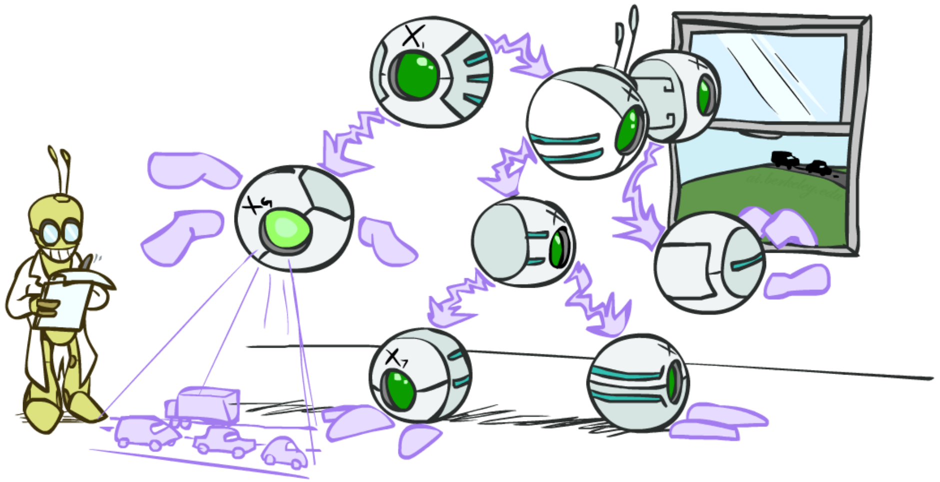
Bayes Nets II: Inference

CSE 473: Introduction to Artificial Intelligence

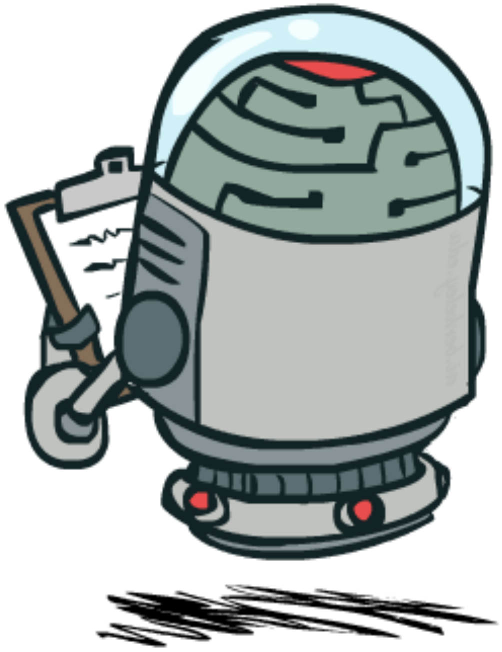


James Weichert (jpw@cs)
cs.uw.edu/473
acknowledgements

Welcome!
What should the CSE 473 logo/mascot be?
sli.do **#cse473**

Mid-Quarter Feedback Survey :)

Administrivia



ANNOUNCEMENTS

- Complete the **Mid-Quarter Feedback** survey!

ASSIGNMENTS

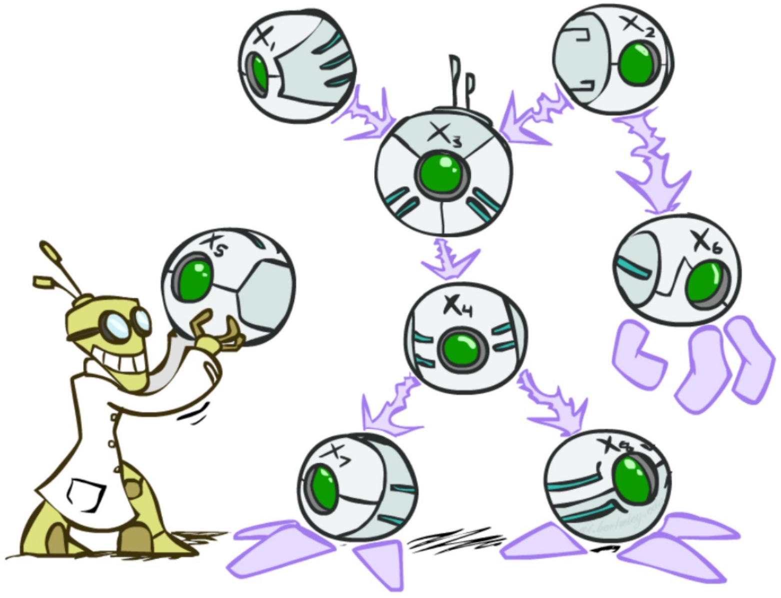
- **Project 3 due** next Thursday (7.30)
- **Homework 3** released today, **due** in two weeks (8.6)

Bayes Nets

FOUNDATIONS

Bayes Nets provide a technique describing **complex joint distributions** using **simple conditional distributions**

- Use local causality / conditional independence
- Trading expressivity of joint distributions for efficiency of conditional distributions



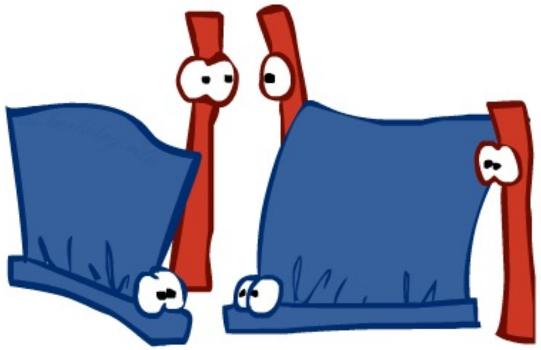
Representation	**Inference**	**Sampling**
Wednesday	*Today*	*Monday*

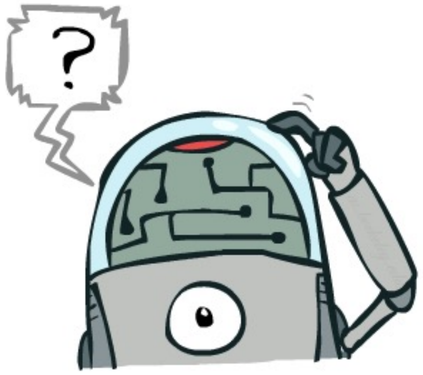
Inference

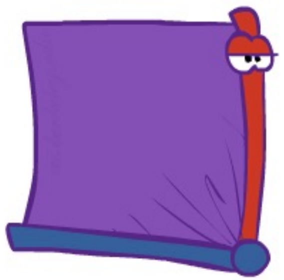
CONTEXT

Inference involves calculating some useful quantity from a joint distribution

- Given posterior probability $P(Q|E_1 = e_1, \dots, E_k = e_k)$
- Most likely explanation is $\text{argmax}_q P(Q = q|E_1 = e_1, \dots, E_k = e_k)$





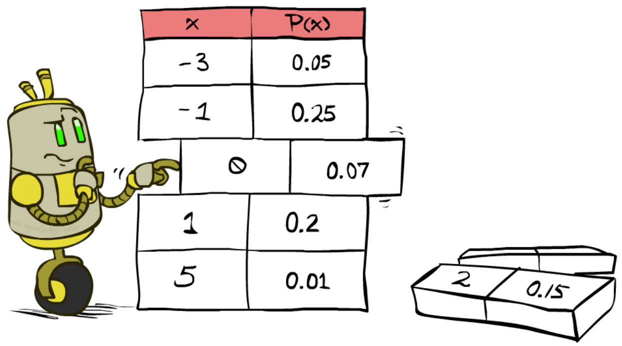


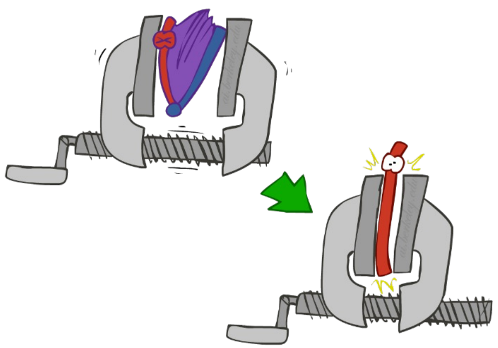
Inference by Enumeration

DEFINITION

- Calculate probability of query Q given:
  - Observations about evidence variables $E_1, \dots, E_k = e_1, \dots, e_k$
  - Hidden (unobserved) variables $H_1, \dots, H_r$

Infer $P(Q|e_1, \dots, e_k)$ from full probability model $P(X_1, \dots, X_n)$





$$P(Q|e_1, \dots, e_k) = \frac{P(Q, e_1, \dots, e_k)}{P(e_1, \dots, e_k)}$$

$$P(Q, h_1, \dots, h_r, e_1, \dots, e_k)$$

Select model entries consistent with evidence

$$P(Q, e_1, \dots, e_k) = \sum_{h_1, \dots, h_r} P(Q, h_1, \dots, h_r, e_1, \dots, e_k)$$

