Announcements

- Midterm: Wednesday 7pm-9pm
  - See midterm prep page (posted on Piazza, inst.eecs page)
  - Four rooms; your room determined by ***last two digits of your SID***:
    - 00-32: Dwinelle 155
    - 33-45: Genetics and Plant Biology 100
    - 46-62: Hearst Annex A1
    - 63-99: Pimentel 1
  - Discussions this week *by topic*
- Survey: complete it before midterm; 80% participation = +1pt

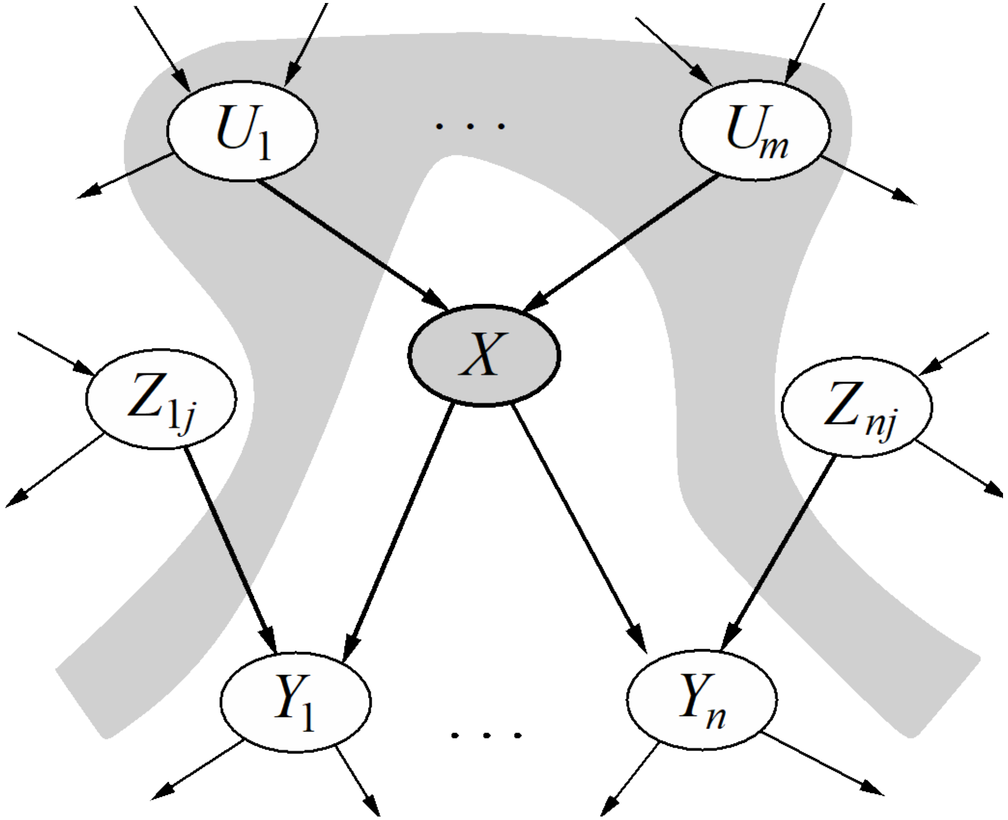
Bayes net global semantics

- Bayes nets encode joint distributions as product of conditional distributions on each variable:

$$P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$$

Conditional independence semantics

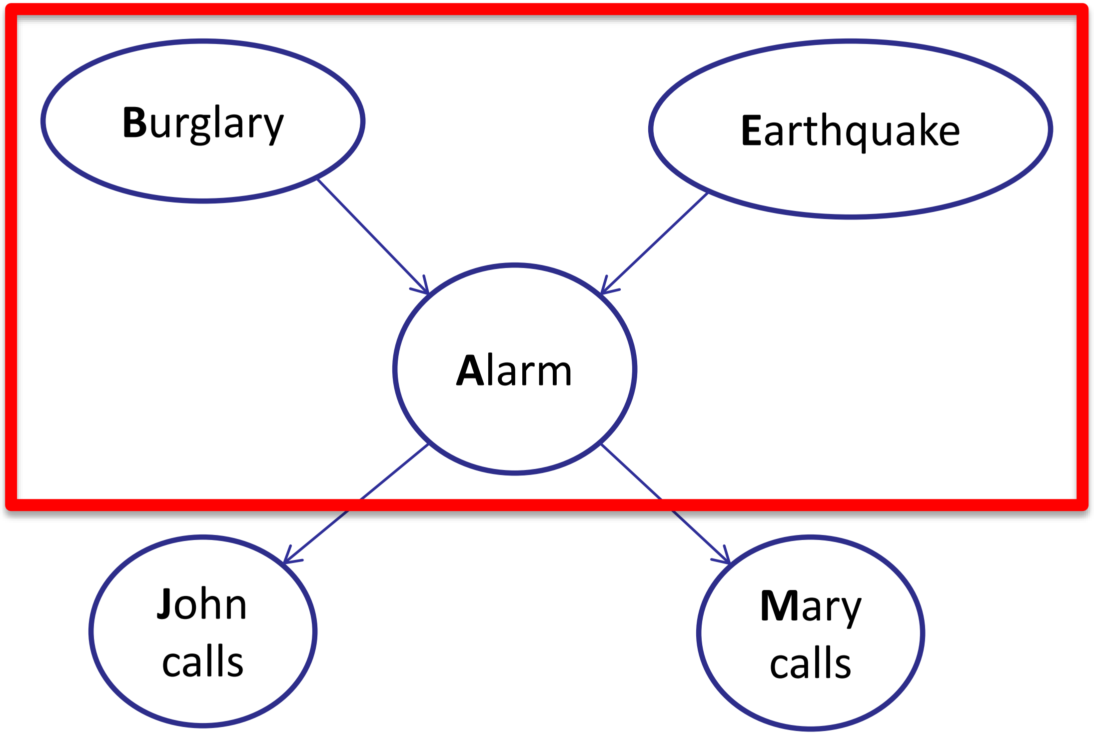
- ***Every variable is conditionally independent of its non-descendants given its parents***
- Conditional independence semantics <=> global semantics



