

Botany 940 Notes

April 12, 2022

1 2022-02-01: brownian motion

[harmon chapters 3-5]

1.1 hamon chapter 3: intro to brownian motion

we consider traits that have continuous distribution — such as body mass. we can model these traits with brownian motion. it is incorrect to equate brownian motion models with models of pure genetic drift. (why?)

when modeling evolution using brownian motion, we usually are considering the dynamics of the **mean character value** \bar{z} in a population. this is the average value of the trait within the population. we denote by $\bar{z}(t)$ the value at time t . we can model this by a brownian motion process. the mean is $\bar{z}(0)$ and the variance, or “evolutionary rate parameter” is denoted by σ^2 . we can simulate change under the brownian motion model by drawing from normal distributions; simply put the changes in a trait value over an interval of time t is drawn from a normal random variable with mean 0 and variance $\sigma^2 t$.

what does this mean: “we assume that mutations are drawn at random from a distribution with mean 0 and mutational variance σ_m^2 .”?

Definition 1 (types of variance). *types of variance under this model:*

- σ_m^2 is the **mutational variance** the variance of the number of new distinct mutations added per generations (*i think*)
- σ_b^2 is **between-population phenotypic variance**, meaning the variance of mean trait values across many independent runs of evolutionary change over a certain time period.
- σ_a^2 is the **additive genetic variance** within each population at some time t . additive genetic variance measures the total amount of genetic variation that acts additively (*i.e. the contributions of each allele add together to predict the final phenotype*)
- σ_w^2 is the total phenotypic variation within a population, including both non-additive genetic effects and environmental effects.

1.2 additive variance equilibrium

Claim 1. *we have $[\sigma_a^2(t)] \approx 2n_e \sigma_m^2$ for large t .*

Proof. over time, σ_a^2 will change. this is due to genetic drift (which tends to decrease σ_a^2) and mutational input (which tends to increase σ_a^2). the expected value of σ_a^2 changes from one generation to the next according to the formula:

$$[\sigma_a^2(t+1)] = \left(1 - \frac{1}{2n_e}\right) [\sigma_a^2(t)] + \sigma_m^2$$

where t is the time in generations. the first term represents the loss due to drift. the second term represents the gain due to new mutations each generation. by induction, we have:

$$[\sigma_a^2(t)] = \left(1 - \frac{1}{2n_e}\right)^t [\sigma_a^2(0) - 2n_e \sigma_m^2] + 2n_e \sigma_m^2 \quad (1)$$

where of course $\sigma_a^2(0)$ is the starting value at time zero. this formula allows us to calculate the expected additive genetic variance at any time provided we know $\sigma_a^2(0)$. since $\left(1 - \frac{1}{2n_e}\right)^t \rightarrow 0$ as $t \rightarrow \infty$, we expect $[\sigma_a^2(t)]$ to approach equilibrium value of $2n_e\sigma_m^2$. \square

1.3 derivation of $\sigma_b^2(t)$

assume σ_a^2 is at equilibrium and thus constant. harmon claims this is justified by equation (1), even though that equation is about the expected value of σ_a^2 . then skipping some calculations,

$$\sigma_b^2(t) = \frac{t\sigma_a^2(t)}{n_e} \quad (2)$$

substituting the equilibrium value $2n_e\sigma_m^2$ in for $\sigma_a^2(t)$ gives

$$\sigma_b^2(t) = 2t\sigma_m^2 \quad (3)$$

this says that the variation between two diverging populations depends on twice the time since they diverged and the rate of mutational input. no dependence on population size or starting state is observed, so under this model, long-term rates of evolution are dominated by the supply of new mutations to a population.

actually (3) is very general, holding under a range of models. unfortunately, we cannot measure σ_m^2 in any natural population.

Definition 2 (heritability). *heritability* is the proportion of total phenotypic variation σ_w^2 within a population that is due to additive genetic effects; namely,

$$h^2 = \frac{\sigma_a^2}{\sigma_w^2}. \quad (4)$$

by the additive variance equilibrium equation (1), we have

$$h^2 = \frac{2n_e\sigma_m^2}{\sigma_w^2} \quad (5)$$

or equivalently,

$$\sigma_m^2 = \frac{h^2\sigma_w^2}{2n_e} \quad (6)$$

and therefore by (3),

$$\sigma_b^2(t) = \frac{h^2\sigma_w^2 t}{n_e}. \quad (7)$$

1.4 brownian motion under selection

recall we wish to model the path followed by population mean trait values under mutation, selection, and drift. the brownian motion model applies even to some cases with selection.

