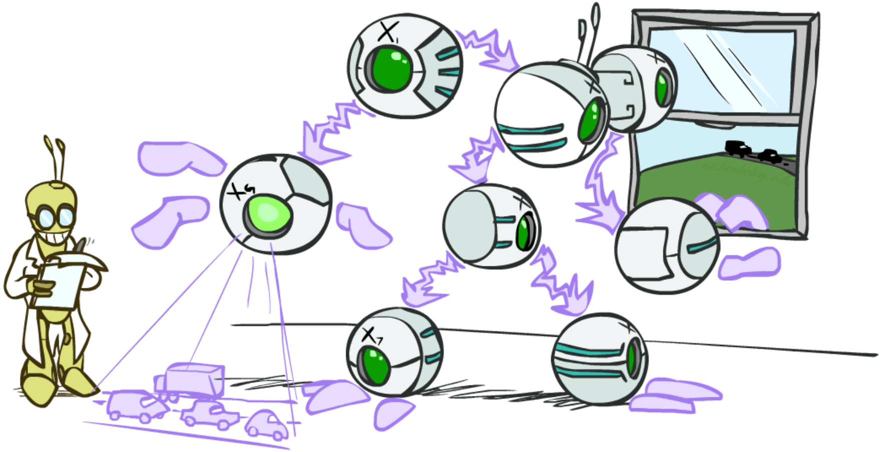
Announcement

- Programming 2 due 11:59pm today!

- Programming 3
  - Due: April 12, 11:59pm
- Homework 3
  - Due: April 14, 11:59pm

Bayes Nets: Exact Inference

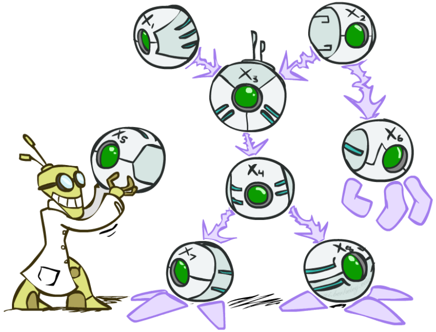


AIMA Chapter 14.4, PRML Chapter 8.4

[Adapted from slides by Dan Klein and Pieter Abbeel at UC Berkeley]

CPT

Bayes' Net Representation



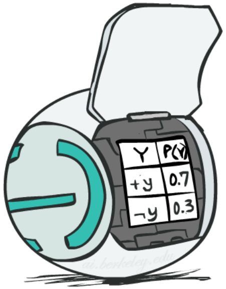
- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

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

- 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:

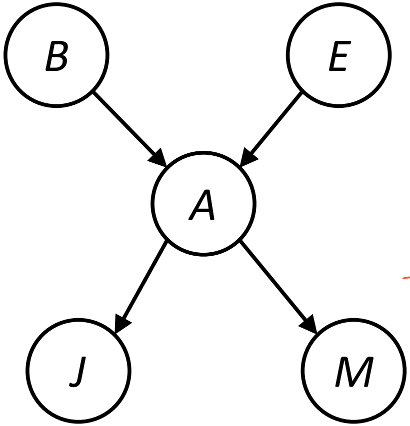
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

wasteful

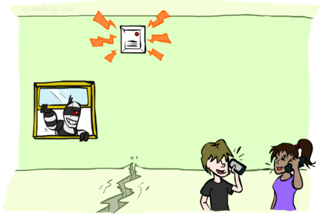


Example: Alarm Network

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

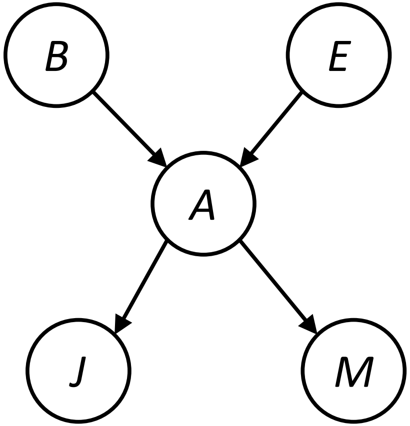
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

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

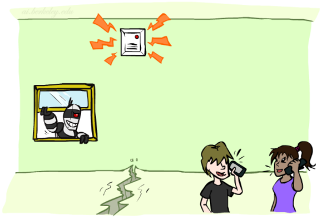
$$P(+b)P(-e)P(+a| + b, -e)P(-j| + a)P(+m| + a) =$$

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

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

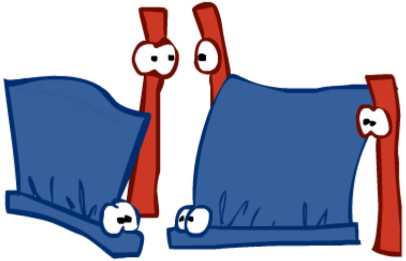
Probabilistic Inference

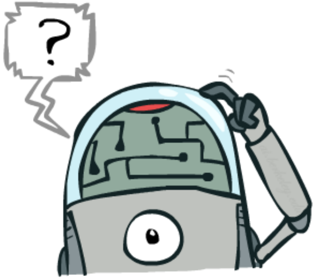
- Enumeration (exact, exponential complexity)
- Variable elimination (exact, worst-case exponential complexity, often better)
- Inference is NP-complete
- Sampling (approximate)

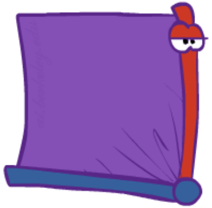
Inference

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

- Examples: Q : cared about
  - Posterior marginal probability
    - $P(Q|e_1,..,e_k)$ → $G_1 = e_1$ ...
    - E.g., what disease might I have?
  - Most likely explanation:
    - $\text{argmax}_q\ P(Q=q|e_1,..,e_k)$
    - E.g., what did he say?







Inference by Enumeration

- General case: value known
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query variable: Q
  - Hidden variables: $H_1 \dots H_r$ unknown.

