

Tree Models [JF:33]

Fredrik Ronquist

October 24, 2005

1 Introduction

A phylogenetic model used in statistical inference contains two essential components: the model of the tree and the model of the evolutionary process occurring on that tree. Much of our discussion of phylogenetic models thus far has focused on the latter component, the model of character evolution. In this lecture, we will move our attention to the tree part of the model.

2 The non-clock tree model

The non-clock model is the standard model used in phylogenetic inference. As mentioned in one of the previous lectures, it can be considered as a branch-breaking model which allows the evolutionary rate to be different for each branch in the tree. Because of this, the non-clock tree model has a large number of free parameters, one for each branch in the tree. If there are n tips in the tree, there are $2n - 3$ branches and hence branch length parameters in the tree model. This means that, for a typical phylogenetic problem, the largest number of free parameters comes from the branch lengths. For instance, consider a GTR model applied to a non-clock tree with 100 tips. There are 8 free parameters in the GTR model (3 free parameters from the stationary state frequencies and 5 from the substitution rates) but an astonishing 197 branch length parameters.

In standard maximum likelihood analysis of phylogeny, branch lengths are maximized and not integrated out of the model. This is somewhat surprising given that there are so many branch length parameters; both from a statistical robustness perspective and from an efficiency perspective it would appear preferable to use integrated likelihood to remove them. The reason that this is not

done is that the ‘obvious’ parametrization of branch lengths does not allow them to be integrated out (Fig. 1). If we focus on just one branch length, there is typically a maximum likelihood peak at a relatively short length and then the likelihood decreases as the length increases. However, there is a boundary value; the likelihood over the branch length in focus can never be lower than the probability of drawing a sequence at random, that is $(1/4)^n$ under the Jukes Cantor model, where n is the length of the sequence. This means that the integrated likelihood over all possible branch length values from 0 to infinity is not bounded; it is infinite. Rather than imposing some arbitrary upper limit on branch lengths, most likelihoodists prefer to bite the bullet and optimize branch lengths instead.

Figure 1: Likelihood as a function of branch length.



In Bayesian phylogenetics, we need to formulate a prior that allows the posterior to be marginalized over branch lengths. In early applications, a typical solution was to assign branch lengths a uniform prior between 0 and some arbitrary cut-off value that was outside the plausible range of branch length values, for instance 10 or 100. However, this occasionally resulted in strange posteriors because so much of the prior was on extremely long branches. Therefore, Bayesian phylogeneticists today typically use exponential priors on branch lengths. The exponential prior goes from 0 to infinity but it dies off fast enough to produce a bounded integral. This results in the likelihood times the prior to be bounded as well.

The exponential branch length prior may seem far from vague. However, if we choose a different parametrization, the perspective becomes different. In particular, consider the possibility of measuring branch length in terms of change probability instead of in terms of the total amount of expected change per site. With the right parameter choice, the exponential prior is actually uniform on the change probability. For the Jukes Cantor model, there is only one change probability, and it goes from 0 to $3/4$. A uniform prior on the change probability can easily be integrated out; that

is, if we measured branches in terms of change probability we could have used integrated likelihood to remove them from the maximum likelihood problem (Fig. 2). This is exactly what we do in Bayesian inference when we marginalize over some other parameter in the model.

Figure 2: Likelihood as a function of change probability (Jukes Cantor model).



3 The clock tree model

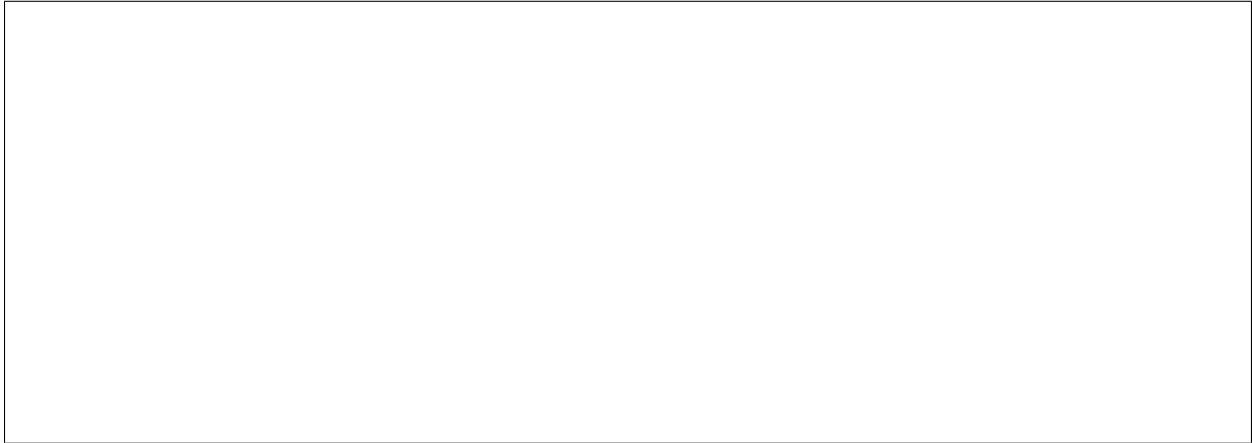
In the simplest form, the clock tree model is simply the other end of the spectrum from the non-clock model; it assumes that the evolutionary rate is the same throughout the entire tree. This results in a drastic reduction in the number of free parameters. A tree with n tips has $2n - 3$ free branch length parameters under the non-clock model but only $n - 1$ under the clock model, that is, roughly half the number of free parameters.

