Bayesian Networks: Inference

Forest Agostinelli
University of South Carolina

Topics Covered in This Class

- **Part 1: Search**
  - Pathfinding
    - Uninformed search
    - Informed search
  - Adversarial search
  - Optimization
    - Local search
    - Constraint satisfaction
- **Part 2: Knowledge Representation and Reasoning**
  - Propositional logic
  - First-order logic
  - Prolog
- **Part 3: Knowledge Representation and Reasoning Under Uncertainty**
  - Probability
  - Bayesian networks
- **Part 4: Machine Learning**
  - Supervised learning
    - Inductive logic programming
    - Linear models
    - Deep neural networks
    - PyTorch
  - Reinforcement learning
    - Markov decision processes
    - Dynamic programming
    - Model-free RL
  - Unsupervised learning
    - Clustering
    - Autoencoders

Outline

- Inference by enumeration
- Variable elimination

Inference by Enumeration

- P(W)?
  - P(sun)=.3+.1+.1+.15=.65
  - P(rain)=1- P(sun)=1-0.65=0.35

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

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

- P(W|winter, hot)?
  - P(W, winter, hot)/P(winter, hot)
  - P(sun|winter,hot)=0.1/(0.1 + 0.05)=2/3
  - P(rain|winter,hot)=0.05/(0.1 + 0.05)=1/3

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

$$P(a|b) = \frac{P(a,b)}{P(b)} \qquad P(t) = \sum_{s} P(t,s)$$

- P(W|winter)?
  - P(W|winter)=P(W,winter)/P(winter)
  - P(sun|winter)=(0.1 + 0.15)/0.5=0.5
  - P(rain|winter)=(0.05 + 0.2)/0.5=0.5

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

  $\left.\right\}$ $X_1, X_2, \dots X_n$ *All variables*

- We want:

** Works fine with multiple query variables, too*

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

- Step 1: Select the entries consistent with the evidence
- Step 2: Sum out H to get joint of Query and evidence
- Step 3: Normalize

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

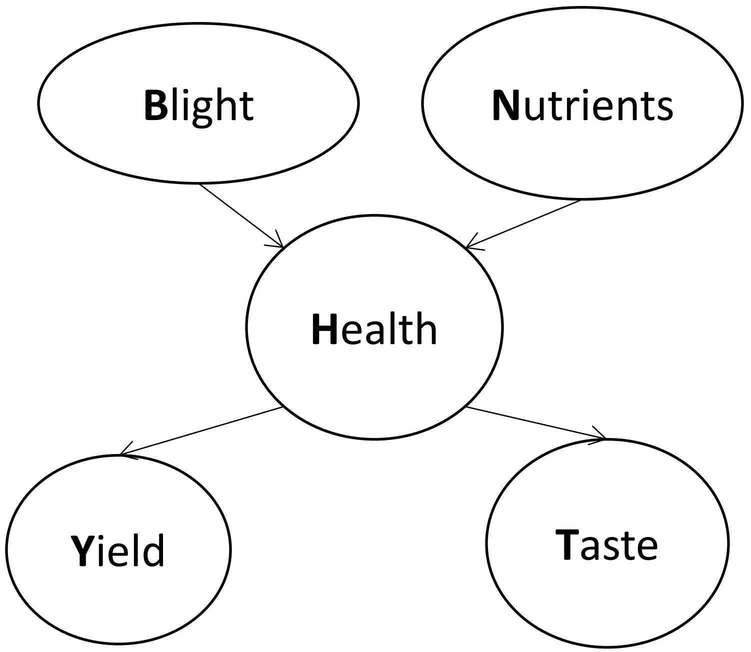
$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