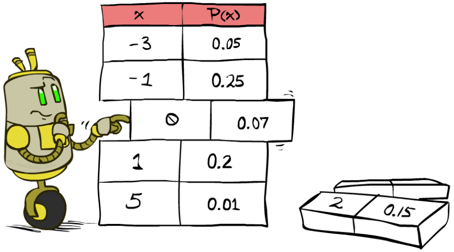
$$\left.\begin{matrix} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{matrix}\right\} X_1, X_2, \dots X_n$$

All variables

- We want:

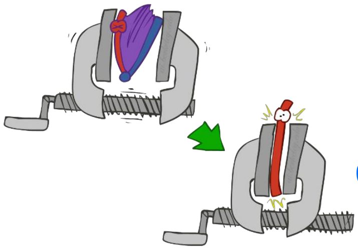
$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence



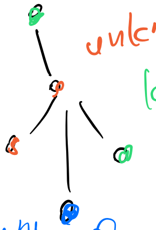
know out inconsistent rows

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



$$\underset{\text{only Q and known}}{P(Q, e_1 \dots e_k)} = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Step 3: Normalize



unknown

known

query Q

hidden

$P(x|w)$

$P(x = \text{big})$

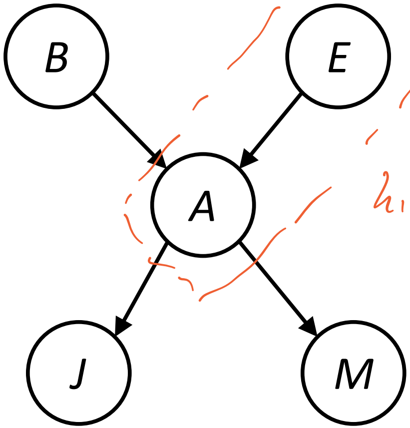
$$\times \frac{1}{Z}$$

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

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

here sum out the hidden variable

Inference by Enumeration in Bayes Net



hidden

- The joint distribution can be computed from a BN by multiplying the conditional distributions
- Then we can do inference by enumeration

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

propositional

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

sum hidden

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

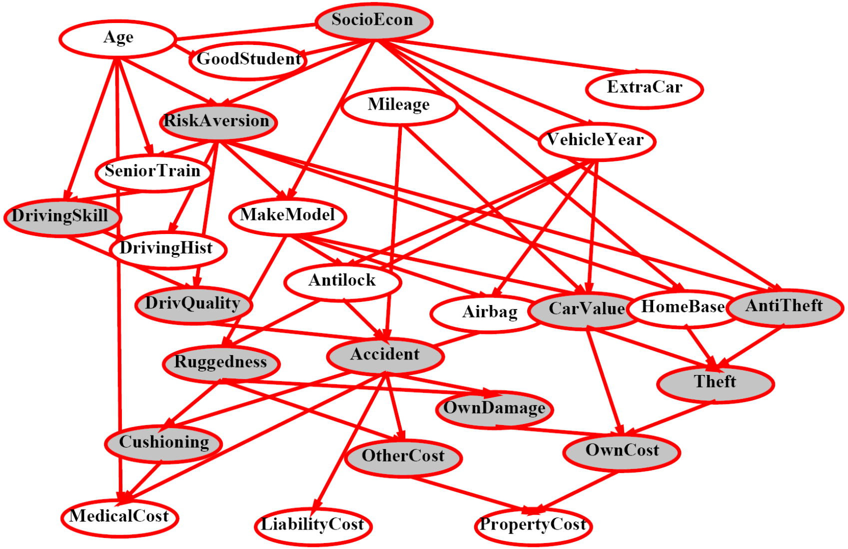
$$P(B)P(+e)P(+a|B,+e)P(+j|+a)P(+m|+a) + P(B)P(+e)P(-a|B,+e)P(+j|-a)P(+m|-a)$$
$$P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$

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

but, BAD

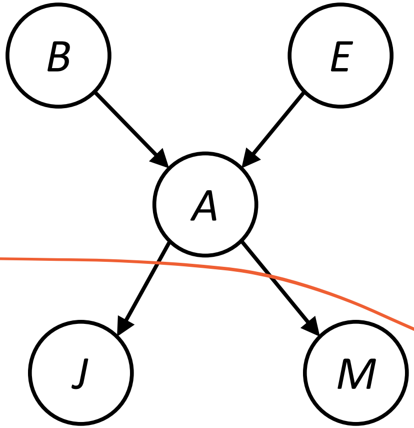
Captial : vector math, sum all val of var in domain

Inference by Enumeration?



$$P(Antilock|observed\ variables) = ?$$

Inference by Enumeration in Bayes Net



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

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

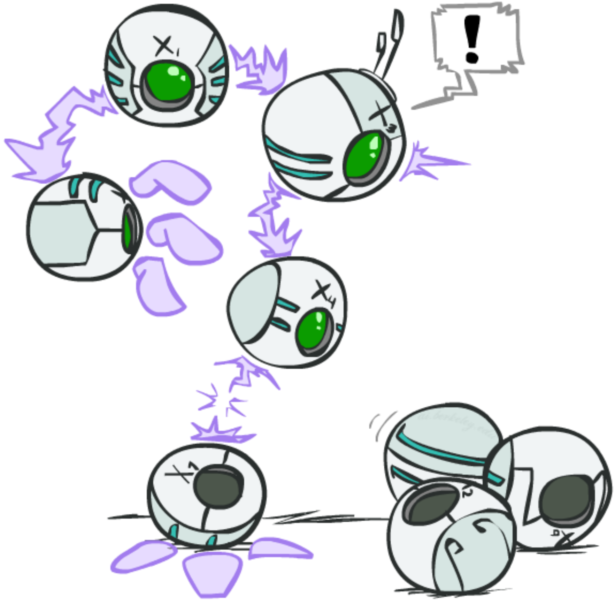
$$= \sum_{e,a} P(B)P(e)P(a|B,e)P(+j|a)P(+m|a)$$

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

$$P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$

Lots of repeated subexpressions!

Variable Elimination



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

Can we do better?

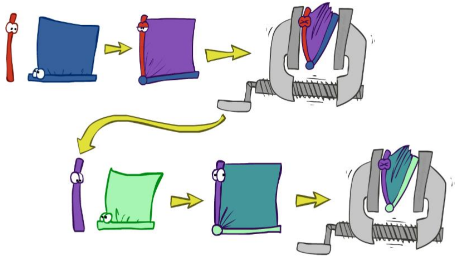
- $\sum_{e,a} P(B)\ P(e)\ P(a|B,e)\ P(j|a)\ P(m|a)$
- $= P(B)P(e)P(a|B,e)P(j|a)P(m|a) + P(B)P(\neg e)P(a|B,\neg e)P(j|a)P(m|a)$

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

  Lots of repeated subexpressions!

