**Comparative methods with sampling error and within-species variation: contrasts revisited and revised**

Joseph Felsenstein[1]

Department of Genome Sciences and Department of Biology, University of Washington, Box 355065, Seattle, Washington 98195-5065

Running Head: Comparative methods with sampling error

*Keywords:* comparative method, contrasts, statistics, phylogeny sampling error, phenotypic variance

[1]E-mail: joe (at) gs.washington.edu

**Elements of the paper:**

Top of Article

Why there is a problem

A correction for within-species variation and measurement error

A model with within-species variation

The model

The contrasts

An example of the contrasts

Estimating the covariances

Tests of covariation

Simulations

Genetics Experiments

A program

Acknowledgments

Literature cited

APPENDIX A

APPENDIX B

Table caption

Table

Figure captions

Abstract. – Comparative methods analyses have usually assumed that the species phenotypes are the true means for those species. In most analyses the actual values used are means of samples of modest size. The covariances of contrasts then involve both the covariance of evolutionary changes and a fraction of the within-species phenotypic covariance, the fraction depending on the sample size for that species. Ives, Midford, and Garland (2007) have shown how to analyze data in this case when the within-species phenotypic covariances are known. The present model allows them to be unknown and to be estimated from the data. A multivariate normal statistical model is used for multiple characters in samples of finite size from species related by a known phylogeny, under the usual Brownian motion model of change and with equal within-species phenotypic covariances. Contrasts in each character can be obtained both between individuals within a species and between species. Each contrast can be taken for all of the characters. These sets of contrasts, each the same contrast taken for different characters, are independent. The within-set covariances are unequal, and depend on the unknown true covariance matrices. An EM algorithm is derived for making an REML estimate of the covariances of evolutionary change and the within-species phenotypic covariances. It is available in the Contrast program of the PHYLIP package. Computer simulations show that the covariances are biased when the finiteness of sample size is not taken into account, and that using the present model corrects the bias. Sampling variation reduces the power of inference of covariation in evolution of different characters. An extension of

this method to incorporate estimates of additive genetic covariances from a simple genetic experiment is also discussed.

Since its introduction (Felsenstein, 1985) the method of contrasts has been widely used to correct comparative methods for covariation of observations due to the phylogeny of the species. It assumes that character change follows a model of correlated Brownian motion, and that the phenotypic means of the characters are observed in each species. But that is a strong assumption: invariably the mean of a species is measured in a finite sample, sometimes quite a small sample. Ricklefs and Starck (1996) have shown an example where the largest contrasts tended to be those from the most closely related species, and they suggested that this was most likely an artifact of sampling error caused by within-species variation.

## Why there is a problem

To see why this artifact occurs, consider a single character in two sister species that are separated by branches whose total length is such that we expect the variance of the differences between species to be $v$. Now suppose that the within-species variance of the character is $\sigma^2$. This variance is either measurement error or simply within-species phenotypic variation, arising from both genetic and environmental causes. If we have sample sizes of $n_1$ and $n_2$ individuals from the two species, the variance of the difference between these two species is

$$v \; + \; \frac{1}{n_1}\,\sigma^2 \; + \; \frac{1}{n_2}\,\sigma^2 \tag{1}$$

If we ignored the within-species variation, we would assume that this contrast had variance $v$. Contrasts are standardized by their standard deviations (more generally,

by a quantity proportional to their standard deviations). So we would standardize by dividing the contrast by $\sqrt{v}$. Thus the effect of the within-species variation is to inflate the variance of the difference by multiplying it by a factor of

$$1 \; + \; \frac{1}{n_1} \, \frac{\sigma^2}{v} \; + \; \frac{1}{n_2} \, \frac{\sigma^2}{v} \tag{2}$$

If the sample sizes $n_1$ and $n_2$ are both large, this may be a minor effect, but often sample sizes may be small, with species in some case being represented by only a single specimen.

