Artificial Intelligence

CS4365 --- Fall 2022

Bayes' Net: Exact Inference

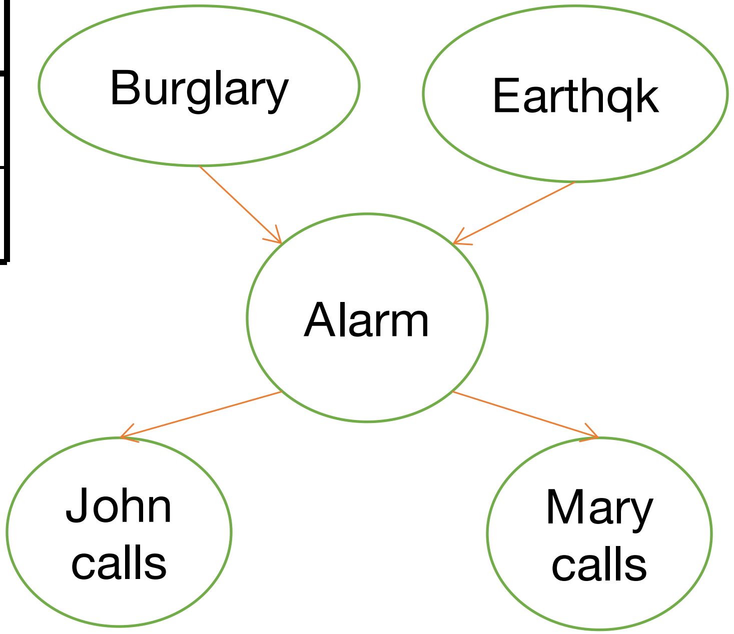
Instructor: Yunhui Guo

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

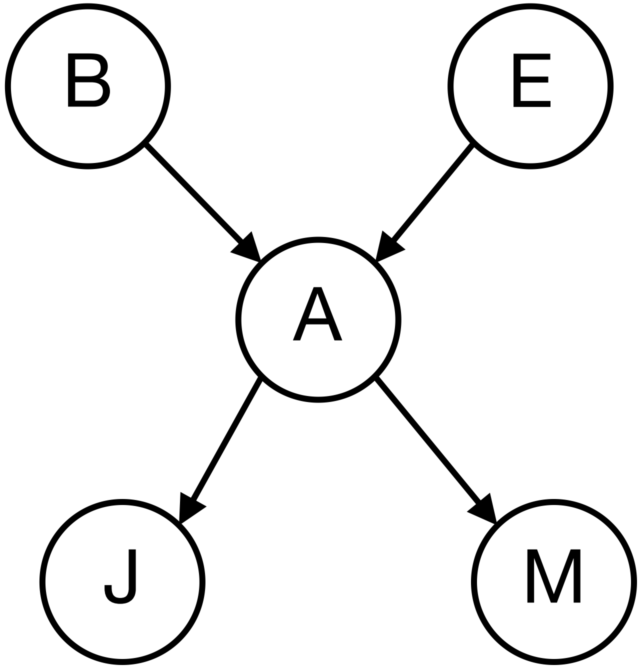
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

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

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

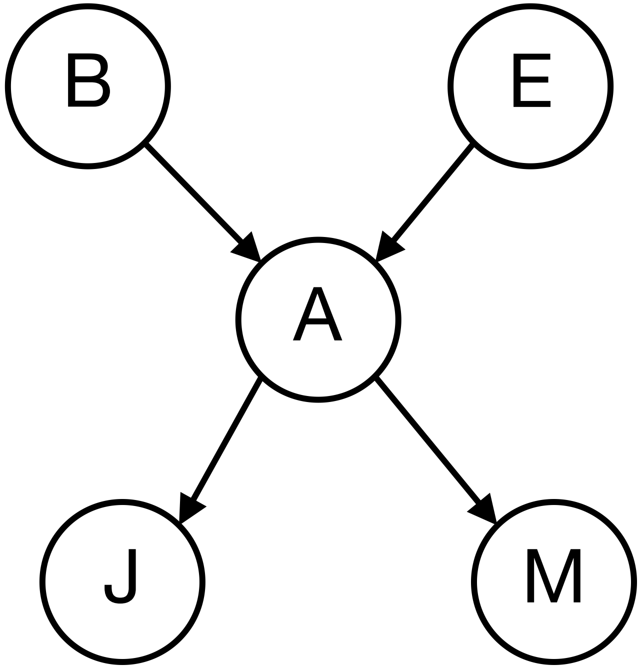
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

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

Inference

- Given a Bayesian network, what questions might we want to ask?

