Today : Bayes Nets

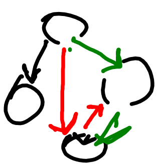
- Inference : Given BN, Data, find likelihood of value of some node

  How difficult is Exact Inference

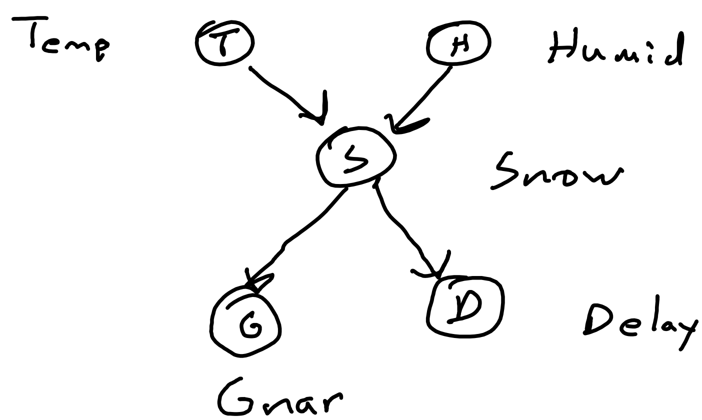
  What are best approx inference algs

- Learning From Data : Given Data

  What BN Explains Best



  - Parameter
  - Structure



Have data for G, D, T, want to infer H

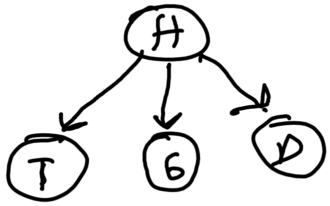
$$P(h \mid g\ d\ t) = \frac{P(h, g, d, t)}{P(g, d, t)} \leftarrow$$

$$\propto P(h, g, d, t)$$

Naive Bayes

$$p(g,d,t) = \sum_h P(h,g,d,t)$$

No delay
No Gran
Cold Temp



NB → small probability of high humidity

Exact Inference → humidity is definitely

$$P(h,g,d,t) = P(h) \cdot \underbrace{P(g|h)P(d|h)P(t|h)}_{\text{Computationally Simple}}$$

Does not take correlation into account

Exact Inference

Have G, D, infer H

$$P(h,g,d) = \sum_t \sum_s P(h,t,s,g,d)$$

$$= \sum_t \sum_s P(h) P(t) P(s|h,t) P(g|s) P(d|s)$$

Exact Answer

Number of additions can be exponential in number of H.V.s

Variable Elimination

$T_1(T)$, $T_2(H)$, $\cancel{T_3(T,H,S)}$, $\cancel{T_4(S,G)}$, $\cancel{T_5(S,D)}$

have G, D

$T_6(S)$, $T_7(S)$

$$T_8(H,T) = \sum_s T_3(T,H,s)\, T_6(s)\, T_7(s)$$

$$T_9(H) = \sum_t T_8(H,t)\, T_1(t)$$

$P(H|g,d)$ $T_9(H)$, $T_2(H)$ multiply together + normalize

What order to eliminate variables?

NP-hard

3SAT - NP-complete

3SAT can be expressed as BN inference

$\therefore$ Exact inference in BN is NP-Hard

Approximate Inference - use sample data

T	H	S	G	D	
0	0	1	1	~~1~~	w
0	0	0	0	0	
1	0	1	1	1	←
0	0	0	0	1	
0	1	1	1	1	←

$G=1$, $D=1$

$P(H=1) = 0.5$

How to generate?

Option 1 : Direct Sampling

Unweighted Particle Filtering

Topological Sort

Sample from each node

Option 2 : Likelihood Weighted

Weighted Particle Filtering

$X_{1:n} \leftarrow$ Topological Sort

$w \leftarrow 1$

for i in $1..n$

 if $o_i =$ missing

  $x_i \leftarrow$ sample from $P(x_i \mid pa_{x_i})$

 else

  $x_i \leftarrow o_i$

  $w \leftarrow w \cdot P(o_i \mid pa_{x_i})$

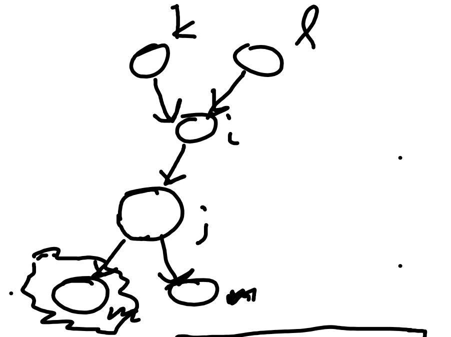
return $(x_{1:n}, w)$

Option 3: Gibbs Sampling

repeat $x \leftarrow x'$

$X_{1:n} \leftarrow$ any ordering

$x'_{1:n} = x_{1:n}$ ← previous sample



for i in 1:n

if o_i = missing

$x'_i \leftarrow$ random sample from $\underline{P(X_i \mid x'_{1:n \setminus i})}$

Markov Blanket of i

else

$\underline{x'_i \leftarrow o_i}$

return $x'_{1:n}$

MCMC

Watch out for correlation

Thinning, accept every nth sample

Learning: Given Data what Bayes Net

"Book of Why"

Parameter Learning

θ_s parameters for $P(S \mid T, H)$

n trials

m positive $X = 1$

$\hat{\theta} =$ m·n m−n m/n ?