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

Example: Plant Health Network

*modified from AIMA

P(B=+b)
0.1



P(N=+n)
0.95

B	N	P(H=+h|B,N)
+b	+n	0.5
+b	-n	0.05
-b	+n	0.99
-b	-n	0.6

H	P(Y=+y|H)
+h	0.7
-h	0.1

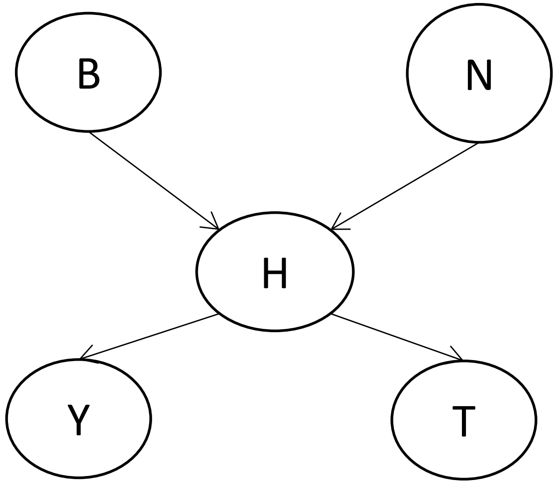
H	P(T=+t|H)
+h	0.9
-h	0.05

Example: Plant Health Network

*modified from AIMA

P(B=+b)
0.1

P(N=+n)
0.95



B	N	P(H=+h|B,N)
+b	+n	0.5
+b	-n	0.05
-b	+n	0.99
-b	-n	0.6

H	P(Y=+y|H)
+h	0.7
-h	0.1

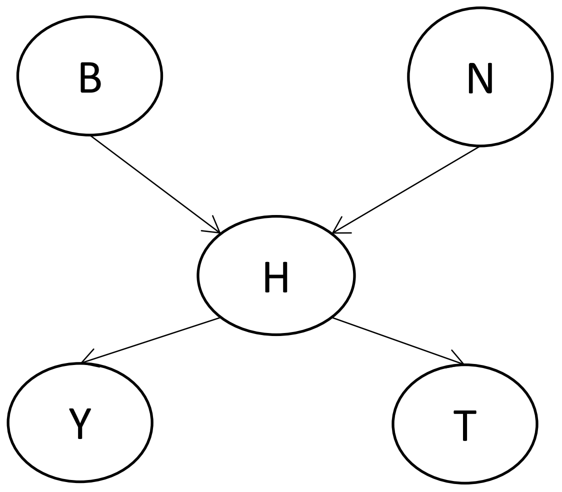
H	P(T=+t|H)
+h	0.9
-h	0.05

$$P(-b,+n,+h,-y,-t) = P(-b)P(+n)P(+h|-b,+n)P(-y|+h)P(-t|+h)$$

$$= 0.9 * 0.95 * 0.99 * 0.3 * 0.1 = 0.025$$

Example: Plant Health Network

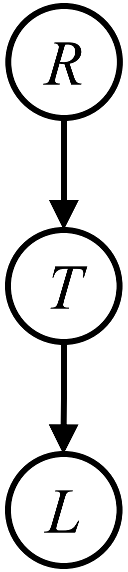
- $P(B|+y,+t) = \frac{P(B,+y,+t)}{P(+y,+t)}$
- $P(B,+y,+t) = \sum_{n,h} P(B,n,h,+y,+t)$
- $= \sum_{n,h} P(B)P(n)P(h|B,n)P(+y|h)P(+t|h)$



- Worst case, for n variables
  - We have to sum out n variables
  - Each is a product of n probability values
  - $O(n2^n)$
- We can take advantage of nested structure to reduce complexity to $O(2^n)$

Example: Traffic Domain

- 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

$$P(L) = ?$$

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

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

Inference by Enumeration: Factors

- Track objects called factors
  - Factors start out as local conditional probability tables

$P(R)$

+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
  - For example, if we know $L = +l$, the initial factors are

$P(R)$

+r	0.1
-r	0.9

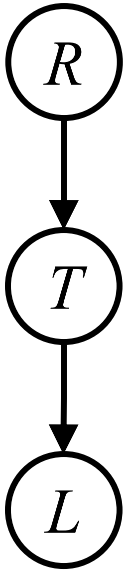
$P(T|R)$

+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

- Join all factors (pointwise multiplication) and then sum out all hidden variables



$P(R)$

+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

Pointwise Multiplication

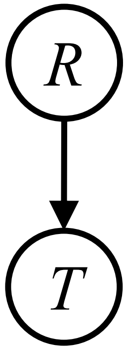
X	Y	$\mathbf{f}(X,Y)$	Y	Z	$\mathbf{g}(Y,Z)$
t	t	.3	t	t	.2
t	f	.7	t	f	.8
f	t	.9	f	t	.6
f	f	.1	f	f	.4

X	Y	Z	$\mathbf{h}(X,Y,Z)$
t	t	t	$.3 \times .2 = .06$
t	t	f	$.3 \times .8 = .24$
t	f	t	$.7 \times .6 = .42$
t	f	f	$.7 \times .4 = .28$
f	t	t	$.9 \times .2 = .18$
f	t	f	$.9 \times .8 = .72$
f	f	t	$.1 \times .6 = .06$
f	f	f	$.1 \times .4 = .04$

Figure 13.12 Illustrating pointwise multiplication: $\mathbf{f}(X,Y) \times \mathbf{g}(Y,Z) = \mathbf{h}(X,Y,Z)$.

