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

Midterm:

- Scores should be out tonight

Assignments

- HW10
  - Plan: Out tonight, due Tue 4/16
- P5
  - Plan: Out after Carnival, due 4/25

Carnival

- See Piazza post for details (no recitation, OH by appointment)

Announcements

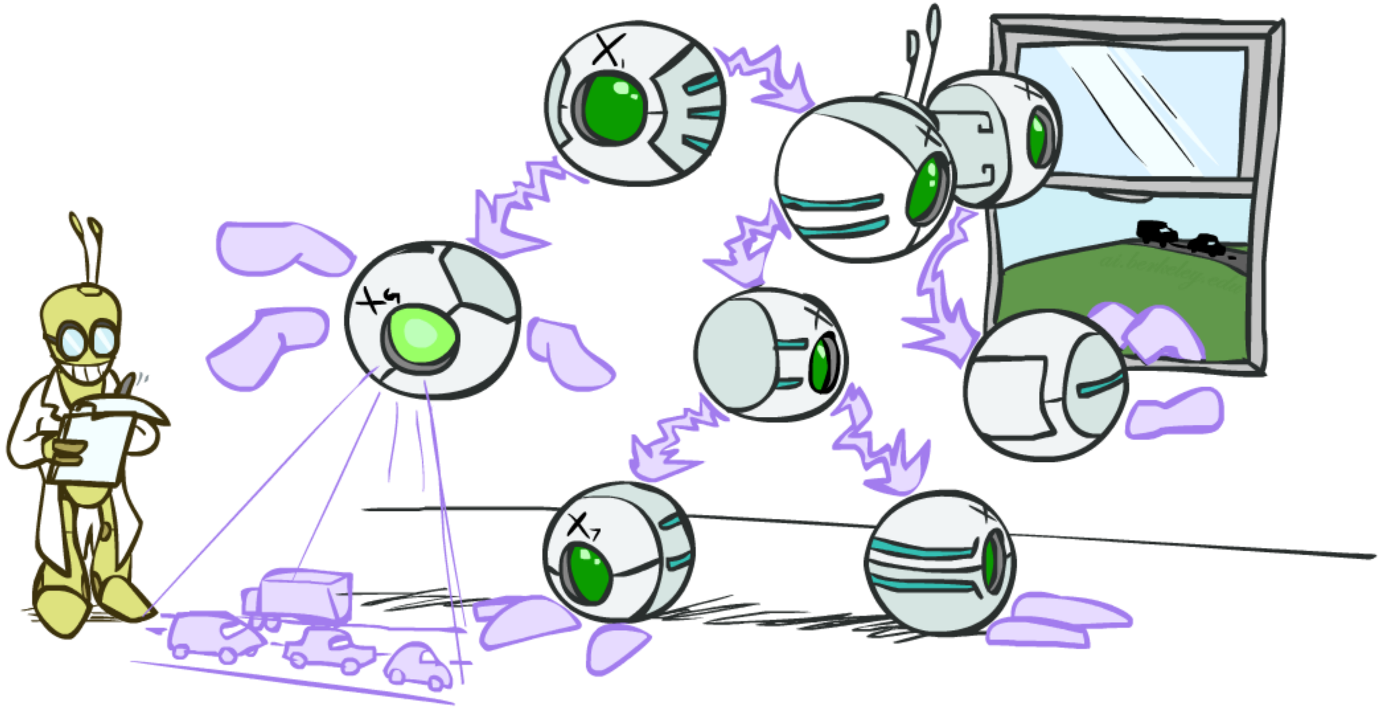
In-class Polls

$$\frac{\#\text{answered}}{\boxed{\#\text{polls}}} \geq 80\%$$

$\underline{5\%}$

AI: Representation and Problem Solving

Bayes Nets Inference



Instructors: Pat Virtue & Stephanie Rosenthal

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?

$P(q)$ $P(q \mid e)$ $P(q_1, q_2 \mid e_1, e_2, e_3)$

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

$P(Q)$ $P(Q \mid e)$ $P(Q_1, Q_2 \mid e_1, e_2, e_3)$

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

$$q^* = \arg\max_{q \in Q} P(q \mid e)$$

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)?

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

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)?

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

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 $P(\ Q \mid e\ )$?

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

A) 0

B) 1

C) 10

D) 30

E) 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 $P(\, Q \mid e\,)$?

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

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

$\sum_{q} P(q, e)$

$\hookrightarrow P(e)$

- 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

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?

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

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

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

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

$P(q_1 \mid e) = \frac{P(q_1, e)}{P(e)}$

$\rightarrow P(q_2 \mid e) = \frac{P(q_2, e)}{P(e)}$

- Q can take on ~~10~~ 2 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 $P(e)$

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

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

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

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

Bayes Nets in the Wild

Example: Speech Recognition

“artificial ……….”

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

$P(F, S, A) = P(F)P(S|F)P(A|F,S)$

Bayes Nets in the Wild

$F \rightarrow S \rightarrow A$

$P(F)P(S|F)P(A|S)$

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

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

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

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

Bayes Nets in the Wild

$\text{First} \rightarrow \text{Second} \rightarrow \text{Audio}$

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

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

Definition of cond. prob.

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

Denominator doesn't affect argmax

Chain rule

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

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

Bayes Net

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

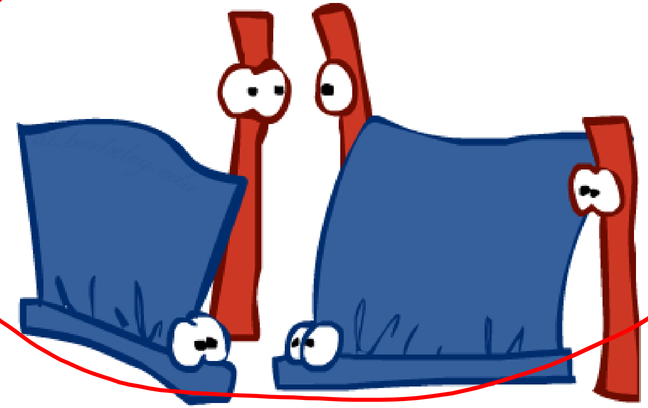
First factor doesn't affect argmax

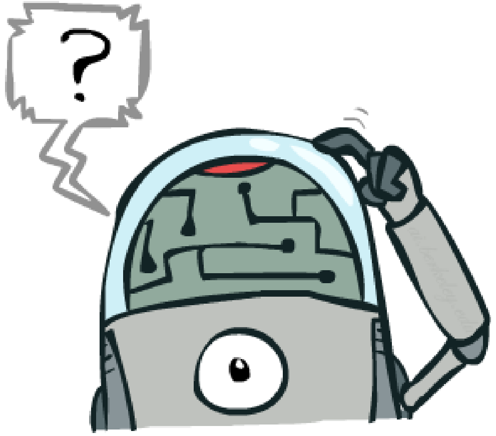
n-gram probability * audio probability

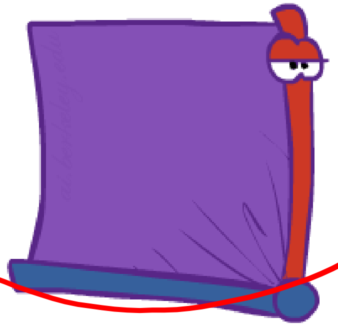
Inference

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?







