Variable Elimination

Notation

Suppose B is a Boolean random variable

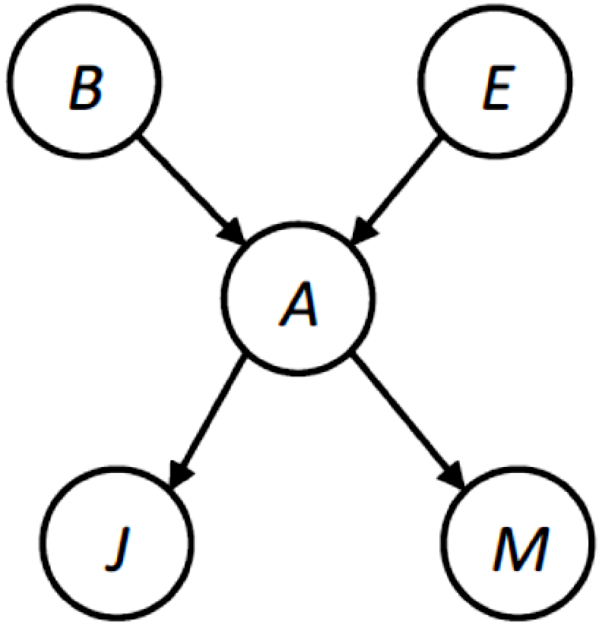
Then all these notations mean the same:

$B=true$ is equivalent to b which is equivalent to $+b$

$B=false$ is equivalent to $\neg b$ which is equivalent to $-b$

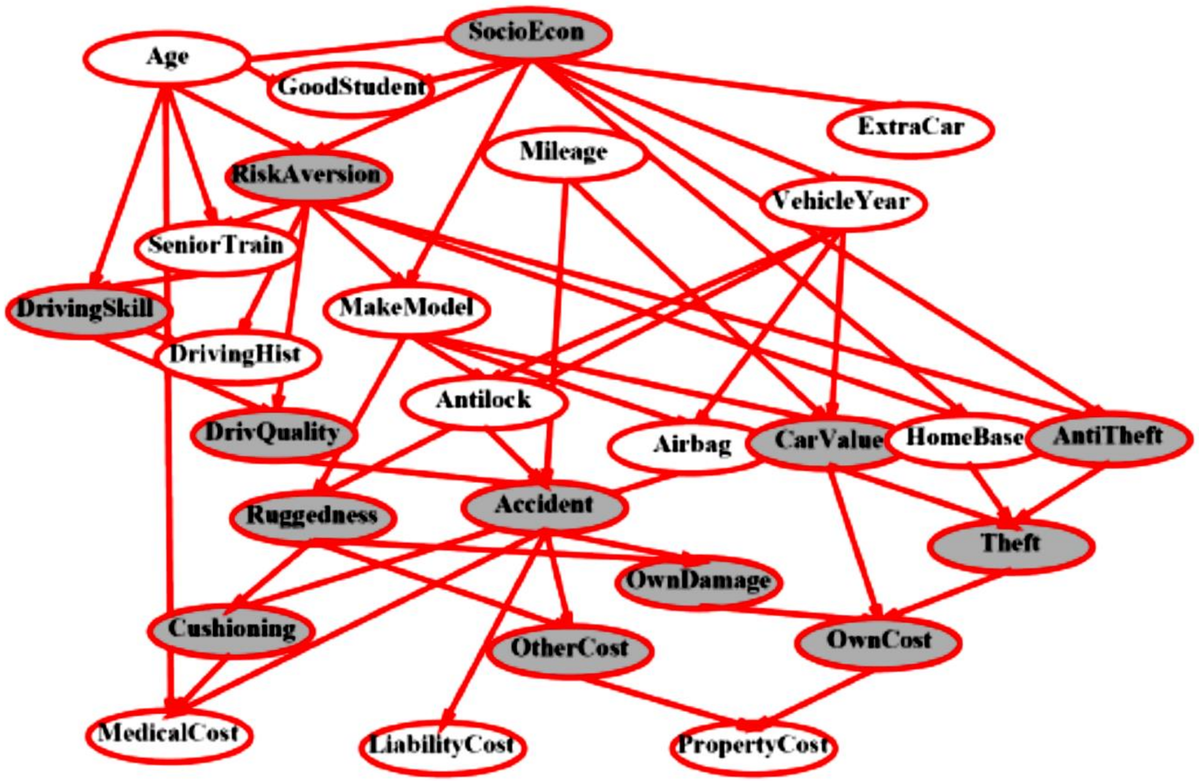
Recall: Inference by Enumeration

- Given unlimited time, inference in BNs is easy
- Recipe:
  - State the marginal probabilities you need
  - Figure out ALL atomic probabilities you need
  - Calculate and combine them
- Example: $P(+b|+j,+m) = \dfrac{P(+b,+j,+m)}{P(+j,+m)}$



$$
\begin{aligned}
P(+b,+j,+m) = \quad & P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+ \\
& P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a)
\end{aligned}
$$

