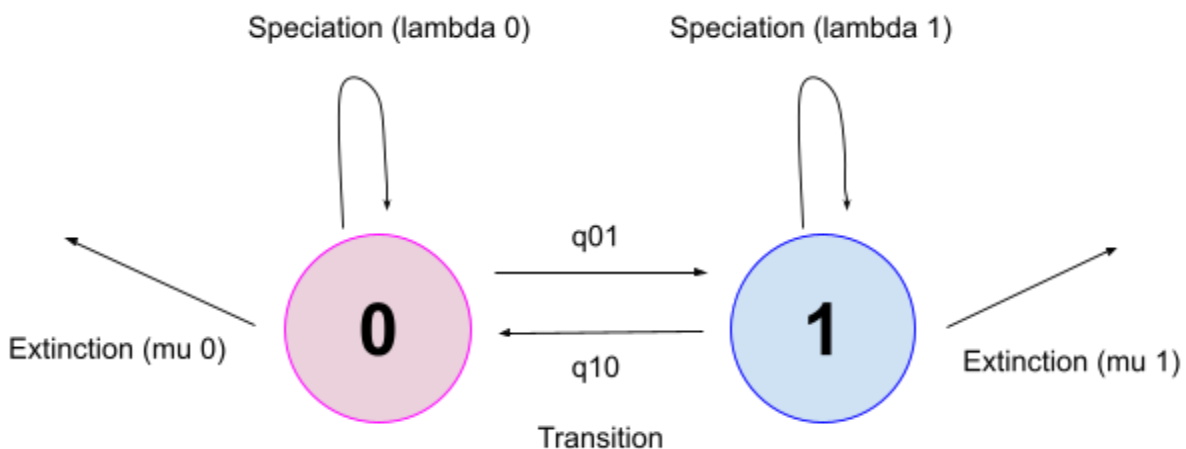


BiSSE and related models are now mostly used to assess if the states of a character are associated with different rates of speciation or extinction.

BiSSE was introduced by Madisson(2009) to fix 2 problems.

- Inferences about character state can be thrown off if the character affects rates of speciation or extinction.
- Inferences based on whether a character affects lineage diversification based on sister clade comparisons can be thrown off by asymmetric transition rates

The theory behind state-dependent diversification models



What does BiSSE models assume

- Only two states
- The state is known for a species

What are ordinary differential equations and why are they related

- The general approach adopted by BiSSE and related models is to derive a set of ordinary differential equations (ODEs) that describe how the probability of observing a descendant clade changes along a branch in the observed phylogeny.
- Each equation in this set describes how the probability of observing a clade changes through time if it is in a particular state over that time period
- The equation is:  $dD_{N,i}(t)/dt$ 
  - Where  $i$  is the state of the lineage at time  $t$ , and  $N$  is the clade descended from that lineage
- ODE describes how the tip state changes over time
- PROBLEM: Must accommodate for the possibility a branch goes extinct before it is observed
  - Needs a set of  $k$  ODEs

- $dE_i(t)/dt$  defines how the probability of eventual extinction from state  $i$  changes over time.

How do we compute a likelihood

- Computing the likelihood proceeds by establishing an initial value problem. We initialize the procedure by observing the character states of some lineages, generally the tip states. Then starting from those probabilities (e.g., species X has state 0 with probability 1 at the present), we describe how those probabilities change over time (described by the ODEs), working our way back until we have computed the probabilities of observing that collection of lineages at some earlier time (e.g., the root).

Problems at the nodes

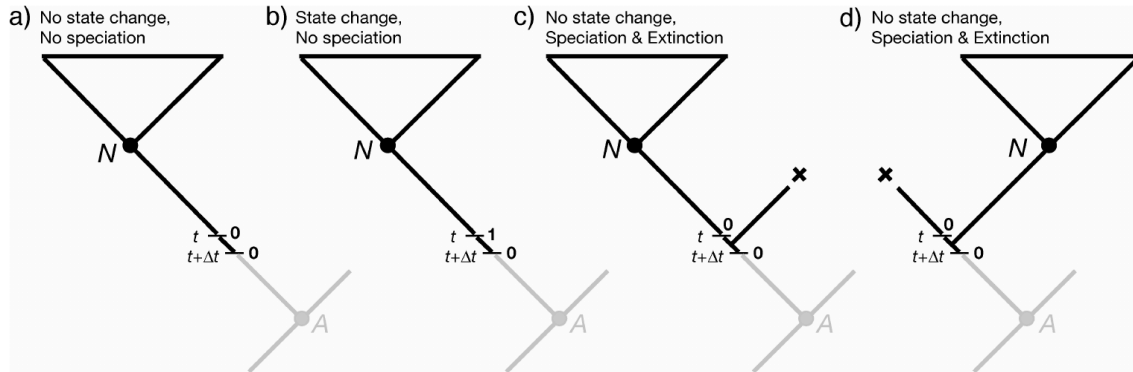
- As we measure from tip to root we also run into nodes. We can fix by:
- Assuming that the parent and daughter lineages have the same state, we multiply together the probabilities that the daughters are state  $i$  and the instantaneous speciation rate  $\lambda_i$  to get the initial value for the ancestral branch subtending that node.