Example

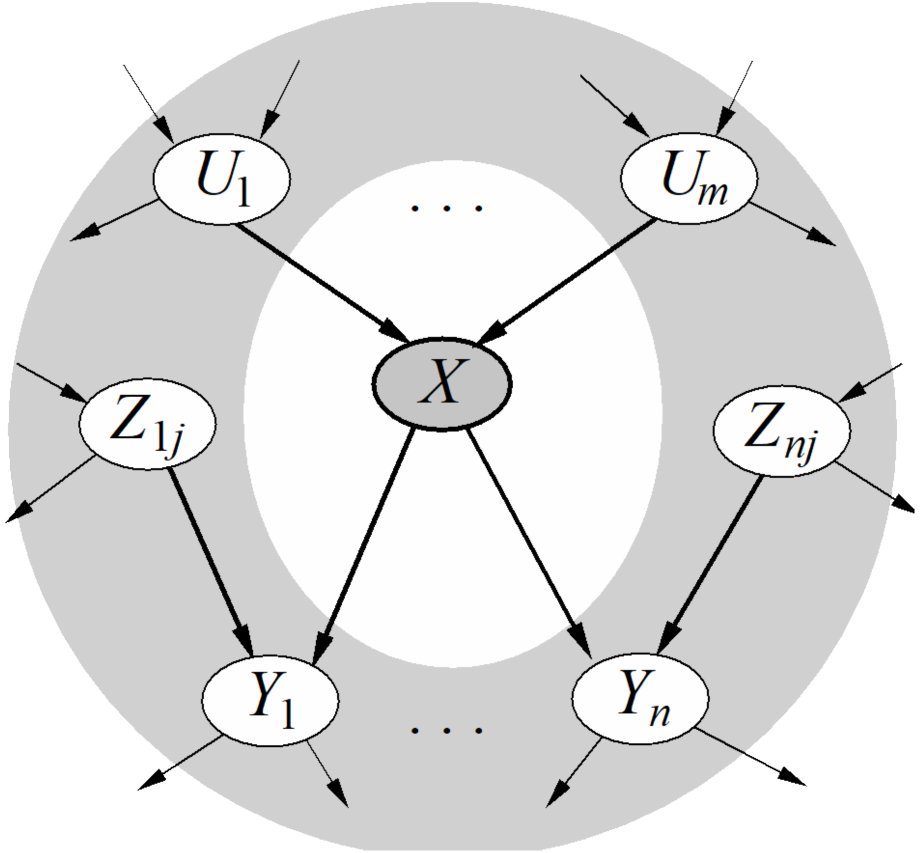
- JohnCalls independent of Burglary given Alarm?
  - Yes
- JohnCalls independent of MaryCalls given Alarm?
  - Yes
- Burglary independent of Earthquake?
  - Yes
- Burglary independent of Earthquake given Alarm?
  - NO!
  - Given that the alarm has sounded, both burglary and earthquake become more likely
  - But if we then learn that a burglary has happened, the alarm is ***explained away*** and the probability of earthquake drops back

V-structure



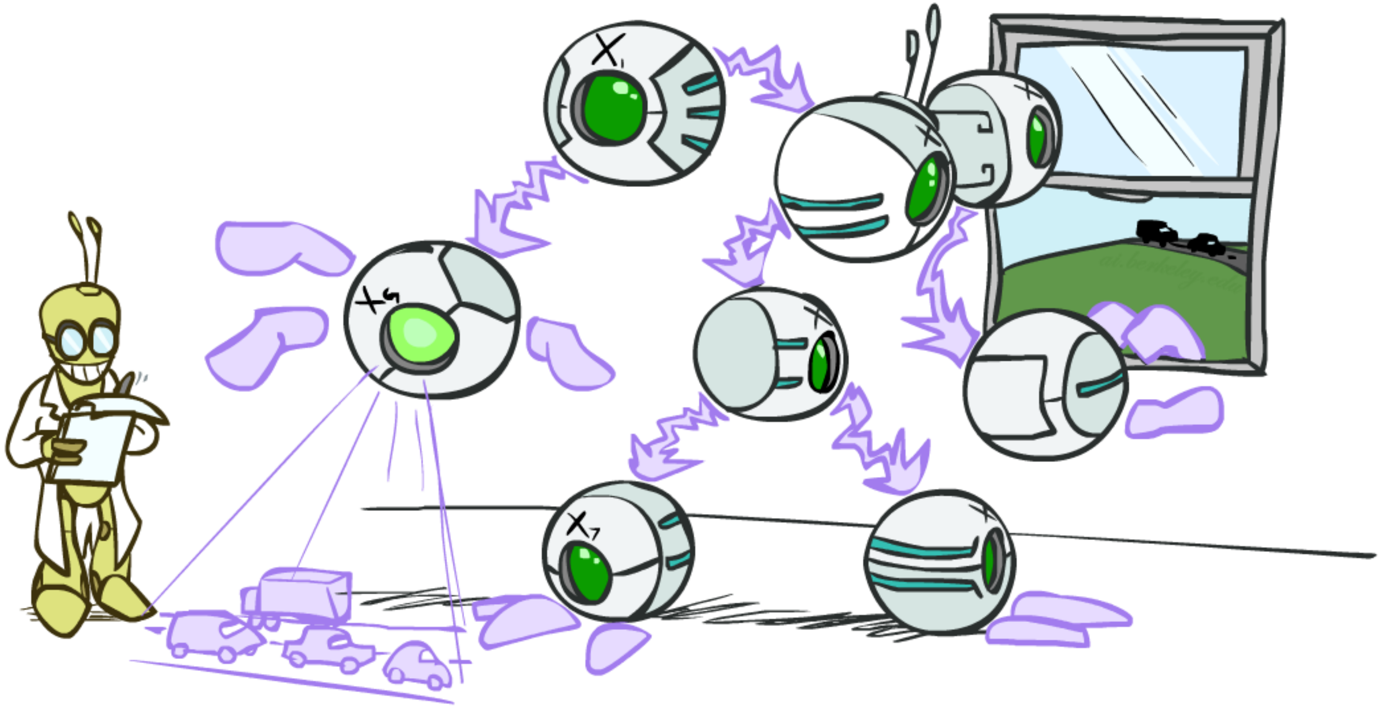
Markov blanket

- A variable's Markov blanket consists of parents, children, children's other parents
- ***Every variable is conditionally independent of all other variables given its Markov blanket***



CS 188: Artificial Intelligence

Bayes Nets: Exact Inference



Instructor: Sergey Levine and Stuart Russell--- University of California, Berkeley