But how about this one using Inference by Enumeration



Notation: *B=true* is same as *b* and same as *+b*

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

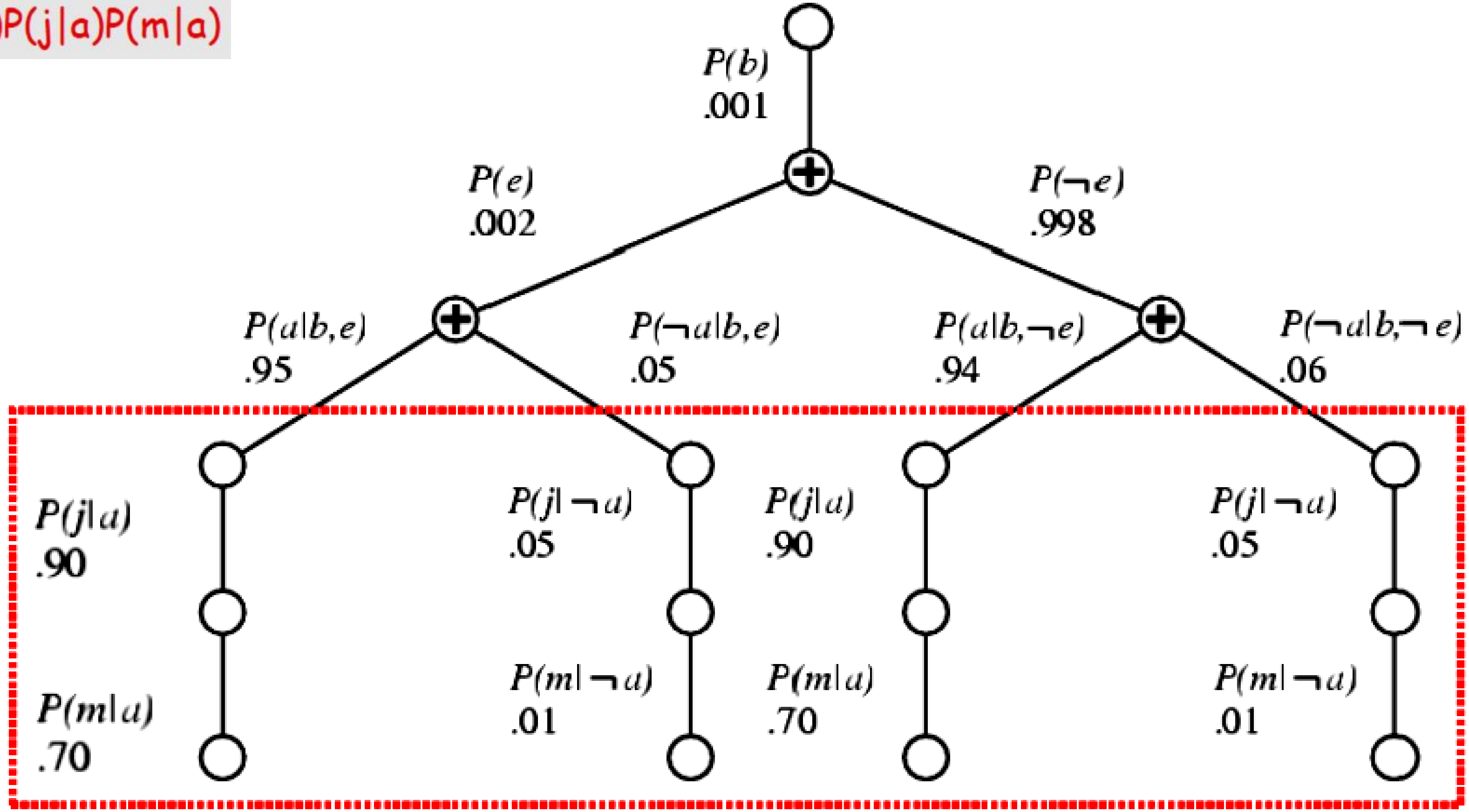
16 mults+3 adds

Factored form:

$P(b|j,m) = \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a|b,e)P(j|a)P(m|a)$

