Lecture 11: Bayes Nets: Inference

Shuai Li
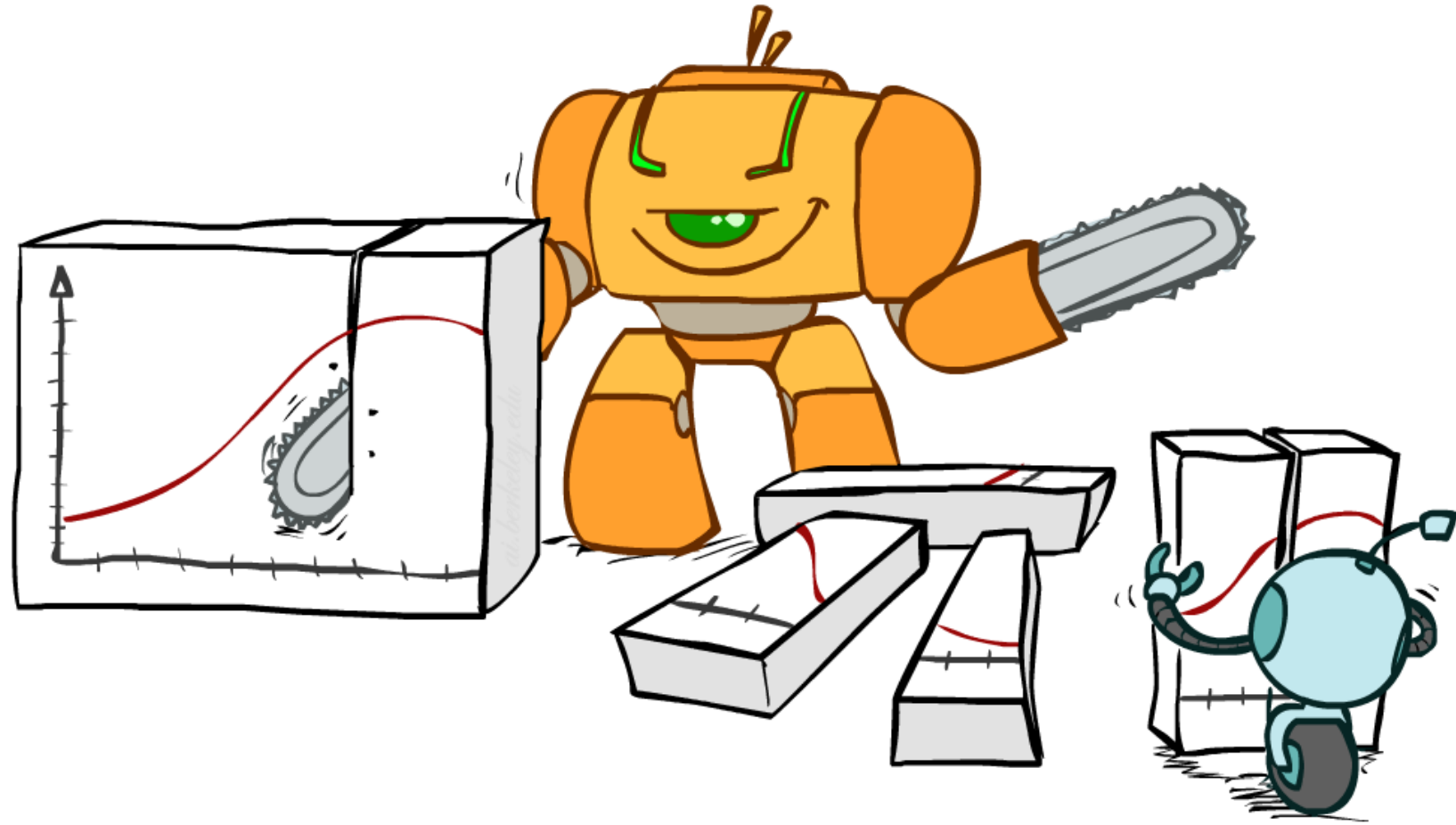
John Hopcroft Center, Shanghai Jiao Tong University

https://shuaili8.github.io

https://shuaili8.github.io/Teaching/CS410/index.html

Bayes Rule

Bayes' Rule

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

That's my rule!

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

- Why is this at all helpful?

  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Foundation of many systems (e.g. ASR, MT)

- In the running for most important AI equation!

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- Example:
  - M: meningitis, S: stiff neck

$$\left. \begin{array}{l} P(+m) = 0.0001 \\ P(+s|+m) = 0.8 \\ P(+s|-m) = 0.01 \end{array} \right\} \text{ Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

  - Note: posterior probability of meningitis still very small
  - Note: you should still get stiff necks checked out! Why?

Quiz: Bayes' Rule

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

- What is P(W | dry) ?

Quiz: Bayes' Rule 2

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

- What is P(W | dry) ?

P(sun|dry) ∝ P(dry|sun)P(sun) = .9*.8 = .72
P(rain|dry) ∝ P(dry|rain)P(rain) = .3*.2 = .06
P(sun|dry)=12/13
P(rain|dry)=1/13

Ghostbusters, Revisited

- Let's say we have two distributions:
  - Prior distribution over ghost location: P(G)
    - Let's say this is uniform
  - Sensor reading model: P(R | G)
    - Given: we know what our sensors do
    - R = reading color measured at (1,1)
    - E.g. P(R = yellow | G=(1,1)) = 0.1

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- We can calculate the posterior distribution P(G|r) over ghost locations given a reading using Bayes' rule:
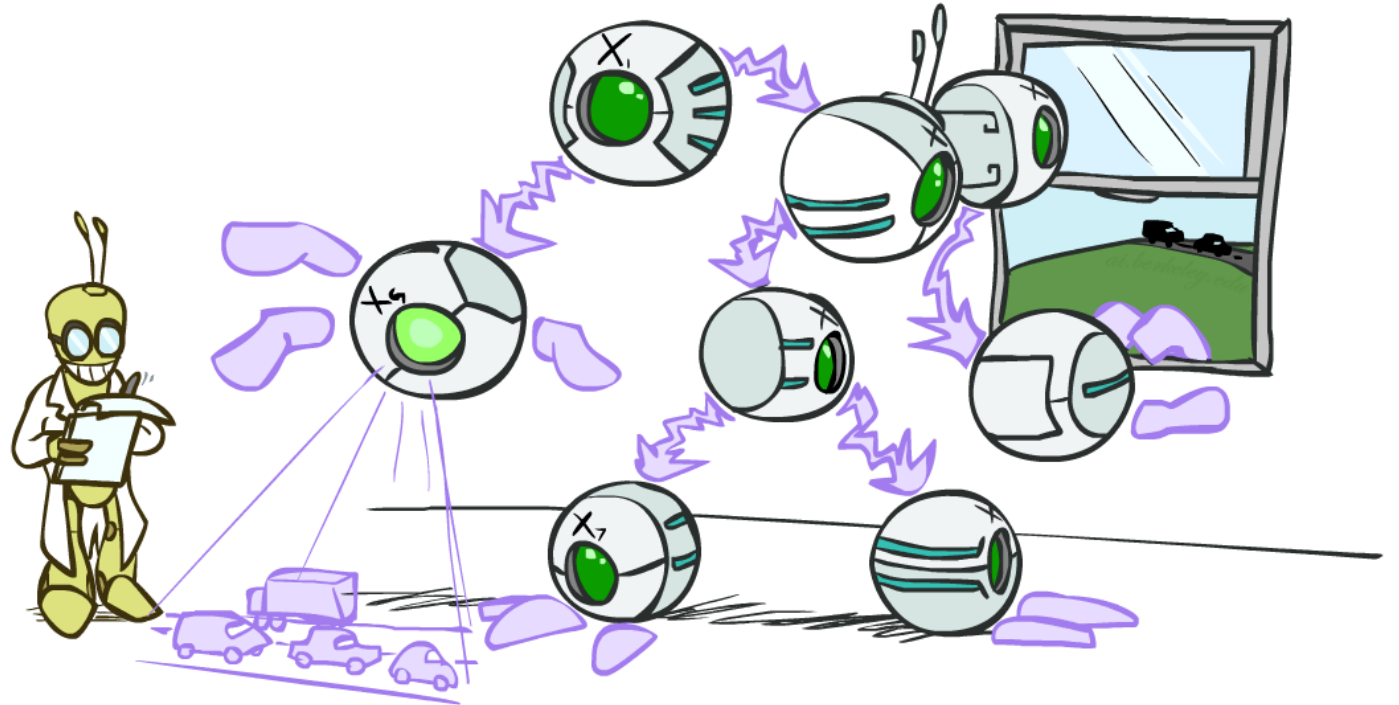
$$P(g|r) \propto P(r|g)P(g)$$

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

[Demo: Ghostbuster – with probability (L12D2)]

Video of Demo Ghostbusters with Probability
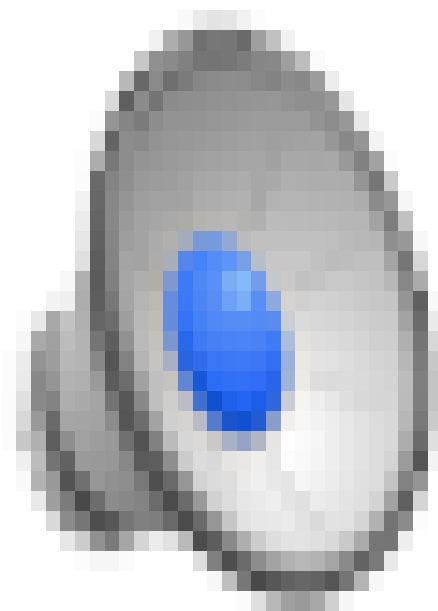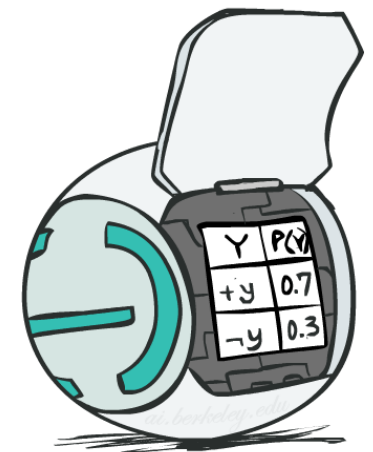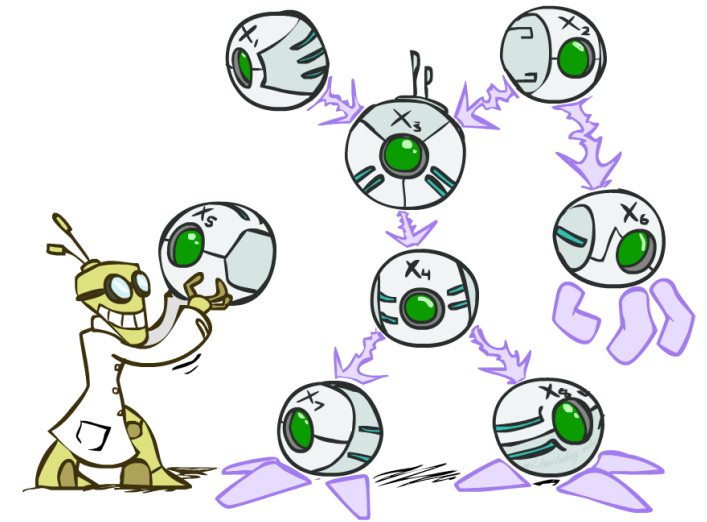
Inference