0	m=2
1	n=6
0	$\frac{2}{6} = \frac{1}{3}$
0	
0	
1	

$P(x=1) = \theta$

$$\hat{\theta} = \underset{\theta}{\arg\max}\, P(D \mid \theta)$$

$$P(D \mid \theta) = \frac{n!}{m!(n-m)!} \underbrace{\theta^m (1-\theta)^{n-m}}$$

$$\propto \theta^m (1-\theta)^{n-m}$$

$$\ell(\theta) = \ln\left(\theta^{m}(1-\theta)^{n-m}\right)$$

log likelihood

$$= m\ln\theta + (n-m)\ln(1-\theta)$$

$$\frac{\partial \ell(\theta)}{d\theta} = \frac{m}{\theta} - \frac{n-m}{\theta} = 0$$

$$\therefore \hat{\theta} = \frac{m}{n}$$

For normal distribution, $N(\mu, \sigma^2)$

$$\hat{\mu} = \frac{\Sigma v_i}{n}$$

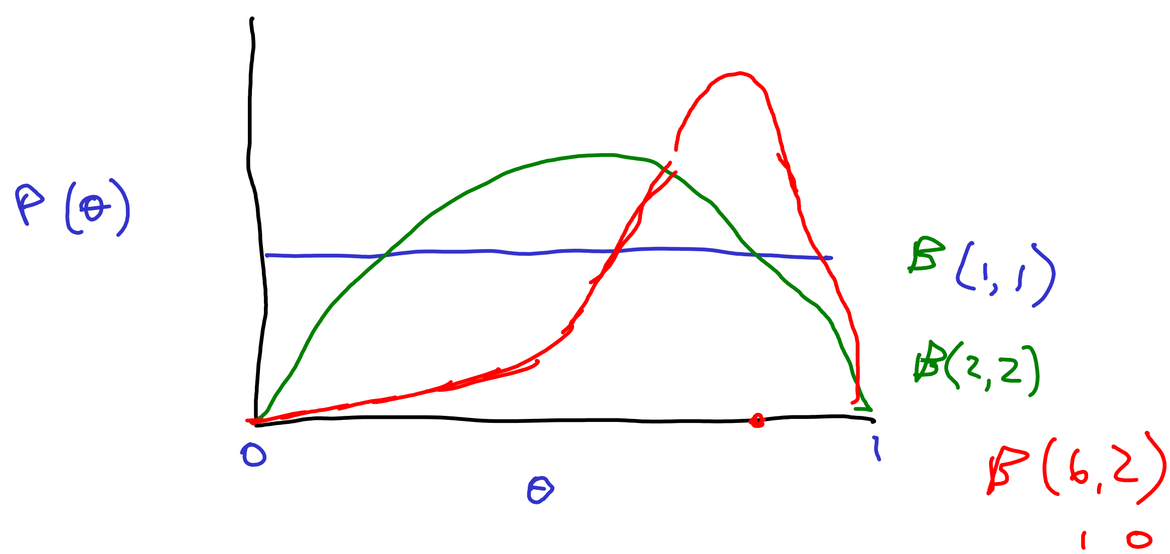
$$\therefore \hat{\sigma}^2 = \frac{\Sigma (v_i - \hat{\mu})^2}{n}$$

What if we don't have much data?

Bayesian Parameter Leaning

Assume a Prior distribution

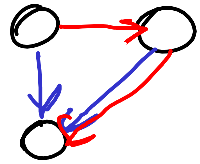
Beta



Nonparametric Learning

$$p(x) = \frac{1}{n}\sum_{i=1}^{n} K(x - o_i)$$

Kernel Function



$$P(G|D) \propto P(G)\,P(D|G)$$
$$= P(G)\int P(D|\theta, G)\,p(\theta|G)\,d\theta$$

$q_i \leftarrow$ instantiations of parents

Look up in Book $\rightarrow$

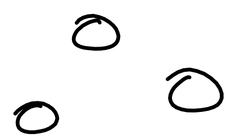
$$P(G|D) = P(G)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij0})}{\Gamma(\alpha_{ij0}+m_{ij0})}\prod^{r_i}\frac{\Gamma(\alpha_{m}\ m)}{\Gamma(\alpha_{m})}$$

$$\ln P(G|D) = \underbrace{\ln(G)}_{\text{prior uniform}} + \underbrace{\sum\sum \ln\frac{\Gamma(\alpha)}{\Gamma(\alpha+m)} + \sum \ln\frac{\Gamma(\alpha+m)}{\Gamma(\alpha}}_{\text{Data}}$$

Bayesian Score.

Can't enumerate all DAGs efficiently

KZ
- start with no edges
- add edges that locally maximize score



local optimization

start with some graph

repeat

$G \leftarrow G'$ in neighborhood of G that maximizes Bayesian Score

- introducing new edge
- removing edge
- reversing edge