11 multiplications + 3 additions

Structure of Factored Form Computation



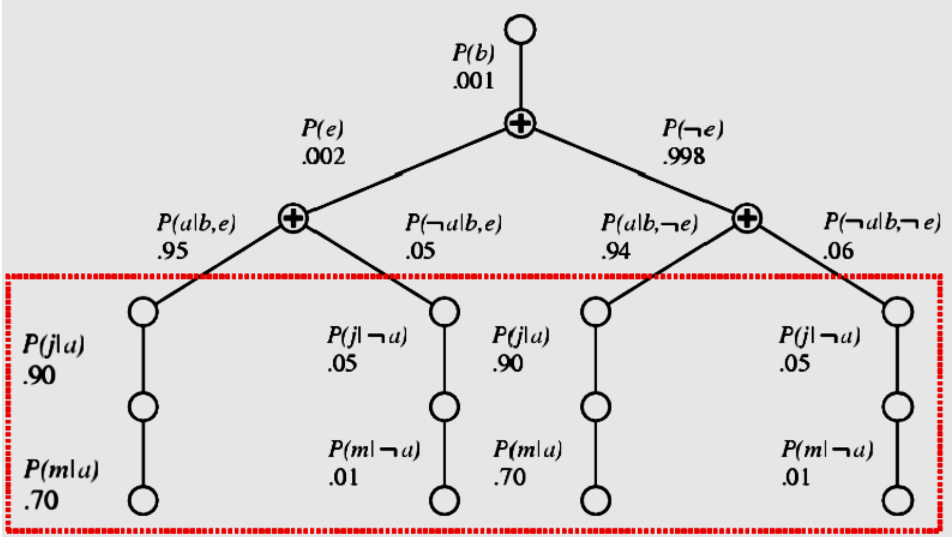
Inference by Enumeration vs. Variable Elimination

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

$$
\begin{aligned}
P(+b,+j,+m) = \quad & P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+ \\
& P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a)
\end{aligned}
$$

Factored Form: $P(b|j,m) = \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a|b,e)P(j|a)P(m|a)$

- Idea: interleave joining and marginalizing!
  - Called “Variable Elimination”
  - Still NP-hard, but usually much faster than inference by enumeration



Variable Elimination contains two main ideas that help address this exponential blowup of the joint distribution.

The 1st idea is about exploiting properties of arithmetic: For example, if you have an expression:

$$yx_1 + yx_2 + yx_3 + yx_4$$

$$\begin{aligned} P(+b,+j,+m) = \; & P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+ \\ & P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+ \\ & P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+ \\ & P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a) \end{aligned}$$

This expression has four multiplications and three additions. However, using the well known properties of arithmetic we can rearrange the expression:

$$y(x_1 + x_2 + x_3 + x_4)$$

$$P(+b,+j,+m) = P(+b)* \left[\begin{aligned} & P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+ \\ & P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+ \\ & P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+ \\ & P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a) \end{aligned} \right]$$

Now we have ONE multiplication and three additions!

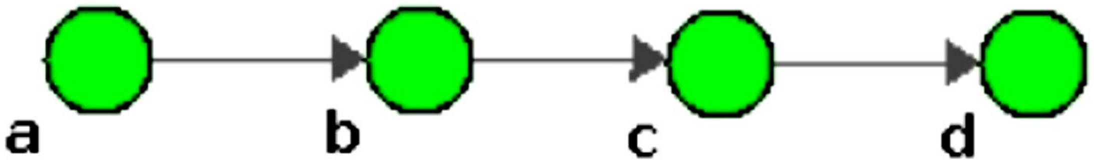
The 2nd idea is about exploiting the structure of the Bayesian Network.
Many of the sub-expressions in the joint only depend on a small number of variables.

$$p(D) = \sum_a \sum_b \sum_c p(a)p(b \mid a)p(c \mid b)p(D \mid c)$$

$$\left.\begin{aligned} p(B) &= \sum_a p(a)p(B \mid a) \\ p(C) &= \sum_b p(b)p(C \mid b) \\ p(D) &= \sum_c p(c)p(D \mid c) \end{aligned}\right\} \quad p(D) = \sum_c p(D|c)\left(\sum_b p(C|b)\left(\sum_a p(a)p(B|a)\right)\right)$$

$$y(\overbrace{x_1 + x_2 + x_3 + x_4})$$

The inner expression p(B) is computed 1st for all values of B and stored and so we compute them once (thus eliminating A). Then p(C) is computed with the values of p(B) and stored. Finally p(D) is computed with the stored values of p(C).



