

Supplementary Information

September 14, 2014

1 Introduction

1.1 Terminology

We consider a continuous biological trait X evolving over a rooted bifurcating phylogenetic tree. At each node n , the biological trait has value x_n . Each node n has two descendant nodes, i and j , unless n is a tip. Each node n has a parent node p , unless n is the root. The branch connecting node n to its parent has length t_n .

1.2 Independent contrasts

Independent contrasts are calculated using an algorithm that traverses the phylogeny from tips to root, iteratively calculating transformed trait values X' and transformed branch lengths T' along the way. Following Felsenstein [1985] the independent contrast associated with node n , IC_n , is defined as:

$$IC_n = \frac{(x'_i - x'_j)}{\sqrt{t'_i + t'_j}} \quad (1)$$

where $x'_n = x_n$ and $t'_n = t_n$ when x_n is known with certainty (for example when node n is a tip on the phylogeny associated with an empirically observed trait

value) and otherwise

$$x'_n = \frac{x'_i/t'_i + x'_j/t'_j}{1/t'_i + 1/t'_j} \quad (2)$$

$$t'_n = t_n + \frac{t'_i t'_j}{t'_i + t'_j} \quad (3)$$

1.3 Maximum likelihood ancestral state reconstruction

Consider a node n connected to its parent p_n by a branch of length t_n . The change in the value of the evolving trait along this branch, $x_n - x_{p_n}$ is normally distributed with mean zero and variance proportional to t_n . The likelihood of the tree with respect to some candidate ancestral state reconstruction \mathbf{X} is given by the product of the normal distributions associated with each branch:

$$L(\mathbf{X}; \mathcal{T}) \propto \prod_n \phi(x_n - x_{p_n}; 0, \sqrt{t_n}) \quad (4)$$

where $\phi(x; \mu, \sigma)$ is the probability density at x of the normal distribution with mean μ and variance σ^2 . The task of maximum likelihood ancestral state reconstruction is to identify the $\hat{\mathbf{X}}$ which maximizes the likelihood function in (4).

Maddison (1994) notes that, in addition to the global maximum likelihood ancestral state reconstruction $\hat{\mathbf{X}}$, there exists a local maximum likelihood ancestral state reconstruction $\tilde{\mathbf{X}}$ such that each \tilde{x}_n is the ancestral state maximizing the likelihood of the subtree rooted at node n . Since (4) is a product of normal distributions, the local likelihood function at node n is also a normal distribution which we denote $N(\tilde{\mu}_n, \tilde{\sigma}_n)$. Evidently the local maximum likelihood estimate of \tilde{x}_n must be equal to $\tilde{\mu}_n$. The parameters $\tilde{\mu}_n$ and $\tilde{\sigma}_n$ for each n can be estimated in a traversal from the tips of the phylogeny to its root. For each tip of the phylogeny with known trait value x_{tip} , we have $\mu_{tip} = x_{tip}$ and $\sigma_{tip} = 0$ if the tip trait value is known with certainty or a nonzero value if the tip trait's distribution is known. For each internal node n with two children i and j connected by branches of length t_i and t_j , the local likelihood function

with respect to x_n is given by the product of the local likelihood functions at i and j suitably weighted by the branch lengths under the Brownian assumption of additive variance [Bromiley, 2003]:

$$\phi(x_n; \tilde{\mu}_n, \tilde{\sigma}_n) = \phi\left(\tilde{\mu}_i - x_n; 0, \sqrt{\tilde{\sigma}_i^2 + t_i}\right) \phi\left(\tilde{\mu}_j - x_n; 0, \sqrt{\tilde{\sigma}_j^2 + t_j}\right) \quad (5)$$

$$= \phi\left(x_n; \tilde{\mu}_i, \sqrt{\tilde{\sigma}_i^2 + t_i}\right) \phi\left(x_n; \tilde{\mu}_j, \sqrt{\tilde{\sigma}_j^2 + t_j}\right) \quad (6)$$

$$= \phi\left(x_n; \frac{\tilde{\mu}_i(\tilde{\sigma}_j^2 + t_j) + \tilde{\mu}_j(\tilde{\sigma}_i^2 + t_i)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)}, \sqrt{\frac{(\tilde{\sigma}_i^2 + t_i)(\tilde{\sigma}_j^2 + t_j)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)}}\right) \quad (7)$$

