Bayesian inference:

Phylogenetic tree = Evolutionary tree
= tree showing evolutionary relationship among species.

Consider n species, with aligned sequences of length N.

For $D_i \in \{0,1\}^{\cancel{N}\,n}$, let $\vec{D} = (D_1, \ldots, D_N)$.

(More generally, $\cancel{D_i \in \{0,1\}^N}$ $D_i \in \{A,C,G,T\}^N$ is a DNA sequence)

Let Ω = phylogenetic trees on n leaves
= trees with internal degree 3 & n leaves.

For $T \in \Omega$, let V_{ext} = leaves (there are n leaves)
& V_{in} = internal vertices.

Each edge in T has a mutation probability $p \in [0, \frac{1}{2}]$
for $e \in E(T)$, let $p(e)$ be its probability.
& $\vec{p} = (p(e))_{e \in E}$

(More generally, can replace $p(e)$ by a 4×4 matrix.)

Given a tree T & probabilities $\vec{p}$ on edges, what's the probability of generating this data $\vec{D}$?

Consider $D_i \in \vec{D}$ so assignment of $\{0,1\}$ to each leaf.

$$D_i : V_{ext} \to \{0,1\}.$$

~~Each~~ An arbitrary internal node chooses a random spin $\{0,1\}$ & then every edge e:

flips with prob. $p(e)$

& stays the same with $1-p(e)$.

Hence,

$$\Pr(D_i \mid T, \vec{p}) = \frac{1}{2} \sum_{\substack{\hat{D}_i \in \{0,1\}^V: \\ \hat{D}_i(V_{ext}) = D_i(V_{ext})}} \prod_{\substack{e=(y,z)\in E(T): \\ \hat{D}_i(y)=\hat{D}_i(z)}} (1-p(e)) \prod_{\substack{e: \\ \hat{D}_i(y)\neq \hat{D}_i(z)}} p(e).$$

Then,

$$\Pr(\vec{D} \mid T, \vec{p}) = \prod_{i=1}^{N} \Pr(D_i \mid T, \vec{p})$$

$$= \exp\left(\sum_{i=1}^{N} \log(\Pr(D_i \mid T, \vec{p})) \right)$$

Bayes theorem:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

$$= \frac{\Pr(A \cap B)}{\Pr(B)}$$

Hence,

$$\Pr(T|\vec{D}) = \frac{\Pr(\vec{D}|T)\,\Pr(T)}{\Pr(\vec{D})}$$

likelihood

prior

normalizing factor

When $p(e)=0$ then no mutations

& $p(e)=\frac{1}{2}$ then endpoints are independent of each other.

Bayes rule implies:

Posterior Probability:
$$\Pr(T|\vec{D}) = \frac{\int_{\vec{p}} \Pr(\vec{D}|T,\vec{p})\,\psi(T,\vec{p})\,dp}{\Pr(\vec{D})}$$

$$= \frac{\int_{\vec{p}} \Pr(\vec{D}|T,\vec{p})\,\psi(T,\vec{p})\,dp}{\sum_{T'} \int_{\vec{p}} \Pr(\vec{D}|T',\vec{p})\,\psi(T',p)\,dp}$$

For tree T, let $\omega(T) = \int_{\vec{p}} \Pr(\vec{D}|T,\vec{p})\,\psi(T,\vec{p})\,dp$

Note, $\psi(T,\vec{p})$ is a prior, & we need that

$$\sum_{T'} \int_{\vec{p}} \psi(T',\vec{p})\,dp = 1$$

For uniform priors, $\psi(T',\vec{p}') = \psi(T,\vec{p})$ then:

$$\Pr(T|\vec{D}) = \frac{\int_{\vec{p}} \Pr(\vec{D}|T,\vec{p})\,dp}{\Pr(\vec{D})} = \frac{\Pr(\vec{D}|T)}{\Pr(\vec{D})}$$

So for uniform priors,

$$\underset{\Pr(T \mid \vec{D})}{\text{Posterior Distribution}} = \underset{\Pr(\vec{D} \mid T)}{\text{likelihoods}}$$

Hence, sampling from posterior distribution is more general than sampling proportional to likelihoods

What priors? MrBayes allows uniform or exponential distributions for priors.

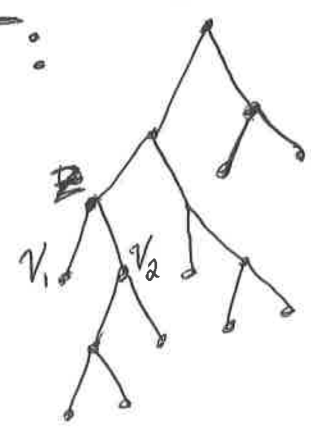
For ψ with a simple formula,

$$w(T) = \int_{\vec{p}} \Pr(\vec{D} \mid T, \vec{p})\, \psi(T, \vec{p})\, dp$$

can be computed efficiently via Dynamic Programming.