Inference Overview

joint $q\ h\ e$

joint $q\ e$

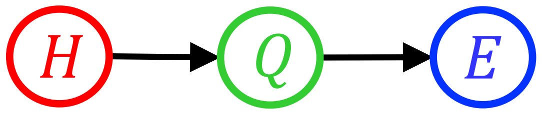
Given random variables Q, H, E (query, hidden, evidence)

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



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

Variable Elimination

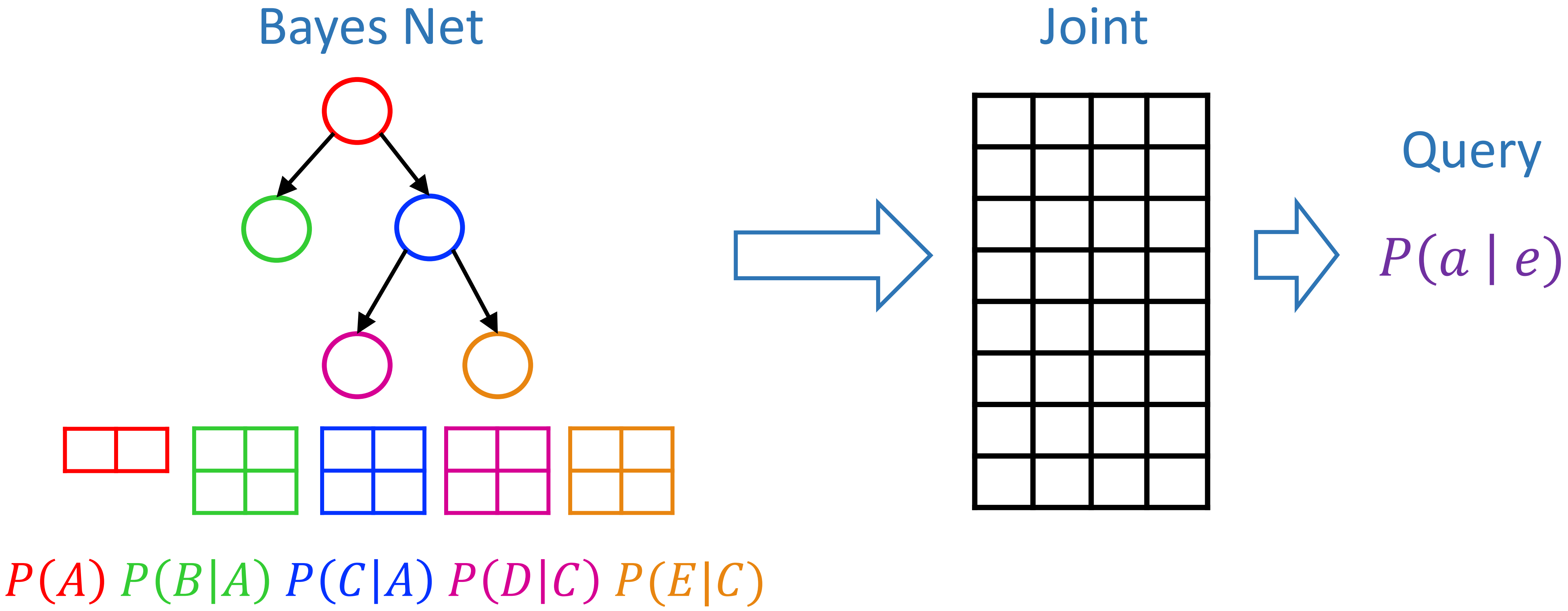
But we can be more efficient

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

Never actually compute joint

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

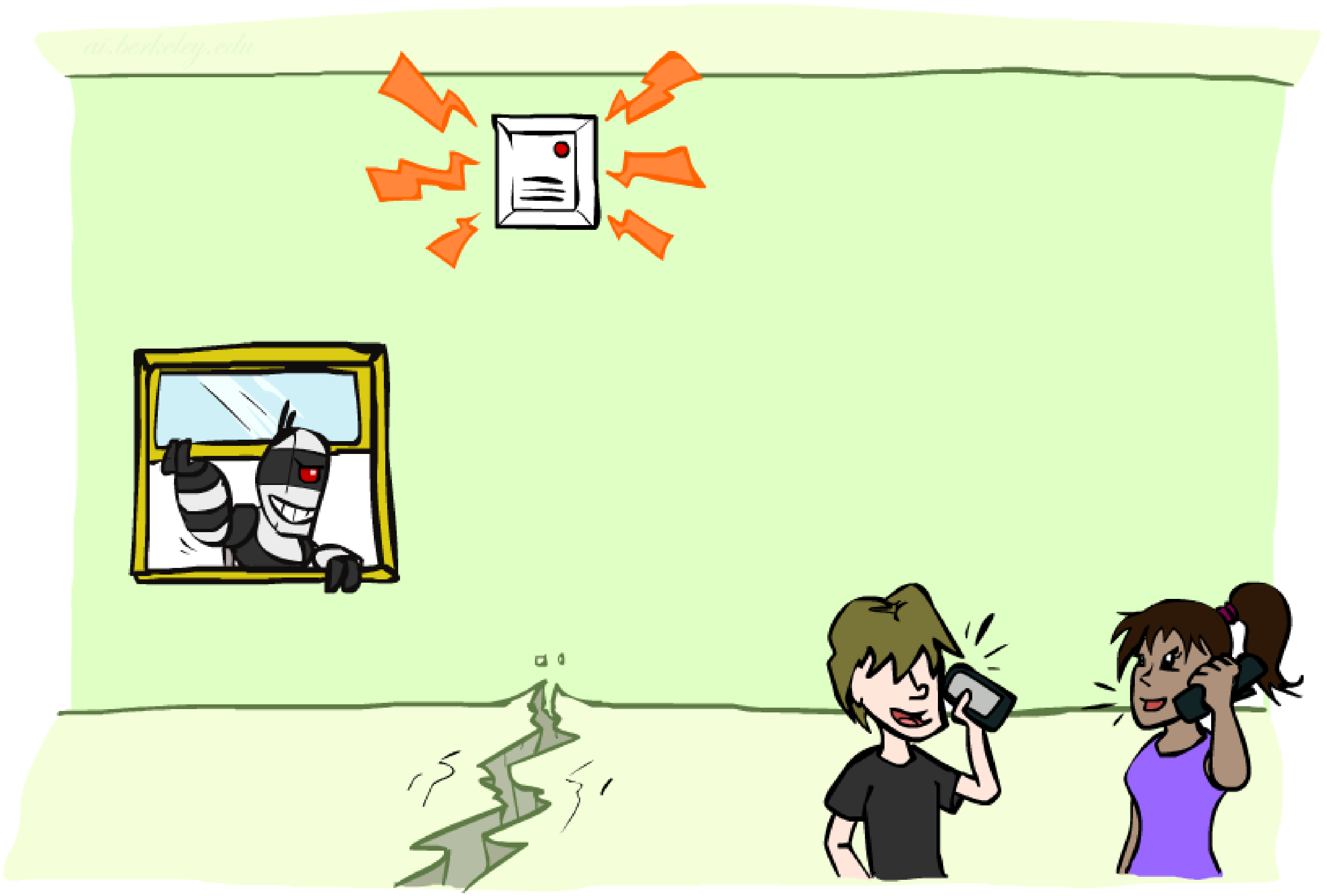
Answer Any Query from Condition Probability Tables