Joining Factors

- In general
  - $\boldsymbol{f}(X_1, \dots, X_j, Y_1, \dots, Y_k) \times \boldsymbol{g}(Y_1, \dots, Y_k, Z_1, \dots, Z_l) = \boldsymbol{h}(X_1, \dots, X_j, Y_1, \dots, Y_k, Z_1, \dots, Z_l)$
  - If all variables are binary
    - $\boldsymbol{f}$ has 2^{j+k} entries
    - $\boldsymbol{g}$ has 2^{k+l} entries
    - $\boldsymbol{h}$ has 2^{j+k+l} entries



$P(R)$

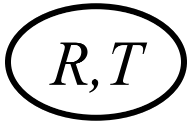
+r	0.1
-r	0.9

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

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

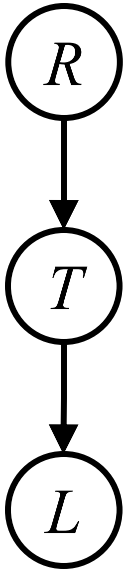
$\Rightarrow$ $P(R,T)$

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



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

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 on R

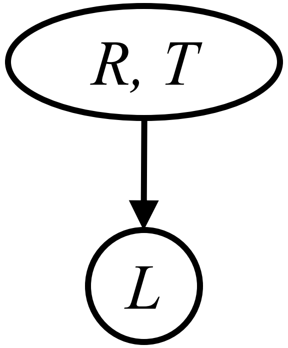
$P(R, T)$

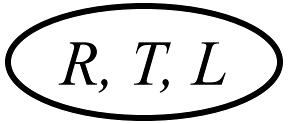
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

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

Join on T





$P(R, T, L)$

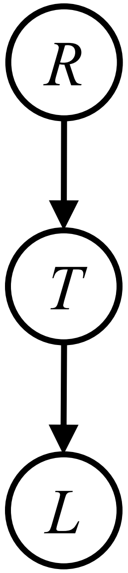
+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

Marginalization

$$P(L) = ?$$

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

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



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

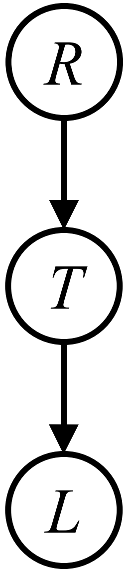
Sum out T →

L

$P(L)$

+l	0.134
-l	0.866

Variable Elimination: Example



$$P(L) = ?$$

- Inference by Enumeration

$$= \sum_t \sum_r P(L|t)P(r)P(t|r)$$

Join on r

Join on t

Eliminate r

Eliminate t

$$= \sum_t P(L|t) \sum_r P(r)P(t|r)$$

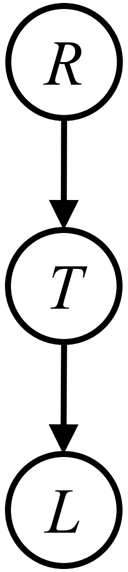
Join on r

Eliminate r

Join on t

Eliminate t

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

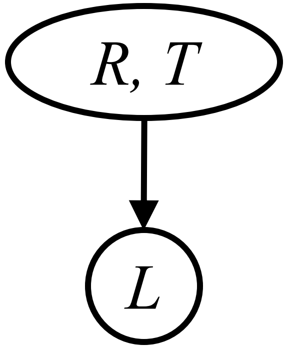
$P(L|T)$

+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

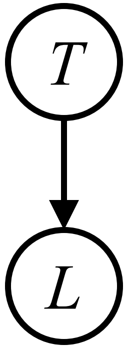


$P(L|T)$

Sum out R

$P(T)$

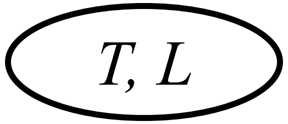
+t	0.17
-t	0.83



$P(L|T)$

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

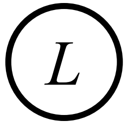
Join T



$P(T, L)$

+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

General Variable Elimination

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

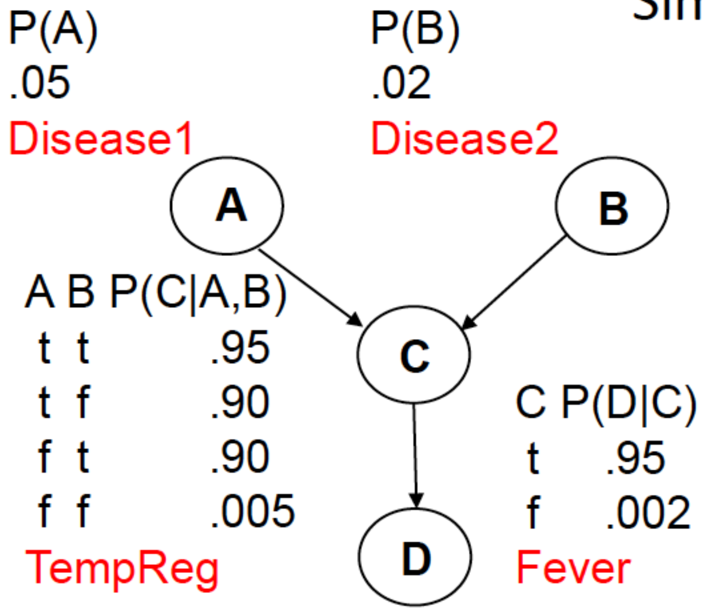
- Start with initial factors:
  - Local CPTs
  - Select entries consistent with evidence

