

Appendix from M. A. Butler and A. A. King, “Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution” (Am. Nat., vol. 164, no. 6, p. 683)

Mathematical Details

The Hansen Model

We begin with a phylogeny with known topology and branch lengths. Let N denote the number of terminal taxa. For definiteness, let us think of these taxa as species. Each species i has a lineage, which is the path of branch segments traversed from the root (most basal node) of the tree to the i th terminus.

Let $X_i(t)$ denote the state of the i th lineage (i.e., the value of the quantitative character of interest) at time t . At the root of the tree, $t = 0$, and at the terminal taxa, $t = T$. Thus $X_i(0)$ is the state of the root of the phylogeny, and $X_i(T)$ is the current state of extant species i .

The model of Hansen (1997) assumes that for each lineage the quantitative character of interest evolves according to an Ornstein-Uhlenbeck (OU) process

$$dX_i(t) = \alpha[\beta_i(t) - X_i(t)]dt + \sigma dB_i(t), \quad (\text{A1})$$

$0 \leq t \leq T$. The function $\beta_i(t)$ represents the optimum trait value and as such identifies the selection regime acting on lineage i over the course of its history. The symbols $dB_i(t)$ denote increments of a standard Brownian motion (BM); heuristically, we can think of them as normal random variables with mean 0 and variance dt . We define the correlation ρ_{ij} of these random terms by the equation

$$\text{Cov}[dB_i(t), dB_j(t)] = \rho_{ij}(t)dt.$$

(Note that ρ_{ij} may depend on time.) It is well known that equation (A1) defines a Gaussian process, and it is elementary to compute its moments. In particular, we have

$$E[X_i(T)|X_i(0) = \theta_0] = \theta_0 e^{-\alpha T} + \int_0^T \alpha e^{-\alpha t} \beta_i(T-t) dt, \quad (\text{A2})$$

$$\text{Cov}[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] = \sigma^2 \int_0^T e^{-2\alpha t} \rho_{ij}(T-t) dt. \quad (\text{A3})$$

These equations are true in more generality than we will require. Let us therefore make assumptions on the forms of $\beta_i(t)$ and $\rho_{ij}(t)$ to adapt these general results to our present needs. In particular, the history of each species is marked by the occurrence of major events, including speciation events and changes in selective regime. We call the times at which these events occur “epochs.” The history of the i th lineage consists of a number, $\kappa(i)$, of sequential branch segments demarcated by epochs $0 = t_i^0 < t_i^1 < t_i^2 < \dots < t_i^{\kappa(i)} = T$ (fig. A1). Following Hansen (1997), we will assume that the selection regime, β_i , acting on lineage i is constant with value β_i^γ over the course of the γ th branch segment, that is, for $t_i^{\gamma-1} < t < t_i^\gamma$. (Throughout the appendix, we use Latin subscripts [i, j , etc.] to denote lineages and Greek superscripts [γ, η , etc.] to refer to epochs.) Under this assumption of piecewise-constant selection regimes, equation (A2) takes the form

$$E[X_i(T)|X_i(0) = \theta_0] = e^{-\alpha T}\theta_0 + \sum_{\gamma=1}^{\kappa(i)} e^{-\alpha T}(e^{\alpha t_i^\gamma} - e^{\alpha t_j^{\gamma-1}})\beta_i^\gamma. \quad (\text{A4})$$

With regard to the correlations $\rho_{ij}(t)$, it is clear that if s_{ij} denotes the time of the speciation event at which lineages i and j diverged, then $\rho_{ij}(t) = 1$ for $t < s_{ij}$. Further, let us assume that after diverging, distinct lineages evolve independently, so that $\rho_{ij}(t) = 0$ for $t > s_{ij}$. Then, on defining $V_{ij} = \text{Cov}[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0]$, equation (A3) becomes

$$V_{ij} = \frac{\sigma^2}{2\alpha} e^{-2\alpha(T-s_{ij})}(1 - e^{-2\alpha s_{ij}}). \quad (\text{A5})$$

Note that the dependence of V_{ij} on σ is particularly simple. In particular, $V_{ij} = \sigma^2 \tilde{V}_{ij}$, where \tilde{V}_{ij} depends only on α . Observe too that expressions (A2)–(A5) are regular at $\alpha = 0$, that is, as $\alpha \rightarrow 0$, $E[X_i(T)|X_i(0) = \theta_0] \rightarrow \theta_0$, and $V_{ij} \rightarrow \sigma^2 s_{ij}$, in agreement with the results for BM.

