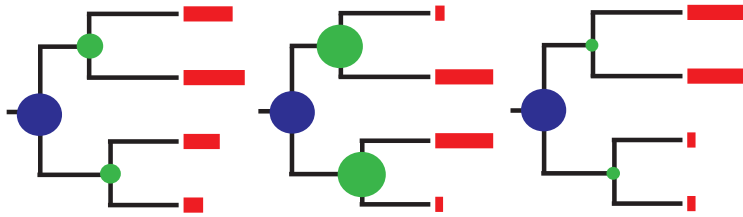


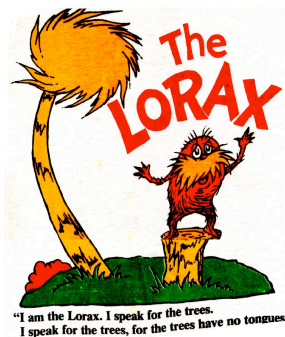
Introduction to Comparative Methods



Lecture 1: March 10, 2009
Luke Harmon

Overview of Comparative Methods

- Learning from the trees
- General statistical framework
- Basic models of growing trees and changing characters
- Details of Birth-Death



Learning from trees

- Why do you want to make trees? What do you hope to learn?

Goals of CM

- Diversification (speciation and extinction)
- Character evolution
- Characters and diversification
- Biogeography
- Testing more complex models

General statistical framework

- We will focus on a model fitting approach
- Develop mathematical models
- Calculate likelihoods
- Compare the fit of a set of competing models

Comparing Likelihoods

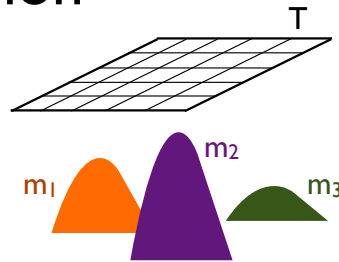
- Likelihood = Probability of obtaining the data given a model and parameter estimates
- $\Pr(D|H)$

Maximum Likelihood

- Find parameter values that maximize the likelihood of a particular model
- Example (class exercise): are you evil?

Akaike Information Criterion

- Compete models with each other
- All models are wrong
- Choose the model that is the most efficient approximation of the truth



Bayesian Statistics

- Use Bayes theorem to calculate posterior probabilities

$$\Pr(H|D) = \frac{\Pr(D|H)\Pr(H)}{\Pr(D)}$$

Likelihood

Prior

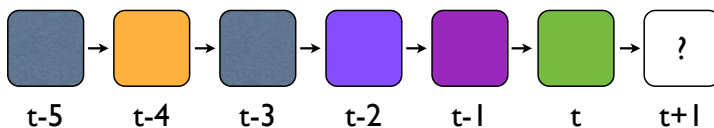
Normalizing
constant

Core Models

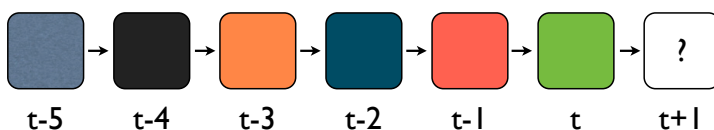
- Birth-death model for diversification
- Mk model for discrete characters
- Brownian motion model for continuous characters

Markov Models

- Most of the models used in comparative biology satisfy the Markov property:
- The conditional probability distribution of future states of the process, given the present state and a constant number of past states, depends only upon the present state and the given states in the past, but not on any other past states



Markov property:
 $Pr(x_{t+1}=i)$ depends only on x_t



Birth-death model

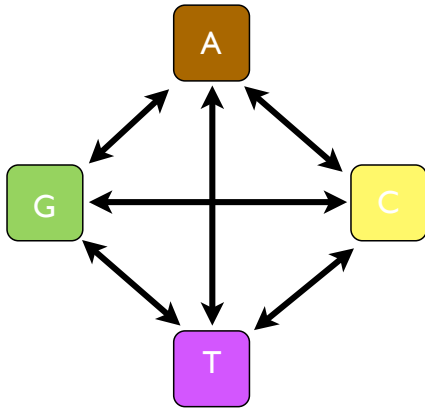
- How do we randomly grow a tree?
- Birth-death model
- Assumptions:
 - Speciation and extinction occur randomly
 - Each lineage has an equal and constant rate of speciation, b , and extinction, d

