Warm-up as you come in

What is the notation behind these generic queries?

- What is the probability of *this* given what I know?
- What are the probabilities of all the possible outcomes (given what I know)?
- Which outcome is the most likely outcome (given what I know)?

Final Exam Review – Piazza Vote on timing

Announcements

Home Stretch

- P4 - Due Thursday 4/6
- HW9 - Due Tues 4/11 (day before Carnival)
- HW10 - Due Tues 4/25 (last week of school)
- P5 – Out 4/11, Due Thursday 4/27 (last week of school)
- Final exam – Thursday 5/4, 5:30-8:30pm

TA Applications! https://www.ugrad.cs.cmu.edu/ta/F23/instructor/

Carnival

- There will be lecture on 4/11 on HMMs
- No Recitation Friday 4/14, videos posted

Bayes Nets

✔ Part I: Representation and Independence

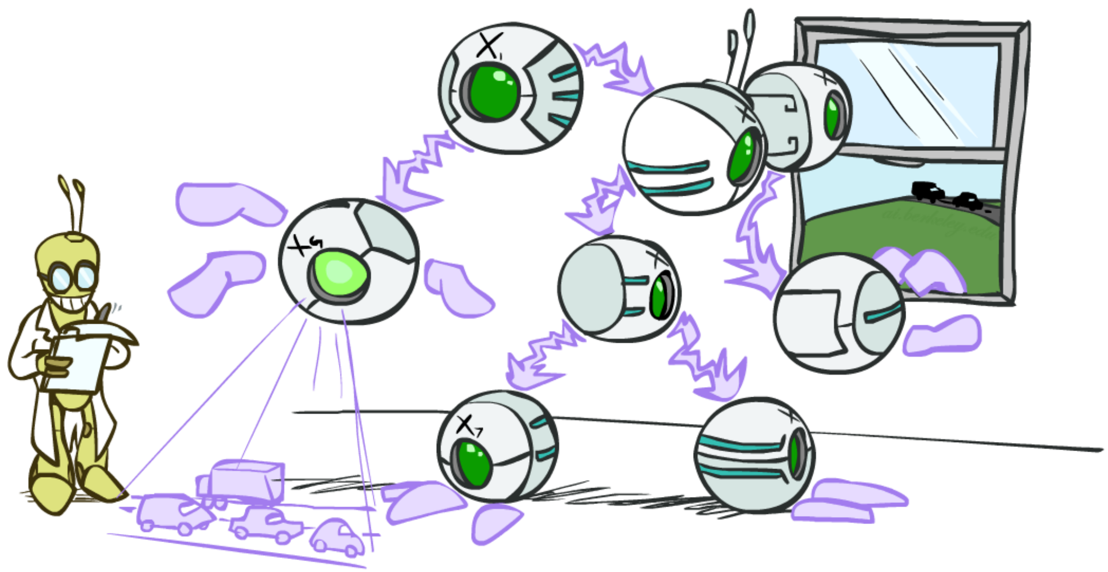
Part II: Exact inference

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

Part III: Approximate Inference

AI: Representation and Problem Solving

Bayes Nets Inference



Instructor: Stephanie Rosenthal

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

Warm-up as you come in

What is the notation behind these generic queries?

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

  $P(q \mid e)$

- 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)?

  $\underset{q \in Q}{\operatorname{argmax}} P(q \mid e)$

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)?

$$\mathrm{argmax}_{q \in Q}\, P(\, q \mid e \,)$$

Queries

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

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

single value

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

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

table

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

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

table

Queries

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

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

joint

$\sum_q P(q, e)$ $\sum_{q,h_1,h_2} P(q, h_1, h_2, e)$

- 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)?

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

Poll 1

If we only have the joint table $P(Q, H_1, H_2, E)$, how many times do we have to compute $P(e)$ to build $P(\,Q \mid e\,)$?

$\sum_q P(q, e)$

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

A) 0

B) 1

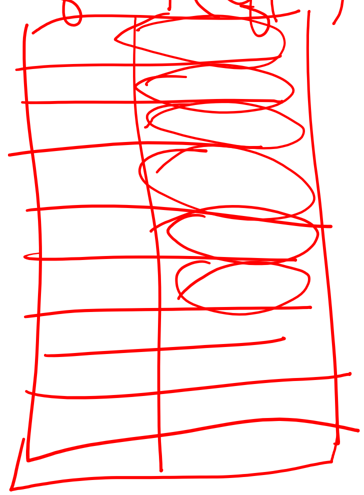
C) 10

D) 30

E) 200

F) 600

$q \quad P(q \mid e)$



- 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 1

If we only have the joint table $P(Q, H_1, H_2, E)$, how many times do we have 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

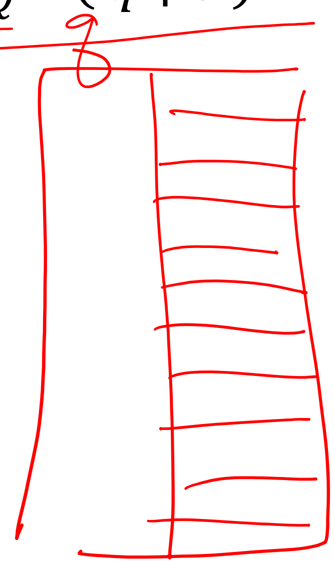
- Q can take on 10 different values
- H_1 can take on 4 different values
- H_2 can take on 5 different values
- E can take on 3 different values

Poll 2

$P(q|e) \propto P(q,e)$

If we only have the joint table $P(Q, H_1, H_2, E)$, how many times do we have to compute $P(e)$ to compute the following?