$$\begin{bmatrix} P(+a) = .7 \\ P(-a) = .3 \end{bmatrix} \begin{bmatrix} P(+b| + a) = .4 \\ P(-b|+a) = .6 \\ P(+b| - a) = .6 \\ P(-b|-a) = .4 \\ \end{bmatrix}$$

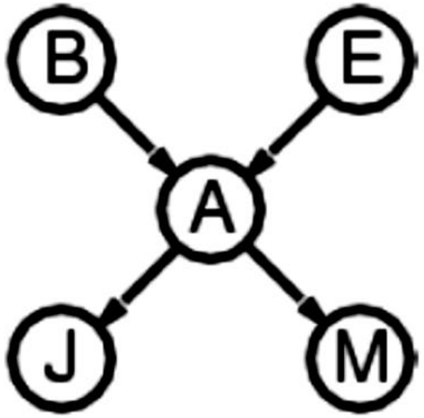
$$p(D) = \sum_c p(D|c)\left(\sum_b p(C|b)\left(\sum_a p(a)p(B|a)\right)\right)$$

$$\begin{bmatrix} .4 * .7 + .6 * 3 \\ .6 * .7 + .4 * .3 \end{bmatrix}$$

In general terms, lets assume a Bayesian network that has the structure of a chain with n variables $X_1 -> ... -> X_n$,where each variable has k possible values. Computing $p(X_{i+1})$ can be defined recursively.

$$p(X_{i+1}) = \sum_{x_i} p(X_{i+1}|x_i)p(x_i)$$

Each recursive step is $O(k^2)$ and we recurse through all n variables in the worst case scenerio, so the total bigO is $O(nk^2)$. This is much better than generating the full k^n probabilities to sum over in the joint distribution.



$$
\begin{aligned}
P(b|j,m) &= \alpha\, P(b,j,m) \\
&= \alpha\, \Sigma_{e,a}\, P(b,j,m,e,a) \\
&= \alpha\, \Sigma_{e,a}\, P(b)\, P(e)\, P(a|b,e)\, P(j|a)\, P(m|a) \\
&= \alpha\, P(b)\, \Sigma_e\, P(e)\, \Sigma_a\, P(a|b,e)P(j|a)P(m|a)
\end{aligned}
$$

- Join all factors containing *a*
- Sum out *a* to get new function of b,e,j,m only

Factors

Factors are Set of Tables.
Examples:

- 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

- 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

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

$P(W|cold)$

T	W	P
cold	sun	0.4
cold	rain	0.6

$P(T,W)$

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

- 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

$P(W|hot)$

$P(W|cold)$

- Specified family: P(y | X)
  - Entries P(y | x) for fixed y, but for all x

$P(rain|T)$

T	W	P
hot	rain	0.2
cold	rain	0.6

$P(rain|hot)$

$P(rain|cold)$

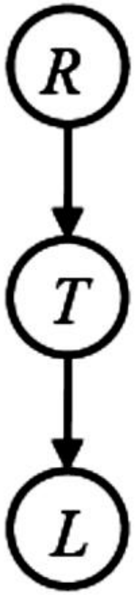
Factor

In general, when we write $P(Y_1 \dots Y_N \mid X_1 \dots X_M)$

- It is a "factor," a multi-dimensional array
- Its values are all $P(y_1 \dots y_N \mid x_1 \dots x_M)$
- Any assigned X or Y is a dimension missing (selected) from the array

Example: Traffic Problem

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



$P(R)$

+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

Variable Elimination 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

- VE: Alternately join factors and eliminate variables

Eliminating a Variable:

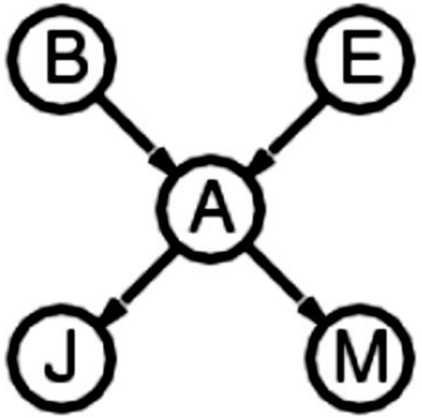
1. *join* all factors containing that variable, multiplying probabilities
2. *sum out* the influence of the variable

$$P(b|j,m) = \alpha \overbrace{P(b)\, \overbrace{\Sigma_e P(e)\, \overbrace{\Sigma_a P(a|b,e)P(j|a)P(m|a)}^{\text{Eliminate a}}}^{\text{Eliminate e}}}^{\text{Remaining factor is a function of b, j, m}}$$

$\downarrow$

Function of b,j,m

Example of VE: P(J)



