

W06 - Probabilistic Models and Inference

$D = \text{data}$

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

$$P(X | D, H) \propto \underbrace{P(D | X, H)} \cdot P(X | H)$$

- Prob of data given X, H
- Likelihood

Until now we have considered cases such that

$D \sim$ one or two random variables

M - single

+ bacterial mutation $D = T$, $M \sim \lambda$

$$P(t_i | M) = \frac{1}{\lambda} e^{-t_i/\lambda}$$

+ linear regression fit $D = F, G$, $M \sim a, b, \sigma$

$$P(g_i, f_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(f_i - a - b g_i)^2}{2\sigma^2}}$$

Now more complicated cases:

$D = \{R_1, \dots, R_M\}$ many variables

$$d_i = R_1^i \dots R_M^i$$

Then to calculate

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

is complicated

→ # data points

↑ # 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 random variable
edges $(N_1) \rightarrow (N_2)$ between nodes

E = enhancer transcription level (trans/not transcribed)

G = gene activated by E

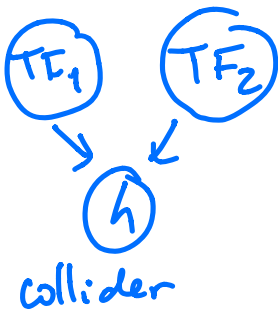
P = protein product



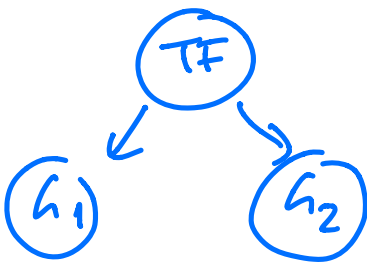


$$\begin{aligned}
 P(E, G, D) &= P(E | G, D) P(G, D) \\
 &= P(E | G, D) \cdot P(G | D) \cdot P(D) \\
 \exists! & \\
 &= P(D | G, E) P(G | E) P(E) \\
 &= P(D | G) P(G | E) \cdot P(E)
 \end{aligned}$$

For this DAG only



$$\begin{aligned}
 P(TF_1, TF_2, G) &= \\
 &= P(G | TF_1, TF_2) \cdot P(TF_1) P(TF_2)
 \end{aligned}$$



$$\begin{aligned}
 P(TF, G_1, G_2) &= P(G_1 | TF) \\
 &P(G_2 | TF) \\
 &P(TF)
 \end{aligned}$$

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

Example for Koller + Friedman (book available as pdf)

