

Warm-up as you walk 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)?

Announcements

Assignments:

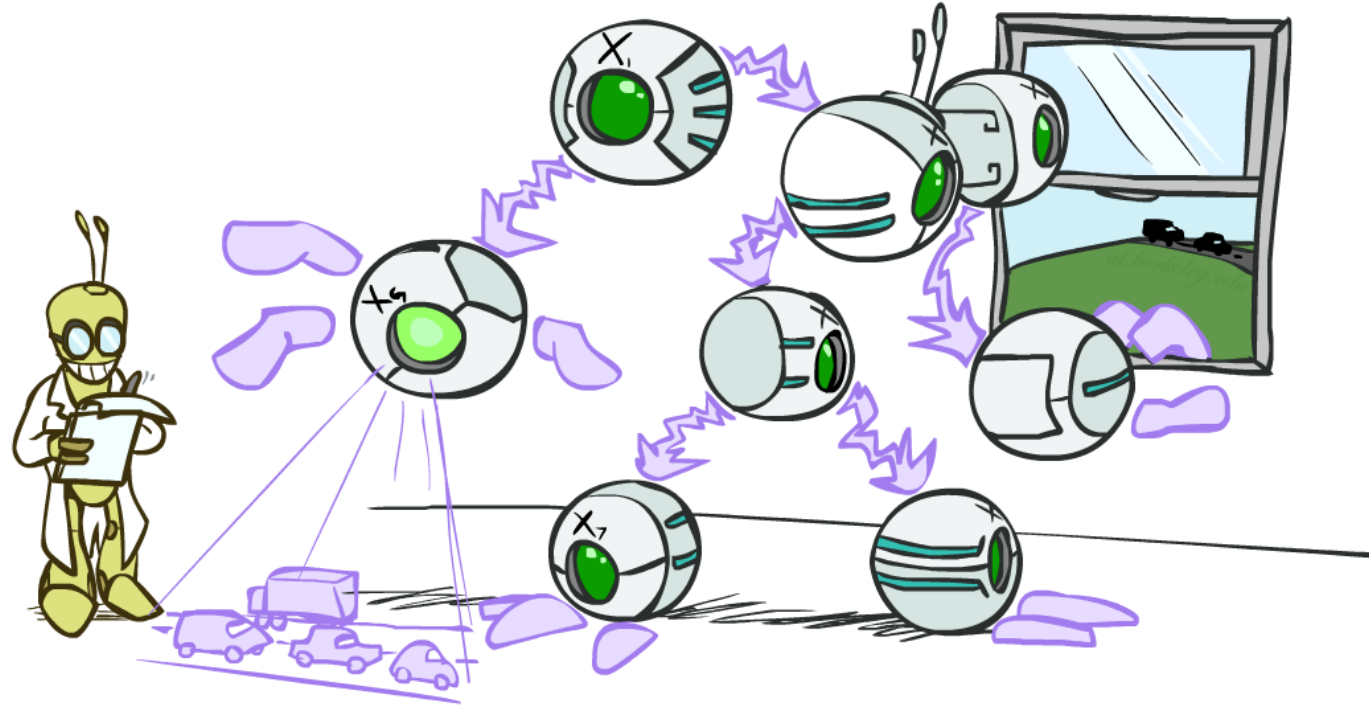
- HW9 (online)
 - Due tonight, 10 pm

Midterm:

- Tue 11/12, in-class
 - See Piazza for details
 - Practice exam
 - Friday: Recitation review
 - Sunday: Review practice exam solutions

AI: Representation and Problem Solving

Bayes Nets Inference



Instructors: Pat Virtue & Fei Fang

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

Warm-up as you walk 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)?

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)?

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

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)?

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

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)?

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

Piazza 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 $P(Q | e)$?

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

- A) 0
- B) 1
- C) 10
- D) 30
- E) 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

Piazza 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 $P(Q | e)$?

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

- A) 0
- B) 1
- C) 10
- D) 30
- E) 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

Piazza 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?

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

- A) 0
- B) 1
- C) 10
- D) 30
- E) 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

Piazza 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?

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

- A) 0
- B) 1
- C) 10
- D) 30
- E) 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

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

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

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

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

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

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.

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

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

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

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

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

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

$P(\mathbf{flavoring} \mid \text{artificial})_{15} * P(\text{audio} \mid \mathbf{flavoring})$

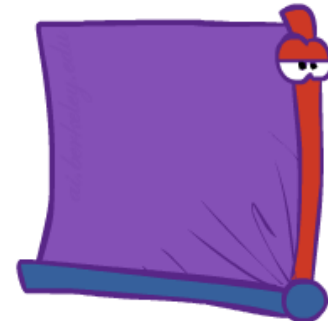
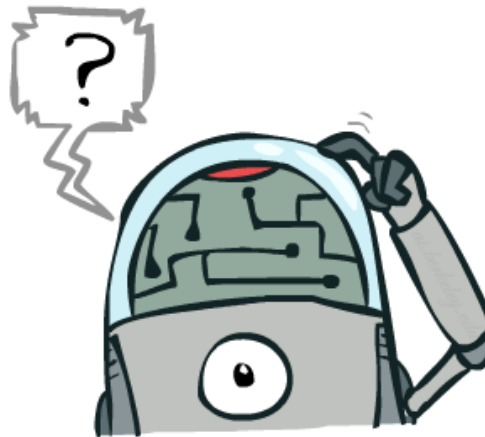
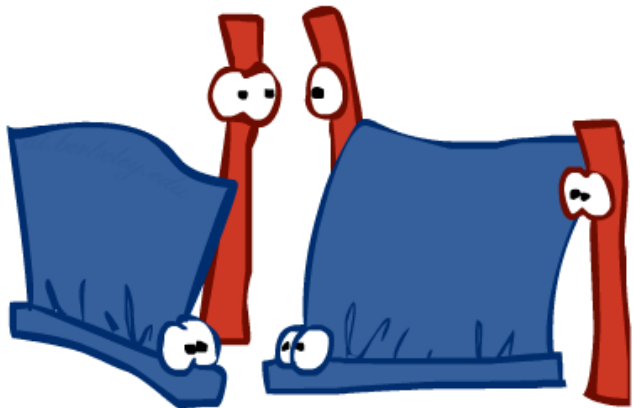
Bayes Nets in the Wild