- **Conditional probability query**: $P(X = x \mid e)$
  - Given instantiations for some of the variables (we'll use e here to stand for the values of all the instantiated variables; it doesn't have to be just one), what is the probability that node X has a particular value x?
- **Maximum a posteriori probability**: $\text{argmax}_q \; P(Q = q | E_1 = e_1 \ldots)$

- **General question**: What's the whole probability distribution over variable X given evidence e, $P(X \mid e)$?

Inference

- Probabilistic inference required for P(outcome|action, evidence)
- Value of information: which evidence to seek next?
- Sensitivity analysis: which probability values are most critical?
- Explanation: why do I need a new starter motor?

Inference

- Inference: calculating some useful quantity from a joint probability distribution
- Exact inference:
  - Inference by enumeration
  - variable elimination
- Approximate inference:
  - Sampling

Inference by Enumeration

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

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

Using the Joint Distribution

- To answer any query involving a conjunction of variables, sum over the variables not involved in the query

- Marginalization:

$$P(y) = \Sigma_{ABC}\, P(a,b,c,y),$$

- $P(y|x) = P(x,y) \,/\, P(x)$

$$= \Sigma_{ABC}\, P(a,b,c,x,y) \,/\, \Sigma_{ABCY}\, P(a,b,c,x,y)$$

Inference by Enumeration

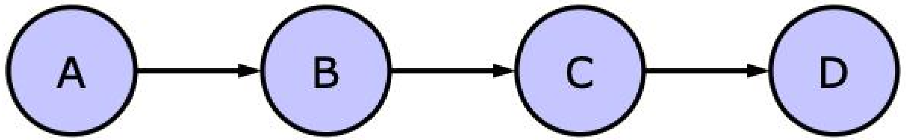
- Step 1:
  - Select the entries consistent with the evidence
- Step 2:
  - Sum out H to get joint of Query and evidence

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

- Step 3:
  - Normalize

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

Simple Case



- $P(d) = \Sigma_{ABC} P(a,b,c,d)$

$= \Sigma_{ABC} P(d|c)P(c|b)P(b|a)P(a)$

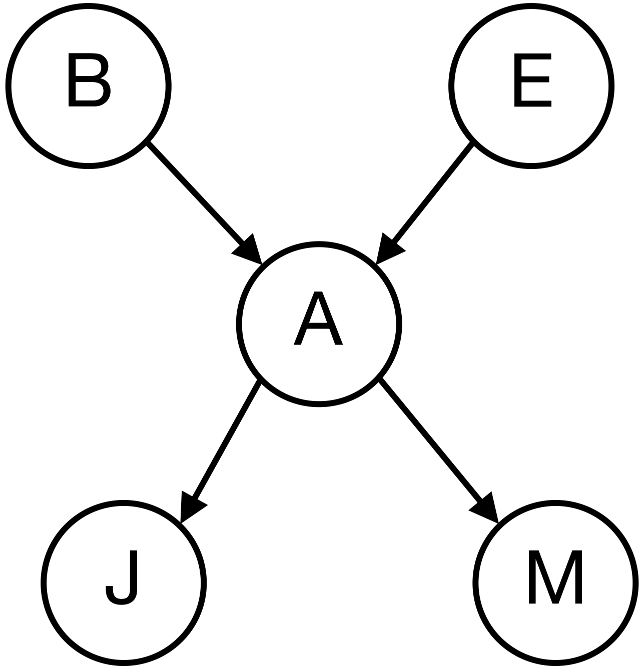
Only need local conditional distributions