recommendation letter

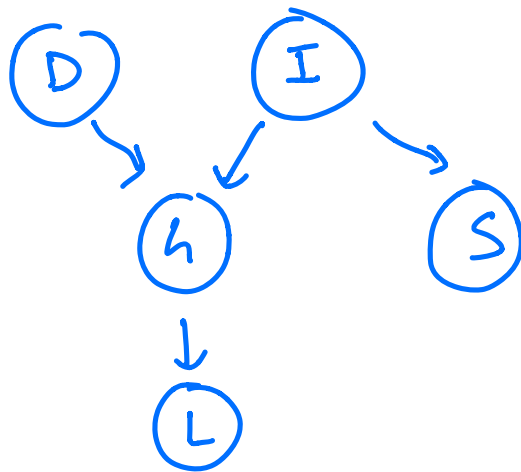
L = strength of letter

D = difficulty of courses

G = grades

SAT = SAT scores

I = intelligence



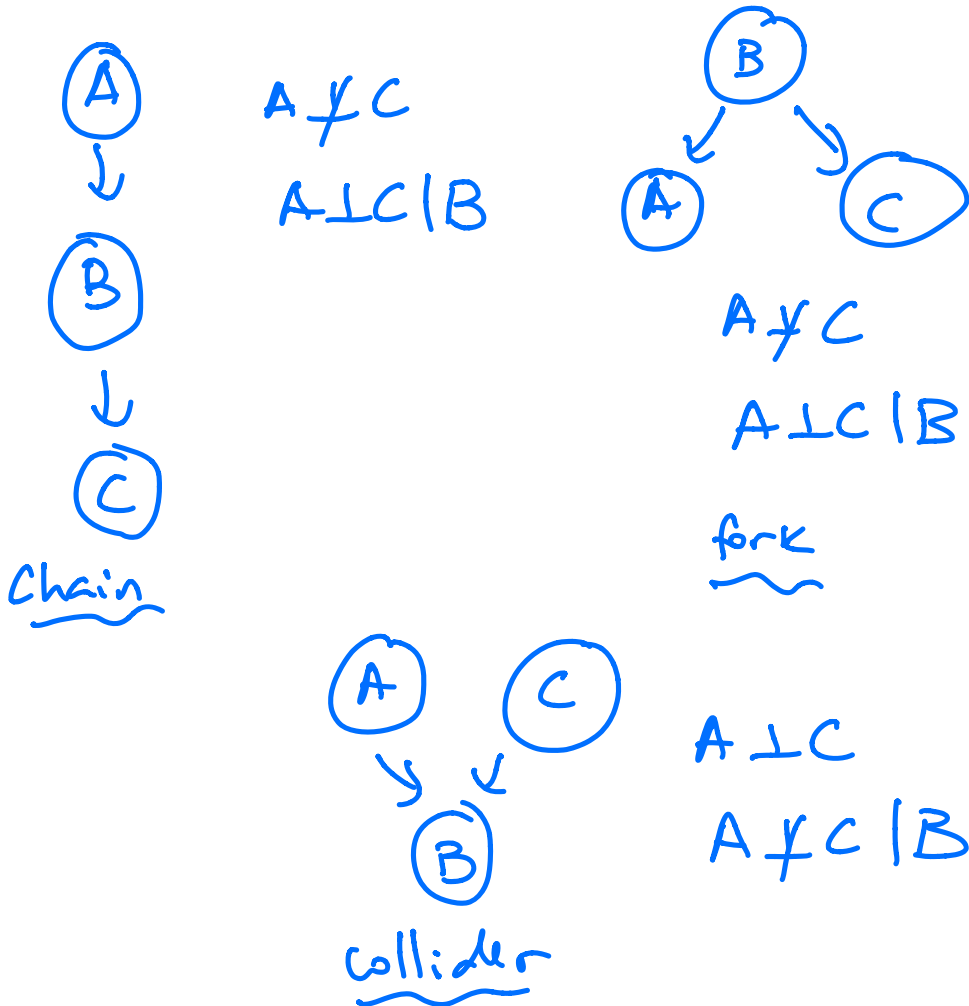
$$\begin{aligned}
 P(LGDIS) &= P(L|GDIS) \cdot P(GDIS) \\
 &= P(L|GDIS) P(G|DIS) P(DIS) \\
 &= P(L|GDIS) \cdot P(G|DIS) \\
 &\quad P(S|DI) \cdot P(DI)
 \end{aligned}$$

always true
(general)

$$\begin{aligned}
 &= P(L|G) \cdot P(G|DI) \\
 &\quad P(S|DI) \cdot P(D|I) \cdot P(I)
 \end{aligned}$$

for this DAG = $P(L|G) \cdot P(G|DI) P(S|DI) \cdot P(D) \cdot P(I)$

- DAGs give us the causal relationship btw variables
 Judea Pearl (CI mer)
- It also give us an efficient way to describe $P(DIM)$
- Three basic architectures

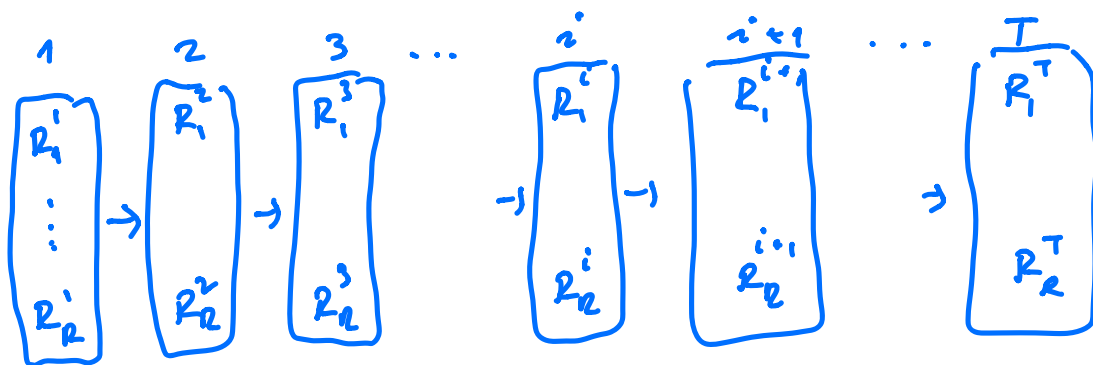


DAG = { chains, colliders, forks }

Hidden Markov Models (HMMs)

a particular type of DAGs.