DP idea:

Consider a tree **T**:
(root it somewhere)



For $v \in V$,
let T_v = tree hanging below v.

For $v \in V$, $s \in \{0,1\}$,

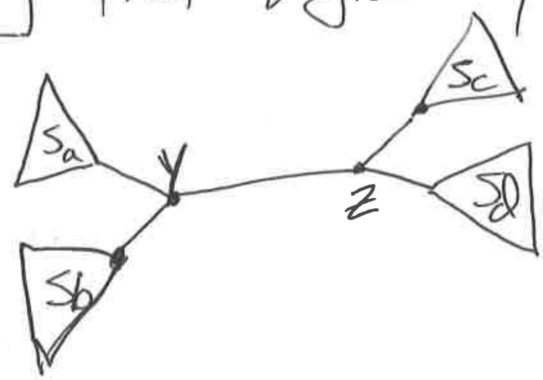
Let $w(T, v, s) = \Pr(\vec{D} \mid T_v, s \text{ at } v)$

Then $w(T, z, 0) = \prod_{i=1}^{2} \left[p \times w(T_{v_i}, v_i, 1) + (1-p) \times w(T_{v_i}, v_i, 0) \right]$

NNI MC: (nearest neighbor interchanges)

From $X_t \in \Omega$,

1. Pick an internal edge (y,z) uniformly at random
2. Let ~~S_y &~~ S_a & S_b be the subtrees hanging off y & S_c & S_d " z (Using that degree of y & z are 3)



3. There are 3 ways of hanging these 4 subtrees S_a, S_b, S_c, S_d off of $e=(y,z)$.
4. Choose one of these 3 at random. Set T' as new tree
5. Set $X_{t+1} = \begin{cases} T' & \text{with prob. } \min\left\{1, \frac{w(T')}{w(X_t)}\right\} \\ X_t & \text{o/w.} \end{cases}$

SPR = subtree pruning & regrafting.

SPR moves:

1. Choose a random internal edge $e = (y, z)$.
2. Two subtrees hanging off it, T_y & T_z.
3. Choose one of these 2 (T_y or T_z) at random.
4. Say T_y, in $T \setminus T_y$, choose random edge e' & insert T_y onto e'.

Fix tree T^* & $\vec{p}^*$

Generate $\vec{D}$ from $\mu_{T^*, \vec{p}^*}$ = random spin at internal node, then flip with prob. $p(e)$, same with $1-p(e)$.

What is mixing time?

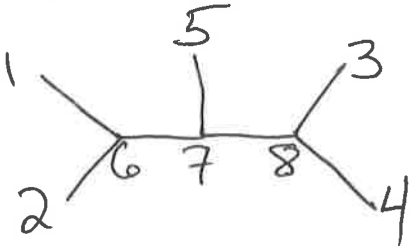
Tree T^* is max likelihood as $N \to \infty$

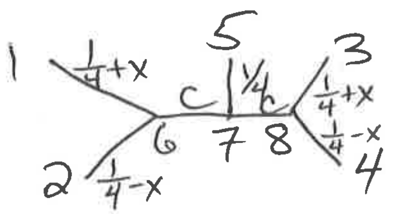
i.e., $\max_T w(T) = w(T^*)$

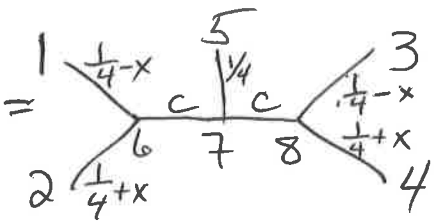
Rapid mixing for all $T^*, \vec{p}^*$?

Consensus tree = summary tree from multiple MCMC runs.

5-taxon tree:

$T^* =$ 

$\vec{P_1} =$ 

$\vec{P_2} =$ 

$$c = \frac{1}{2}\left(1 - \sqrt{\frac{1-16x^2}{1+16x^2}}\right)$$

Let $\mu = \frac{1}{2}\left(\mu_{T^*, \vec{P_1}} + \mu_{T^*, \vec{P_2}}\right)$

$\exists x_0 > 0, \forall x < x_0,$ with Prob. $\geq 1 - \exp(-\Omega(N))$

$$T_{mix} = \exp(\Omega(N))$$

for NNI chain.

Open: What about MC^3 = Metropolis-coupled MCMC,

Can we construct a bad example?