1.4.1 random directional selection

we assume directional selection, but with the strength and direction of selection varying randomly from one generation to the next. each generation, we model selection as being drawn from a normal distribution with mean 0 and variance σ_s^2 .

in this case, for some reason, we have

$$\sigma_b^2 = \left(\frac{h^2\sigma_w^2}{n_e} + \sigma_s^2 \right) t \quad (8)$$

in this case, if variation in selection is much greater than variation due to drift, then for some reason we have

$$\sigma_b^2 \approx \sigma_s^2 \quad (9)$$

1.4.2 random stabilizing selection

on the other hand, if the trait is under stabilizing selection for a particular optimal value, where the position of the optimal value changes randomly according to a brownian motion process with variance σ_e^2 , then

$$\sigma_b^2 \approx \sigma_e^2$$

under both of these models, the pattern of trait evolution through time still follows a brownian motion model, even though the changes are dominated by selection rather than drift. therefore brownian motion evolution does not assume that characters are not under selection.

1.4.3 constant directional selection

not going to go through the equations on this one. if a trait is under directional selection of strength s in both populations, then we get

$$\sigma_b^2(t) = \frac{h^2 \sigma_w^2 t}{n_e} \quad (10)$$

this does not depend on s and in fact is the same equation as (7) which was derived earlier. thus we can't tell by looking at two living populations whether the trait we are measuring was subject to selection, or something like that.

1.5 brownian motion on a tree

i am mostly familiar with this section from the diffusion group. a new thing is the **phylogenetic variance-covariance matrix \mathbf{c}** . this matrix has a special structure. for phylogenetic trees with n species, this is an $n \times n$ matrix, with each row and column corresponding to one of the n taxa in the tree. along the diagonal are the total distances of each taxon from the root of the tree, while the off-diagonal elements are the total branch lengths shared by particular pairs of taxa. so it is a matrix $(\sigma^2 x_{ij})$ where x_{ii} is the length of the path $p(i, r)$ from leaf i to the root, and x_{ij} is the length of those edges shared by paths $p(i, r)$ and $p(j, r)$.

example: suppose we have a phylogenetic tree with two leaves 1 and 2 with lengths t_2, t_3 respectively, and a root edge of length t_1 . we can describe the trait values for the two species as a single draw from a multivariate normal distribution. each trait has the same expected value $\bar{z}(0)$, which is the starting value of the trait at the root, and two traits have the following covariance matrix:

$$\begin{bmatrix} \sigma^2(t_1 + t_2) & \sigma^2 t_1 \\ \sigma^2 t_1 & \sigma^2(t_1 + t_3) \end{bmatrix} = \sigma^2 \begin{bmatrix} t_1 + t_2 & t_1 \\ t_1 & t_1 + t_3 \end{bmatrix} = \sigma^2 \mathbf{c}$$

we see what \mathbf{c} is in this case.

2 2022-02-01: fitting brownian motion models to single characters

[harmon chp 4]

2.1 log transforms

if your data is a continuous trait and you think percentage changes matter more than absolute changes, then you should do a log-transform of your data.

2.2 estimating rates using independent contrasts

phylogenetical independent contrasts (pics) are a way to estimate the rate of character change across a phylogeny. there is an algorithm, the algorithm of independent contrasts, from felsenstein. it is used to estimate the rate of evolution under a brownian model. it uses some of the following ideas. the **raw contrast** is the difference between the value of the character at two chosen tips: $c_{ij} = x_i - x_j$. under the brownian model, c_{ij} has expectation zero and variance proportional to $v_i + v_j$. the **standardized contrast** is $s_{ij} = \frac{c_{ij}}{\sqrt{v_i + v_j}}$. it follows that $s_{ij} \sim n(0, \sigma^2)$.

2.3 estimating rates using maximum likelihood

another way to estimate evolutionary rate is by finding the maximum-likelihood parameter values for a brownian motional model. our data is character values at the tips of the tree which we assume was generated with the brownian model. there is a standard formula for the likelihood of drawing from a multivariate normal distribution; remember we wish to find the parameter values that maximize this function; this is done using optimization algorithms.

actually, in some cases we have analytic solutions.

restricted maximum likelihood (reml) refers to an approach which uses ml after transforming the data set to remove nuisance parameters, such as the root state.

2.4 bayesian approach to evolutionary rates

this approach uses explicit priors for parameter values, and then runs an mcmc to estimate posterior distributions of parameter estimates.

2.5 fitting brownian models to multiple characters

motivation: a wide variety of hypotheses can be framed as tests of correlations between continuously varying traits across species. for example, is the body size of a species related to its metabolic rate? how does the head length of a species relate to overall size, and do deviations from this relationship relate to an animal's diet?

we make a distinction between **standard correlation** and **evolutionary correlation**. the first means that one trait predicts the value of another. the second is when two traits evolve together due to a process like mutation, drift, or selection. if there is an evolutionary correlation between two characters, that means we can predict the magnitude and direction of changes in one character given knowledge of evolutionary changes in another.

phylogenetic relatedness alone can lead to a relationship between two variables that are not in face evolving together. in that case, there is a confounding variable – the clade the related species belong to.

