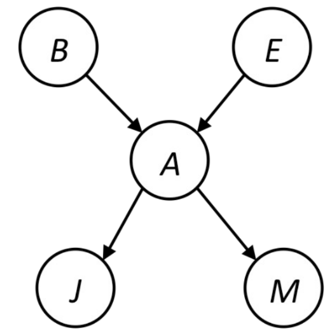
Choose E (inner most sum)

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

$$= \alpha P(B) \sum_e t_2(B, e, j, m)$$



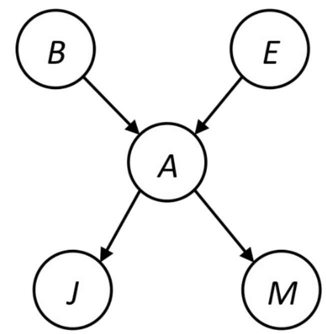
Choose E (inner most sum)

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!

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

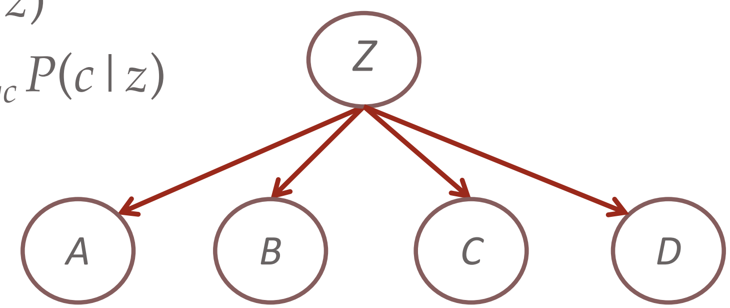
$$P(B|j, m) = \alpha P(B, j, m)$$

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

Order matters

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



- Elimination Order: Z, C, B, A
  - $P(D) = \alpha \sum_{a,b,c,z} P(a \mid z)\, P(b \mid z)\, P(c \mid z)\, P(D \mid z)\, P(z)$
  - $= \alpha \sum_a \sum_b \sum_c \sum_z P(a \mid z)\, P(b \mid z)\, P(c \mid z)\, P(D \mid z)\, P(z)$
  - 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

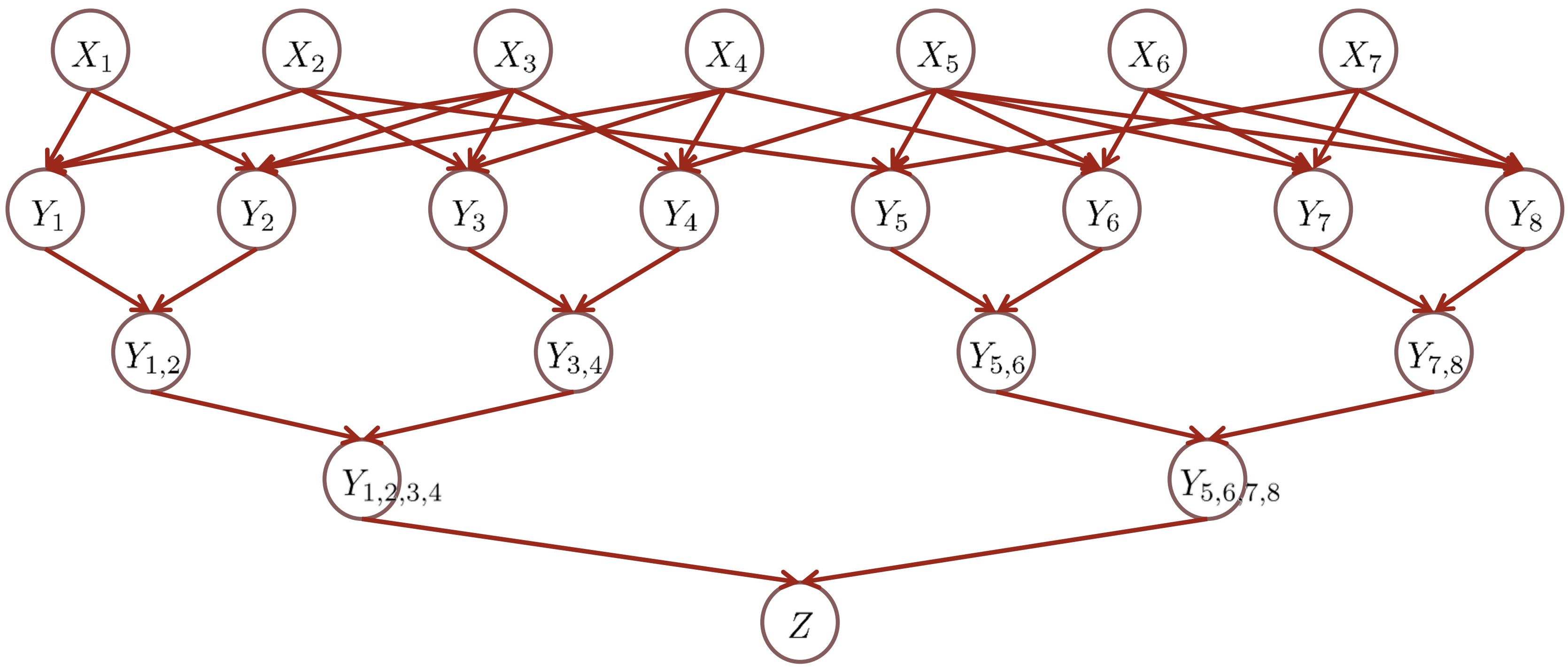
- The computational and space complexity of variable elimination is determined by the largest factor (and it's space that kills you)

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

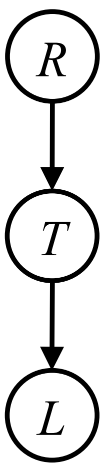
- Does there always exist an ordering that only results in small factors?
  - No!

VE: Computational and Space Complexity

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



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

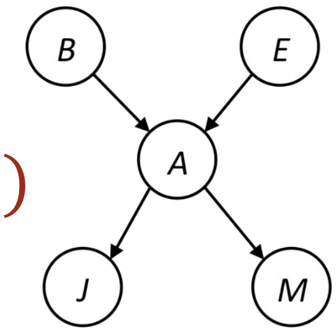
Join on t

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

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

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 (next lecture): Approximate Inference

Post-Lecture Poll

- Which one of the following statements is true?
  a) The variable elimination algorithm is slower than inference by enumeration
  b) The two algorithms are equally fast
  c) The variable elimination algorithm is faster than inference by enumeration