

W06 - Probabilistic Models and Inference

$D = \text{data}$

$H, X = \text{hypothesis and parameters} = M = \{X, H\}$

$$P(X | D, H) \propto \underbrace{P(D | X, H)}_{\cdot \text{Prob of data}} \cdot P(X | H)$$

- Prob of data
given X, H
- Likelihood

Until now we have considered cases such that

$D \sim \text{one or two random variables}$

$M - \text{single}$

$$\begin{aligned} &+ \text{exponential distribution } D=T, \quad P(t_i | M) = \frac{1}{\lambda} e^{-t_i/\lambda}, \\ &\quad M \sim \lambda \end{aligned}$$

$$\begin{aligned} &+ \text{linear regression fit } D=F_1 G \quad P(g_i^i, f_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(f_i - a - b g_i)^2}{2\sigma^2}} \\ &\quad M \sim a, b, \sigma \end{aligned}$$

Now more complicated cases:

$D = \{R_1, \dots, R_n\} \text{ many variables}$

$d_i = R_1^i, \dots, R_m^i \quad \# \text{data point}$

Then to calculate

$$P(R_1^1, \dots, R_M^1, R_1^2, \dots, R_N^2, \dots, R_1^n, \dots, R_N^n | M) = ?$$

is complicated $\uparrow \text{# of variables per data point.}$

How to express complicated relationships btw
many random variables?

DAG = directed acyclic graphs
Probabilistic graphical models (PGM)
Bayesian Networks (BN)

DAG nodes (N) each node represents a
 edges $(N_1 \rightarrow N_2)$ random variable
 between nodes

E = enhancer transcription level (trans/not trans)

G = gene activated by E

P = protein product



$$P(E, G, D) = \begin{cases} P(E|G, D) P(G, D) \\ = P(E|G, D) \cdot P(G|D) \cdot P(D) \\ = P(D|G, E) P(G|E) P(E) \\ \vdots \\ = P(D|G) P(G|E) \cdot P(E) \end{cases}$$

*For this
DAG only*

$$P(TF_1, TF_2, h) = \\ = P(h|TF_1, TF_2) \cdot P(TF_1) P(TF_2)$$

$$P(TF, G_1, G_2) = P(G_1|TF) \\ P(G_2|TF) \\ P(TF)$$

DAG gives us a restricted form of the joint distributions of all random variables ~ causality [Next Friday]

Example from Koller + Friedman (book available as pdf)