2.6 modeling the evolution of correlated characters

multivariate brownian motion. each trait evolves under a brownian motion, but these brownian motions are not necessarily independent, as specified by some covariance matrix \mathbf{r} .

recall that correlations at the tips depend also on branch lengths, specified by some matrix \mathbf{c} (see chapter 3). then the variance-covariance matrix is the kroenker product $\mathbf{r} \otimes \mathbf{c}$, an $nr \times nr$ matrix, where n is the number of leaves of the tree and r the number of traits.

here again, there is a formula for the likelihood of observing data \mathbf{x}_{nr} given the input parameters. also, we can estimate these parameters analytically with formulas similar to the univariate case.

2.7 testing for evolutionary correlations

fit a model with correlation. fit a model without correlation. calculate the optimal log-likelihood in both cases. apply log-likelihood test.

there is also a bayesian approach that i didn't read.

2.8 testing with traditional approaches (pic, pgls)

pgls is **phylogenetic general least squares**. in general least squares, we construct a model of the relationship between the column vectors $x, y \in \mathbb{R}^n$ of trait values and the whose correlation we wish to test:

$$y = x_d b + \epsilon$$

where $x_d = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, b is 2×1 , and the error vector ϵ is assumed to be multivariate normal with mean

zero and some covariance matrix ω . in the brownian model, we assume that the residuals have variances and covariances following the structure of the phylogenetic tree: that is, given by the matrix \mathbf{c} . we carry out standard least square to estimate model parameters. the first term in b is the phylogenetic mean $\bar{z}(0)$, and the second term is an estimate for the slope of the relationship between y and x .

another way to think about a pgl model is that we are treating x as a fixed property of species. the deviation of y from what is predicted by x is what evolves under a brownian motion model.

3 2022-02-01: fitting brownian motion models ot multiple characters

[harmon chapter 5]

many hypotheses can be framed as tests of correlations between continuously varying traits across species. for example is the body size of a species related to the metabolic rate? multivariate gaussians. kroenecker product of the \mathbf{r} and \mathbf{c} matrices. too much typing. didn't take better notes.

4 2022-02-08: continuous traits: ou models

[harmon chapter 6, butler et al 2004]

4.1 beyond brownian motion

[harmon chapter 6]

this chapter considesr ways in which comparative methods can move beyond simple brownian motion models:

1. by transforming the variance-covariance matrix
2. by incorporating variation in rates of evolution
3. by accounting for evolutionary constrains
4. by modeling adaptive radiation and ecological opportunity.

4.1.1 transforming the covariance matrix

three statistical models to test whether data deviates from a constant-rate process evolving on a phylogenetic tree. these involve three kinds of pagel tree transformations, each acting on the covariance matrix \mathbf{c} . they are λ, δ, κ .

- for pagel's λ , you just multiply all the off-diagonal entries of \mathbf{c} by $\lambda \in [0, 1]$. this is like shrinking all the internal branches by a factor of λ but leaving the leaf edges unchanged.
- pagel's δ is for capturing variation in rates of evolution over time. it is done by raising all the entries of \mathbf{c} to the power of δ . depending on whether $\delta > 1$ or $\delta < 1$, this has the effect of stretching or compressing the node heights, with deep and shallow branches affected differently.
- pagel's κ transformation is used to capture patterns of speciation change in trees. in this, all the branch lengths are raised to the power of $\kappa \geq 0$. this transformation isn't really used anymore.

4.1.2 variation in rates of trait evolution across clades

brownian motion assumes that the rate of change σ^2 is constant. there are three ways to test for differences in rate of evolution across clades:

1. compare the magnitude of independent contrasts across clades. the key idea is that when rates are high, we expect to see large independent contrasts in that part of the tree; in particular, the square of an independent contrast is an estimate of the brownian motion rate parameter σ^2 .
2. use model comparison approaches to compare the fit of single- and multiple-rate models to data on trees
3. use a bayesian approach combined with reversible-jump machinery to try to drift the places on the tree where the rate shifts have occurred.

i didn't really understand the last two.

4.1.3 non-brownian evolution under stabilizing selection

ornstein-uhlenbeck (ou) model.

4.1.4 early burst models

test for adaptive radiations by looking for bursts of trait evolution deep in the tree. the simplest method is to use a time-inhomogeneous brownian motion model; that is, a model in which the rate parameter σ^2 varies through time, such as

$$\sigma^2(t) = \sigma_0^2 e^{bt}, \quad (b < 0)$$

this model also generates multivariate normal distributions of tip values. in particular, given model parameters $(\bar{\zeta}_0, \sigma^2, b)$, we can obtain a vector of means and a covariance matrix for the tips. from this, we can use the multivariate gaussian distribution function to calculate a likelihood, which can then be used in an ml or bayesian statistical framework.