Variable elimination: The basic ideas

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

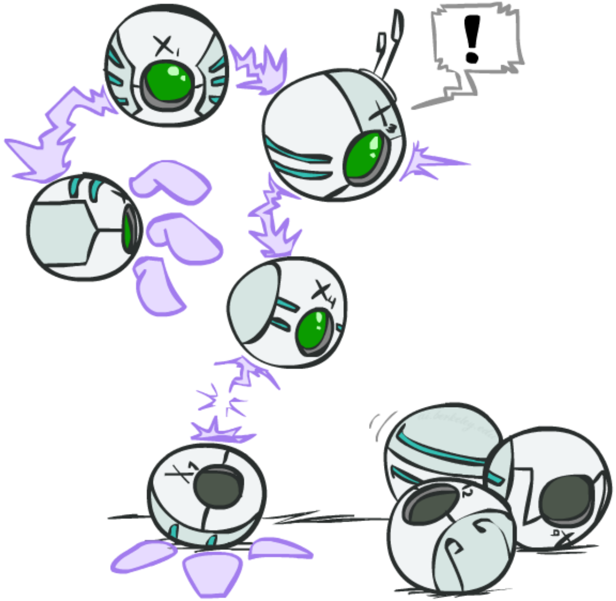


- Do the calculation from the inside out
  - I.e., sum over *a* first, the sum over *e*
  - Problem: $P(a \mid B,e)$ isn't a single number, it's a bunch of different numbers depending on the values of *B* and *e*
  - Solution: use arrays of numbers (of various dimensions) with appropriate operations on them; these are called ***factors***

B, e

vectors.

Operations on Factors



Factors

- A factor is a multi-dimensional array to represent $P(Y_1 \ldots Y_N \mid X_1 \ldots X_M)$
  - If a variable is assigned (represented with lower-case), its dimension is missing (selected) from the array

- Joint distribution: P(X,Y)
  - Entries P(x,y) for all x, y
  - Sums to 1

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

fixed

- Selected joint: P(x,Y)
  - A slice of the joint distribution
  - Entries P(x,y) for fixed x, all y
  - Sums to P(x)

$$P(cold, W)$$

T	W	P
cold	sun	0.2
cold	rain	0.3

Factors

- A factor is a multi-dimensional array to represent $P(Y_1 \ldots Y_N \mid X_1 \ldots X_M)$
  - If a variable is assigned (represented with lower-case), its dimension is missing (selected) from the array

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

$$P(W|cold)$$

T	W	P
cold	sun	0.4
cold	rain	0.6

assign

condition

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

$$P(W|T)$$

T	W	P
hot	sun	0.8
hot	rain	0.2
cold	sun	0.4
cold	rain	0.6

$P(W|hot)$

$P(W|cold)$

Factors

- A factor is a multi-dimensional array to represent $P(Y_1 \ldots Y_N \mid X_1 \ldots X_M)$
  - If a variable is assigned (represented with lower-case), its dimension is missing (selected) from the array
    - Specified family: P(y | X)
      - Entries P(y | x) for fixed y, but for all x
      - Sums to … who knows!

$P(rain|T)$

T	W	P
hot	rain	0.2
cold	rain	0.6

$P(rain|hot)$

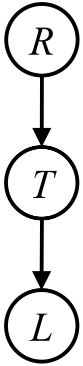
$P(rain|cold)$

Selected.

assign outcome.

Example: Traffic Domain

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



$P(R)$

+r	0.1
-r	0.9

2d

$P(T|R)$

+r	+t	0.8	1
+r	-t	0.2	
-r	+t	0.1	1
-r	-t	0.9	

$P(L|T)$

2d

+t	+l	0.3	1
+t	-l	0.7	
-t	+l	0.1	1
-t	-l	0.9	

$$P(L) = ?$$

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

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

Running Example: Traffic Domain

- Initial factors are local CPTs (one per node)

$P(R)$

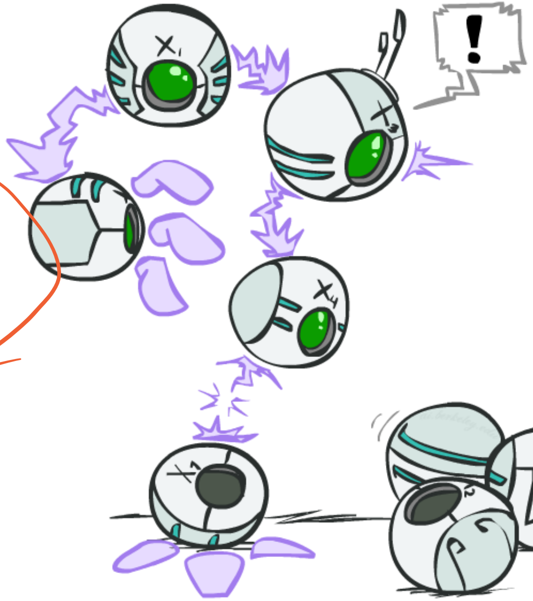
+r	0.1
-r	0.9

$P(T|R)$

+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

$P(T|R)$

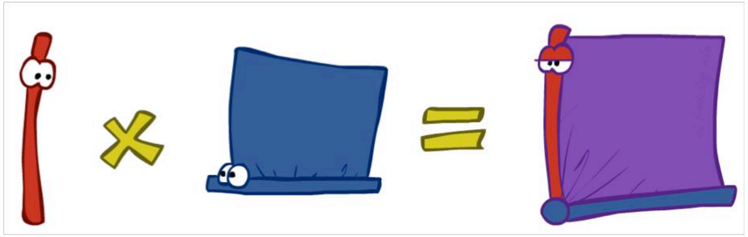
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(+\ell|T)$

+t	+l	0.3
-t	+l	0.1

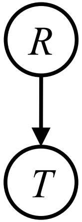
Operation 1: Join Factors

- First basic operation: joining factors
  - Just like a database join
  - Given multiple factors, build a new factor over the union of the variables involved
  - Each entry is computed by pointwise products