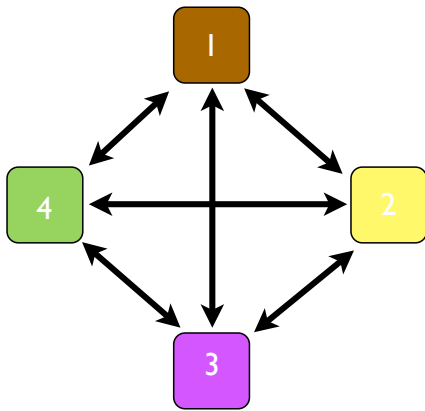
Pure-birth model

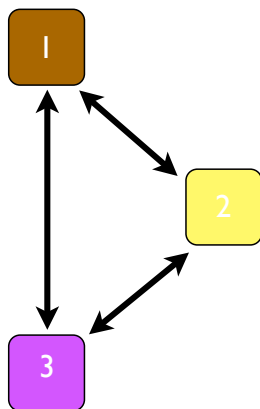
- Pure-birth model is a special case of the bd model where $d=0$
- Speciation only model

Mk Model

- This is a general version of the GTR model for sequence evolution
- Characters can exhibit one of a set of possible *states*
- Transition rates among these states is governed by a rate matrix (Q)



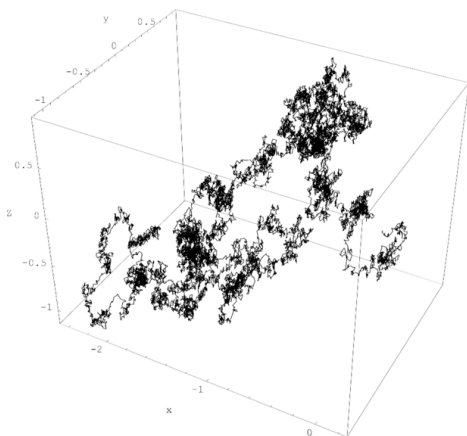






Brownian Motion

- A model for the evolution of continuously-valued characters
- States change continuously through time
- After some time, expected character states follow a normal distribution



The Birth-Death Model

- We will talk about the expected distributions of two things
 - Species Diversity
 - Waiting Times

Species Diversity

- What is the expected distribution of species diversity at time t ?

Expected species diversity under a birth-death model:

$$E[N_t] = N_o e^{(b-d)t}$$

N_t = species diversity after time t

N_o = starting species diversity

b = birth rate

d = death rate

t = time

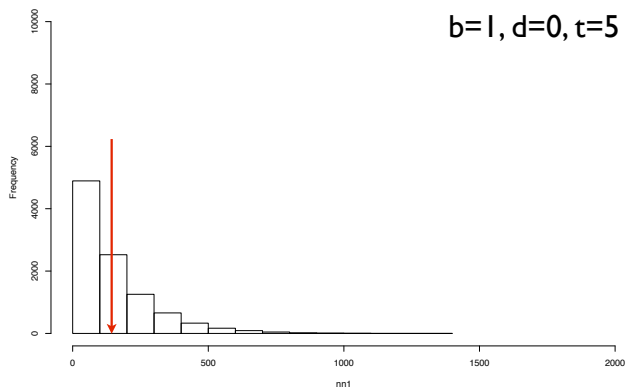
Species Diversity

- Consider a pure-birth model with $b=1$
- We begin with a single lineage, $N_0=1$
- How many lineages do we expect to have at time $t=10$?

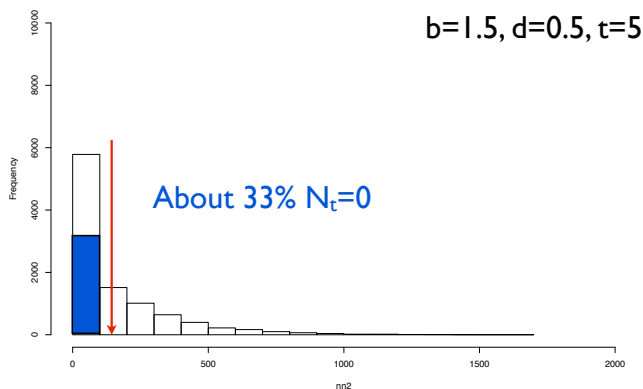
$$E[N_t] = N_0 e^{(b-d)t}$$

$$E[N_t] = 1 e^{(1-0)10}$$

$$E[N_t] = e^{10} = 22027$$



$$E[N_t] = N_0 e^{(b-d)t} = 148.4$$



$$E[N_t] = N_0 e^{(b-d)t} = 148.4$$

Species Diversity

- What is the expected distribution of species diversity at time t ?

