Lecture 20

... and last!

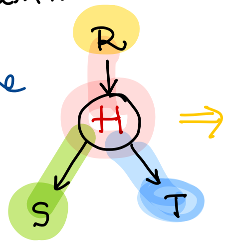
- Prob reasoning

  Q's / Announcements
- Statistics post 391

  Course evals

Probabilistic Reasoning

Ex. Patient + doctor



Structure $\downarrow$ Bayesian network

$\Rightarrow$ $R \perp\!\!\!\perp T \mid H$

$R \perp\!\!\!\perp S \mid H$

$S \perp\!\!\!\perp T \mid H$

directed, acyclic graph (DAG)

(no directed cycle)

unobserved

H = HIV infection $\in \{H, \bar{H}\}$ ← infected

T = HIV test $\in \{T, \bar{T}\}$ ← positive

observed

S = symptoms $\in \{S, \bar{S}\}$

R = risky behavior $\in \{R, \bar{R}\}$

$P(H \mid R) = 0.1$

$P(H \mid \bar{R}) = 10^{-3}$

$P(R) = 10^{-2}$ ← no parent

$P(S \mid H) = 0.8$

$P(S \mid \bar{H}) = 0.1$

$\left.\begin{array}{l} P(T \mid H) = .99 \\ P(T \mid \bar{H}) = 0.03 \end{array}\right\} P_{T \mid \text{"parent"}}$

parameters of B.N. $\{ P_{X \mid \text{parents}(X)}$ for all $X \}$

$S = S_R \times S_H \times S_S \times S_T = \{ (r, h, s, t),\ r \in \{R, \bar{R}\},\ h \in \{H, \bar{H}\} \dots \}$

sample space

$|S| = 2^4$

$P_{\underset{Y}{R}H\underset{X}{ST}}(r, h, s, t) = P_R(r)\, P_{HST \mid R}(h, s, t \mid r)$

$= P_R(r)\, P_{H \mid R}(h \mid r)\, P_{ST \mid HR}(s, t, h, r)$

$= P_R(r)\, P_{H \mid R}(h, r)\, P_{S \mid H\cancel{R}}(s \mid h, r)\, P_{T \mid \cancel{SHR}}(t \mid s, h, r)$

$P_{T \mid H}$

$P_{T \mid HR}$

$P_{X \mid Y} = \dfrac{P_{XY}}{P_Y}$ ← joint, ← marginal

↑ conditional prob.

$P_{XY} = P_Y P_{X \mid Y}$ Chain Rule

$P_{Y \mid X} = \dfrac{P_Y P_{X \mid Y}}{P_X}$

Bayes' Rule

INDEPENDENCE

$X \perp\!\!\!\perp Y \Leftrightarrow$

$P_{XY} = P_X P_Y$

$\boxed{P_{X \mid Y} = P_X}$

$X \perp\!\!\!\perp Y \mid Z \Leftrightarrow$

$P_{XY \mid Z} = P_{X \mid Z} \cdot P_{Y \mid Z}$

$P_{X \mid YZ} = P_{X \mid Z}$

X conditionally indep of Y given Z

1. Gr Model vs Table

\# params in table ($S_R \times S_H \times S_S \times S_T = S$) $\quad |S| = 16$

$16 - 1 = 15$ parameters $= p'$

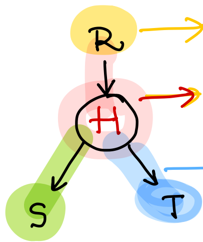
\# —"— Bayes Net $\quad p = 1+2+2+2 = 7 < 15$

1. $p < p' \Rightarrow$ Var (B.N.) lower

But B.N. can have bias !!

2. Data for estimating params

3. Inference more efficient in BN.



$\to P_R \leftarrow$ public health agency

$\to$ medical study

$\to P_{T|H}$ pharma company

$\to$ med study

inference in table

$$P_{H|R,S} = \frac{P_{RHS}}{P_{R,S}}$$

$$\boxed{P_{RHST}(r,h,s,t)} \xrightarrow{\text{Table}} P_{RHS} = \sum_{t=T,\bar{T}} P_{RHST}(\cdot\cdot)$$

V = set of variables

$I \in V$ variables of interest

$O \subset V$ —"— observed

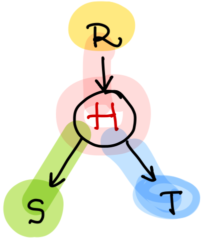
$$P_{I|O} = \frac{P_{IO}}{P_O}$$

$$P_{IO} = \overbrace{\sum_{v_1} \sum_{v_2 \cdots \in V \setminus \{I \cup O\}}}^{2^{|V \setminus \{I \cup O\}|}} P_V(\cdots)$$

$$P_O = \overbrace{\sum_{i_1} \sum_{i_2}}^{2^{|I|} \text{ terms}} P_{IO}(i_1, i_2, \cdot)$$