When the two species are not closely related, $v$ will be large and the inflation may still be small, but when they are closely related, the small value of $v$ means that the contrast may have a much larger variance than expected. The contrast will have been standardized by being divided by far too small a quantity. This is presumably the reason that Ricklefs and Starck (1996) found that the outliers in their statistical analysis tended to be the contrasts associated with close relatives. It should be evident that this incorrect standardization can result in improper weighting of evidence from different parts of the tree.

### A correction for within-species variation and measurement error

The problem posed by within-species variation and measurement error has been discussed by a number of people. Lynch (1991) gave a model that could be used for within-species variation. It assumed that the species means were subject to an additional, nonphylogenetic variation whose covariance matrix could be estimated.

The primary motivation for that model was a nonphylogenetic burst of adaptation specific to each species, making more explicit the model implicit in the work of Cheverud et al. (1985) It is not intended to model variation from individual to individual within species, but Lynch's model would work for within-species variation if the sample size for each species was one individual. Lynch mentioned the possibility of using estimates of the sampling error of the species means to correct his inferences. Housworth et al. (2004) have further discussed the possibility of extending Lynch's model to take account of within-species variation. Christman et al. (1997) used an extension of Lynch's model to cope with sampling variation, implementing it in MATLAB. They did not give their detailed estimation equations. Harmon and Losos (2005) have used computer simulation to check the effect of finite intraspecific sample size on comparative studies. They found that it could cause inflation of type I error rates, and they suggested using an ANOVA in advance to check whether sample size was too small.

Ives, Midford, and Garland (2007) have addressed the problem and have shown how to correct for the effects of within-species variation and measurement error. Their method allows for each character to have variation within species and allows for the measurements of different characters within species to be correlated. They show how to compute ML, REML, and GLS (Generalized Least Squares) estimates of the covariances of evolutionary change, allowing for the phenotypic covariances within species as well. Their models are essentially identical to those given in this

paper.

The present model extends their methods in one major way. They have assumed that the phenotypic covariances within species are known, rather than to be estimated from the data set. They argue that this is not a serious restriction in practice with real data. The method given here is more general than theirs in that it estimates both the within-species phenotypic covariances and the between-species phylogenetic covariances from the same data set, thus allowing for our uncertainty about the within-species covariances. The computational scheme presented here also shows how the method of contrasts can be extended within and between species in a self-consistent way. The EM algorithm developed here for inference of covariances can readily be adapted to testing hypotheses about constraints on the covariances. We will see in the discussion below that the present framework can also be extended to accomodate quantitative genetics experimental designs, creating the possibility of detecting whether the evolutionary change between species reflects natural selection rather than genetic drift.

## A model with within-species variation

In this paper I will modify Lynch's model to allow for sample sizes greater than 1 and within-species phenotypic variance, and I will provide a computational strategy. As we shall see, although contrasts can be used within species and between species, the resulting contrasts cannot be scaled so as to all have the same variances, at least not without knowing the parameters in advance. Alternatively, it would be

possible to use the covariances of the original observations, without taking contrasts, as Grafen (1989) did with his "phylogenetic regression" as an alternative to the standard contrasts method.

The formulas presented here can be used either way. The present computational scheme is an REML (reduced maximum likelihood) rather than an ML (maximum likelihood) method. Lynch's method was an iterative weighted least squares method which converges to a direct maximum likelihood analysis of the full multivariate normal model. My method transforms the data by taking between-species and within-species contrasts, and doing this once for each character. For each character we omit the grand mean of all observations. It is thus a REML analysis. As we shall see, it can be extended to take into account of within-species genetic experiments of a simple type (such as half-sib designs). The method is thus able to correct for estimates of genetic covariation, treating within and between-species analyses in a unified way.

This paper presents the computational strategy, some computer simulation experience with the behavior of the estimates, and will discuss extension of the method to allow for estimates of genetic covariation.

## The model

The model used here will have a component for phenotypic variation from individual to individual within the species. Such variation can arise from sampling or from measurement error. The relationship of the present paper to previous

treatments of Lynch's model is somewhat circular: Housworth et al. (2004) mention that Lynch's Phylogenetic Mixed Model can be modified to take account of within-species variation as well, and they cite a prepublication communication of the present results as an alternative to their suggestions.