from which the values of $\tilde{x}_n = \tilde{\mu}_n$ and $\tilde{\sigma}_n$ can be read off. Note that Equation 6 arises from the symmetry of the normal distribution, and the fact that the distributions can be shifted with respect to their means in order to be likelihoods around $\tilde{\mu}_i$ and $\tilde{\mu}_j$ rather than around zero. This reformulation allows application of the formulae for products of normal distributions presented by Bromiley [2003]. Further, note that the variance term $\tilde{\sigma}_i^2$ is transformed by adding t_i when one might expect it to be transformed by adding $t_i v$, here v is a constant term for the rate of evolution. In practice this constant term does not affect the value of the point estimate for the ancestral state at an internal nodes, only the height of the likelihood function at that point, so it can be disregarded here. A maximum likelihood estimate for the rate of evolution can easily be obtained from the ancestral state estimates themselves and is used in calculation of likelihood scores (see below) but is safely ignored, for convenience, from the formula above. The same argument holds for t_j .

Similarly we denote the maximum likelihood function at node n with respect to x_n as $N(\hat{\mu}_n, \hat{\sigma}_n)$. As mentioned above, $\hat{\mu}_{root} = \tilde{\mu}_{root}$ and $\hat{\sigma}_{root} = \tilde{\sigma}_{root}$. For other nodes of the phylogeny for which the maximum likelihood function must be estimated, we follow a similar logic (conditioning on the children of the focal

node) to obtain:

$$\phi(x_n; \hat{\mu}_n, \hat{\sigma}_n) = \phi(\tilde{\mu}_n - x_n; 0, \tilde{\sigma}_n) \phi(x_n - \hat{\mu}_{p_n}; 0, \sqrt{t_n}) \quad (8)$$

$$= \phi\left(x_n; \frac{\tilde{\sigma}_n^2 \hat{\mu}_{p_n} + t_n \tilde{\mu}_n}{\tilde{\sigma}_n^2 + t_n}, \sqrt{\frac{t_n \tilde{\sigma}_n^2}{t_n + \tilde{\sigma}_n^2}}\right) \quad (9)$$

from which the values of $\hat{x}_n = \hat{\mu}_n$ and $\hat{\sigma}_n$ can be read off.

We note that Maddison [[Maddison, 1991](#)] has described a similar two-pass algorithm based on the quadratic function describing the sum of squared deviations in trait value on the phylogeny, a method yielding identical local and global ancestral state estimates resulting from somewhat more complex formulae. The distributional approach described above has the benefit of directly yielding confidence intervals based on $\hat{\sigma}_n$ for each node n .

2 Results

2.1 Independent contrasts can be expressed in terms of the local maximum likelihood ancestral state reconstruction

Consider Equation (3). When calculating independent contrasts by traversing from the tips to the root of a phylogeny, each branch is extended by a factor of

$$\frac{t'_i t'_j}{t'_i + t'_j} \quad (10)$$

a term which is strikingly similar in form to the variance of the local maximum likelihood function at node n defined in Equation (7) as $\tilde{\sigma}_n^2$:

$$\frac{(\tilde{\sigma}_i^2 + t_i)(\tilde{\sigma}_j^2 + t_j)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)} \quad (11)$$

Indeed if we grant that each tip t of the phylogeny has a fixed point estimate of x_t , with $\tilde{\sigma}_t = 0$, then for any node whose children are tips we have

$$\frac{t'_i t'_j}{t'_i + t'_j} = \frac{(\tilde{\sigma}_i^2 + t_i)(\tilde{\sigma}_j^2 + t_j)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)} \quad (12)$$

and this identity will hold for all branches since the left side of the equation, like the right side, is additive down the the phylogeny and stored as a constant factor in the transformed branch lengths (under independent contrasts) or in the variance of the local maximum likelihood function (under maximum likelihood reconstruction), resulting in the equality:

$$t'_n = t_n + \tilde{\sigma}_n^2 \quad (13)$$

Note that the formula for x'_n given in Equation (2) simplifies to:

$$x'_n = \frac{x'_i t'_j + x'_j t'_i}{t'_i + t'_j} \quad (14)$$

which is strikingly similar in form to the mean of the local maximum likelihood function at node n defined in Equation (7):

$$\tilde{\mu}_n = \frac{\tilde{\mu}_i(\tilde{\sigma}_j^2 + t_j) + \tilde{\mu}_j(\tilde{\sigma}_i^2 + t_i)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)} \quad (15)$$

By substituting according to equation (13) we obtain:

$$\tilde{\mu}_n = \frac{\tilde{\mu}_i t'_j + \tilde{\mu}_j t'_i}{t'_i + t'_j} \quad (16)$$

Again, since $\tilde{\mu}_t = x'_t$ for any tip with fixed trait value, by induction on Equation (14) it follows that

$$x'_n = \tilde{\mu}_n \quad (17)$$

for general n .