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

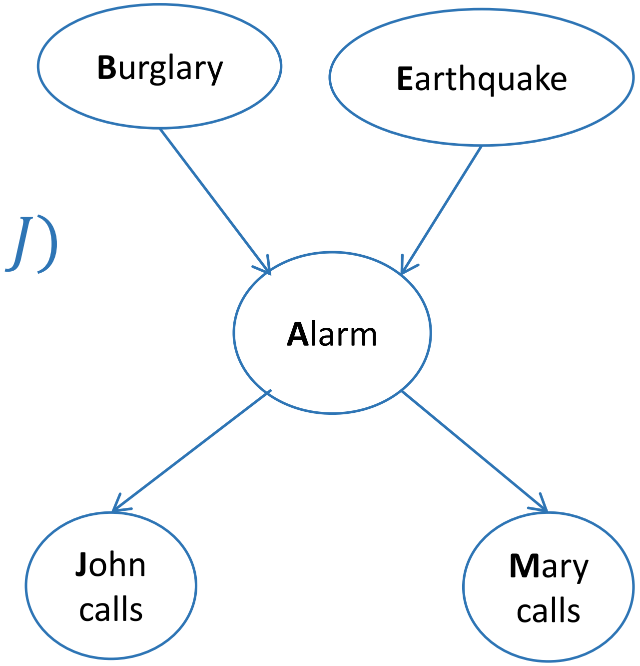
- $P(X_1 \ldots X_2) = \underline{\prod_i P(X_i} |\, X_1 \ldots X_{i-1})$

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

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

Bayes nets

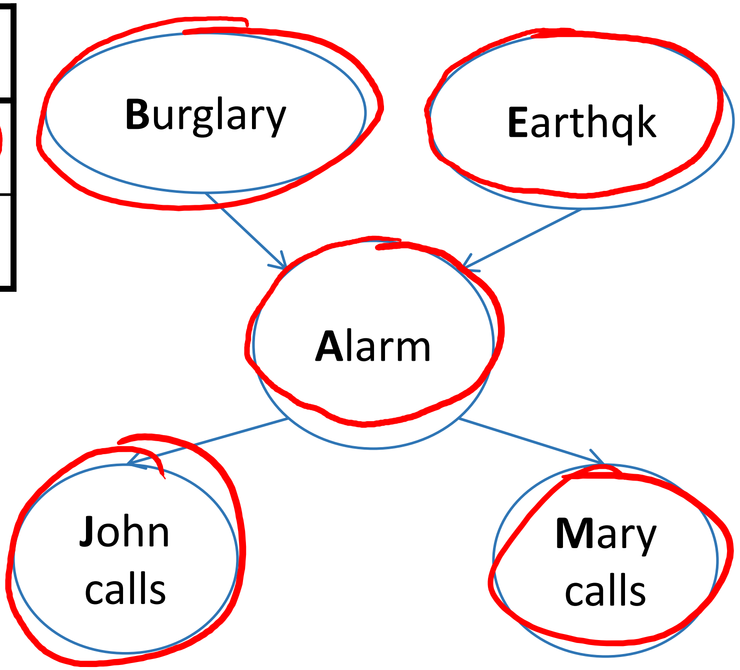
- $P(X_1 \ldots X_2) = \underline{\prod_i P(X_i} |\, Parents(X_i))$



Example: Alarm Network

$$\underbrace{P(+b, -e, -a, -m, -j) =}$$

B	P(B)
+b	0.001
-b	0.999

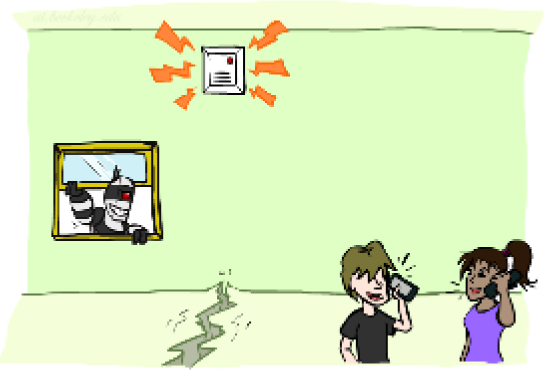


E	P(E)
+e	0.002
-e	0.998

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



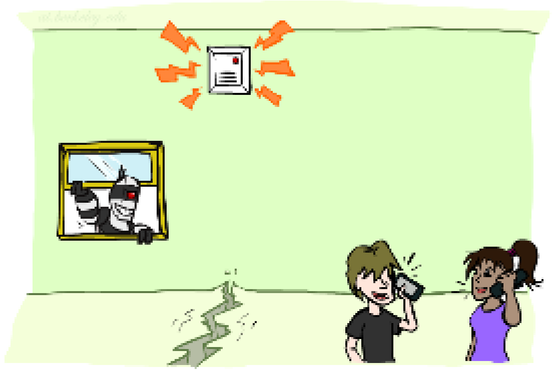
Example: Alarm Network

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

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

$P(+b)\,P(-j|-a)$

0.00095



Example: Alarm Network

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

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

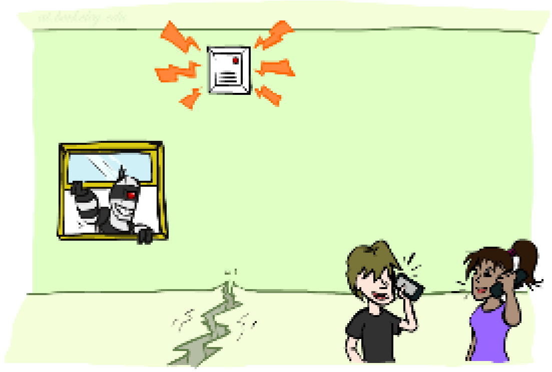
$$P(+b,-e,-a,-j,-m) = P(-e) * P(-a|+b,-e) * P(+b)\, P(-j|-a) * P(-m|-a)$$
$$= 0.998 * 0.06 * 0.0095 * 0.99$$

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

Factors

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

$$f_1(+b) \cdot f_2(-e) \cdot f_3(-a,+b,-e) \cdot f_4(-j,a) \cdot f_5(-m,-a)$$



Factors

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

$$P(+b,-e,-a,-j,-m) = f_2(-e) * f_3(-a,+b,-e) \left(* \boxed{f_1(+b)} * f_4(-j,-a) \right) * f_5(-m,-a)$$
$$= 0.998 * 0.06 \left(* \boxed{0.001} * 0.95 \right) * 0.99$$

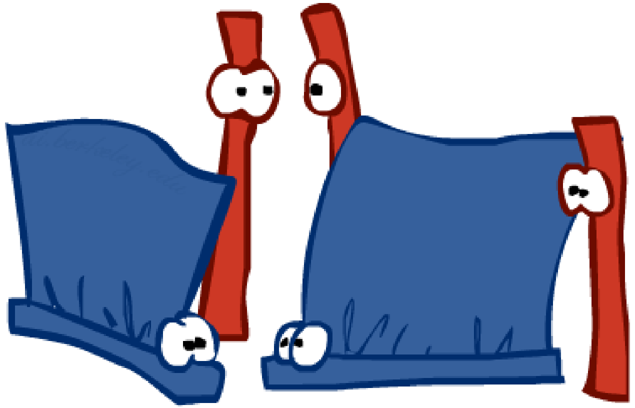
$$P(+b,-e,-a,-j,-m) = f_2(-e) * f_3(-a,+b,-e) * \boxed{f_6(+b,j,-a)} * f_5(-m,-a)$$
$$= 0.998 * 0.06 * \boxed{0.0095} * 0.99$$