I will explain the model in terminology somewhat different from Lynch's; he used terms (such as "additive value") from the analogous quantitative genetics context from which his computational methods were derived. This has the disadvantage that users may not realize that the additive genetic variation of traits within species is not necessarily the cause of the "additive" variations between species. Lynch warned against misinterpretation of this terminology. To help avoid the problem I will instead call the between-species covariances "phylogenetic" and the within-species covariances "phenotypic".

Let us assume a Brownian motion model of evolutionary change with correlated change of $p$ characters along a phylogeny of $s$ species. Within each species the characters will also covary across individuals in a sample from a multivariate normal distribution, with the covariances of characters assumed to be the same within all species. This assumption ought to be controversial, as natural selection and genetic drift can alter genetic covariances as gene frequencies change between populations; it becomes increasingly suspect as we treat widely-diverged species. Nevertheless, we need the assumption to carry through our analysis. Similar assumptions are made by Lynch (1991) and by Ives, Midford, and Garland (2007).

The within- and between-species covariances are allowed to be completely different. I have explained elsewhere (Felsenstein, 1988, 2002a, 2004) why we should not simply assume that these two sets of covariances are proportional to each other. We will assume that a total of $n$ individuals have been measured, with $n_i$ of them coming from species $i$. As is usual in Brownian motion models, the expected change of each character in a lineage is taken to be zero, so that the expectation of the character in each species is given by the overall expectation. The joint distribution of all characters in all individuals is multivariate normal. The distribution is therefore determined by its expectations and covariances; we need to estimate them.

In a lineage the covariance of evolutionary change between characters $k$ and $\ell$ per unit branch length will be denoted $a_{k\ell}$. The covariance of the total changes of these two characters along a lineage will be obtained by multiplying this by the length of the branch. If the branches of the phylogeny from the root up to the most recent common ancestor of the species for individual $i$ and the species for individual $j$ are of total length $t_{ij}$, then the covariance of character $k$ in individual $i$ and character $\ell$ in individual $j$ due to shared evolutionary change will be (Felsenstein, 2004)

$$\mathrm{Cov}\left[\, x_{ik}, x_{j\ell} \,\right] = t_{ij}\, a_{k\ell}. \tag{3}$$

We stack the rows of the array $x_{ij}$, so that we make a vector $\mathbf{y}$ with the elements

$$\mathbf{y}^T = (x_{11},\, x_{12},\, \ldots,\, x_{1p},\, x_{21},\, x_{22},\, \ldots,\, x_{2p},\;\; \ldots,\;\; x_{n1},\, x_{n2},\, \ldots,\, x_{np})\,, \tag{4}$$

with all of the phenotypes for individual 1 followed by all those for individual 2, and so on. We will assume that the individuals are themselves arranged so that members

of the same species are adjacent. Then we can write the covariance matrix of the elements of $\mathbf{y}$ compactly as a Kronecker product of matrix $\mathbf{T}$ and matrix $\mathbf{A}$:

$$\mathrm{Cov}[\,\mathbf{y}\,] \;=\; \mathbb{E}[\,\mathbf{y}\mathbf{y}^T\,] \;=\; \mathbf{T}\otimes\mathbf{A} \tag{5}$$

The Kronecker product of these two matrices is an $np\times np$ matrix which arranges the products $t_{ij}a_{k\ell}$ in the appropriate order to be the covariance matrix of the stacked vector $\mathbf{y}$.

The variation within species is incorporated by adding to the covariance an additional term for the within-species phenotypic covariance of character $k$ in individual $i$ with character $\ell$ in that same individual. This is done using the "Kronecker delta" $\delta_{ij}$ which is simply the bookkeeping device which is 1 when $i=j$ and 0 otherwise (it has no subtle connection to the Kronecker product):

$$\mathrm{Cov}\,[\,x_{ik},x_{j\ell}\,] \;=\; t_{ij}\,a_{k\ell} \;+\; \delta_{ij}\,p_{k\ell} \tag{6}$$

This adds a within-species covariance $p_{k\ell}$ between characters of the same individual. The result can be expressed in matrix terms by

$$\mathrm{Cov}[\,\mathbf{y}\,] \;=\; \mathbb{E}[\,\mathbf{y}\mathbf{y}^T\,] \;=\; \mathbf{T}\otimes\mathbf{A} \;+\; \mathbf{I}\otimes\mathbf{P} \tag{7}$$

where $\mathbf{I}$ is the $n\times n$ identity matrix.

