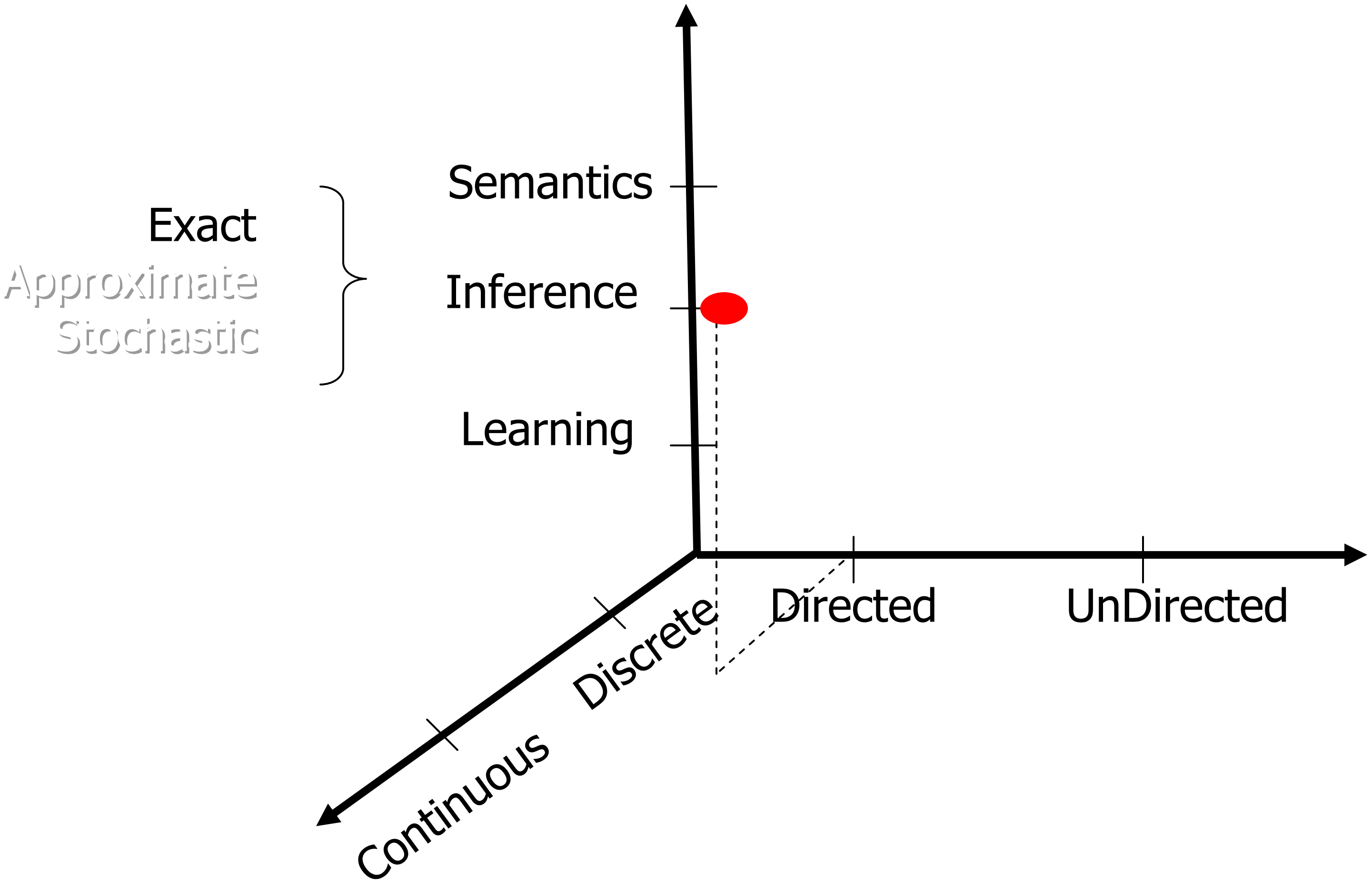
Belief Network Inference: Bucket Elimination

KF, Chapter 8
(RN, Chapter 14.4)

Some material taken from C Guesterin (CMU)

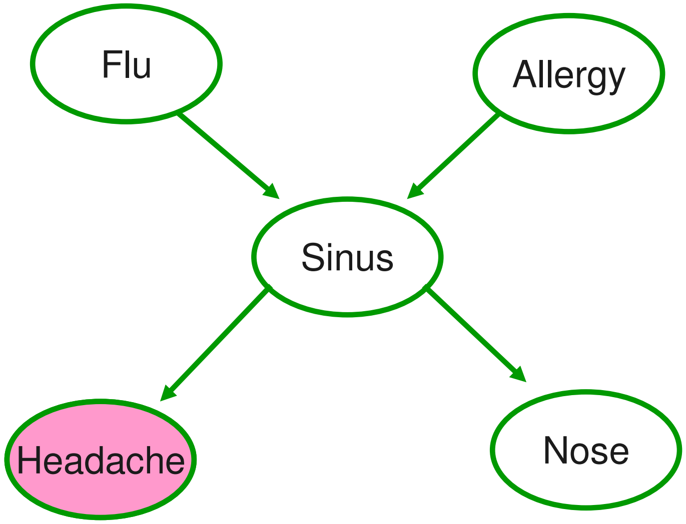
Space of Topics



Outline

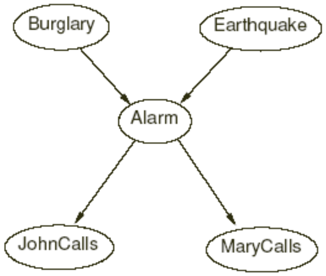
- Framework
- PolyTree algorithm
  - Message Propagation
- Complexity
- Bucket Elimination
- Other types of Inference tasks
  - Complexity

Inference in graphical models: Typical queries 1: Posterior



- Conditional probabilities
  - Distribution of some var(s). given evidence
  - Eg: Given "Headache = +", what is prob of Flu
  - P(+f | +h) ?

Inference in Belief Net

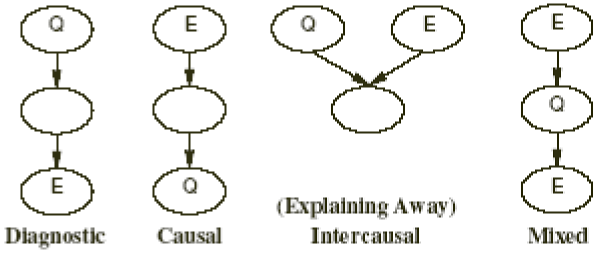


- Given: Fixed Belief Net

$$\mathcal{BN} = \left\{ \begin{array}{ll} \mathcal{N} & \text{Nodes} \equiv \text{Variables} \\ \mathcal{A} & \text{Arcs} \equiv \text{Dependencies} \\ \mathcal{C} & \text{CPtable} \equiv \text{``weights''} \end{array} \right\}$$

- Belief Assessment: Compute P(H | **E**)
  - H is single node (eg, *Meningitis*)
  - **E** is set of evidence nodes
    BloodTest = +, Jaundice = –, Smoker = –, Temp = 99.8, ...
- Other Tasks:
  - MPE (Most Probable Explanation)
  - MAP (Maximum a Posteriori)

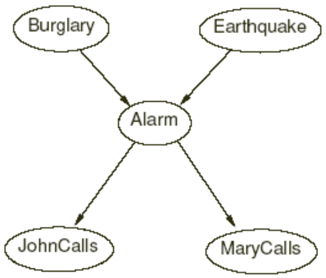
Types of Reasoning



- **Typical case:** P(QueryVar | EvidenceVars = vals)
  - Eg: P(+Burglary | +JohnCalls, ¬MaryCalls)
- **Diagnostic**: from effect to (possible) causes
  - P(+Burglary | +JohnCalls) = 0.016
- **Causal**: from cause to effects
  - P(+JohnCalls | +Burglary) = 0.86
- **InterCausal**: between causes of common effect
  - P(+Burglary | + Alarm) = 0.376
  - P(+Burglary | +Alarm, +Earthquake) = 0.003

  Earthquake EXPLAINS alarms, and so

  Earthquake **EXPLAINS AWAY** burglary



- **Mixed**: combinations of . . .
  - P(Alarm | JohnCall, ¬Earthquake) = 0.03

Approaches to Belief Assessment

- Exact, Guaranteed
  - PolyTree Algorithm
  - Inherent complexity. . .
  - Clustering Approach
  - Bucket Elimination
  - CutSet Approach

- Approximate, Guaranteed
  - Algorithm Modification
  - Value Merging
  - Node Merging
  - Arc Removal

- Stochastic (Approximate)
  - Logic Sampling
  - Likelihood Sampling

PolyTree Framework

See Dynamic Bayesian Nets!

- *Def'n:* BN is PolyTree *iff*
  ∃ ≤ 1 (undirected) path between any pair of nodes X, Y ∈ N
  ("Singly Connected")

⇒ Linear Time algorithm!