recommendation letter

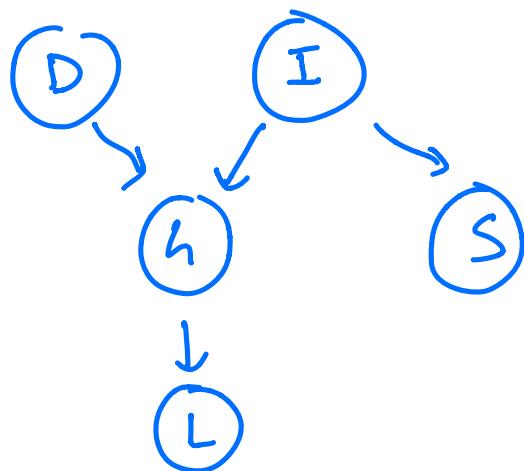
L = strength of letter

D = difficulty of courses

G = grades

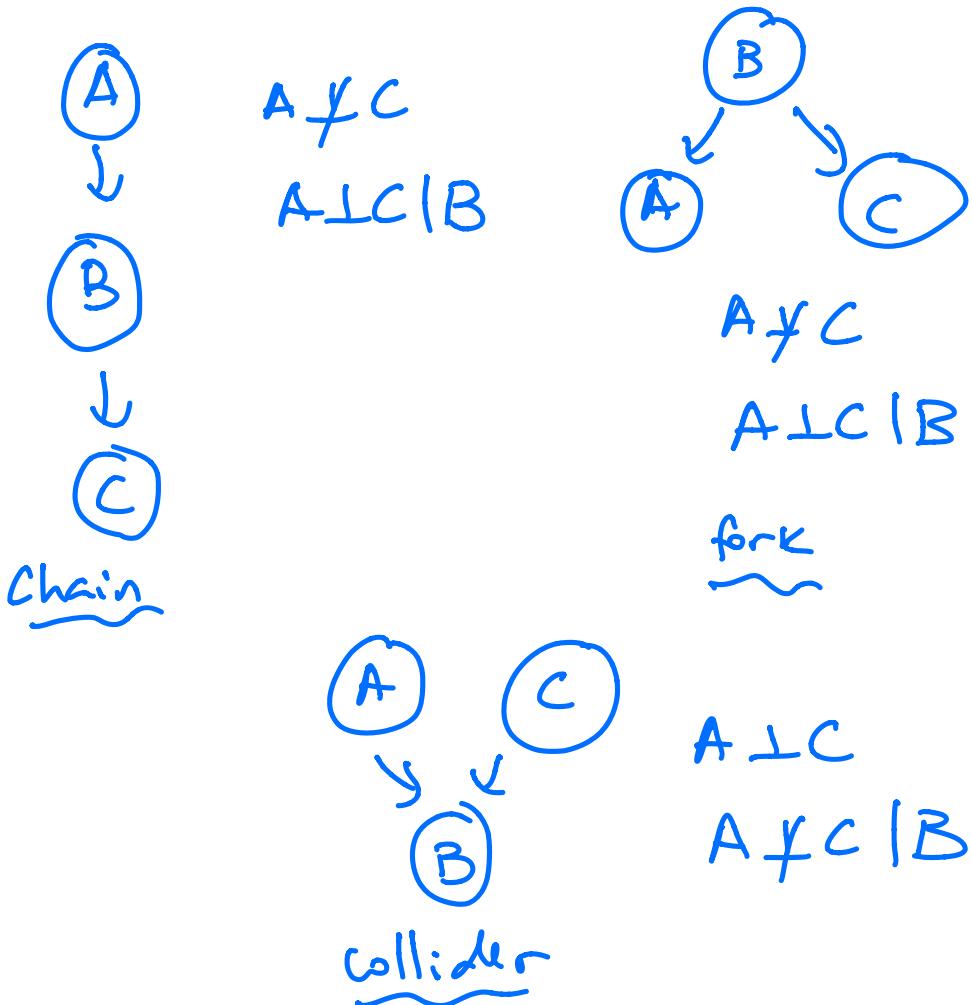
SAT = SAT scores

I = intelligence



$$\begin{aligned} P(L|GDISS) &= P(L|GDI) \cdot P(GDISS) \\ &= P(L|GDI) P(G|DI) P(DI) P(S|DI) \\ &= P(L|GDI) \cdot P(G|DI) \\ &\quad P(S|DI) \cdot P(DI) \cdot P(I) \\ \text{always true} \\ (\text{general}) &= P(L|GDI) \cdot P(G|DI) \\ &\quad P(S|DI) \cdot P(DI) \cdot P(I) \\ \text{for this DAG} &= P(L|G) \cdot P(G|DI) P(S|I) \cdot P(D) \cdot P(I) \end{aligned}$$

- DAGs give us the causal relationship b/w variables
Judea Pearl (CI prior)
- It also give us an efficient way to describe P(D|M)
- Three basic architectures

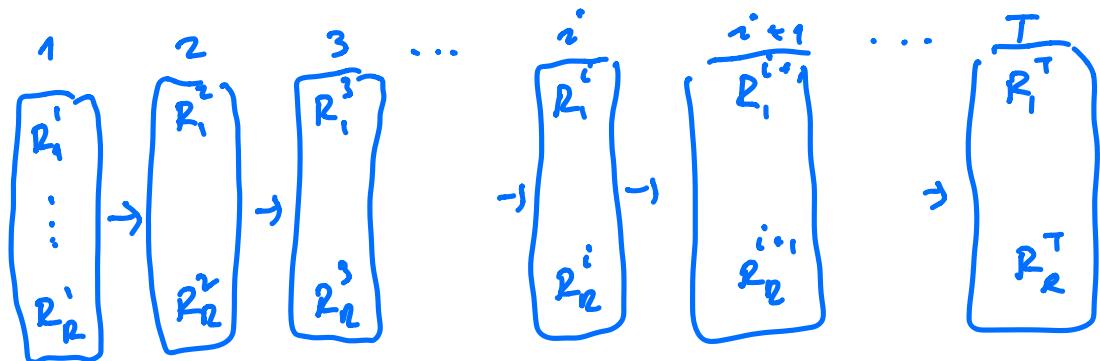


DAG = {chains, colliders, forks}

Hidden Markov Models (HMMs)

a particular type of DAGs.

"a series" in time, location,



$$P(R_1^1, \dots, R_n^1, R_1^2, \dots, R_n^2, \dots, R_1^i, \dots, R_n^i, R_1^{i+1}, \dots, R_n^{i+1}, \dots, R_1^T, \dots, R_n^T)$$