- Example: join on R

become 1 big var



$P(R)$

+r	0.1
-r	0.9

$\times$

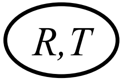
$P(T|R)$

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

$\rightarrow$

$P(R,T)$ 2d

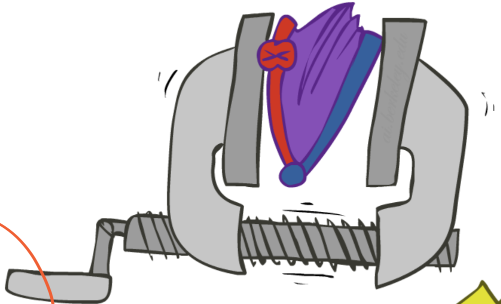
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81



$$\forall r,t: \quad P(r,t) = P(r) \cdot P(t|r)$$

Operation 2: Eliminate

- Second basic operation: eliminating a variable
  - Take a factor and sum out (marginalize) a variable



- Example:

$P(R,T)$

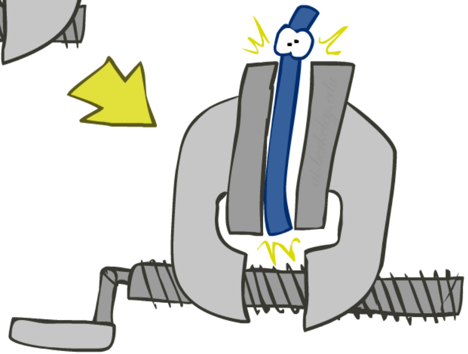
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$\xrightarrow{\text{sum } R}$

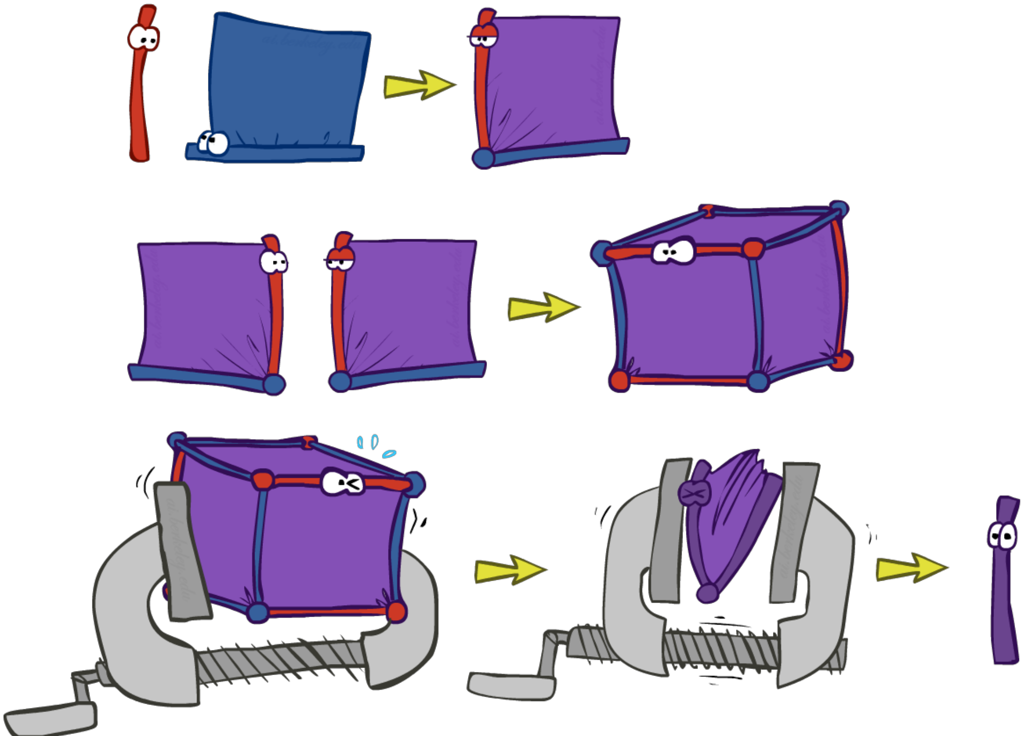
$P(T)$

+t	0.17
-t	0.83

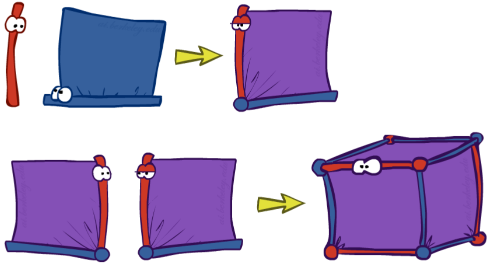
Sum = 1

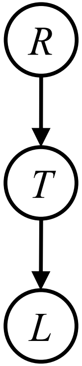


Inference by Enumeration in BN = Multiple Join + Multiple Eliminate



Computing *P*(*L*): 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

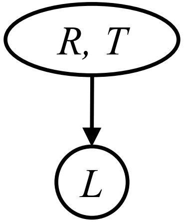
Join

$P(R, T)$

+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

$P(R, T, L)$

+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

Computing $P(L)$: 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

A factor of exponential size!