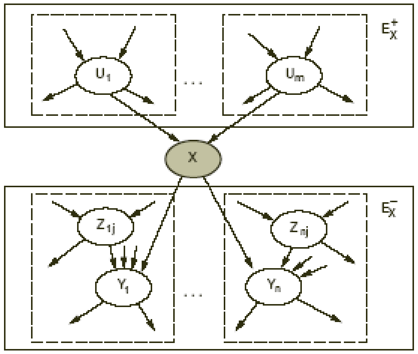
Idea: Write $P(X \mid E) = \ldots P(X \mid E_X^+)\, P(E_X^- \mid X)$

where $\left\{ \begin{array}{c} E_X^+ \\ E_X^- \end{array} \right\}$ is evidence $\left\{ \begin{array}{c} \textit{above} \\ \textit{below} \end{array} \right\}$ X

Then

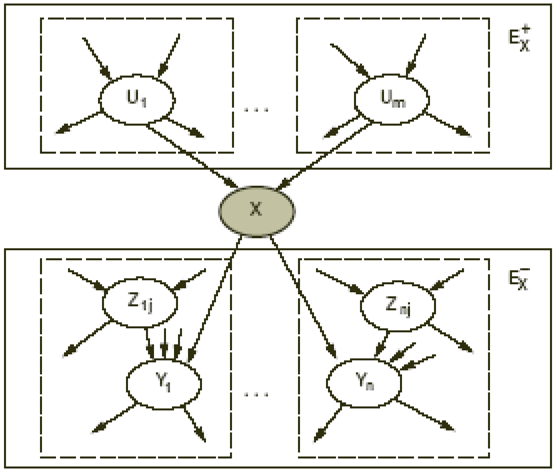
- compute $P(X \mid E_X^+)$ as simple function of
  $P(U \mid E_U^+),\ P(E_U^- \mid U)$
  for parents U
- compute $P(E_X^- \mid X)$ as simple function of
  $P(Y \mid E_Y^+),\ P(E_Y^- \mid Y)$
  for children, co-parents Y

("One pass up, one pass down")



PolyTree Inference

- Consider node X
  parents: $U_1, \ldots, U_m$
  children: $Y_1, \ldots, Y_n$
  co-parents: $Z_{11}, \ldots, Z_{nj}$

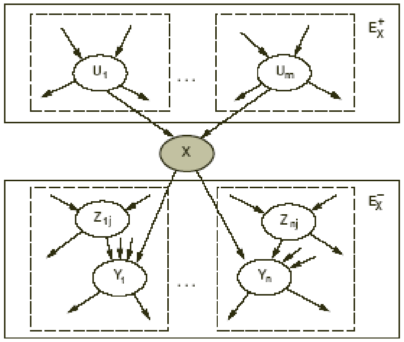


Evidence $\left\{ \begin{array}{c} E_X^+ \\ E_X^- \end{array} \right\}$ is $\left\{ \begin{array}{c} \textit{above} \\ \textit{below} \end{array} \right\}$ X

Write:
$$
\begin{aligned}
P(X \mid E) &= P(X \mid E_X^+, E_X^-) \\
&= \frac{P(E_X^- \mid X, E_X^+) P(X \mid E_X^+)}{P(E_X^- \mid E_X^+)} \\
&= \alpha P(E_X^- \mid X) P(X \mid E_X^+)
\end{aligned}
$$

where $\alpha = \frac{1}{P(E_X^- \mid E_X^+)}$ is normalizing constant

PolyTree Inference (con't)



- $P(X \mid E) = \alpha P(E_X^- \mid X) P(X \mid E_X^+)$
- $P(X \mid E_X^+) = \sum_{\mathbf{u}} P(X \mid \mathbf{u}) \prod_i P(U_i \mid E_{u_i/X})$

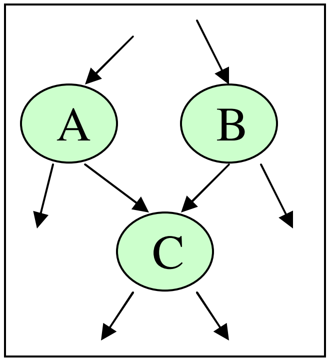
Note: $P(X \mid \mathbf{u}) =$ CPtable
$E_{u_i/X}$ is evidence connected to node U_i
EXCEPT via path from X
$P(U_i \mid E_{u_i/X})$ is recursive call

- $P(E_X^- \mid X) = \beta \prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i) \sum_{\mathbf{z}_i} P(y_i \mid X, \mathbf{z}_i) \prod_j P(z_{ij} \mid E_{Z_{ij}/Y_i})$

Note: β is normalizing constant
$P(y_i \mid X, \mathbf{z}_i) =$ CPTable
$P(E_{Y_i}^- \mid y_i)$ is recursive call of $P(E_X^- \mid X)$
$P(z_{ij} \mid E_{Z_{ij}/Y_i})$ is recursive call of $P(X \mid E_{X/})$

- Can also be used to compute $P(X_1, \ldots \mid E)$ in linear time.

Inference not "Detachable"



- Reasoning is "detached" if, given can compute

  P(C | **E**) = f(α, β)

  from α = P(A | **E**) β = P(B | **E**)
- True IF A, B independent (ie, if poly-tree)
- But FALSE in general...

  if uninstantiated (undirected) cycle !

⇒ Need to preserve "justifications" for

  P(B | **E**) P(A | **E**)

  (to see if they interact)
- Approaches:
  - Remove Cycle (clustering)
  - Remove "uninstantiated" cycle (cutset)

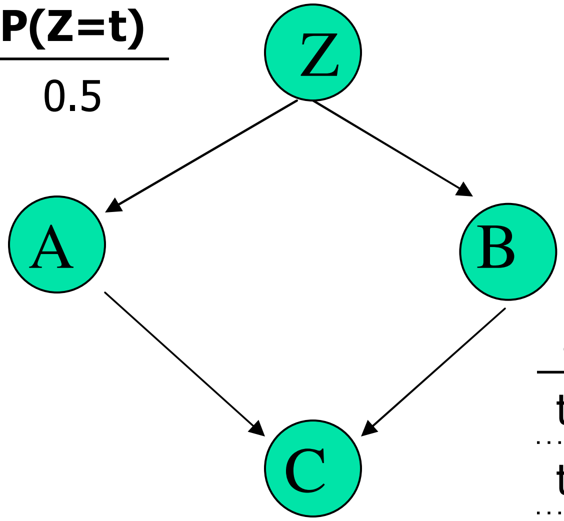
Why Reasoning is Hard

- BN reasoning may look easy:

  Just "propagate" information from node to node

P(Z=t)
0.5

z	P(A=t|Z=z)
t	1.0
f	0.0



z	P(B=t|Z=z)
t	0.0
f	1.0

a	b	P(C=t|a,b)
t	t	1.0
t	f	0.0
f	t	0.0
f	f	0.0

- Challenge: What is $P(C=t)$?

$A = Z = \neg B \quad P(A = t) = P(B=t) = \ldots = ½$

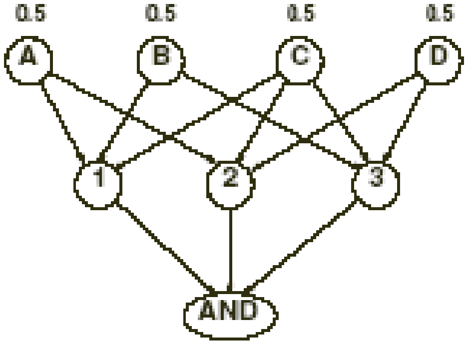
So… ? $P(C= t) = P(A = t, B = t)$

$= P(A = t) \times P(B = t) = ½ \times ½ = ¼$

- Wrong: $P(C = t) = 0$!

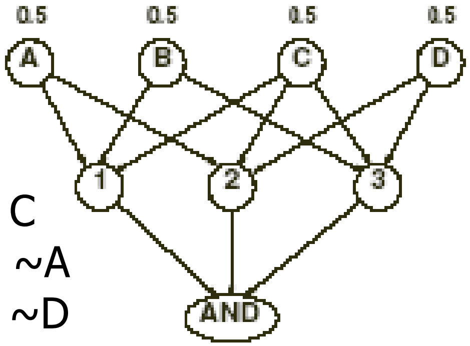
  Need to maintain dependencies! $P (A = t, B = t) = P (A = t) \times P (B = t \mid A = t)$

Encoding 3-SAT



1. A v B v C
2. C v D v ~A
3. B v C v ~D

Inherent Complexity



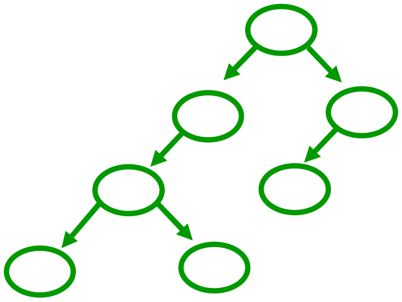
1. A v B v C
2. C v D v ~A
3. B v C v ~D

- Worst case:
  - NP-hard to get exact answer (#P-complete)
  - NP-hard to get answer within 0.5
  - Cannot get relative error within $2^{n^{1-\varepsilon}}$ unless P = NP
  - Cannot stochastically approximate 1-bit, unless P=RP
- Efficient algorithm . . .
  - for "PolyTree": Poly time
    - ≤ 1 path between any two nodes
  - if CPtable "bounded" (sub-exp time) wrt $\lambda = M/m$
    M = largest CPtable entry; m = smallest

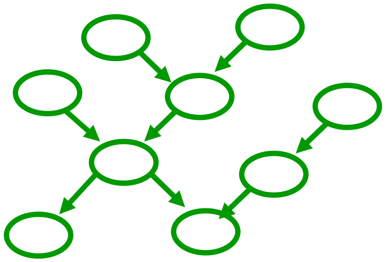
Inference is #P-complete… hopeless?