- 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

Quick Quiz

Simple Example



} **Query Variables A, B**

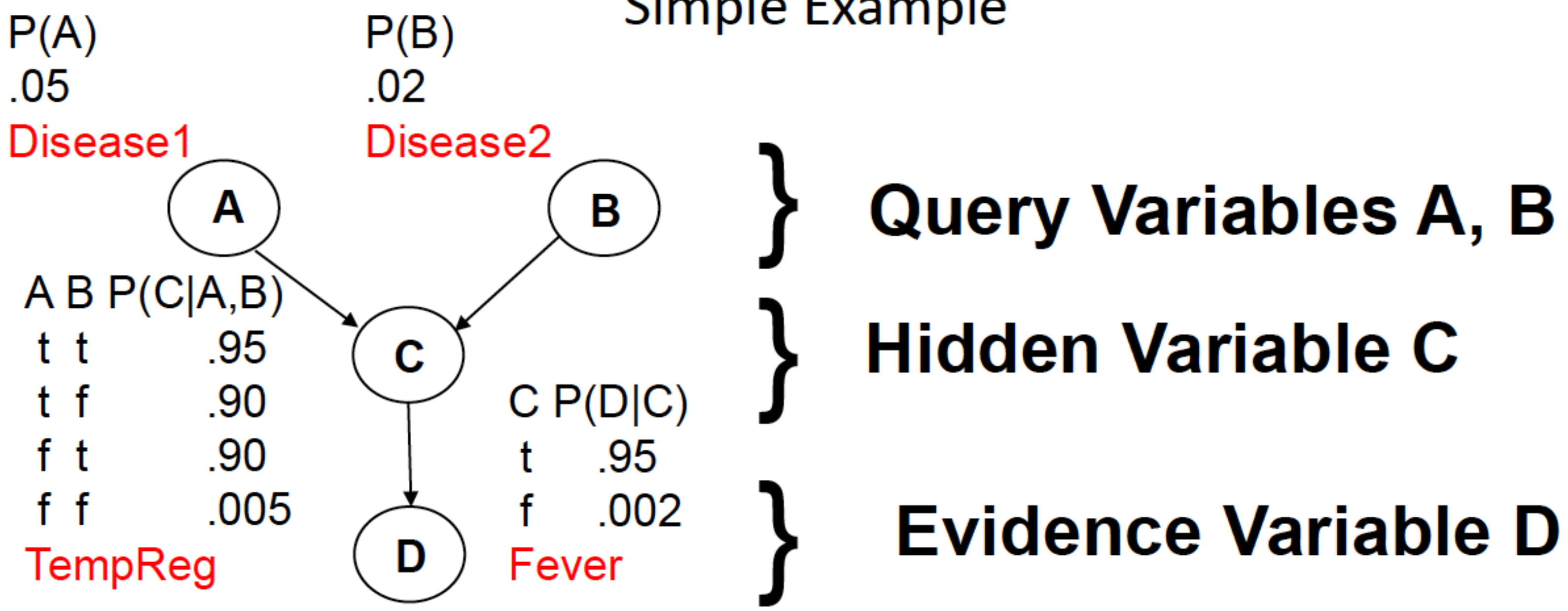
} **Hidden Variable C**

} **Evidence Variable D**

- Compute (do not compute normalizing factor)
  - $P(+a, +b | +d)$
- *Not a biologically correct model of how diseases cause fever