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

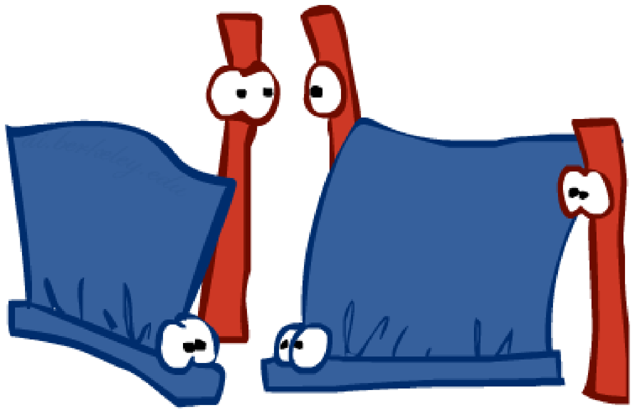
Part III: Approximate Inference

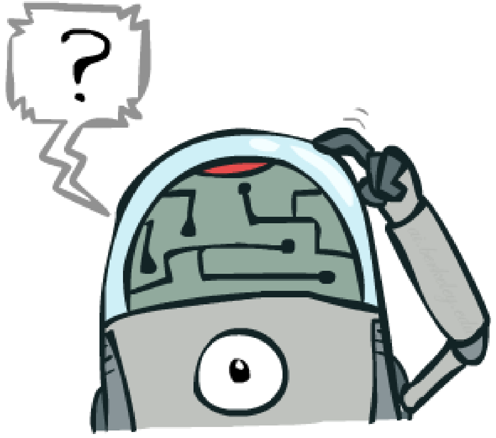
Later: Learning Bayes nets from data

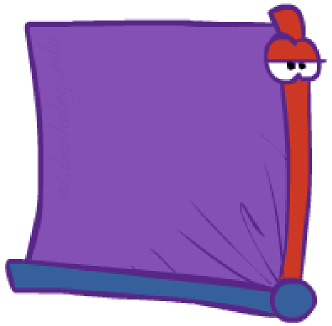
Inference

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

- Examples:
  - Posterior marginal probability
    - $P(Q|e_1,..,e_k)$
    - E.g., what disease might I have?
  - Most likely explanation:
    - $\text{argmax}_{q,r,s}\ P(Q=q,R=r,S=s|e_1,..,e_k)$
    - E.g., what did he say?

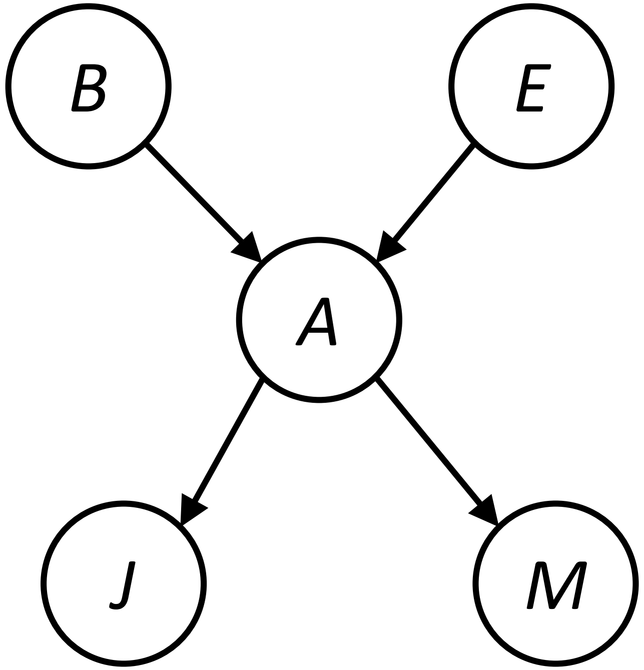






Inference by Enumeration in Bayes Net

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



Can we do better?

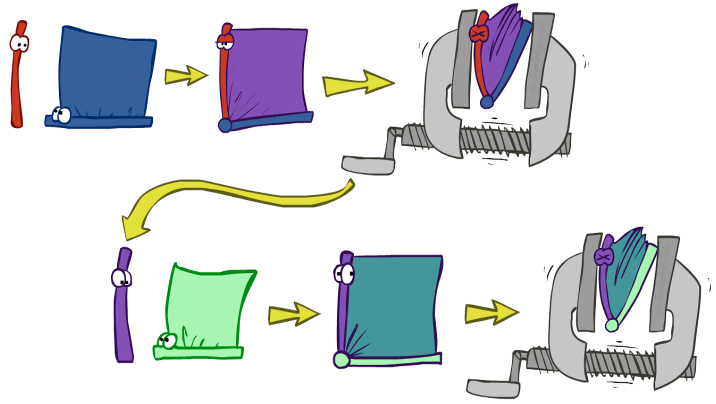
- Consider **uwy + uwz + uxy + uxz + vwy + vwz + vxy +vxz**
  - 16 multiplies, 7 adds
  - Lots of repeated subexpressions!
- Rewrite as **(u+v)(w+x)(y+z)**
  - 2 multiplies, 3 adds
- $\sum_{e,a} P(B)\ \mathrm{P}(\mathrm{e})\ \mathrm{P}(\mathrm{a}|B,e)\ P(j|a)\ P(m|a)$
- $= P(B)P(e)P(\mathrm{a}|B,e)P(j|a)P(m|a) + P(B)P(\neg e)P(\mathrm{a}|B,\neg e)P(j|a)P(m|a)$

  $+ P(B)P(e)P(\neg \mathrm{a}|B,e)P(j|\neg a)P(m|\neg a) + P(B)P(\neg e)P(\neg \mathrm{a}|B,\neg e)P(j|\neg a)P(m|\neg a)$