$$\begin{aligned} \text{second}^* &= \operatorname{argmax}_{\text{second}} P(\text{second} \mid \text{artificial}, \text{audio}) \\ &= \operatorname{argmax}_{\text{second}} \frac{P(\text{second}, \text{artificial}, \text{audio})}{P(\text{artificial}, \text{audio})} \\ &= \operatorname{argmax}_{\text{second}} P(\text{second}, \text{artificial}, \text{audio}) \\ &= \operatorname{argmax}_{\text{second}} P(\text{artificial}) P(\text{second} \mid \text{artificial}) P(\text{audio} \mid \text{artificial}, \text{second}) \\ &= \operatorname{argmax}_{\text{second}} P(\text{artificial}) P(\text{second} \mid \text{artificial}) P(\text{audio} \mid \text{second}) \\ &= \operatorname{argmax}_{\text{second}} P(\text{second} \mid \text{artificial}) P(\text{audio} \mid \text{second}) \\ &\quad \text{n-gram probability} * \text{audio probability} \end{aligned}$$

Inference

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

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



Inference Overview

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

We know how to do inference on a joint distribution

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

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

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



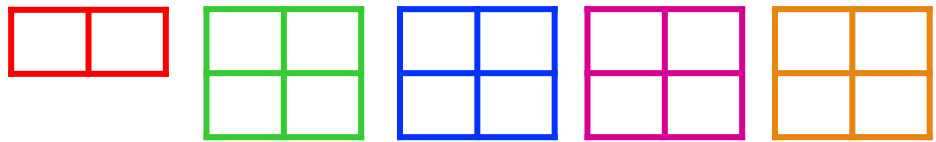
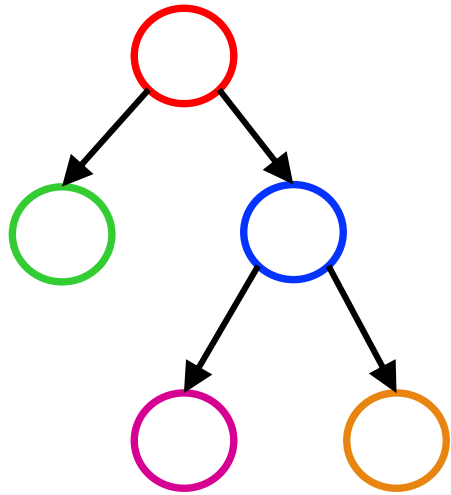
But we can be more efficient

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

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

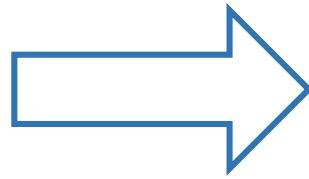
Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Joint



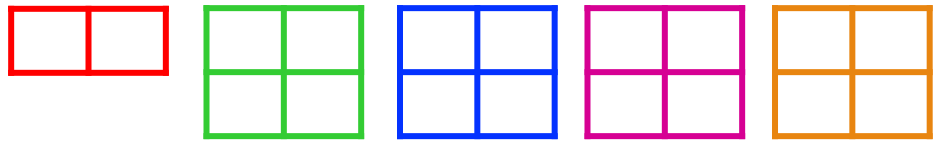
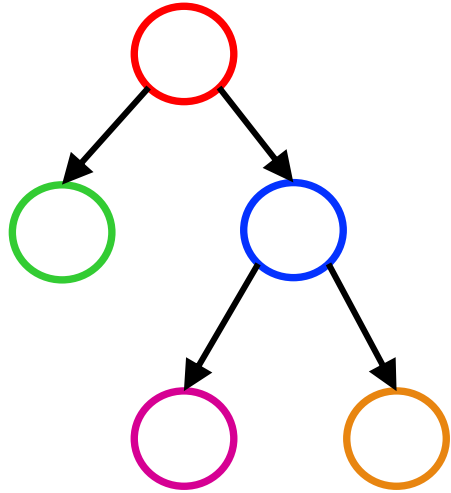


