Introduction to Comparative Methods
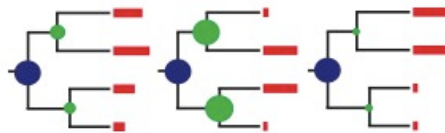
Lecture 1: March 10, 2009
Luke Harmon

Goals of CM

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

Overview of Comparative Methods

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

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

Learning from trees

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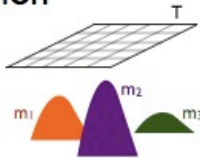
Comparing Likelihoods

- 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?

Core Models

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

Akaike Information Criterion

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

m_1 m_2 m_3 T

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

Bayesian Statistics

- Use Bayes theorem to calculate posterior probabilities
- $Pr(H|D) = \dfrac{Pr(D|H)Pr(H)}{Pr(D)}$

Likelihood
Prior
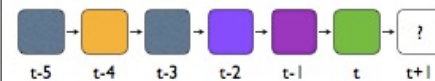Normalizing constant

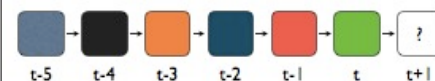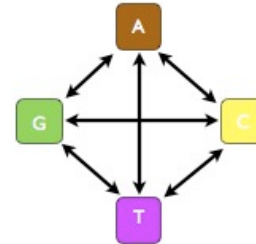t-5 t-4 t-3 t-2 t-1 t t+1

Markov property:
$Pr(x_{t+1}) = i$ depends only on x_t
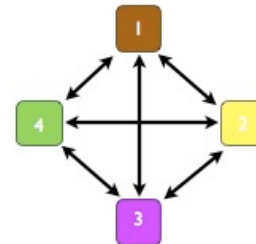
t-5 t-4 t-3 t-2 t-1 t t+1

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)
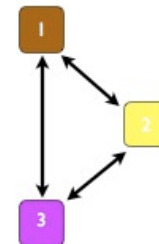
Brownian Motion

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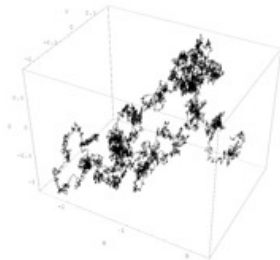
The Birth-Death Model

- 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:

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

N_t = species diversity after time t
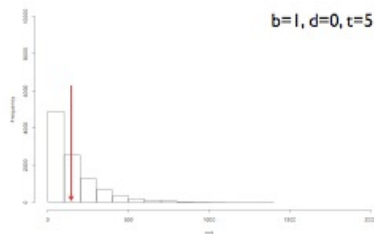N_0 = starting species diversity
b = birth rate
d = death rate
t = time

Species Diversity

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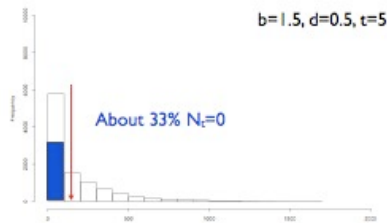- How many lineages do we expect to have at time t=10?

$$E[N_t] = N_o e^{(b-d)t}$$
$$E[N_t] = 1 \ e^{(1-0)10}$$
$$E[N_t] = e^{10} = 22027$$

Species Diversity

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

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

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

This also gives us a *likelihood*

---

b=1, d=0, t=5

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


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.

---

b=1.5, d=0.5, t=5

About 33% $N_t=0$

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

Distributions associated with the bd model

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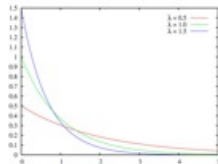- Species Diversity
- Waiting Times

Slide 1

Waiting times

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

Slide 2

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

t_w

$t = 0$

Slide 3
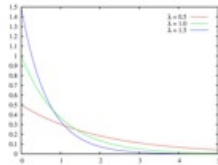
Waiting times

- Exponential distributions are governed by one parameter, λ
- $E[X] = 1/\lambda$

Expected value of x; mean of the distribution

Slide 4

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

t_w

Slide 5

PURE BIRTH EXAMPLE

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

$t = 0$

Slide 6

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

t_w

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

t_w
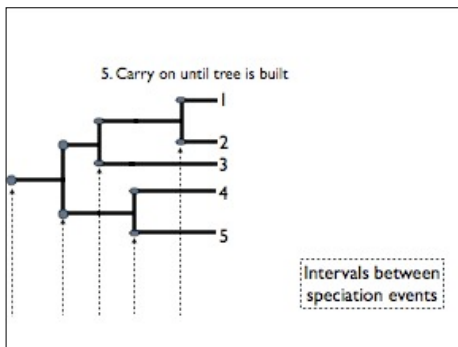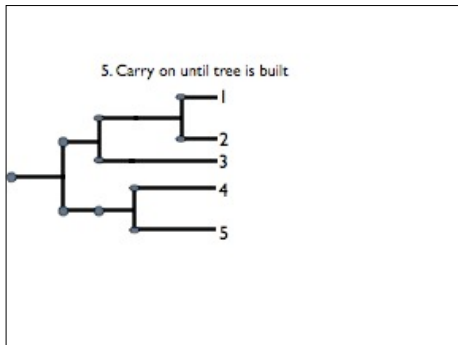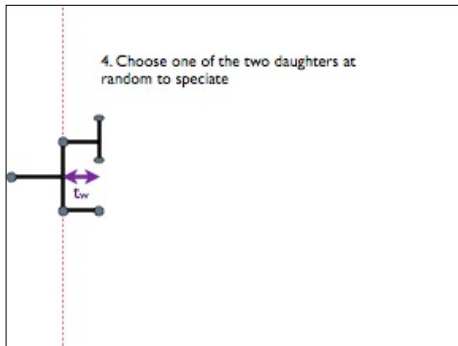
Growing birth-death trees

- 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
- Pr[speciation] = b/(b+d); Pr[extinction]=d/(b+d)

5. Carry on until tree is built

1
2
3
4
5

5. Carry on until tree is built

1
2
3
4
5

Intervals between speciation events