4.1.5 peak shift models

punctuated trait shifts. has to do with multivariate models.

4.2 phylogenetic comparative analysis: modeling adaptive evolution

[butler and king 2004]

ou model incorporates both selection and drift. this paper develops the method for one quantitative character. this allows us to test hypotheses regarding adaptation in different selective regimes against data using ml.

if two species are in different selective regimes, we expect their mean phenotypes to differ. but in brownian models, all lineages share the same expected mean phenotype. also we should expect variance in mean phenotypes to remain bounded, but in brownian motion models, this variance grows without bound.

ou model includes brownian motion as a special case. consider the evolution of a quantitative character x along the branch of a phylogenetic tree. we can decompose the change in x into deterministic and stochastic parts:

$$dx(t) = \alpha [\theta - x(t)] dt + \sigma db(t)$$

where α measures the strength of selection and σ measures the intensity of the random fluctuations in the evolutionary process. the parameter θ gives the optimum trait value.

there are two evolutionary interpretations:

- $x(t)$ represents the species' mean phenotype, with θ taking the role of a local maximum in a fitness landscape,
- something else

each lineage in the tree evolves according to its own ou process, which requires that there be one optimum per branch of the phylogeny.

5 uyeda (2013) a novel bayesian method for inferring...'

5.1 abstract and intro

develops reversible-jump bayesian method of fitting multi-optima ou models to phylogenetic comparative data. goal: estimate the placement and magnitude of adaptive shifts from the data.

adaptive landscape. microevolution assumes it's fixed, but this is not reasonable for macroevolution.

software often requires optima to be "painted" onto branches according to researcher's hypotheses.

5.2 methods

ou parameters

$$d\bar{z} = \alpha(\theta - \bar{z}) + \sigma dw$$

here α is the *rate of adaptation*, measured in inverse time units; θ is the optimum value of the process; σ^2 is the per unit time magnitude of uncorrelated diffusion (measured in units of squared trait units per time unit).

the phylogenetic half-life is $\alpha^{-1} \log(2)$, which is measured in time units. this is an easier way to interpret α . it is the amount of time it takes for the expected trait value to get halfway to the phenotypic optimum. if this is small (i.e. α is large) then covariance between species erodes exponentially fast, so that the results will resemble a bm process. otoh, if α is small (ie much shorter than the youngest split on a phylogeny) then the model will resemble a white noise process (i.e., residual trait values are completely uncorrelated). [i don't understand that].

a useful compound parameter for ou process is: $v_y = \sigma^2/2\alpha$, which is **stationary variance**, the equilibrium variance of an ou process evolving around a stationary optimum θ .

multi-optima ou models are a modification of standard ou model in which adaptive optimum θ is allowed to vary accross the phylogeny according to discrete shifts in adaptive regimes. however, the number of shifts and their locations on the phylogeny are not fixed. instead the algorithm presented in this paper estimates those things.

5.3 reversible-jump model

start with a fully resolved phylogeny with n taxa. the optima at the root is θ_0 , and the vector of optima on the tree is $\bar{\theta} = (\theta_0, \dots, \theta_k)$ where k is the number of shifts between adaptive optima. the locations of shifts are mapped onto the phylogeny as locations $\mathbf{l} = \{l_1, \dots, l_k\}$. given these parameters, the distribution of tip states y is multivariate normal with expectation

$$[y|y_0, \alpha, \mathbf{l}, \bar{\theta}] = \mathbf{w}\bar{\theta}$$

where y_0 is the root state and \mathbf{w} a matrix of weights used to calculate the weighted average of adaptive optima, discounted by an exponentially decreasing function that depends on the rate parameter α and the time since the species evolved under a given optima. the full explanation and derivation is not given in this paper.

the covariance matrix for y is given by

$$y_i = \frac{\sigma^2}{2\alpha} [1 - e^{-2\alpha t_{0,i}}]$$

$$y_{ij} = e^{-\alpha t_{ij}} \frac{\sigma^2}{2\alpha} [1 - e^{-2\alpha t_{0,ij}}]$$

where $t_{0,i}$ is the time from the root to species i ; t_{ij} is the total time separating species i and j ; and $t_{0,ij}$ is the total time separating the root and the mrca of species i and j .

what is a pruning algorithm (“we used a pruning algorithm to speed computation and calculate conditional likelihood” p904)?

the reversible-jump framework uses metropolis-hastings algorithm to explore models with varying dimensionality through the course of the mcmc. dimensionality here refers to the number of optima shifts which occur. the amount of time the mcmc spends in a given model is proportional to its posterior probability.

adaptive optima are constructed in a way that makes them unique (i.e. different from the other optima). this was done because it’s easier for the detailed balance condition to be shown to be true in that case.