Recall: Bayes' Net Representation

- A directed, acyclic graph, one node per random variable

- A conditional probability table (CPT) for each node

  - A collection of distributions over X, one for each combination of parents'' values

  $$P(X|a_1 \ldots a_n)$$

- Bayes' nets implicitly encode joint distributions

  - As a product of local conditional distributions

  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

  $$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

Inference

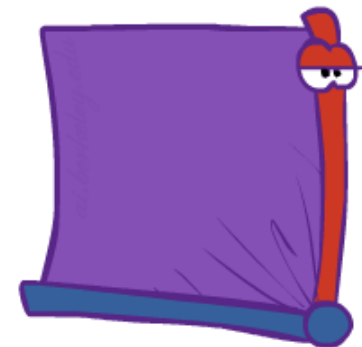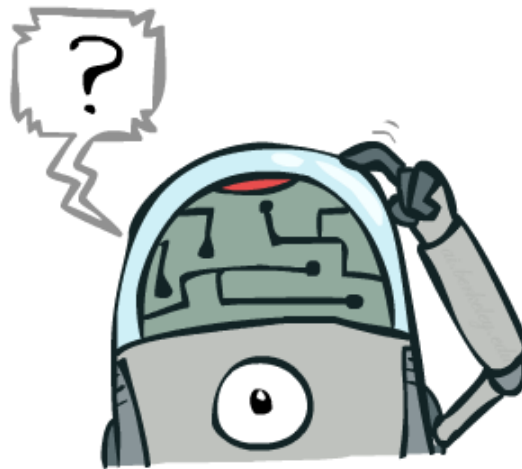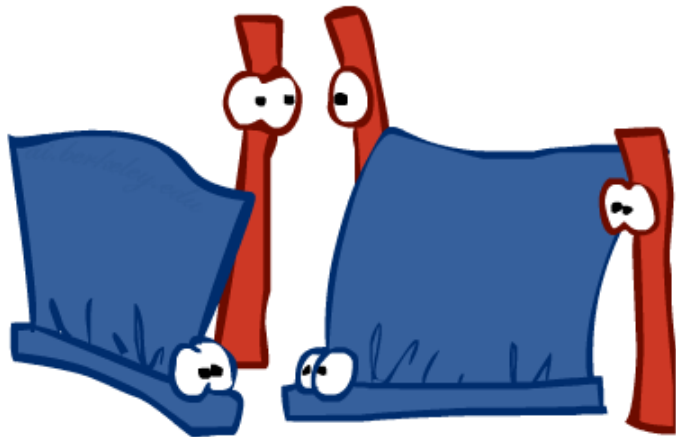- Inference: calculating some useful quantity from a joint probability distribution

- Posterior probability

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

- Most likely explanation:

$$\text{argmax}_q \ P(Q = q|E_1 = e_1 \ldots)$$

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)?

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

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

Inference by Enumeration in Joint Distributions

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$ $\left.\begin{array}{c} \\ \\ \\ \end{array}\right\}$ $X_1, X_2, \ldots X_n$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

    All variables

■ We want:

Works fine with multiple query variables, too

$$P(Q|e_1 \ldots e_k)$$

■ Step 1: Select the entries consistent with the evidence

■ Step 2: Sum out H to get joint of Query and evidence

$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(\underbrace{Q, h_1 \ldots h_r, e_1 \ldots e_k}_{X_1, X_2, \ldots X_n})$$

■ Step 3: Normalize

$$\times \frac{1}{Z}$$
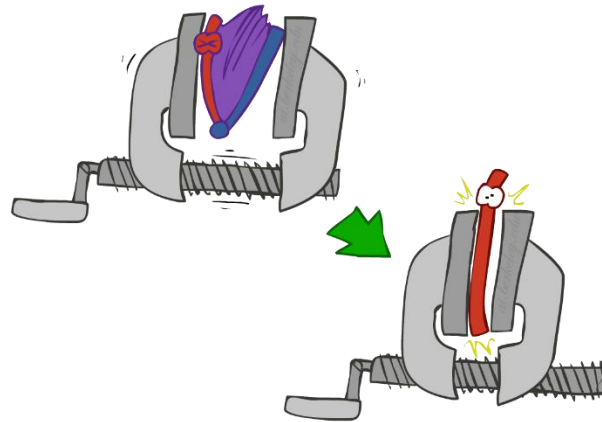
$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration in Bayes' Net

- Given unlimited time, inference in BNs is easy

$$P(B \mid +j, +m) \quad \propto_B P(B, +j, +m)$$

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

$$= \sum_{e,a} P(B)P(e)P(a|B,e)P(+j|a)P(+m|a)$$
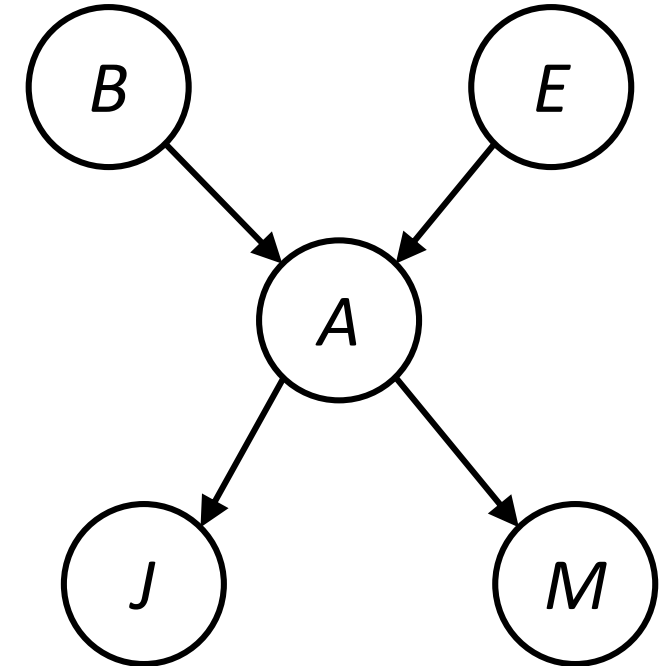
$$=P(B)P(+e)P(+a|B,+e)P(+j|+a)P(+m|+a) + P(B)P(+e)P(-a|B,+e)P(+j|-a)P(+m|-a)$$
$$P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$
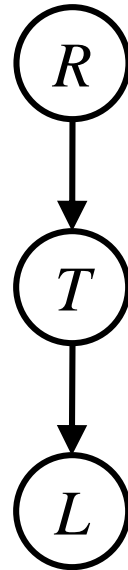
Example: Traffic Domain

- **Random Variables**
  - R: Raining
  - T: Traffic
  - L: Late for class!

$$P(L) = ?$$

$$= \sum_{r,t} P(r, t, L)$$

$$= \sum_{r,t} P(r)P(t|r)P(L|t)$$

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Inference by Enumeration: Procedural Outline

- Track objects called factors
- Initial factors are local CPTs (one per node)

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

- Any known values are selected
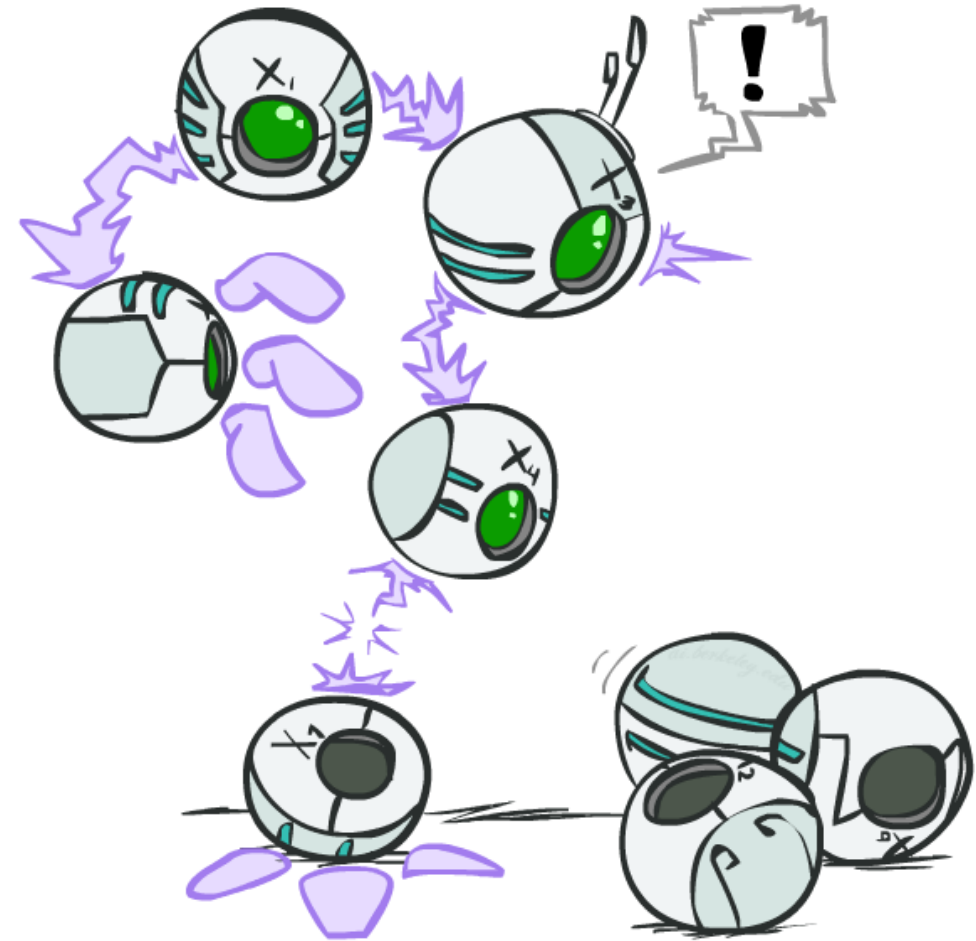 - E.g. if we know $L = +\ell$, the initial factors are

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

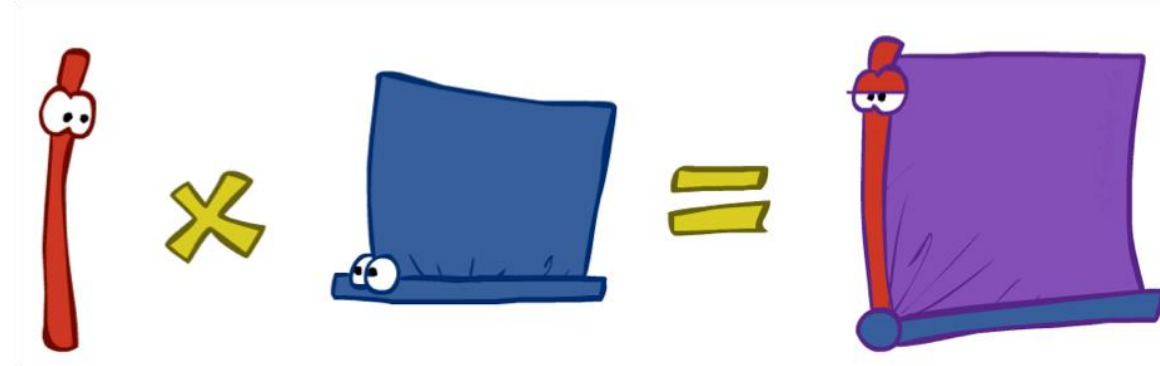| +r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(+\ell|T)$
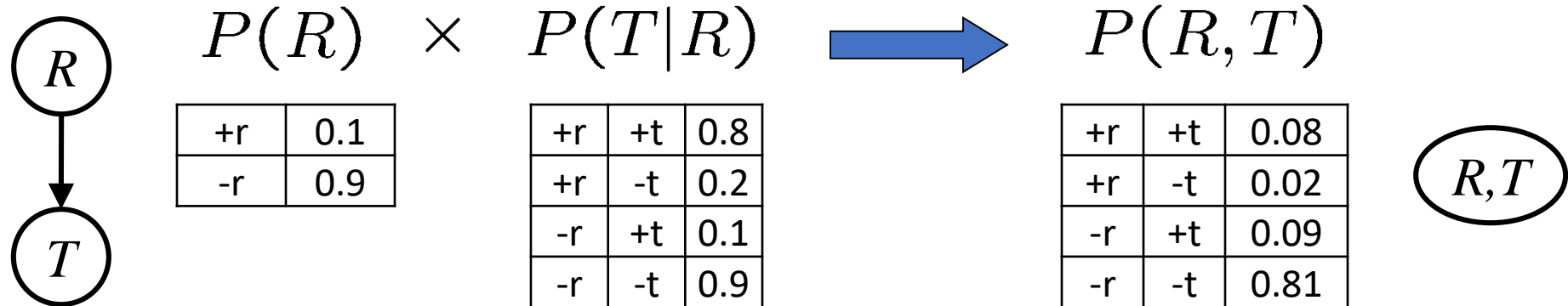
+t	+l	0.3
-t	+l	0.1