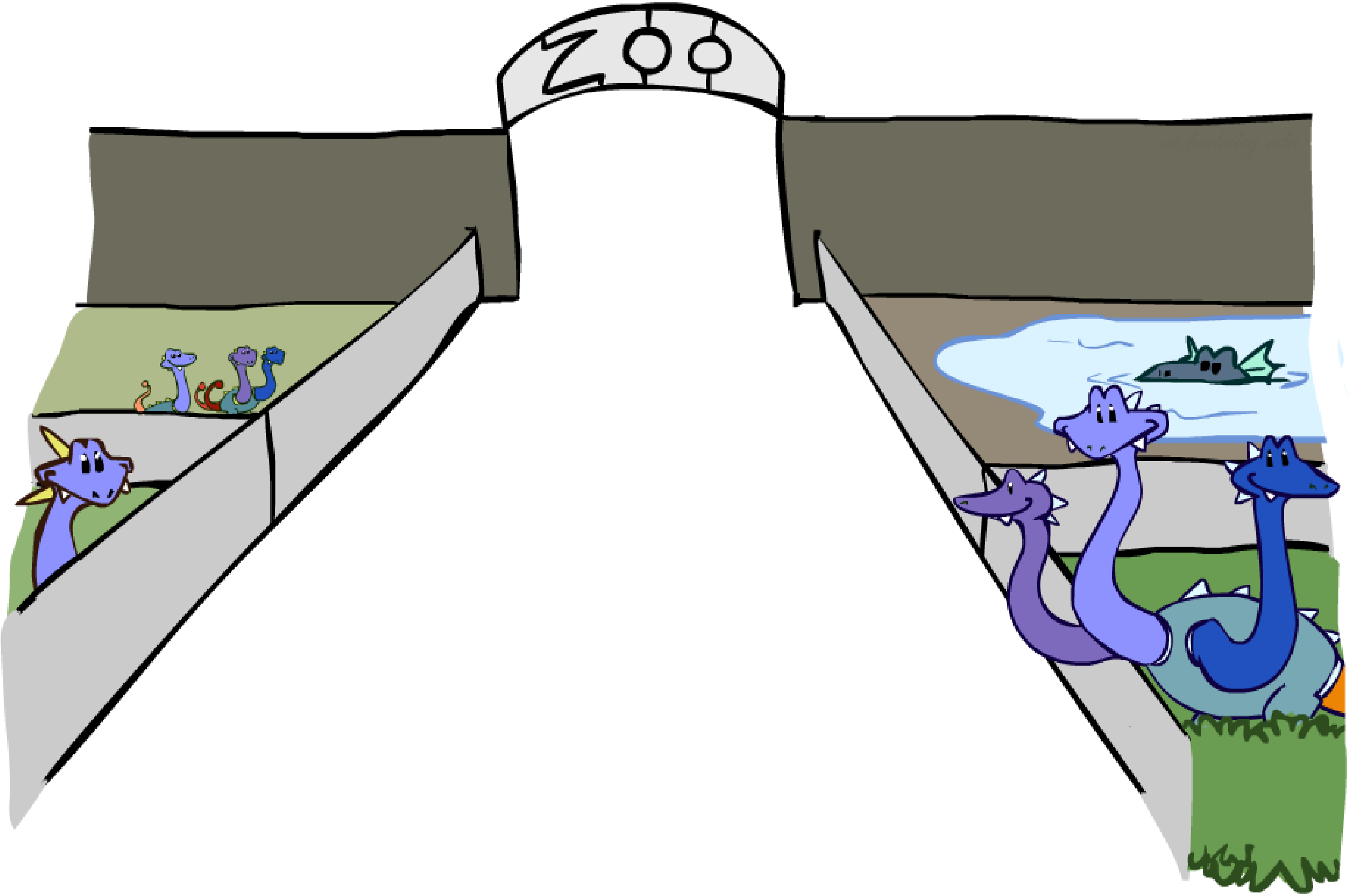
  Lots of repeated subexpressions!

Variable elimination: The basic ideas

- Move summations inwards as far as possible
  - $P(B \mid j, m) = \alpha \sum_{e,a} P(B)\, P(e)\, P(a|B,e)\, P(j|a)\, P(m|a)$
  - $= \alpha\, P(B) \sum_{e} P(e) \sum_{a} P(a|B,e)\, P(j|a)\, P(m|a)$
- 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***



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

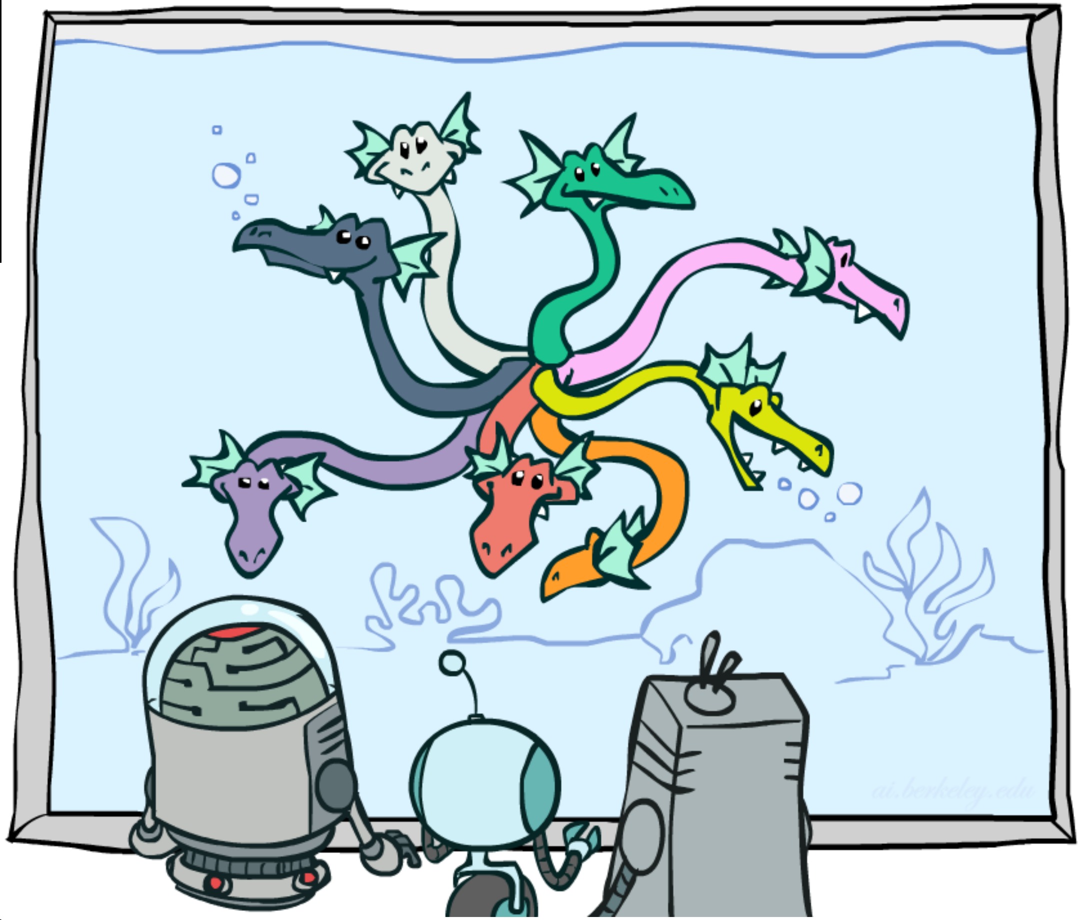
P(A,J)

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)