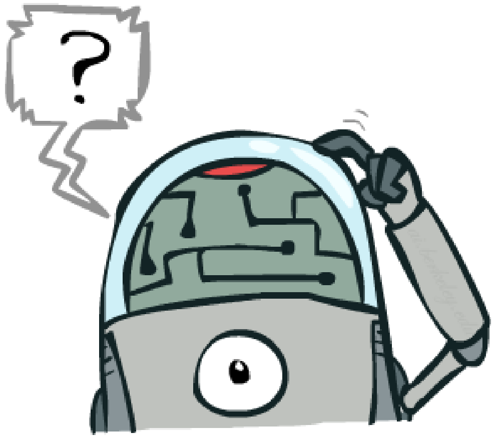
- Multiplication order can change (commutativity)
- Multiplication pairs don't have to make sense (associativity)

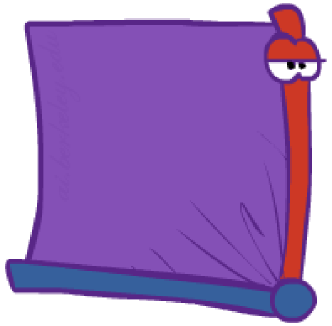
Factor Tables

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

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



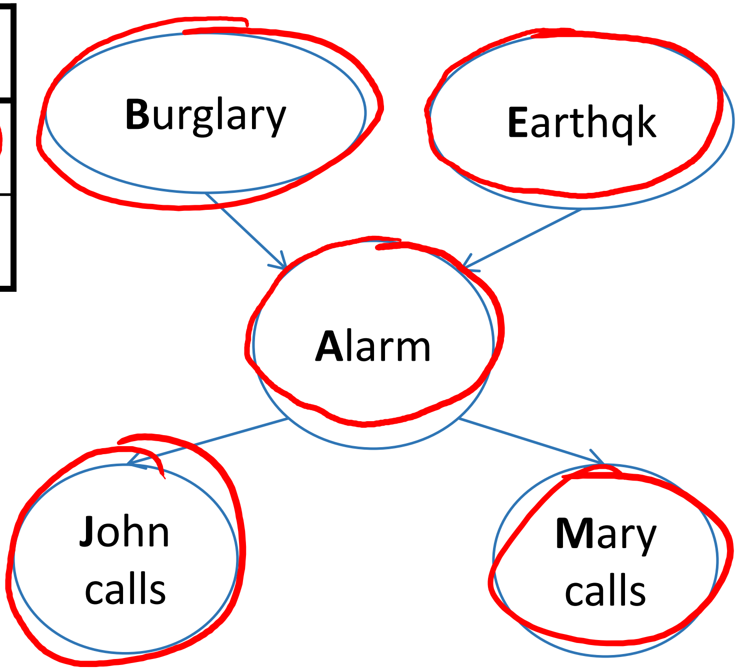




Example: Alarm Network

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

B	P(B)
+b	0.001
-b	0.999

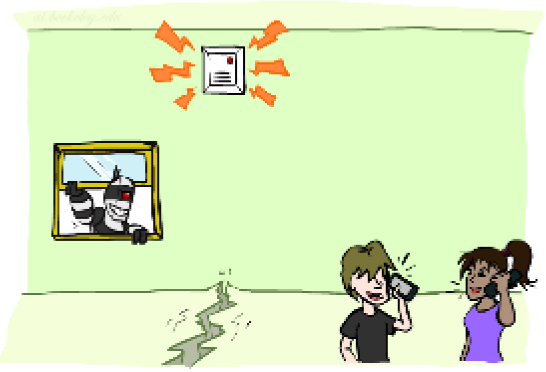


E	P(E)
+e	0.002
-e	0.998

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



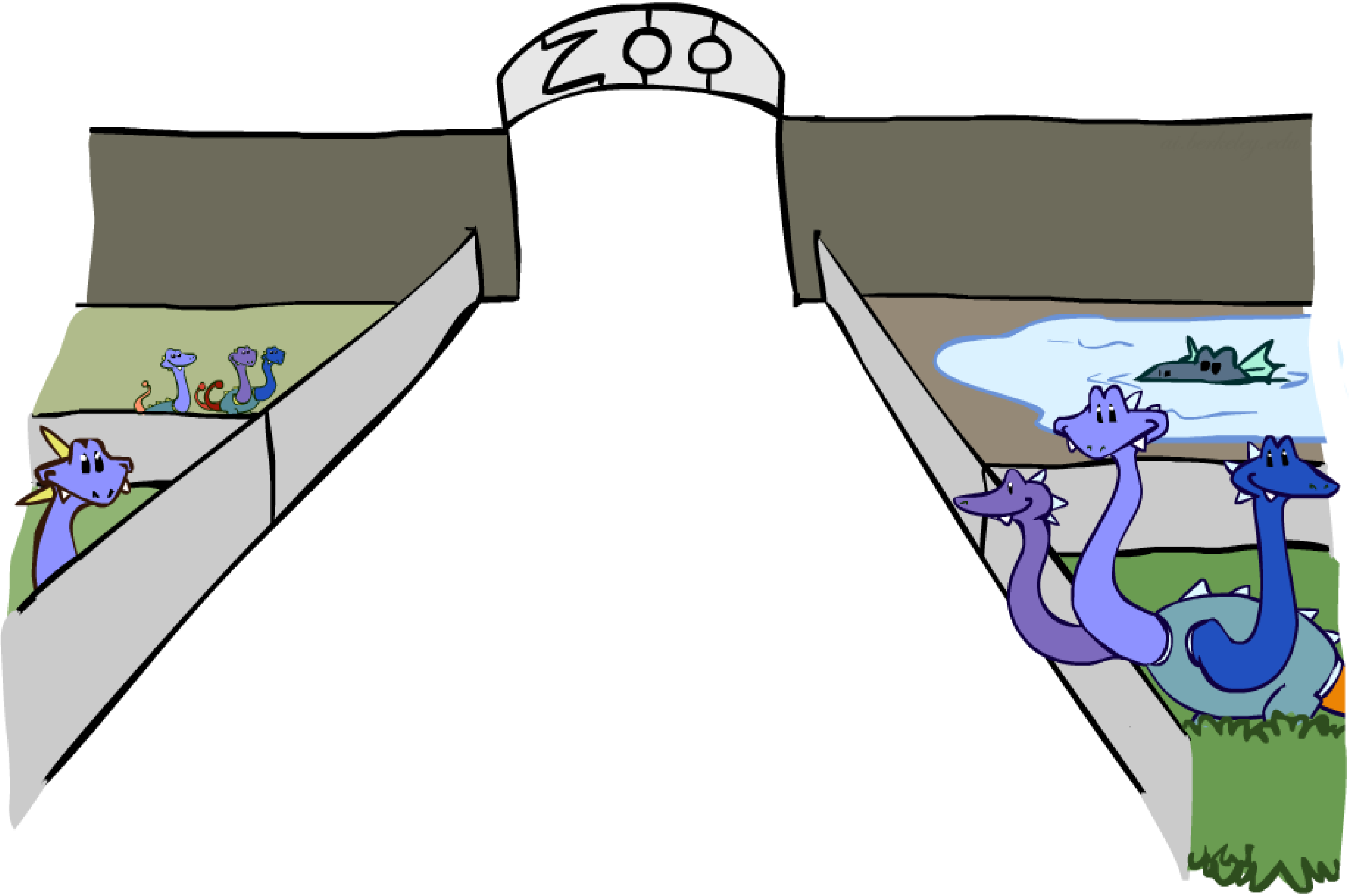
Factor Tables

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

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

- Compute each entry in the output table, one at a time
- Multiplication order can change (commutativity)
- Multiplication pairs don’t have to make sense (associativity)