- Procedure: Join all factors, then sum out all hidden variables

Operation 1: Join Factors

- First basic operation: joining factors

- Combining factors:
  - Just like a database join
  - Get all factors over the joining variable
  - Build a new factor over the union of the variables involved

- Example: Join on R

$$P(R) \quad \times \quad P(T|R) \quad \longrightarrow \quad P(R,T)$$

+r	0.1
-r	0.9

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

R,T

- Computation for each entry: pointwise products $\forall r, t: \quad P(r,t) = P(r) \cdot P(t|r)$

Example: Multiple Joins

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$
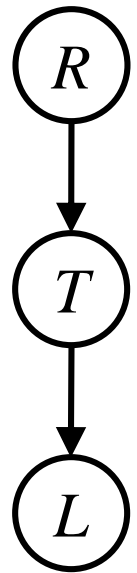
+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Join R →

$P(R,T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Join T →
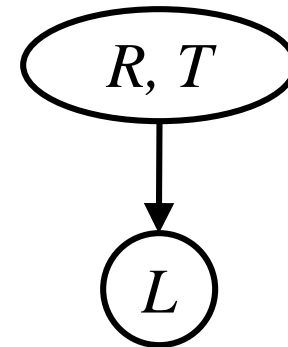
R, T, L

$P(R,T,L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

18
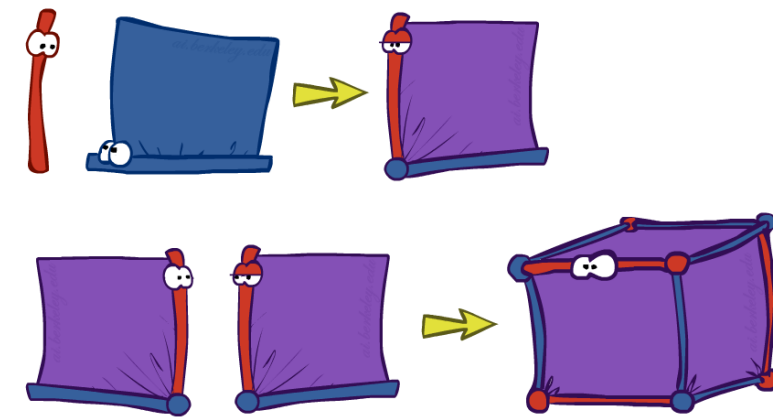
Operation 2: Eliminate

- Second basic operation: marginalization

- Take a factor and sum out a variable

  - Shrinks a factor to a smaller one

  - A projection operation

- Example:

$$P(R, T)$$
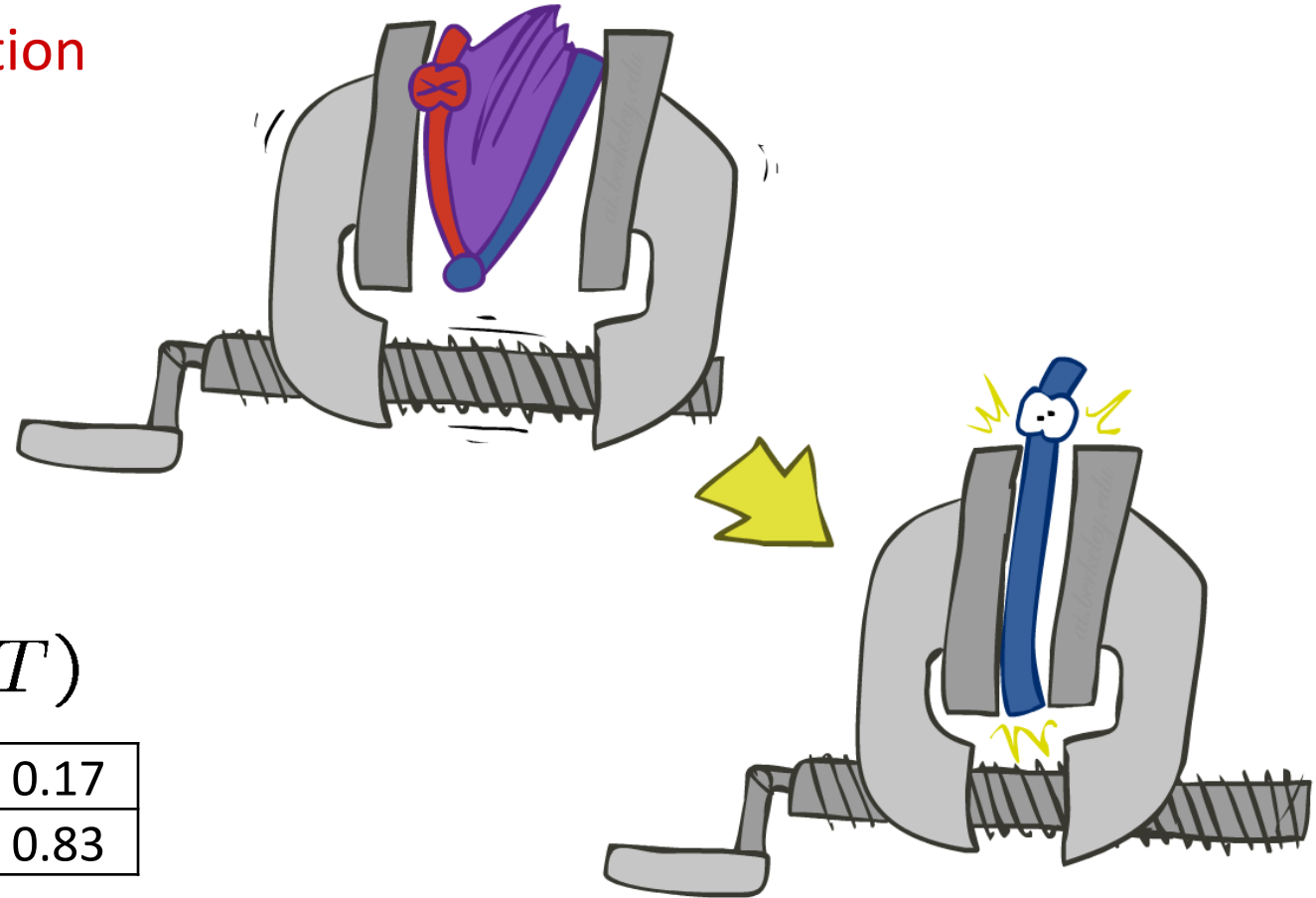
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