Sum out hidden variables to calculate joint probability

Normalize by joint probability of the evidence

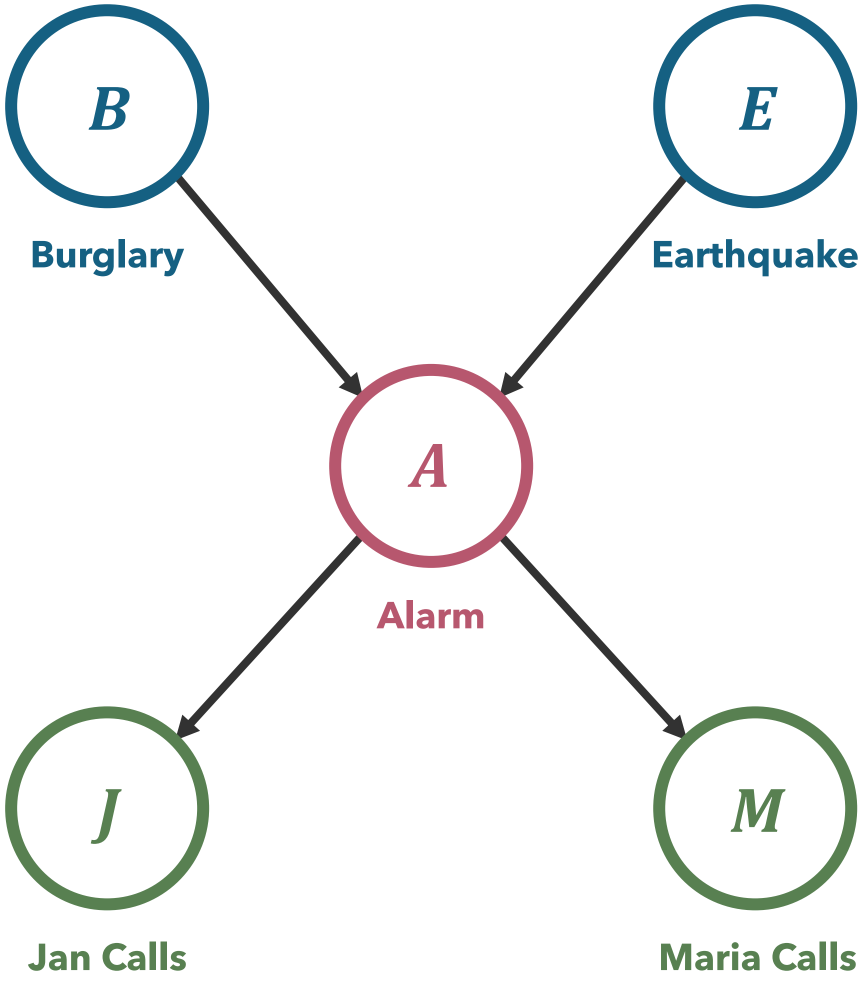
Bayes Net Inference by Enumeration

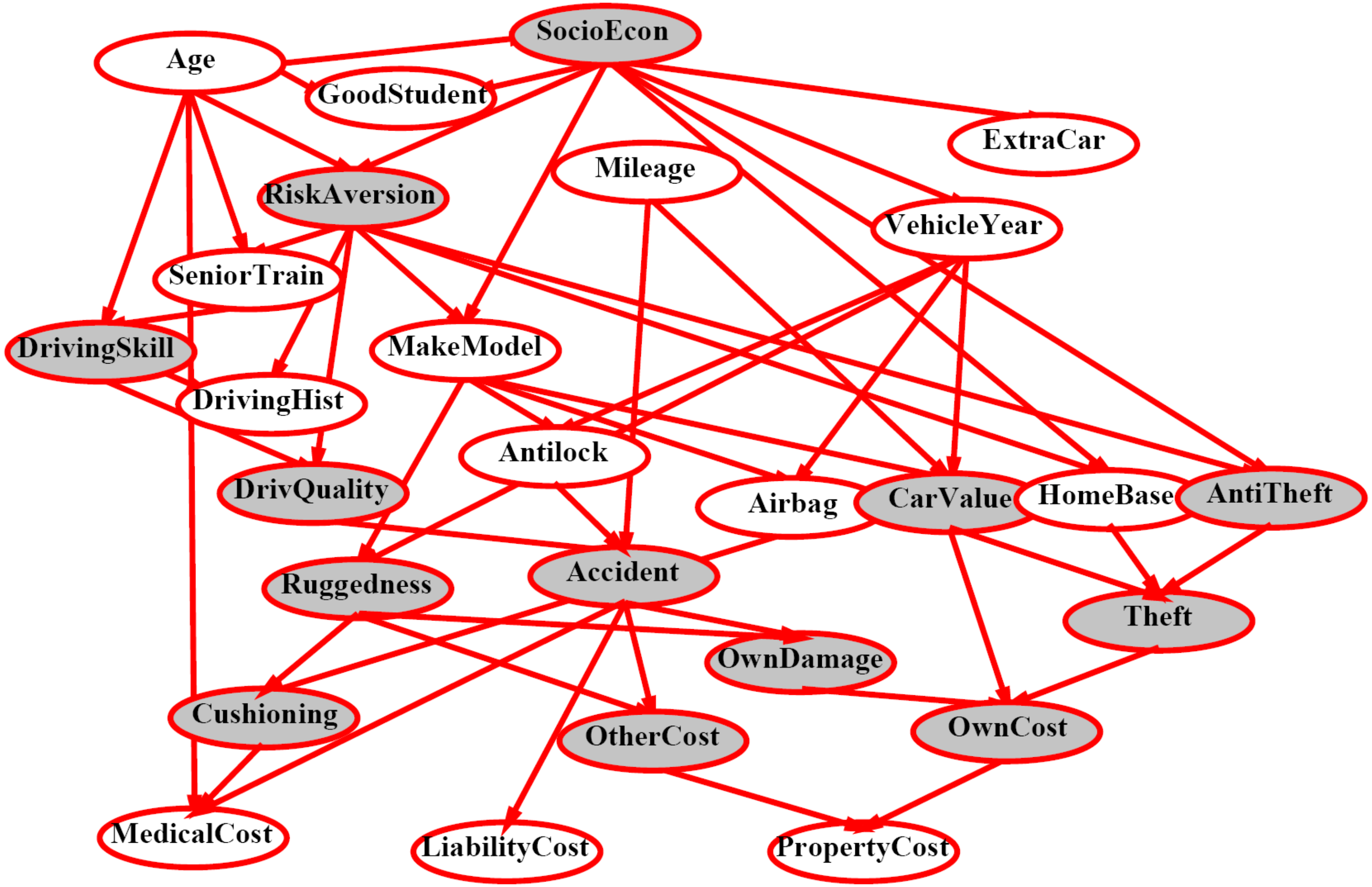
FOUNDATIONS

- Goal:
  - $P(\boldsymbol{Q}|\boldsymbol{e}) = \alpha \sum_{\boldsymbol{h}} P(\boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{e})$
- In a Bayes Net:
  - $P(\boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{e}) = \prod_i P(X_i|\text{parents}(X_i))$
  - For example:

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

  - Therefore, inference involves computing ***sums of products***
- Problem:
  - Sums of ***exponentially many*** **products**!





$P(\textbf{Antilock}|\textbf{observed variables})$?

Can We Do Better?

CONTEXT

- Consider $uwy + uwz + uxy + uxz + vwy + vwz + vxy + vxz$
  - 16 multiplications, 7 additions
  - Lots of repeated subexpressions!
- Rewrite as $(u + v)(w + x)(y + z)$
- Bayes Net Example
  - $\sum_{e,a} P(B) \cdot P(e) \cdot P(a|B,e) \cdot P(j|a) \cdot P(m|a)$
  - $= \boxed{P(B) \cdot P(e)} \cdot P(a|B,e) \cdot \boxed{P(j|a) \cdot P(m|a)} + \boxed{P(B) \cdot P(\neg e)} \cdot P(a|B,\neg e) \cdot \boxed{P(j|a) \cdot P(m|a)}$
    $+ \boxed{P(B) \cdot P(e)} \cdot P(\neg a|B,e) \cdot \boxed{P(j|\neg a) \cdot P(m|\neg a)} + \boxed{P(B) \cdot P(\neg e)} \cdot P(\neg a|B,\neg e) \cdot \boxed{P(j|\neg a) \cdot P(m|\neg a)}$

