1 Botany 940 Notes

1.1 2021-02-01: Brownian Motion

1.1.1 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.*

1.1.2 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 $2N_e\sigma_m^2$. $\qquad\square$

1.1.3 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),

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

1.2 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.

1.2.1 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

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

1.2.2 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 thorough 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.

1.2.3 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.

1.3 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 Fitting Brownian Motion Models to Single Characters

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.

Restricted 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.

3 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.

3.1 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 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.

3.2 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.

3.3 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 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.