CSC 665: Artificial Intelligence

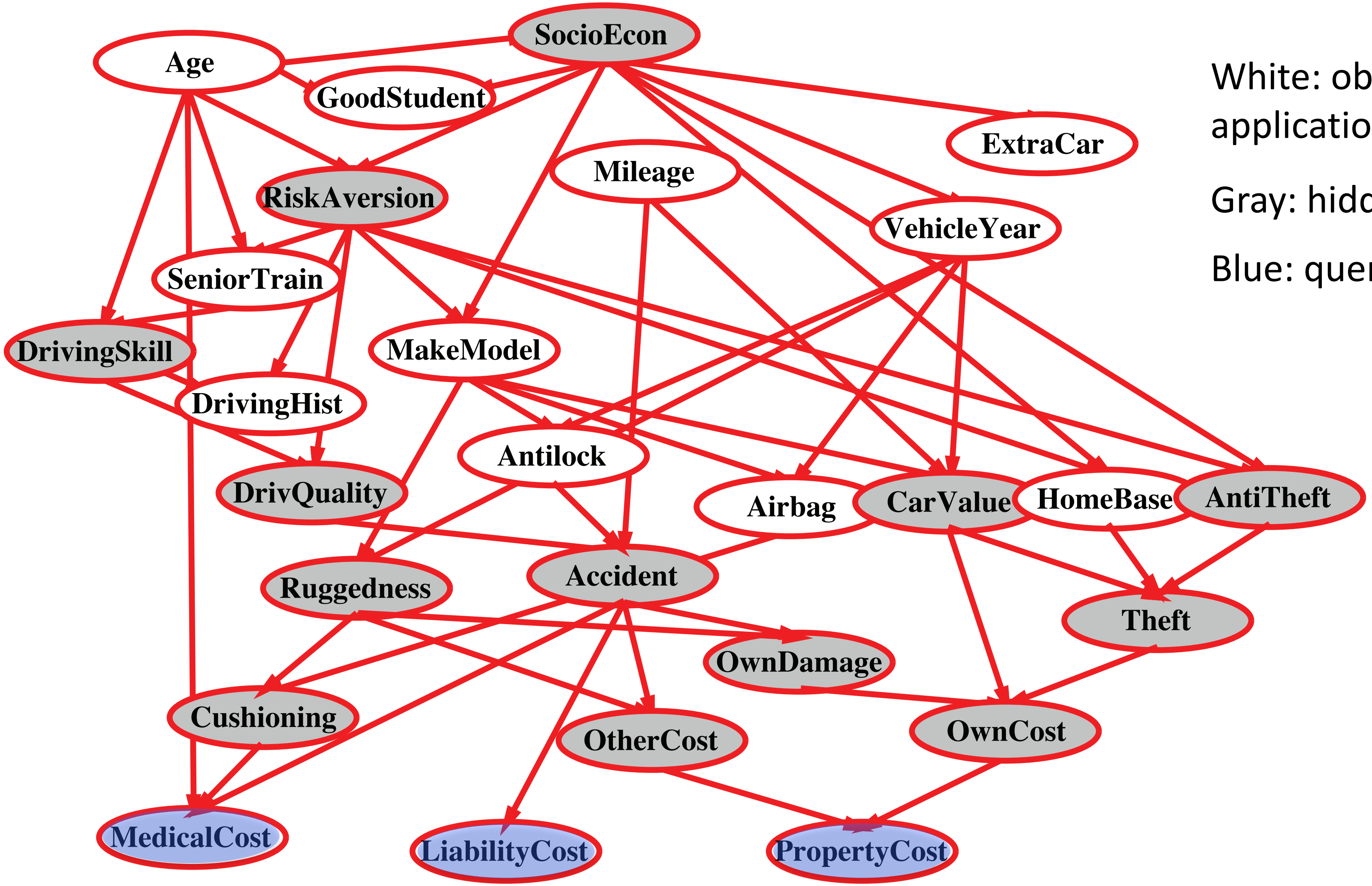
Bayes Nets: Inference

Instructor: Pooyan Fazli
San Francisco State University

Today

- Inference in Bayes Net
  - Inference by Enumeration
  - Variable Elimination

Example Bayes Net: Driver Insurance



White: observe on application form (evidence)

Gray: hidden variables

Blue: query variables

Inference

- Inference: Compute a desired probability from other known probabilities (e.g. conditional from joint)

- Example:
  - Compute conditional probability

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

  - from

    $$P(Q, H_1 \dots H_r, E_1 = e_1 \dots E_k = e_k)$$

Inference by Enumeration

Inference by Enumeration

- P(W)?