$= \Sigma_A \Sigma_B \Sigma_C P(d|c)P(c|b)P(b|a)P(a)$

$= \Sigma_C P(d|c)\Sigma_B P(c|b)\Sigma_A P(b|a)P(a)$

Inference by Enumeration in Bayes' Net

- Given unlimited time, inference in BNs is easy
- Inference by enumeration:



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

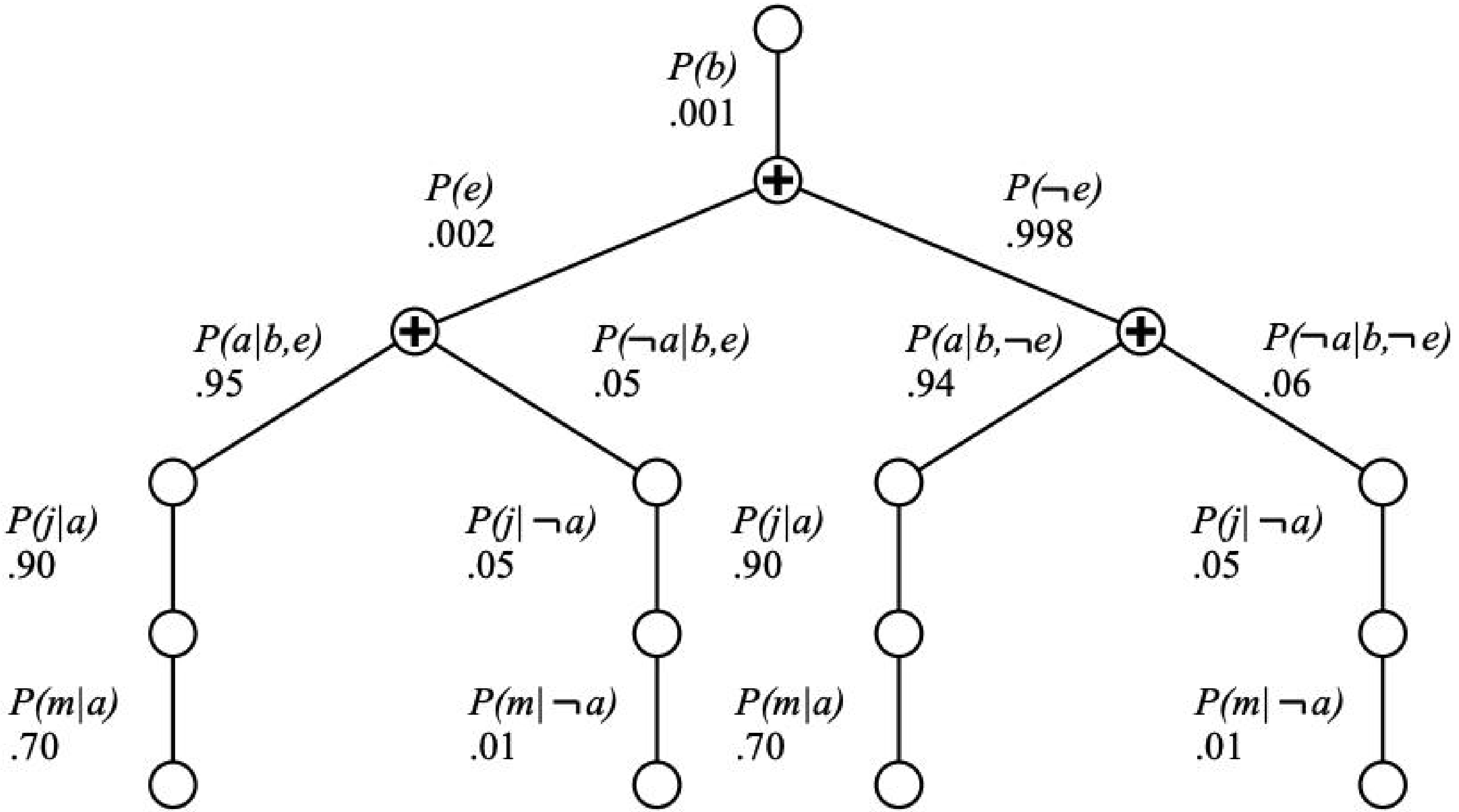
$$= \sum_{e,a} P(B, e, a, +j, +m)$$

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

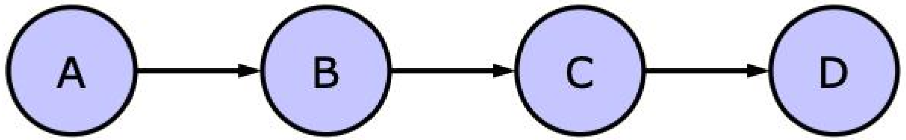
$$= P(B)P(+e)P(+a|B,+e)P(+j|+a)P(+m|+a) + P(B)P(+e)P(-a|B,+e)P(+j|-a)P(+m|-a)$$
$$P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$

Evaluation tree

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



Simple Case

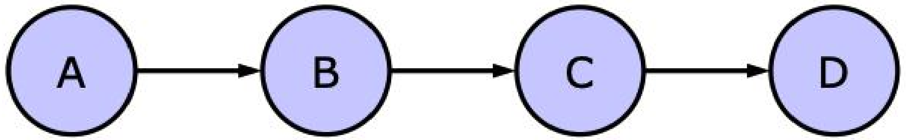


- $P(d) = \Sigma_A \Sigma_B \Sigma_C\, P(d|c)P(c|b)P(b|a)P(a)$

  $= \Sigma_C\, P(d|c)\Sigma_B P(c|b)\underbrace{\Sigma_A P(b|a)P(a)}$

$$\begin{array}{cc} P(b_1|a_1)P(a_1) & P(b_1|a_2)P(a_2) \\ P(b_2|a_1)P(a_1) & P(b_2|a_2)P(a_2) \end{array}$$

Simple Case



- $P(d) = \Sigma_A \Sigma_B \Sigma_C\, P(d|c)P(c|b)P(b|a)P(a)$

$$= \Sigma_C\, P(d|c)\Sigma_B P(c|b)\underbrace{\Sigma_A P(b|a)P(a)}$$

$$\Sigma_A\, P(b_1|a)P(a)$$

$$\Sigma_A\, P(b_2|a)P(a)$$

Factor

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

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

- Number of capitals = dimensionality of the table

$P(T, W)$

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

$P(cold, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

Factor

- 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