- In general: yes, hopeless!
  - Even approximate!
- In practice
  - Exploit structure
  - (Certain) Easy structure ⇒ simple inference
  - Many effective approximation algorithms
    - some with guarantees

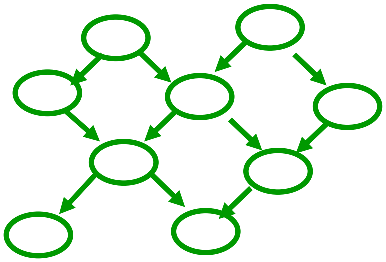
Trees



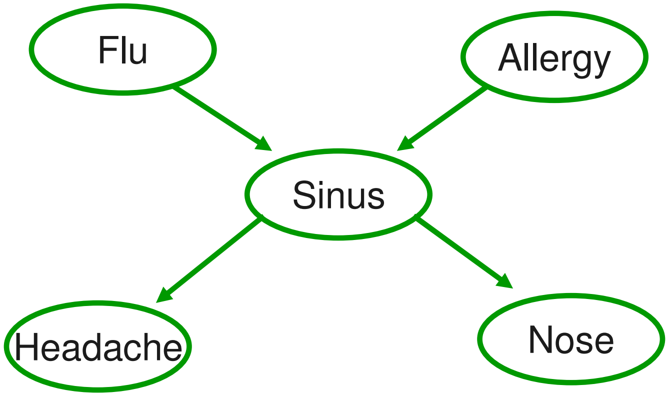
Poly-Trees



Low-treewidth Graphs



General probabilistic inference



- Query: P(X | **e**)

- Use cond. prob.:

$$P(X \mid e) = \frac{P(X, e)}{P(e)}$$

← Normalization constant

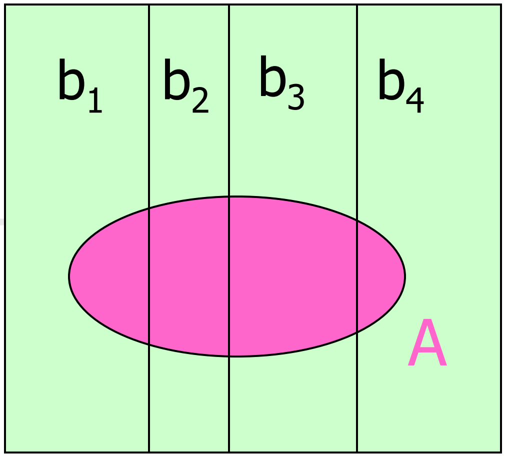
- Normalization:

P(X|**e**) $\propto$ P(X, **e**)

P(X=t, **e**) = 0.4
P(X=f, **e**) = 0.1

Normalize:
P(X=t|**e**) = 0.4/(0.4+0.1) = 0.8

Factoids



- $P(A = a) = \sum_b P(A = a, B = b)$

$$P(B=b \mid A=+) = \frac{P(B=b, A=+)}{P(A=+)}$$

$$= \frac{P(B=b, A=+)}{\sum_{b'} P(B=b', A=+)} = \frac{P(B=b, A=+)}{P(B=+, A=+) + P(B=-, A=+)}$$

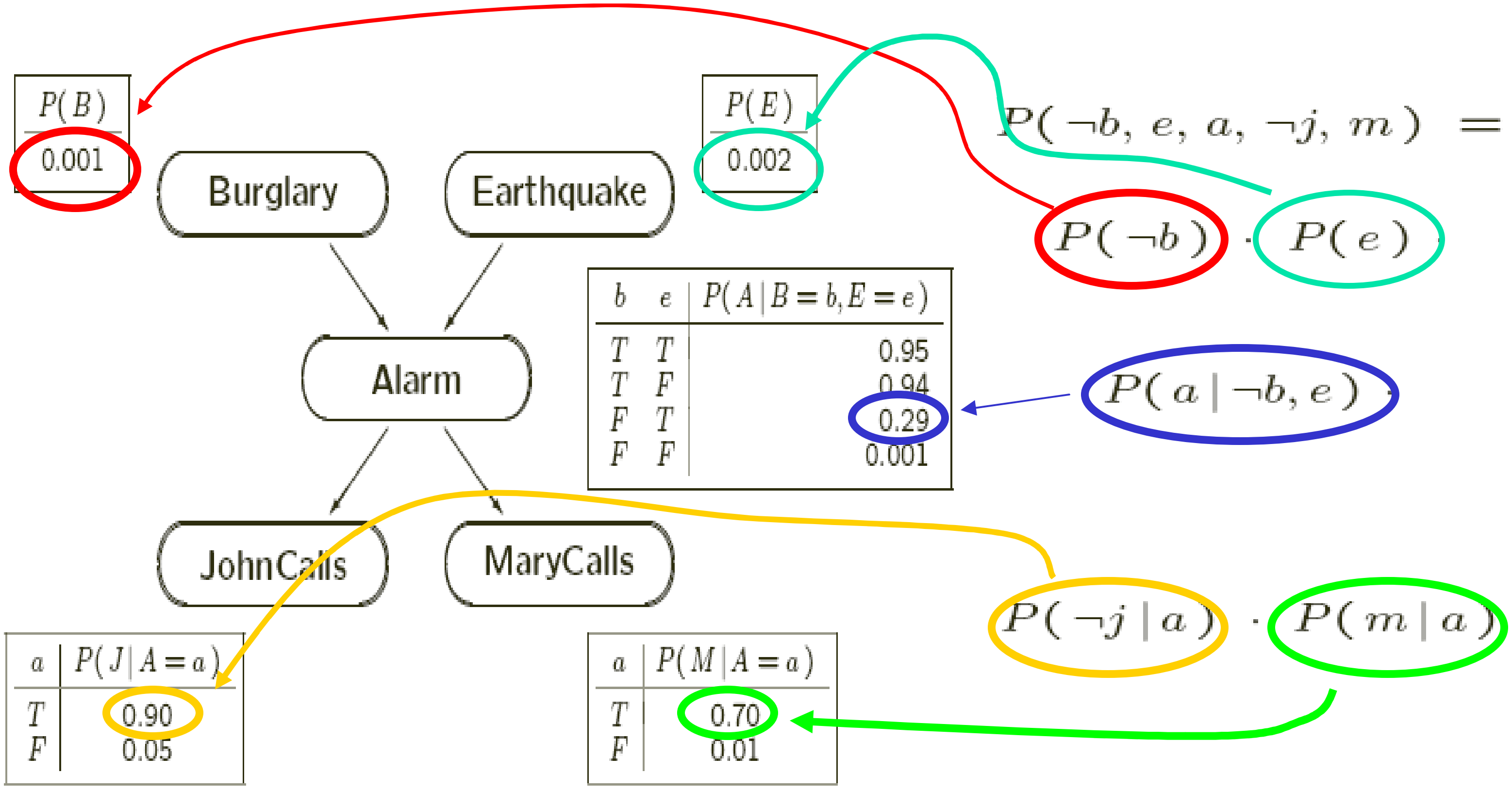
$$\Rightarrow \text{Just need } \left\{ \begin{matrix} P(B=+,\ A=+) \\ P(B=-,\ A=+) \end{matrix} \right\}$$

ie, only need UNCONDITIONAL queries

Recovering Joint

$$\begin{array}{lllll} P(\neg b,\, e,\, a,\, \neg j,\, m) \quad = & & & & \\ P(\neg b) & P(e \mid \neg b) & P(a \mid e, \neg b) & P(\neg j \mid a, e, \neg b) & P(m \mid \neg j, a, e, \neg b) \\ P(\neg b) & P(e) & P(a \mid e, \neg b) & P(\neg j \mid a) & P(m \mid a) \\ 0.99 \times & 0.02 \times & 0.29 \times & 0.1 \times & 0.70 \end{array}$$