What do we get:

- Once we reach the root we get  $k$  probabilities.
  - these  $k$  probabilities represent the probability of observing the phylogeny conditional on the root being in each of the state.
- The overall likelihood of the tree is a weighted average of  $k$  probabilities.

Derivation for the binary state birth-death process

- Consider a (time-independent) birth-death process with two possible states (a binary character), with diversification rates  $\{\lambda_0, \mu_0\}$  and  $\{\lambda_1, \mu_1\}$ .
- CLADE PROB. described as  $D_{Ni}$ 
  - For example: We define  $D_{N,0}(t)$  as the probability of observing lineage N descending from a particular branch at time  $t$ , given that the lineage at that time is in state 0.
- There is a possibility of observing the lineage at some earlier time point, to compute:  $D_{N,0}(t+\Delta t)$ , this represents all possible events that could occur within the interval  $\Delta t$ . There are four possibilities if  $\Delta t$  is small
- 1) nothing happens;
- 2) a transition occurs, so the state changes  $0 \rightarrow 1$ ;
- 3) a speciation event occurs, and the right descendant subsequently goes extinct before the present, or;
- 4) a speciation event occurs and the left descendant subsequently goes extinct before the present.



There are multiple equations that represent the four different outcomes. Lets summarize them into a differential equation:

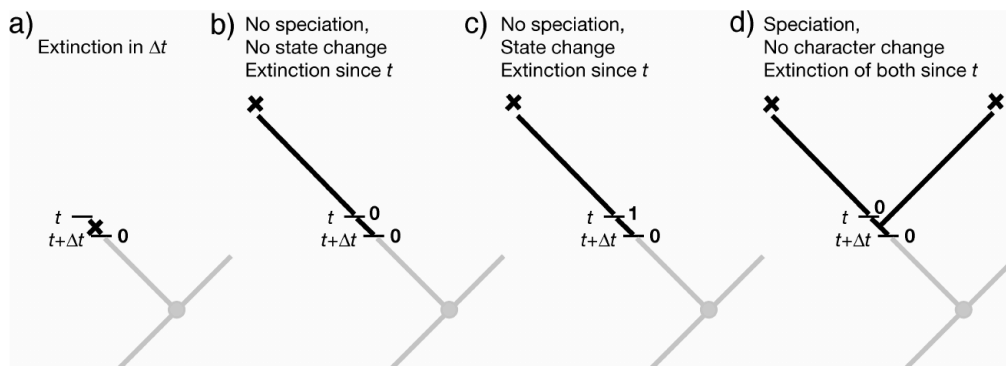
$$\frac{dD_{N,i}(t)}{dt} = -(\lambda_i + \mu_i + q_{ij}) D_{N,i}(t) + q_{ij} D_{N,j}(t) + 2\lambda_i E_i(t) D_{N,i}(t)$$

But to solve that equation we need extinction probabilities

- Define  $E_0(t)$  as the probability that a lineage in state 0 at time  $t$  goes extinct before the present.

To determine the extinction probability at an earlier point,  $E_0(t+\Delta t)$ , we can again enumerate all the possible events in the interval  $\Delta t$ :

1. the lineage goes extinct within the interval;
2. the lineage neither goes extinct nor speciates, resulting in a single lineage that must eventually go extinct before the present;
3. the lineage neither goes extinct nor speciates, but there is a state change, resulting in a single lineage that must go extinct before the present, or;
4. the lineage speciates in the interval, resulting in *two* lineages that must eventually go extinct before the present.