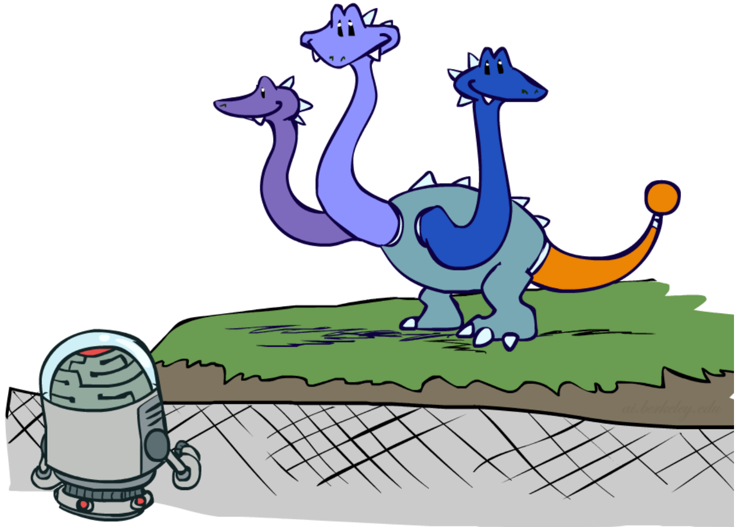
A \ J	true	false
true	0.09	0.01



Number of variables (capitals) = dimensionality of the table

Factor Zoo II

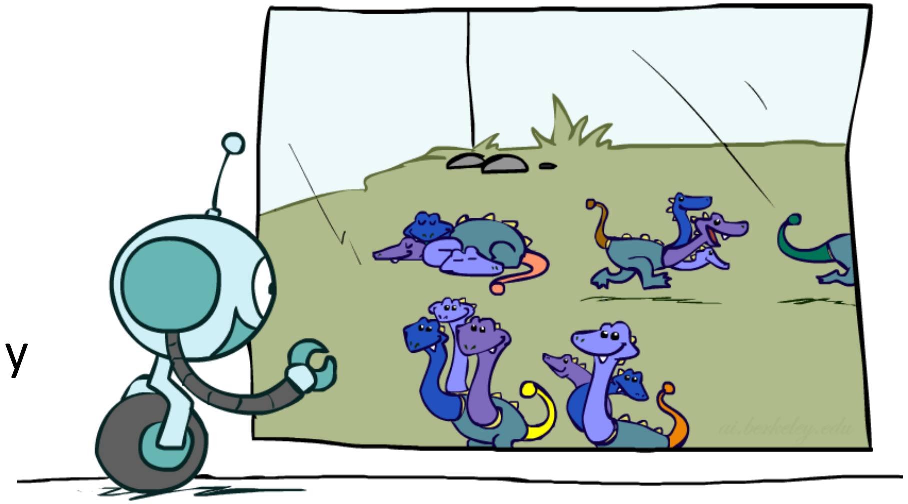
- Single conditional: P(Y | x)
  - Entries P(y | x) for fixed x, all y
  - Sums to 1



$P(J|a)$

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|



$P(J|A)$

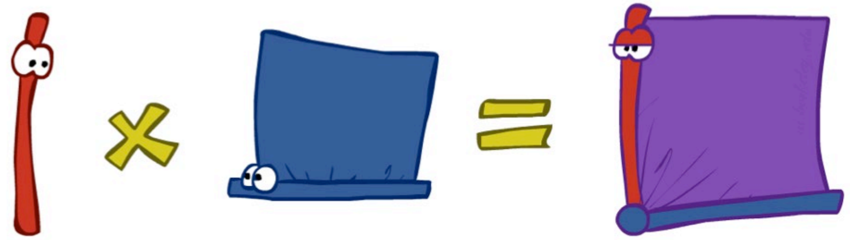
A \ J	true	false
true	0.9	0.1
false	0.05	0.95

} - $P(J|a)$

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

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

$P(A)$

true	0.1
false	0.9

x

$P(J|A)$

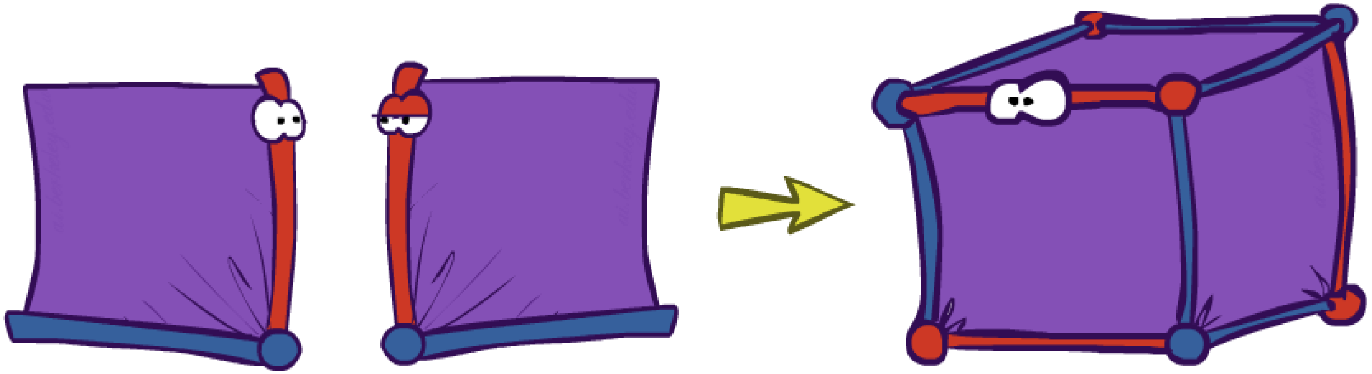
A \ J	true	false
true	0.9	0.1
false	0.05	0.95

=

$P(A,J)$

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Example: Making larger factors



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

P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

x

P(A,M)

A \ M	true	false
true	0.07	0.03
false	0.009	0.891

=

P(A,J,M)