W	P(W)
sun	0.65
rain	0.35

- P(W | winter)?

W	P(W | winter)
sun	c
rain	d

- P(W | winter, hot)?

W	P(W | winter, hot)
sun	e
rain	f

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- P(W)?

W	P(W)
sun	0.65
rain	0.35

- P(W \| winter)?

W	P(W | winter)
sun	0.5
rain	0.5

- P(W \| winter, hot)?

W	P(W | winter, hot)
sun	e
rain	f

S	T	W	P
~~summer~~	~~hot~~	~~sun~~	~~0.30~~
~~summer~~	~~hot~~	~~rain~~	~~0.05~~
~~summer~~	~~cold~~	~~sun~~	~~0.10~~
~~summer~~	~~cold~~	~~rain~~	~~0.05~~
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- P(W)?

W	P(W)
sun	0.65
rain	0.35

- P(W | winter)?

W	P(W | winter)
sun	0.5
rain	0.5

- P(W | winter, hot)?

W	P(W | winter, hot)
sun	0.667
rain	0.333

S	T	W	P
~~summer~~	~~hot~~	~~sun~~	~~0.30~~
~~summer~~	~~hot~~	~~rain~~	~~0.05~~
~~summer~~	~~cold~~	~~sun~~	~~0.10~~
~~summer~~	~~cold~~	~~rain~~	~~0.05~~
winter	hot	sun	0.10
winter	hot	rain	0.05
~~winter~~	~~cold~~	~~sun~~	~~0.15~~
~~winter~~	~~cold~~	~~rain~~	~~0.20~~

Inference by Enumeration

- General case:

$$P(Q, H_1 \dots H_r, E_1 = e_1 \dots E_k = e_k)$$

- We want:

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

- Step 1: Sum out H to get joint of Query and evidence

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

- Step 2: Normalize

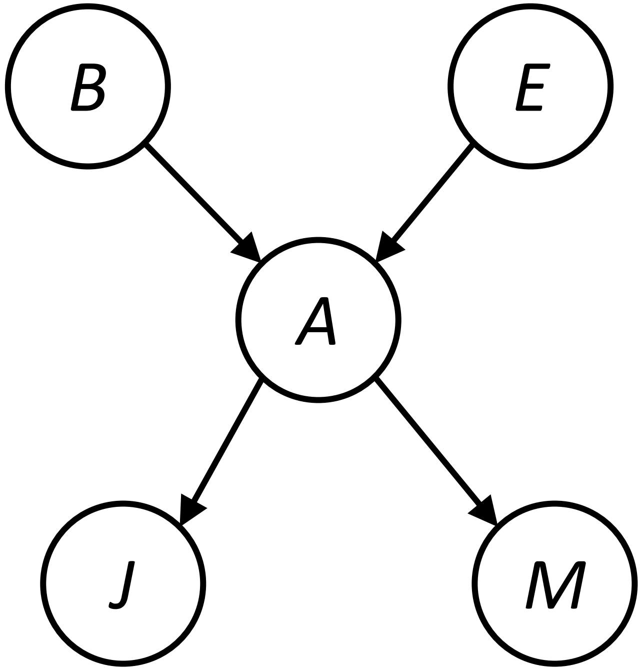
$$P(Q \mid e_1 \dots e_k) = \underbrace{\frac{P(Q, e_1 \dots e_k)}{P(e_1 \dots e_k)}}_{\text{Definition of conditional probability}} = \frac{P(Q, e_1 \dots e_k)}{\underbrace{\sum_q}_{\text{Marginalization over Q}} P(Q, e_1 \dots e_k)}$$

Definition of conditional probability

Marginalization over Q

All we need to compute is the joint probability of the query and the evidence!