sum R

$$P(T)$$

+t	0.17
-t	0.83

Multiple Elimination

R, T, L

$P(R, T, L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

Sum out R

T, L

$P(T, L)$

+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

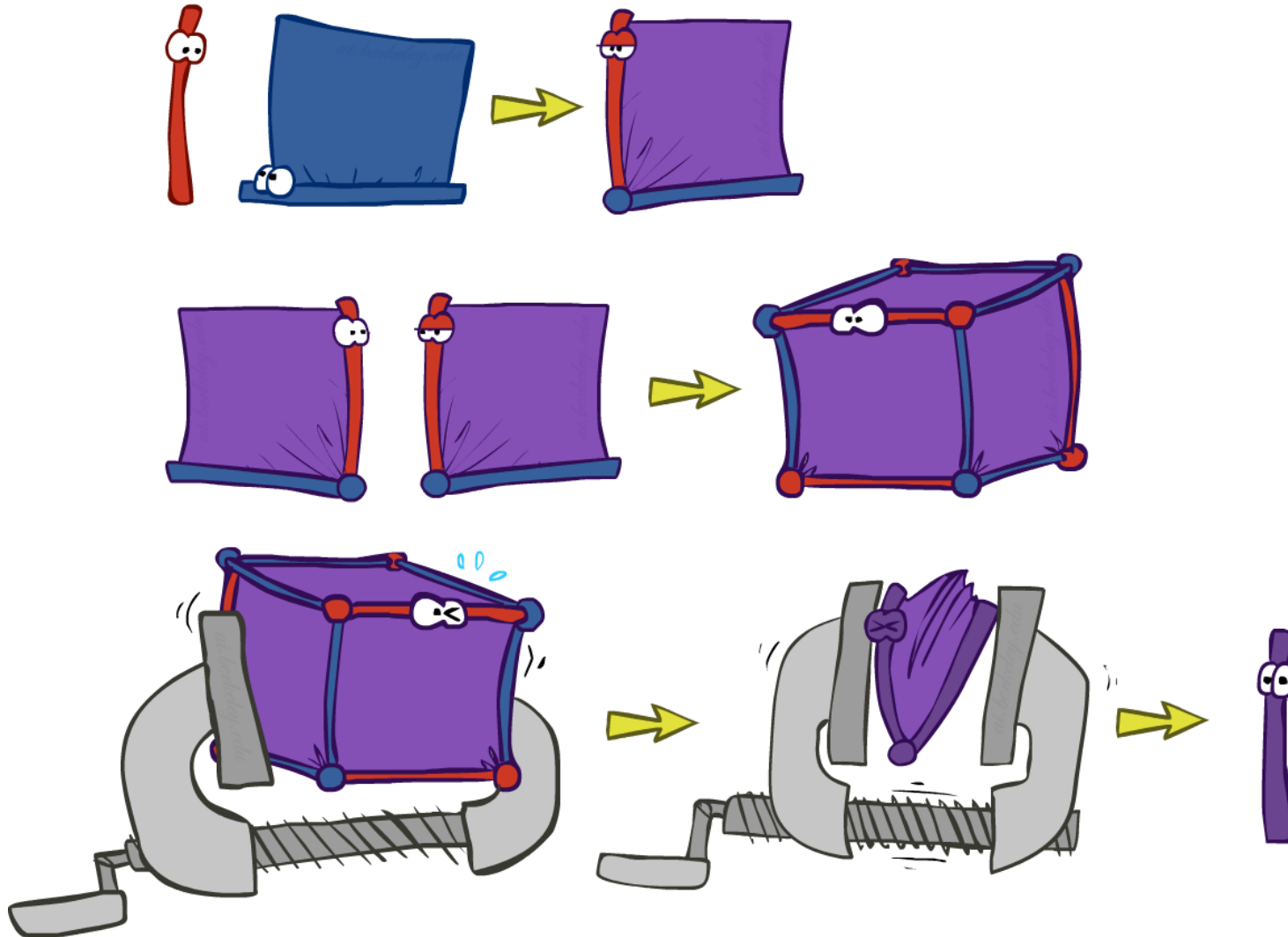Sum out T

L

$P(L)$

+l	0.134
-l	0.866
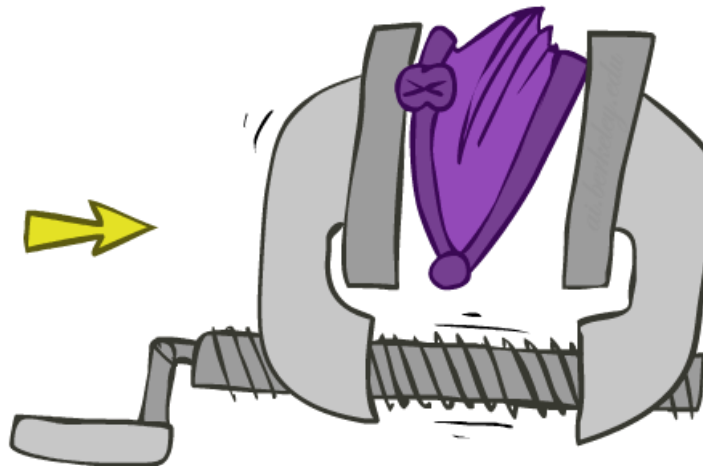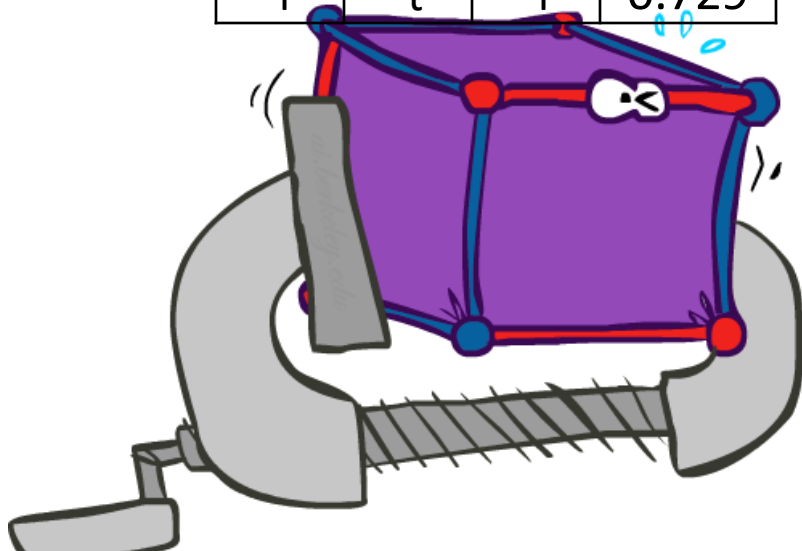
Thus Far: Multiple Join, Multiple Eliminate (= Inference by Enumeration)

$P(R)$

$P(T|R)$ $\Longrightarrow$ $P(R, T, L)$ $\Longrightarrow$ $P(L)$

$P(L|T)$

Thus Far: Multiple Join, Multiple Eliminate (= Inference by Enumeration)

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

$$P(B \mid j, m) = \alpha\, P(B, j, m)$$
$$= \alpha \sum_{e,a} P(B, e, a, j, m)$$
$$= \alpha \sum_{e,a} P(B)\, P(e)\, P(a \mid B, e)\, P(j \mid a)\, P(m \mid a)$$
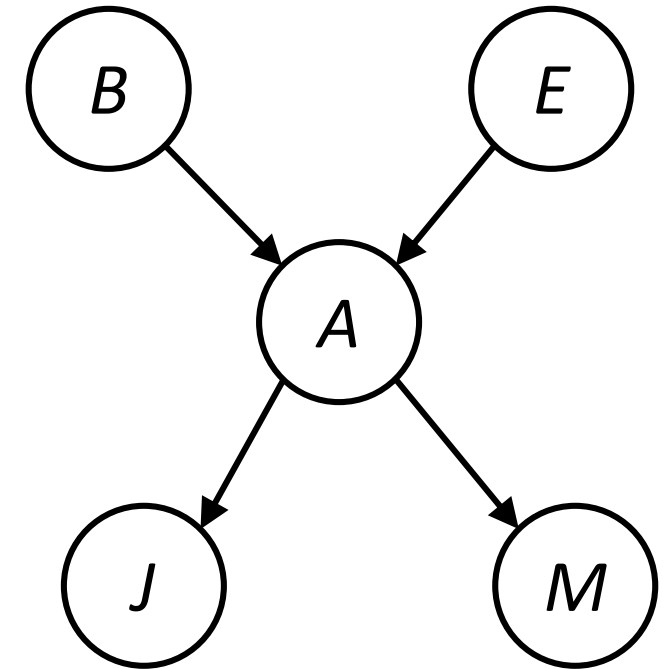
- So inference in Bayes nets means computing sums of products of numbers: sounds easy!!

- Problem: sums of *exponentially many* products!

Can we do better?

- Consider
  - $x_1 y_1 z_1 + x_1 y_1 z_2 + x_1 y_2 z_1 + x_1 y_2 z_2 + x_2 y_1 z_1 + x_2 y_1 z_2 + x_2 y_2 z_1 + x_2 y_2 z_2$
  - 16 multiplies, 7 adds
  - Lots of repeated subexpressions!
- Rewrite as
  - $(x_1 + x_2)(y_1 + y_2)(z_1 + z_2)$
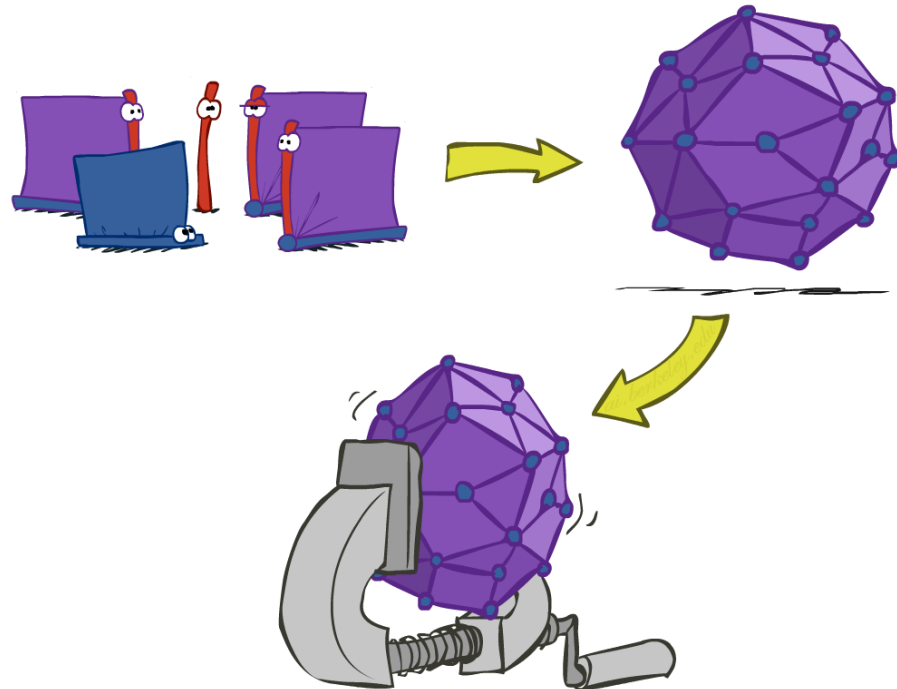 - 2 multiplies, 3 adds

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

- Lots of repeated subexpressions!

Variable Elimination

Inference by Enumeration vs. Variable Elimination

- Why is inference by enumeration so slow?
  - You join up the whole joint distribution before you sum out the hidden variables

- **Idea: interleave joining and marginalizing!**
  - Called "Variable Elimination"
  - Still NP-hard, but usually much faster than inference by enumeration
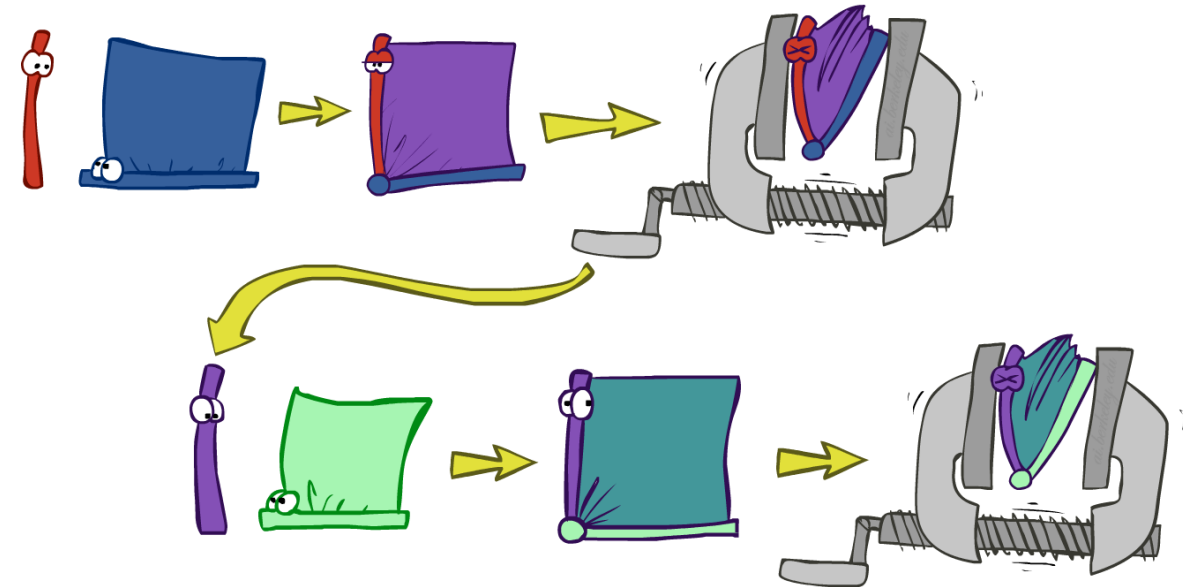
Inference Overview

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

- 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

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

$$= \alpha \, [P(h_1) \, P(q|h_1) \, P(e|q) + P(h_2) \, P(q|h_2) \, P(e|q)]$$

$H \rightarrow Q \rightarrow E$
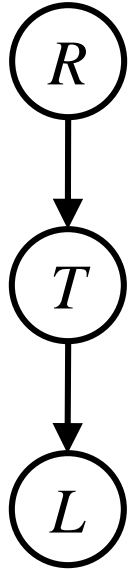
- But we can be more efficient