Factor Zoo



Factor Zoo I

Joint distribution: P(X,Y)

- Entries P(x,y) for all x, y
- |X|x|Y| matrix
- Sums to 1

P(A,J)

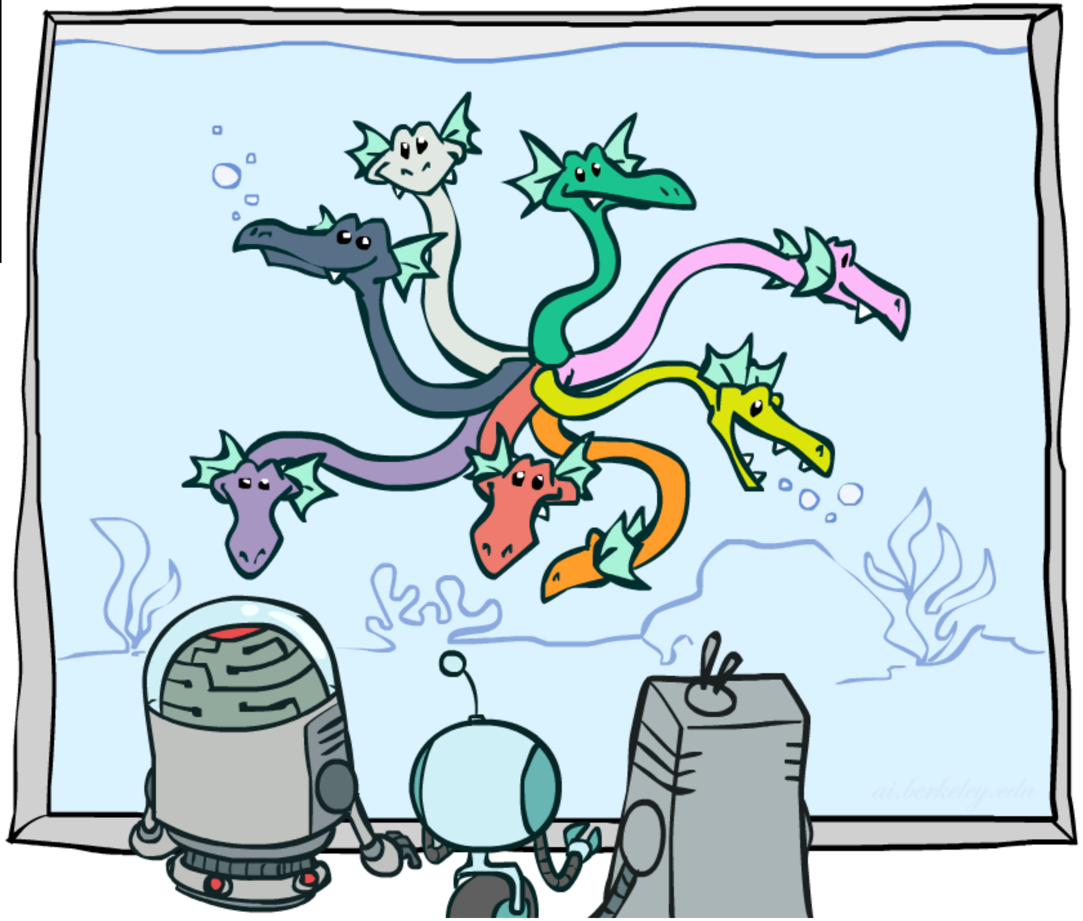
A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Projected joint: P(x,Y)

- A slice of the joint distribution
- Entries P(x,y) for one x, all y
- |Y|-element vector
- Sums to P(x)

P(a,J)

A \ J	true	false
true	0.09	0.01

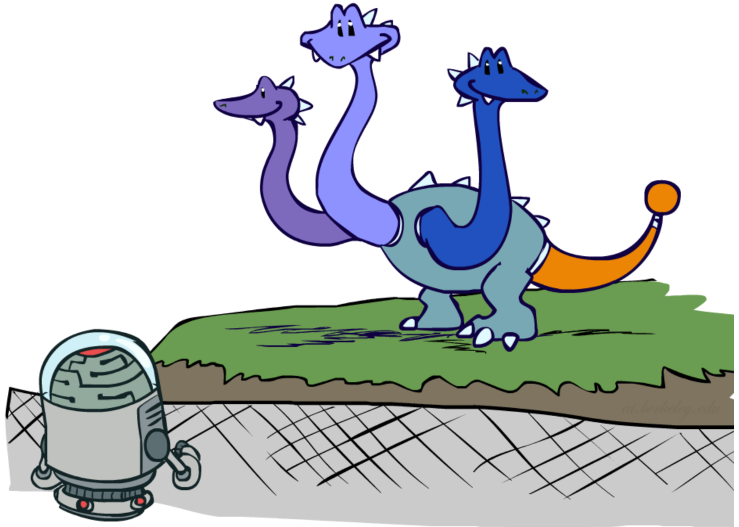


Number of variables (capitals) = dimensionality of the table

Factor Zoo II

Single conditional: P(Y | x)

- Entries P(y | x) for fixed x, all y
- Sums to 1



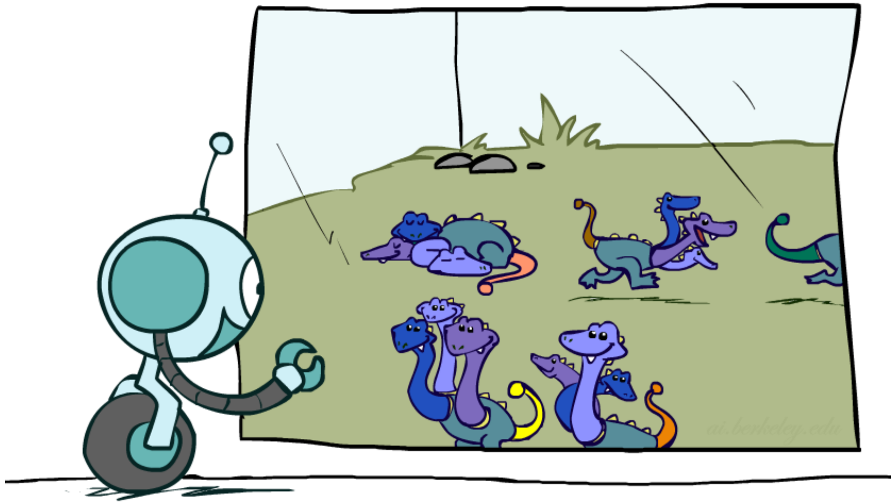
$P(J|a)$

A \ J	true	false
true	0.9	0.1

Family of conditionals:

P(X |Y)