5.4 simulation study

variation in body sizes in 226 extant chelonians (turtles and tortoises).

i am a little confused about what uyeda and harmon (2013) are doing in the simulation study of chelonian body sizes, and have tried to write out the basic steps of their approach. if i am understanding correctly, the basic procedure is as follows:

1. the input for the simulation study is a phylogenetic species tree t with the 226 extant chelonians as leaves, along with (a) the average body size of the taxa at each leaf of t , and (b) some prior distribution on the space of possible parameters k, l, θ, α , and σ^2 .

2. the mcmc algorithm jumps around the space of possible combinations of parameters; the markov chain is constructed such that its stationary distribution is the posterior probability. therefore in the long run, the proportion of time the markov chain spends in a particular state converges to the posterior probability of that state. after running the markov chain for 500,000 or so steps, a sufficient degree of convergence has taken place, and a sample is taken.

3. repeat step 2 several thousand times obtain several thousand samples.

4. the sample distribution is taken to be our estimate of the posterior distribution. the posterior distribution is a distribution *on the space of parameters*.

5. provided that chelonian evolution is well approximated by the ou process, the posterior distribution is thought to say something informative about where/when optima shifts occurred, intensity of selection, etc (e.g. figure 6)

is this a correct understanding of the basic procedure?

6 omeara et al 2006: testing for different rates of continuous trait evolution using likelihood

6.1 basic properties of character evolution on trees

disparity is commonly measured as variance of the states of the taxa. so higher disparity for a character means the taxa are less similar for the character. bm is the standard model for continuous character evolution. of course disparity depends on both variance of bm (the “rate of character evolution”) and the underlying phylogenetic tree (i.e. divergence times).

suppose we have a tree with n leaves. let c be the $n \times n$ matrix whose ij -th entry is the distance from the root to the mrca of leaves i and j . then

$$[\text{disparity}] = \sigma^2 \left[\frac{1}{n} \text{tr}(c) - \frac{1}{n^2} \mathbf{1}^\top c \mathbf{1} \right]$$

note that extinctions can reduce the age of mrca of a clade. the ml estimator of the rate of phenotypic evolution under bm is

$$\hat{\sigma}^2 = \frac{[x - [x]]^\top c^{-1} [x - [x]]}{n}$$

where x is the column vector of tip observations.

the **noncensored approach** is where each branch is painted with a rate parameter σ^2 . there is a ml estimator for this in which the rates are allowed to vary on different places of the tree. i think harmon went over this and it’s in my notes above. the **censored approach** makes no assumptions about where or how the change has happened somewhere on that branch.

7 2022-02-20: discrete character evolution

7.1 hamon chapter 7: models of discrete character evolution

snakes are lizards that lost their limbs.

Mk Model: analogue to JC model. For evolution of discrete characters that have a set number of fixed states. The Extended Mk Model does sth similar

8 2022-02-21: pagel and meade 2006

this paper is about bayesian analysis of correlated evolution of discrete characters by reversible-jump markov mcmc.

synopsis: the rj markov chain visits these models in proportion to their posterior probabilities, thereby directly estimating the support for the hypothesis of correlated evolution. in addition, the rj markov chain simultaneously estimates the posterior distributions of the rate parameters of the model of trait evolution.

they implement the method by analyzing the question of whether mating system and advertisement of estrus in females have coevolved in the old world monkeys and great apes. in this case, there are two binary traits

- advertisement of estrus
- mating system (multimale mating or not)

which yields four possible states: 00,01,10,11. if the traits have evolved independently, then the rate of change between the two states of one variable will not depend on the state of the other. only states which share exactly one component can transition to each other (e.g. 00 can transition to 10 or 01 but not to 11). there are 8 possible transitions, and the rates for a given transition may or may not equal the rate of a different transition. there are 4140 distinct models wherein each rate is associated with anywhere from 1 to 8 possible rate classes. for example, (1,1,1,1,1,1,1,1) is one possible model, in which all of the rates are the same (i.e. are in rate class 1); under (1,1,1,2,2,2,2,2), the first three transitions share the same rate which is different from the last five; under (1,2,3,4,5,6,7,8) the transition rates are all different.

if a markov chain following this acceptance rule is allowed to run for long enough, it reaches a stationary distribution. at stationarity, a properly constructed chain wanders through the universe of possible states, visiting better and worse models in proportion to their posterior probabilities, rather than inexorably moving toward better outcomes, as an optimizing approach, such as maximum likelihood, would do. the stationary distribution of the chain thereby simultaneously samples the posterior distribution of models of trait evolution, the parameters of these models, and the likelihood of the data.

they describe a number of moves on the space $\{(x_1, \dots, x_8) : x_i \in [8]\}$ of rate models. note that 15 of the 4140 rate models correspond to independent evolution.

they use the “bayes factors” (bf) test, defined as

$$\text{bf}_{ij} = \frac{l[d|m_i]}{l[d|m_j]}$$

where i, j are models. it’s just the ratio of marginal likelihoods. if $\text{bf}_{ij} > 1$ then the average likelihood of the data under model i is greater than that for model j . bayes factors can be estimated with mcmc. there is some criteria for determining if a bf is sufficiently large to be strong or very strong evidence of correlated evolution.