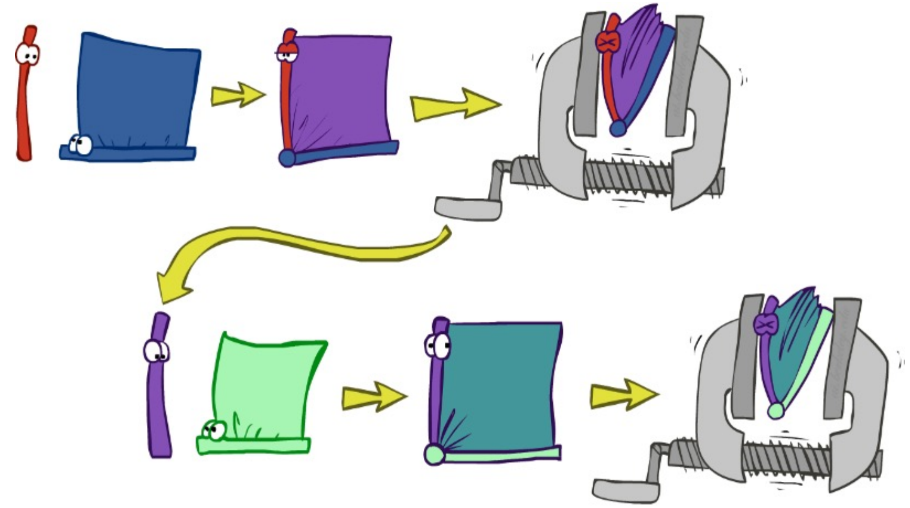
Questions?

live and on sli.do #cse473

Variable Elimination

FOUNDATIONS

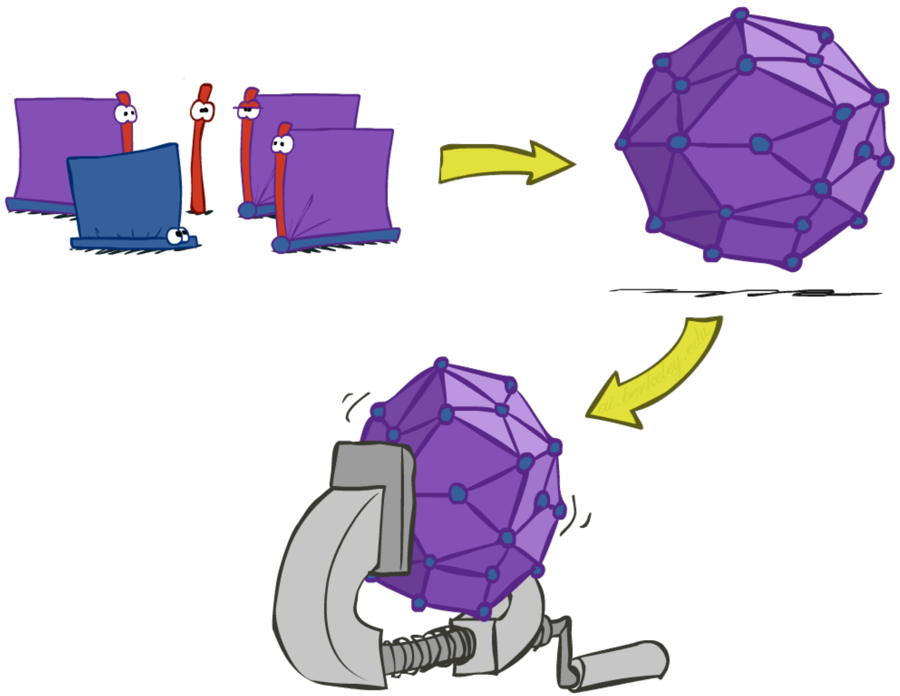
- Idea: Move summations inwards as far as possible
  - $\mathrm{P}(\mathrm{B}|\mathrm{j},\mathrm{m}) = \alpha \sum_{\mathrm{e,a}} \mathrm{P}(\mathrm{B}) \cdot \mathrm{P}(\mathrm{e}) \cdot \mathrm{P}(\mathrm{a}|\mathrm{B},\mathrm{e}) \cdot \mathrm{P}(\mathrm{j}|\mathrm{a}) \cdot \mathrm{P}(\mathrm{m}|\mathrm{a})$
    $= \alpha \cdot P(B) \cdot \sum_e P(e) \cdot \sum_a P(a|B,e) \cdot P(j|a) \cdot P(m|a)$
- Perform calculations from inside out
  - Sum over a first, then sum over e
  - Problem:
    - $P(a|B,e)$ is a series of numbers depending on the value of B
  - Solution:
    - Use arrays of numbers with appropriate operations
    - These are called **factors**