Sum out R

T, L

$P(T, L)$

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

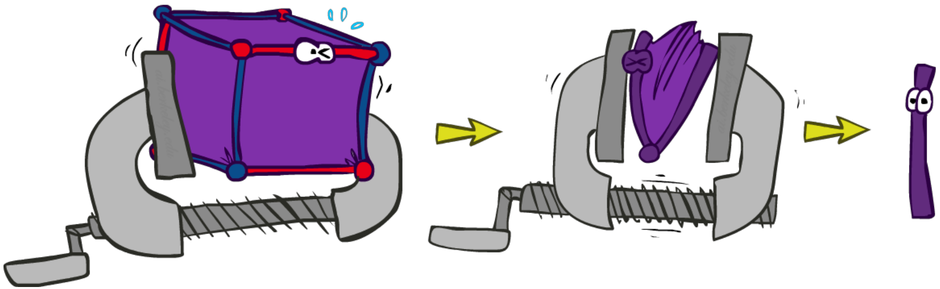
Sum = 1

Sum out T

L

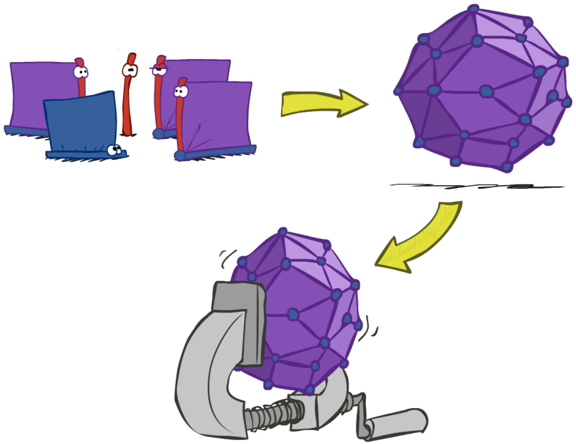
$P(L)$

+l	0.134
-l	0.866

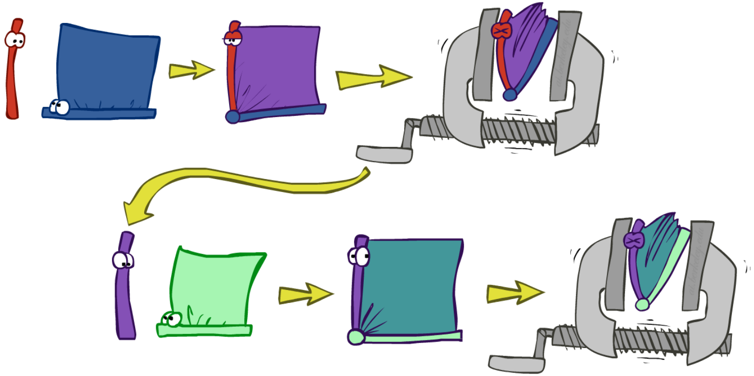


Inference by Enumeration vs. Variable Elimination

- 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 elimination!
  - Called "Variable Elimination"
  - Still NP-hard, but usually much faster than inference by enumeration



Inference by Enumeration

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

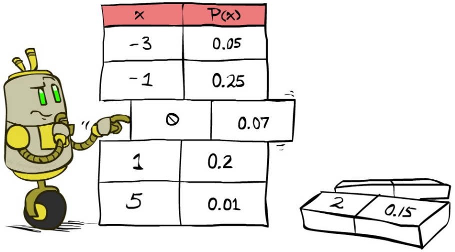
$\left.\begin{array}{l} E_1 \ldots E_k = e_1 \ldots e_k \\ Q \\ H_1 \ldots H_r \end{array}\right\} \begin{array}{c} X_1, X_2, \ldots X_n \\ \textit{All variables} \end{array}$

- We want:

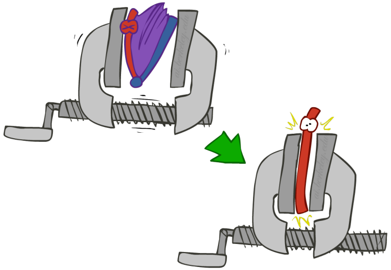
** Works fine with multiple query variables, too*

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

- Step 1: Select the entries consistent with the evidence



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



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

- Compute joint
- Sum out hidden variables

- Step 3: Normalize

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

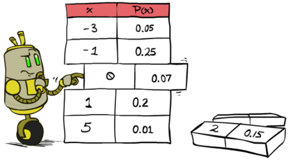
$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

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

Variable Elimination

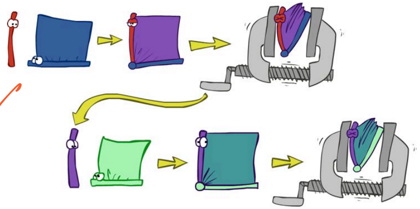
- Query: $P(Q|E_1 = e_1, \ldots E_k = e_k)$

- 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

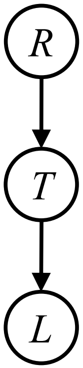
eliminate hidden



- Join all remaining factors and normalize

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

Traffic Domain



$$P(L) = ?$$

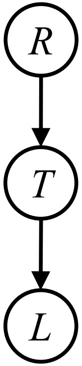
- Inference by Enumeration

$$= \underbrace{\sum_t \underbrace{\sum_r \underbrace{P(L|t)\underbrace{P(r)P(t|r)}_{\text{Join on r}}}_{\text{Join on t}}}_{\text{Eliminate r}}}_{\text{Eliminate t}}$$

- Variable Elimination

$$= \underbrace{\sum_t \underbrace{P(L|t) \underbrace{\sum_r \underbrace{P(r)P(t|r)}_{\text{Join on r}}}_{\text{Eliminate r}}}_{\text{Join on t}}}_{\text{Eliminate t}}$$

Marginalizing Early! (aka VE)



$P(R)$

+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

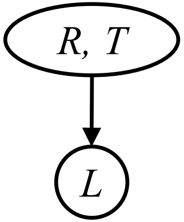
$P(L|T)$

+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



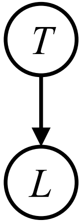
$P(L|T)$

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

Sum out R

$P(T)$

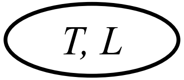
+t	0.17
-t	0.83



$P(L|T)$

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

Join T



$P(T,L)$

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

Sum out T

$P(L)$

+l	0.134
-l	0.866

target

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:

select evidence

$P(+r)$

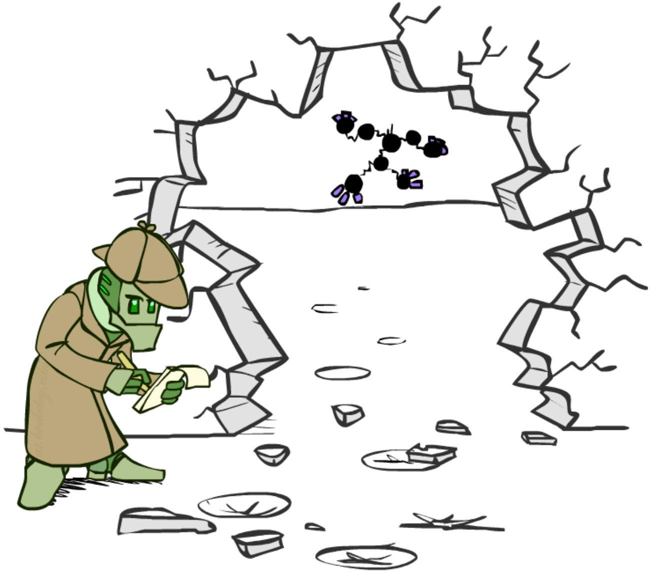
+r	0.1

$P(T|+r)$

+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

(There)