In addition, 99.86% of the models at convergence included at least one rate in the zero bin. BF of about 28 gives strong evidence of correlated evolution.

Figure 7 is helpful. Interpreted adaptively, mating system changes first in evolution, and this selects for a change in the female display strategy. The alternative hypothesis that females alter their displaying strategy and that this selects for a multimale mating system is not supported.

8.1 Question

I once (naively) implemented a Metropolis-Hastings MCMC algorithm in an attempt to uniformly sample paths from one corner of a hypercube to another. As the dimension of hypercube increased, it would tend to get “stuck,” endlessly rejecting proposed moves. This limited the ability to explore the sample space.

Are there general criteria for predicting whether MCMC is likely to be effective for a given sample space?

9 2022-02-28: Ives and Garland 2009: Phylogenetic Logistic Regression for Binary Dependent Variables

This paper develops statistical methods for phylogenetic logistic regression in which the dependent variable is binary and not independent between species. That is, species more closely related are more likely to have the same character value. These methods give a way to estimate the phylogenetic signal of binary traits.

Takeaway: phylogenetic logistic regression should be used rather than standard logistic regression.

Phylogenetic signal is the tendency for related species to resemble each other. Multispecies datasets (e.g. for the relation of brain size to body size) should be analyzed with methods that account for this. A tree is **hierarchical** if it is not a star tree. Under Brownian Motion assumption of trait evolution, there will be some amount of phylogenetic signal. But this resemblance between species violates one or more assumptions of most common statistical methods, ie independence.

If the only concern is hypothesis testing, then it is possible to apply conventional statistical methods (eg ANOVA), apply F ratios, and compare those test statistics with null distributions that have been derived by simulated or randomizing data in accordance with a specified phylogenetic tree and assumed model of evolution.

The method of phylogenetically independent contrasts (PIC) assumes a known phylogenetic tree and an assumed a model of character evolution can be used to transform tip data to make them have equal expected variances and remove correlations related to phylogenetic signal. [Not really understand PIC](#). PIC is a special case of phylogenetic least squares methods. Under Brownian motion model of character evolution, PGLS and PIC calculations give the same parameter estimates and statistical tests.

For this paper, there is a binary dependent variable Y and zero, one or more independent variables X , which can be either continuous or discrete. The approach involves generalized linear models (GLMs) that can be used to analyze data from any exponential family of distributions. This method does not require apriori assignment of phylogenetic signal. However this means that the statistical tests will not be the same since there is additional uncertainty from this estimation in the model.

The model presented in this paper gives a plausible model that can be used for statistical analyses. There are two steps (multivariate case): First, values of Y evolve up the phylogenetic tree with asymptotic probability of being in state 1 equal to μ and transition rate α . This gives a Y value of 0 or 1 to each species. Second, the value of Y_i for each species evolves toward either 0 or 1 depending on the values of the independent variables X^j and the regression coefficients b_j . No assumption is made about the distribution of independent variables, or even that they contain phylogenetic signal.

Output: estimates of regression coefficients that account for phylogenetic correlations. An estimate of the strength of phylogenetic signal in the residual variation.

I am still trying to understand the multivariate case in the Ives and Garland paper. The process is divided into two components. In the first component, a binary character evolves along the phylogenetic tree according to some Markov process starting at the root and having transition rates $\alpha = (\alpha_0, \alpha_1)$ and stationary distribution μ which is some function of these rates. This yields a vector Y of binary states corresponding to the leaves of the tree. (So far so good). In the second component, “we assume the value of Y_i evolves toward either 0 or 1 depending on the values of [the species-specific independent variables], with the rate of evolution no longer depending on the transition rate α but instead depending on the regression coefficients b_j for the independent variables X^j ”.

Does this second step change the values of Y obtained in the first step?

10 2022-02-28: Hipp et al 2017: Sympatric parallel Diversification of major oak clades

Used 300 samples from 146 species of oak trees using RAD-seq sequencing. An ML phylogeny was inferred and analyzed alongside other types of data to reconstruct the evolutionary history of American oaks.