$P(J)$

$= \Sigma_{M,A,B,E}\ P(J,M,A,B,E)$

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

$= \Sigma_A P(J|A)\ \Sigma_M P(M|A)\ \Sigma_B P(B)\ \Sigma_E P(A|B,E)P(E)$

$= \Sigma_A P(J|A)\ \Sigma_M P(M|A)\ \Sigma_B P(B)\ f1(A,B)$

$= \Sigma_A P(J|A)\ \Sigma_M P(M|A)\ f2(A)$

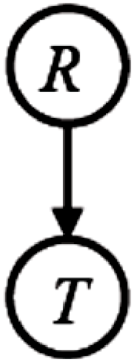
$= \Sigma_A P(J|A)\ f3(A)$

$= f4(J)$

Variable Elimination Algorithm: Details

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



$P(R)$

+r	0.1
-r	0.9

$\times$ $P(T|R)$

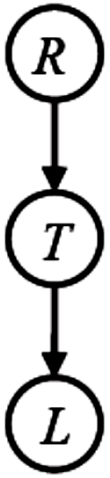
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(R,T)$

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

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

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

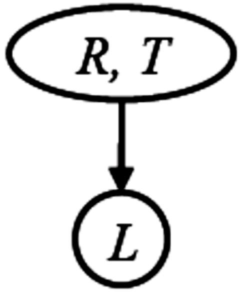
$P(L|T)$

+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

$P(R,T,L)$

+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

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 $\longrightarrow$

$P(T)$

+t	0.17
-t	0.83

Multiple Elimination

R, T, L T, L 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

$P(T, L)$

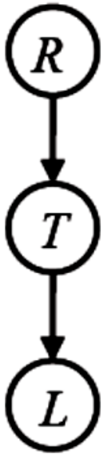
+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

Sum out T

$P(L)$

+l	0.134
-l	0.886

VE in Action - Marginalizing Early



$P(R)$

+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

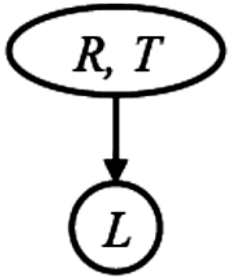
$P(L|T)$

+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



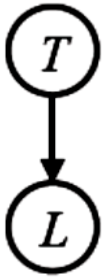
$P(L|T)$

+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	0.886

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

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

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

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

$P(+r, L)$

+r	+l	0.026
+r	-l	0.074

Normalize $\Rightarrow$

$P(L | +r)$

+l	0.26
-l	0.74

- To get our answer, just normalize this!

- That's it!

General Variable Elimination

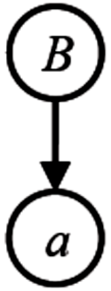
- Query: $P(Q|E_1 = e_1, \dots E_k = e_k)$
- Start with initial factors:
  - Local CPTs (but instantiated by evidence)
- While there are still hidden variables (not Q or evidence):
  - Pick a hidden variable H
  - Join all factors mentioning H
  - Eliminate (sum out) H
- Join all remaining factors and normalize

VE applied to Bayes Rule

Start / Select

$P(B)$

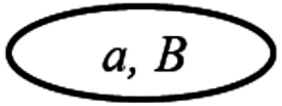
B	P
+b	0.1
¬b	0.9



$P(A|B) \rightarrow \boldsymbol{P(a|B)}$

B	A	P
+b	+a	0.8
~~+b~~	~~¬a~~	~~0.2~~
¬b	+a	0.1
~~¬b~~	~~¬a~~	~~0.9~~

Join on B



$P(a, B)$

A	B	P
+a	+b	0.08
+a	¬b	0.09

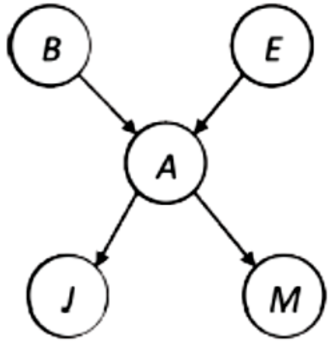
Normalize

$P(B|a)$

A	B	P
+a	+b	8/17
+a	¬b	9/17

VE applied to Alarm Example

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



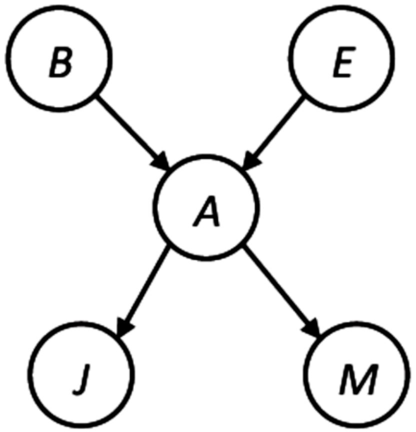
$$P(B) \qquad P(E) \qquad P(A|B,E) \qquad P(j|A) \qquad P(m|A)$$

Choose A

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

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

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

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

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

1.