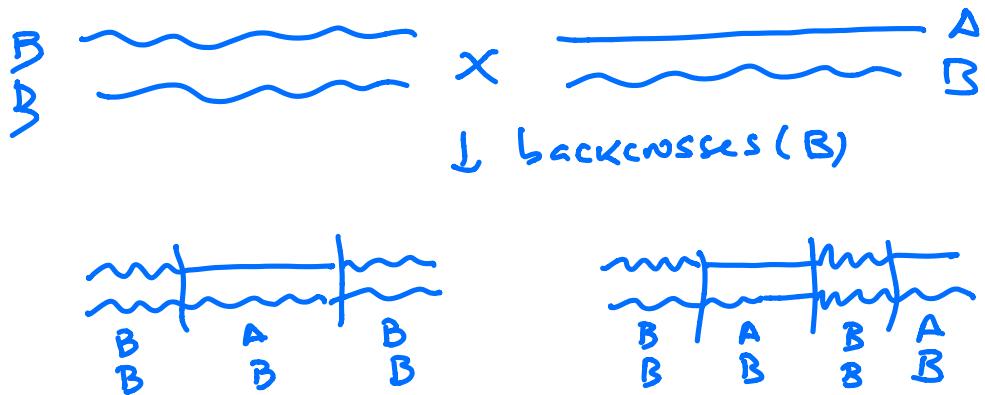
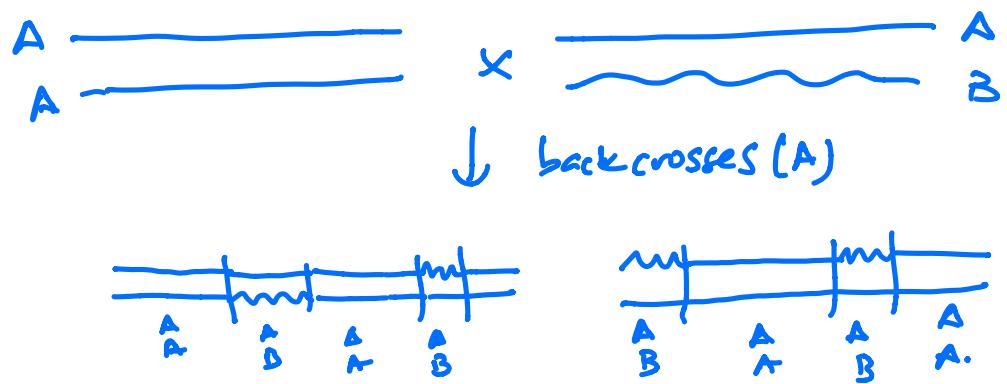
HMM if $\underbrace{P(R_1^1, \dots, R_n^1 | R_1^2, \dots, R_n^2)}_{\text{Markov.}}$

$$P(R_1^1, \dots, R_n^1, \dots, R_1^N, \dots, R_n^N) = P(R_1^1, \dots, R_n^1 | R_1^2, \dots, R_n^2) \\ P(R_1^2, \dots, R_n^2 | R_1^3, \dots, R_n^3) \dots P(R_1^N, \dots, R_n^N | R_1^1, \dots, R_n^1)$$

$$P(R_1^2, \dots, R_n^2 | R_1^1, \dots, R_n^1)$$

$$P(R_1^1, \dots, R_n^1) \quad (\text{special case})$$

Our biological example



Experiment : Take backcrosses ♂ flies
do RNA-seq.

	A genome	<u> </u> q q q	
(known)	B genome	<u> </u> m m m m m m m m	
($\begin{matrix} AAB \\ ABB \end{matrix}$)	Z	B A $\begin{matrix} ABA \\ BBB \end{matrix}$? ?	
ancestor		<u> </u> q q q	
G		<u> </u> c c ?	
genotype			.
[unclear]			
X		a a a	
reads		c c c	
(known variants)		c c t	
			c g c

for each position $i = 1, \dots, L$

$$x^i = (n_k^i, n_c^i, n_S^i, n_t^i)$$

$$h^i \in \left(\begin{matrix} a & a & a & a & c & c & c & g & g & t \\ a & c & g & t & c & s & t & g & t & c \end{matrix} \right) \text{ (10 poss)}$$