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



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

Poll 2

If we only have the joint table $P(Q, H_1, H_2, E)$, how many times do we have 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 $P(e)$

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

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

$$Z = P(e)$$

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

$$\alpha = 1/P(e)$$

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

$$\propto P(Q, e)$$

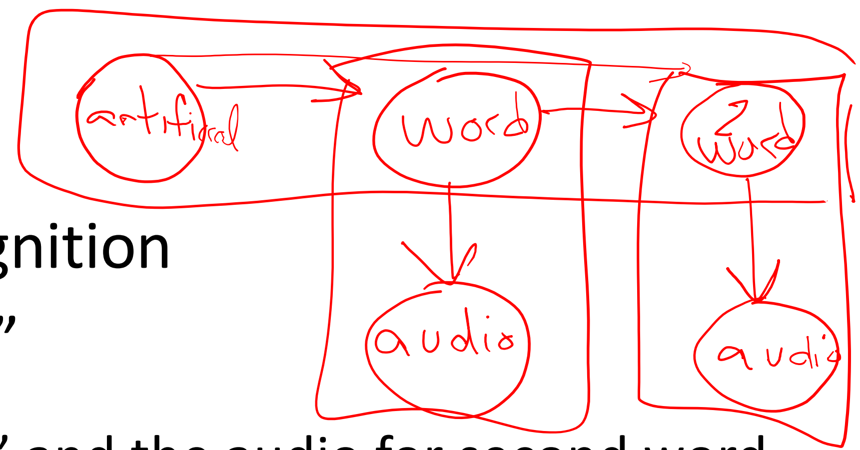
Bayes Nets in the Wild

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?

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

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

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

P(artificial)

<u>Break down problem</u>

n-gram probability * audio probability

prior posterior

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

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

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

Bayes Nets in the Wild

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

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

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

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

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

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

n-gram probability * audio probability

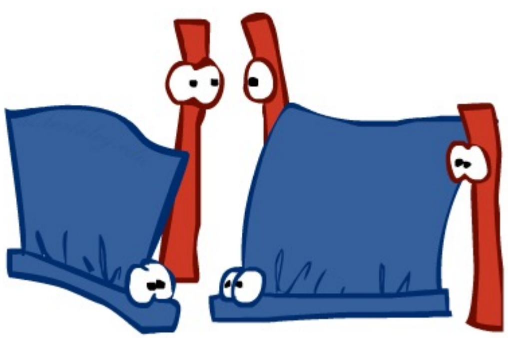
chain rule

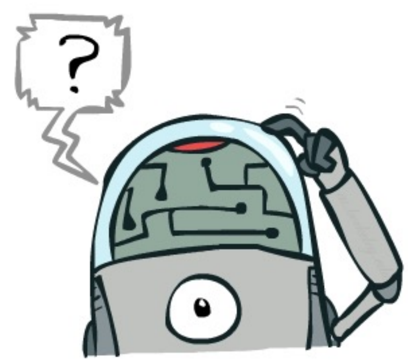
bayes net

Inference

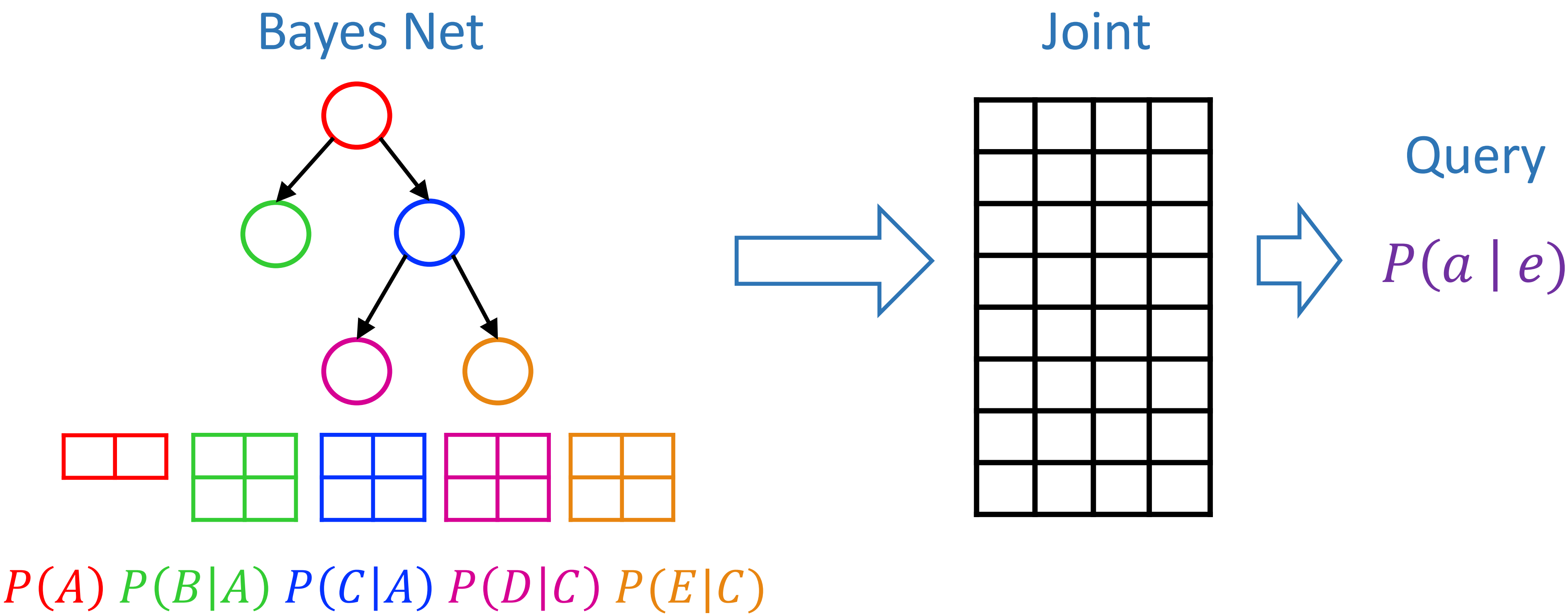
Inference: calculating some useful quantity from a probability model (joint probability distribution)

- 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?