Node independent of predecessors, given parents



$P(B)$
0.001

$P(E)$
0.002

b	e	$P(A \mid B=b, E=e)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

a	$P(J \mid A=a)$
T	0.90
F	0.05

a	$P(M \mid A=a)$
T	0.70
F	0.01

$$P(\neg b,\, e,\, a,\, \neg j,\, m) \; =$$

$$P(\neg b) \cdot P(e)$$

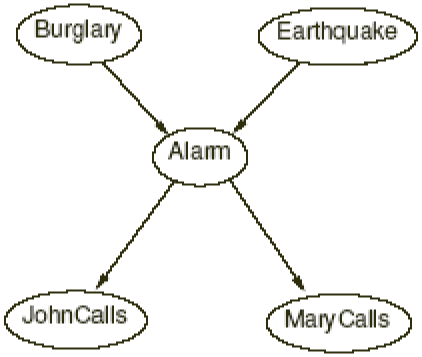
$$P(a \mid \neg b, e)$$

$$P(\neg j \mid a) \cdot P(m \mid a)$$

Exact Inference: Re-arrange Sums

$P(A = a) = \sum_b P(A = a, B = b)$

$P(+B, +J, +M)$



$$= \sum_e \sum_a P(+B, E = e, A = a, +J, +M)$$

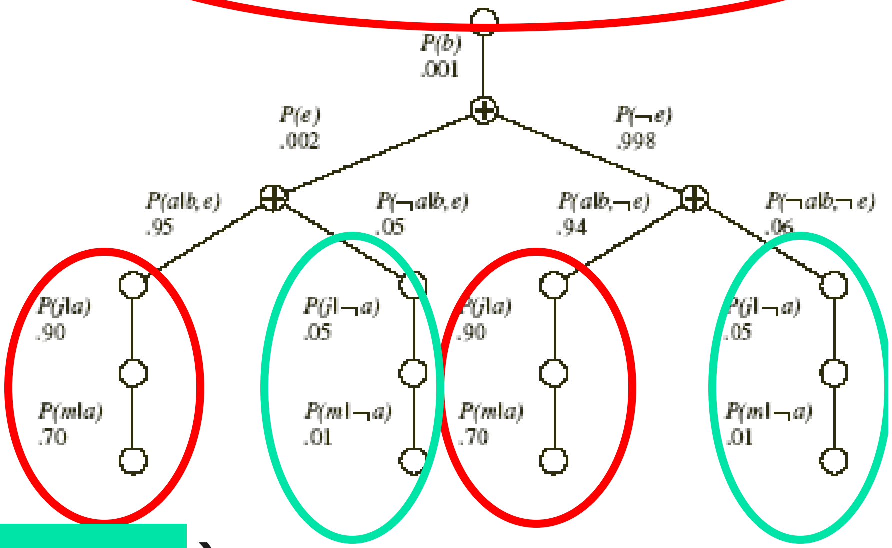
$$= \sum_e \sum_a P(+B) P(e) P(a \mid +B, e) P(+J \mid a) P(+M \mid a)$$

$$= P(+B) \sum_e P(e) \sum_a P(a \mid +B, e) P(+J \mid a) P(+M \mid a)$$

Still Duplicated Computation!

$P(+b, +j, +m) =$

$P(+b) \sum_e P(e) \sum_a P(a \mid +b, e) P(+j \mid a) P(+m \mid a)$



- Enumeration is *inefficient*:

  ... as repeated computation

  Computes $P(+j \mid -a) P(+m \mid -a)$)

  for each value of E: { +e, –e }
- Better to have DAG…

  re-use COMMON SUBEXPRESSION !

BucketElimation Algorithm

- Notation
  - Table
  - Join
  - Elim
  - Projection
- **Bucket-Elimination** Algorithm
  - Initialization
  - Iteration

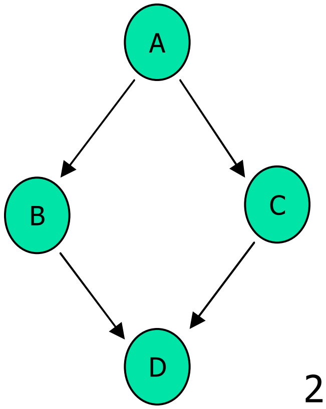
Table

- Table: $f(a,b) = \lambda \langle A,B \rangle.$

a	b	f(A=a, B=b)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09

- λ **S** . M
  - maps NAMED variables **S** = $\langle$vars$\rangle$ to $\Re$ based on table **M**
  - (λ **S** . **M**)[1, 0] = 0.70
- Like CPtable for node…

a	$\theta_{B=1\|A=a}$	$\theta_{B=0\|A=a}$
1	0.30	0. 70
0	0.91	0.09



Simple Belief Net

$f_A(a) = \lambda \langle A \rangle.$

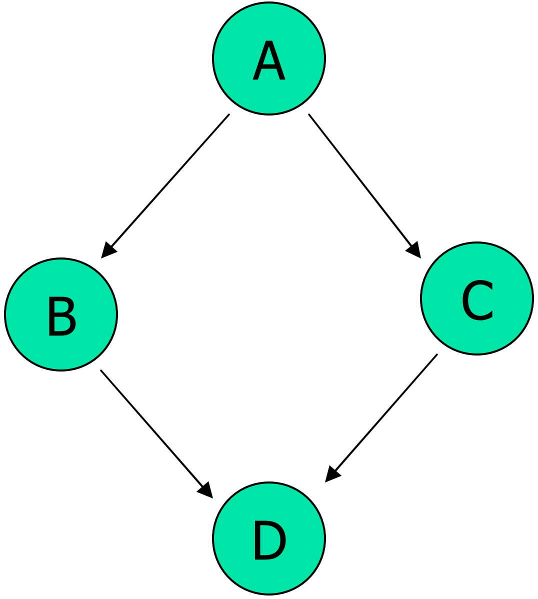
a	f(a)
1	0.30
1	0.70

$f_C(a,c) = \lambda \langle A,C \rangle.$

a	c	f(a, c)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09

$f_B(a,b) = \lambda \langle A,B \rangle.$

a	b	f(a,b)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09



$f_D(b,c,d) = \lambda \langle B,C,D \rangle.$

b	c	d	f(b, c,d)
1	1	1	0.30
1	0	0	0.70
…	.	.	…
0	0	1	0.09

Join operator: $\bowtie$

$f(a,b) = \lambda \langle A,B \rangle.$

a	b	f(A=a, B=b)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09

$g(a,c) = \lambda \langle A,C \rangle.$

a	c	g(A=a, C=c)
1	1	0.22
1	0	0.78
0	1	0.99
0	0	0.01

$$h(\mathbf{x}) = (f \bowtie g)(\mathbf{x})$$
$$= f(\mathbf{x}|_{\mathrm{Vars}(f)}) \times g(\mathbf{x}|_{\mathrm{Vars}(g)})$$

$h(a,b,c) = \lambda \langle A,B,C \rangle.$

a	b	c	h(A=a, B=b,C=c)
1	1	1	0.30 x 0.22
1	1	0	0.30 x 0.78
1	0	1	0.70 x 0.22
1	0	0	0.70 x 0.78
0	1	1	0.91 x 0.99
0	1	0	0.91 x 0.01
0	0	1	0.09 x 0.99
0	0	0	0.09 x 0.01

Join operator: ⋈

$f(a,b) = \lambda \langle A,B \rangle.$

a	b	f(A=a, B=b)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09

$g(a,c) = \lambda \langle A,C \rangle.$

a	c	g(A=a, C=c)
1	1	0.22
1	0	0.78
0	1	0.99
0	0	0.01

$$h(\mathbf{x}) = (f \bowtie g)(\mathbf{x})$$
$$= f(\mathbf{x}|_{\mathrm{Vars}(f)}) \times g(\mathbf{x}|_{\mathrm{Vars}(g)})$$

$h(a,b,c) = \lambda \langle A,B,C \rangle.$

a	b	c	h(A=a, B=b,C=c)
1	1	1	0.30 x 0.22
1	1	0	0.30 x 0.78
1	0	1	0.70 x 0.22
1	0	0	0.70 x 0.78
0	1	1	0.91 x 0.99
0	1	0	0.91 x 0.01
0	0	1	0.09 x 0.99
0	0	0	0.09 x 0.01

Elim (Elimination)

$f(a,b) = \lambda \langle A,B \rangle.$

a	b	f(A=a, B=b)
1	1	0.30
1	0	0.70
0	1	0.91
0	0	0.09

$\text{elim}_A[f]\ (b) = \lambda \langle B \rangle.$

b	$f_{-A}(\ B=b)$
1	0.30+0.91
0	0.70+0.09

$$\boxed{\text{elim}_X[f](\ \underline{y}\ ) = \sum_X f(\underline{Y}=\underline{y},\ x)}$$

Projection

for C= 0

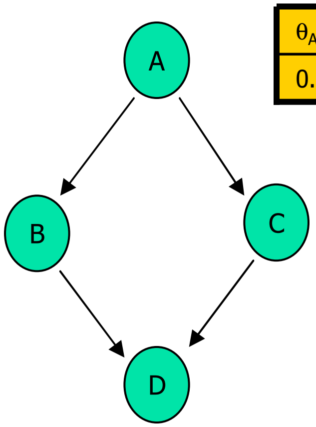
$h_{ABC}(a,b,\cancel{C}) = \lambda \langle A,B,\cancel{C} \rangle.$

a	b	c	h(A=a, B=b,C=c)
1	1	1	0.066
1	1	0	0.234
1	0	1	0.154
1	0	0	0.546
0	1	1	0.9009
0	1	0	0.0091
0	0	1	0.0891
0	0	0	0.0009