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

$$= \alpha \, P(e|q) \, [P(h_1) P(q|h_1) + P(h_2) P(q|h_2)]$$

$$= \alpha \, P(e|q) \, P(q)$$

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

Traffic Domain

$P(L) = ?$

$R \rightarrow T \rightarrow L$

• 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

Marginalizing Early (= Variable Elimination)
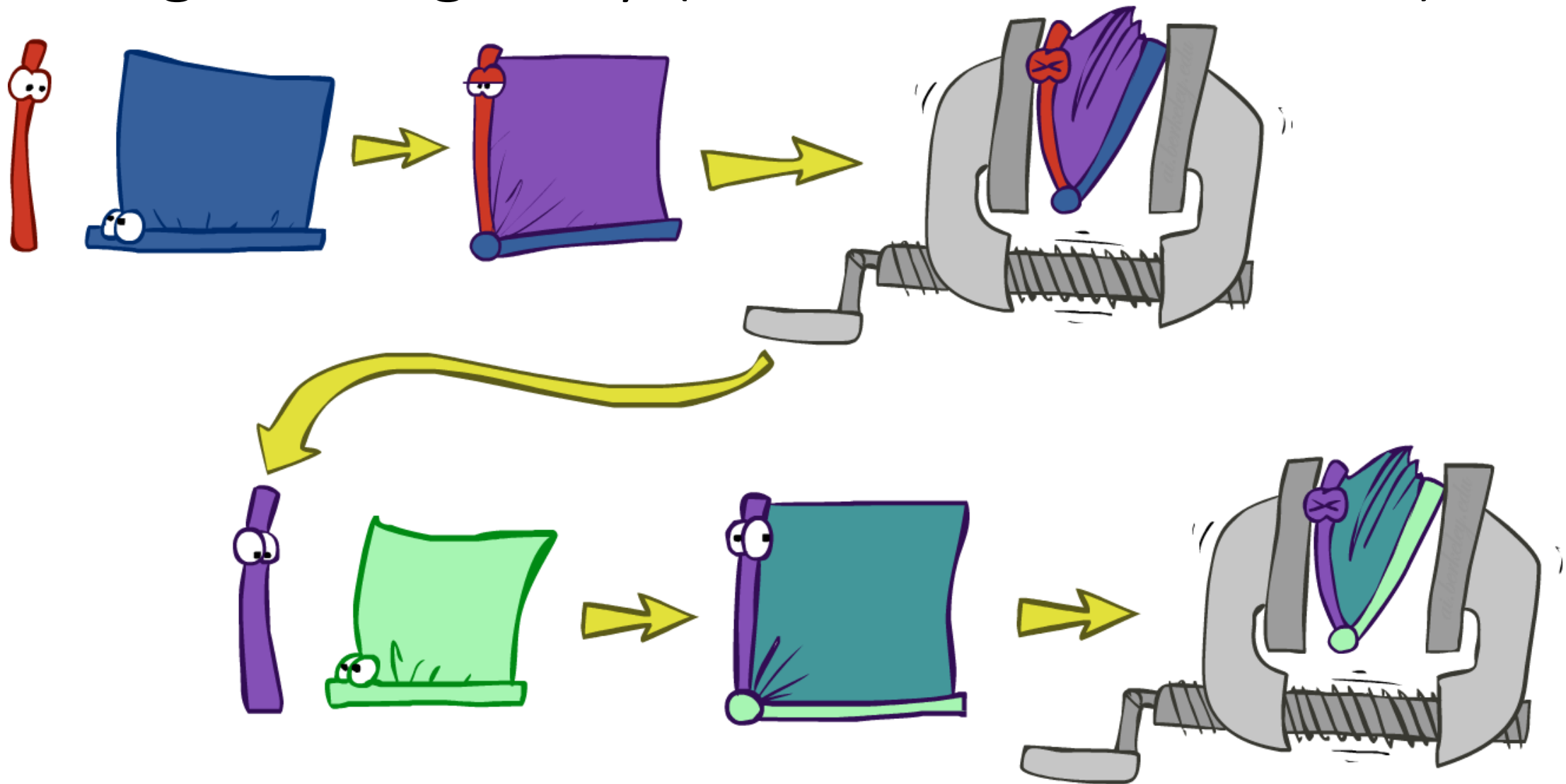
Answer Any Query from Bayes Net (Previous)

Bayes Net

Joint

Query

$$P(a \mid e)$$
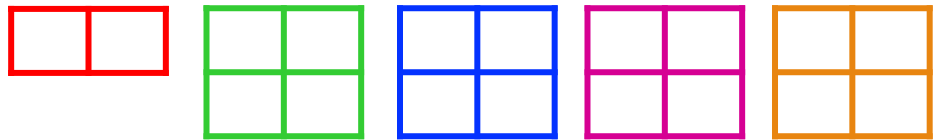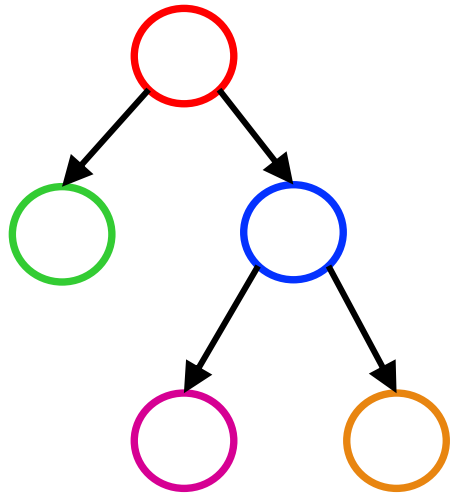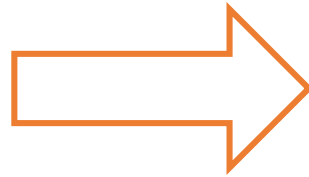
$P(A) \; P(B|A) \; P(C|A) \; P(D|C) \; P(E|C)$

Next: Answer Any Query from Bayes Net

Bayes Net

Query

$$P(a \mid e)$$

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

Marginalizing Early! (aka VE)

$P(R)$

+r	0.1
-r	0.9

Join R →

$P(R, T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

Sum out R →

$P(T)$

+t	0.17
-t	0.83

Join T →

Sum out T →

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

R, T → L

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

T → L

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

T, L

$P(T, L)$

+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

L

$P(L)$

+l	0.134
-l	0.866

32
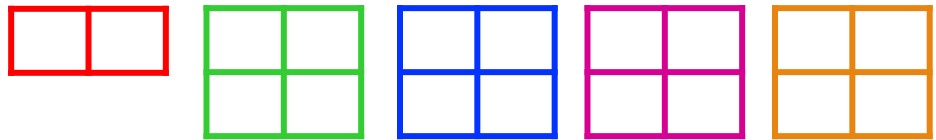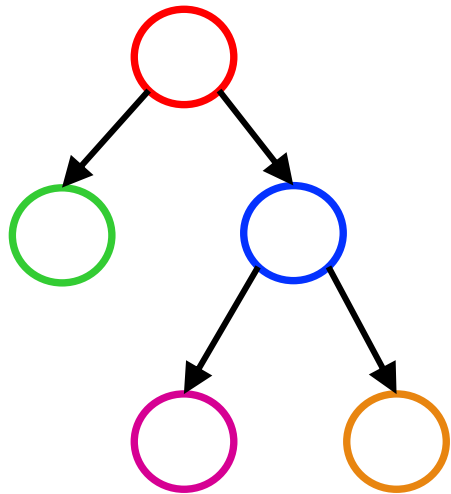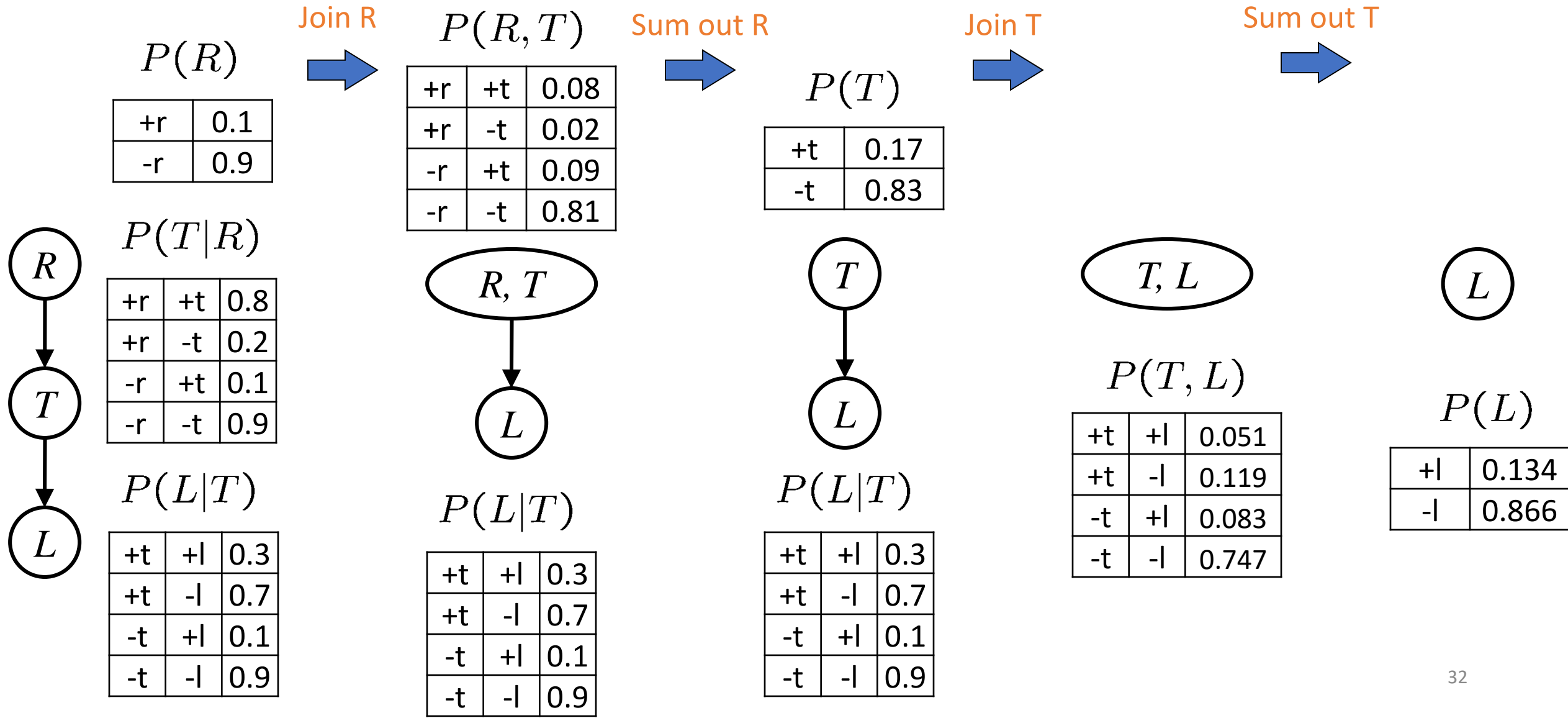
Evidence

- If evidence, start with factors that select that evidence
  - No evidence uses these initial factors:

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

- Computing $P(L| + r)$, the initial factors become:

$P(+r)$

+r	0.1

$P(T| + r)$

+r	+t	0.8
+r	-t	0.2

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

- We eliminate all vars other than query + evidence
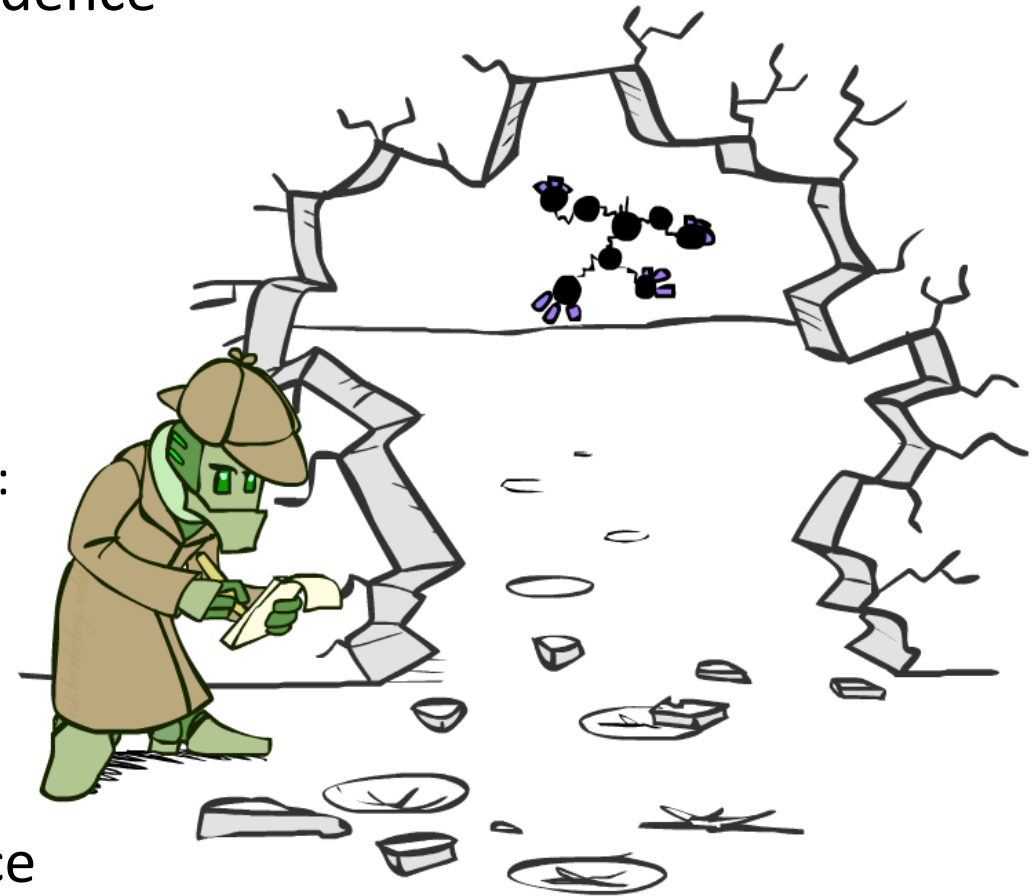
Evidence II

- Result will be a selected joint of query and evidence
  - E.g. for P(L | +r), we would end up with:

$$P(+r, L)$$

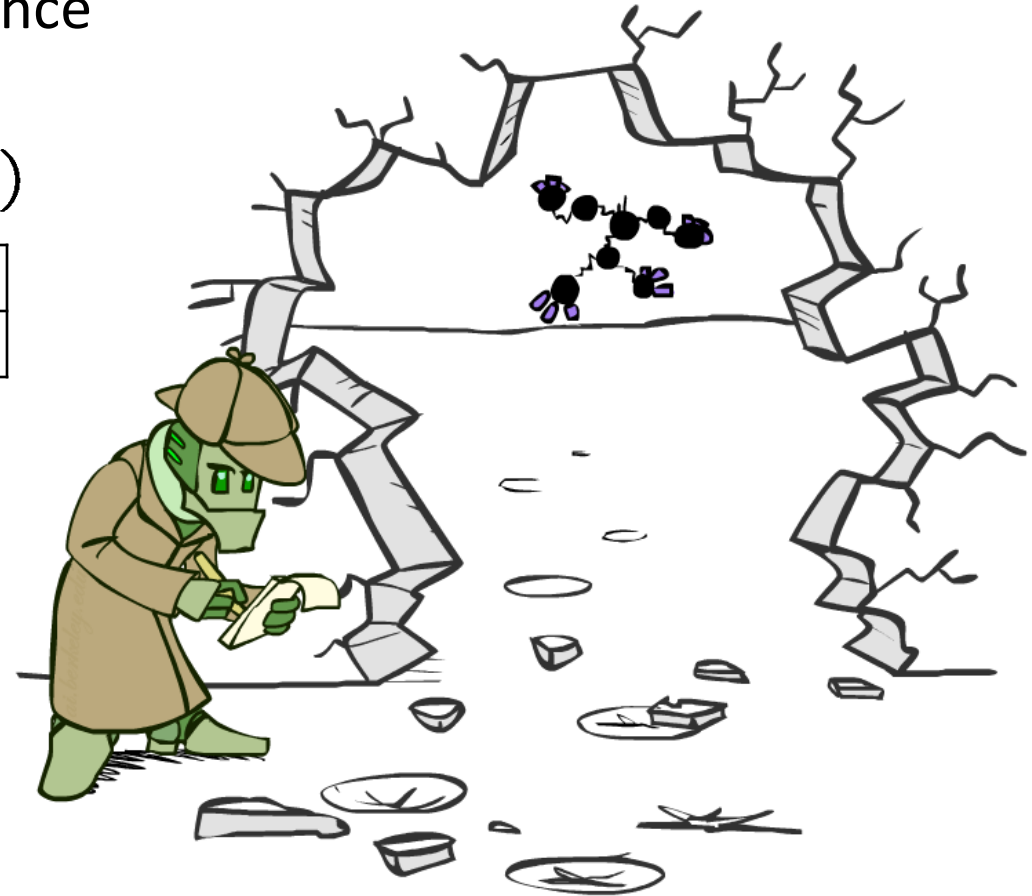|		
+r	+l	0.026
+r	-l	0.074

Normalize →

$$P(L| + r)$$

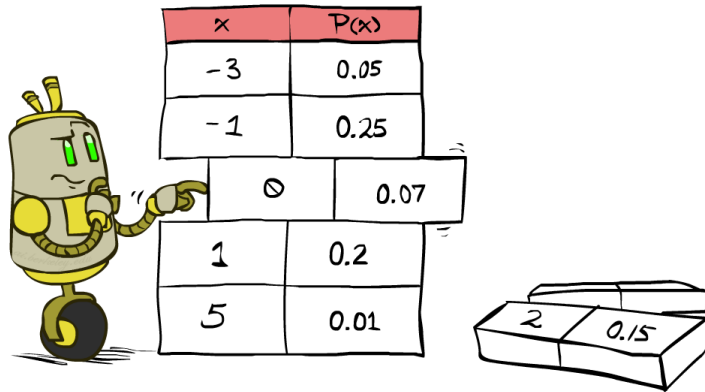|	
+l	0.26
-l	0.74

- To get our answer, just normalize this!
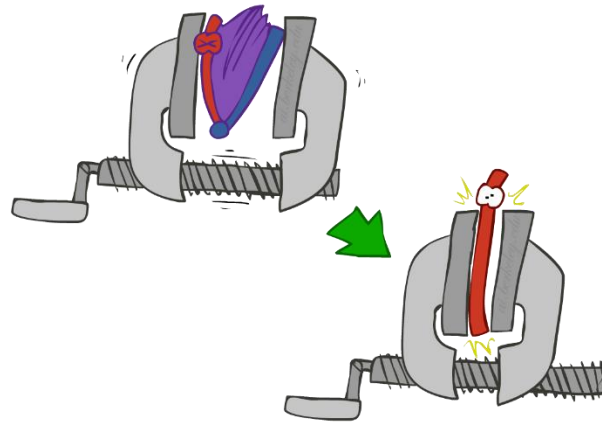
- That's it!

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

$\left. \right\}$ $X_1, X_2, \ldots X_n$

All variables

- We want:

$$P(Q|e_1 \ldots e_k)$$

* Works fine with multiple query variables, too

- Step 1: Select the entries consistent with the evidence

x	P(x)
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01
2	0.15

- Step 2: Sum out H to get joint of Query and evidence

- Compute joint

$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(\underbrace{Q, h_1 \ldots h_r, e_1 \ldots e_k}_{X_1, X_2, \ldots X_n})$$

- Sum out hidden variables

- Step 3: Normalize

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

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Variable Elimination

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$ $\Big\}$ $X_1, X_2, \ldots X_n$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$ $\Big\}$ *All variables*

- We want:

$$P(Q|e_1 \ldots e_k)$$

- **Step 1: Select the entries consistent with the evidence**

x	P(x)
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01
2	0.15

- **Step 2: Sum out H to get joint of Query and evidence**
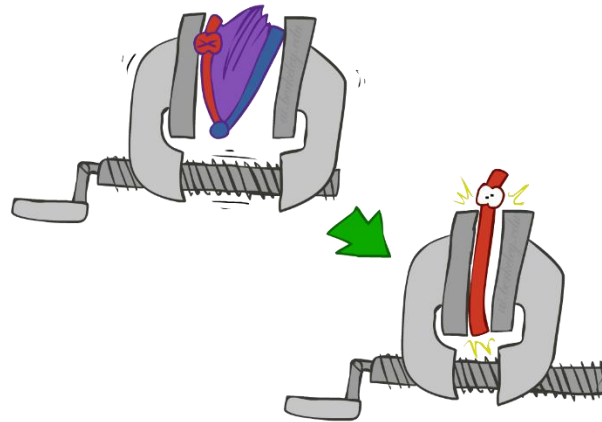
$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(\underbrace{Q, h_1 \ldots h_r, e_1 \ldots e_k}_{X_1, X_2, \ldots X_n})$$

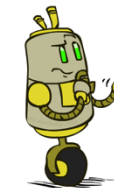- Interleave joining and summing out

- **Step 3: Normalize**

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

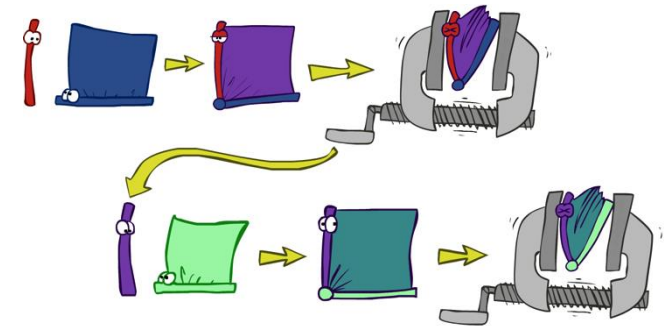$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

General Variable Elimination

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

- 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

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

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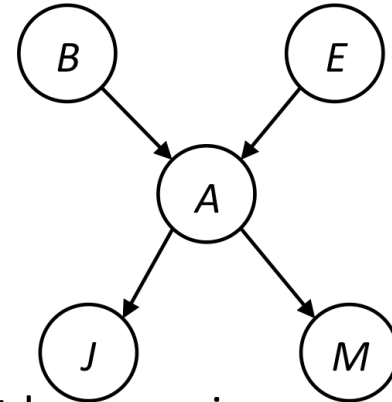
$$P(B|j,m) \propto P(B,j,m)$$

marginal can be obtained from joint by summing out

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

use Bayes' net joint distribution expression

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

use x*(y+z) = xy + xz

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

joining on a, and then summing out gives f_1

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

use x*(y+z) = xy + xz

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

joining on e, and then summing out gives f_2

$$= P(B)f_2(j,m|B)$$

All we are doing is exploiting uwy + uwz + uxy + uxz + vwy + vwz + vxy +vxz = (u+v)(w+x)(y+z) to improve computational efficiency!