Query

$P(a | e)$

Next: Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Query

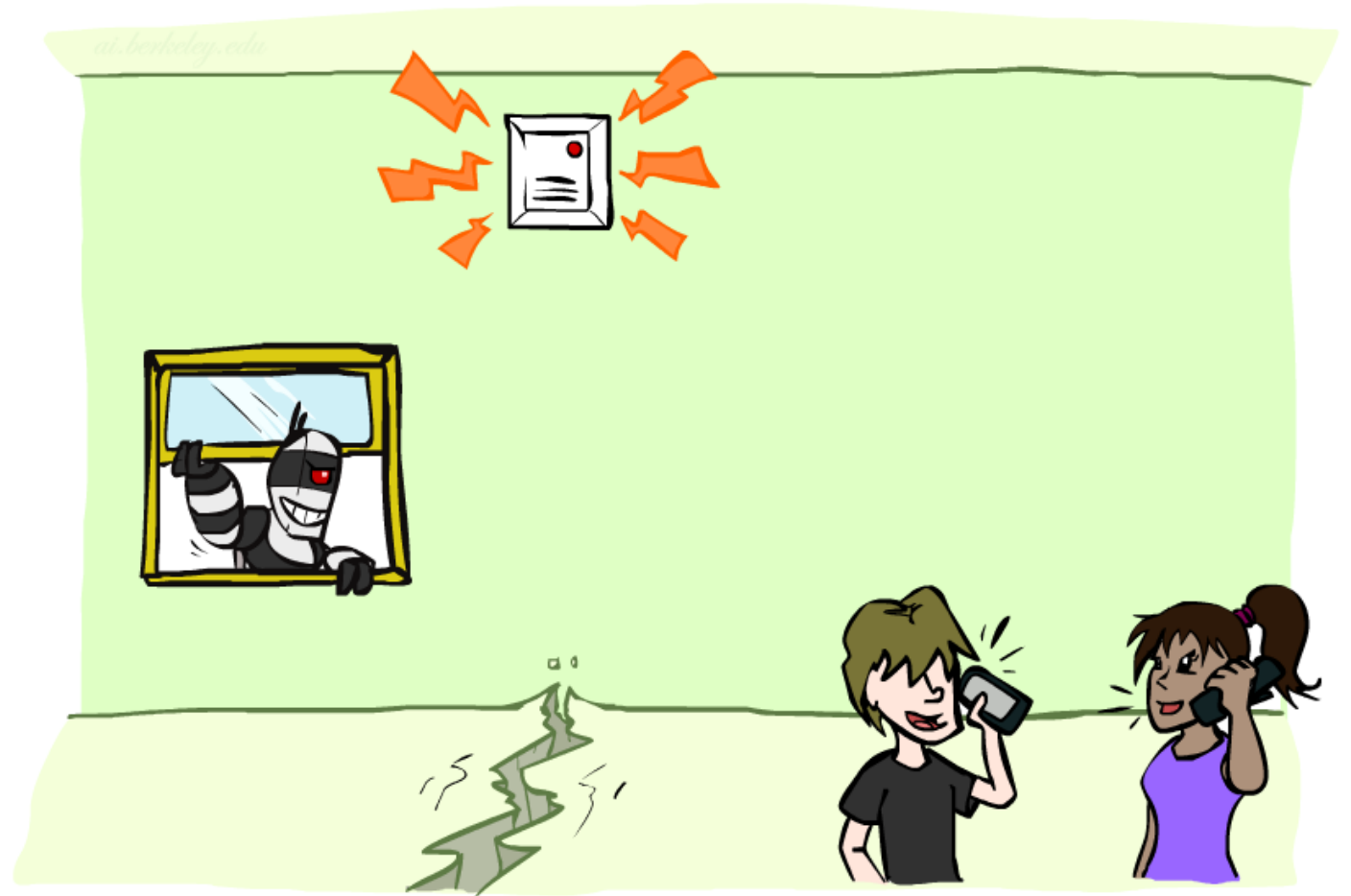


$P(a | e)$

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

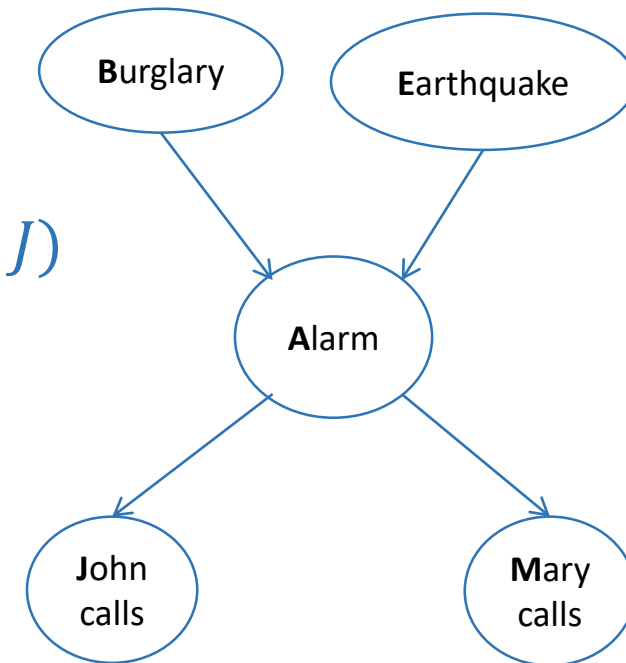
$$\blacksquare P(X_1 \dots X_n) = \prod_i P(X_i | X_1 \dots X_{i-1})$$

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

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

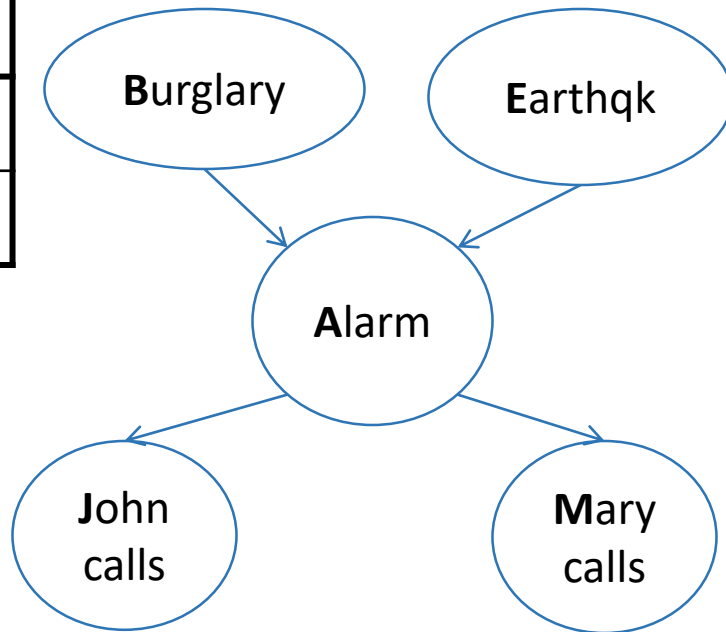
Bayes nets

$$\blacksquare P(X_1 \dots X_n) = \prod_i P(X_i | \text{Parents}(X_i))$$



Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999

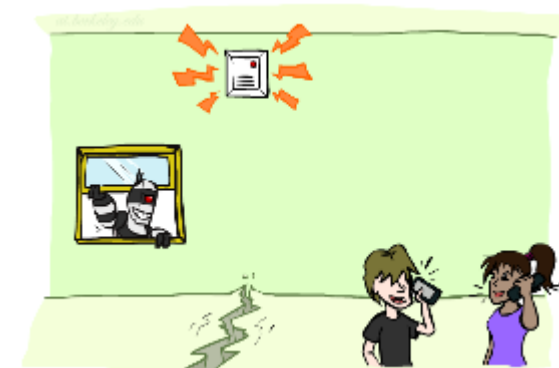


E	P(E)
+e	0.002
-e	0.998

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

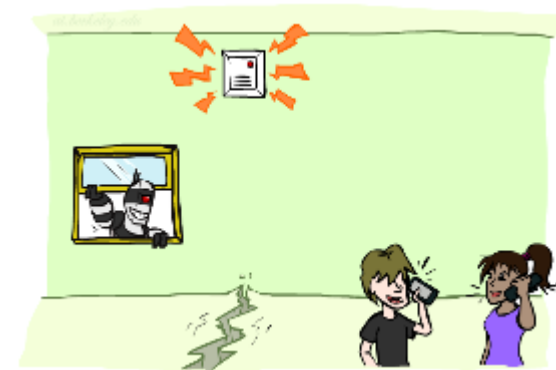
B	E	A	P(A B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999