We find a sympatric parallel diversification in niche, leaf habit, and diversification rates. We find that oaks adapted rapidly to niche transitions.

Key part: This paper investigated the number of transitions in leaf habit for the phylogeny as a whole. Leaf habit is treated as a binary trait. Phylogenetic regression was used to assess the partial effects of various variables (temperature seasonality, moisture stress, etc) on variance in leaf habit. Analyses were performed with phylogenetic logistic regression, the technique introduces by today's other paper: Ives and Garland 2009.

Results for this: Mexican oak clades transition between deciduous and evergreen habit twice as frequently as the remainder of the oaks. The phylogenetic logistic regression model and the "generalized estimating equations" (a method from a different paper) both point to two models as the "best fit": one in which all the factors are included, and one in which only moisture stress and mean temperature in the warmest quarter are included. They don't agree on which is best, however. Don't know which method is better.

11 Questions

What is a jackknife dataset?

I don't understand how to interpret bootstrap support (for example in figure 2 of Hipp 2017). As well as the bootstrap confidence intervals and bootstrap p values in Ives and Garland.

What is parallel sympatric diversification?

12 2022-03-03: Revell 2013: Two new graphical models for mapping trait evolution on phylogenies

Two different graphical models are presented for visualizing phenotypic evolution on the tree.

1. Method 1: a new approach for plotting the posterior density of stochastically mapped character histories for a binary (two-state) phenotypic trait on a phylogeny
2. Method 2: a closely related technique that uses ancestral character estimation to visualize historical character states for a continuous trait along the branches of the tree.

The second method has the downside of ignoring the uncertainty about ancestral traits. In this paper, a "traitgram" is presented which allows visualization of ancestral state uncertainty.

An interesting equation from Felsenstein is used in both papers

$$\hat{a} = \frac{\frac{x_i}{v_i} + \frac{x_j}{v_j}}{\frac{1}{v_i} + \frac{1}{v_j}}$$

to interpolate the value \hat{a} of a trait on an edge by its values x_i and x_j at phylogenetic distances v_i and v_j away. This can be interpreted as a weighted harmonic mean (as on the wikipedia page for harmonic mean, with $w_i = x_i/v_i$). Is there a simple intuition why this is the appropriate way to interpolate character values?

Also, was BUCKy was named after the badger?

Stochastic character mapping involves sampling possible histories of a discretely valued character from their Bayesian posterior distribution. When you have thousands of such character maps, you can estimate the history of a discretely-valued character trait which evolved on the tree. Unfortunately, it's hard to visualize this on a figure for publication.

13 2022-03-03: Kriebel 2014: Discovery of unusual anatomical and continuous characters in the evolutionary history of *Conostegia*

Conostegia is a genus of flowering plants native to the Americas. This paper finds it to be paraphyletic. Unusual characters were documented and their evolution studied in light of the phylogeny. The importance of documenting anatomical and continuous characters is demonstrated.

Use of both binary and continuous characters. The continuous characters used were: herkogamy and seed size.

Two methods used to reconstruct ancestral states: SIMMAP – stochastic character mapping ACE – ancestral character estimation

13.1 Phylogenetic Analyses

MAFFT to make alignments. ML analyses done using RAxML. Bayesian analyses done on partitioned dataset using MrBayes. Convergence analyses using Tracer v1.5 and AWTY. Bayesian concordance analyses (BCA) implemented in BUCKy,

13.2 Presentation about Kriebel

Researchers were interested in determining if the genus *Conostegia* is monophyletic, also how did the traits evolve and what are the relationships within *Conostegia*?

Sampled 31 species of *Conostegia*.

Used a combined dataset of both plastid and nuclear markers. 15 characters scored: 13 binary (e.g. calyx teeth present or absent) and 2 continuous (e.g. seed length).

You can have discordances between different trees as a result of ILS, hybridization, etc. Bucky evaluates discordant trees.

There were some differences between concatenated vs concordance trees, suggesting the possibility of ancient hybridization or chloroplast capture.

14 The Dark Side 2022-04-12

Some of the challenges and limitations of phylogenetic comparative methods. Roadmap

1. Intro
2. Unsolved challenge to correlation tests for categorical characters (Maddisio and FitzJohn 2014)
3. Model inadequacy and mistaken inference (Rabosky and Goldberg 2014)
4. Tying the readings together

14.1 Intro

Comparative methods has limitations – how can we overcome these limitations to better our understanding of evolutionary biology. We can look at huge datasets to test hypotheses about adaptation, ancestral states, rates of evolution and diversification, etc.

Two major limitations:

1. State-dependent models consider only a relatively small set of possible models
2. If we apply the tests to empirical phylogenetic trees, even with made-up data or neutral characters, we almost always find statistically significant results. There are Type-I errors.