$h_{A,B,-c}(a,b) = \lambda \langle A,B \rangle.$

a	b	$h_{A,B,-C}$(A=a, B=b)
1	1	0.234
1	0	0.546
0	1	0.0091
0	0	0.0009

Bucket-Elimination: Set-up



$\theta_{A=1}$	$\theta_{A=0}$
0.4	0.6

a	$\theta_{B=1\|A=a}$	$\theta_{B=0\|A=a}$
1	0.325	0.675
0	0.440	0.550

a	$\theta_{C=1\|A=a}$	$\theta_{C=0\|A=a}$
1	0.200	0.800
0	0.367	0.633

b	c	$\theta_{D=1\|B=b,C=c}$	$\theta_{D=0\|B=b,C=c}$
1	1	0.300	0.700
1	0	0.333	0.667
0	1	0.250	0.750
0	0	0.450	0.550

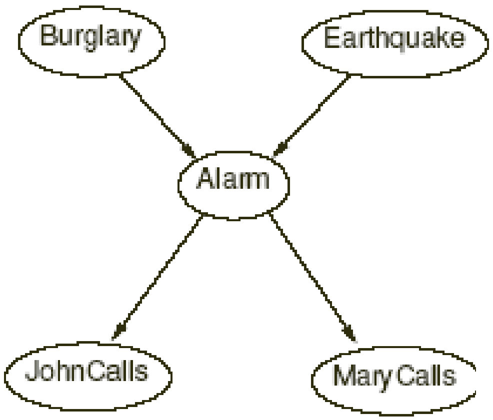
- Given
  - specific structure
  - specific CPtable entries
  - Fixed ordering over variables: $\pi_0 = \langle A,B,C,D \rangle$
- Create |Vars|+1 buckets
  - $b_{\{\}}$, b_A, b_B, b_C, b_D

$f_B(b) = \lambda\langle b\rangle.$

b	f(b)
0	0.999
1	0.001

$f_E(e) = \lambda\langle e\rangle.$

e	f(e)
0	0.998
1	0.002



$f_A(a,e,b) = \lambda\langle A,E,B\rangle.$

a	e	b	f(a, e, b)
1	1	1	0.95
1	1	0	0.29
:	:	:	:
0	0	1	0.06
0	0	0	0.999

$f_J(j,a) = \lambda\langle J,A\rangle.$

j	a	f(j,a)
1	1	0.90
1	0	0.05
0	1	0.10
0	0	0.95

$f_M(m,a) = \lambda\langle M,A\rangle.$

m	a	f(m,a)
1	1	0.70
1	0	0.01
0	1	0.30
0	0	0.99

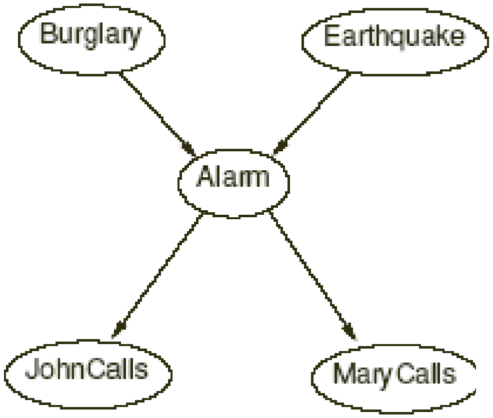
$-b, \quad +j, \quad +m$

$f_{-b}() = \lambda\langle\rangle.$

b	f(b)
0	0.999
~~1~~	~~0.001~~

$f_E(e) = \lambda\langle e\rangle.$

e	f(e)
0	0.998
1	0.002



$f_{A,-b}(a,e) = \lambda\langle A,E\rangle.$

a	e	~~b~~	f(a, e, b)
~~1~~	~~1~~	~~1~~	~~0.95~~
1	1	~~0~~	0.29
:	:	~~:~~	:
~~0~~	~~0~~	~~1~~	~~0.06~~
0	0	~~0~~	0.999

$f_{+j}(a) = \lambda\langle A\rangle.$

~~j~~	a	f(j,a)
~~1~~	1	0.90
~~1~~	0	0. 05
~~0~~	~~1~~	~~0.10~~
~~0~~	~~0~~	~~0.95~~

$f_{+m}(a) = \lambda\langle A\rangle.$

~~m~~	a	f(m,a)
~~1~~	1	0.70
~~1~~	0	0. 01
~~0~~	~~1~~	~~0.30~~
~~0~~	~~0~~	~~0.99~~

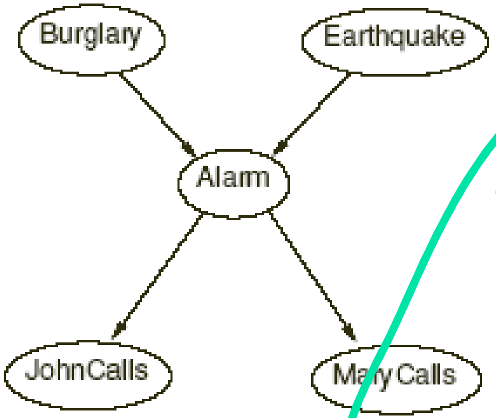
$\boxed{-b,\quad +j,\quad +m}$

$f_{-b}() = \lambda \langle\rangle.$

b	f(-b)
0	0.999

$f_E(e) = \lambda \langle e \rangle.$

e	f(e)
0	0.998
1	0.002



$f_{A,-b}(a,e) = \lambda \langle A,E \rangle.$

a	e	f(a, e, -b)
1	1	0.29
:	:	:
0	0	0.999

$f_{+j}(a) = \lambda \langle A \rangle.$

a	f(+j,a)
1	0.90
0	0.05

$f_{+m}(a) = \lambda \langle A \rangle.$

a	f(+m,a)
1	0.70
0	0.01

$b_{\{\}}$	b_B	b_E	b_A	b_J	b_M
$f_{\{\},1}() = \theta_{-b}$	-	$f_{E,1}(e) = \theta_e$	$f_{A,1}(a,e) = \theta_{a\|-b,e}$ $f_{A,2}(a) = \theta_{+j\|a}$ $f_{A,3}(a) = \theta_{+m\|a}$	-	-

“Variable Elimination”: Factors

P(-b, +j, +m) =

$$\underbrace{P(-b)}_{B} \quad \underbrace{\sum_e P(e)}_{E} \quad \underbrace{\sum_a P(a \mid -b, e)}_{A} \quad \underbrace{P(+j \mid a)}_{J} \quad \underbrace{P(+m \mid a)}_{M}$$

- Store intermediate results (factors) to avoid recomputation
  - Factor for M:
    2-element vector

$$\lambda_M(A) = \begin{pmatrix} P(+M \mid +A) \\ P(+M \mid -A) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}$$

$$\lambda_M(+A) = 0.70; \quad \lambda_M(-A) = 0.01$$

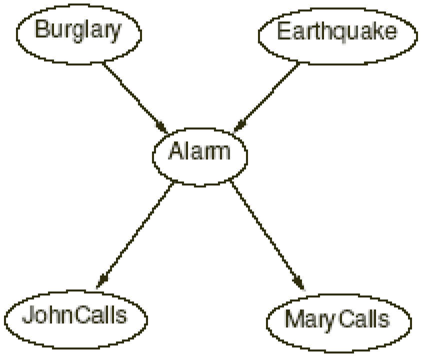
  - Factor for J:

$$\lambda_J(A) = \begin{pmatrix} P(+J \mid +A) \\ P(+J \mid -A) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}$$

  - Factor for A:
    ≡ 4-element vector

$$\lambda_A(A, E) = \begin{pmatrix} P(+A \mid -B, +E) \\ P(+A \mid -B, -E) \\ P(-A \mid -B, +E) \\ P(-A \mid -B, -E) \end{pmatrix} = \begin{pmatrix} 0.29 \\ 0.001 \\ 0.71 \\ 0.999 \end{pmatrix}$$

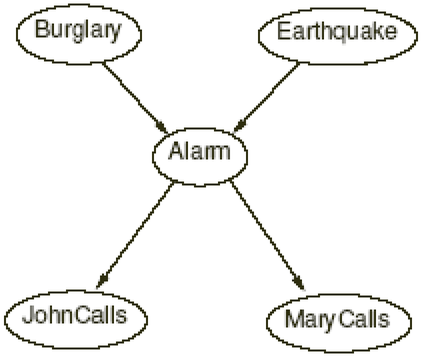
Bucket-Elimination Alg



- Compute TABLE from each CPtable entry… instantiating bound variables
  - For P(M=1): $f_{M: \text{“}M=1\text{”}}(a)$ [projection]
- Store each CPtable into bucket of HIGHEST index

$b_{\{\}}$	b_B	b_E	b_A	b_J	b_M
$f_{\{\},1}() = \theta_{-b}$	-	$f_{E,1}(e) = \theta_e$	$f_{A,1}(a,e) = \theta_{a\|-b,e}$ $f_{A,2}(a) = \theta_{+j\|a}$ $f_{A,3}(a) = \theta_{+m\|a}$	-	-