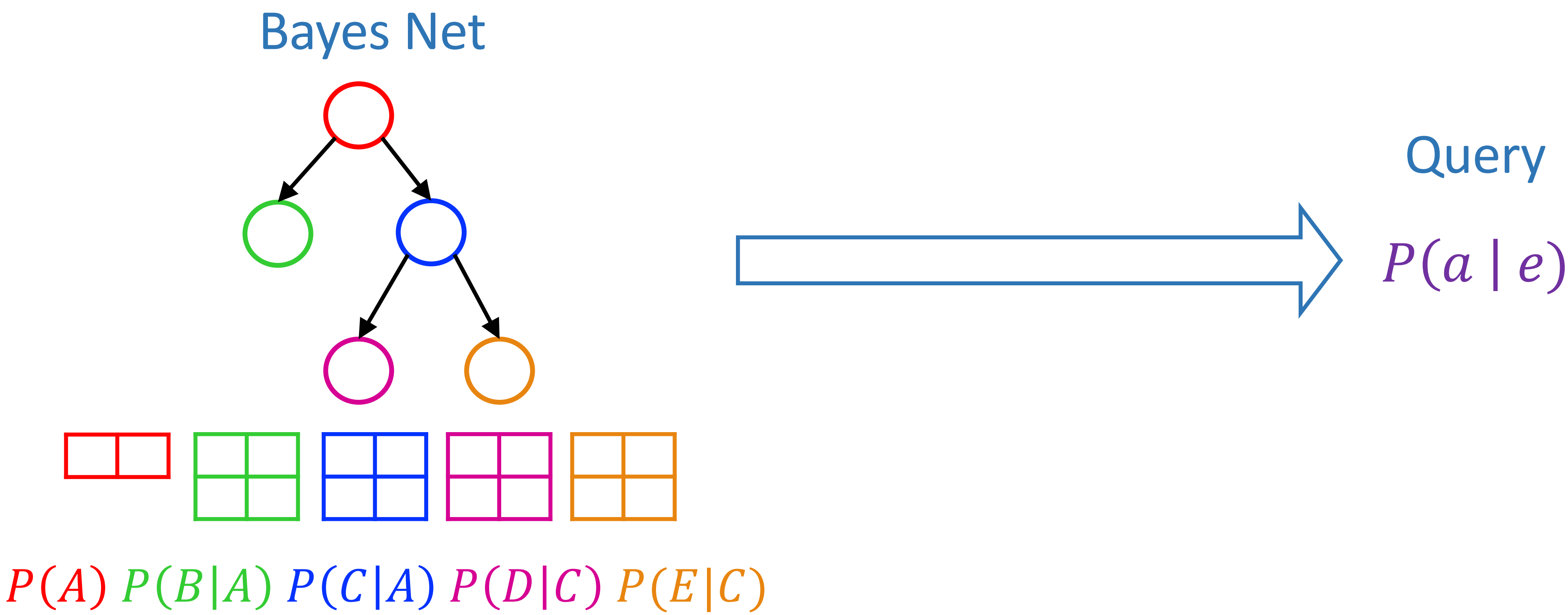




Answer Any Query from Bayes Net

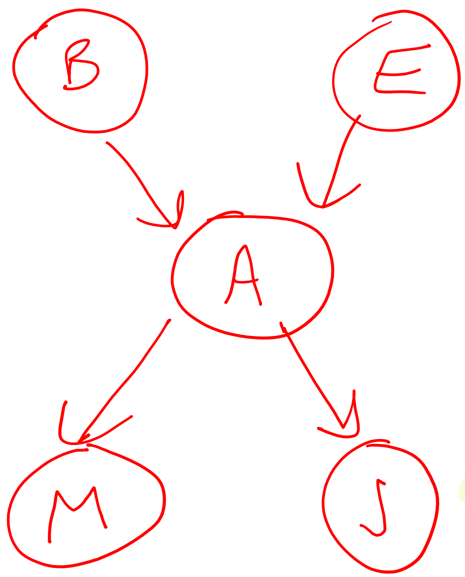


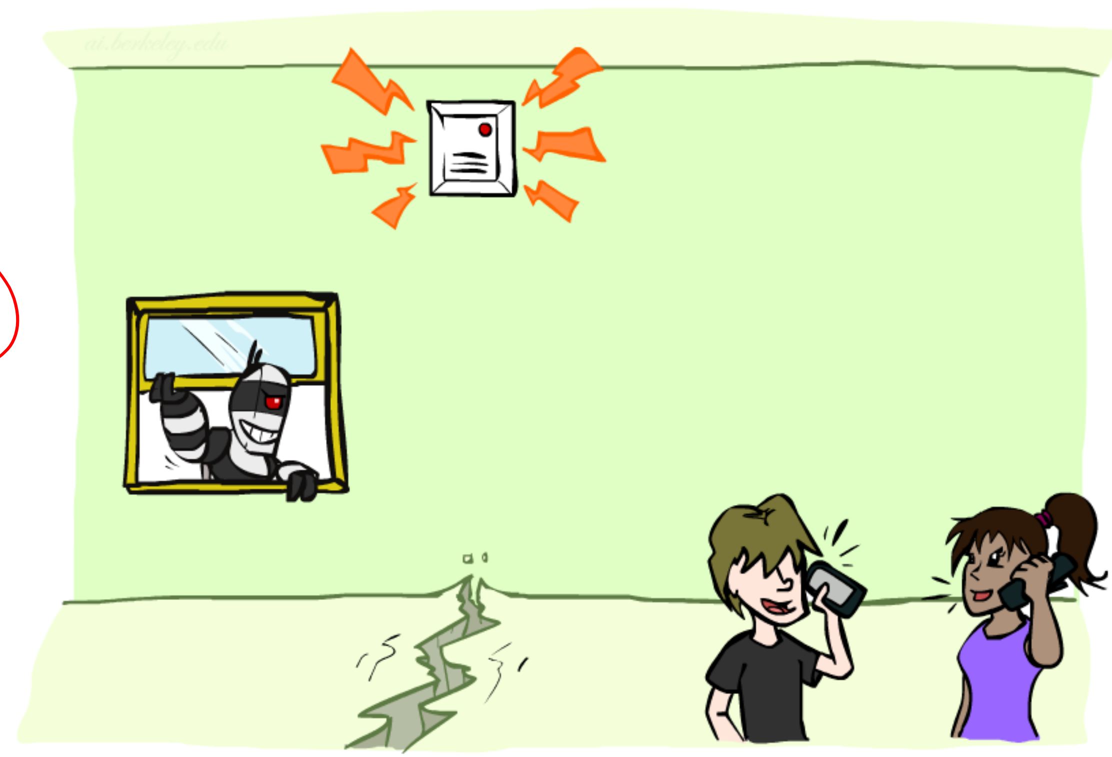
Next: Answer Any Query from Bayes Net



Example: Alarm Network

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





Example: Alarm Network

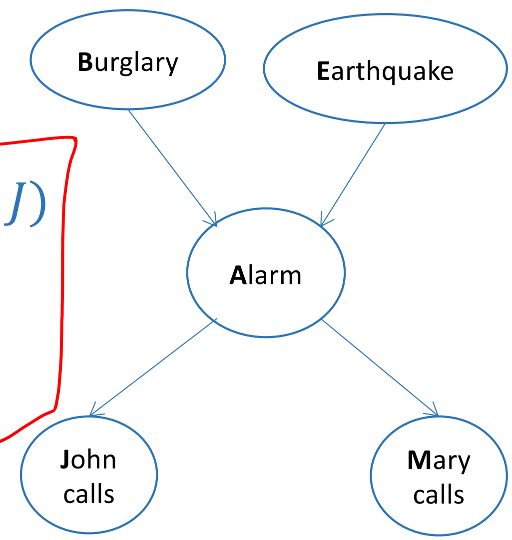
Joint distribution factorization example

Generic chain rule

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

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