$$\Pr(N_t = n, n > 0) = \sum_{j=1}^{\min(a,n)} \binom{a}{j} \binom{n-1}{j-1} A^{a-j} [(1-A)(1-B)]^j B^{n-j}$$
$$\Pr(N_t = 0) = A$$

$$A = \left\{ \frac{d[e^{(b-d)t} - 1]}{be^{(b-d)t} - d} \right\}^a \quad B = \frac{bA}{d}$$

This also gives us a *likelihood*

Species Diversity and Rates of Diversification

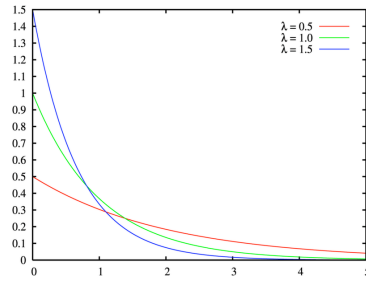
- Using the equations from the previous slide, one can obtain maximum-likelihood estimates of b and d given ages and diversities for a set of clades
- See Magallón and Sanderson 2001 Evolution 55:1762.

Distributions associated with the bd model

- We will talk about the expected distributions of two things
 - Species Diversity
 - **Waiting Times**

Waiting times

- Under a birth-death model, the time intervals between successive events are always drawn from exponential distributions

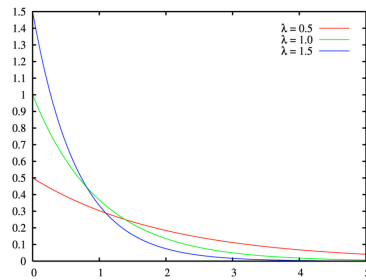


Waiting times

- Exponential distributions are governed by one parameter, λ

- $E[X] = 1/\lambda$

Expected value of x ;
mean of the distribution



PURE BIRTH EXAMPLE

Starting with a single ancestor at
time $t = 0$

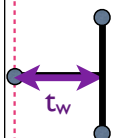
$t = 0$

1. Draw waiting time from exponential distribution with $\lambda = b$

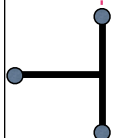


$t = 0$

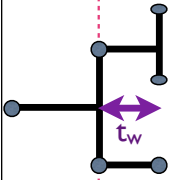
2. A speciation event occurs at time $t_0 + t_w$; add a new species at that time



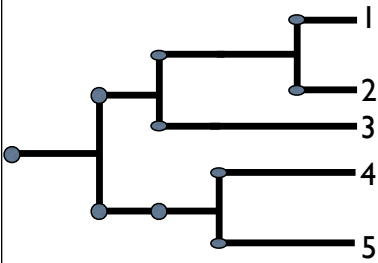
3. The next waiting time is also drawn from an exponential distribution, but now with $\lambda = 2b$



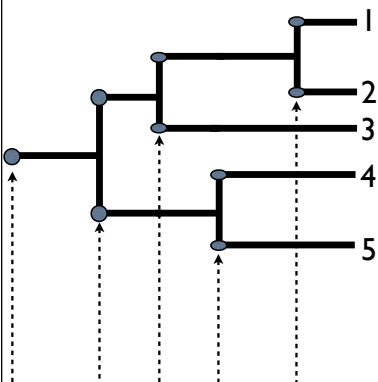
4. Choose one of the two daughters at random to speciate



5. Carry on until tree is built



5. Carry on until tree is built



Intervals between
speciation events

Growing birth-death trees

- Same protocol, except waiting times are drawn from exponential distribution with $\lambda = n(b + d)$, where n is the number of living species at that time
- Choose a random lineage
- Decide whether you have a speciation or extinction event
- $\text{Pr}[\text{speciation}] = b/(b+d)$; $\text{Pr}[\text{extinction}] = d/(b+d)$