Note that the matrix $\mathbf{T}$ expresses the phylogeny of the individuals. Individuals in a sample from the same species are considered to be separated from each other on it by branches of zero length (see Figure 1). This implicitly assumes that we have sampled individuals independently from a single population. If we have samples from

multiple populations for a species, the situation is more complex. I have elsewhere outlined how one would construct a within-species comparative method (Felsenstein, 2002b). It is not easily integrated into the present framework, because it assumes that phenotypes are selected towards a local optimum in each population rather than undergoing an unconstrained Brownian motion.

(Insert Figure 1 about here)

We will assume that the tree $\mathbf{T}$ is known, and ask how the covariances $\mathbf{A}$ of phylogenetic change can be estimated, as well as the within-species phenotypic covariances $\mathbf{P}$. Lynch (1990) gave a method based on the "mixed model" in wide use in animal breeding. The method presented here will instead be based on having contrasts among all individuals, calculated so as to remove all covariances due to phylogeny. Each of these contrasts is computed for all characters. Unlike the original contrasts method, the method does not scale the contrasts so that they have equal variance, but instead standardizes the contrast coefficients to be an orthonormal transformation. We will see that we can then use an EM algorithm to estimate the covariances.

I will assume that all characters are available for all individuals. It would be possible to develop a likelihood inference method for cases where there are missing values; since even inference of simple covariance matrices becomes difficult with missing data, this complication will not be covered here.

Again, I emphasize that the use of contrasts is a convenience, that the whole analysis could be done with the original phenotype values, and if a REML analysis were used for that, the results would be equivalent.

## The contrasts

The contrasts are obtained from the tree $T$ in the usual way (Felsenstein, 1985), except that they are not normalized to have variance 1, but rather to have the sum of squares of their coefficients add to 1. Figure 1 will serve as an example for the calculation of the contrasts.

### *Within-species contrasts*

Contrasts for each character are first formed within each species that has multiple individuals sampled. For example, species B has 4 individuals. A simple way to form the orthonormal contrasts for each character within this species is to first take the difference for the character between individuals 1 and 2, then the difference of individual 3 from the mean of those two, then the difference of individual 4 from the mean of the first three. These contrasts are then multiplied by appropriate constants to make the sum of squares of their coefficients add to 1. Thus the contrasts will

end up as

$$\begin{aligned} c_1 &= f_1 \left( x_{B1} - x_{B2} \right) \\ \\ c_2 &= f_2 \left( x_{B3} - \left( \tfrac{1}{2} x_{B1} + \tfrac{1}{2} x_{B2} \right) \right) \\ \\ c_3 &= f_3 \left( x_{B4} - \left( \tfrac{1}{3} x_{B1} + \tfrac{1}{3} x_{B2} + \tfrac{1}{3} x_{B3} \right) \right) \end{aligned} \tag{8}$$

The normalization constants $f_1$, $f_2$, and $f_3$ must be chosen so that the sum of squares of the coefficients in each contrast is 1:

$$\begin{aligned} f_1^2 + f_1^2 &= 1 \\ \\ f_2^2 + \tfrac{1}{4} f_2^2 + \tfrac{1}{4} f_2^2 &= 1 \\ \\ f_3^2 + \tfrac{1}{9} f_3^2 + \tfrac{1}{9} f_3^2 + \tfrac{1}{9} f_3^2 &= 1 \end{aligned} \tag{9}$$

which yields immediately

$$\begin{aligned} f_1 &= \sqrt{\tfrac{1}{2}} \\ \\ f_2 &= \sqrt{\tfrac{2}{3}} \\ \\ f_3 &= \sqrt{\tfrac{3}{4}} \end{aligned} \tag{10}$$

and, more generally

$$f_m = \sqrt{\frac{m}{m+1}} \tag{11}$$

This is a convenient set of within-species contrasts. For each contrast, one computes it for each character. The calculation of the contrasts in effect makes an orthonormal

transformation of the original character values. The $i$th such contrast is independent of the $j$th such contrast, but the $i$th contrast for one character has a covariance with the $i$th contrast for another character, which is the same as the phenotypic covariance of those characters.