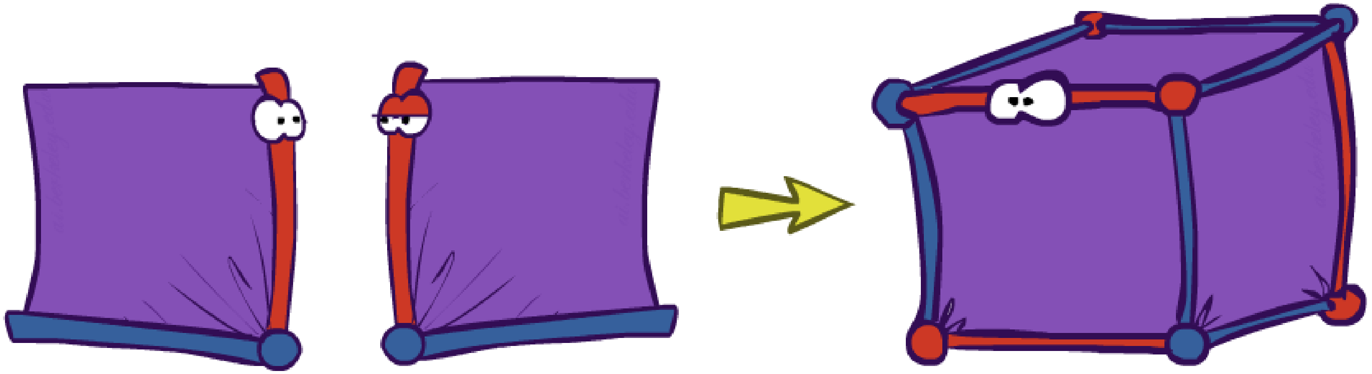
A=false

J \ M	true	false
true		18
false		

A=true

J \ M	true	false
true		
false		.0003

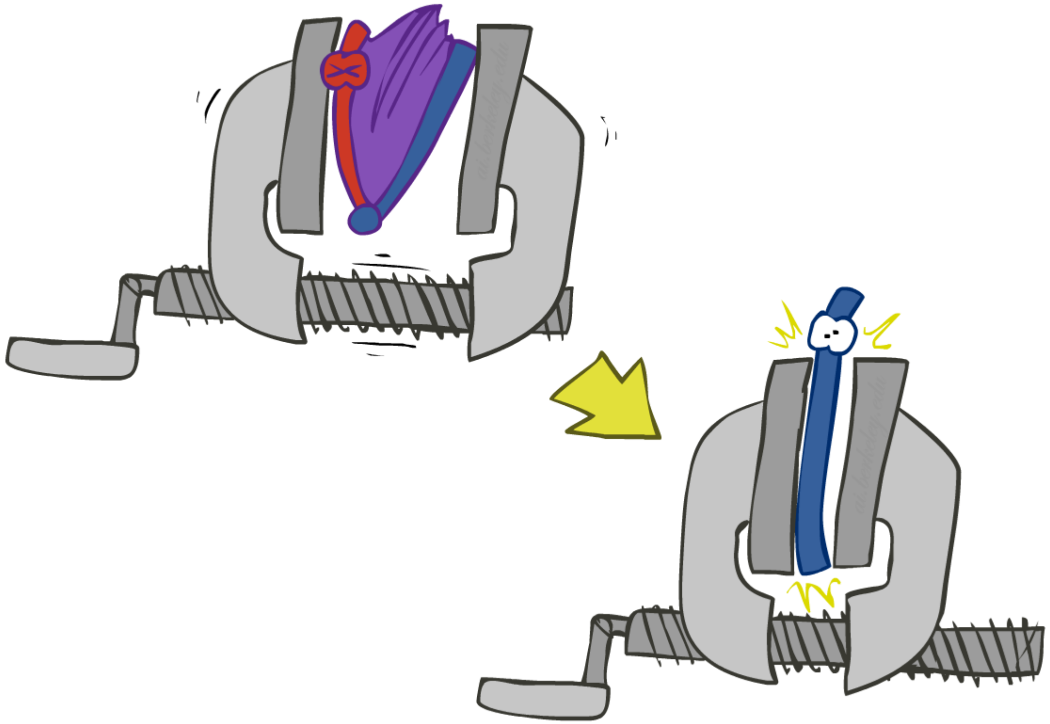
Example: Making larger factors



- Example: *P*(*U*,*V*) x *P*(*V*,*W*) x *P*(*W*,*X*) = *P*(*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