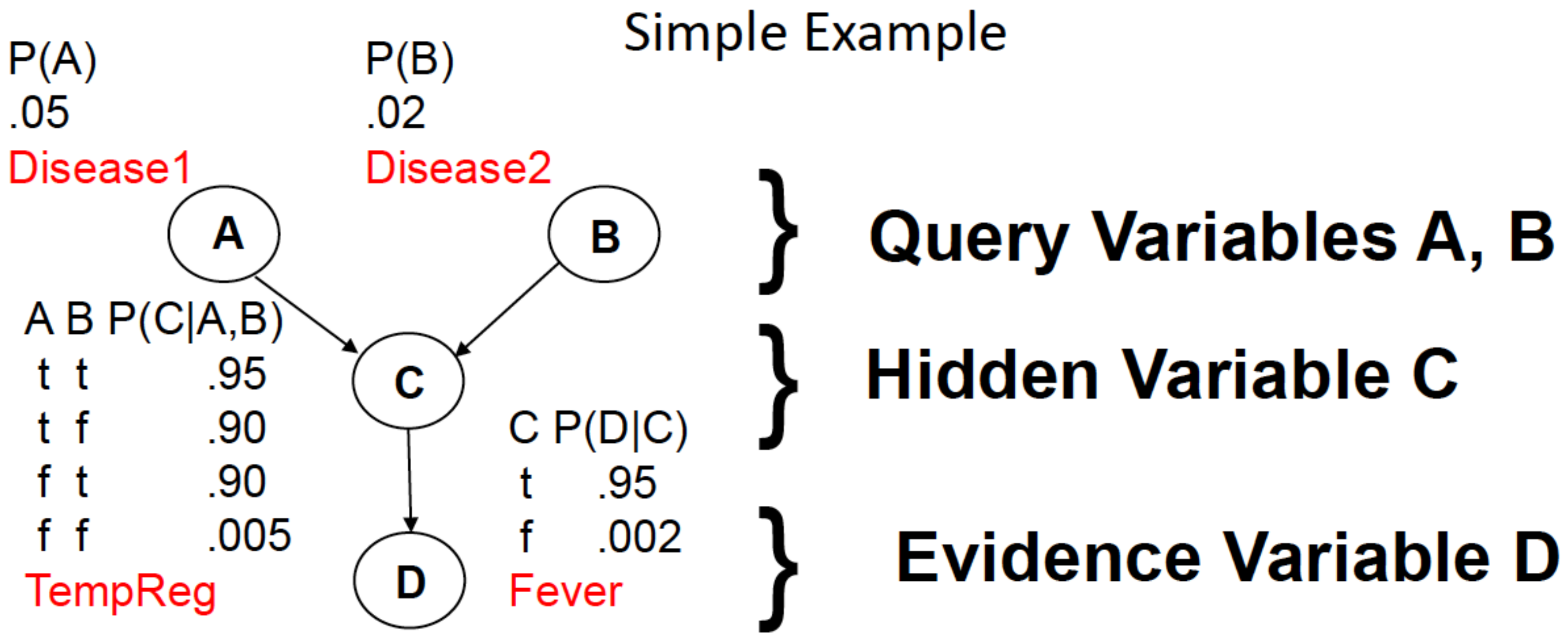
Quick Quiz

Simple Example



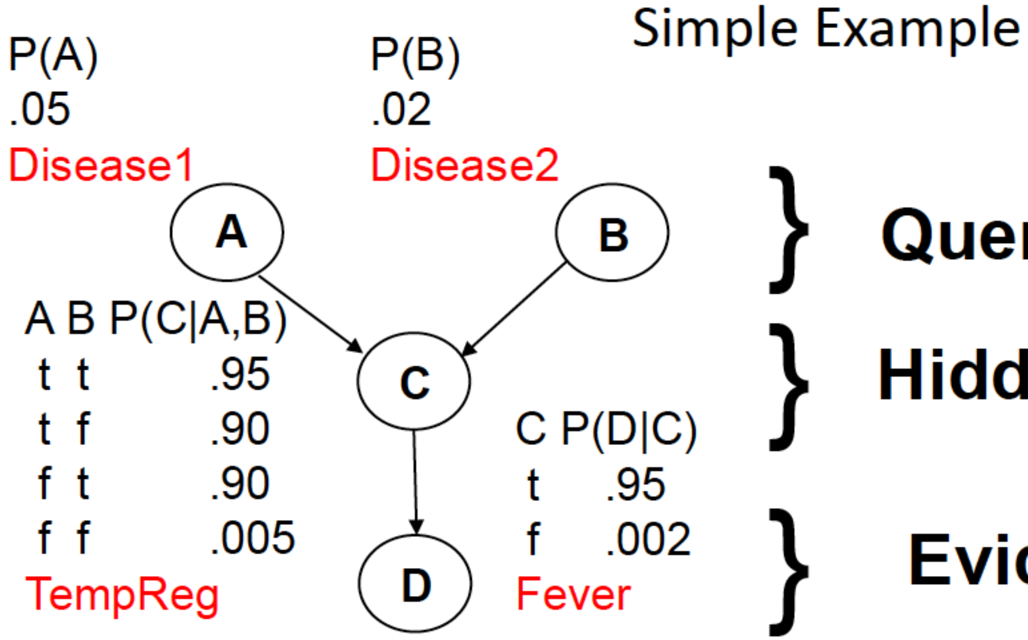
- $P(+a, +b|+d)\ \alpha\ P(+a, +b, +d)$
- $P(+a, +b, +d) = \sum_c P(+a)P(+b)P(c|+a, +b)p(+d|c)$
- $= P(+a)P(+b)\sum_c P(c|+a, +b)p(+d|c)$
- $0.05 * 0.02 * (0.95 * 0.95 + 0.05 * 0.002) = 0.000903$