Example



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

P(Bl+j, +m) = P(B, +j, +m) / sum(P(Bi, +j, +m))

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

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

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

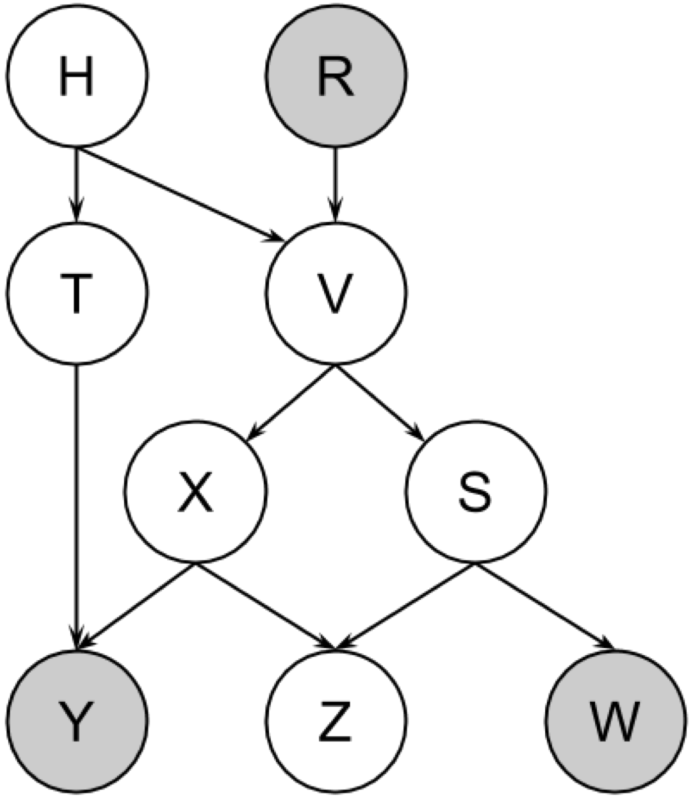
$$+\ P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$

Question

To compute the query P(X|r,w,y) in this Bayes Net, which variables are summed over in inference by enumeration?

Answer: H, S, T, V, Z

Query variables: X
Evidence: E, W, Y
Hidden variables: H, S, T, V, Z



Variable Elimination

Factors

Factors

$$P(T, W)$$

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

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

$$P(cold, W)$$

T	W	P
cold	sun	0.2
cold	rain	0.3

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

Factors

- 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

- P(X |Y)
  - Multiple conditionals
  - Entries P(x | y) for all x, y
  - Sums to |Y| <—2

$$P(W|T)$$

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

- 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

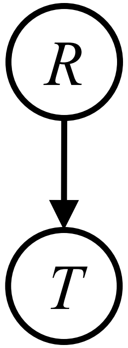
$P(rain|hot)$

$P(rain|cold)$

Operations on Factors

Operation 1: Join Factors

- First basic operation: joining factors
  - Just like a database join
  - Build a new factor over the union of the variables involved
- Example: Join on R



$P(R)$

+r	0.1
-r	0.9

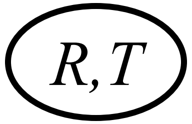
$\times$

$P(T|R)$

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

$P(R,T)$

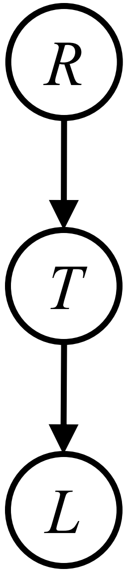
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81



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

Example: Multiple Joins

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

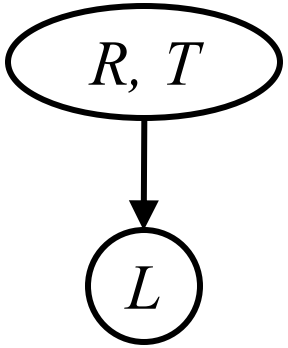
Join R

$P(R,T)$

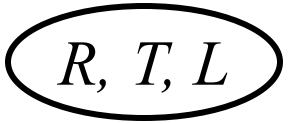
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

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



Join T



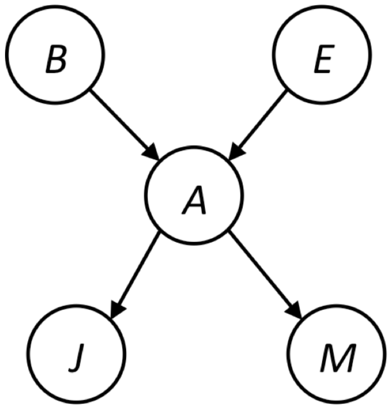
$P(R,T,L)$

+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

Example: Multiple Joins

If a variable appears in front of the conditioning bar in any of the factors participating in the join, it'll be in front of the conditioning bar in the resulting factor. Otherwise it'll end up behind the conditioning bar.