Again convert into differential equations:

$$\frac{dE_i(t)}{dt} = \mu_i - (\lambda_i + \mu_i + q_{ij}) E_i(t) + q_{ij} E_j(t) + \lambda_i E_i(t)^2$$

Initial values: tips and sampling

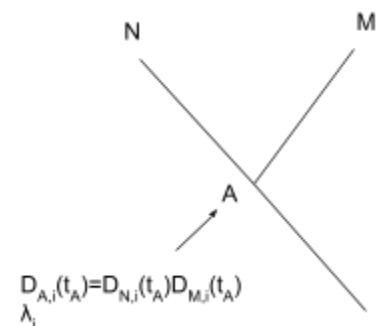
- The equations above describe how to get the answer at time  $t+\Delta t$  assuming we already have the answer at time  $t$ . How do we start this process?
  - The answer is with our character state observations, which are generally the tip state values.
- For example: If species  $s$  has state  $i$ 
  - $D_{s,i}(0)=1$  (probability is 1 at time 0 [the present] because we observed it for sure)
  - $E_i(0)=0$  (probability 0 of being extinct at the present).
  - For all states other than  $i$ ,  $D_{s,j}(0)=0$  and  $E_j(0)=1$ .
- We can adjust these initial conditions to allow for incomplete sampling.
  - If a proportion  $p$  of species are included on the tree, we would instead set  $D_{s,i}(0)=p$  (probability of having state  $s$  and of being on the tree) and  $E_i(0)=1-p$  (probability of absent, due to sampling rather than extinction).
  - This simple form of incomplete sampling assumes that any species is equally likely to be on the tree

At Nodes

- The differential equations for speciation and extinction are the BiSSE ODEs, describing probabilities along the branches of a phylogeny.
- What about the nodes:
  - BiSSE assumes the ancestor (called  $A$ ) and descendants (called  $N$  and  $M$ ) have the same state (i.e., there is no cladogenetic character change).
  - The initial value for the ancestral branch going into a node (at time  $t_A$ ) is then the product of the final values for each of the daughter branches coming out of that node, times the instantaneous speciation rate (to account for the observed speciation event):
- $D_{A,i}(t_A) = D_{N,i}(t_A) D_{M,i}(t_A) \lambda_i$

At the root

- After performing the differential equations for speciation and extinction at the tips and account for the nodes we arrive at the root. with the  $D$  values (called  $D_{R,i}$ ), one for each state.
- Need to combine these to get the likelihood
  - $\text{Likelihood}(\text{tree, tip states} \mid \text{model}) = (\sum_i) D_{R,i} p_{R,i}$



Basically... State  $A$  is the average of the states  $N$  and  $M$  were in to account for speciation but leaves out cladogenetic change

- Sometimes a fixed approach is used, assuming that the prior root state probabilities are either all equal, or are the same as the observed tip state frequencies, or are the equilibrium state frequencies under the model parameters.
  - These assumptions do not have a real basis, however (unless there is some external data that supports them), and they can cause trouble (Goldberg and Igić 2008).
- An alternative is to use the BiSSE probabilities themselves to determine the root state weightings, essentially adjusting the weightings to be most consistent with the data and BiSSE parameters (FitzJohn et al. 2009).
  - Perhaps better is to treat the weightings as unknown parameters to be estimated. These estimates are usually quite uncertain, but in a Bayesian framework, one can treat the  $p_{R,i}$  as nuisance parameters and integrate over them.

#### Equations for the multi-state birth-death process

- The entire derivation above can easily be expanded to accommodate an arbitrary number of states (FitzJohn 2012). The only extra piece is summing over all the possible state transitions.
- The resulting differential equations within the branches are:

$$\frac{dD_{N,i}(t)}{dt} = - \left( \lambda_i + \mu_i + \sum_{j \neq i}^k q_{ij} \right) D_{N,i}(t) + \sum_{j \neq i}^k q_{ij} D_{N,j}(t) + 2\lambda_i E_i(t) D_{N,i}(t)$$

$$\frac{dE_i(t)}{dt} = \mu_i - \left( \lambda_i + \mu_i + \sum_{j \neq i}^k q_{ij} \right) E_i(t) + \sum_{j \neq i}^k q_{ij} E_j(t) + \lambda_i E_i(t)^2$$

#### What I did to learn/Practice

- Built a BiSSE model from Luke Harmons R script
  - <https://lukejharmon.github.io/ilhabela/2015/07/05/BiSSE-and-HiSSE/>
- Working through RevBayes tut. Next
  - <https://revbayes.github.io/tutorials/sse/bisse.html>