Quick Quiz



- Compute (do not compute normalizing factor)
  - $P(-a, +b | +d)$

Quick Quiz



} **Query Variables A, B**

} **Hidden Variable C**

} **Evidence Variable D**

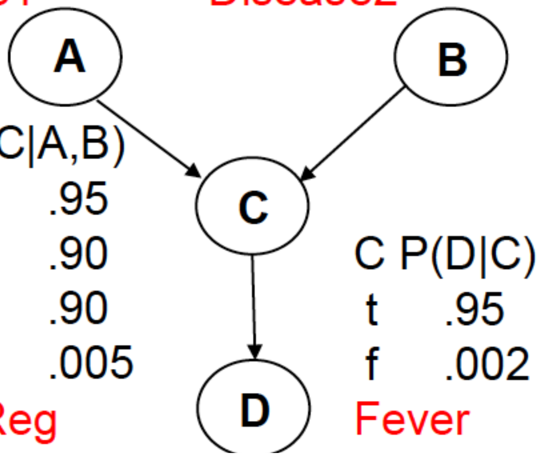
- $P(-a,+b|+d)\ \alpha\ P(-a,+b,+d) = \sum_c P(-a)P(+b)P(c|-a,+b)p(d|c)$
- $P(-a,+b,+d) = \sum_c P(-a)P(+b)P(c|-a,+b)p(+d|c)$
- $= P(-a)P(+b)\sum_c P(c|-a,+b)p(+d|c)$
- $0.95 * 0.02 * (0.90 * 0.95 + 0.1 * 0.002) = 0.0162$

Quick Quiz

Simple Example

P(A): .05 — Disease1

P(B): .02 — Disease2



A	B	P(C|A,B)
t	t	.95
t	f	.90
f	t	.90
f	f	.005

TempReg

C	P(D|C)
t	.95
f	.002

Fever

} **Query Variables A, B**

} **Hidden Variable C**

} **Evidence Variable D**

$$P(a,b|d) = \alpha\, P(a)P(b)\ \Sigma_c\ P(c|a,b)P(d|c) = \alpha\, P(a)P(b)\{\, P(c|a,b)P(d|c)+P(\neg c|a,b)P(d|\neg c)\,\}$$
$$= \alpha\ .05\text{x}.02\text{x}\{.95\text{x}.95+.05\text{x}.002\} \approx \alpha\ .000903 \approx .014$$

$$P(\neg a,b|d) = \alpha\, P(\neg a)P(b)\ \Sigma_c\ P(c|\neg a,b)P(d|c) = \alpha\, P(\neg a)P(b)\{\, P(c|\neg a,b)P(d|c)+P(\neg c|\neg a,b)P(d|\neg c)\,\}$$
$$= \alpha\ .95\text{x}.02\text{x}\{.90\text{x}.95+.10\text{x}.002\} \approx \alpha\ .0162 \approx .248$$

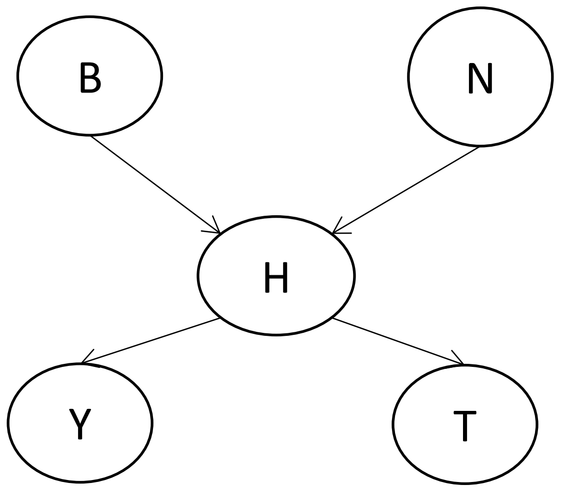
$$P(a,\neg b|d) = \alpha\, P(a)P(\neg b)\ \Sigma_c\ P(c|a,\neg b)P(d|c) = \alpha\, P(a)P(\neg b)\{\, P(c|a,\neg b)P(d|c)+P(\neg c|a,\neg b)P(d|\neg c)\,\}$$
$$= \alpha\ .05\text{x}.98\text{x}\{.90\text{x}.95+.10\text{x}.002\} \approx \alpha\ .0419 \approx .642$$

