Botany 940 Notes

February 14, 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 $\mathbb{E}\left[\sigma_a^2(t)\right] \approx 2N_e\sigma_m^2$ 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:

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

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:

$$\mathbb{E}\left[\sigma_a^2(t)\right] = \left(1 - \frac{1}{2N_e}\right)^t \left[\sigma_a^2(0) - 2N_e\sigma_m^2\right] + 2N_e\sigma_m^2 \tag{1}$$

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 $\left(1 - \frac{1}{2N_e}\right)^t \to 0$ as $t \to \infty$, we expect $\mathbb{E}\left[\sigma_a^2(t)\right]$ 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,

$$\sigma_B^2(t) = \frac{t\sigma_a^2(t)}{N_e} \tag{2}$$

Substituting the equilibrium value $2N_e\sigma_m^2$ in for $\sigma_a^2(t)$ gives

$$\sigma_B^2(t) = 2t\sigma_m^2 \tag{3}$$

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,*

$$h^2 = \frac{\sigma_a^2}{\sigma_w^2}. \tag{4}$$

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

$$h^2 = \frac{2N_e\sigma_m^2}{\sigma_w^2} \tag{5}$$

or equivalently,

$$\sigma_m^2 = \frac{h^2\sigma_w^2}{2N_e} \tag{6}$$

and therefore by (3),

$$\sigma_B^2(t) = \frac{h^2\sigma_w^2 t}{N_e}. \tag{7}$$

1.4 Brownian Motion under Selection

Recall we with 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

$$\sigma_B^2 = \left(\frac{h^2 \sigma_w^2}{N_e} + \sigma_s^2 \right) t \tag{8}$$

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

$$\sigma_B^2 \approx \sigma_s^2 \tag{9}$$

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 thorugh 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

$$\sigma_B^2(t) = \frac{h^2 \sigma_w^2 t}{N_e} \tag{10}$$

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 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 idependent contrasts (PICs) are a way to estimate the rate of character change across a phylogeny. There is an algorithm, the algorithm of independent contrats, 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 $s_{ij} = \frac{c_{ij}}{v_i + v_j}$. 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.

Restrictied Maximum Liklihood (REML) refers to an approach which uses ML after transforming the data set to remove nuisance paramters, 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 **R**.

Recall that correlations at the tips depend also on branch lengths, specified by some matrix C (see chapter 3). Then the variance-covariance matrix is the Kroeneker 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 $\overline{z}(0)$, and the second term is an estimate for the slope of the relationship between y and x.

Another way to think about a PGLS 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 speciational 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 multipe-rate modelds 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 deepin the tree. The simplest method is to use a time-inhomogenous 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:

$$dX(t) = \alpha \left[\theta - X(t)\right] dt + \sigma dB(t)$$

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

Devolops 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 shorted 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 $\bar{\theta} = (\theta_0, \ldots, \theta_K)$ where K is the number of shifts between adaptive optima. The locations of shifts are mapped onto the phylogeny as locations $\mathbf{L} = \{L_1, \ldots, L_K\}$. Given these parameters, the distribution of tip states Y is multivariate normal with expectation

$$\mathbb{E}\left[Y|Y_0, \alpha, \mathbf{L}, \bar{\theta}\right] = \mathbf{W}\bar{\theta}$$

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

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

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

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 $K, \mathbf{L}, \theta, \alpha$, and σ^2.

2. The MCMC algorithm jumps around the space of possible combinations of parameters; the Markov chain is constructed such that 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 suffcient 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

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

Note that extinctions can reduce the age of MRCA of a clade. The ML estimator of the rate of phenotypic evolution under BM is

$$\widehat{\sigma}^2 = \frac{\left[X - \mathbb{E}\left[X\right] \right]^\top C^{-1} \left[X - \mathbb{E}\left[X\right] \right]}{N}$$

where X is the column vector of tip observations.

The **noncensored approach** is where each branch is painted with a rate parameter σ^2.