Choose A

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

$P(j|A)$

A	J	P(J|A)
+a	+j	0.9
-a	+j	0.05

$P(m|A)$

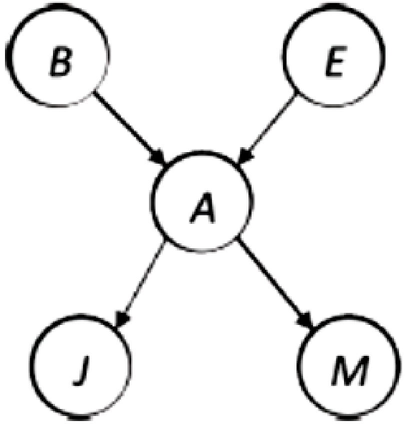
A	M	P(M|A)
+a	+m	0.7
-a	+m	0.01

***P(j,m|A)*:**

P(j,m|+a) =.63
P(j,m|-a) =.005

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

B	E	A	$P(j,m,A\|B,E)$	$P(j,m\|B,E)$
+b	+e	+a	. 95*.9*.7=.5985	.5985+.000025 =.599
+b	+e	-a	.05*.05*.01=.000025	
+b	-e	+a	.94*.9*.7=.5922	.5922+.00003 =.592
+b	-e	-a	.06*.05*.01=.00003	
-b	+e	+a	.29*.9*.7=.1827	.1827+.000355 =.183
-b	+e	-a	.71*.05*.01=.000355	
-b	-e	+a	.001*.9*.7=.00063	.00063+.0005 =.001
-b	-e	-a	.999*.1*.01=.0005	



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

Choose E

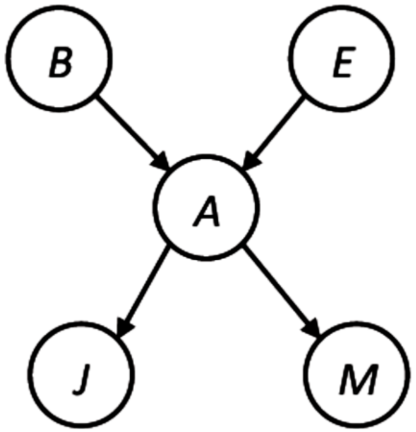
$$P(E) \quad P(j,m|B,E) \xrightarrow{\times} P(j,m,E|B) \xrightarrow{\Sigma} P(j,m|B)$$

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

Finish with B

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

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

Choose A

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

Choose E

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

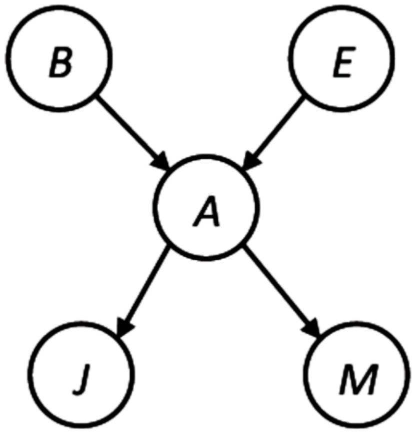
2.

$P(E)$

E	P(E)
+e	0.002
-e	0.998

B	E	$P(j,m\|B,E)$	= $P(j,m,E\|B)$	= $P(j,m\|B)$
+b	+e	.599	.599*.002 = .0012	.0012+.591 = .59
+b	-e	.592	.592*.998 = .591	
-b	+e	.183	.183*.002 = .0004	.0004+.001 = .001
-b	-e	.001	.001*.998 = .001	

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

Choose A

1. $P(A|B,E)$, $P(j|A)$, $P(m|A)$ $\xrightarrow{\times}$ $P(j,m,A|B,E)$ $\xrightarrow{\sum}$ $P(j,m|B,E)$

Choose E

2. $P(E)$, $P(j,m|B,E)$ $\xrightarrow{\times}$ $P(j,m,E|B)$ $\xrightarrow{\sum}$ $P(j,m|B)$

Finish with B

3. $P(B)$, $P(j,m|B)$ $\xrightarrow{\times}$ $P(j,m,B)$

3.