$I = \{i_1, i_2, --\}$

Assume: all $v \in V$ binary $\Rightarrow n$



Query: $P_{H|R,S,T} = ?$

$r = R$
$s = S$
$t = T$

$$P_{H|R,S,T} = \frac{P_{RHST}}{P_{RST}} = \frac{P_R P_{H|R} P_{S|H} P_{T|H}}{\dots\dots} = .967$$

$$P_{RST} \uparrow \quad P_{RHST}(R,H,S,T) + P_{RHST}(R,\bar{H},S,T) \leftarrow \text{Total Prob}$$

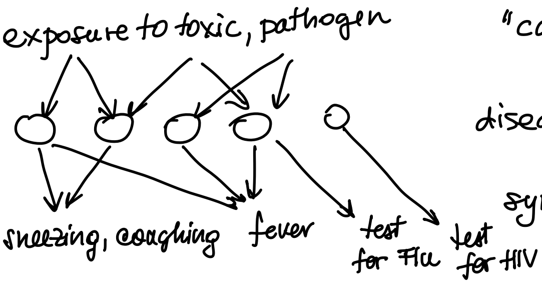
$$P_{H|R,S} = \frac{P_{RHS}}{P_{R,S}} = \frac{P_R P_{H|R} P_{S|H}}{P_R\left(P_{H|R}(H|R)P_{S|H}(S|H) + P_{H|R}(\bar{H}|R)P_{S|H}(S|\bar{H})\right)} \quad \sim 6 \text{ op}$$

Rule 1 numerator = 1 term, denominator = sum

Rule 2 $\rightarrow$ keep them $\exists$ table
$\leftarrow$ use Bayes Rule

T.B. CONTINUED...

a real database



"causes" $\sim 10^2$

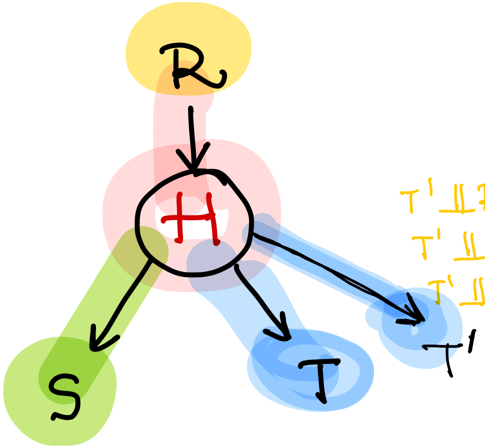
diseases $\sim 10^{2-3}$

symptoms $\sim 10^3$

Alarm network

ICU patient on respirator

variables 42



$T' \perp\!\!\!\perp R \mid H$

$T' \perp\!\!\!\perp T \mid H$

$T' \perp\!\!\!\perp S \mid H$

Observed $R, S, T, \bar{T}'$

wanted $H \rightarrow P_{H|R,S,T,\bar{T}'}$

$$P_{H|RST\bar{T}'} = \frac{P_{RHST\bar{T}'}}{P_{RST\bar{T}'}} = 0.23$$

$$P_{RHST\bar{T}'} = P_R \cdot P_{H|R} \cdot P_{T|H} P_{S|H} P_{\bar{T}'|H}$$

$$P_{RST\bar{T}'} = P_{RHST\bar{T}'} + P_{R\bar{H}ST\bar{T}'}$$

wanted $P_R = ?$ ← read out from P_R table

table P_{RHST} ⇒ $P_R = \sum_{T,\bar{T}} \sum_{S,\bar{S}} \sum_{H,\bar{H}} P_{RHST}(r,h,s,t)$

wanted $P_H = ?$ ← table —"— ⇒ same # operations

$$P_H = P_R P_{H|R} + P_{\bar{R}} P_{H|\bar{R}}$$

$$= \sum_{r} \sum_{s} \sum_{t} P_{RHST}(r,h,s,t)$$

wanted $P_S = ?$ ← table : same # operations

$$P_S = P_{S|H} P_H + P_{S|\bar{H}} P_{\bar{H}} = P_{S|H} P_H + P_{S|\bar{H}} (1 - P_H)$$

Past exams ← TBP (Friday)

Points for participation ← —"—

hw ← (HW7 by Tue?)

OH times ← TBP (Fri)

Grading alg ✓

Proportions: HW ≈ 47%

Q ≈ 13%

P = 12%

Ex = 28%

100%

EXACT FORMULA

$$\frac{\sum h_j W_H + \sum g_j W_Q}{\sum_j h_j^0 \cdot W_H + \sum g_j^0 \cdot W_Q}$$

$W_Q \sim 13$

$W_H \sim 47$

$$\cancel{\frac{\sum h_j}{\sum h_j^0} W_H + \frac{\sum g_j}{\sum g_j^0} W_Q}$$

Exam logistics: ✓

Topics: ← TBP (Friday)

* Tips for studying
* Topics