Example: Alarm Network

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

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



Example: Alarm Network

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

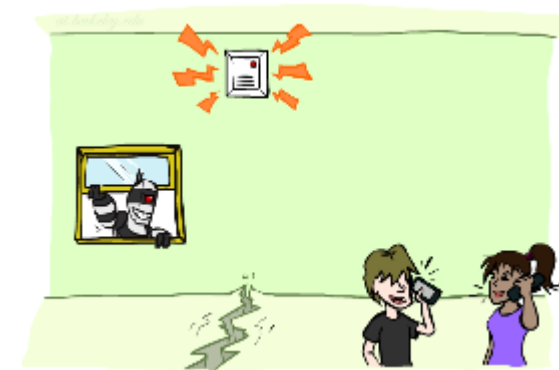
$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(-e) * P(-a|+b, -e) * P(+b) * P(-j|-a) * P(-m|-a) \\ &= 0.998 * 0.06 * 0.001 * 0.95 * 0.99 \end{aligned}$$

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

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

Factors

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



Factors

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_1(+b) * f_2(-e) * f_3(-a, +b, -e) * f_4(-j, -a) * f_5(-m, -a) \\ &= 0.001 * 0.998 * 0.06 * 0.95 * 0.99 \end{aligned}$$

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_2(-e) * f_3(-a, +b, -e) * f_1(+b) * f_4(-j, -a) * f_5(-m, -a) \\ &= 0.998 * 0.06 * 0.001 * 0.95 * 0.99 \end{aligned}$$

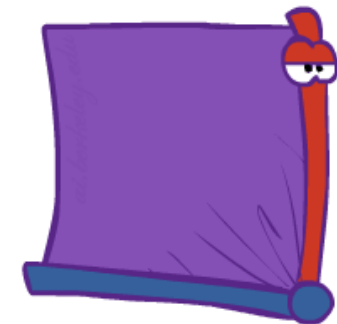
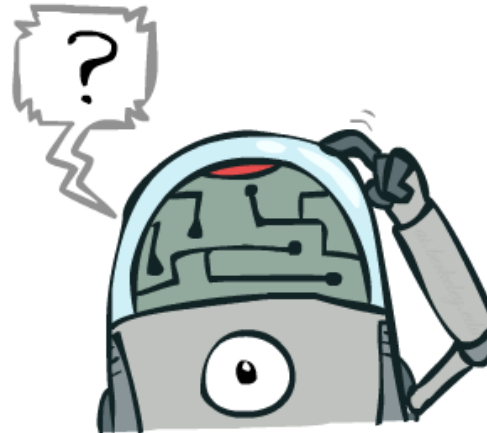
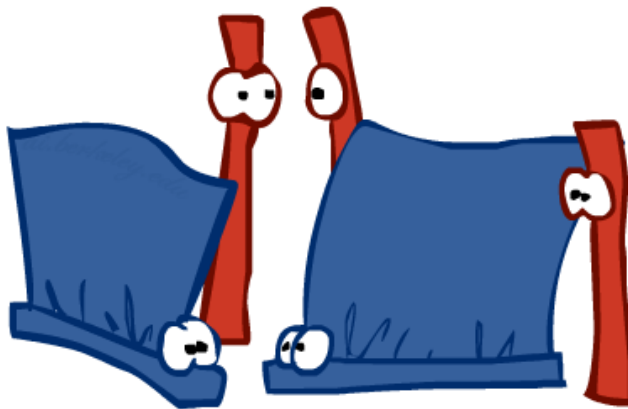
$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_2(-e) * f_3(-a, +b, -e) * f_6(+b, j, -a) * f_5(-m, -a) \\ &= 0.998 * 0.06 * 0.0095 * 0.99 \end{aligned}$$

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

Factor Tables

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

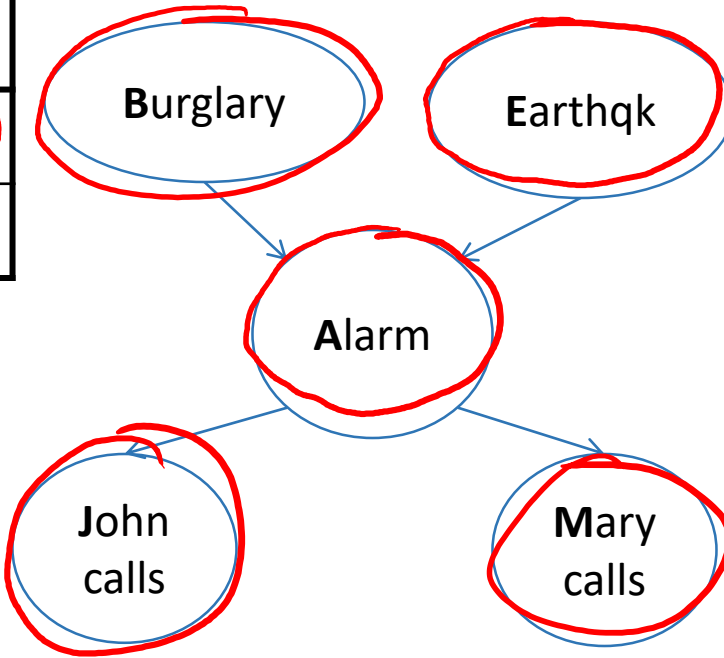
$$P(B, E, A, J, M) = P(B) * P(E) * P(A|B, E) * P(J|A) * P(M|A)$$



Example: Alarm Network

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

B	P(B)
+b	0.001
-b	0.999

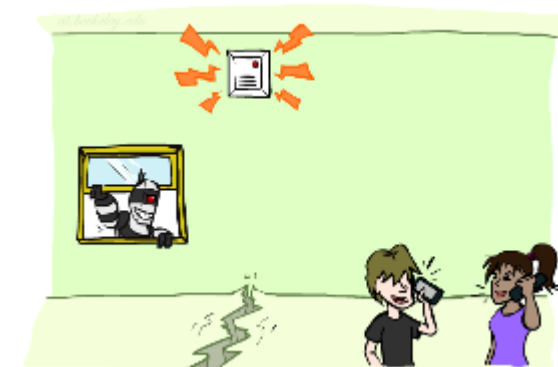


E	P(E)
+e	0.002
-e	0.998

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
<u>+b</u>	<u>-e</u>	<u>-a</u>	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