$P(B)$

B	P(B)
+b	.001
-b	.999

$P(j,m|B)$

(+b)	.59
(-b)	.001

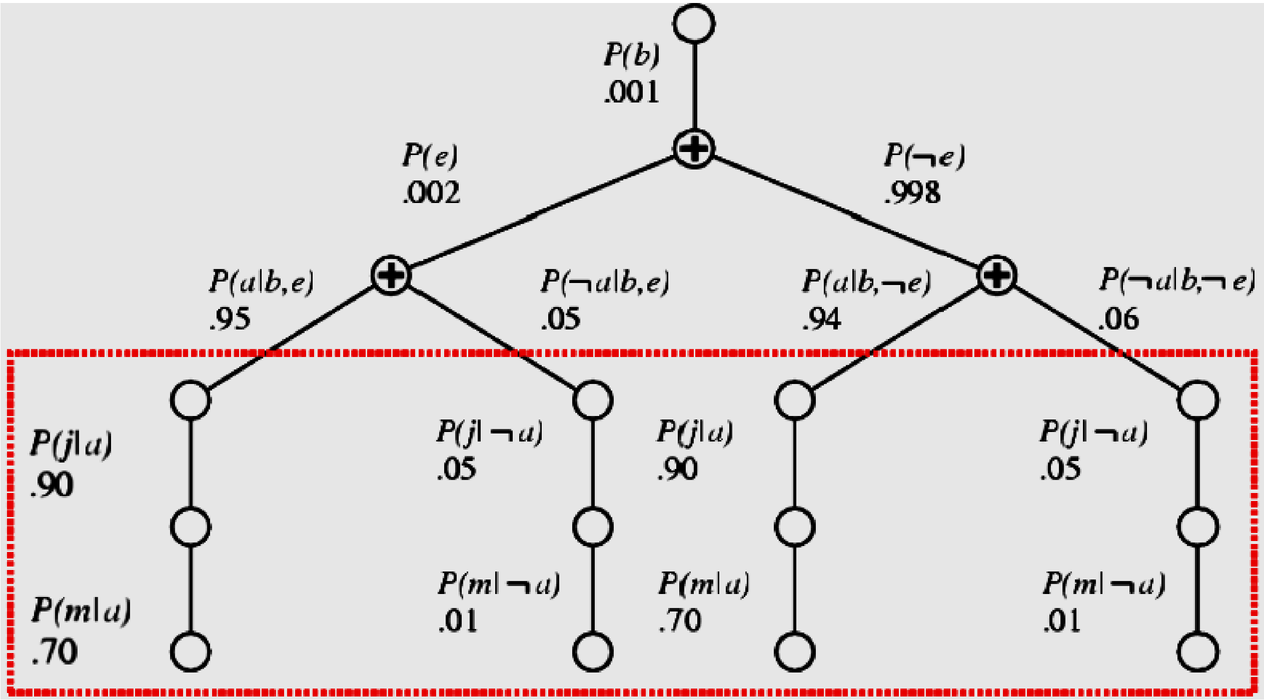
=

$P(j,m,B)$

.592*.001=.0006 (+b)
.001*.999=.001 (-b)

4. Normalize → $P(B|j,m)$ =

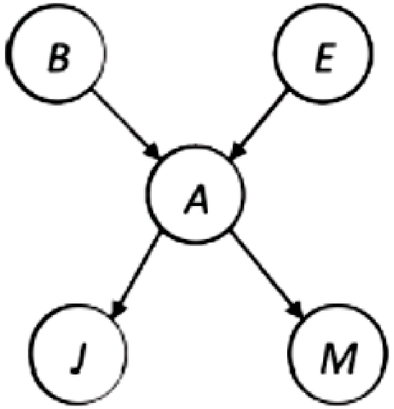
P(+b|j,m)	.375
P(-b|j,m)	.625



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

$$\alpha \mathbf{P}(B) \overbrace{\sum_e P(e) \underbrace{\sum_a \underbrace{\mathbf{P}(a|B,e)P(j|a)P(m|a)}_{P(j,m,A|B,E)}}_{P(j,m|B,E)}}^{P(j,m|B)}$$

Alarm Example in Equations



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

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

$$
\begin{aligned}
P(B|j,m) &\propto P(B,j,m) \\
&= \sum_{e,a} P(B,j,m,e,a) && \text{marginal can be obtained from joint by summing out} \\
&= \sum_{e,a} P(B)P(e)P(a|B,e)P(j|a)P(m|a) && \text{use Bayes' net joint distribution expression} \\
&= \sum_{e} P(B)P(e)\sum_{a} P(a|B,e)P(j|a)P(m|a) && \text{use } x*(y+z) = xy + xz \\
&= \sum_{e} P(B)P(e)f_1(B,e,j,m) && \text{joining on a, and then summing out gives } f_1 \\
&= P(B)\sum_{e} P(e)f_1(B,e,j,m) && x*(y+z) = xy + xz \\
&= P(B)f_2(B,j,m) && \text{joining on e, and then summing out gives } f_2
\end{aligned}
$$