$$P(\neg a,\neg b|d) = \alpha\, P(\neg a)P(\neg b)\ \Sigma_c\ P(c|\neg a,\neg b)P(d|c) = \alpha\, P(\neg a)P(\neg b)\{\, P(c|\neg a,\neg b)P(d|c)+P(\neg c|\neg a,\neg b)P(d|\neg c)\,\}$$
$$= \alpha\ .95\text{x}.98\text{x}\{.005\text{x}.95+.995\text{x}.002\} \approx \alpha\ .00627 \approx .096$$

$$\alpha \approx 1\ /\ (.000903+.0162+.0419+.00627) \approx 1\ /\ .06527 \approx 15.32$$

Example: Plant Health Network

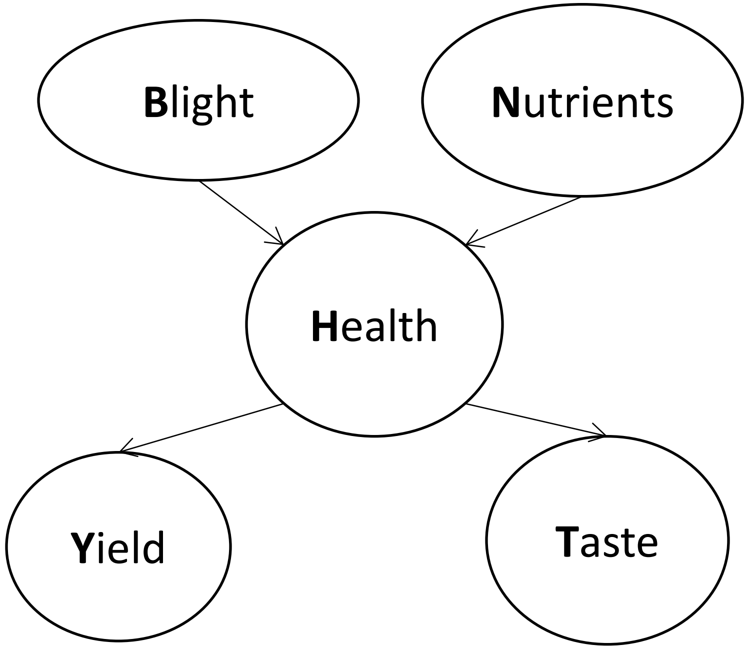
- $P(B| + y, +t) = \frac{P(B,+y,+t)}{P(+y,+t)}\ \alpha\ P(B, +y, +t)$
- $P(B, +y, +t) = \sum_{n,h} P(B, n, h, +y, +t)$
- $= \sum_{n,h} P(B)P(n)P(h|B, n)P(+y|h)P(+t|h)$
- $= \sum_{n,h} f_1(B)f_2(N)f_3(H, B, N)f_4(H)f_5(H)$
- $= f_1(B) \sum_n f_2(N) \sum_h f_3(H, B, N)f_4(H)f_5(H)$
- $f_6(B, N) = \sum_h f_3(H, B, N)f_4(H)f_5(H)$
  - $= f_3(+h, B, N)f_4(+h)f_5(+h) + f_3(\neg h, B, N)f_4(\neg h)f_5(\neg h)$
- $f_1(B) \sum_n f_2(N)f_6(B, N)$
- $f_7(B) = \sum_n f_2(N)f_6(B, N)$
  - $= f_2(+n)f_6(B, +n) + f_2(-n)f_6(B, -n)$
- $f_1(B)f_7(B)$



Example: Plant Health Network

$f_1(H)$

P(B=+b)
0.1



$f_2(H)$

P(N=+n)
0.95

$f_3(H)$

B	N	P(H=+h|B,N)
+b	+n	0.5
+b	-n	0.05
-b	+n	0.99
-b	-n	0.6

$f_4(H)$

H	P(Y=+y|H)
+h	0.7
-h	0.1

$f_5(H)$

H	P(T=+t|H)
+h	0.9
-h	0.05

Example: Plant Health Network

- $f_6(B,N) = \sum_h f_3(H,B,N) f_4(H) f_5(H)$
  - $= f_3(+h,B,N) f_4(+h) f_5(+h) + f_3(\neg h,B,N) f_4(\neg h) f_5(\neg h)$