"a series" in time, location, ...



$$P(R_1^1 \dots R_2^1, R_1^2 \dots R_2^2, \dots, R_1^T \dots R_2^T)$$

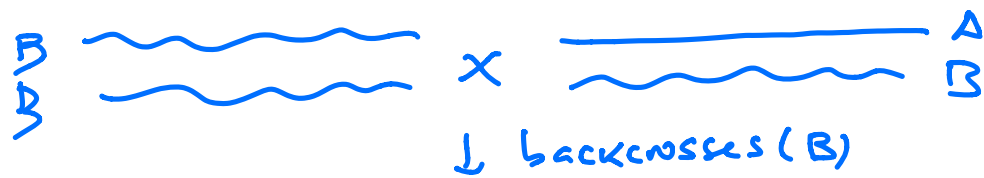
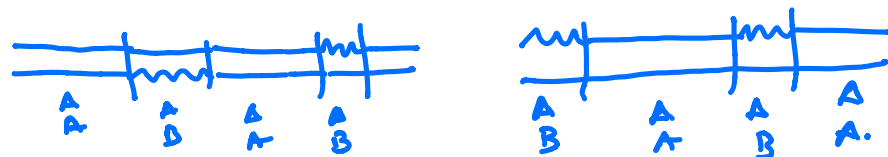
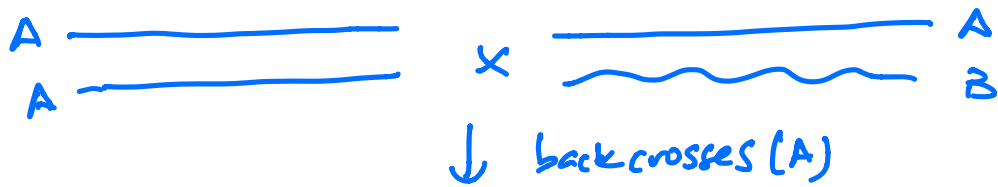
HMM if $P(R_1^i \dots R_2^i | R_1^{i-1} \dots R_2^{i-1})$ Markov.

$$P(R_1^1 \dots R_2^1 \dots R_1^N \dots R_2^N) = P(R_1^T \dots R_2^T | R_1^{T-1} \dots R_2^{T-1}) \\ P(R_1^{T-1} \dots R_2^{T-1} | R_1^{T-2} \dots R_2^{T-2})$$

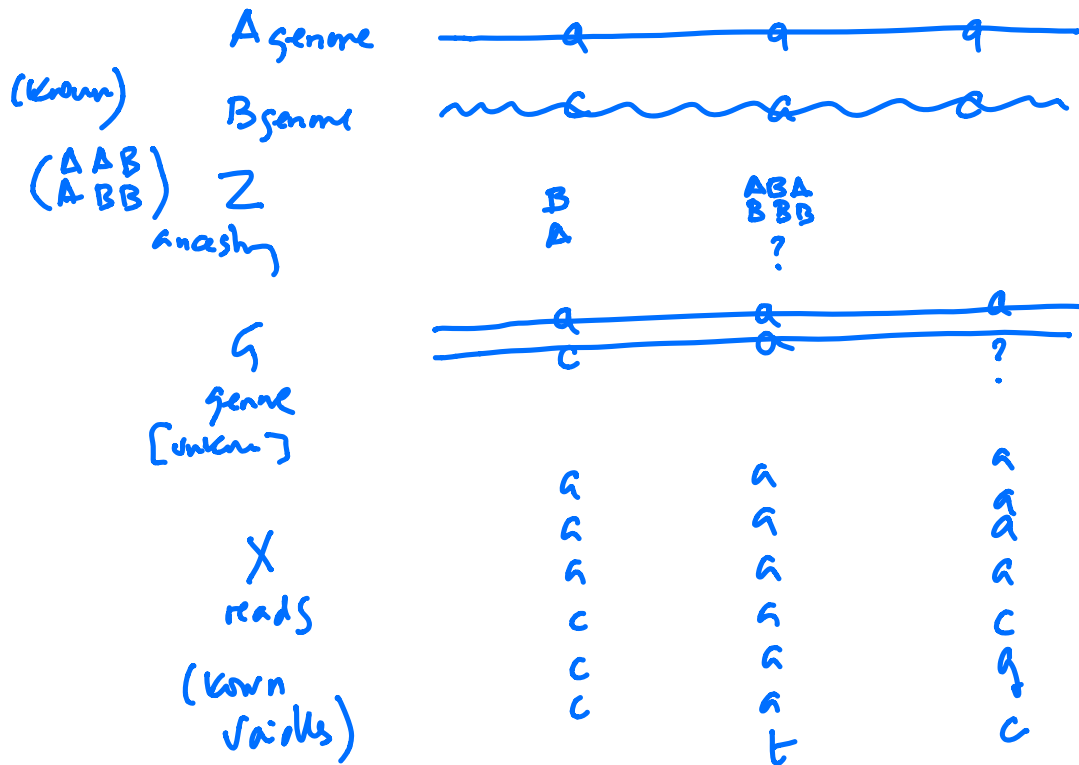
$$P(R_1^2 \dots R_2^2 | R_1^1 \dots R_2^1)$$