BE Alg, con't



- Process buckets, from highest to lowest
  - $g_X := \text{elim}_X[\ f_{X,1} \bowtie f_{X,2} \bowtie \ldots \bowtie f_{X,k}\ ]$
  - g_x is function of $\cup_i \text{Vars}(\ f_{X,I}\ ) - \{X\}$
    Let highest index by "Y"

  Store g_X into b_Y

- Process b_A
  - $g_A(e) = \text{elim}_A[\ f_{A,1} \bowtie f_{A,2} \bowtie f_{A,3}]$
  - add to b_E …

$b_{\{\}}$	b_B	b_E	b_A	b_J	b_M
$f_{\{\},1}() = \theta_{-b}$	-	$f_{E,1}(e) = \theta_e$	~~$f_{A,1}(a,e) = \theta_{a\|-b,e}$~~ ~~$f_{A,2}(a) = \theta_{+j\|a}$~~ ~~$f_{A,3}(a) = \theta_{+m\|a}$~~	-	-

$$\boxed{f_{E,2}(e) = \text{elim}_A[\ f_{A,1} \bowtie f_{A,2} \bowtie f_{A,3}]}$$

Variable Elimination: "Join"

$$P(-b, +j, +m) = \underbrace{P(-b)}_{B}\ \underbrace{\textstyle\sum_e P(e)}_{E}\ \textstyle\sum_a \underbrace{\underbrace{P(a \mid -b, e)}_{A}\ \underbrace{P(+j \mid a)}_{J}\ \underbrace{P(+m \mid a)}_{M}}_{[A,J,M]}$$

- Compute "join"

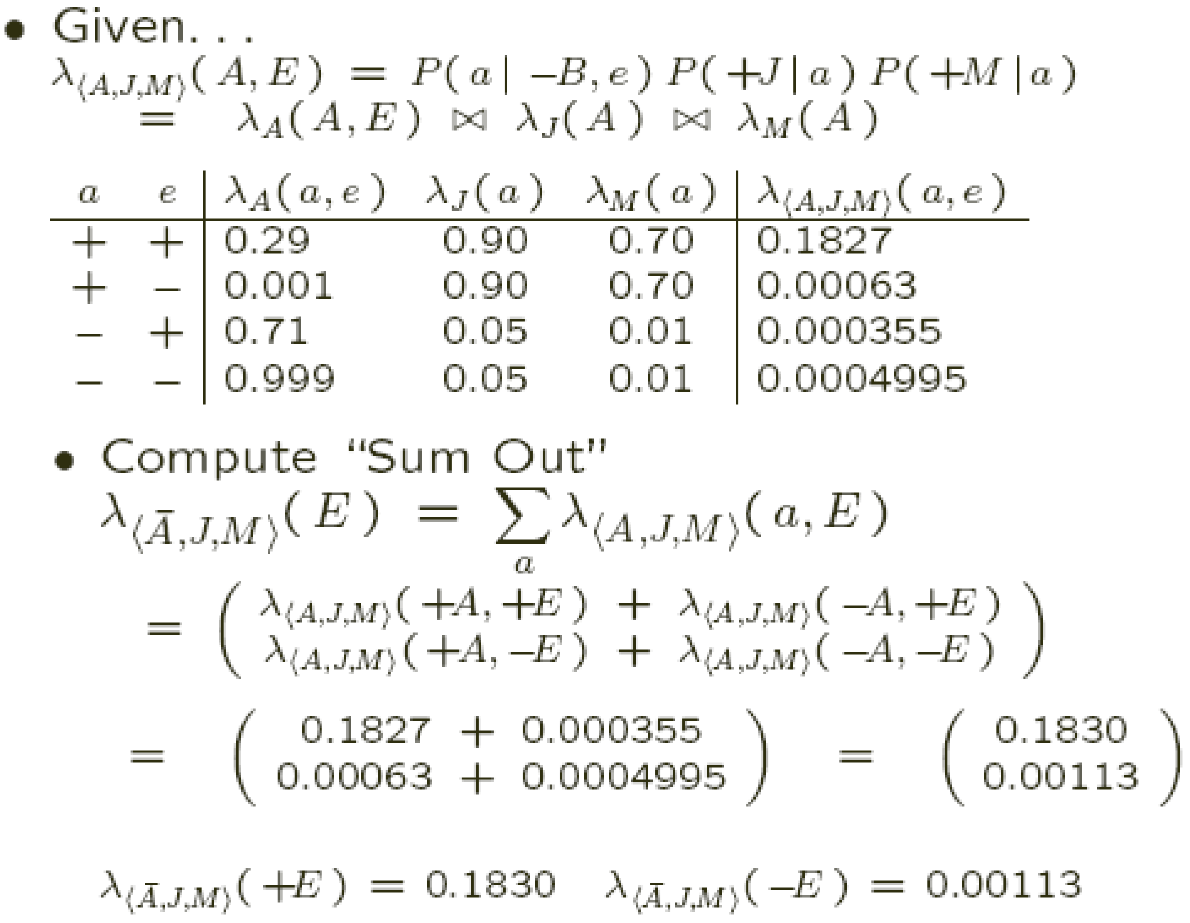
$$\begin{aligned}\lambda_{\langle A,J,M\rangle}(a,e) &= P(a \mid -B, e)\, P(+J \mid a)\, P(+M \mid a) \\ &= \lambda_A(a,e) \bowtie \lambda_J(a) \bowtie \lambda_M(a)\end{aligned}$$

a	e	$\lambda_A(a,e)$	$\lambda_J(a)$	$\lambda_M(a)$	$\lambda_{\langle A,J,M\rangle}(a,e)$
$+$	$+$	0.29	0.90	0.70	$0.29 \times 0.90 \times 0.70 = 0.1827$
$+$	$-$	0.001	0.90	0.70	$0.001 \times 0.90 \times 0.70 = 0.00063$
$-$	$+$	0.71	0.05	0.01	$0.71 \times 0.05 \times 0.01 = 0.000355$
$-$	$-$	0.999	0.05	0.01	$0.999 \times 0.05 \times 0.01 = 0.0004995$

Variable Elimination: "Sum Out"

P(-b, +j, +m) =

$P(-b) \; \sum_e P(e) \; \sum_a P(a \mid -b, e) \; P(+j \mid a) \; P(+m \mid a)$



- Given. . .

$$\lambda_{\langle A,J,M\rangle}(A, E) = P(a \mid -B, e)\, P(+J \mid a)\, P(+M \mid a)$$
$$= \lambda_A(A, E) \bowtie \lambda_J(A) \bowtie \lambda_M(A)$$

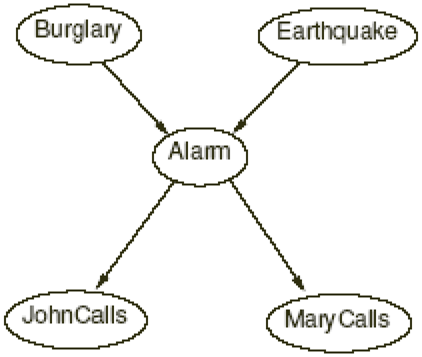
a	e	$\lambda_A(a,e)$	$\lambda_J(a)$	$\lambda_M(a)$	$\lambda_{\langle A,J,M\rangle}(a,e)$
+	+	0.29	0.90	0.70	0.1827
+	–	0.001	0.90	0.70	0.00063
–	+	0.71	0.05	0.01	0.000355
–	–	0.999	0.05	0.01	0.0004995

- Compute "Sum Out"

$$\lambda_{\langle \bar{A},J,M\rangle}(E) = \sum_a \lambda_{\langle A,J,M\rangle}(a, E)$$
$$= \begin{pmatrix} \lambda_{\langle A,J,M\rangle}(+A,+E) + \lambda_{\langle A,J,M\rangle}(-A,+E) \\ \lambda_{\langle A,J,M\rangle}(+A,-E) + \lambda_{\langle A,J,M\rangle}(-A,-E) \end{pmatrix}$$
$$= \begin{pmatrix} 0.1827 + 0.000355 \\ 0.00063 + 0.0004995 \end{pmatrix} = \begin{pmatrix} 0.1830 \\ 0.00113 \end{pmatrix}$$

$\lambda_{\langle \bar{A},J,M\rangle}(+E) = 0.1830 \quad \lambda_{\langle \bar{A},J,M\rangle}(-E) = 0.00113$

BE Alg, con't



- Process buckets, from highest to lowest
  - $g_X := \text{elim}_X[\ f_{X,1} \bowtie f_{X,2} \bowtie \ldots \bowtie f_{X,k}\ ]$
  - g_x is function of $\cup_i \text{Vars}(\ f_{X,I}\ ) - \{X\}$
    Let highest index by "Y"

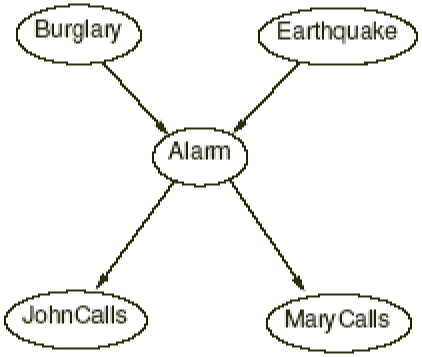
  Store g_X into b_Y

- Process b_E
  - $g_E() = \text{elim}_E[\ f_{E,1} \bowtie f_{E,2}]$
  - add to $b_{\{\}}$ …