Inference by Enumeration vs. Variable Elimination

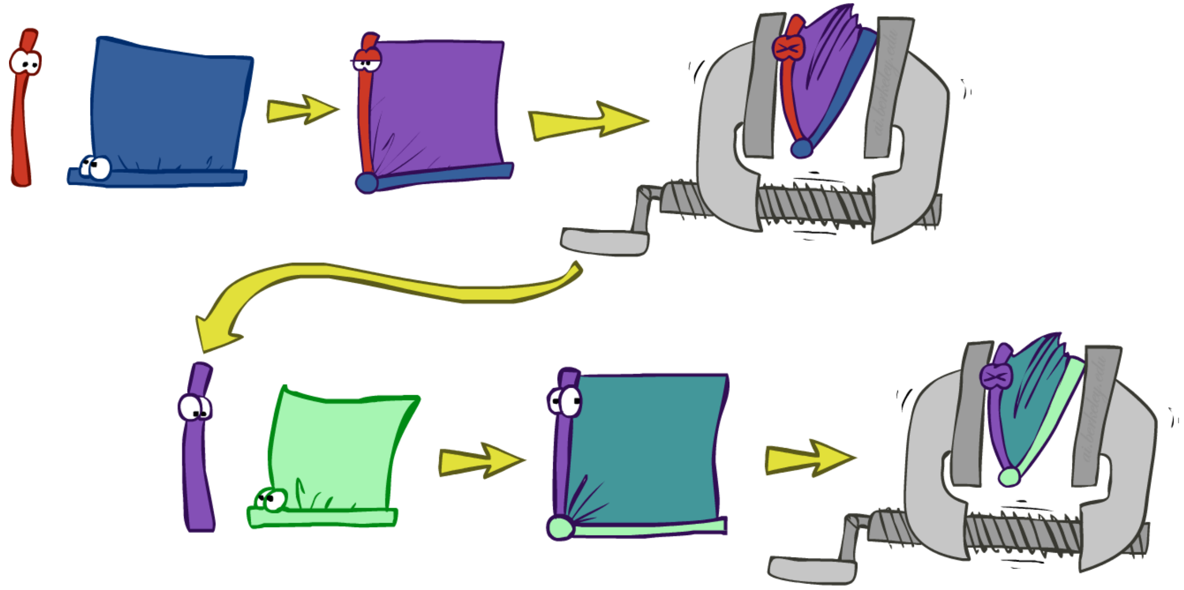
CONTEXT

Inference by Enumeration



join, then marginalize

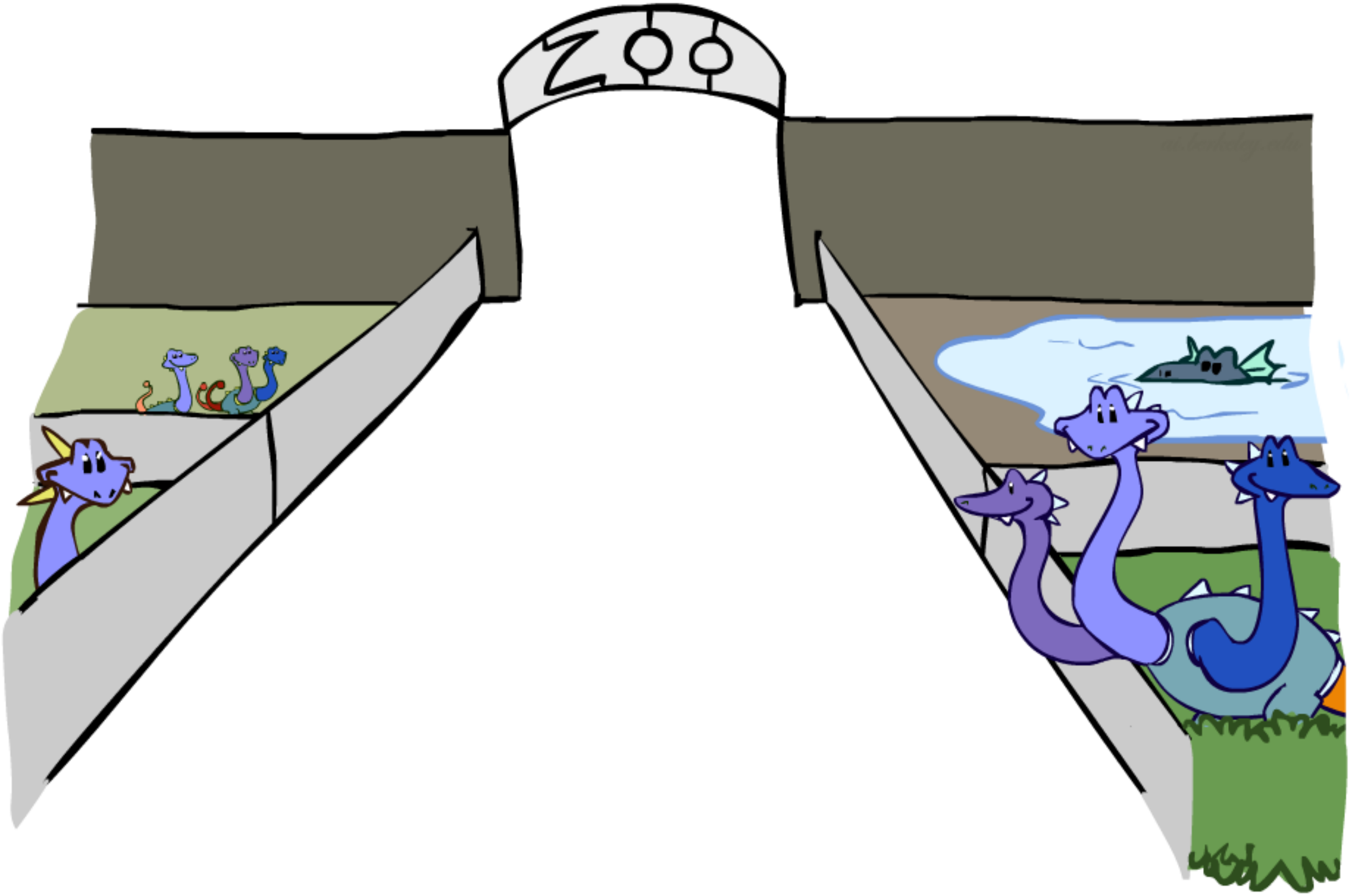
Variable Elimination



marginalize as you go

Factor Zoo

EXAMPLE



Joint Factors

FOUNDATIONS

- Joint Distribution $P(X, Y)$
  - Entries $P(x, y)$ for all x, y
  - $|X| \times |Y|$ matrix
  - $\sum_{x,y} P(x, y) = 1$

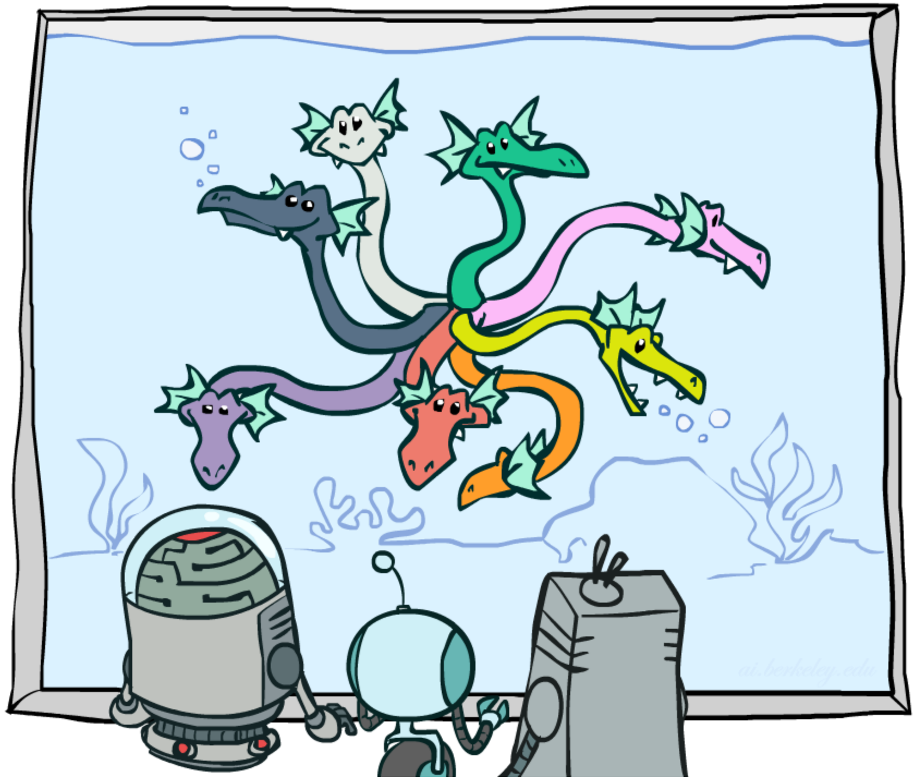
$P(T, W)$

T/W	sun	rain
hot	0.4	0.1
cold	0.2	0.3

- Projected Joint Distribution $P(x, Y)$
  - Slice of the joint distribution
  - Entries $P(x, y)$ for fixed x, all y
  - $|Y|$-element vector
  - $\sum_y P(x, y) = P(x)$

$P(\text{cold}, W)$

T/W	sun	rain
cold	0.2	0.3



Number of variables (capital letters) = **dimensionality**

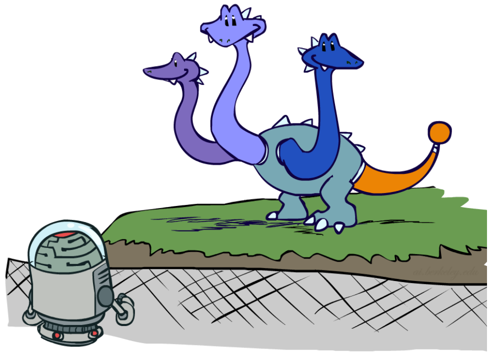
Conditional Factors

FOUNDATIONS

- Single Conditional $P(X|y)$
  - Entries $P(x|y)$ for fixed y, all x
  - $|X|$-element vector
  - $\sum_x P(x|y) = 1$

$P(W|\text{cold})$

T/W	sun	rain
cold	0.4	0.6



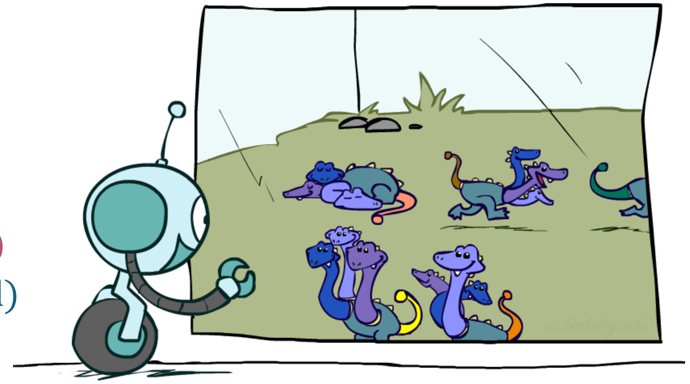
- Family of Conditionals $P(X|Y)$
  - Entries $P(x|y)$ for all x, y
  - $|X| \times |Y|$ matrix
  - $\sum_{x,y} P(x|y) = |Y| \cdot \sum_x P(x|y) = |Y|$

$P(W|T)$

T/W	sun	rain
hot	0.8	0.2
cold	0.4	0.6

$P(W|\text{hot})$

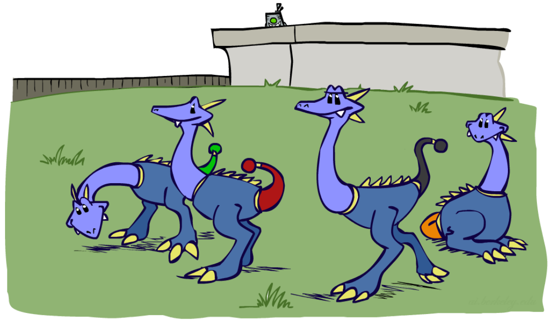
$P(W|\text{cold})$



- Specified Family $P(x|Y)$
  - Entries $P(x|y)$ for fixed x, all y
  - $|Y|$-element vector
  - $\sum_y P(x|y) = ?$

$P(\text{sun}|T)$

T/W	sun
hot	0.8
cold	0.4

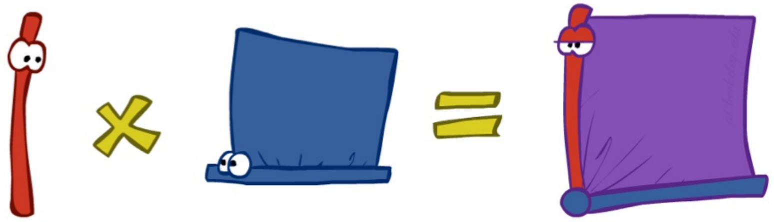


Questions?

live and on sli.do #cse473

Operation 1: Pointwise Product

DEFINITION



- Join factors using a **pointwise product**
  - Like a database join, not a matrix multiplication!
  - New factor has the union of variables of the original factors
  - Each entry is the product of the corresponding entries from the original factors