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

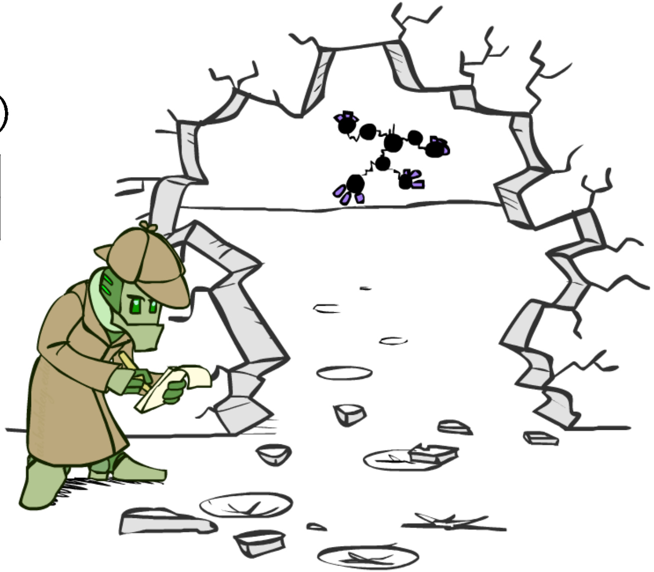
$$P(L | +r)$$

+l	0.26
-l	0.74

de : l . nom

- To get our answer, just normalize this!

- That 's it!

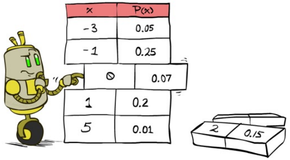


General Variable Elimination

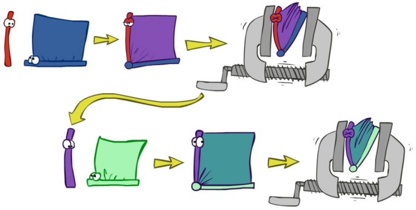
- Query: $P(Q|E_1 = e_1, \dots E_k = e_k)$

  may be multi

- 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



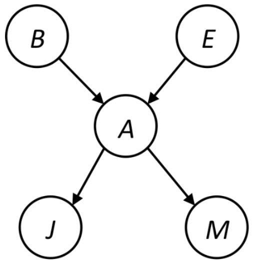
- Join all remaining factors and normalize

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

same thing:
only diff: normalized

Example

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



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

selected.

Choose A

eliminate A

$P(A|B,E)$

$P(j|A)$

$P(m|A)$

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

1, 2d

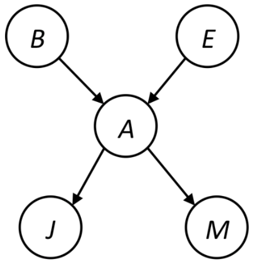
$P(j,m,A|B,E)$

summed.

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

Example

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



Choose E

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

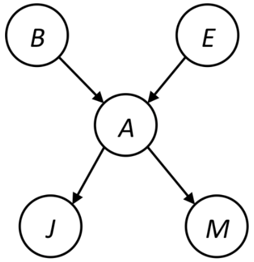
$P(B) \qquad P(j, m|B)$

Finish with B

$$\left.\begin{matrix} P(B) \\ P(j, m|B) \end{matrix}\right. \xrightarrow{\;\times\;} P(j, m, B) \xrightarrow{\text{Normalize}} P(B|j, m)$$

Same Example in Equations

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



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

$$
\begin{aligned}
P(B|j,m) &\propto P(B,j,m) \\
&= \sum_{e,a} P(B,j,m,e,a) \\
&= \sum_{e,a} P(B)P(e)P(a|B,e)P(j|a)P(m|a) \\
&= \sum_{e} P(B)P(e)\sum_{a} P(a|B,e)P(j|a)P(m|a) \\
&= \sum_{e} P(B)P(e)f_1(B,e,j,m) \\
&= P(B)\sum_{e} P(e)f_1(B,e,j,m) \\
&= P(B)f_2(B,j,m)
\end{aligned}
$$

don't

marginal can be obtained from joint by summing out

use Bayes' net joint distribution expression

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

joining on a, and then summing out gives f_1

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

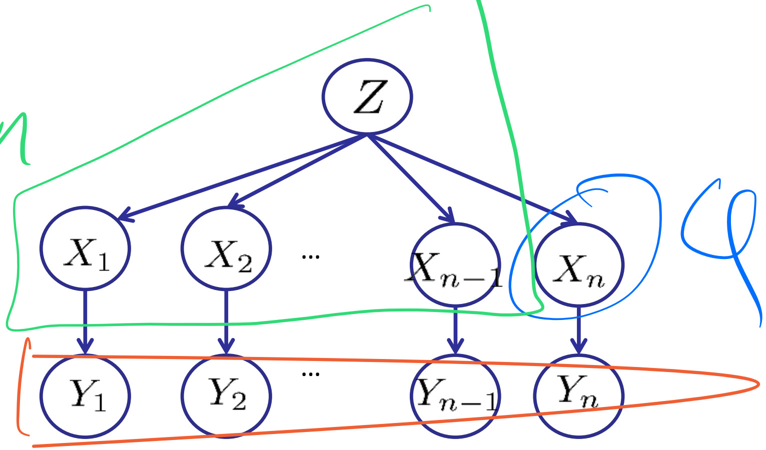
joining on e, and then summing out gives f_2

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

Variable Elimination Ordering

- Query: $P(X_n | y_1, \ldots, y_n)$
- Two different orderings: Z, X_1, …, X_{n-1} and X_1, …, X_{n-1}, Z.
- What is the size of the maximum factor generated for each of the orderings?

Hidden



q

known.