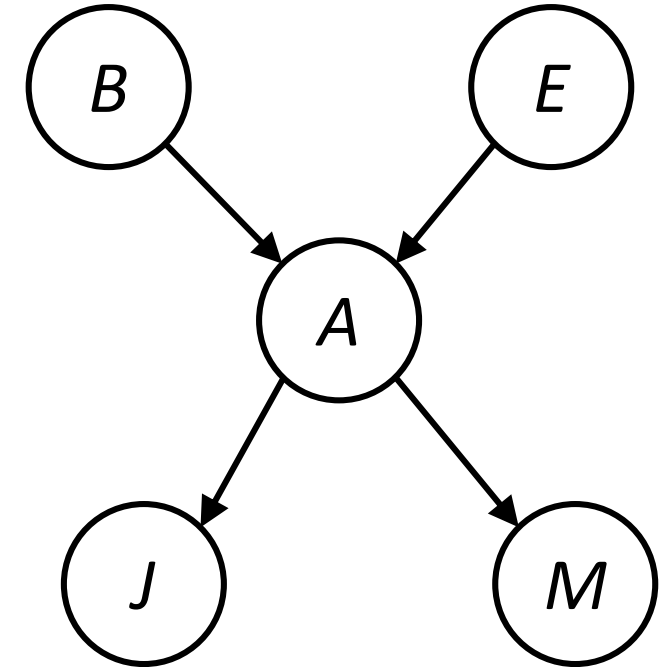
Reminder of inference by enumeration:

- 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

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

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

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



Can we do better?

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

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

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

- Lots of repeated subexpressions!

Variable elimination: The basic ideas

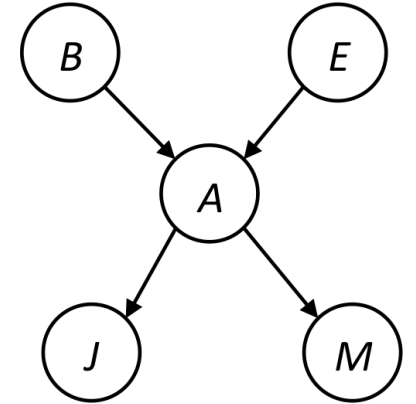
Move summations inwards as far as possible

$$\begin{aligned} P(B \mid j, m) &= \alpha \sum_e \sum_a P(B, e, a, j, m) \\ &= \alpha \sum_e \sum_a P(j|a) P(e) P(m|a) P(a|B, e) P(B) \\ &= \alpha P(B) \sum_e P(e) \sum_a P(j|a) P(m|a) P(a|B, e) \end{aligned}$$

Variable Elimination

```
function VariableElimination( $Q$  ,  $e$ ,  $bn$ ) returns a distribution over  $Q$   
   $factors \leftarrow []$   
  for each  $var$  in ORDER( $bn.vars$ ) do  
     $factors \leftarrow [MAKE-FACTOR(var,  $e$ ) |  $factors$ ]$   
    if  $var$  is a hidden variable then  
       $factors \leftarrow SUM-OUT(var, factors)$   
  return NORMALIZE(POINTWISE-PRODUCT( $factors$ ))
```

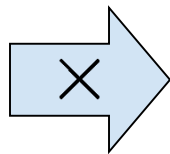
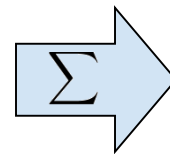
Example



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

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

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

Example

$P(B)$	$P(E)$	$P(j,m B,E)$
--------	--------	--------------

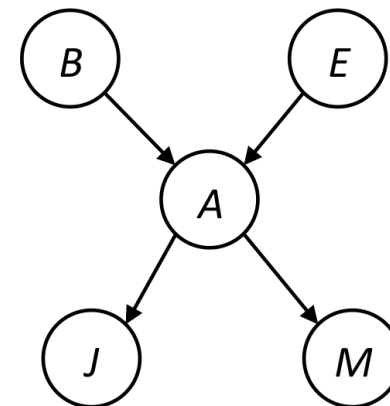
Choose E

$$\begin{array}{c} P(E) \\ P(j,m|B,E) \end{array} \xrightarrow{\times} P(E,j,m|B) \xrightarrow{\Sigma} P(j,m|B)$$

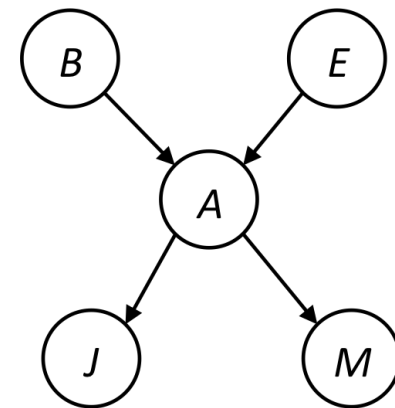
$P(B)$	$P(j,m B)$
--------	------------

Finish with B

$$\begin{array}{c} P(B) \\ P(j,m|B) \end{array} \xrightarrow{\times} P(j,m,B) \xrightarrow{\text{Normalize}} P(B | j,m)$$



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



Order matters

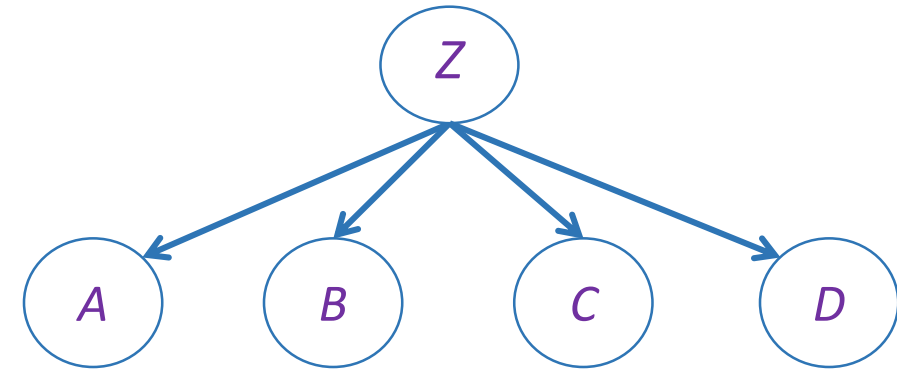
- Order the terms D, Z, A, B C

- $P(D) = \alpha \sum_{z,a,b,c} P(D|z) P(z) P(a|z) P(b|z) P(c|z)$
- $= \alpha \sum_z P(D|z) P(z) \sum_a P(a|z) \sum_b P(b|z) \sum_c P(c|z)$
- Largest factor has 2 variables (D,Z)