Example (cont'd)

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

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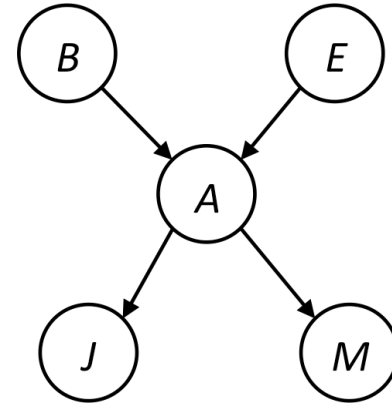
Choose A

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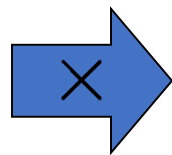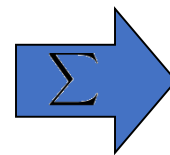
$P(B)$	$P(E)$	$P(j,m|B,E)$

Example (cont'd)

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

Choose E

$P(E)$

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

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

Finish with B

$P(B)$

$P(j,m|B)$ $\times$ $P(j,m,B)$ Normalize $P(B|j,m)$
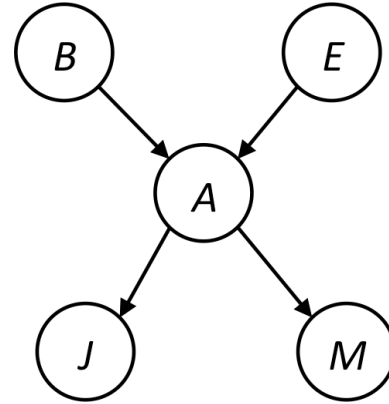
Another Variable Elimination Example

Query: $P(X_3|Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$

Start by inserting evidence, which gives the following initial factors:

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

Eliminate X_1, this introduces the factor $f_1(y_1|Z) = \sum_{x_1} P(x_1|Z)P(y_1|x_1)$, and we are left with:

$P(Z), P(X_2|Z), P(X_3|Z), P(y_2|X_2), P(y_3|X_3), f_1(y_1|Z)$

Eliminate X_2, this introduces the factor $f_2(y_2|Z) = \sum_{x_2} P(x_2|Z)P(y_2|x_2)$, and we are left with:
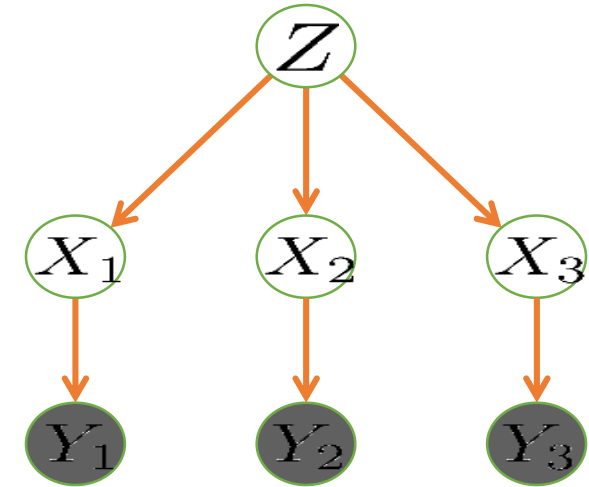
$P(Z), P(X_3|Z), P(y_3|X_3), f_1(y_1|Z), f_2(y_2|Z)$

Eliminate Z, this introduces the factor $f_3(y_1, y_2, X_3) = \sum_z P(z)P(X_3|z)f_1(y_1|Z)f_2(y_2|Z)$, and we are left with:

$P(y_3|X_3), f_3(y_1, y_2, X_3)$

No hidden variables left. Join the remaining factors to get:

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

Normalizing over X_3 gives $P(X_3|y_1, y_2, y_3) = f_4(y_1, y_2, y_3, X_3)/\sum_{x_3} f_4(y_1, y_2, y_3, x_3)$

Computational complexity critically depends on the largest factor being generated in this process. Size of factor = number of entries in table. In example above (assuming binary) all factors generated are of size 2 --- as they all only have one variable (Z, Z, and X₃ respectively).

Variable Elimination Ordering

- For the query $P(X_n | y_1, \ldots, y_n)$ work through the following two different orderings as done in previous slide: $Z, X_1, \ldots, X_{n-1}$ and $X_1, \ldots, X_{n-1}, Z$. What is the size of the maximum factor generated for each of the orderings?
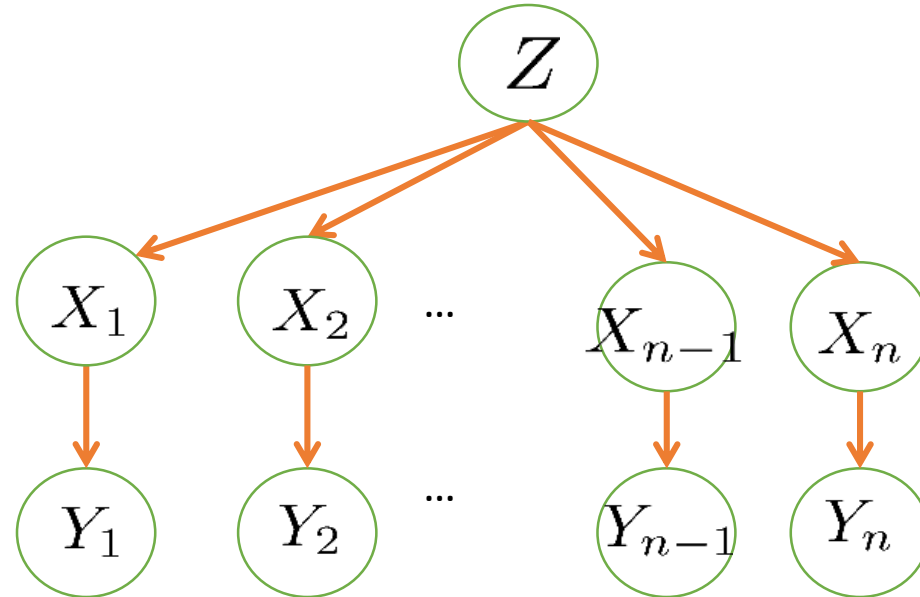
- Answer: 2^n versus 2 (assuming binary)

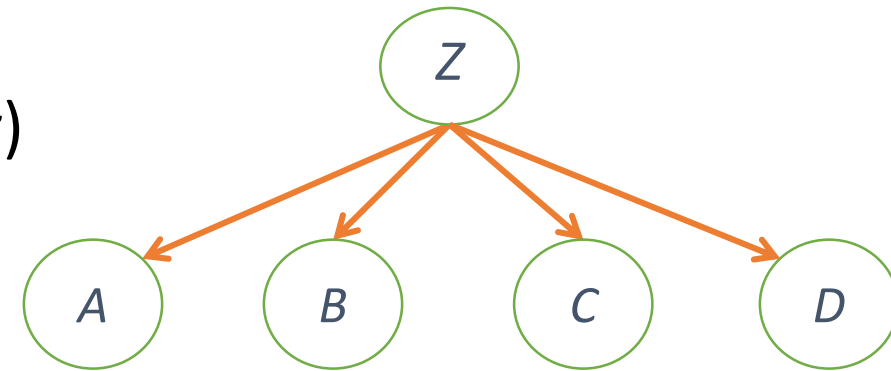- In general: the ordering can greatly affect efficiency

Detail of size 4

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

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

- 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!

Worst Case Complexity?

- CSP:

$$(x_1 \lor x_2 \lor \neg x_3) \land (\neg x_1 \lor x_3 \lor \neg x_4) \land (x_2 \lor \neg x_2 \lor x_4) \land (\neg x_3 \lor \neg x_4 \lor \neg x_5) \land (x_2 \lor x_5 \lor x_7) \land (x_4 \lor x_5 \lor x_6) \land (\neg x_5 \lor x_6 \lor \neg x_7) \land (\neg x_5 \lor \neg x_6 \lor x_7)$$

$$P(X_i = 0) = P(X_i = 1) = 0.5$$

$$Y_1 = X_1 \lor X_2 \lor \neg X_3$$

...

$$Y_8 = \neg X_5 \lor X_6 \lor X_7$$

$$Y_{1,2} = Y_1 \land Y_2$$
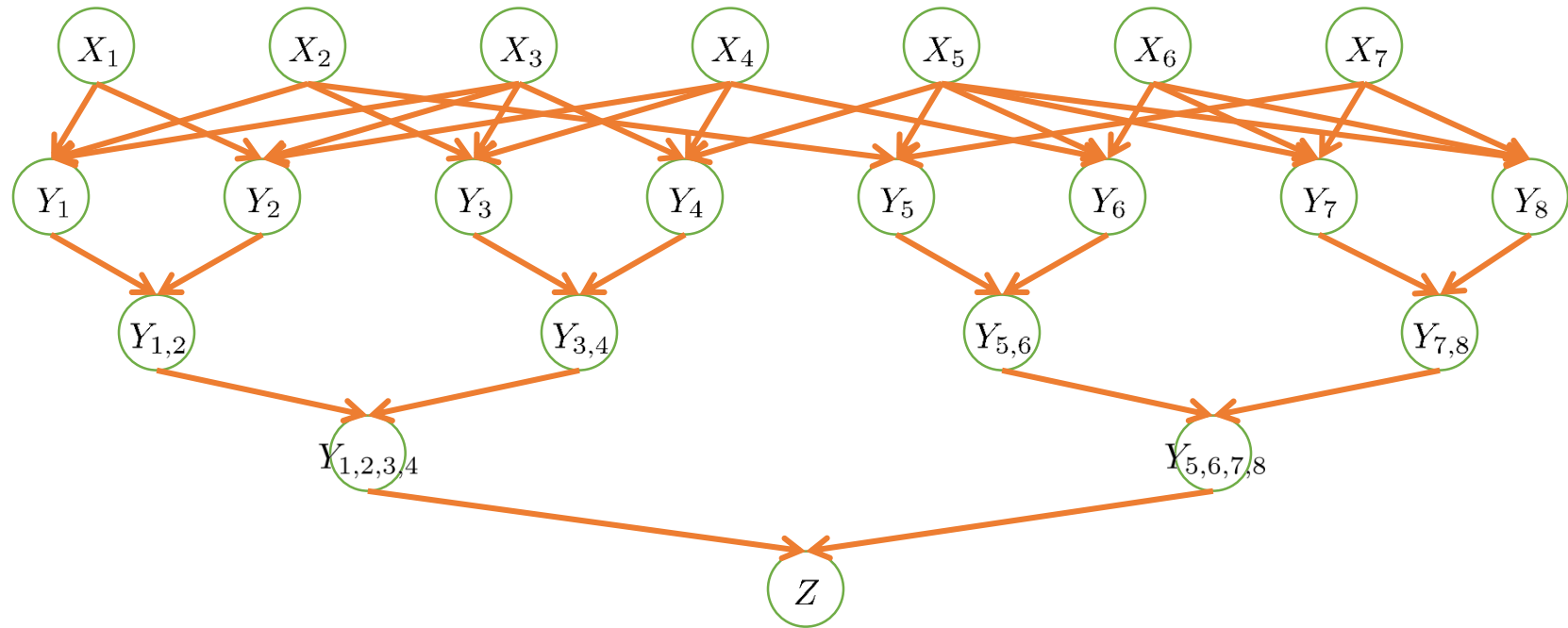
...

$$Y_{7,8} = Y_7 \land Y_8$$

$$Y_{1,2,3,4} = Y_{1,2} \land Y_{3,4}$$

$$Y_{5,6,7,8} = Y_{5,6} \land Y_{7,8}$$

$$Z = Y_{1,2,3,4} \land Y_{5,6,7,8}$$

- If we can answer P(z) equal to zero or not, we answered whether the 3-SAT problem has a solution
- Hence inference in Bayes' nets is NP-hard. No known efficient probabilistic inference in general

"Easy" Structures: Polytrees

- A polytree is a directed graph with no undirected cycles

- For poly-trees you can always find an ordering that is efficient
  - Try it!!

- Cut-set conditioning for Bayes' net inference

  - Choose set of variables such that if removed only a polytree remains
  - Exercise: Think about how the specifics would work out!