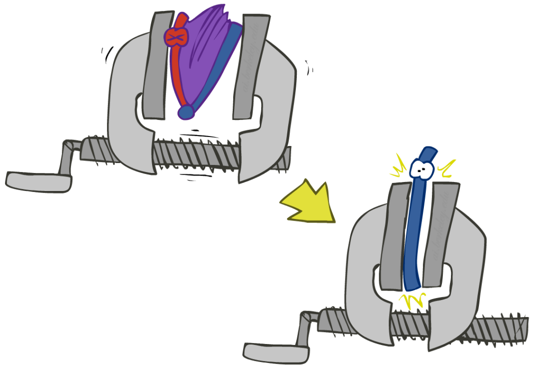
Sum out J →

$P(A)$

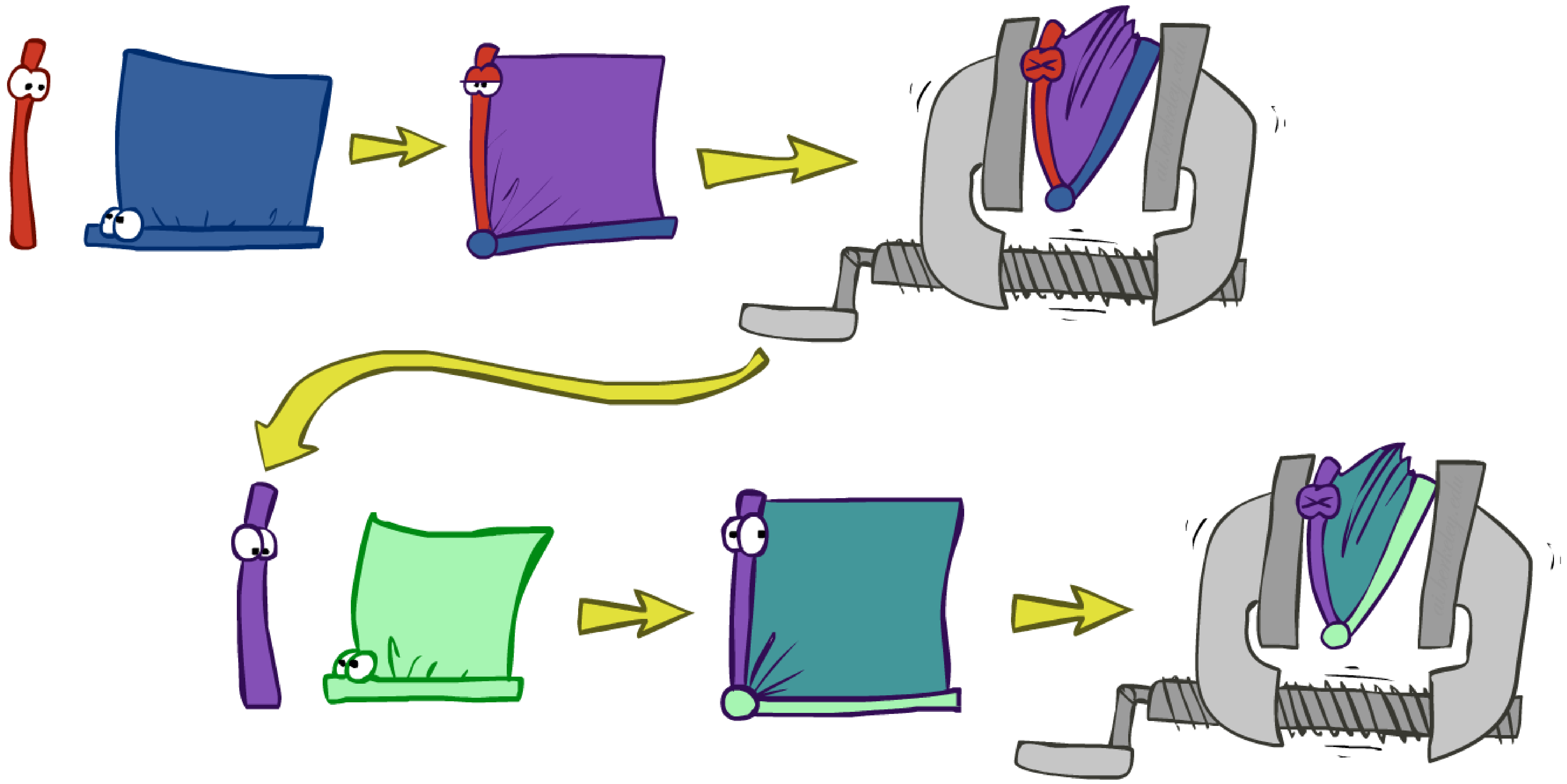
true	0.1
false	0.9

Summing out from a product of factors

- Project the factors each way first, then sum the products
- Example: $\sum_a P(a|B,e) \times P(j|a) \times P(m|a)$
- $= P(a|B,e) \times P(j|a) \times P(m|a) +$
- $P(\neg a|B,e) \times P(j|\neg a) \times P(m|\neg a)$

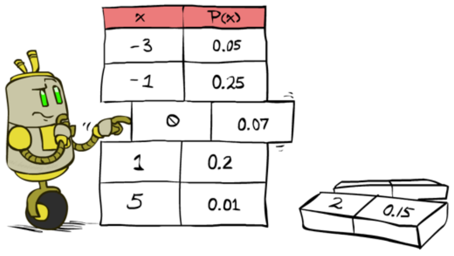


Variable Elimination

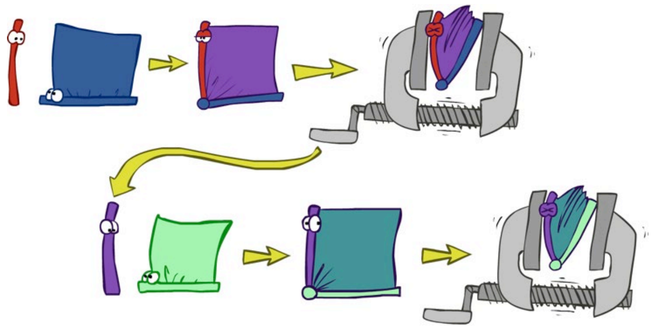


Variable Elimination

- 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**

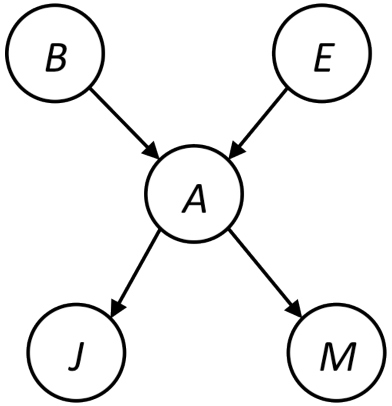
        factors ← SUM-OUT(*var*,*factors*)

return NORMALIZE(POINTWISE-PRODUCT(factors))

Example

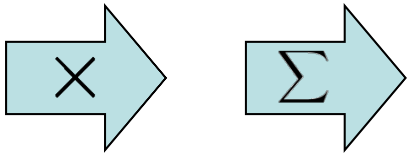
Query $P(B \mid j,m)$

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



Choose A

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

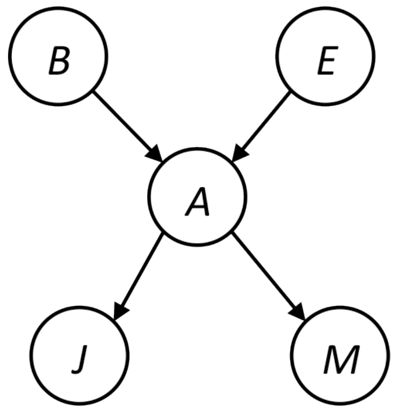


$P(j,m \mid B,E)$

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

Example

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



Choose E

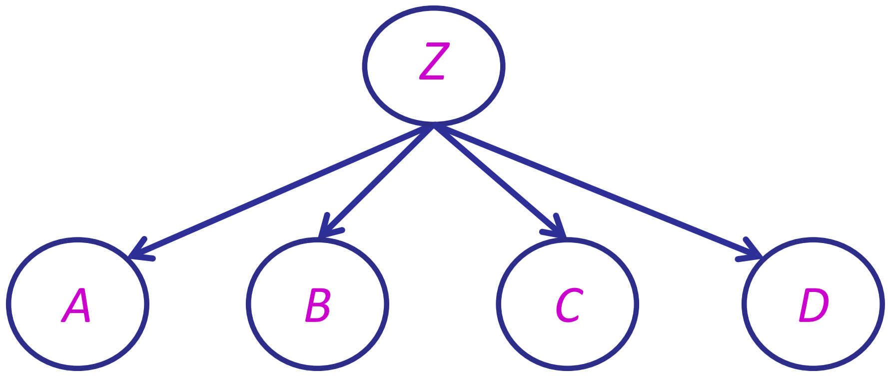
$P(E)$
$P(j,m|B,E)$ $\times$ → Σ → $P(j,m|B)$

