Review:

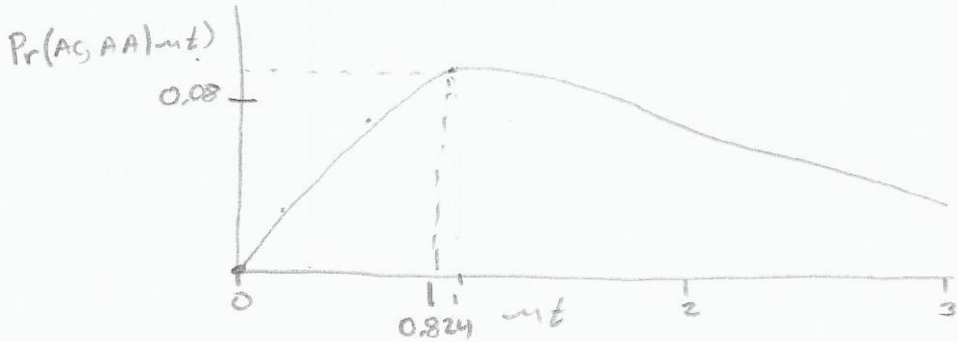
we have discussed how to compute likelihoods.

Likelihood: Pr(data | model)

Example: $AA \xrightarrow{\mu t} AC$

For Jukes-Cantor:

$$Pr(AC, AA|\mu t) = Pr(A|A, \mu t) \cdot Pr(A|C, \mu t)$$
$$= p \cdot \frac{(1-p)}{3}$$
$$= \left(\tfrac{3}{4} e^{-\frac{4}{3}\mu t} + \tfrac{1}{4}\right) \cdot \left(\frac{1 - [\frac{3}{4} e^{-\frac{4}{3}\mu t} + \frac{1}{4}]}{3}\right)$$



So what is μt?

Maximum likelihood: model that maximizes Pr(data|model), so $\mu t = 0.824$ in this case.

Bayes Theorem: $$Pr(model | data) = \frac{Pr(data|model) \cdot Pr(model)}{Pr(data)}$$
$$= \frac{Pr(data|model) \cdot Pr(model)}{\sum_{model'} Pr(data|model') \cdot Pr(data)}$$

Advantages: tells us what we really want, which is Pr(model | data)

Disadvantages: 1) Hard to compute (not so big a problem anymore)
2) What is prior, Pr(model)?

How to compute?

Markov Chain Monte Carlo (MCMC)

1) Start with model m_1
2) Pick a new model, m_2. If using Metropolis method, choose m_2 such that proposal rate $m_1 \to m_2 = m_2 \to m_1$. Typically local steps (i.e. m_2 similar to m_1)
3) Compute $R = \frac{Pr(m_2|data)}{Pr(m_1|data)} = \frac{Pr(data|m_2) \cdot Pr(m_2)}{Pr(data|m_1) \cdot Pr(m_1)}$
4) If $R \geq 1$, move to model m_2. Otherwise if $R < 1$, move to m_2 with probability R
5) Repeat step 2

If MCMC is repeated for enough steps, the probability that the chain will be on model m_i is $Pr(m_i \mid data)$.

To see this, consider two models m_i and m_j. Let $Pr(m_i \mid data) \geq Pr(m_j \mid data)$. Let f_i be the fraction of the time the chain is at m_i.

At equilibrium:

$$f_i \cdot Pr(m_i \rightarrow m_j) = f_j \cdot Pr(m_j \rightarrow m_i).$$

Now $Pr(m_i \rightarrow m_j) = R = \dfrac{Pr(m_j \mid data)}{Pr(m_i \mid data)}$

$Pr(m_j \rightarrow m_i) = 1$

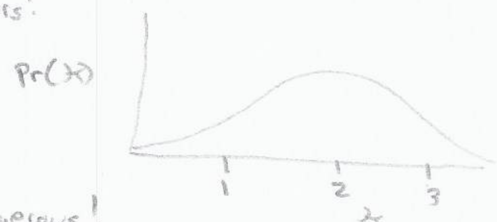
so $f_i \cdot \dfrac{Pr(m_j \mid data)}{Pr(m_i \mid data)} = f_j$, so $\dfrac{f_i}{f_j} = \dfrac{Pr(m_i \mid data)}{Pr(m_j \mid data)}$

Therefore, the chain samples models according to their posterior probability if the MCMC is run long enough.

How do we know how long is "long enough"?

About the prior, $Pr(model) \longrightarrow$ what does that mean?

Sometimes we can use existing knowledge to estimate "reasonable" priors:



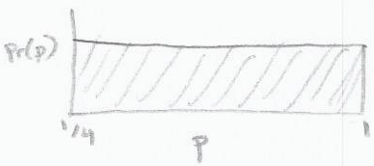
But can be dangerous!

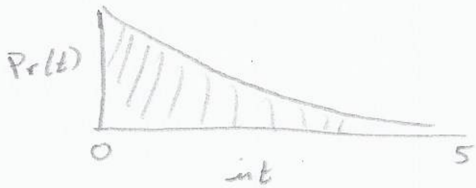
Jukes-Cantor: $p = \frac{3}{4} e^{-\frac{4}{3}\mu t} + \frac{1}{4}$

In our example, $Pr(\mu t \mid AC, AA) = Pr(\mu t) \cdot Pr(AC, AA \mid \mu t)$

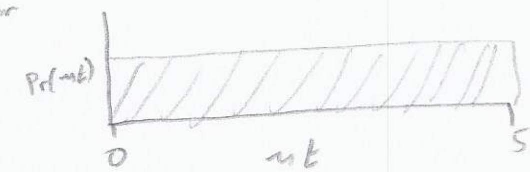
Let's say we choose a "flat" prior:

Flat prior on p

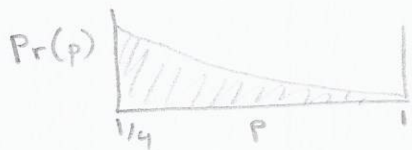




Flat prior on μt



$\Longrightarrow$



Hopefully you choose reasonable priors that have finite definite integrals over parameter space. Hopefully you have enough data that results aren't too sensitive to prior. Remember: Bayesian approaches ... probabilities over models, not just one best model