Selective Optima

We view the phylogeny as given. That is, the epochs t_i^γ and branching times s_{ij} are assumed known. To specify a model of evolution along this phylogeny, it remains to estimate the parameters. However, as written, the model has one β_i^γ for each branch and hence more parameters than termini. To reduce the number of parameters that must be estimated, we assume that only a small number, r , distinct selective regimes have operated and that each is defined by a single optimum θ_k , $k = 1, \dots, r$. That is, we replace each branch optimum β_i^γ with the θ_k corresponding to the selective regime operative on that branch. In the terminology of our examples, in which we painted particular branches with the color of their hypothesized selective regime, β_i^γ is the color of the branch ending in epoch t_i^γ and the set $\theta_1, \dots, \theta_r$ is the palette of colors from which we can choose. To express this idea mathematically, we posit that each branch optimum β_i^γ depends linearly on $\theta_1, \theta_2, \dots, \theta_r$:

$$\beta_i^\gamma = \sum_{k=1}^r \beta_{ik}^\gamma \theta_k. \quad (\text{A6})$$

For our purposes, we can assume that each β_{ik}^γ is either 0 or 1, and for each γ and i , there is an index k' such that $\beta_{ik'}^\gamma = 1$ and $\beta_{ik}^\gamma = 0$ for all $k \neq k'$. Of course, self-consistency requires that $\beta_{ik}^\gamma = \beta_{jk}^\gamma$ whenever lineage i and j share the branch ending in epoch $t_i^\gamma = t_j^\gamma$. To return to the painting analogy, the matrix β_{ik}^γ is the mapping of the colors onto the branches.

The parameters $\theta_0, \theta_1, \dots, \theta_r$, together with the selection strength, α , and the drift strength, σ , must be estimated. Maximum likelihood (ML) estimation of these parameters takes a particularly simple form, as we shall now show.

Maximum Likelihood Estimates

In the following, it will be convenient to make use of matrix notation. Accordingly, we collect our random variables $X_i(t)$ in the vector $\mathbf{X}(t)$ and our quantitative data in the vector \mathbf{x} , the components of which are $x_i = x_i(T)$, the observed values of the quantitative character in each species i , $i = 1, \dots, N$, at the end of an evolutionary process of length time = T . Inspection of equation (A4) in light of equation (A6) shows that

$$E[\mathbf{X}(T)|\mathbf{X}(0) = \theta_0] = \mathbf{W}(\alpha)\theta,$$

where $\theta = (\theta_0, \theta_1, \dots, \theta_r)'$ and the weight matrix \mathbf{W} has entries

$$\begin{aligned} W_{i0} &= e^{-\alpha T}, \\ W_{ik} &= e^{-\alpha T} \sum_{\gamma=1}^{\kappa(i)} \beta_{ik}^\gamma (e^{\alpha t_i^\gamma} - e^{\alpha t_j^{\gamma-1}}), \end{aligned} \quad (\text{A7})$$

for $i = 1, \dots, N$ and $k = 1, \dots, r$. It follows from the multivariate normality of $\mathbf{X}(T)$ that the likelihood of the parameters α , σ , and θ , given the data $\mathbf{X}(T) = \mathbf{x}$, is

$$\mathcal{L}(\alpha, \sigma, \theta | \mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \sigma^{2N} \det \tilde{\mathbf{V}}}} \exp \left[-\frac{(\mathbf{x} - \mathbf{W}\theta)' \tilde{\mathbf{V}}^{-1} (\mathbf{x} - \mathbf{W}\theta)}{2\sigma^2} \right].$$

Maximizing \mathcal{L} is equivalent to minimizing $U = -2 \log \mathcal{L}$. We have

$$U(\alpha, \sigma, \theta | \mathbf{x}) = N \log 2\pi\sigma^2 + \log \det \tilde{\mathbf{V}} + \frac{1}{\sigma^2} (\mathbf{x} - \mathbf{W}\theta)' \tilde{\mathbf{V}}^{-1} (\mathbf{x} - \mathbf{W}\theta).$$

Taking derivatives of U with respect to σ and θ , we see that U has a minimum only if $\sigma = \hat{\sigma}$ and $\theta = \hat{\theta}$, where

$$\begin{aligned} \hat{\sigma}^2(\alpha, \theta) &= \frac{1}{N} [\mathbf{x} - \mathbf{W}(\alpha)\theta]' \tilde{\mathbf{V}}^{-1}(\alpha) [\mathbf{x} - \mathbf{W}(\alpha)\theta], \\ \hat{\theta}(\alpha) &= (\mathbf{W}' \tilde{\mathbf{V}}^{-1} \mathbf{W}) \mathbf{W}' \tilde{\mathbf{V}}^{-1} \mathbf{x}. \end{aligned} \tag{A8}$$

Finding ML estimates for α , σ , and θ is thus equivalent to minimizing the function of one parameter

$$U(\alpha, \hat{\sigma}(\alpha), \hat{\theta}(\alpha)) = N[1 + \log 2\pi\hat{\sigma}^2(\alpha, \hat{\theta}(\alpha))] + \log \det \tilde{\mathbf{V}}(\alpha). \tag{A9}$$