- 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

$P(W|hot)$

$P(W|cold)$

Factor

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

$$P(rain|T)$$

T	W	P
hot	rain	0.2
cold	rain	0.6

$P(rain|hot)$

$P(rain|cold)$

Factor

- In general, when we write $P(Y_1 \ldots Y_N \mid X_1 \ldots X_M)$
- It is a "factor," a multi-dimensional array
- Its values are $P(y_1 \ldots y_N \mid x_1 \ldots x_M)$
- Any assigned (=lower-case) X or Y is a dimension missing (selected) from the array

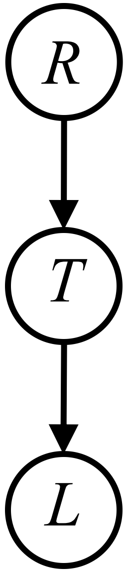
Example: Traffic Domain

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

- P(L) = ?

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

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



$P(R)$

+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 = +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

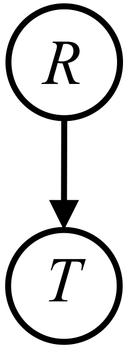
$P(+\ell|T)$

+t	+l	0.3
-t	+l	0.1

- Procedure: Join all factors, then eliminate 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



$P(R)$

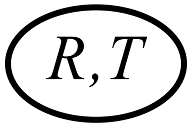
+r	0.1
-r	0.9

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

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

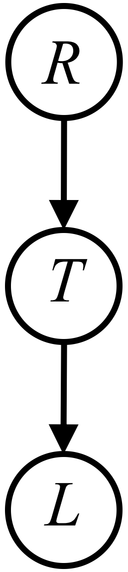
$P(R,T)$

+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



$P(R)$

+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

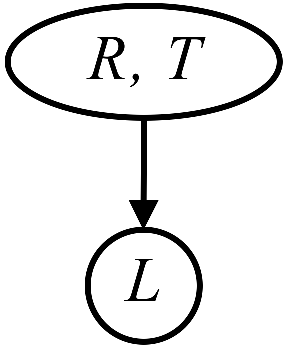
Join R

$P(R, T)$

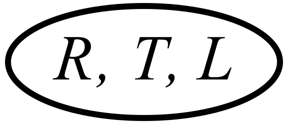
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