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



Operation 2: Eliminate

- Second basic operation: marginalization
- Take a factor and sum out a variable
  - Shrinks a factor to a smaller one
- 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

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

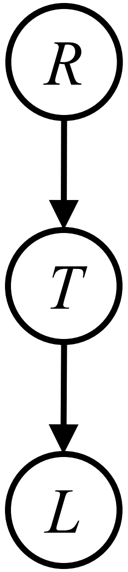
$P(L)$

+l	0.134
-l	0.886

Traffic Domain

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

$$P(L) = ?$$



$P(R)$

+r	0.1
-r	0.9

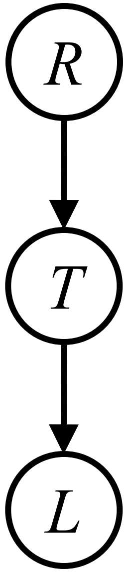
$P(T|R)$

+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 vs Variable Elimination



$$P(L) = ?$$

- 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

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

Inference by Enumeration vs Variable Elimination

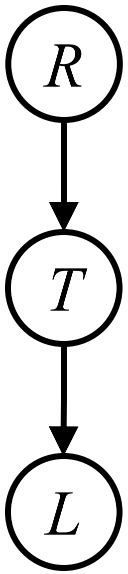
- Example: it takes 6 multiplications to evaluate the expression
  - ab + ac + ad + eh + fh + gh.

- How can this expression be evaluated more efficiently?
  - Factor out a and h giving a(b + c + d) + h(e + f + g)
  - This takes only 2 multiplications (same number of additions as above)

Variable Elimination

$P(R)$

+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

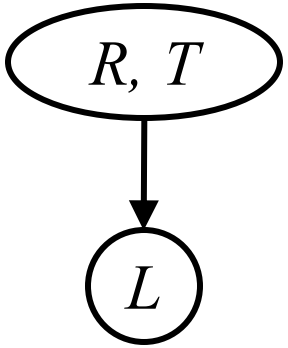
$P(L|T)$

+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



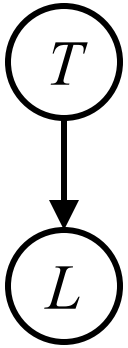
$P(L|T)$

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

Sum out R

$P(T)$

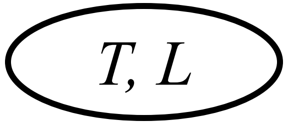
+t	0.17
-t	0.83



$P(L|T)$

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

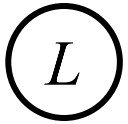
Join T



$P(T,L)$

+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

Evidence

- To compute $P(L)$, the initial factors are:

$P(R)$

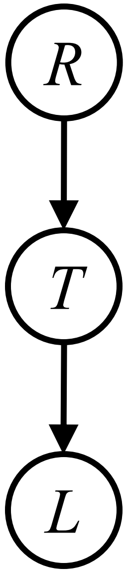
+r	0.1
-r	0.9

$P(T|R)$

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

$P(L|T)$

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



- To compute $P(L|+r)$, the initial factors will be:

$P(+r)$

+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

Evidence II

- Result will be a 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!

General Variable Elimination

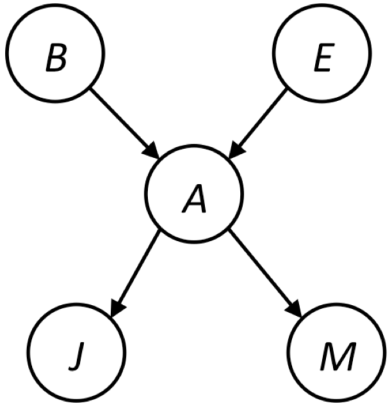
- Query: $P(Q|E_1 = e_1, \dots E_k = e_k)$

- 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

Variable Elimination: Example



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

Initial Factors: $\boxed{P(B) \quad P(E) \quad P(A|B,E) \quad P(j|A) \quad P(m|A)}$

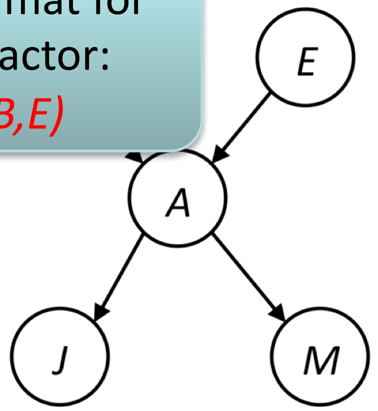
Choose A:

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

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

Another format for the new factor:
$f_1(j,m|B,E)$

Variable Elimination: Example



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

Another format for the new factor: *$f_1(j,m|B,E)$*

Choose E

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

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

Another format for the new factor: *$f_2(j,m|B)$*

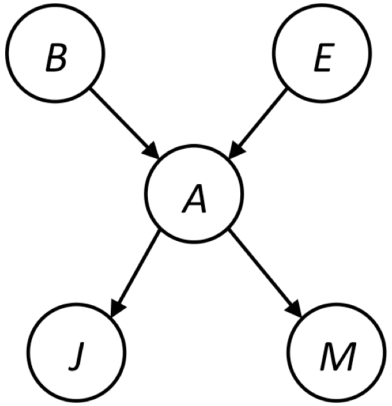
Finish with B

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

Same Example in Equations

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

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



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

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:

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

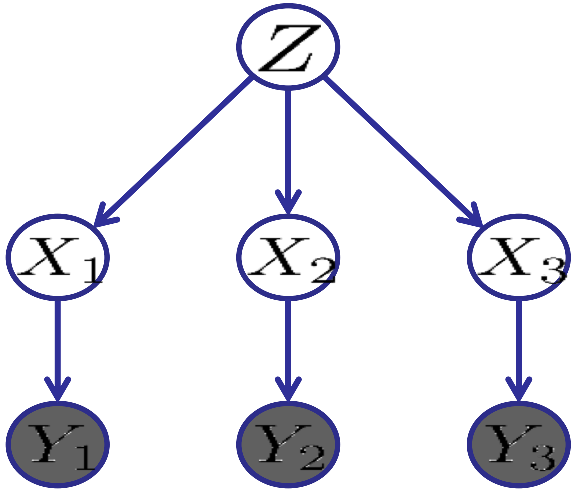
Eliminate Z, this introduces the factor $f_3(y_1, y_2, X_3) = \sum_z P(z)P(X_3|z)f_1(y_1|Z)f_2(y_2|Z)$, and we are left with:

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

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

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

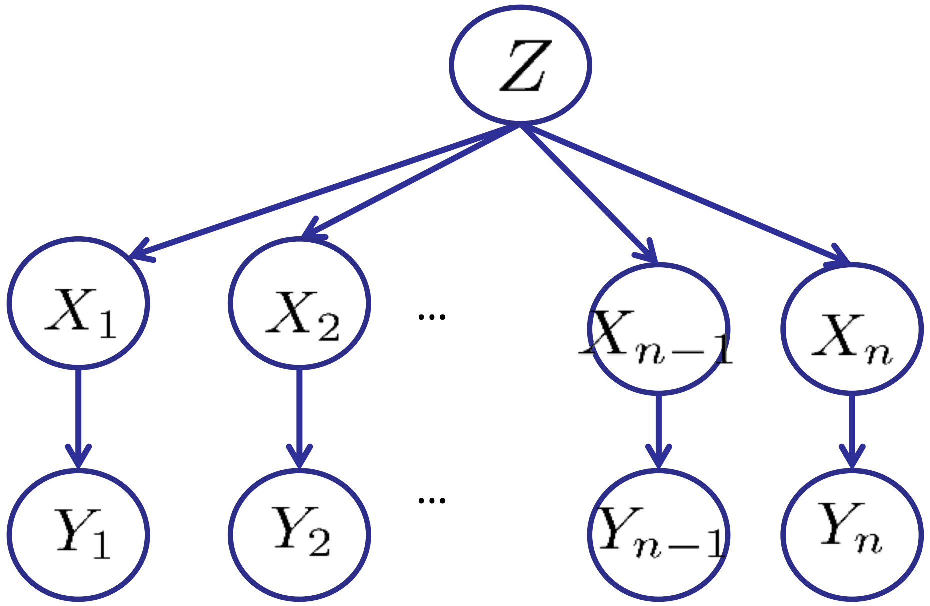
Normalizing over X_3 gives $P(X_3|y_1, y_2, y_3) = f_4(y_1, y_2, y_3, X_3) / \sum_{x_3} f_4(y_1, y_2, y_3, x_3)$



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 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 2^1 (assuming binary)

- In general: the ordering can greatly affect efficiency.

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!

Bayes Nets

- ✓ Representation
- ✓ Conditional Independences
- Probabilistic Inference
  - ✓ Enumeration (exact, exponential complexity)
  - ✓ Variable elimination (exact, worst-case exponential complexity, often better)
  - ✓ Inference is NP-complete

Reading

- Read Section 14.4 in the AIMA textbook (Third Edition)