- Order the terms A, B C, D, Z

- $P(D) = \alpha \sum_{a,b,c,z} P(a|z) P(b|z) P(c|z) P(D|z) P(z)$
- $= \alpha \sum_a \sum_b \sum_c \sum_z P(a|z) P(b|z) P(c|z) P(D|z) P(z)$
- 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

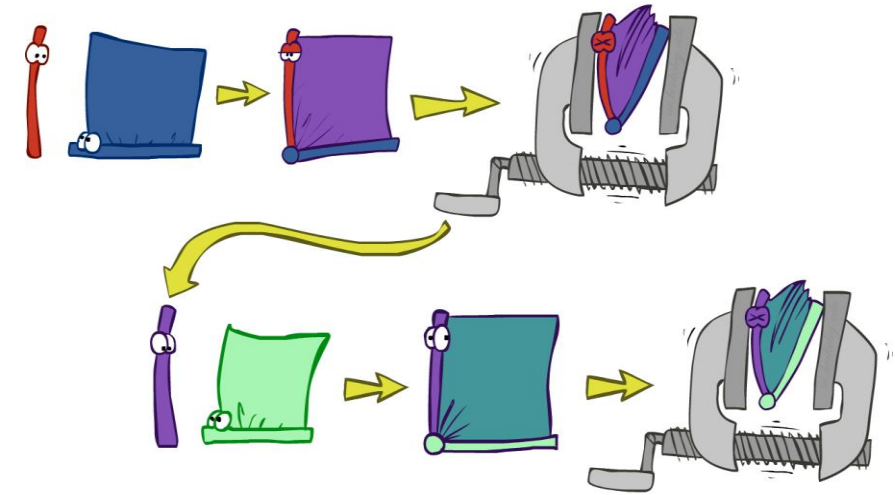
Does there always exist an ordering that only results in small factors?

- No!

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**

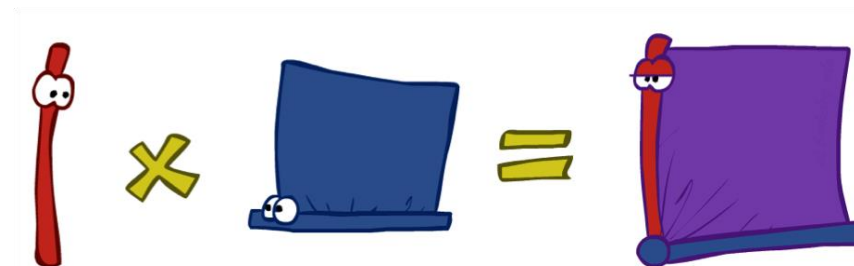
Operation 1: Pointwise product

First basic operation: **pointwise product** of factors
(similar to a **database join**, **not** matrix multiply!)

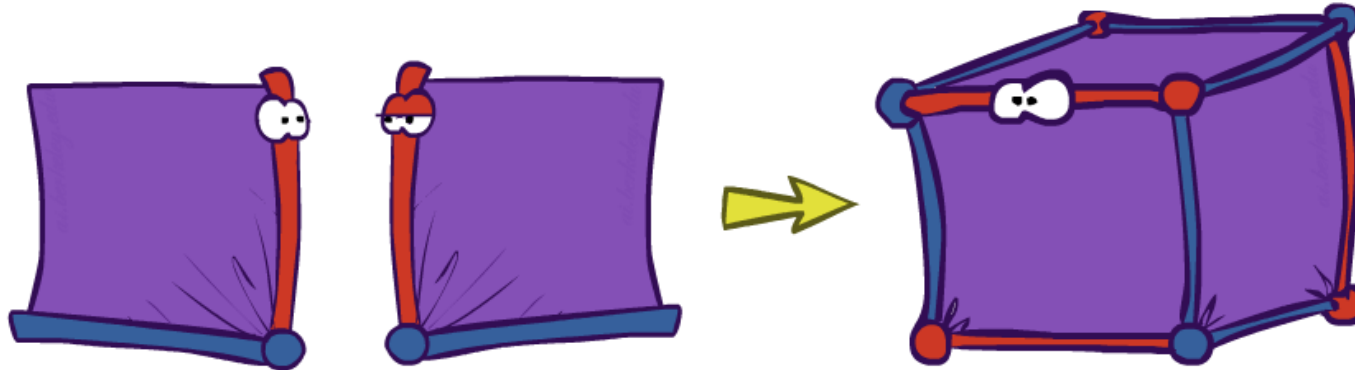
- New factor has **union** of variables of the two original factors
- Each entry is the product of the corresponding entries from the original factors

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

$P(A)$		$\times$	$P(J A)$			$=$	$P(A,J)$		
A \ J	true		false	A \ J	true		false		
true	0.1		0.9	0.1	true		0.09	0.01	
false	0.9		0.05	0.95	false		0.045	0.855	