$$P(R_1^1 \dots R_2^1) \text{ (special case)}$$

Our biological example



Experiment: Take backcrosses of ♂ flies
do RNA-seq.



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

$$x^i = (n_a^i, n_c^i, n_g^i, n_t^i)$$

$$h^i \in \begin{pmatrix} a & a & a & a & c & c & c & g & g & t \\ a & c & g & t & c & g & t & g & t & t \end{pmatrix} \quad (10 \text{ pairs})$$

$$z^i \in \begin{matrix} A & A & B \\ B & A & B \end{matrix}$$

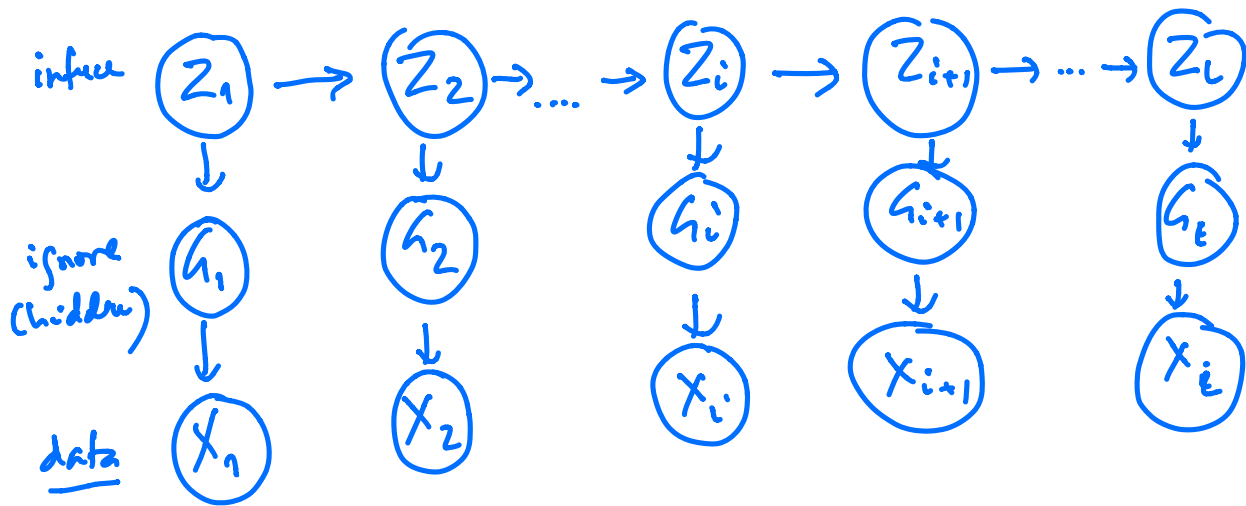
3 possibilities

x^i = known variables

h^i = unknown / don't care

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

Model? an HMM.



$$\begin{aligned}
 P(X_1, h_1, Z_1, X_2, h_2, Z_2, \dots, X_L, h_L, Z_L) &= \\
 &= P(X_L, h_L, Z_L \mid X_{L-1}, h_{L-1}, Z_{L-1}) \\
 &\quad P(X_{L-1}, h_{L-1}, Z_{L-1} \mid X_{L-2}, h_{L-2}, Z_{L-2}) \\
 &\quad \vdots \\
 &\quad P(X_2, h_2, Z_2 \mid X_1, h_1, Z_1) \\
 &\quad P(X_1, h_1, Z_1)
 \end{aligned}$$

$$\begin{aligned}
 P(X_i, h_i, Z_i \mid X_{i-1}, h_{i-1}, Z_{i-1}) &= \\
 &= P(Z_i \mid Z_{i-1}) P(h_i \mid Z_i) P(X_i \mid h_i)
 \end{aligned}$$