Equations (13) and (17) provide a fundamental connection between maxi-

imum likelihood ancestral state reconstruction and independent contrasts, permitting us to represent an independent contrast at node n in terms of the local maximum likelihood function. By substituting into (1) we obtain:

$$IC_n = \frac{(\tilde{\mu}_i - \tilde{\mu}_j)}{\sqrt{t_i + \tilde{\sigma}_i^2 + t_j + \tilde{\sigma}_j^2}} \quad (18)$$

2.2 The sum of squared independent contrasts over a phylogeny is identical to the sum of squared deviations over a phylogeny imputed by the global maximum likelihood ancestral state reconstruction

Since the local likelihood function at node n is a normal distribution with mean $\tilde{\mu}_n$, the local maximum likelihood ancestral state estimate is $\tilde{x}_n = \tilde{\mu}_n$, a quantity which generates the following sum of squared deviations in the evolving trait on branches leading from n to its children:

$$\tilde{S}S_n = \frac{(\tilde{\mu}_i - \tilde{\mu}_n)^2}{\tilde{\sigma}_i^2 + t_i} + \frac{(\tilde{\mu}_j - \tilde{\mu}_n)^2}{\tilde{\sigma}_j^2 + t_j} \quad (19)$$

By substituting (15) we obtain:

$$\tilde{S}S_n = \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2}{t_i + \tilde{\sigma}_i^2 + t_j + \tilde{\sigma}_j^2} \quad (20)$$

$$= IC_n^2 \quad (21)$$

Since, at the root, the local likelihood and sum of squares is equal to the global likelihood and sum of squares, and given that the sum of squared deviations derived from the pair of branches descending from any node is identical to the squared independent contrast, we obtain:

$$\sum_n \frac{(\tilde{\mu}_n - \tilde{\mu}_{p_n})^2}{\tilde{\sigma}_n^2 + t_n} = \sum_n \frac{(\hat{\mu}_n - \hat{\mu}_{p_n})^2}{t_n} = \sum_n IC_n^2 \quad (22)$$

2.3 Regression estimators derived from maximum likelihood ancestral state reconstruction and from independent contrasts yield identical estimates of slope and correlation coefficient for bivariate traits evolving under Brownian motion

Reduced major axis and ordinary least squares estimators for slope and correlation coefficient depend solely on the variance and covariance of the variables subject to regression analysis. The variance of trait \mathbf{X} given a maximum likelihood ancestral state reconstruction $\hat{\mathbf{X}}$ is given by

$$\sigma_{\hat{\mathbf{X}}}^2 = \frac{1}{2t-2} \sum_n \frac{(\hat{\mu}_{\mathbf{X}_n} - \hat{\mu}_{\mathbf{X}_{p_n}})^2}{t_n} \quad (23)$$

where the phylogeny has t tips and $2t-2$ is the number of edges. The variance based on independent contrasts is:

$$\sigma_{\text{IC}_{\mathbf{X}}}^2 = \frac{1}{t-1} \sum_n \text{IC}_{\mathbf{X}_n}^2 \quad (24)$$

because two branches are consumed by each independent contrast. For this reason $\sigma_{\text{IC}_{\mathbf{X}}}^2$ is exactly twice as large as $\sigma_{\hat{\mathbf{X}}}^2$. It can be shown using Equation (18) and the same line of reasoning that the covariance $\sigma_{\text{IC}_{\mathbf{X}\mathbf{Y}}}^2$ is also exactly twice as large as $\sigma_{\hat{\mathbf{X}}\hat{\mathbf{Y}}}^2$.