$b_{\{\}}$	b_B	b_E	b_A	b_J	b_M
$f_{\{\},1}() = \theta_{-b}$	-	~~$f_{E,1}(e) = \theta_e$~~ ~~$f_{E,2}(e) = \ldots$~~	$f_{A,1}(a,e) = \theta_{a\|-b,e}$ $f_{A,2}(a) = \theta_{+j\|a}$ $f_{A,3}(a) = \theta_{+m\|a}$	-	-

$$\boxed{f_{\{\},2}() = \text{elim}_E[\ f_{E,1} \bowtie f_{E,2}]}$$

BE Alg, con't



- Process buckets, from highest to lowest
  - $g_X := \text{elim}_X[\ f_{X,1} \bowtie f_{X,2} \bowtie \ldots \bowtie f_{X,k}\ ]$
  - g_x is function of $\cup_i \text{Vars}(\ f_{X,I}\ ) - \{X\}$
    Let highest index by "Y"

  Store g_X into b_Y

- Process $b_{\{\}}$
  - $g_{\{\}}() = [\ f_{\{\},1} \bowtie f_{\{\},2}]$
  - Return $g_{\{\}|}$ …

$b_{\{\}}$	b_B	b_E	b_A	b_J	b_M
~~$f_{\{\},1}() = \theta_{-b}$~~ ~~$f_{\{\},2}() = \ldots$~~	-	$f_{E,1}(e) = \theta_e$ $f_{E,2}(e) = \ldots$	$f_{A,1}(a,e) = \theta_{a\|-b,e}$ $f_{A,2}(a) = \theta_{+j\|a}$ $f_{A,3}(a) = \theta_{+m\|a}$	-	-

Return $f_{\{\},1} \bowtie f_{\{\},2}$

Bucket Elimination Algorithm

Given:

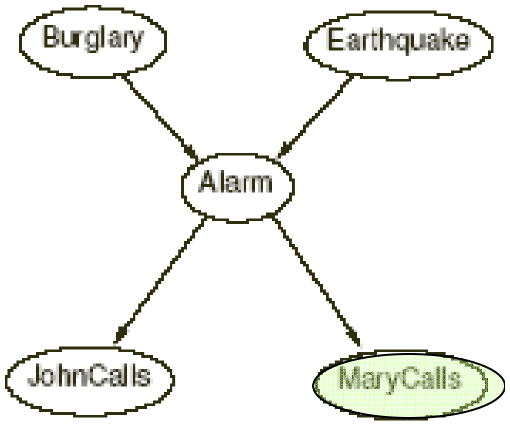
- Belief Net BN = $\langle$ **N**, A, θ $\rangle$
- Order of nodes $\pi = \langle X_1, \dots, X_{|\mathbf{N}|} \rangle$
- Evidence (nodes $\{E_i\} \subset$ **N**, values $\{e_i\}$)
- (Single) Query node $X \in$ **N**

Compute: $P(X \mid E_1 = e_1, \dots)$

by computing $P(X = x, E_1 = e_1, \dots) \; \forall \; x$

- Step#1: Initialize |N| +1 "buckets"
  - . . . bucket b_i for variable X_i
  - Each "instantiated form of CPtables" is function of variables
  - Store in bucket with highest index
- Step#2: Process each bucket
  - . . . from highest index down
  - to eliminate associated variable
- Step#3: Read off answer
  - . . . in "top" bucket, $b_{\{\}}$

Remove “Dead Variables”



$P(+b, +j\ ) =$

$= \sum_e \sum_a \sum_m P(+b, E=e, A=a, +j, M=m)$

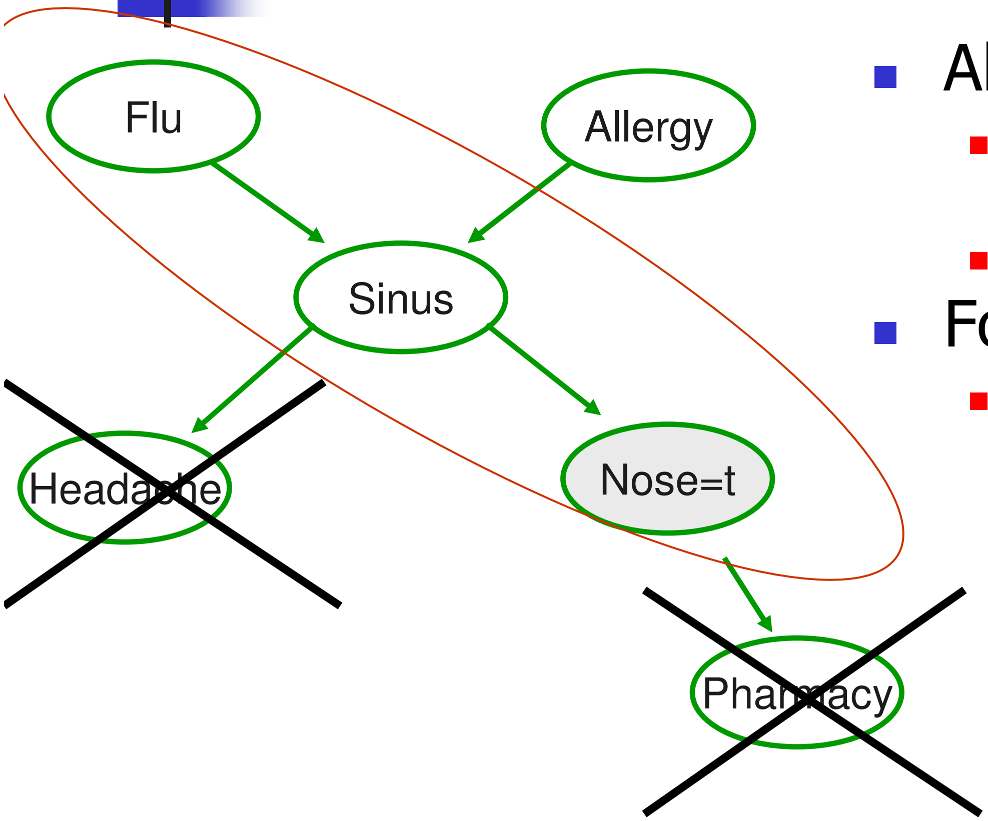
$= \sum_e \sum_a \sum_m P(+b)\ P(E=e)\ P(a|+b,e)\ P(+j|a)\ P(m|a)$

$= P(+b) \sum_e P(e) \sum_a P(a|+b,e)\ P(+j|a) \sum_m P(m|a)$

- Note for any $A=a$, $\sum_m P(\ M=m \mid a\ ) = 1$

  $\Rightarrow$ can remove this node!
- In general: need to keep only nodes ABOVE query, evidence notes
  (Remove any nodes below)

Prune Irrelevant Variables



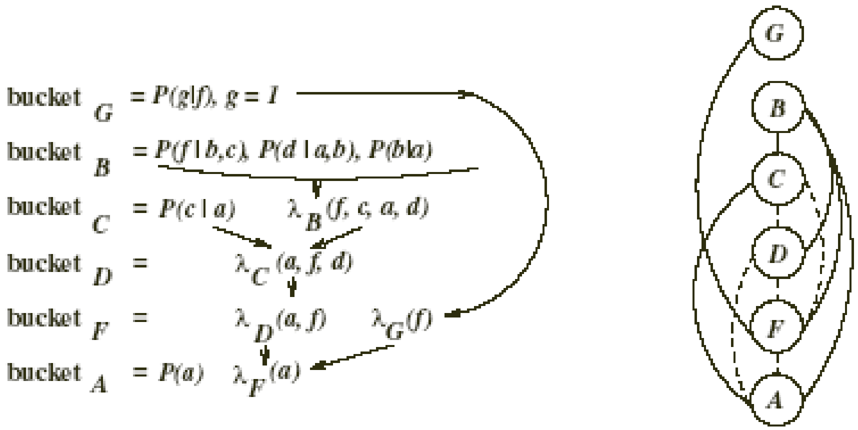
- Alg:
  - Find trails between Query and Evidence vars
  - Remove all descendants
- For P(F, N=t)
  - Remove: Headache Pharmacy

Prune all non-ancestors of query variables.
In general: Prune all nodes not on active trail between evidence and query vars

Effectiveness of Bucket Elimination

- Returns Correct answer,
  given any order of nodes.
- But COMPLEXITY is very sensitive to Node Order!