The Model parameters ?

$$\underline{P(Z_i | Z_{i-1})}$$

$$P\left(\begin{matrix} A \\ B \end{matrix} \middle| \begin{matrix} A \\ B \end{matrix}\right) = P$$

$$P\left(\begin{matrix} A \\ A \end{matrix} \middle| \begin{matrix} A \\ B \end{matrix}\right) = P_A$$

$$P\left(\begin{matrix} B \\ B \end{matrix} \middle| \begin{matrix} A \\ B \end{matrix}\right) = P_B$$

$$P_A + P_B = 1 - P$$

$$P\left(\begin{matrix} A \\ A \end{matrix} \middle| \begin{matrix} A \\ A \end{matrix}\right) = \text{continue} = P$$

$$P\left(\begin{matrix} A \\ B \end{matrix} \middle| \begin{matrix} A \\ A \end{matrix}\right) = \text{break} = 1 - P$$

$$P\left(\begin{matrix} B \\ B \end{matrix} \middle| \begin{matrix} A \\ A \end{matrix}\right) = \text{double break} = 0$$

$$P\left(\begin{matrix} B \\ B \end{matrix} \middle| \begin{matrix} B \\ B \end{matrix}\right) = \text{only} = P$$

$$P\left(\begin{matrix} A \\ B \end{matrix} \middle| \begin{matrix} B \\ B \end{matrix}\right) = \text{break} = 1 - P$$

$$P\left(\begin{matrix} A \\ A \end{matrix} \middle| \begin{matrix} B \\ B \end{matrix}\right) = 0$$

Ⓟ ? how to set it



$P^n (1-P) \rightarrow$ geometric dist

$$\langle n \rangle = \frac{P}{1-P} = \frac{L}{2} \quad \text{if 1 break expected}$$

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

$$P(1 + L/2) = L/2$$

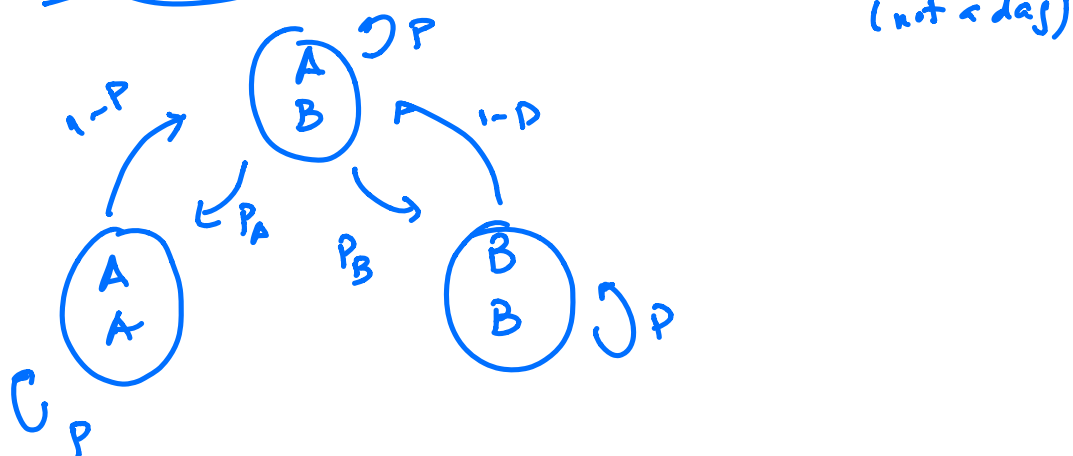
$$P^* = \frac{L/2}{1 + L/2} = \frac{L}{L+2}$$

Maximum likelihood value of P

$$P_{ML} = \frac{L/2}{1 + L/2}$$

if you expect 1 break, at about $L/2$

A graphical representation of the states



Inference

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

$$= \sum_{z_1, z_2} \dots \sum_{z_L, z_L} P(X_1, z_1, \dots, X_L, z_L | P, HMM)$$

$(3.10)^L$ terms!

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

these HMM algorithms are called dynamic programming.

Richard Bellman (1954)

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