Unlike non-clock trees, clock trees are always rooted. Thus, there is no need to specify an outgroup; indeed, a clock tree analysis may result in a rooting that is inconsistent with any previous notions about outgroups. A clock tree can be forced to have a particular outgroup only if the topology is constrained.

Calculating the likelihood of a clock tree proceeds in exactly the same fashion as for a non-clock tree. However, optimizing branch lengths is a little different because branch lengths are dependent on each other. Thus, we cannot focus on a single branch length and optimize it independently of all other branches. Instead, we have to optimize branch lengths by readjusting one speciation event at a time, looking at the effects on the three adjacent branches of that speciation event (Fig. 3).

In Bayesian phylogenetics, several issues arise with clock trees. In particular, the prior on branch

Figure 3: Optimizing branch lengths in a clock tree



lengths and topology must be specified slightly differently than for non-clock trees. Several of the early Bayesian phylogenetics papers used clock trees but were never explicit about the prior they assumed. It seems that some of them implicitly assumed a $\text{Uni}(0, \infty)$ prior on tree height. This is of course an improper prior and because the likelihood function does not decrease to 0 when the tree grows in size, the posterior will be improper as well. Felsenstein complains bitterly about this (pp. 295-298) but it is sufficient to assume some upper limit on tree height to fix the problem, and this upper limit is unlikely to be reached in analyses starting from a moderate tree height, that is, the upper limit need not be specified for the analysis to work properly for most real data sets.

However, one must agree with Felsenstein that the improper prior is logically unsatisfactory. A more elegant prior on clock trees uses an exponential or gamma distribution for the tree height. The relative speciation times, call them $b = \{b_1, b_2, b_3, \dots\}$, given a topology τ can now be associated with a uniform prior, that is $p(b|\tau) \propto 1$. Interestingly, this induces a prior probability distribution on topologies, which puts equal probability on all unique labeled histories.

Calculating the Hastings ratio for clock trees can be tricky. For instance, the original formulation of the clock version of the LOCAL (Larget and Simon, 1999) contains three separate errors in the calculation of the Hastings ratio. In using Green's formula, one has to remember that it is the relative speciation times (b), not the individual branch lengths, which should be left unaffected by any stretching or shrinking induced by the proposal.

4 The birth-death process

The prior on clock trees described above arises from a special case of the Yule model, which assumes that species evolve by a birth and death process in which the birth and death rates are constant through time. Assume that the instantaneous birth rate is λ and the death rate is μ . Then the probability that a lineage leaves at least one descendant after time t in the Yule model, that is, the probability that it survives until time t , is:

$$s(t) = \text{Prob}(n > 0|t) = \frac{\lambda - \mu}{\lambda - \mu e^{(\mu-\lambda)t}}$$

where n is the number of descendants. The probability that there will be exactly one descendant after time t is:

$$p_1(t) = \text{Prob}(n = 0|t) = s(t)^2 e^{(\mu-\lambda)t}$$

We can now calculate the probability of a set of relative speciation times $b = \{b_1, b_2, b_3, \dots, b_{n-2}\}$ given the birth rate λ , the death rate μ , and the tree height T for a set of s species as:

$$p(b|T, \lambda, \mu) \propto \prod_i \frac{\lambda p_1(b_i)}{\nu_T}$$

where

$$\nu_T = 1 - s(T)e^{(\mu-\lambda)T} = \int_0^T \lambda p_1(t) dt$$

If we set $\lambda = \mu$, that is, a situation where the birth and death rates are equal, we get

$$p(b|T, \lambda, \mu) \propto \prod_i \frac{1 + \mu}{(1 + \mu b_i)^2}$$

If, instead, we set $\mu = 0$ (no extinction) we get

$$p(b|T, \lambda, \mu) \propto 1$$

that is, the prior probability distribution we discussed in the previous section on clock trees. In other words, the uniform prior probability distribution on branching times is the one resulting from a pure birth model.