- Multiple conditionals
- Entries P(x | y) for all x, y
- Sums to |Y|



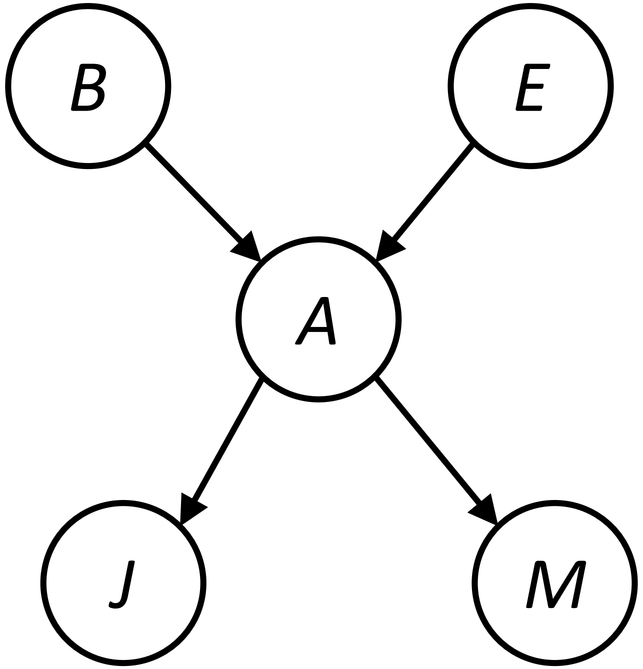
$P(J|A)$

A \ J	true	false
true	0.9	0.1
false	0.05	0.95

$\}$ - $P(J|a)$

$\}$ - $P(J|\neg a)$

Inference by Enumeration in Bayes Net



Reminder of inference by enumeration:

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

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

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

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

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

- Lots of repeated subexpressions!

Variable elimination: The basic ideas

Move summations inwards as far as possible

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

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

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

Variable elimination: The basic ideas

Move summations inwards as far as possible

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

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

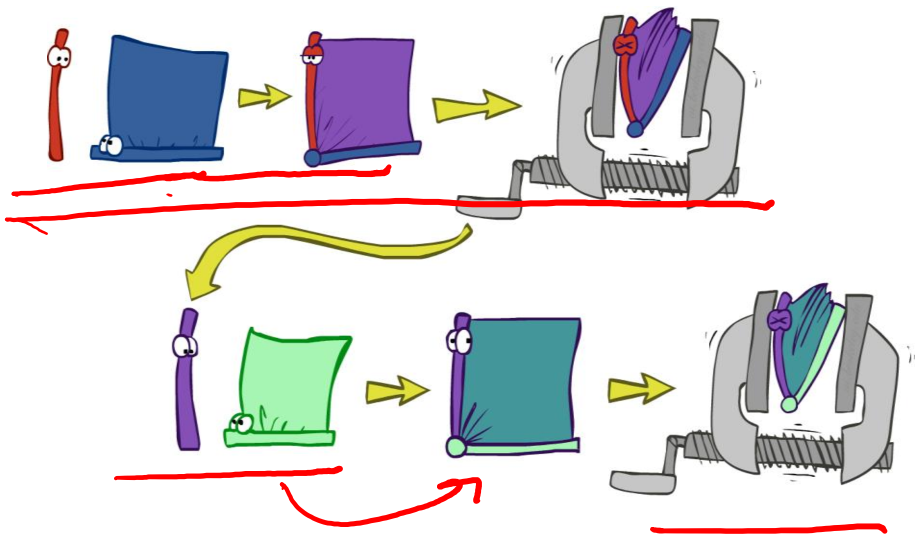
$$= \alpha\, P(B) \sum_e P(e) \sum_a P(j|a)\, P(m|a)\, P(a|B,e)$$

Variable elimination: The basic ideas

Move summations inwards as far as possible

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

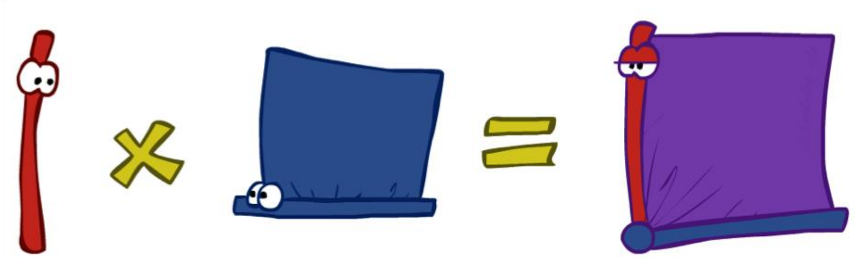
  $= \alpha\, P(B) \sum_e P(e) \sum_a P(a|B,e)\, P(j|a)\, P(m|a)$



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

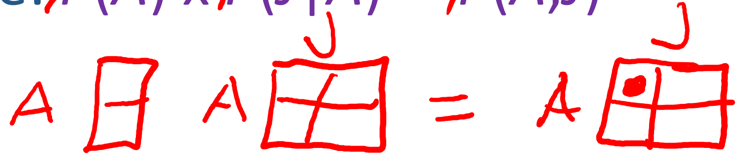
Operation 1: Pointwise product



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: $\overset{f}{\cancel{P}}(A) \times \overset{f}{\cancel{P}}(J|A) = \overset{f}{\cancel{P}}(A,J)$



$P(A)$

true	0.1
false	0.9

x

$P(J|A)$

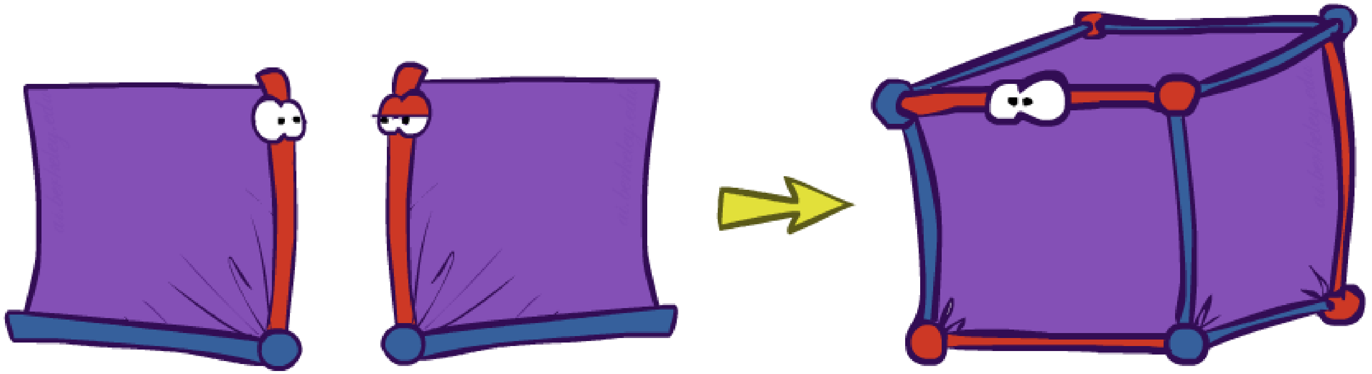
A \ J	true	false
true	0.9	0.1
false	0.05	0.95

=

$P(A,J)$

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Example: Making larger factors