- Example:

$$P(W|T) \times P(T) = P(W,T)$$

$P(T,W)$

T	$P(T)$
hot	0.5
cold	0.5

×

$P(W|T)$

T/W	sun	rain
hot	0.8	0.2
cold	0.4	0.6

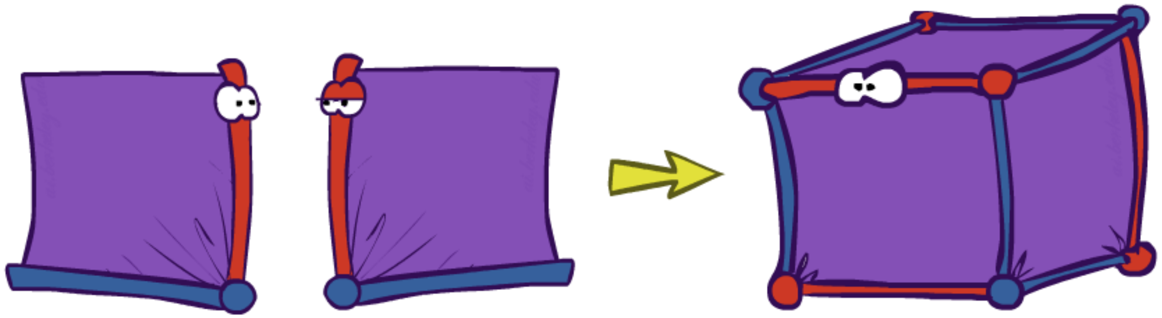
→

$P(T,W)$

T/W	sun	rain
hot	0.4	0.1
cold	0.2	0.3

Making Larger Factors

EXAMPLE



- Alarm Example:

$$P(A,J) \times P(A,M) = P(A,J,M)$$

$P(A,J)$

A/J	true	false
true	0.09	0.01
false	0.045	0.855

$\times$

$P(A,M)$

A/M	true	false
true	0.07	0.03
false	0.009	0.891

$P(A,J,M)$

$A = \text{false}$

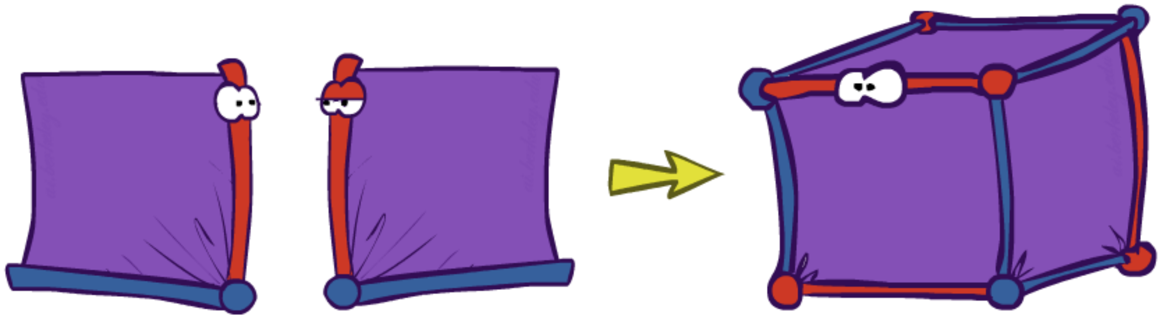
J/M	true	false
true		
false		0.762

$A = \text{true}$

J/M	true	false
true		
false		0.0003

Beware of Large Factors!

CONTEXT



- Example:
  - Variables U, V, W, X, each with range 10
  - $P(U,V)$, $P(V,W)$, $P(W,X)$ each have 10×10 entries

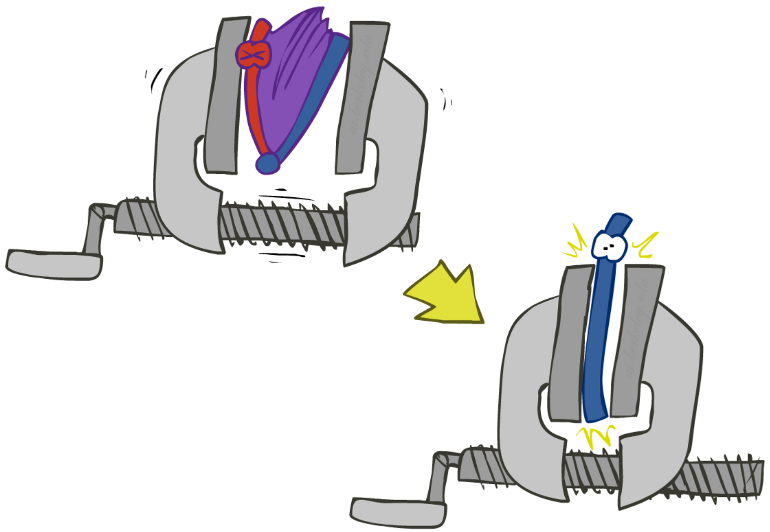
$$P(U,V) \times P(V,W) \times P(W,X) = P(U,V,W,X)$$

  - $P(U,V,W,X)$ has $10^4 = 10{,}000$ entries!

Operation 2: Marginalize

DEFINITION

- Sum out (**marginalize**) to eliminate a variable from a factor
  - Shrinks a factor to a smaller one
  - A *projection* operation



- Example:

$$\sum_t P(W,t) = P(W)$$

$P(W,T)$

T/W	**sun**	**rain**
hot	0.4	0.1
cold	0.2	0.3

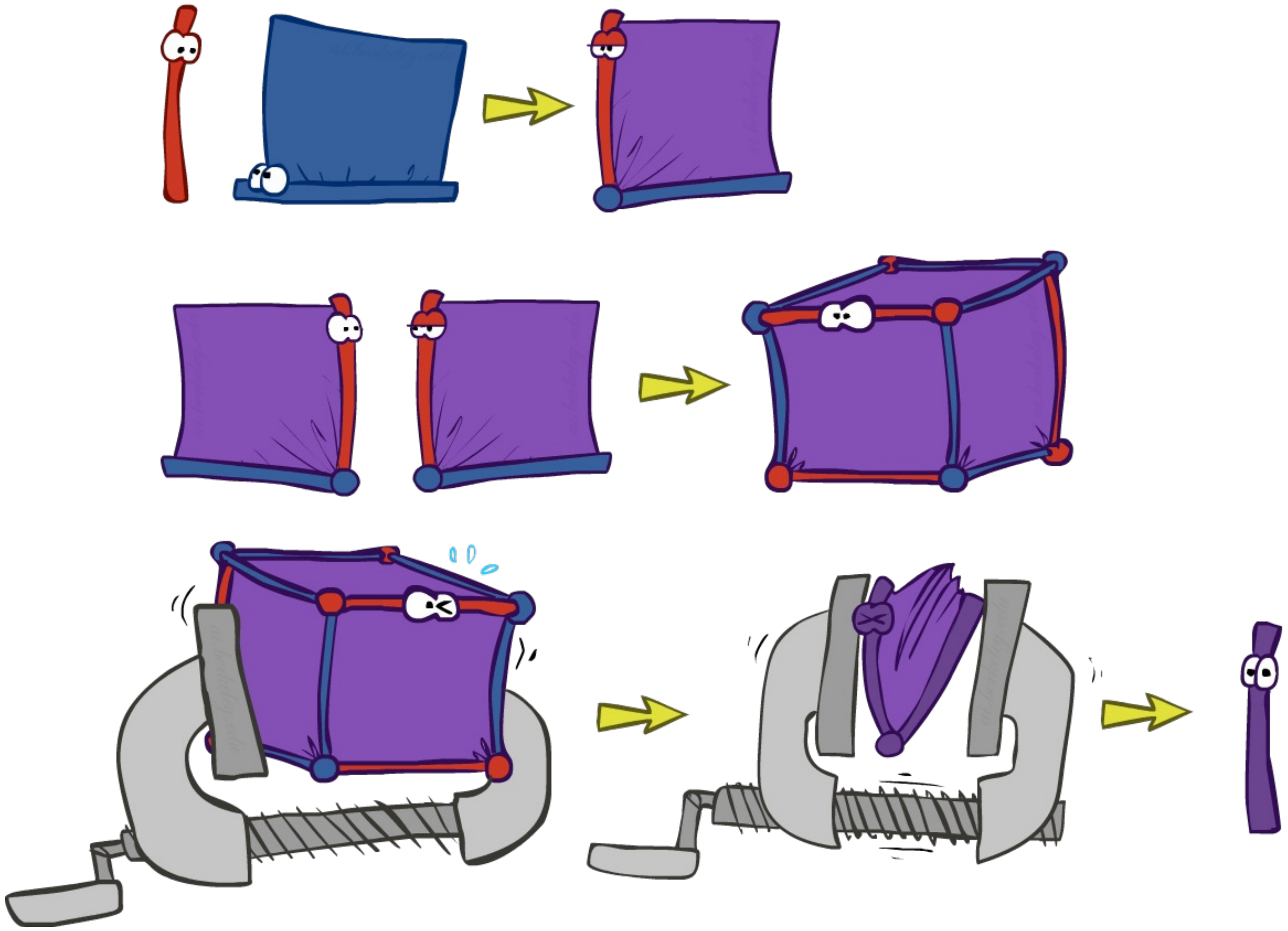
→

$P(W)$

W	$P(W)$
sun	0.6
rain	0.4

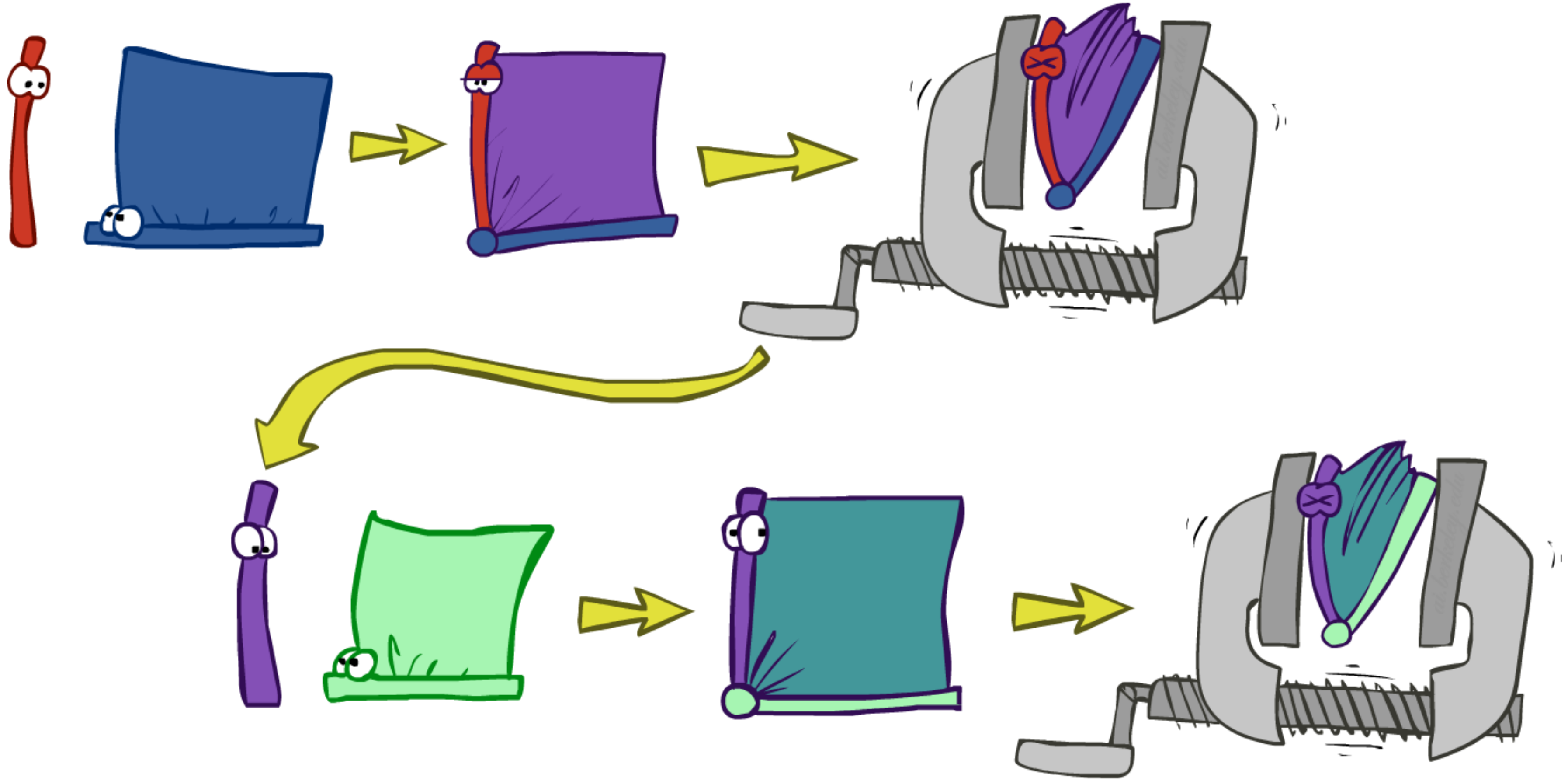
‘Long’ Way: Inference by Enumeration

CONTEXT



Shortcut: Variable Elimination

CONTEXT



Variable Elimination Implementation

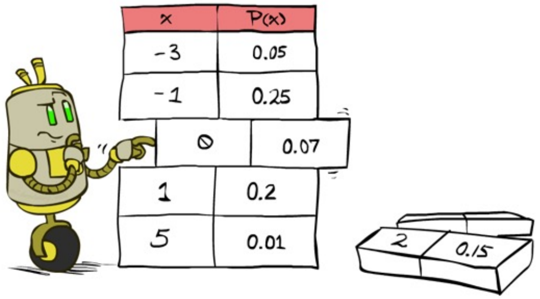
DEFINITION

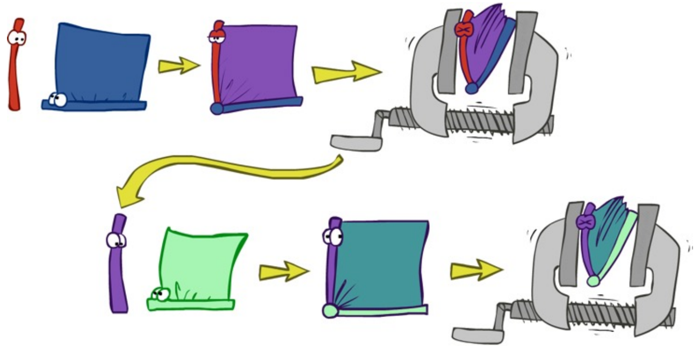
Query: $P(Q|E_1 = e_1, \ldots, E_k = e_k)$

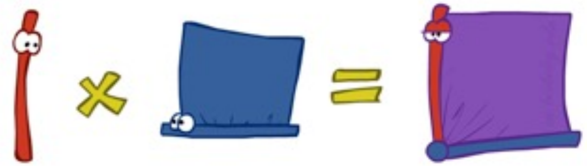
1. Start with initial factors
   - Local CPTs instantiated by evidence

2. While there are still hidden variables:
   - Pick a hidden variable H_j
   - Eliminate (sum out) H_j from the product of all factors mentioning H_j

3. Join remaining factors and normalize







Questions?

live and on sli.do #cse473

Alarm Variable Elimination (1)

EXAMPLE

- **Query:** $P(B|j,m)$
  - Initial Factors:

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

- Variable Elimination Round 1
  - Choose A, join factors with A:

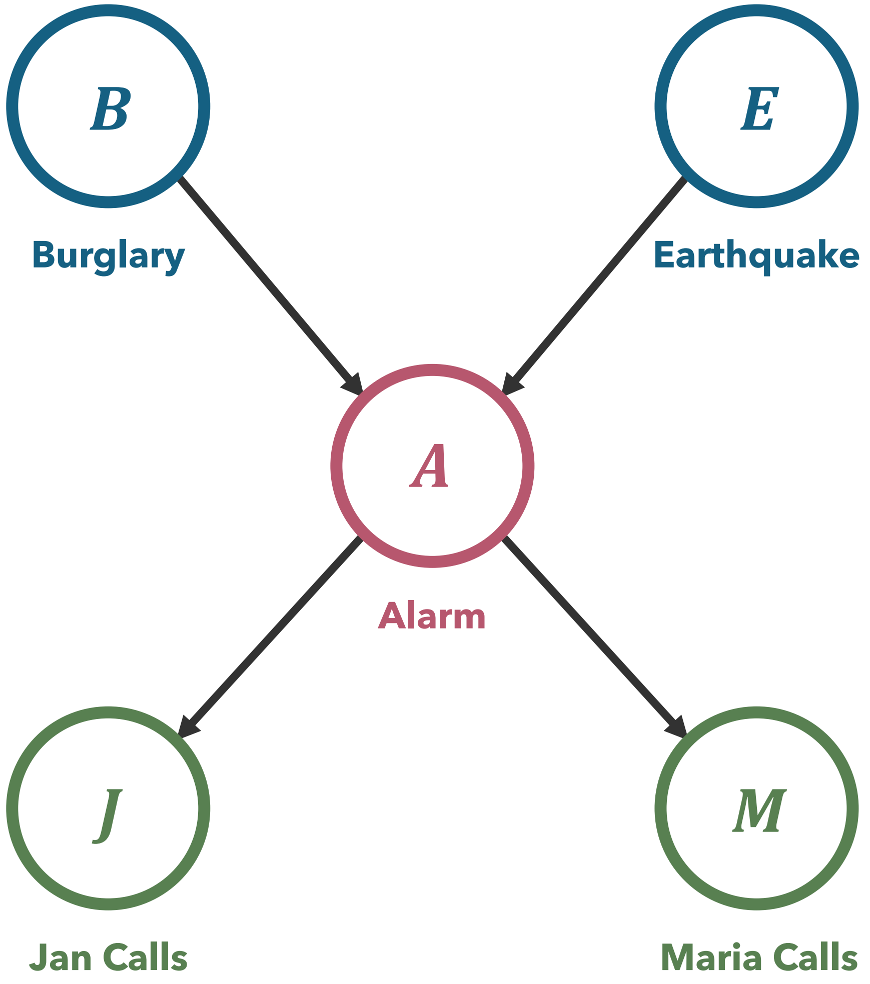
$$P(A|B,E) \times P(j|A) \times P(m|A) = P(j,m,A|B,E)$$