$p(y_1 | x_1)$

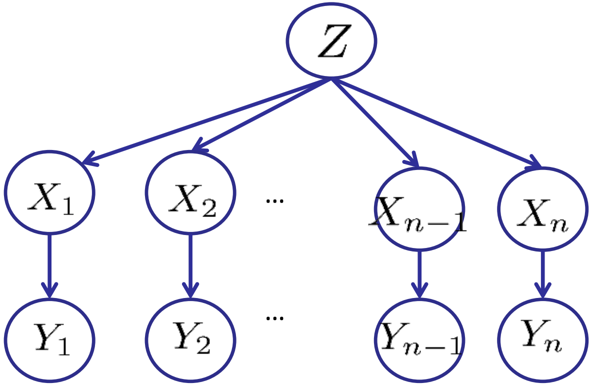
$p(y_1 | z)$

$p(y_1, x_1 | z)$

$p(y_1 | z)$

Variable Elimination Ordering

- $Z, X_1, \ldots, X_{n-1}$



$$P(Z)P(X_1|Z)P(X_2|Z),\ldots,P(X_n|Z)$$

$P(X_1 \cdots x_{n-1}\ Z)$

$$\Downarrow$$

$$f_1(X_1, X_2, \ldots, X_n)$$

$$\Downarrow$$

$$f_2(y_1, X_2, \ldots, X_n)$$

$$\Downarrow$$

$$f_3(y_1, y_2, \ldots, X_n)$$

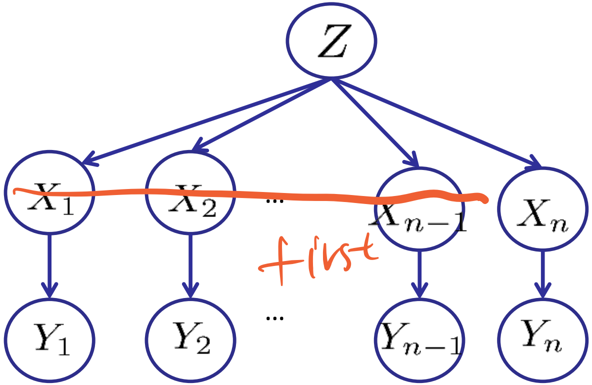
$$\vdots$$

$$2^{n+1}$$

Variable Elimination Ordering

- $X_1, \ldots, X_{n-1}, Z$

good ordering



$$P(X_1|Z)P(y_1|X_1)$$

$$f_1(Z, y_1)$$

$$f_1(Z,y_1), f_2(Z,y_2), \ldots, f_{n-1}(Z,y_{n-1}), P(Z), P(X_n|Z)$$

$$f_n(X_n, y_1, \ldots, y_{n-1})$$

known

22

VE: Computational Complexity

- The size of the largest factor determines the time and space complexity of VE
- The elimination ordering can greatly affect the size of the largest factor.
  - E.g., previous slide's example 2^{n+1} vs. 2^2

vector elimination

- Does there always exist an ordering that only results in small factors?
  - No!

NP-hard

Reduction from 3SAT

$$(x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1 \vee x_3 \vee \neg x_4) \wedge (x_2 \vee \neg x_2 \vee x_4) \wedge (\neg x_3 \vee \neg x_4 \vee \neg x_5) \wedge (x_2 \vee x_5 \vee x_7) \wedge (x_4 \vee x_5 \vee x_6) \wedge (\neg x_5 \vee x_6 \vee \neg x_7) \wedge (\neg x_5 \vee \neg x_6 \vee x_7)$$

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

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

$$\ldots$$

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

$$Y_{1,2} = Y_1 \wedge Y_2$$

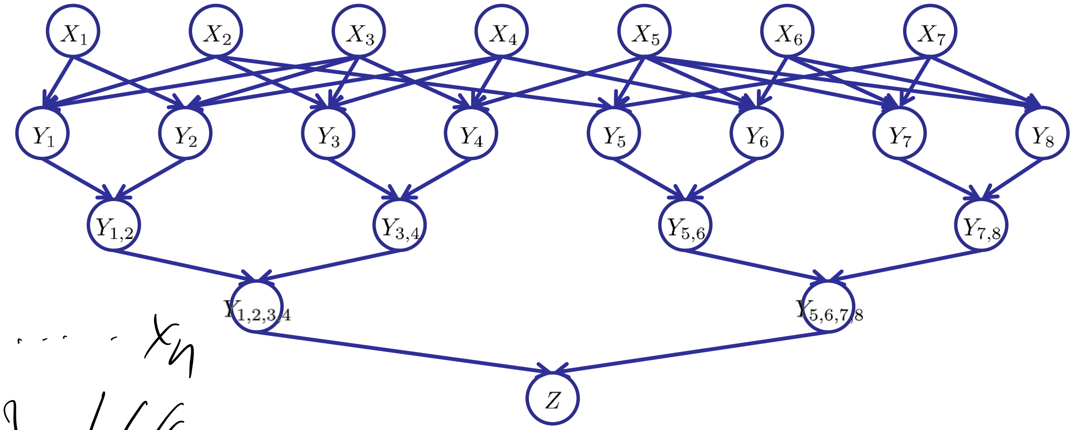
$$\ldots$$

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

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

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

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



$Y_1 \; . \; . \; . \; . \; . \; X_n$

Z

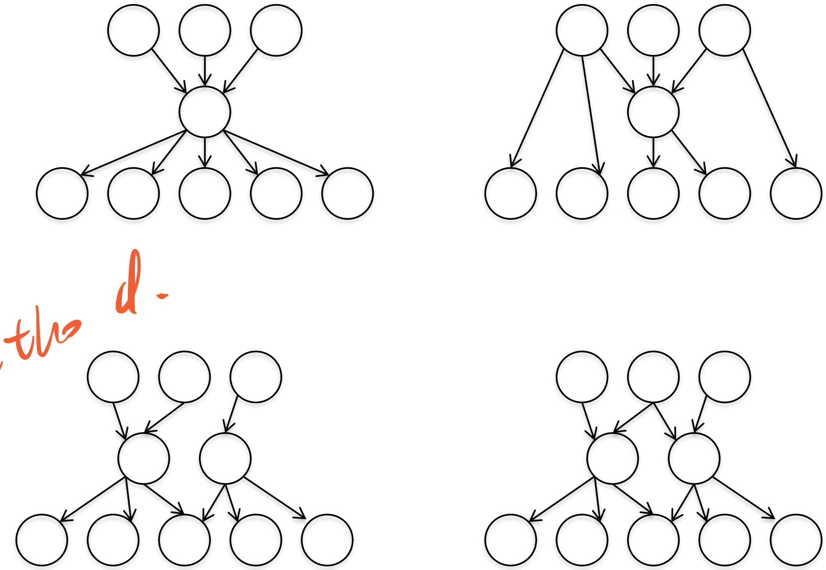
$P(z)$

- 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.