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

YAE – aka Yet Another 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 $f_1(Z, y_1) = \sum_{x_1} p(x_1|Z)p(y_1|x_1)$, and we are left with:

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

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

$$p(Z)f_1(Z, y_1)f_2(Z, y_2)p(X_3|Z)p(y_3|X_3)$$

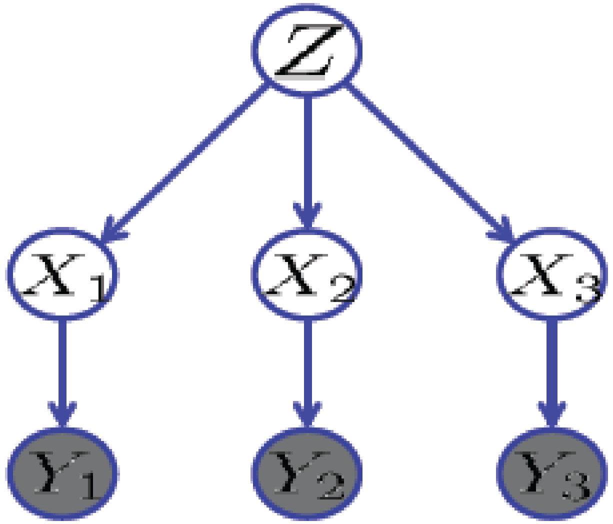
Eliminate Z, this introduces the factor $f_3(y_1, y_2, X_3) = \sum_z p(z)f_1(z, y_1)f_2(z, y_2)p(X_3|z)$, and we are left:

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

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

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

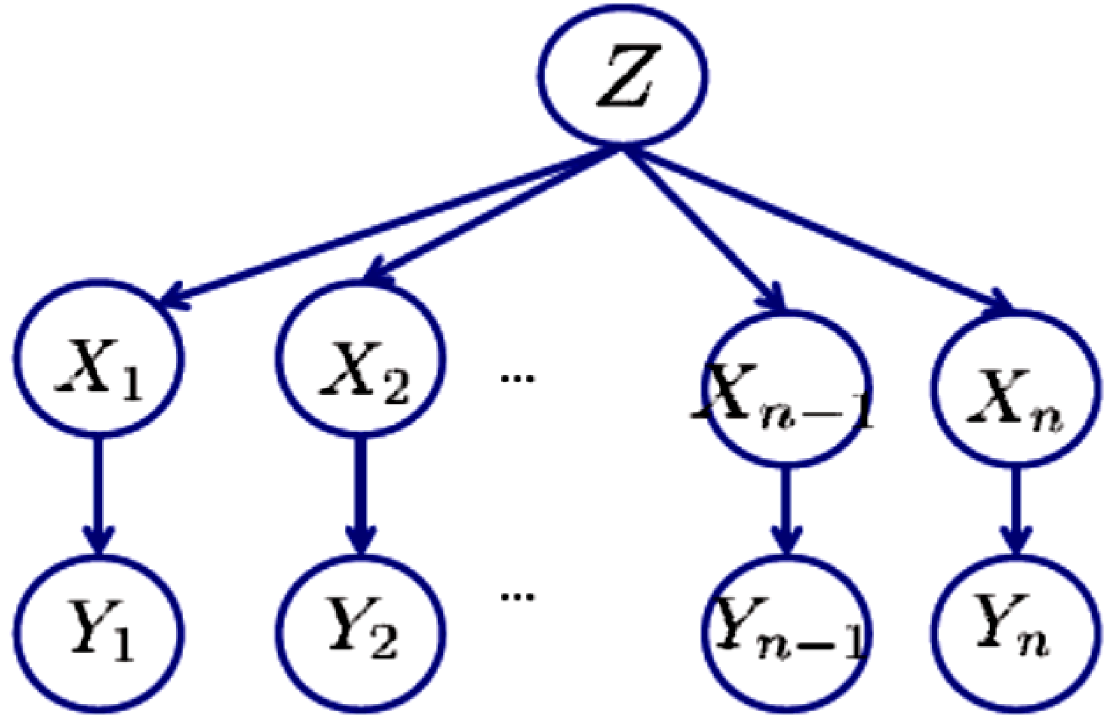
Normalizing over X_3 gives $P(X_3|y_1, y_2, y_3)$.



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 X_3 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.

Variable Elimination 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

- Best elimination ordering? NP-hard problem