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



Bayes nets

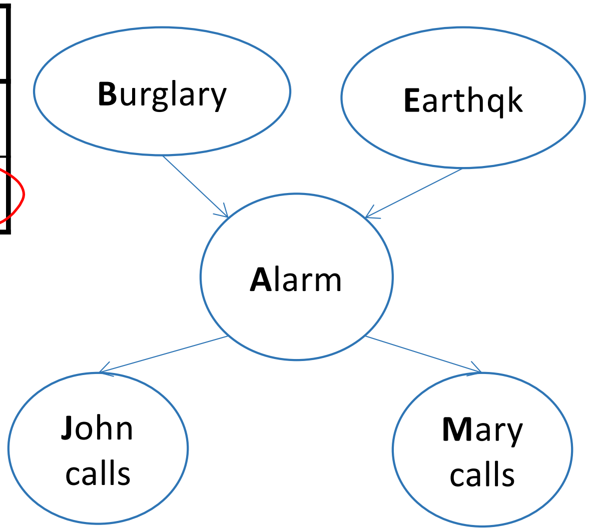
- $P(X_1 \ldots X_2) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

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

4 multiplies

B	P(B)
+b	0.001
-b	0.999

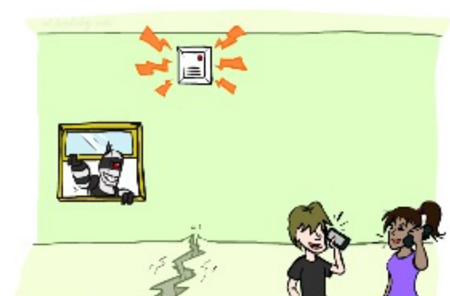


E	P(E)
+e	0.002
-e	0.998

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

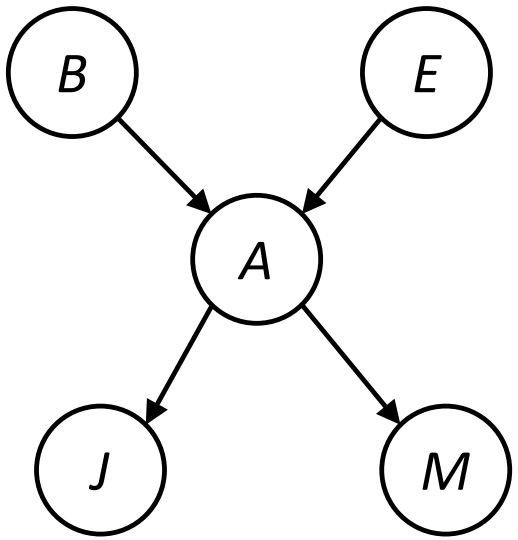
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99