Polytrees

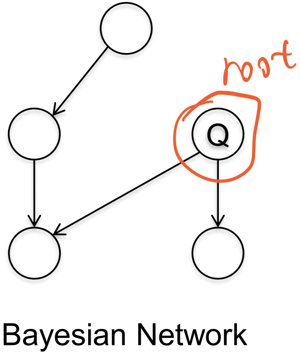
- A polytree is a directed graph with no undirected cycles

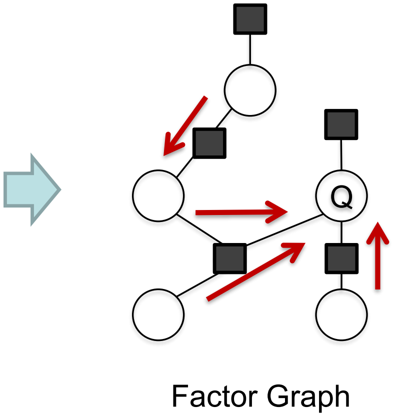
can always find efficient method.



Variable Elimination on Polytrees

- For poly-tree BNs, the complexity of VE is ***linear in the BN size*** (number of CPT entries) with the following elimination ordering:
  - Convert to a factor graph
  - Take Q as the root
  - Eliminate from the leaves towards the root



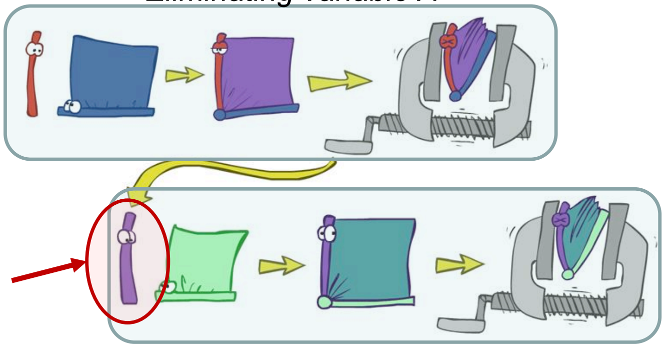


Variable Elimination on Polytrees

- VE for poly-tree BNs is equivalent to
  - Sum-product message passing algorithm or belief propagation algorithm

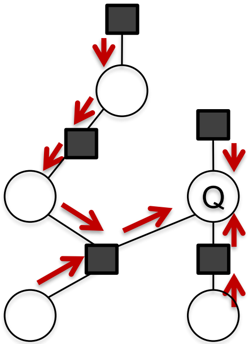
  (i.e., passing messages/beliefs from leaf nodes to the root node)

Eliminating variable A



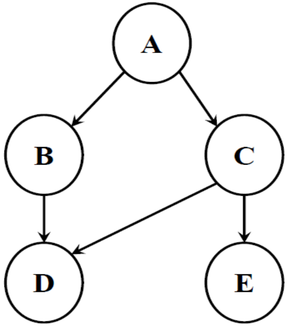
Message from A, which is used in eliminating B

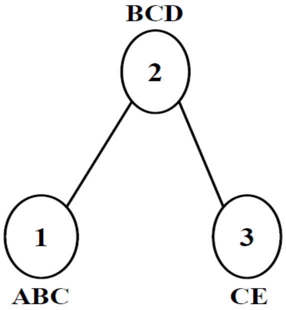
Eliminating variable B



Message Passing on General Graphs

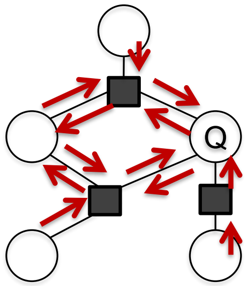
- Exact inference: Junction Tree Algorithm
  - Group individual nodes to form cluster nodes in such a way that the resulting network is a polytree (called a junction tree or join tree)
  - Run a sum-product-like algorithm on the junction tree.
  - *Intractable* on graphs with large cliques (i.e., large tree-width).





Message Passing on General Graphs

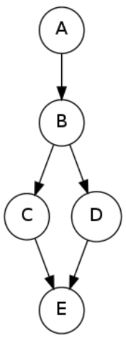
- Approximate inference: Loopy Belief Propagation
  - Simply pass the messages on the general graph
    - Will not terminate with loops
    - Run until convergence (not guaranteed!)
  - *Approximate* but *tractable* for large graphs.
  - Sometime works well, sometimes not at all.



Summary

- Exact inference of Bayesian networks
  - Enumeration
    - exponential complexity
  - Variable Eliminating
    - worst-case exponential complexity, often better
  - VE on polytrees
    - linear complexity
    - = message passing
  - Message passing on general graphs
    - Junction Tree Algorithm
    - Loopy Belief Propagation: no longer exact

Assume the following Bayes Net and corresponding CPTs. In this exercise, we are given the query P(C|e=1), and we will complete the tables for each factor generated during the elimination process.



$P(c, e=1)$

$P(e=1 \mid c, D)$

After introducing evidence, we have the following probability tables.

A	$P(A)$
0	0.900
1	0.100

B	A	$P(B\|A)$
0	0	0.700
1	0	0.300
0	1	0.500
1	1	0.500

B=0

B=1

C	B	$P(C\|B)$
0	0	0.400
1	0	0.600
0	1	0.400
1	1	0.600

D	B	$P(D\|B)$
0	0	0.300
1	0	0.700
0	1	0.100
1	1	0.900

C	D	$P(e=1\|C,D)$
0	0	0.400
1	0	0.600
0	1	0.400
1	1	0.200

Three steps are required for elimination, with the resulting factors listed below:
Step 1: eliminate A.We get the factor $f_1(B) = \sum_a$ P(a)P(B|a).
Step 2: eliminate B.We get the factor $f_2(C,D) = \sum_b$ P(C|b)P(D|b)$f_1(b)$.
Step 3: eliminate D.We get the factor $f_3(C, e=1) = \sum_d$ P(e=1|C,d)$f_2(C,d)$.

P(c=0, e=0)

P(C=1, e=0)

P(C=0, e=1)

P(C=1, e=1),

Complete the tables below for the factors generated during elimination.

B	$f_1 : P(B)$
0	Blank 1
1	Blank 2

C=0

C	D	$f_2 : P(C,D)$
0	0	Blank 3
1	0	Blank 4
0	1	0.306
1	1	0.458

C	$f_3 : P(C, e=1)$
0	Blank 5
1	0.177

Sum up D: +d, -d

C=1

Blank 1 = 0.680
Blank 2 = 0.320
Blank 3 = 0.0944
Blank 4 = 0.1416
Blank 5 =

After getting the final factor P(C, e=1), a final renormalization step needs be carried out to obtain the conditional probability P(C|e=1). Please fill into the table below the final conditional probability.

C	$P(C \mid e=1)$
0	Blank 6
1	Blank 7

normalize!

Blank 6 =
Blank 7 =