  - Marginalize A:

$$\sum_{a \in A} P(j,m,A|B,E) = P(j,m|B,E)$$

  - Remaining Factors:

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



Alarm Variable Elimination (2)

EXAMPLE

- **Query:** $P(B|j,m)$
  - Current Factors:

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

- Variable Elimination Round 2
  - Choose E, join factors with E:

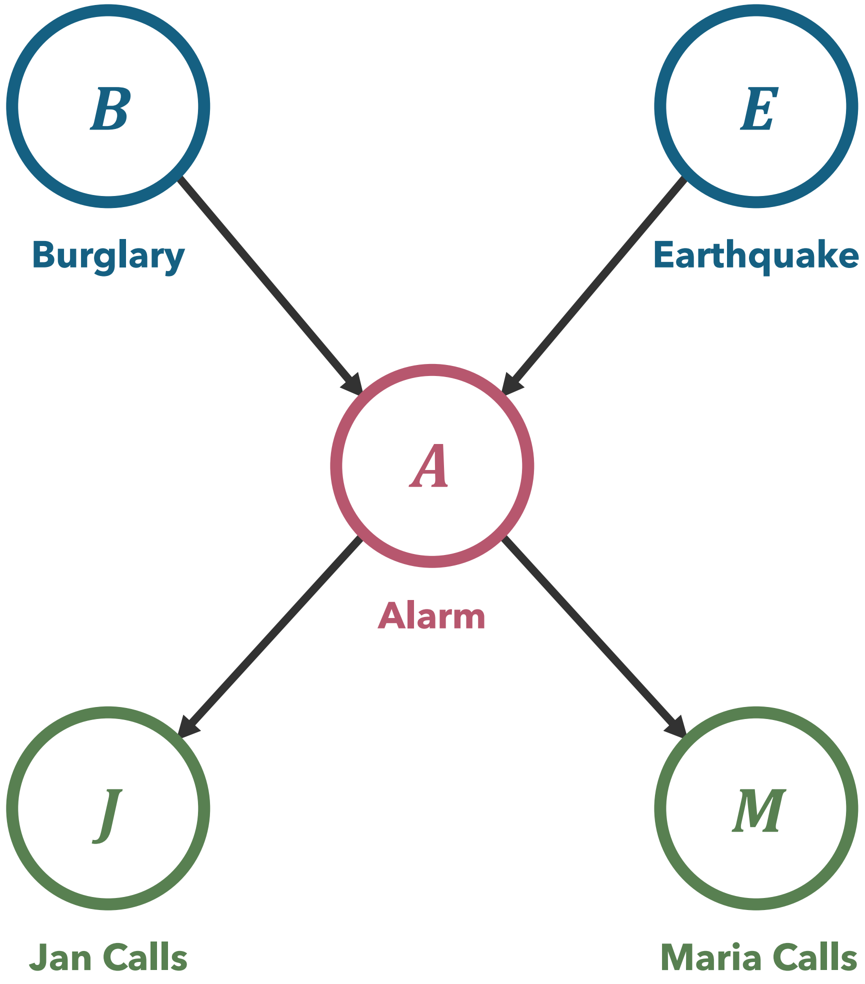
$$P(E) \times P(j,m|B,E) = P(j,m,E|B)$$

  - Marginalize E:

$$\sum_{e \in E} P(j,m,E|B) = P(j,m|B)$$

  - Remaining Factors:

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



Alarm Variable Elimination (3)

EXAMPLE

- **Query:** $P(B|j,m)$
  - Current Factors:

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

  - Note $P(B|j,m) \propto P(j,m,B)$

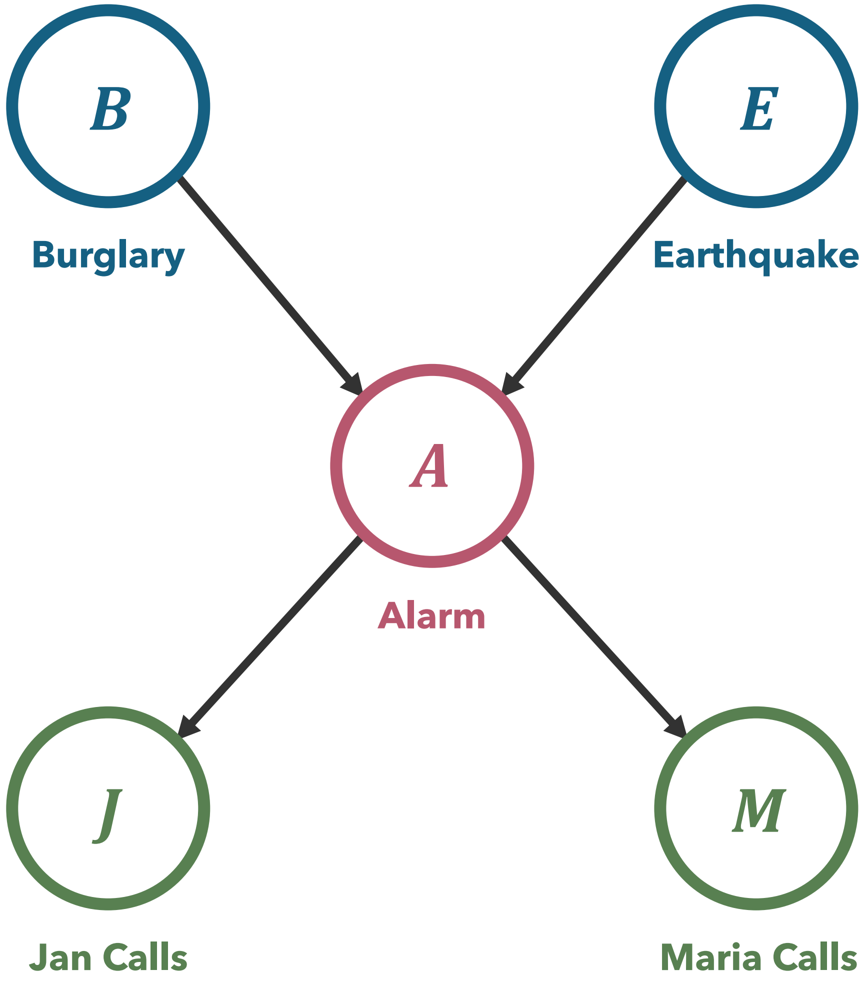
- Variable Elimination Round 3
  - Choose B, join factors with B:

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

  - Normalize by $P(j,m)$:

$$\alpha = \sum_{b \in B} P(j,m,b) = P(j,m)$$

  - Final Factor: $P(B|j,m) = \frac{1}{\alpha} \cdot P(j,m,B)$



Common Cause Variable Elimination

EXAMPLE

- **Query:** $P(X_3|y_1, y_2, y_3)$
  - Initial Factors:

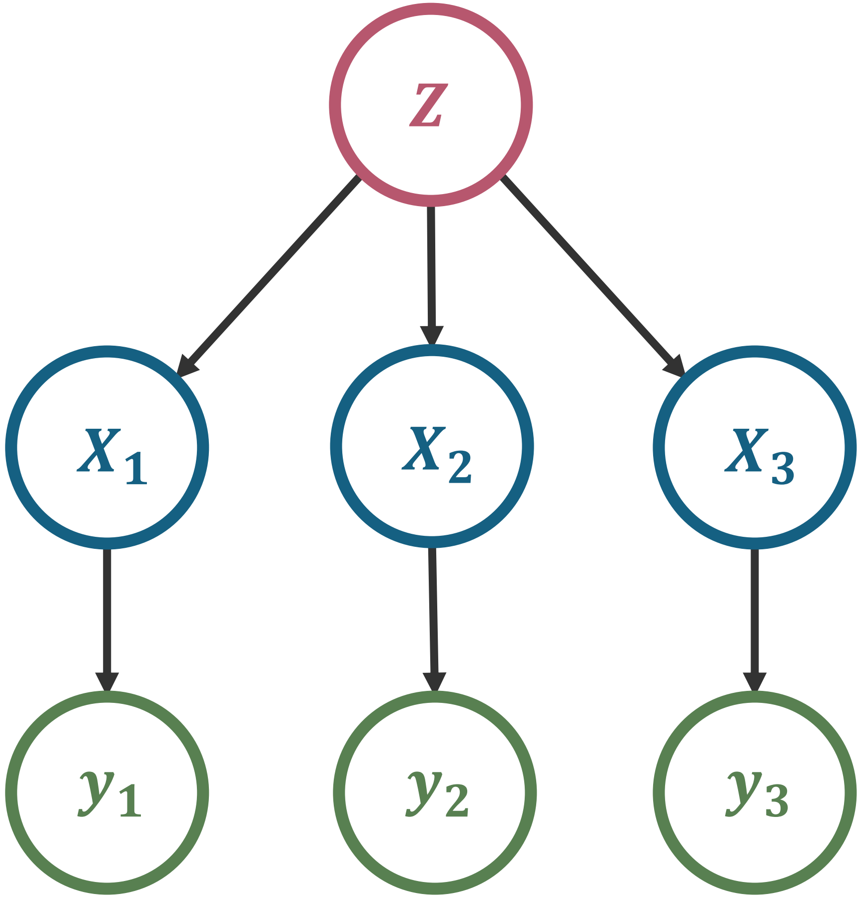
$P(Z)$	$P(X_1|Z)$	$P(X_2|Z)$	$P(X_3|Z)$	$P(y_1|X_1)$	$P(y_2|X_2)$	$P(y_3|X_3)$