Example: $P(J|A)\ \text{x}\ P(M|A) = P(J,M|A)$

$P(J|A)$

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

=

$P(J,M|A)$

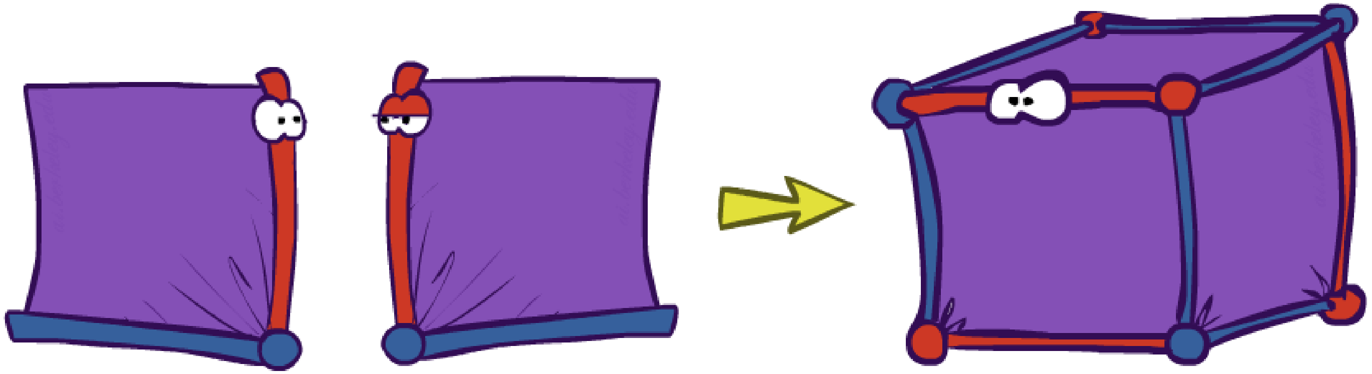
A=false

J \ M	true	false
true		18
false		

A=true

J \ M	true	false
true		
false		.0003

Example: Making larger factors



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

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

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

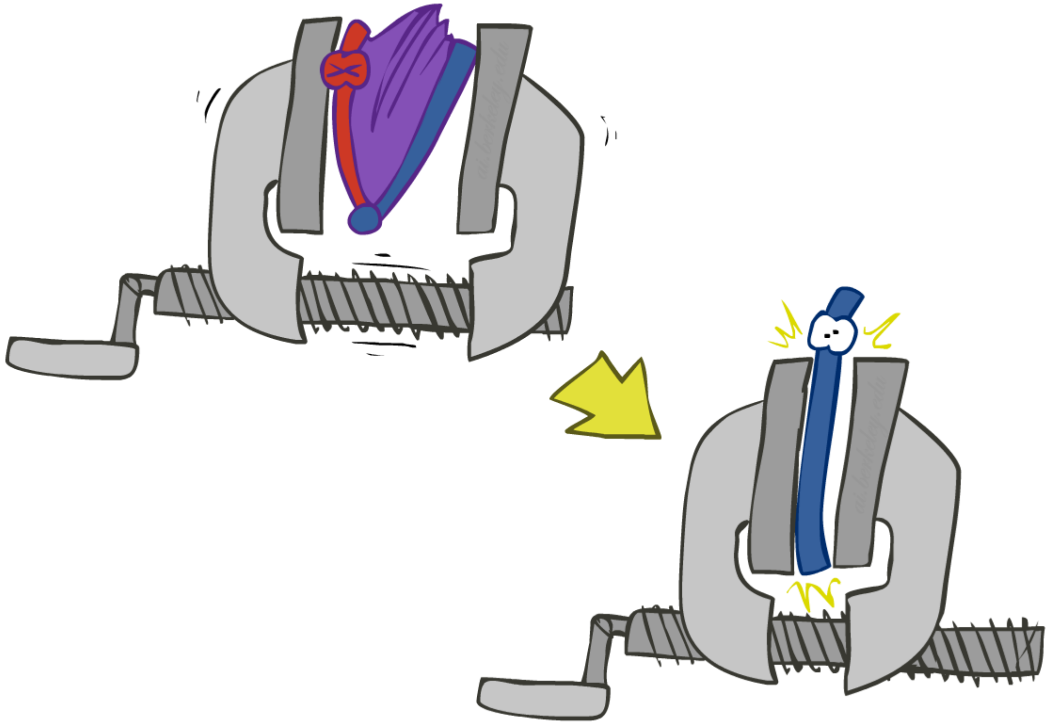
Factor blowup can make VE very expensive

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

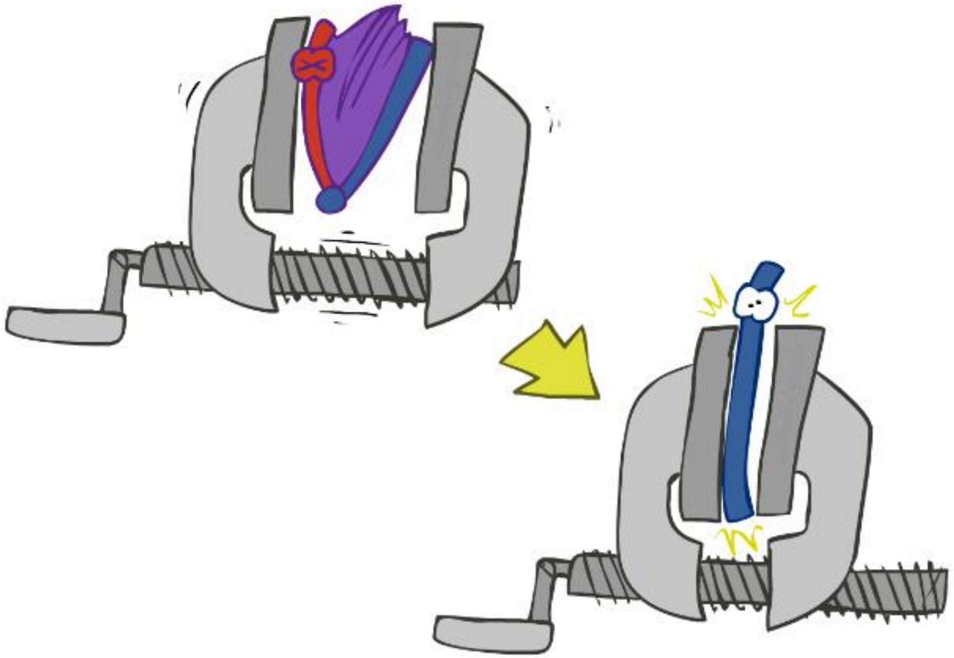
$$= P(a|B,e)\ P(j|a)\ P(m|a) + P(\neg a|B,e)\ P(j|\neg a)\ P(m|\neg a)$$