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

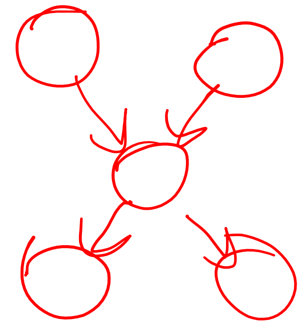


$$P(B \mid j, m) = \alpha\, P(B, j, m)$$
$$= \alpha \textstyle\sum_{e,a} P(B, e, a, j, m)$$
$$= \alpha \textstyle\sum_{e,a} P(B)\, P(e)\, P(a \mid B,e)\, P(j \mid a)\, P(m \mid a)$$

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

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

Can we do better?



$$\sum_e \sum_a \underbrace{P(B)\, P(e)\, P(a \mid B, e)\, P(j \mid a)\, P(m \mid a)}_{4}$$

+b	
-b	

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

- Lots of repeated subexpressions!

Can we do better?

$x_1(y_1(\quad)+y_2(\quad)) + x_2(y_1(\quad)+y_2(\quad))$

Consider

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

$2 \times 8 + 7$

Rewrite as

- $(x_1 + x_2)(y_1 + y_2)(z_1 + z_2)$

Inference Overview

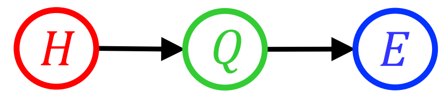
Given random variables Q, H, E (query, hidden, evidence)

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 in to CPT factors



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

Variable Elimination

But we can be more efficient

$F_1(q)$

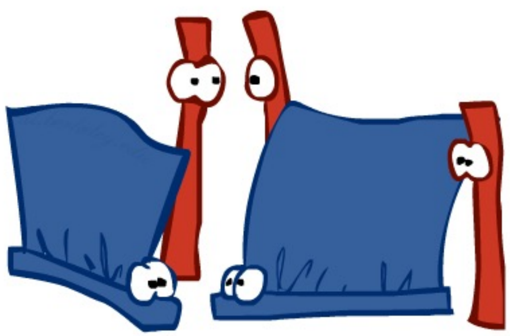
$$P(q|e) = \alpha\, P(e|q) \boxed{\sum_{h\in\{h_1,h_2\}} P(h) P(q|h)} \leftarrow P(q)$$
$$= \alpha\, P(e|q)\, [P(h_1)P(q|h_1) + P(h_2)P(q|h_2)]$$
$$= \alpha\, P(e|q)\, P(q)$$

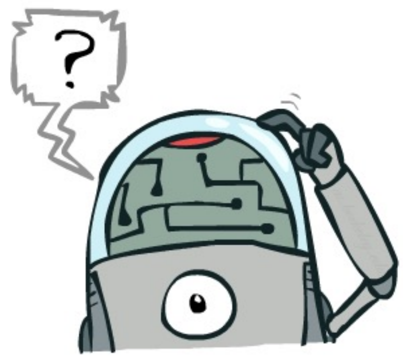
Now just extend to larger Bayes nets and a variety of queries

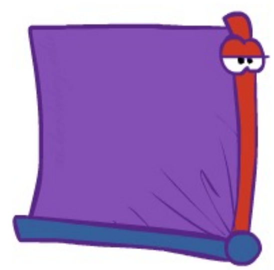
Factor Tables

$$\begin{aligned} P(+b,-e,-a,-j,-m) &= P(+b) \ast P(-e) \ast P(-a|+b,-e) \ast P(-j|-a) \ast P(-m|-a) \\ &= 0.001 \ast 0.998 \ast 0.06 \ast 0.95 \ast 0.99 \end{aligned}$$

$$\begin{aligned} P(+b,-e,-a,-j,-m) &= P(-e) \ast P(-a|+b,-e) \ast P(+b)\,P(-j|-a) \ast P(-m|-a) \\ &= 0.998 \ast 0.06 \ast 0.0095 \ast 0.99 \end{aligned}$$







Example: Alarm Network

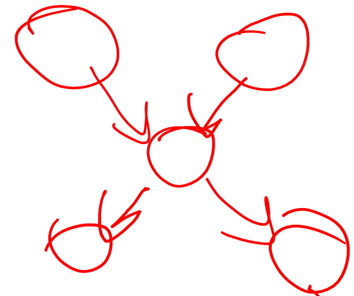
$$P(+b,-e,-a,-j,-m) = P(+b) \ * \ P(-e) \ * \ P(-a|+b,-e) \ * \ P(-j|-a) \ * \ P(-m|-a)$$
$$= 0.001 \ * \ 0.998 \ * \ 0.06 \ * \ 0.95 \ * \ 0.99$$

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

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

- 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

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

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

Variable elimination: The basic ideas

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

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

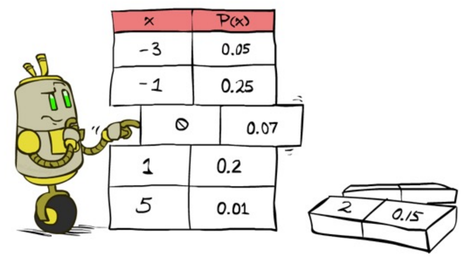
$$= \alpha \sum_e \sum_a P(j|a)\, P(e)\, P(m|a)\, P(a|B,e)\, \boxed{P(B)}$$

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

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

Variable Elimination

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



Start with initial factors:

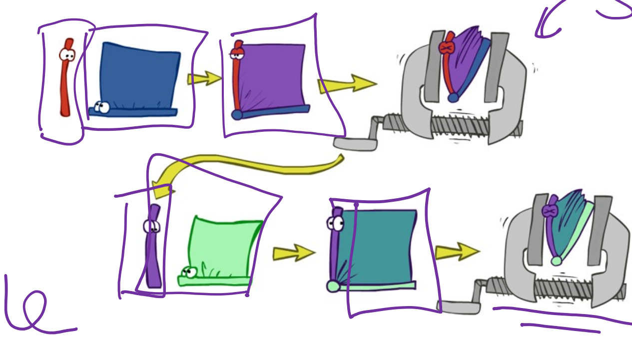
- Local CPTs (but instantiated by evidence)

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

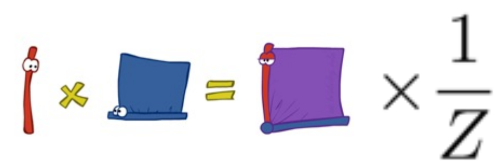
table

factor table

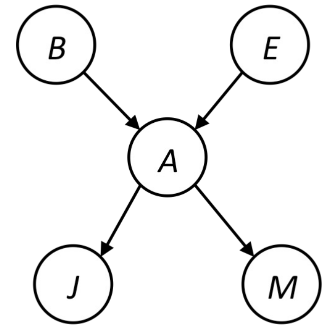


sum

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.

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

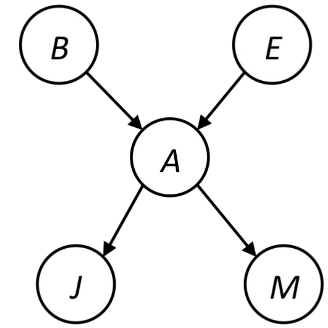
Example

+b	
-b	

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

constant constant

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



Choose A (inner most sum)

Create a table $t_1 = P(A|B,e)P(j|A)P(m|A)$

How many entries does each table for each value of A have?

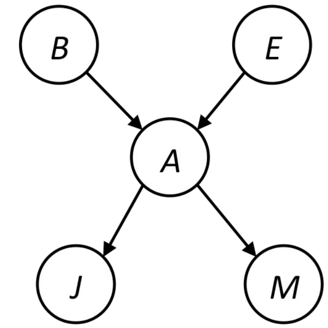
$$t_1 = \alpha P(B) \sum_{e} P(e) \sum_{a} t_1(A,B,e,j,m)$$

2 × 2 × 2

+b -b +e -e

+b -b +e -e +a -a

Example



Query *P(B | 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?

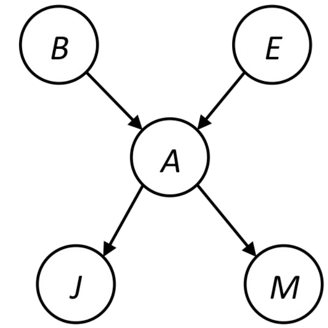
$$4 = \alpha P(B) \sum_{e} P(e) f_1(B, e, j, m)$$

$t_1(+b, +e, +a) + t_1(+b, +e, -a)$

+b	+e	
+b	-e	
-b	+e	
-b	-e	

Example

$$= \alpha P(B) \sum_e \underbrace{P(e)}_{2} \underbrace{f_1(B,e,j,m)}_{4}$$



Choose E (inner most sum)

Create a table $t_2 = P(e)\ f_1(B,e,j,m)$

How many entries does each table for each value of E have?

$$4 \quad t_2 = \alpha P(B) \sum_e t_2(B,e,j,m)$$

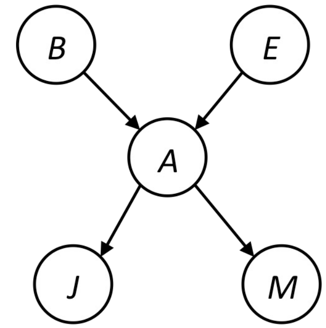
+b	+e	

+e	
-e	

×

+b	+e	
+b	-e	
-b	+e	
-b	-e	

Example



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

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?

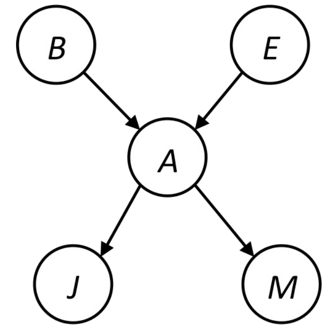
$$= \alpha P(B) f_2(B, j, m)$$

2

+b	
-b	

+b	+e	
+b	-e	
-b	+e	
-b	-e	

Example



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

+b	
-b	

+b	
-b	

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

How many entries does this probability table have? 2

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

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

$P(B,j,m) =$

+b	
-b	

$/ P(j,m) = P(B|j,m)$

$P(j,m) = \Sigma$

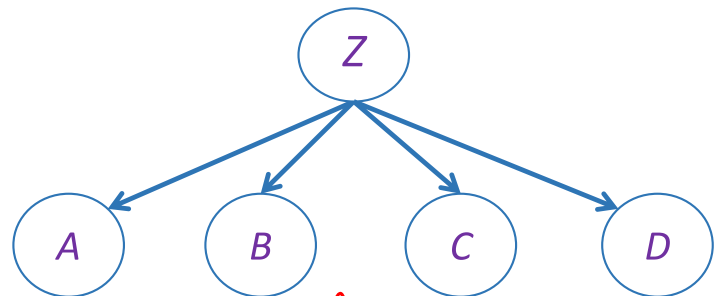
$$P(B|j,m) = \alpha P(B,j,m)$$

Order matters

$\sum_c P(c,z) = P(c_1,z) + P(c_2,z) + P(c_3,z)$

$P(z)$

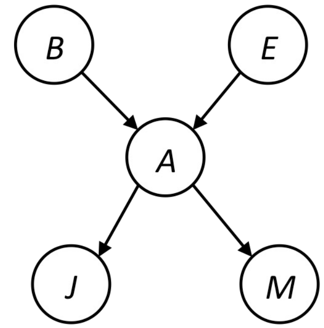
- Elimination Order: C, B, A, Z
  - $P(D) = \alpha \sum_{z,a,b,c} \boxed{P(D|z)\, P(z)\, P(a|z)\, P(b|z)\, P(c|z)}$
  - $\quad = \alpha \sum_z P(D|z)\, P(z) \sum_a P(a|z) \sum_b P(b|z) \sum_c P(c|z)$
  - Largest factor has 2 variables (D,Z)



- 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

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.