Example: Making larger factors



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

$P(J/A)$

$A \setminus J$	true	false
true	0.99	0.01
false	0.145	0.855

$\times$

$P(M/A)$

$A \setminus M$	true	false
true	0.97	0.03
false	0.019	0.891

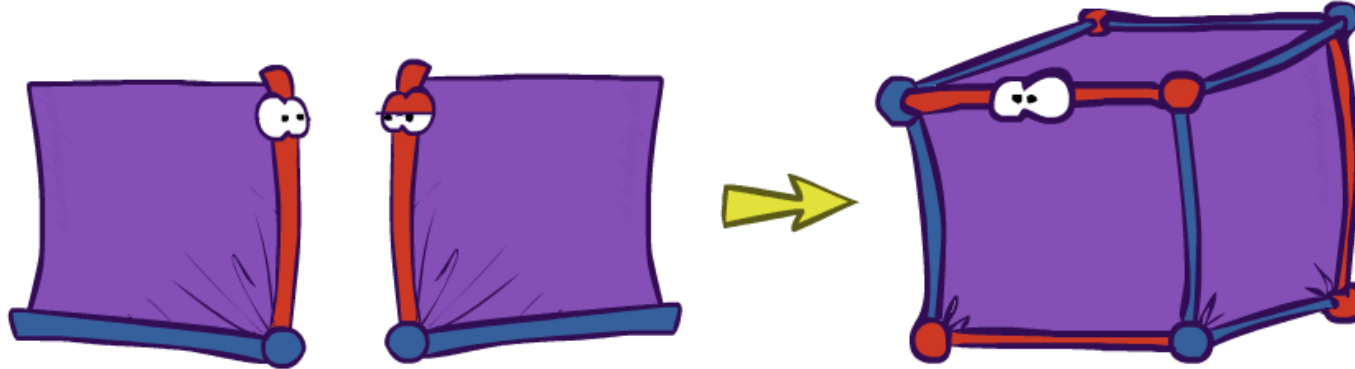
$=$

$P(J,M/A)$

	$J \setminus M$		
	true	false	
$J \setminus M$	true	false	
true			18
false		.0003	

A=false
A=true

Example: Making larger factors



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

Sizes: $[10,10] \times [10,10] \times [10,10] = [10,10,10,10]$

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

$P(A, J)$

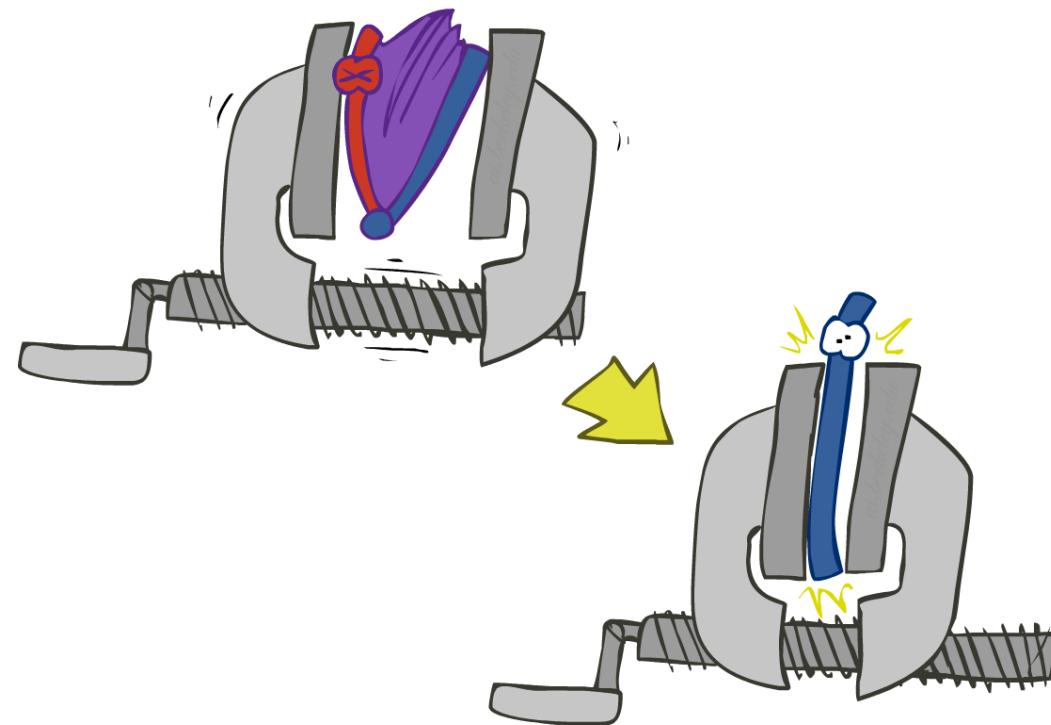
$A \setminus J$	true	false
true	0.09	0.01
false	0.045	0.855