The birth-death model can also be combined with a sampling probability, that is, we can accommodate the common situation that not all tips are present in the sample we are analyzing. See Yang and Rannala (1997) for details. Felsenstein claims that Yang and Rannala's sampling model is different from randomly sampling n individuals from the survivors of the birth and death process but I am uncertain whether this statement is correct.

There are several interesting aspects of the birth-death model. One is that it allows us to estimate birth and death rates by looking at patterns of branch lengths. Many deep nodes suggest a group that has a high extinction rate; many shallow nodes indicate a high speciation rate.

5 The coalescent model

The coalescent model is very similar to the birth-death model but it differs in several small details. It will be covered in detail in coming lectures.

6 Relaxed clock tree models

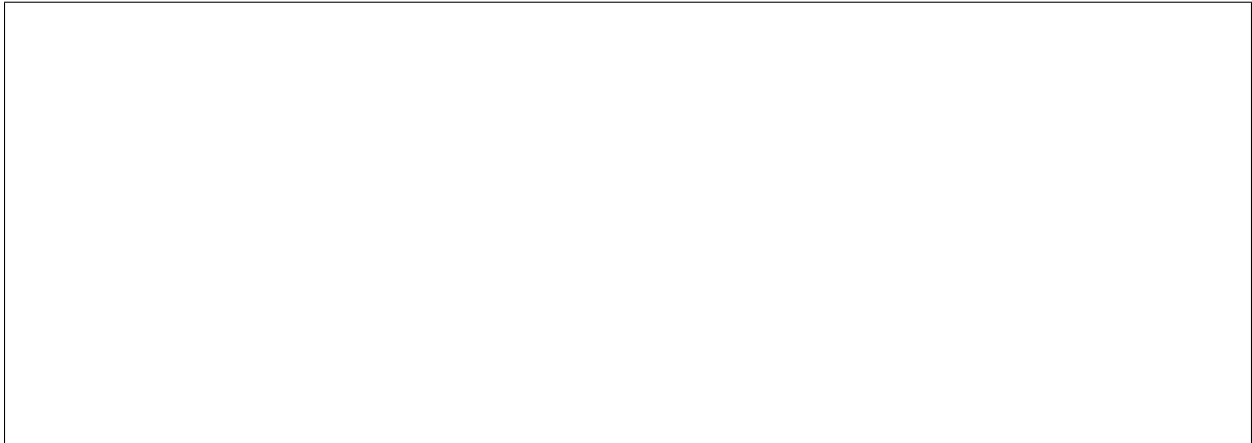
Despite the large reduction in the number of parameters, it is often possible to reject the clock model in favor of the non-clock model with real data because of significant rate heterogeneity. In these situations, one can explore the middle ground of relaxed clock models. They allow some variation in evolutionary rates across branches but not as extreme as the non-clock model. We have covered these models previously and they will not be discussed further here.

7 Tree balance and speciation rate

There has been ample discussion in the literature about tree balance, that is, how asymmetric trees are, and what it tells us about evolution (Fig. 4). The Yule process tends to generate symmetric trees and several empirical studies suggest that real trees are more asymmetric than expected under this model. Several other models have been explored in the search for a satisfactory explanation

of the preponderance of unbalanced trees but none is entirely satisfactory. It seems clear, though, that an important reason for the lack of balance is that speciation and extinction rates are inherited and can vary across the tree. For instance, several studies have shown that the origination of key ecological traits is associated with increased numbers of extant species. Despite the apparent misfit between the Yule process and some of the patterns we see in real trees, it is nevertheless a useful null model and exploratory tool in examining speciation and extinction patterns. In this respect, it resembles the clock model.

Figure 4: Tree balance



8 Study Questions

1. How many free parameters are there in a non-clock model for n taxa? In a clock model for n taxa?
2. Why is integrated likelihood rarely used to deal with branch length parameters even though they are nuisance parameters?
3. Describe an appropriate prior for Bayesian analysis of clock trees
4. What are the advantages of an exponential prior over a uniform prior on branch lengths?
5. What is tree balance?