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

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

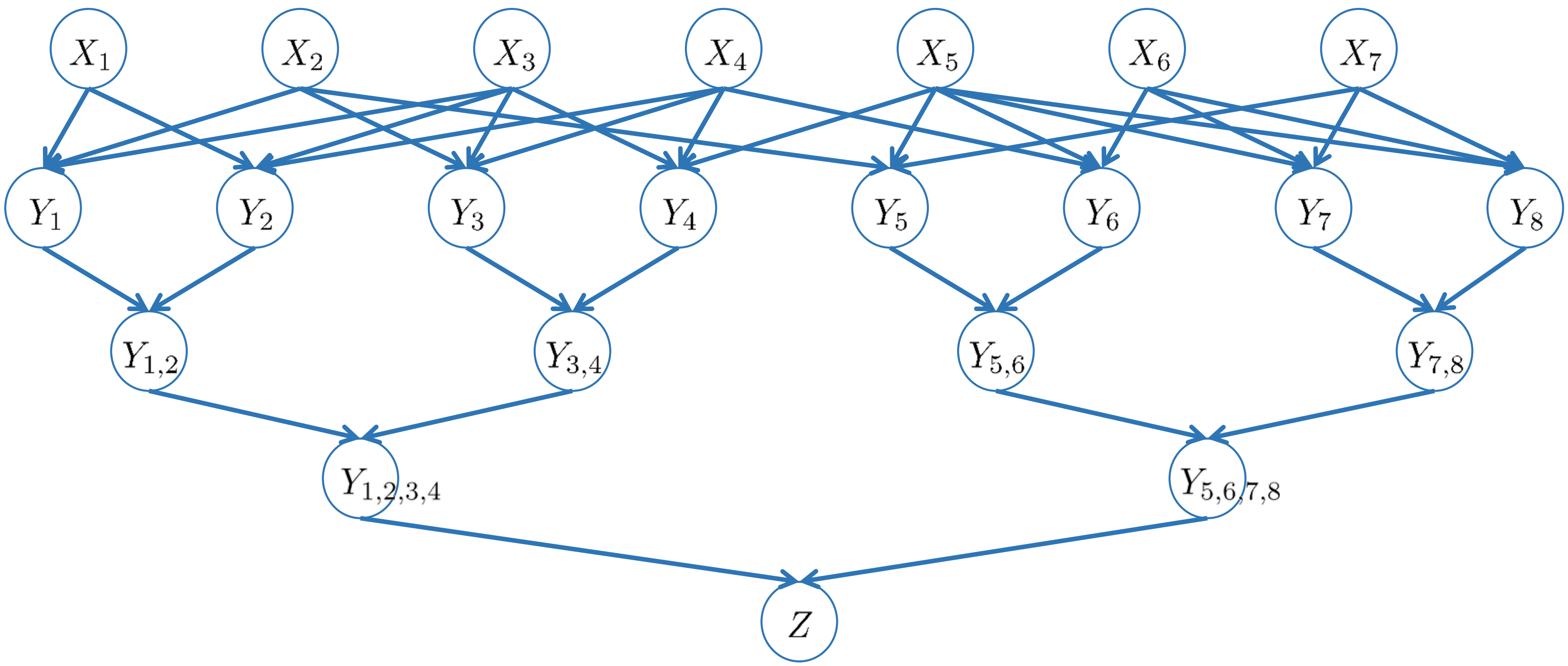
Does there always exist an ordering that only results in small factors?

- 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

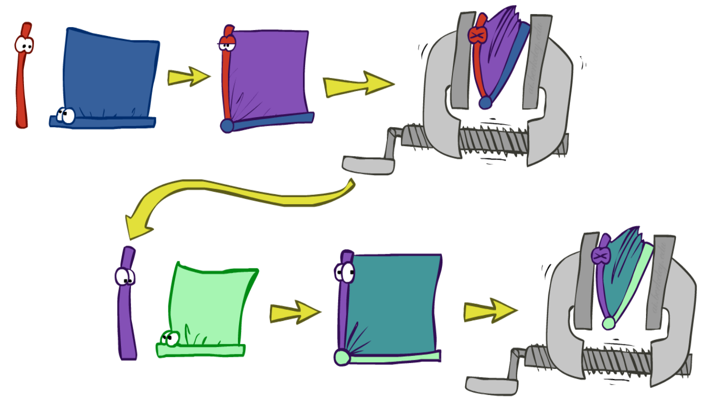
Additional Variable Elimination Slides

Variable elimination: The basic ideas

Move summations inwards as far as possible

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

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



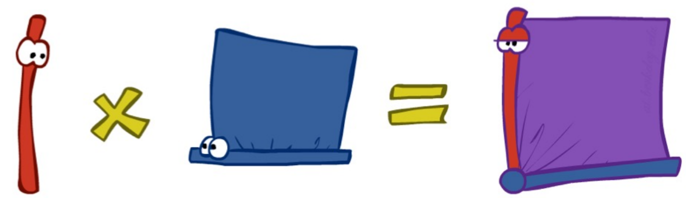
Do the calculation from the inside out

- I.e., sum over *a* first, then sum over *e*
- Problem: $P(a|B,e)$ 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