Eg: $\pi_2 = \langle A, F, D, C, B, G \rangle$:



bucket $_G$ = $P(g|f)$, $g = 1$

bucket $_B$ = $P(f|b,c)$, $P(d|a,b)$, $P(b|a)$

bucket $_C$ = $P(c|a)$ $\lambda_B(f, c, a, d)$

bucket $_D$ = $\lambda_C(a, f, d)$

bucket $_F$ = $\lambda_D(a, f)$ $\lambda_G(f)$

bucket $_A$ = $P(a)$ $\lambda_F(a)$

Note: π_1 has complexity $O(|Dom|^3)$

π_2 has complexity $O(|Dom|^4)$

- In gen'l: time/space complexity is
  Exponential in size of largest bucket

Time/Space Complexity

- Given BN and ordering π,
  bucket for X_i includes X_j if
  - X_j is parent/child/coparent of X_j
    (or X_j is p/c/c-p of X_k which is p/c/c-p of earlier X_m) and
  - X_j is before X_j

  (Ie, after connecting earlier neighbors of processed var)
- Called: "**Induced width**" of var $X \subset N$
- $w^*(\pi)$ = induced width of ordering π
  = maximum induced width of any variable
  - "induced width of ordered moral graph"
- $\Rightarrow$ Time/Space Complexity of BucketElimination:
  EXPONENTIAL in $w^*(\pi)$

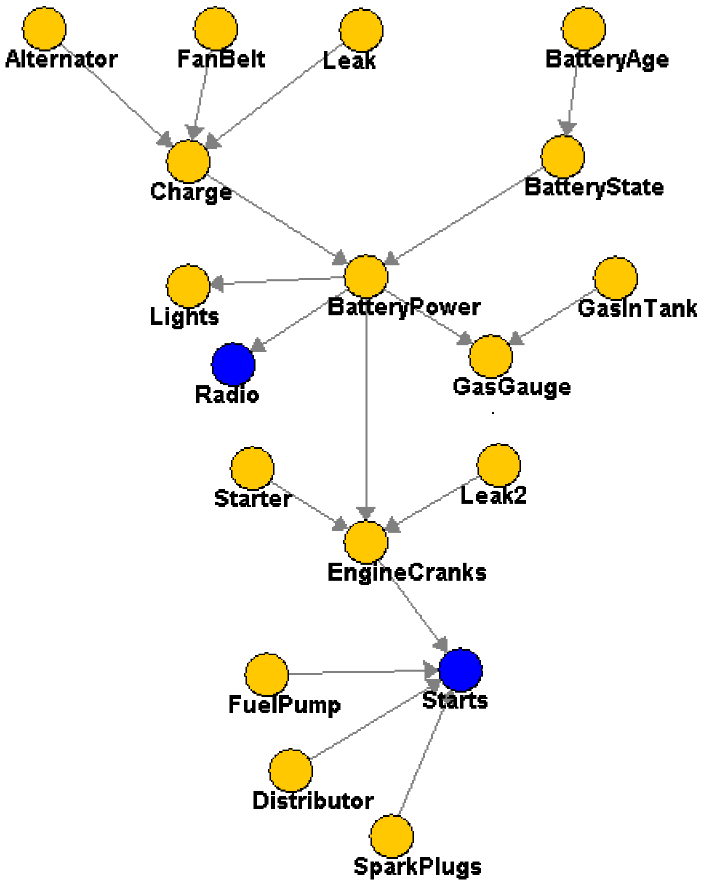
- NP-hard to find best ordering π^*. . .

Comments on Bucket Elimination

- Observed Variables " $X_i = e_i$":
  + Put at end of ordering (process first)
  or... do not use bucket $\beta[X_i]$
  + Ignore when computing induced width
- Can skip bucket $\beta[X_i]$ if it contains *no*
  - evidence variables
  - query variable,
  - newly-computed $\lambda_X(Y)$
- If BN is poly-tree, and π respects topol. ordering, then BucketElim $\approx$ PolyTree alg !

- BucketElim $\equiv$
  - SPI [D'Ambrosia, et al.]
  - Join-Tree Clustering [Lauritzen,Speigelhalter'88] ("Directional Version")

Complexity of Variable Elim: (Poly)-tree graphs



Variable elimination order:

Start from "leaves" inwards:

- Start from skeleton!
- Choose a "root"… any node
- Find topological order for root
- Eliminate variables in reverse order

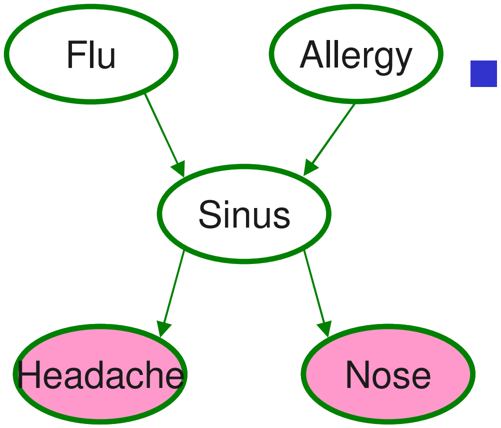
Running time:

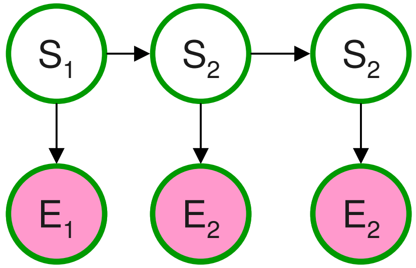
- Exponential in #parents
- For poly-tree: linear is #vars!!
  … linear in CPTable size!

Outline

- Framework
- PolyTree algorithm
  - Message Propagation
- Complexity
- Bucket Elimination
- Other types of Inference tasks
  - Complexity

Inference in graphical models: Typical queries 2 – Maximization

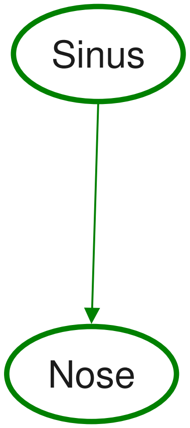




- Most probable explanation (MPE)
  - Most likely assignment to all hidden vars given evidence
  - $\text{argmax}_{f,a,s,n}$ $P(F=f, A=a, S=s \mid N=n, H=t)$
  - $\text{argmax}_{s1,s2,s3}$
    $P(S_1=s_1, S_2=s_2 , S_3=s_3 \mid e_1,e_2,e_3)$
- Maximum a posteriori (MAP)
  - Most likely assignment to some hidden vars given evidence
  - argmax_f $P(F=f \mid H=t)$
  - argmax_{s3} $P(S_3=s_3 \mid e_1,e_2,e_3)$

Is MAP Monotonic?

s	t	f
	.4	.6



P(N|S)

s\n	t	f
t	.9	.1
f	.5	.5

- $\text{argmax}_s\ P(S=s) = [-s]$
- $\text{argmax}_{s,n}\ P(S=s,N=n) = [+s, +n]$

Other BN Tasks

- **MPE (Most Probable Explanation):**
  Given evidence **E = e** $(E_1 = e_1, \ldots, E_m = e_m)$
  - find assignment **x** that maximizes $P(\mathbf{x} \mid \mathbf{E} = \mathbf{e})$

    $$= \arg\max_x \prod_{i=1..m} P(x_i \mid e, pa_i)$$
  - Alg ≈ like BucketElim for BeliefAssessment
    but replace $\sum$ with max
- **MAP (Maximum a Posteriori):**
  Given evidence **E = e**

  and set of hypothesis $H_1, \ldots, H_k$
  - find assignment to HYPOTHESIS h that maximizes
    $P(h \mid \mathbf{E} = \mathbf{e}) = \arg\max_h \prod_{i=1..m} P(x_i \mid \mathbf{e}, pa_i)$

Complexity of Inference

- Probabilistic inference
  - general graphs: #P-complete
  - poly-trees and low tree-width: easy
- Approximate probabilistic inference
  - Absolute error: $\left|\hat{P}-P\right| \le \varepsilon \ldots \quad \mathrm{NP-hard}\ \forall \varepsilon < 0.5$
  - Relative error: $\dfrac{\left|\hat{P}-P\right|}{P} \le \varepsilon \ldots \quad \mathrm{NP-hard}\ \forall \varepsilon > 0$
- Most probable explanation (MPE)
  - general graphs: NP-complete
  - poly-trees and low tree-width: easy
- Maximum a posteriori (MAP)
  - general graphs: $\mathrm{NP}^{\mathrm{PP}}$-complete
  - poly-trees and low tree-width: NP-hard

Summary

- Types of queries
  - probabilistic inference
  - most probable explanation (MPE)
  - maximum a posteriori (MAP)
    - MPE and MAP are truly different (diff answers, diff complexities)

- Hardness of inference
  - Exact and approximate inference are NP-hard
  - MPE is NP-complete
  - MAP is much harder (NP^{PP}-complete)

- Variable Elimination algorithm
  - Eliminate a variable:
    - Combine factors that include this var into single factor
    - Marginalize var from new factor
  - Efficient algorithm ("only" exponential in induced-width, not #Vars)

- Elimination order is important!
  - NP-complete problem
  - Many good heuristics

Probabilistic Inference Tasks, in Gen’l

- **Simple queries**: compute posterior marginal P(X | **E** = **e**)
  - P(NoGas | Gauge = empty, Lights = on, Starts = false)

- **Conjunctive queries**:
  P(X, Y | **E = e**) = P(X | **E = e**) P(Y | X, **E = e**)

- **Optimal decisions**:
  - decision networks include utility information.
  - Probabilistic inference required for P(outcome | action, **E = e**)

- **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?