$f_4(H)$

H	P(Y=+y|H)
+h	0.7
-h	0.1

$f_5(H)$

H	P(T=+t|H)
+h	0.9
-h	0.05

Join

H	P(+y,+t|H)
+h	0.63
-h	0.005

$f_3(H,B,N)$

B	N	P(H=+h|B,N)
+b	+n	0.5
+b	-n	0.05
-b	+n	0.99
-b	-n	0.6

Join

H	B	N	P(H,+y,+t|B,N)
+h	+b	+n	0.315
+h	+b	-n	0.0315
+h	-b	+n	0.6237
+h	-b	-n	0.378
-h	+b	+n	0.0025
-h	+b	-n	0.00475
-h	-b	+n	0.00005
-h	-b	-n	0.002

Sum

$f_6(B,N)$

B	N	P(+y,+t|B,N)
+b	+n	0.3175
+b	-n	0.03625
-b	+n	0.62375
-b	-n	0.38

Example: Plant Health Network

- $f_7(B) = \sum_n f_2(N) f_6(B,N)$
  - $= f_2(+n) f_6(B,+n) + f_2(-n) f_6(B,-n)$

$f_2(N)$

P(N=+n)
0.95

$f_6(B,N)$

B	N	P(+y,+t|B,N)
+b	+n	0.3175
+b	-n	0.03625
-b	+n	0.62375
-b	-n	0.38

Join

B	N	P(N,+y,+t|B)
+b	+n	0.301625
+b	-n	0.0018125
-b	+n	0.5925625
-b	-n	0.019

Sum

$f_7(B)$

B	P(+y,+t|B)
+b	0.3034375
-b	0.6115625

$$P(+y,+t|B,N)P(N) = \frac{P(+y,+t,N|B)}{P(N)} P(N) = P(+y,+t,N|B)$$

Example: Plant Health Network

- $f_1(B)f_7(B)$

$f_1(B)$

P(B=+b)
0.1

$f_7(B)$

B	P(+y,+t|B)
+b	0.3034375
-b	0.6115625

Join

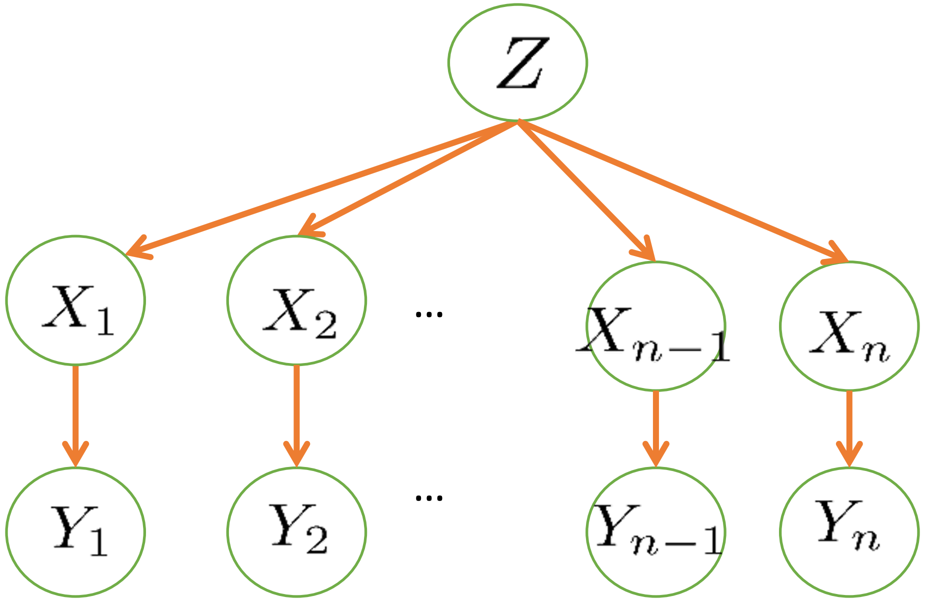
B	P(B,+y,+t)
+b	0.03034375
-b	0.55040625

Normalize

B	P(B|+y,+t)
+b	0.052249
-b	0.947751

Variable Ordering

- For the query $P(X_n | y_1,\ldots,y_n)$ work through the following two different orderings: 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 Ordering

- 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.
- There does not always exist an ordering that results in small factors
- Determining the optimal ordering is intractable

Polytrees

- A polytree is a directed graph with no undirected cycles
- For poly-trees you can always find an ordering that is efficient

Approximate Inference

- One can approximate inference using sampling
  - Section 13.4 in the 4th edition of Artificial Intelligence: A Modern Approach
- It is generally fast but can be wrong

Next Time

- Inductive logic programming