Due to cancelling out of the denominators in the variance terms, the correlation coefficient based on independent contrasts, ρ_{IC} , is identical to that based on maximum likelihood ancestral state reconstruction, ρ_{ML} :

$$\rho_{\text{IC}} = \frac{\sigma_{\text{IC}_{\mathbf{X}\mathbf{Y}}}^2}{\sigma_{\text{IC}_{\mathbf{X}}}\sigma_{\text{IC}_{\mathbf{Y}}}} = \frac{2\sigma_{\hat{\mathbf{X}}\hat{\mathbf{Y}}}^2}{\sqrt{2}\sigma_{\hat{\mathbf{X}}}\sqrt{2}\sigma_{\hat{\mathbf{Y}}}} = \rho_{\text{ML}} \quad (25)$$

Similarly, the ordinary least squares regression slope estimators based on independent contrasts and maximum likelihood ancestral state reconstruction are

also identical:

$$\beta_{\text{IC}} = \frac{\sigma_{\text{IC}_{\mathbf{x}\mathbf{y}}}}{\sigma_{\text{IC}_{\mathbf{x}}}} = \frac{\sqrt{2}\sigma_{\hat{\mathbf{x}}\hat{\mathbf{y}}}}{\sqrt{2}\sigma_{\hat{\mathbf{x}}}} = \beta_{\text{ML}} \quad (26)$$

as are the reduced major axis regression slope estimators:

$$\beta_{\text{IC}} = \text{sign}(\rho_{\text{IC}}) \frac{\sigma_{\text{IC}_{\mathbf{y}}}}{\sigma_{\text{IC}_{\mathbf{x}}}} = \text{sign}(\rho_{\text{ML}}) \frac{\sqrt{2}\sigma_{\hat{\mathbf{y}}}}{\sqrt{2}\sigma_{\hat{\mathbf{x}}}} = \beta_{\text{ML}} \quad (27)$$

2.4 Independent contrasts for Brownian motion with a directional tendency

The standard Brownian motion model of continuous character evolution has zero mean such that the expected value of a trait after a period of evolution of duration t is equal to the value of the trait prior to the period of evolution. A directional tendency to the evolutionary process can be modelled in terms of a nonzero mean M , such that the expected value of a trait after a period of evolution of duration t is equal to tM . By modifying Equation (7) appropriately it is trivial to incorporate the directional tendency into the model described previously. In a traversal from the tips of the tree to the root we define the local likelihood function for each internal node:

$$\tilde{\mu}_n = \frac{(\tilde{\mu}_i - t_i M)(\tilde{\sigma}_j^2 + t_j) + (\tilde{\mu}_j - t_j M)(\tilde{\sigma}_i^2 + t_i)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)} \quad (28)$$

$$\tilde{\sigma}_n = \sqrt{\frac{(\tilde{\sigma}_i^2 + t_i)(\tilde{\sigma}_j^2 + t_j)}{(\tilde{\sigma}_i^2 + t_i) + (\tilde{\sigma}_j^2 + t_j)}} \quad (29)$$

and calculate a phylogenetically independent contrast incorporating directional tendency:

$$IC_n^2 = \frac{(\tilde{\mu}_i - \tilde{\mu}_n - t_i M)^2}{\tilde{\sigma}_i^2 + t_i} + \frac{(\tilde{\mu}_j - \tilde{\mu}_n - t_j M)^2}{\tilde{\sigma}_j^2 + t_j} \quad (30)$$

Substituting (28) into (30) we obtain:

$$IC_n = \frac{(\tilde{\mu}_i - \tilde{\mu}_j) - M(t_i - t_j)}{\sqrt{t_i + \tilde{\sigma}_i^2 + t_j + \tilde{\sigma}_j^2}} \quad (31)$$

For the sake of completeness, we here also define the maximum likelihood function for each node, which can be calculated in a second traversal from the root of the phylogeny to its tips in order to obtain maximum likelihood ancestral states under a directional tendency:

$$\hat{\mu}_n = \frac{\tilde{\sigma}_n^2(\hat{\mu}_{p_n} + t_n M) + t_n \tilde{\mu}_n}{\tilde{\sigma}_n^2 + t_n} \quad (32)$$

$$\hat{\sigma}_n = \sqrt{\frac{t_n \tilde{\sigma}_n^2}{t_n + \tilde{\sigma}_n^2}} \quad (33)$$

Given some value of M it is thus possible to calculate a set of phylogenetically independent contrasts for a trait evolving with directional tendency. The value of M is typically not known, but is easily estimated from the data by conducting a linear search to identify the \hat{M} which minimizes the sum of squared contrasts or maximizes the global likelihood of the model.

References

- [Bromiley, P. \(2003\). Products and convolutions of Gaussian distributions. Technical Report Tina Memo No. 2003-003, Medical School, University of Manchester, UK.](#)
- [Felsenstein, J. \(1985\). Phylogenies and the comparative method. *Am. Nat.*, 125\(1\):1–15.](#)
- [Maddison, W. \(1991\). Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Sys. Biol.*, 40\(3\):304–314.](#)