+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 $\Rightarrow$

$P(T)$

+t	0.17
-t	0.83

Multiple Elimination

R, T, L → Sum out R → *T, L* → Sum out 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

$P(T,L)$

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

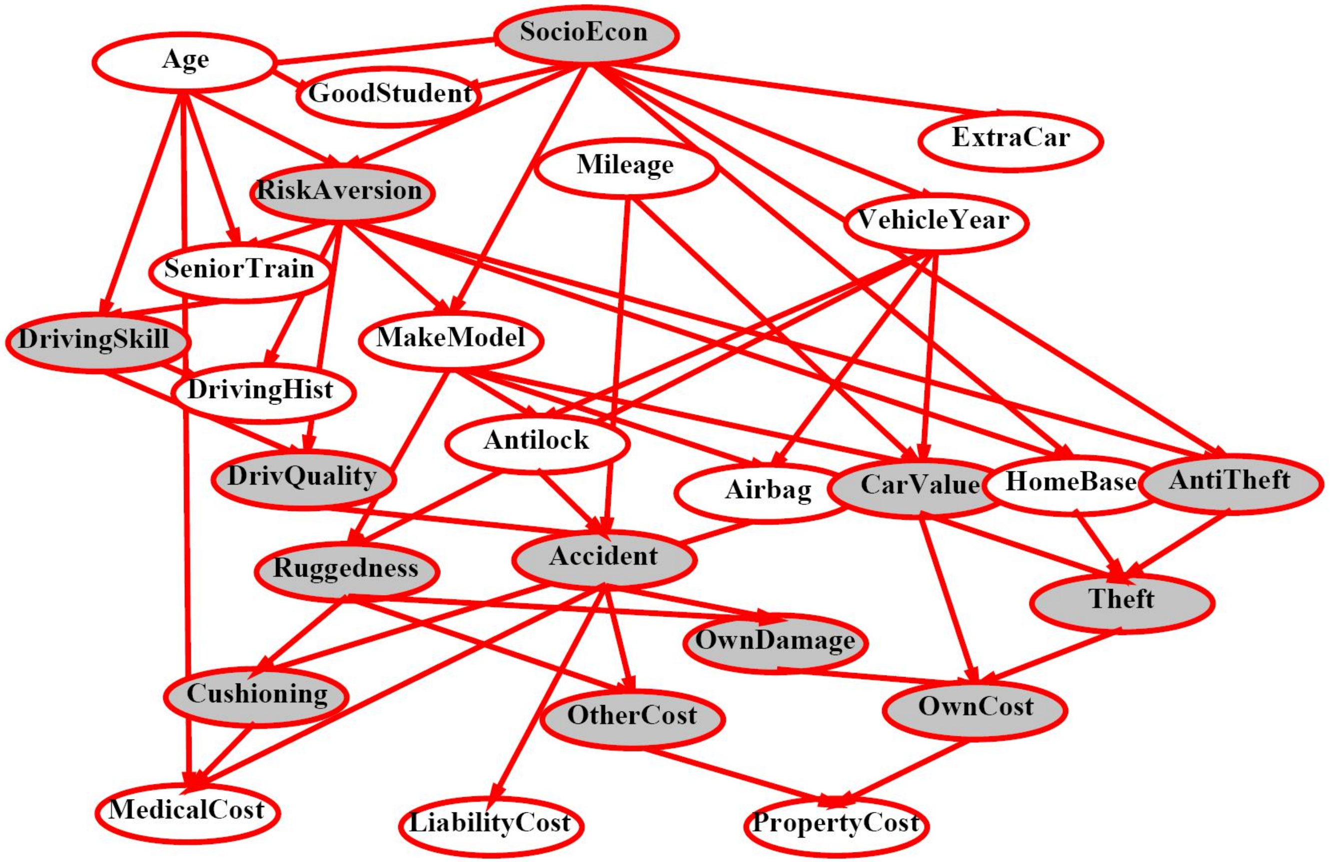
Sum out T

$P(L)$

+l	0.134
-l	0.886

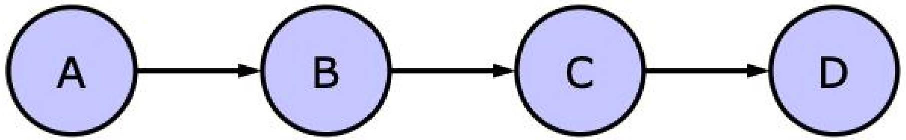
- Multiple Join, Multiple Eliminate (= Inference by Enumeration)

Inference by Enumeration?



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

Simple Case



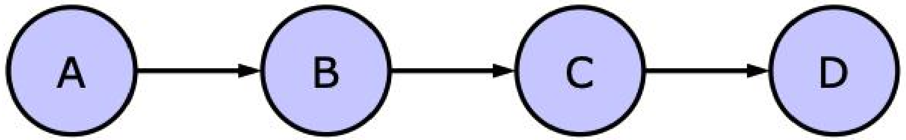
- $P(d) = \Sigma_A \Sigma_B \Sigma_C \, P(d|c)P(c|b)P(b|a)P(a)$

$$= \Sigma_C \, P(d|c)\Sigma_B P(c|b)\Sigma_A P(b|a)P(a)$$

$$\Sigma_A \, P(b_1|a)P(a)$$

$$\Sigma_A \, P(b_2|a)P(a)$$

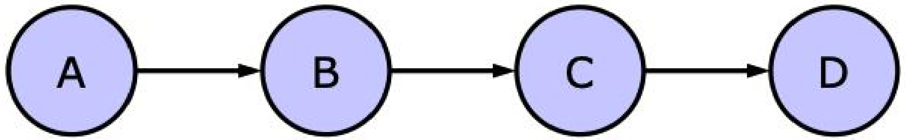
Simple Case



- $P(d) = \Sigma_A \Sigma_B \Sigma_C\ P(d|c)P(c|b)P(b|a)P(a)$