There are some ways to deal with these issues, according to Harmon

1. Compare statistical support with the support one obtains from random data
2. Make sure the events we test are **replicated** in our data. This means we need to see the events in *more than one clade*.
3. Explicitly consider the possibility that some unmeasured character is what is influencing diversification rates.

14.2 Paper 1: The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters

Main points of the paper:

- We might mistakenly conclude that there is an adaptive or functional relationship when a trait in a lineage is accompanied by a second trait's evolutionary change
- Methods will indicate statistically significant association even if codistribution arose only once
- Flaws due to simplified models (sth)
- Upshot: even if the models adjusted to predict such distributions, they would still be susceptible to "pseudoreplication"

Figure 1 in the paper is pretty clear. In (1a), we see replicated codistribution. In (1b) we see replicated bursts. Both of these cases should be taken as evidence of a functional/adaptive relationship, and the number of samples is 6, as this is the number of times the coincidence arose. In (1c) and (1d) we see examples where coincidences are NOT good evidence because the number of samples is essentially one – codistribution in these cases is more likely due to one case where the traits coincided, and then were inherited together. This results in their descendants being **pseudoreplicates**, because they appear to be examples of replication, but the "replication" is not independently-occurring (due to the shared phylogeny). Of course something could be happening in (1c) and (1d), but we just don't have good statistical evidence in those cases.

A cool example of (1b) is plant carnivory and evolution of pitcher-like pitfalls to trap prey. Floral structure vs pollinator visitors. Vanilla plants: island habitat vs leaf loss. Cats: pleiotropy (one gene has multiple functions) – whiskers vs penis hooks. Maybe hard to find examples where trait evolves independently > 6 times; herbaceous vs woody is an example. There are about 13 times where aquatic plants transitioned to terrestrial.

Figure 3 gives more examples which are less clear-cut, and as-yet unanswered. For example, does replication across *two* clades constitute good statistical evidence? For example the examples of codistribution in (3d) and (1b) look very similar, so where is the quantitative threshold for statistical significance between these two cases?

14.3 Paper 2: Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation

Main points

- Associations between character states ...

Figure 1 and Figure 2. Phylogeny of extant whales. Is body size associated with diversification? Figure 2 involved resampling the tree topology to obtain pseudo p-values for whether the results were statistically significant. Neutral traits found to be highly significant. Should start with a hypothesis rather than generating one after looking at the data. Should test the shape of the tree against neutral characters. Some tree shapes are likely to give rise to spurious significance results for random neutral characters.

Significance needs to be understood in relative, rather than absolute terms, i.e. such as by comparing a character with other characters hypothesized to be neutral.

15 Hidden Markov Models

I am presenting on the Beaulieu et al. paper entitled “Identifying Hidden Rate Changes in the Evolution of a Binary Character.” The link to the paper is

<https://doi.org/10.1093/sysbio/syt034>

TLDR: The paper introduces a method of identifying different rates of evolution in a binary character along different branches of a phylogeny and applies this method to explore a flowering plant clade comprised of 35,000 species.

Character Trait: An example of a binary character considered in this paper is whether a plant species is woody (W) or herbaceous (H). This character trait can exhibit tremendous intraclade variation, with apparently multiple shifts between these growth forms.

Problem: Erroneous assumptions about transition rates along different edges of a species tree could lead to estimation error. This is especially true for large phylogenetic trees (e.g. trees with tens of thousands of tips) where variation in transition rates is to be expected.

15.1 Methods: Hidden Rates Model

Definition 3 (Time-Homogenous Model). *The time-homogenous model (TH model) involves estimating ancestral states using maximum-likelihood under the assumption that the binary trait of interest evolves with constant transition rates which are the same everywhere on the species tree. That is, at any time and at any point on an edge of the species tree, the rate of transition from W to H or H to W is governed by the rate matrix*

$$Q = \begin{bmatrix} 0 & q_{W \rightarrow H} \\ q_{H \rightarrow W} & 0 \end{bmatrix}$$

Remark 1. *The above model assumes, for example, that if the transition rate from H to W is fast in one place on the tree, then it must be fast everywhere on the tree. Such assumptions are unrealistic. This is the problem the HRM is intended to address.*

The HRM model assumes that the transition rate along an edge of the species tree may belong to any one of a certain number of “rate classes”. For example, there might be two rate classes, fast (F) and slow (S). The rate classes are treated as a “hidden variable” in the sense that it is assumed that we cannot observe it directly. This gives a single model governed by the 4×4 rate matrix given by equation (3) in the paper, as we have four states: $\{WS, HS, WF, HF\}$.

Next they discuss a biological interpretation of this four-state model. The takeaway is that the additional states allow the HRM to paint a phylogeny with areas where transitions happen frequently or infrequently due to unmeasured characters that affect the rates.