- Variable Elimination
  - Eliminate X_1: $f_1(Z, y_1) = \sum_{x_1} P(x_1|Z) \cdot P(y_1|x_1)$
  - Eliminate X_2: $f_2(Z, y_2) = \sum_{x_2} P(x_2|Z) \cdot P(y_2|x_2)$
  - Eliminate Z: $f_3(X_3, y_1, y_2) = \sum_z P(z) \cdot f_1(Z, y_1) \cdot f_2(Z, y_2) \cdot P(X_3|z)$
  - Join remaining factors:

$f_4(X_3, y_1, y_2, y_3) = P(X_3, y_1, y_2, y_3) = P(y_3|X_3) \cdot f_3(X_3, y_1, y_2)$

  - Normalize over X_3:

$$P(X_3|y_1, y_2, y_3) = \frac{f_4(X_3, y_1, y_2, y_3)}{\sum_{x_3} f_4(x_3, y_1, y_2, y_3)}$$



Questions?

live and on sli.do #cse473

Try It Yourself

EXAMPLE

- Evidence
  - $C = c_1, B = b_2$
- Initial Factors

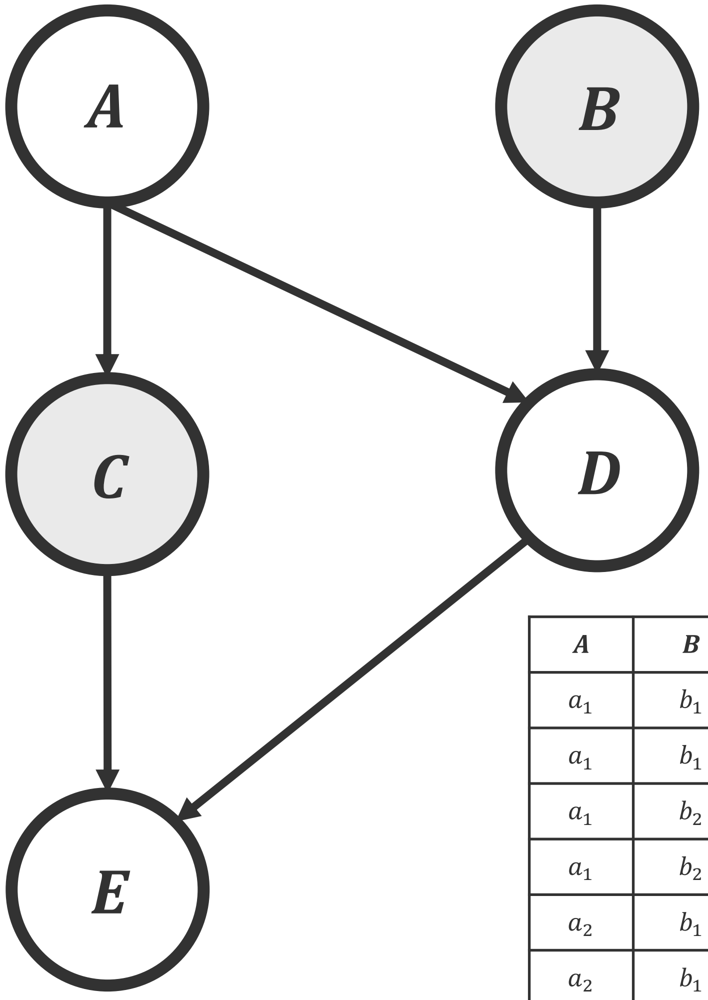
$P(A)$	$P(b_2)$	$P(c_1 \vert A)$	$P(D \vert A, b_2)$	$P(E \vert c_1, D)$

- What is $\boldsymbol{P(A, b_2, D)}$?
- What is $\boldsymbol{P(c_1 \vert E)}$?

A	$P(A)$
a_1	0.2
a_2	0.8

B	$P(B)$
b_1	0.6
b_2	0.4

A/C	c_1	c_2
a_1	0.3	0.7
a_2	0.5	0.5



C	D	E	$P(E \vert C, E)$
c_1	d_1	e_1	0.5
c_1	d_1	e_2	0.5
c_1	d_2	e_1	0.7
c_1	d_2	e_2	0.3
c_2	d_1	e_1	0.6
c_2	d_1	e_2	0.4
c_2	d_2	e_1	1
c_2	d_2	e_2	0

A	B	D	$P(D \vert A, B)$
a_1	b_1	d_1	0
a_1	b_1	d_2	1
a_1	b_2	d_1	0.9
a_1	b_2	d_2	0.1
a_2	b_1	d_1	0.25
a_2	b_1	d_2	0.75
a_2	b_2	d_1	0.5
a_2	b_2	d_2	0.5

Order Matters

FOUNDATIONS

- Query: $P(C)$
- With elimination ordering A, B, Z:
  - $P(C) = \alpha \sum_{z,b,a} P(z) \cdot P(a|z) \cdot P(b|z) \cdot P(C|z)$

    $= \alpha \boxed{\sum_z P(z) \cdot \boxed{[\sum_b P(b|z) \cdot \boxed{[\sum_a P(a|z) \cdot P(C|z)]}]}}$
  - Largest factor has two variables (C, Z)
- With elimination ordering Z, A, B:
  - $P(C) = \alpha \sum_{b,a,z} P(z) \cdot P(a|z) \cdot P(b|z) \cdot P(C|z)$

    $= \alpha \boxed{\sum_b \boxed{[\sum_a \boxed{[\sum_z P(z) \cdot P(a|z) \cdot P(b|z) \cdot P(C|z)]}]}}$

$f_1(C, Z)$

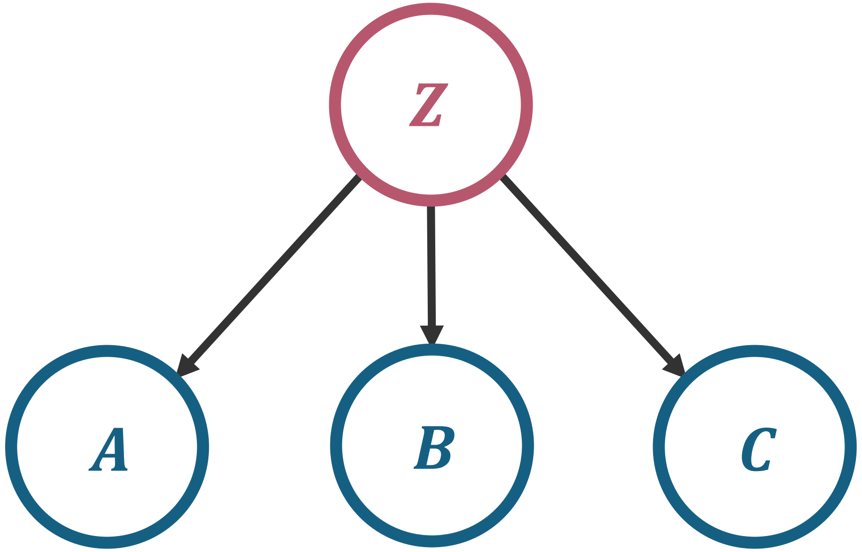
$f_2(C, Z)$

$f_3(C)$

$f_1(A, B, C)$

$f_2(B, C)$

$f_3(C)$



Properties of Variable Elimination

DEFINITION

The computational and space complexity of variable elimination is determined by the **largest factor**

- Elimination ordering greatly affects the size of the largest factor
  - In the shallow tree scenario:
    - d^n when eliminating the root first
    - d^k when eliminating leaves first, where k is the number of variables involves in the query
- Does there always exist an ordering that only results in small factors?
  - No. Such an ordering cannot be guaranteed. Sometimes the network structure will require a large factor

QUESTIONS

Questions?

live and on sli.do #cse473

that's it for today!

SUMMARY

- Implications of independence and conditional independence
- Bayes nets encode joint distributions by tracking conditional probabilities

UPCOMING

- How to compute conditional probabilities of queries given evidence?

REMINDERS

- Complete **Practice Problem 12**
- Fill out the **Mid-Quarter Feedback Survey**
- Do **Project 3** (due 7.30)
- Do **Homework 3** (due 8.6)