Nonlinear optimization of this sort is a well-studied problem, and numerous public-domain algorithms exist for its solution. Having determined $\hat{\alpha}$ by minimization of equation (A9), one computes $\hat{\theta}$ and $\hat{\sigma}$ by means of equation (A8). Pseudocode for the fitting algorithm is given below in “Maximum Likelihood Parameter Estimation Algorithm”; computer code for use with the free software packages R (<http://www.r-project.org>), OCTAVE (<http://www.octave.org>), and the commercial program MATLAB (<http://www.mathworks.com>) is available on the authors’ Web site (<http://www.tiem.utk.edu/~king>).

Maximum Likelihood Parameter Estimation Algorithm

Given:

1. Data on N extant species

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}.$$

2. A phylogeny relating all extant species, that is, epochs t_i^γ and branching times s_{ij} .
3. An association painting evolutionary optima onto branches, that is, β_{ik}^γ .

Procedure:

1. Guess α .
2. Compute the $N \times (R + 1)$ weight matrix \mathbf{W} from equation (A7).
3. Compute the scaled covariance matrix $\tilde{\mathbf{V}}$ by means of

$$\tilde{V}_{ij} = \frac{e^{-2\alpha(T-s_{ij})} (1 - e^{-2\alpha s_{ij}})}{2\alpha}.$$

4. Compute $\hat{\theta}$ and $\hat{\sigma}$ by means of equation (A8).
5. Compute U using equation (A9).
6. Return to step 2 until the value $\hat{\alpha}$, which minimizes U , is found.
7. For $\alpha = \hat{\alpha}$, compute $\hat{\theta}$ and $\hat{\sigma}$ by means of equation (A8).

Notes on the Implementation

Although the formalism described above mathematically specifies the parameter estimation and model selection procedures, the implementation of the method on a digital computer poses a small number of difficulties. Estimation of model parameters requires one to minimize a scalar function U (eq. [A9]) over a single scalar parameter α subject to the constraint $\alpha \geq 0$. Numerous algorithms for constrained minimization exist in the literature and as computer programs. Numerical difficulties arise, however, in the computation of U for very small and for very large values of α . As $\alpha \rightarrow 0$ —that is, as the OU model approaches BM—the optima θ_k , $k = 1, \dots, r$, become progressively less identifiable. That is, the likelihood profile in the corresponding direction becomes flat. Of course, it is then very difficult to estimate reliably these parameters. As α grows very large, on the other hand—that is, as selection becomes very strong—the influence of all selective regimes other than those in which the terminal twigs lie becomes progressively weaker. Hence, the estimates of the θ_k associated with these regimes become less and less reliable.

To cope with these difficulties, we adopted the following procedure. First, we performed the numerical minimization on $\log \alpha$ to facilitate estimation when α is small. To avoid the numerical difficulties associated with large and small α , we imposed the constraint that $0.001 < \alpha < 20$ in the optimization algorithm itself. The bootstrap confidence intervals of table 4 were computed using those replicates for which the OU(7) model was preferred to the BM model (using both SIC and AIC; in 84 of 10,000 replicates, BM was preferred). Replicates that resulted in estimates up against either of the constraints were discarded; when the algorithm returns an estimated α against one of these constraints, it is to be interpreted as an indeterminate and therefore invalid result. Of the 10,000 replicates, 236 were such that the lower constraint was effective, and 42 were such that the upper constraint was effective.

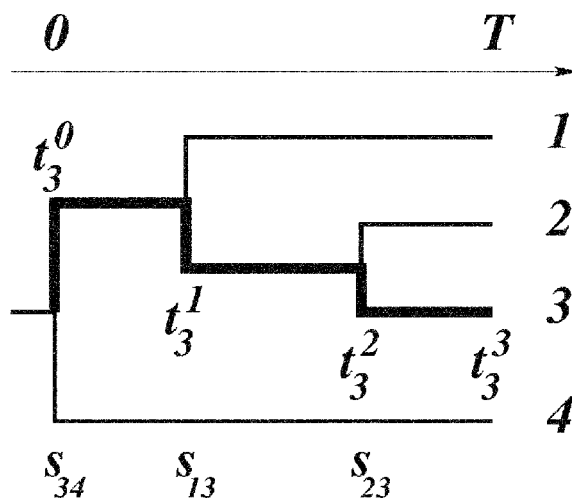


Figure A1: Schematic phylogenetic tree. The epochs t_i^γ and divergence times s_{ij} for lineage $i = 3$ are labeled. In this example, the history of lineage 3 is composed of $\kappa(3) = 3$ branch segments.