Variable Elimination: The basic ideas

- Move summations inwards as far as possible
  - $P(B \mid j, m) = \alpha \sum_e \sum_a P(B)\, P(e)\, P(a|B,e)\, P(j|a)\, P(m|a)$

 $= \alpha\, P(B) \sum_e P(e) \sum_a P(a|B,e)\, P(j|a)\, P(m|a)$
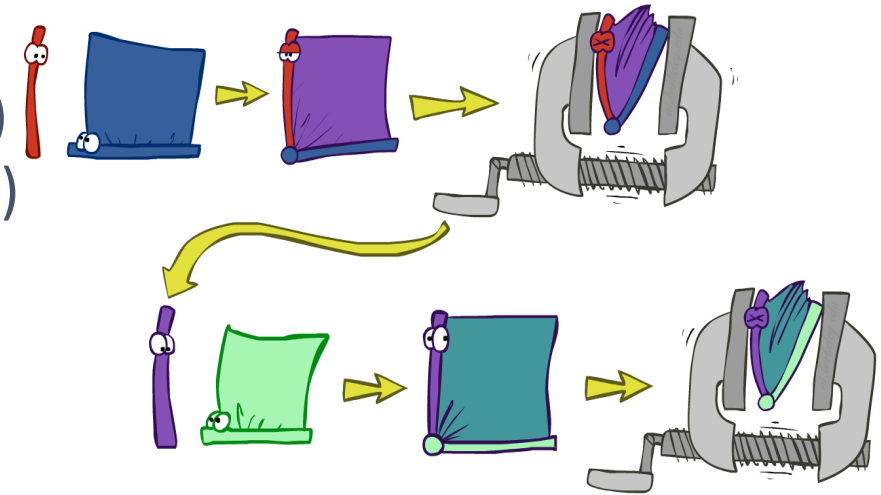
- Do the calculation from the inside out
  - I.e., sum over *a* first, then sum over *e*
  - Problem: $P(a|B,e)$ isn't a single number, it's a bunch of different numbers depending on the values of *B* and *e*
  - Solution: use arrays of numbers (of various dimensions) with appropriate operations on them; 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: $P(A) \times P(J|A) = P(A,J)$
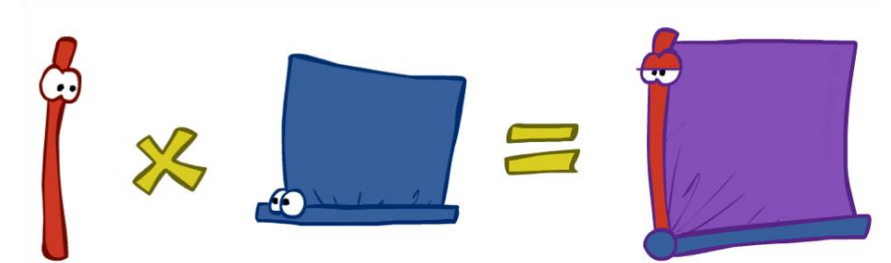
$P(A)$

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
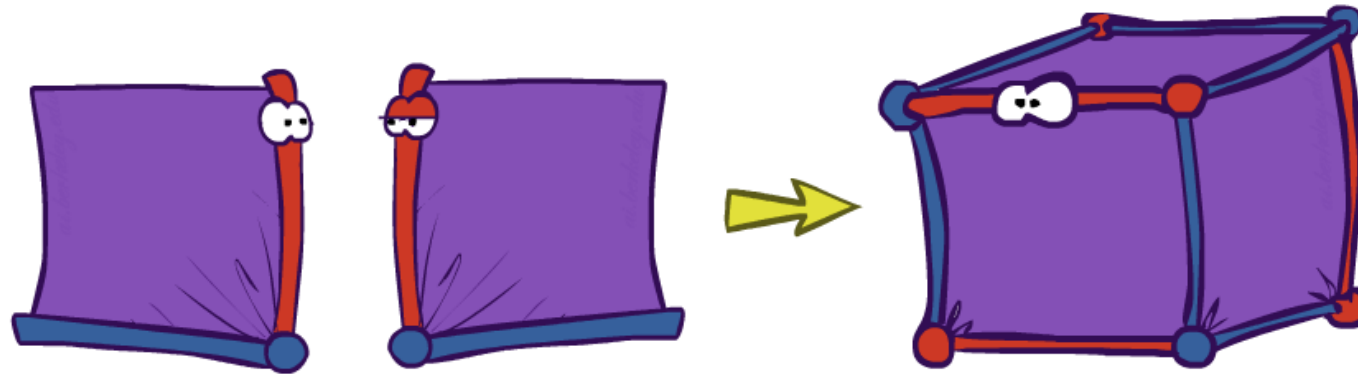
- Example: $P(J|A)$ x $P(M|A)$ = $P(J,M|A)$

$P(J/A)$

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

=

$P(J,M/A)$

J \ M	true	false
true		18

A=false

J \ M	true	false
true		
false		.0003

A=true

Example: Making larger factors 2
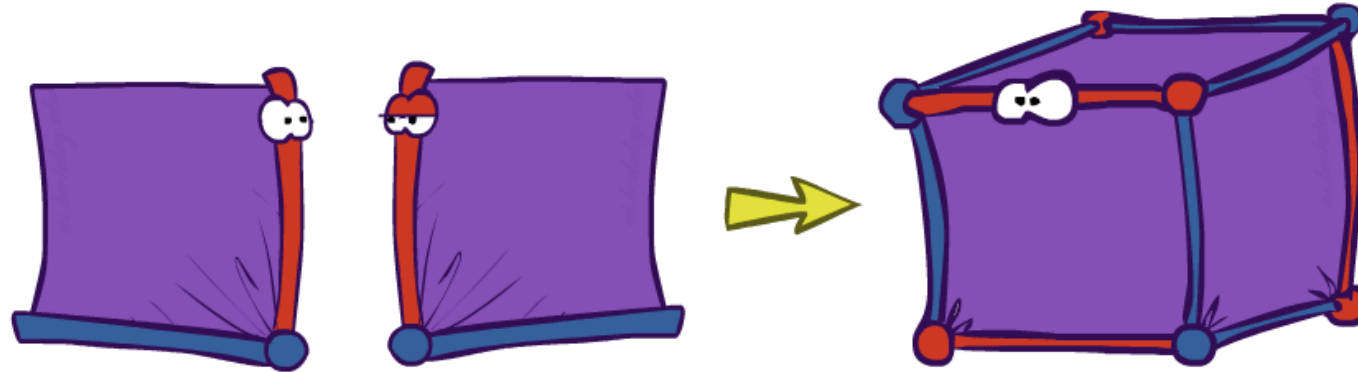
- Example: $f_1(U,V)$ x $f_2(V,W)$ x $f_3(W,X)$ = $f_4(U,V,W,X)$

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

- i.e., 300 numbers blows up to 10,000 numbers!

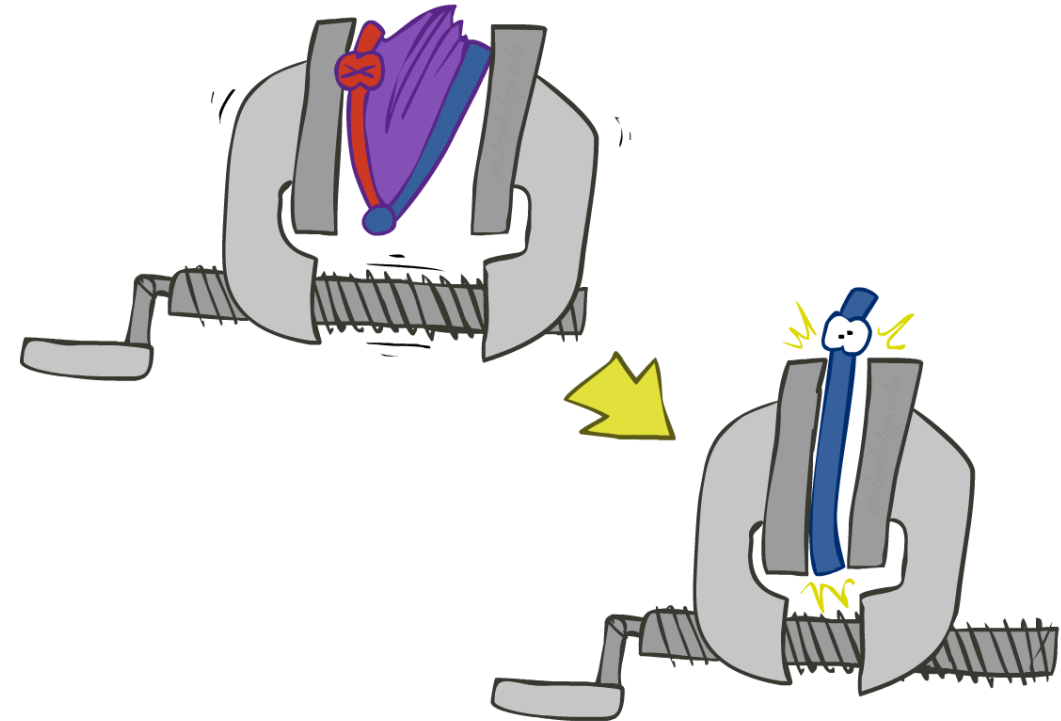- Factor blowup can make VE very expensive

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

$P(A,J)$

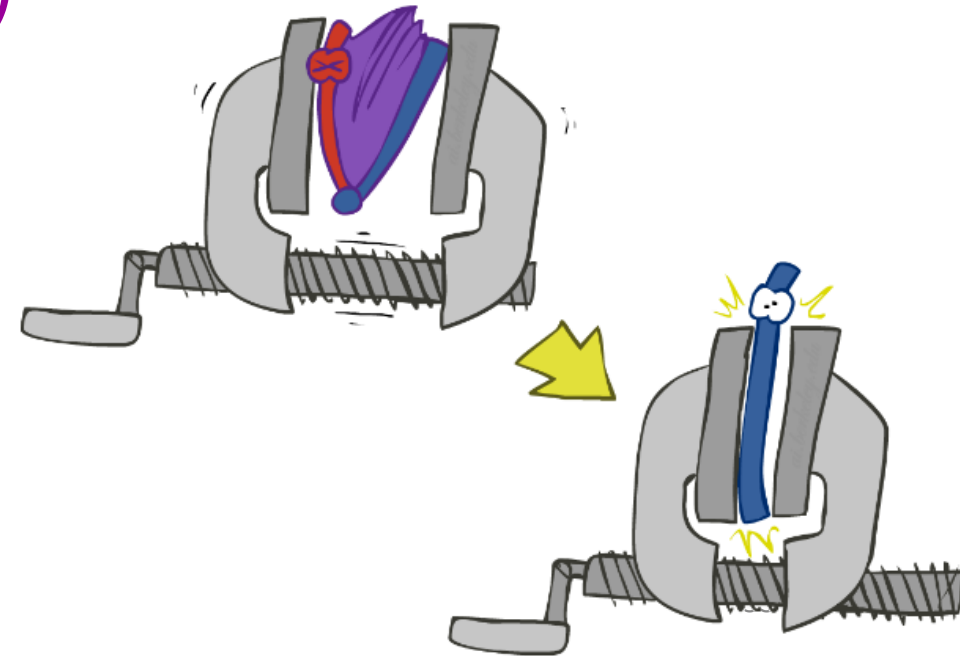| A \ J	true	false
true	0.09	0.01
false	0.045	0.855

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

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

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

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

Summary

- Bayes

Shuai Li

https://shuaili8.github.io

Questions?