$$= \Sigma_C\ P(d|c)\Sigma_B P(c|b)\underbrace{\Sigma_A P(b|a)P(a)}_{f(b)}$$

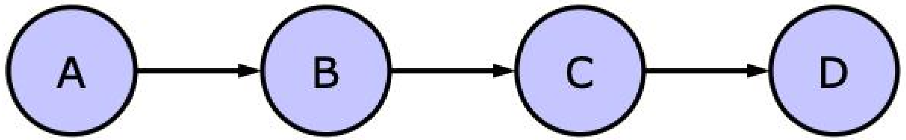
Simple Case



- $P(d) = \sum_A \sum_B \sum_C P(d|c)P(c|b)P(b|a)P(a)$

$$= \sum_C P(d|c)\underbrace{\sum_B P(c|b)f(b)}_{f(c)}$$

Simple Case



- $P(d) = \Sigma_A \Sigma_B \Sigma_C \, P(d|c)P(c|b)P(b|a)P(a)$

  $= \Sigma_C \, P(d|c)f(c)$

Variable Elimination

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

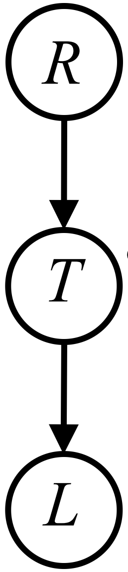
- Idea: interleave joining and marginalizing!
  - Called “Variable Elimination”
  - Still NP-hard, but usually much faster than inference by enumeration

Summary

- Multiple Join, Multiple Eliminate (= Inference by Enumeration)

- Marginalizing Early (= Variable Elimination)

Traffic Domain



$$P(L) = ?$$

- Inference by Enumeration

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

Eliminate r

Eliminate t

Variable Elimination

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

Eliminate t