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

Finish with B

$P(B)$
$P(j,m|B)$ $\times$ → $P(j,m,B)$ Normalize → $P(B \mid j,m)$

Order matters



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

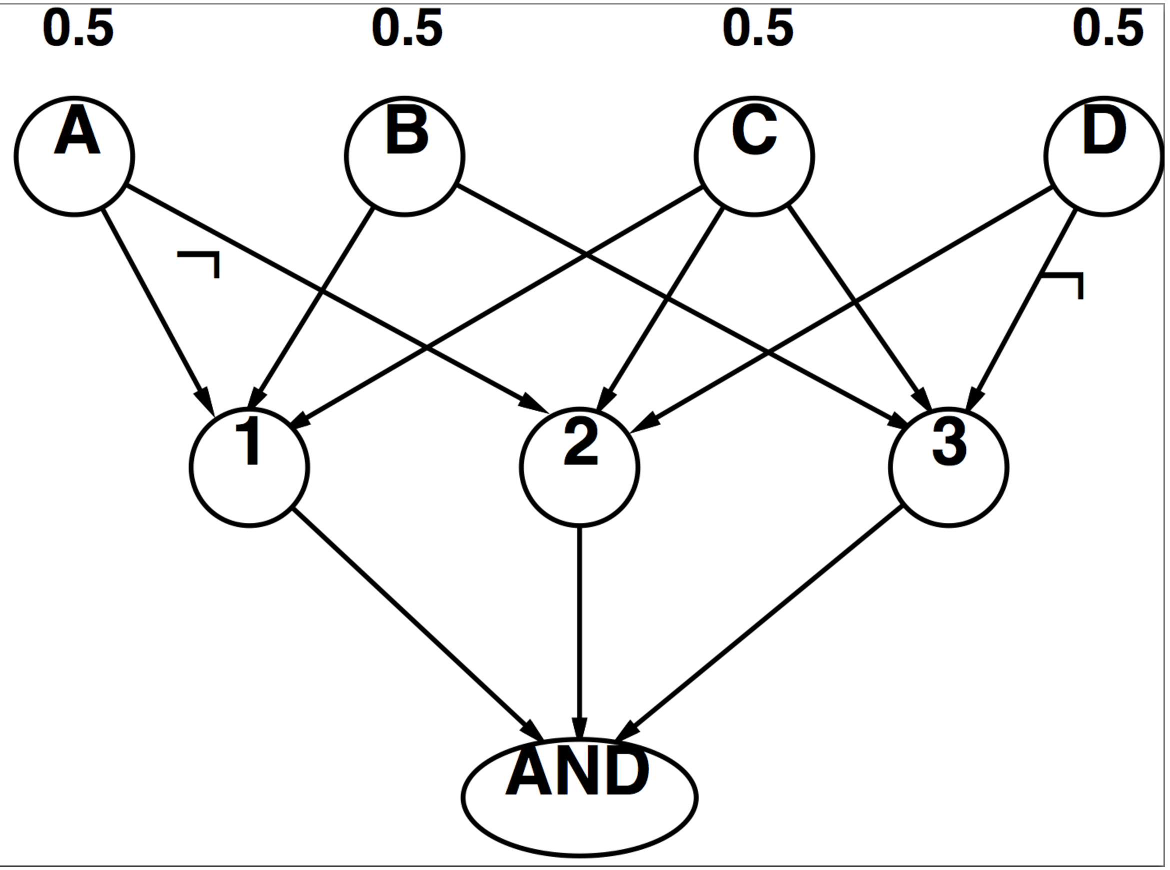
VE: Computational and Space Complexity

- 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!

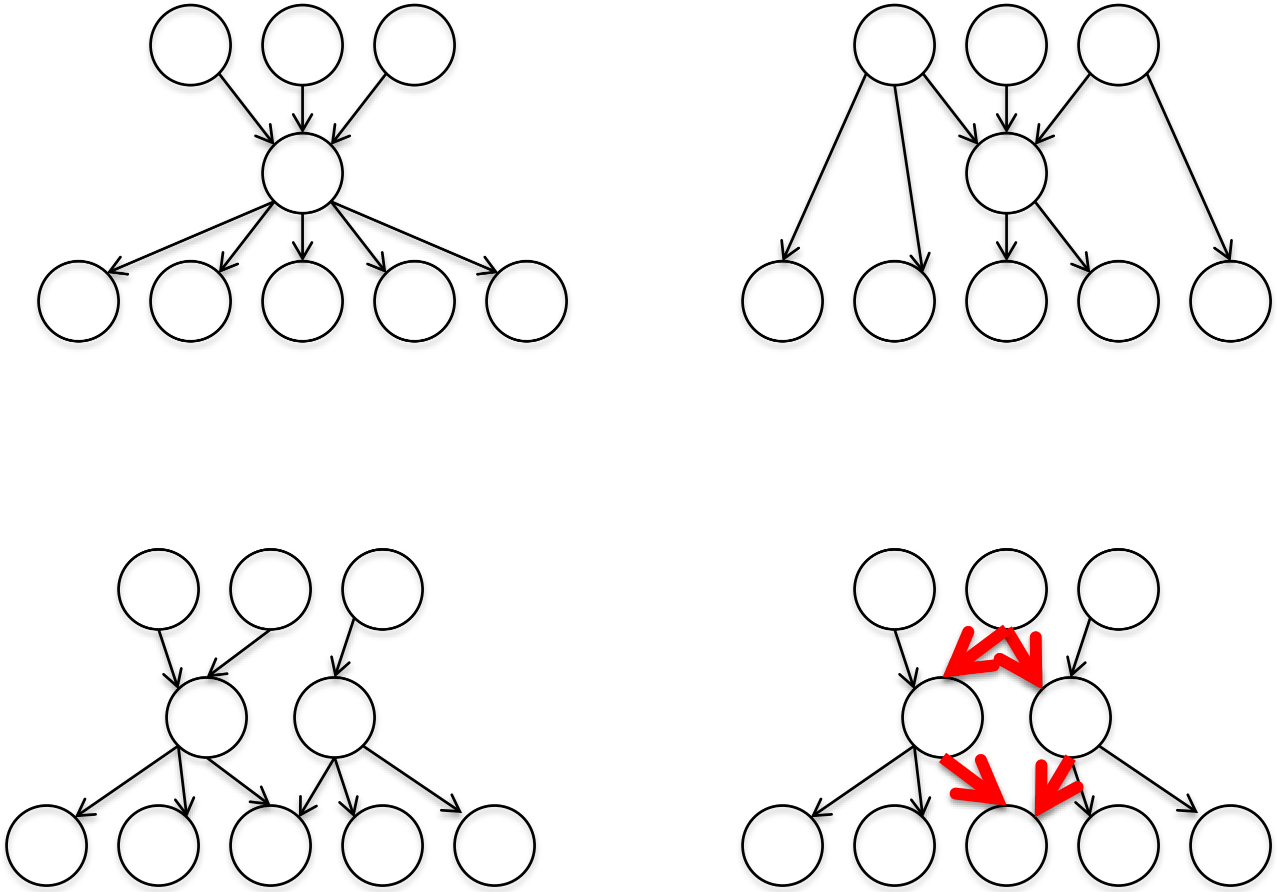
Worst Case Complexity? Reduction from SAT



- CNF clauses:
  1. A v B v C
  2. C v D v ¬A
  3. B v C v ¬D
- *P*(AND) > 0 iff clauses are satisfiable
  - => NP-hard
- P(AND) = S x 0.5^n where S is the number of satisfying assignments for clauses
  - => #P-hard

Polytrees

- A polytree is a directed graph with no undirected cycles
- For poly-trees the complexity of variable elimination is ***linear in the network size*** if you eliminate from the leave towards the roots
  - This is essentially the same theorem as for tree-structured CSPs



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

Later: Learning Bayes nets from data