Sum out J



$P(A)$

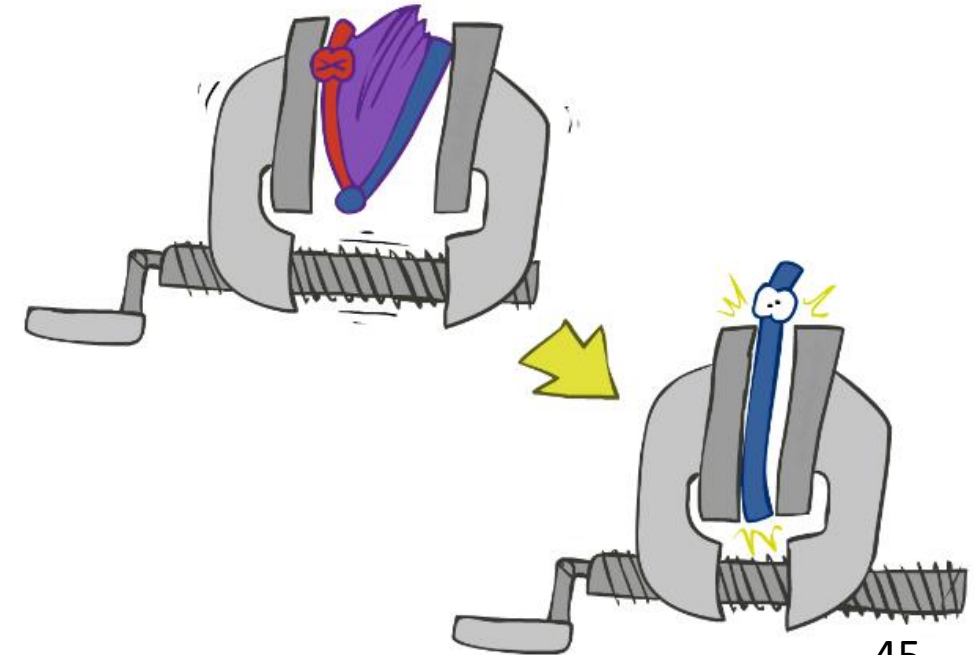
	0.1
true	0.1
false	0.9



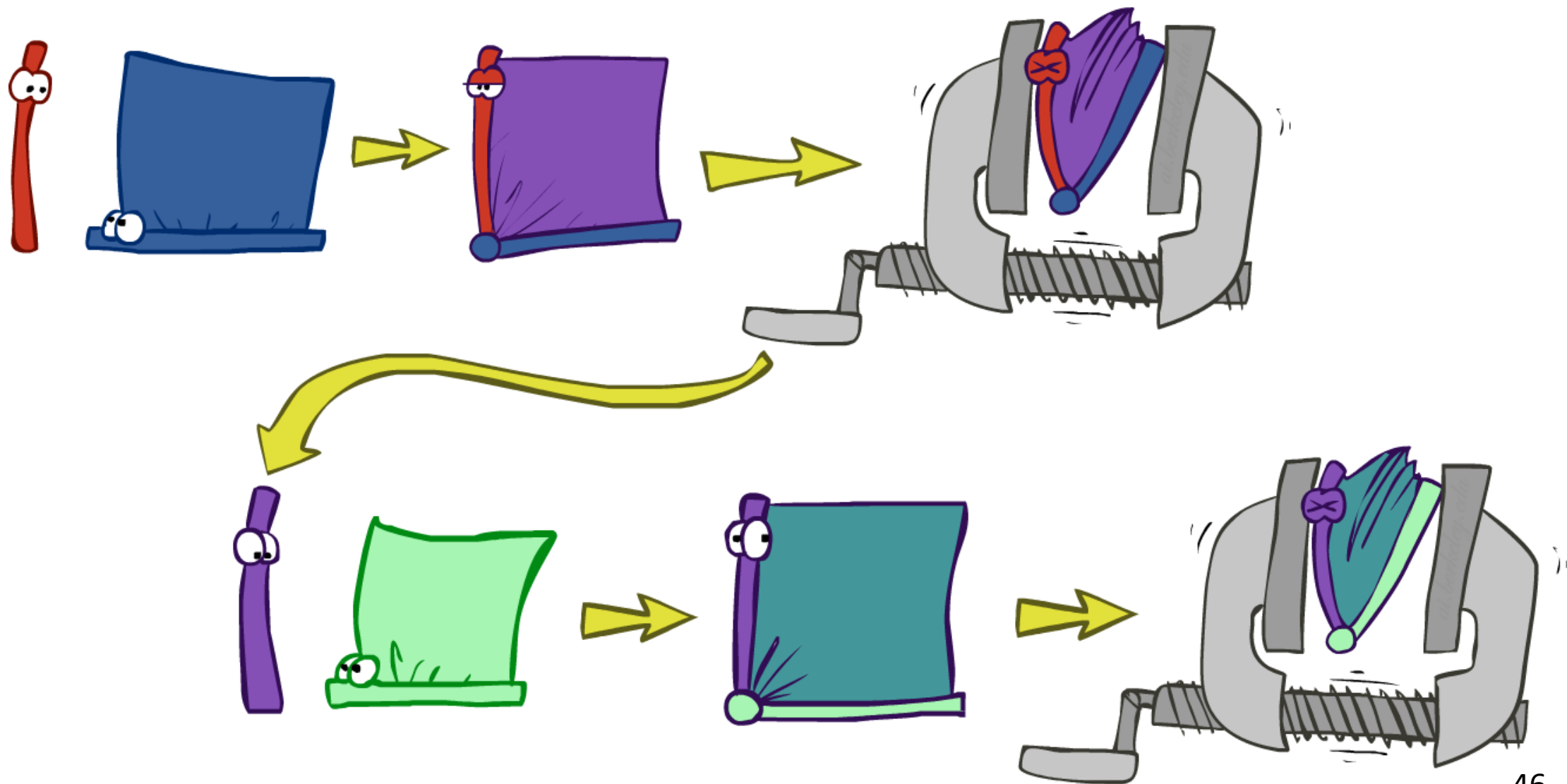
Summing out from a product of factors

Project the factors each way first, then sum the products

- Example: $\sum_a P(a|B,e) P(j|a) P(m|a)$
 $= P(a|B,e) P(j|a) P(m|a) + P(\neg a|B,e) P(j|\neg a) P(m|\neg a)$
 $= P(a,j,m|B,e) + P(\neg a,j,m|B,e)$
 $= P(j,m|B,e)$



Variable Elimination



Variable Elimination

- Query: $P(Q|E_1=e_1, \dots, E_k=e_k)$

Start with initial factors:

- Local CPTs (but instantiated by evidence)

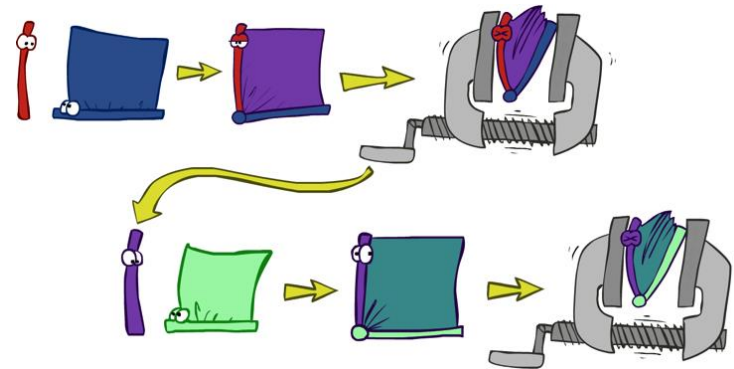


x	P(x)
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01

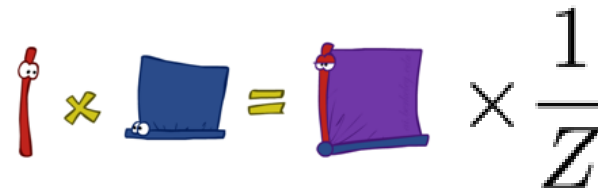
2 0.15

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

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



Join all remaining factors and normalize


$$\text{stick figure} \times \text{blue pillow} = \text{purple pillow} \times \frac{1}{Z}$$

Bayes Nets

✓ Part I: Representation

✓ 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