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

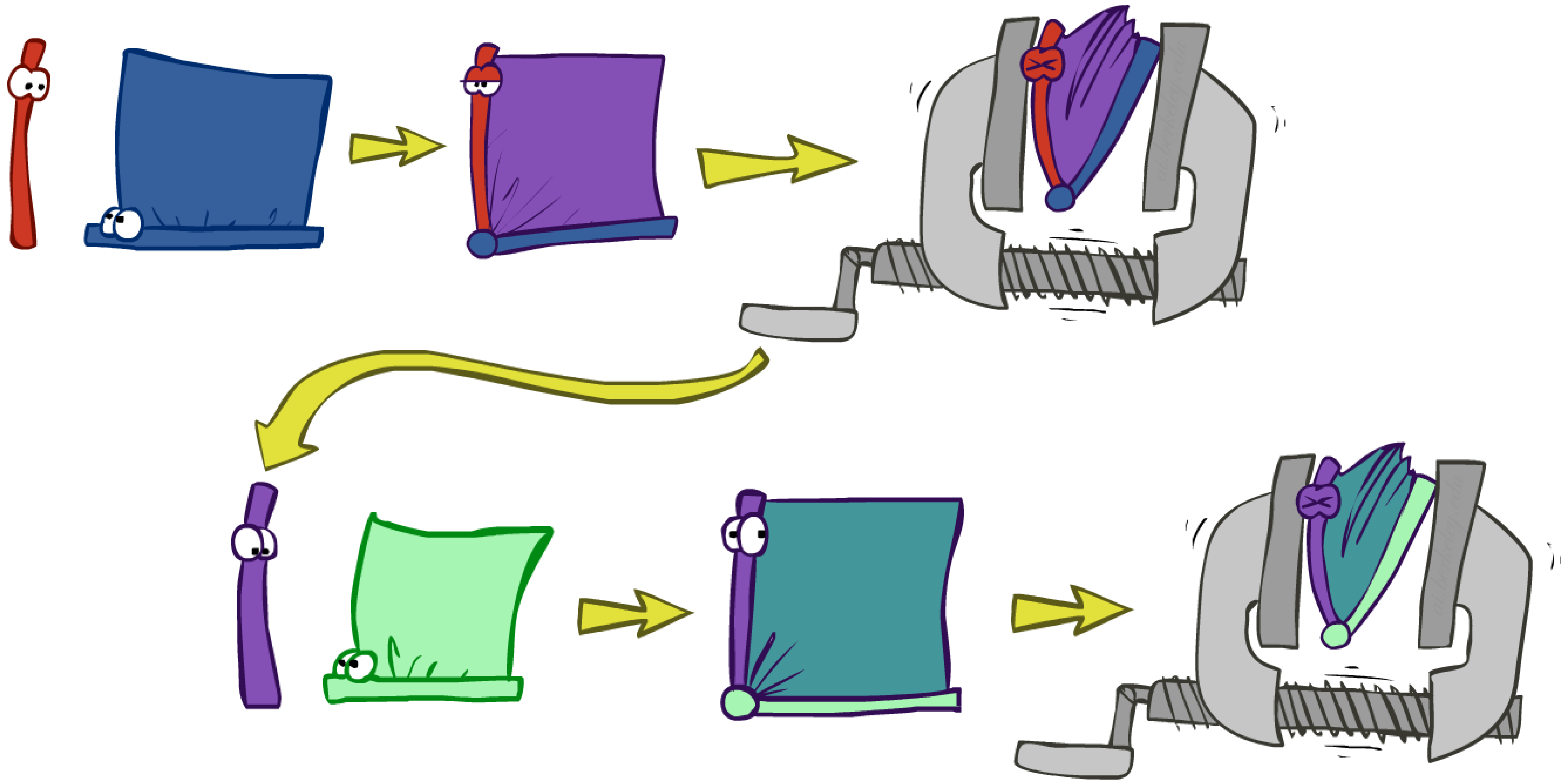
$$= P(j,m|B,e)$$

$P(B) \sum_e P(e) \sum_a$

$P(B|j\,m)$

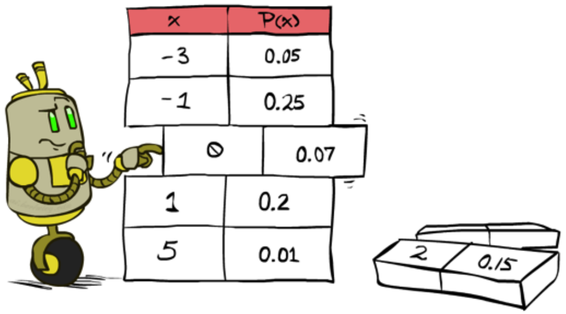


Variable Elimination

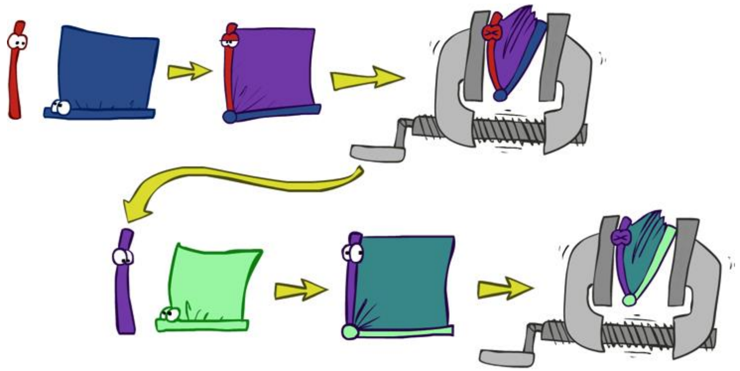


Variable Elimination

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



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



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

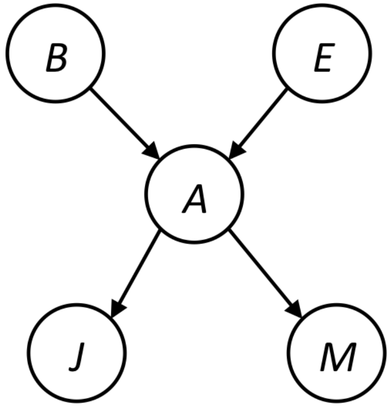
Join all remaining factors and normalize

 $\times \frac{1}{Z}$

Variable Elimination

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

Example



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

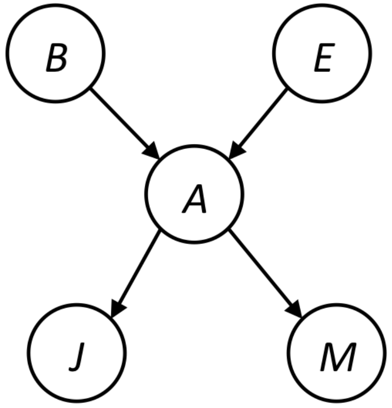
$$P(B) \quad P(E) \quad P(A|B,E) \quad P(j|A) \quad P(m|A)$$

Choose A

$$\left.\begin{array}{l} P(A|B,E) \\ P(j|A) \\ P(m|A) \end{array}\right. \xrightarrow{\times} P(A,j,m|B,E) \xrightarrow{\Sigma} P(j,m|B,E)$$

$$P(B) \quad P(E) \quad P(j,m|B,E)$$

Example



$\boxed{P(B) \quad P(E) \quad P(j,m|B,E)}$

Choose E

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

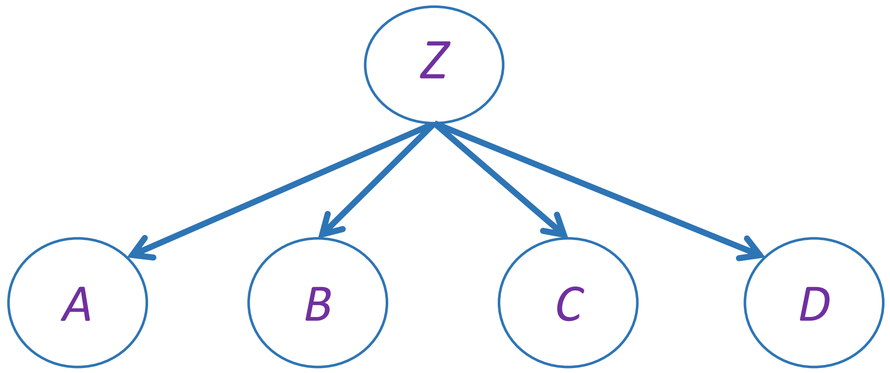
$\boxed{P(B) \quad P(j,m|B)}$

Finish with B

$P(B)$, $P(j,m|B)$ $\xrightarrow{\times}$ $P(j,m,B)$ $\xrightarrow{\text{Normalize}}$ $P(B \mid j,m)$

Order matters

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

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