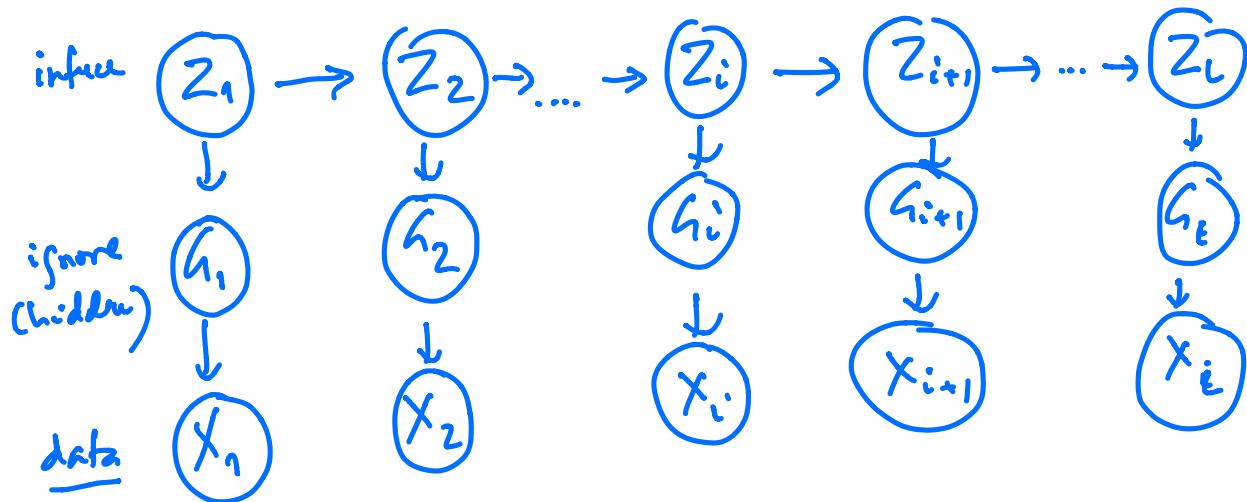
$Z^i \notin \begin{matrix} A & A & B \\ B, A, B \end{matrix}$ 3 possibility

x^i = known surely

h^i = unknown / don't care

Z^i = unknown, want to do inference on.

Model? an HMM.



$$P(x_1 h_1 z_1, x_2 h_2 z_2, \dots, x_L h_L z_L) =$$

$$= P(x_L h_L z_L | x_{L-1}, h_{L-1}, z_{L-1})$$

$$P(x_{L-1}, h_{L-1}, z_{L-1} | x_{L-2}, h_{L-2}, z_{L-2})$$

⋮

$$P(x_2 h_2 z_2 | x_1 h_1 z_1)$$

$$P(x_1 h_1 z_1)$$

$$P(x_i h_i z_i | x_{i-1}, h_{i-1}, z_{i-1}) =$$

$$= P(z_i | z_{i-1}) P(h_i | z_i) P(x_i | h_i)$$

The Model parameters ?

$$\underbrace{P(z_i | z_{i-1})}$$

$$P(A|A) = \text{contine} = p$$

$$P(B|A) = \text{break} = 1-p$$

$$P(B|B) = \text{double break} = 0$$

$$\overline{P(B|B)} = \text{only} = p$$

$$P(A|B) = \text{break} = 1-p$$

$$P(A|B) = 0$$

$$P_A + P_B = 1-p$$

(P)? now to set n

$$\underbrace{\begin{array}{c} A \\ A \end{array} \dots \begin{array}{cc} A & A \\ A & B \end{array}}_n \quad p^n(1-p) \rightarrow \text{geometric dist} \\ \langle n \rangle = \frac{p}{1-p} = \frac{L}{2} \quad \text{if 1 break expected}$$

$$p = (1-p)^{L/2}$$

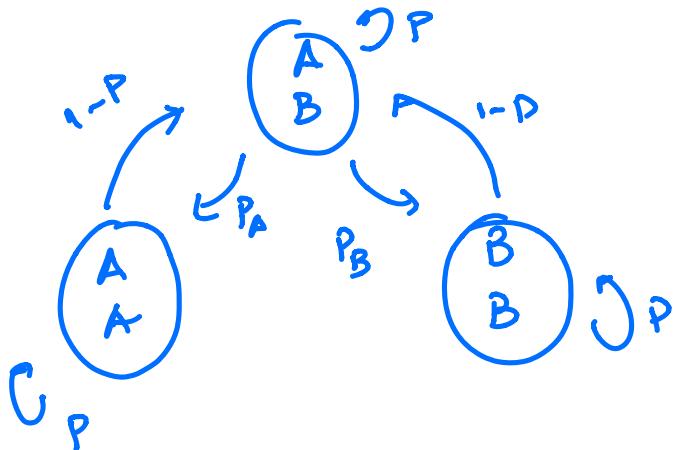
$$P(L+L/2) = L/2 \quad \hat{p} = \frac{L/2}{1+L/2} = \frac{L}{L+2}$$

Maximum likelihood rule of P

$$P_{ML} = \frac{L/2}{1+L/2} \quad \text{if you expect 1 break, at about } L/2$$

A graphical representation of the states

(not a dag)



Inference

$$P(D|HMM) = P(X_1 \dots X_L | P, HMM)$$

$$= \sum_{G_1 Z_1} \dots \sum_{G_L Z_L} P(X_1 G_1 Z_1 \dots X_L G_L Z_L | P, HMM)$$

$(3 \cdot 10)^L$ terms !

Next : an efficient algorithm to calculate
 $P(D | HMM)$ in $\Theta(L)$ [1 loop]

these HMM algorithms are called dynamic programs.

Richard Bellman (1954)

. then $P(Z_i | \text{Data}, HMM)$