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: $P(A) \times P(J|A) = P(A,J)$

$P(A)$

true	0.1
false	0.9

x

$P(J|A)$

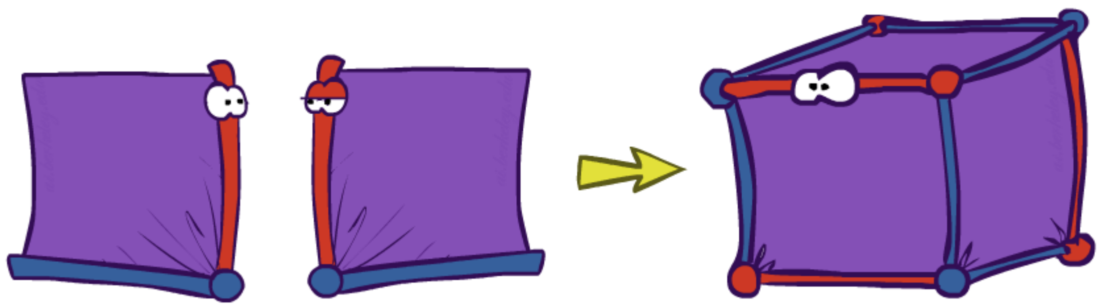
A \ J	true	false
true	0.9	0.1
false	0.05	0.95

=

$P(A,J)$

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Example: Making larger factors



Example: $P(J|A) \times P(M|A) = P(J,M|A)$

$P(J|A)$

A \ J	true	false
true	0.99	0.01
false	0.145	0.855

x

$P(M|A)$

A \ M	true	false
true	0.97	0.03
false	0.019	0.891

=

$P(J,M|A)$

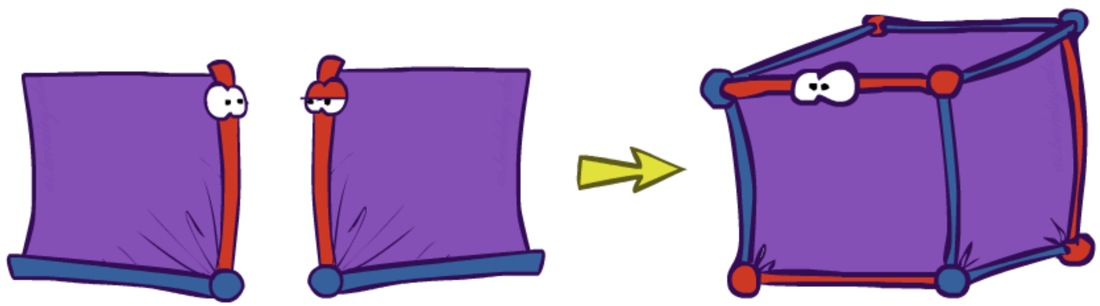
A=false

J \ M	true	false
true		18
false		

A=true

J \ M	true	false
true		
false		.0003

Example: Making larger factors



Example: $f_1(U,V)$ x $f_2(V,W)$ x $f_3(W,X)$ = $f_4(U,V,W,X)$

Sizes: [10,10] x [10,10] x [10,10] = [10,10,10,10]

I.e., 300 numbers blows up to 10,000 numbers!

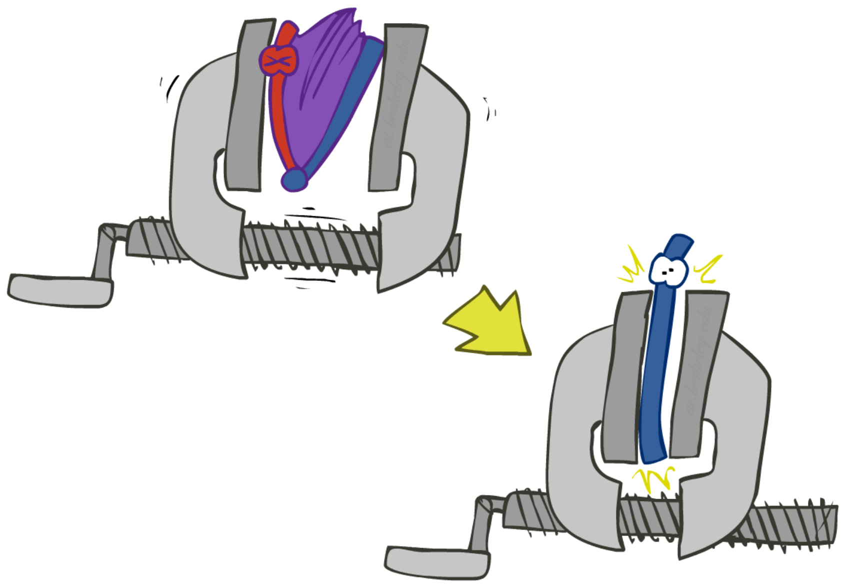
Factor blowup can make VE very expensive

Operation 2: Summing out a variable

Second basic operation: ***summing out***
(or eliminating) a variable from a factor

- Shrinks a factor to a smaller one

Example: $\sum_j P(A,J) = P(A,j) + P(A,\neg j) = P(A)$



$P(A,J)$

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Sum out J

$P(A)$

true	0.1
false	0.9

Summing out from a product of factors

Project the factors each way first, then sum the products

- Example: $\sum_a P(a \mid B,e)\ P(j \mid a)\ P(m \mid a)$

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

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

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