We will see a similar pattern in the contrasts between species. There too, the different contrasts are independent of each other, but the same contrast taken on two different characters has a covariance equal to the covariance of the original values of the characters.

There are many other ways that orthonormal contrasts could be taken within the species; for the calculations in this paper, all are equivalent. They correspond to different trees connecting the observations within species, each tree having all of its internal branches of length zero, and all external branches of equal length. All of these will leave us, for each character, with one linear combination of the individual values that is orthogonal to all of these within-species contrasts. That linear combination is, of course, the species mean of the character. The vector of species mean phenotypes has covariances composed of two components: the covariances of the true species mean phenotypes, plus a fraction of the within-species phenotypic covariances $\mathbf{P}$.

*Between-species contrasts*

Imagine taking the tree $\mathbf{T}$ and forming contrasts according to the usual contrasts method, with the exception that each individual is a tip, connected to the node

for its species by a branch of length 0. In the present case, each contrast is not scaled so that its variance is 1, but so that the sum of squares of the coefficients of the individual measurements is 1. There is a simple algorithm for doing this, an extension of the usual contrasts algorithm. This algorithm forms a set of contrasts among the individual measurements. We compute each contrast for each character.

The contrasts thus come in sets. Each set is the result of computing the same contrast for all of the characters. The contrasts in the different sets are independent. We will use these sets of contrasts to estimate the phylogenetic and phenotypic covariances in a way that takes into account the effect of the within-species phenotypic covariances on their distribution.

We start with the contrasts between individuals within species. Using the contrasts outlined above, there will be $n - s$ of these within-species contrasts in all. Each has expectations 0 in all characters and covariance matrix $\mathbf{P}$ between characters. We are then left with the means of each character in each species.

### *A recursive algorithm*

There is a convenient algorithm for computing the between-species contrasts. To carry out the algorithm for a single character, imagine computing three quantities at each node of the tree. We start, not at the tips where the individual measurements are, but at the nodes for each species (in the case of a species with a sample size of 1, we do start at that individual). One of the three quantities is the mean phenotype of the character at the node ($x$), another ($\delta v$) is proportional to the variance of that

mean caused by the phylogenetic component of variance. Thus at the start it is zero, as no phylogenetic component contributes to the differences between measurements within a species. The third quantity ($s$) is the sum of squares of the coefficients used to compute the mean at that node from the values in the $n_i$ individuals. The necessary contrasts can be computed using these three quantities.

At the start, at the node for a species, the mean phenotype $x$ is the mean of all individuals for that species. As the mean was computed as the mean of the values from the $n_i$ individuals in the species, the sum of squares of the coefficients of the individual values in that mean is $s_i = n_i(1/n_i^2) = 1/n_i$. The mean at any interior node of the tree will be a linear combination of the means of the species above it, and hence also a linear combination of the phenotypes of the individuals of that species, and we will compute the sum of squares of the coefficients for those individual measurements.

For interior nodes of the tree these three quantities are updated down the tree recursively, the values at each node being computed from the values in its immediate descendants. I will show how this can be done for a bifurcating tree. Multifurcations can be dealt with in an entirely analogous fashion, most easily by inserting fictional branches of zero length into the tree and treating each multifurcation as a series of bifurcations (Felsenstein, 1985, p. 10; Purvis and Garland, 1993).

Suppose that we are computing the values at a node, whose immediate descendants are nodes $L$ (left) and $R$ (right). The branch lengths leading from

the node up to nodes $L$ and $R$ are $v_L$ and $v_R$. The equations for the values of the three quantities at the node are the weighted average of the two descendant values;

$$x \;=\; \frac{\frac{1}{v_L+\delta v_L}\,x_L \;+\; \frac{1}{v_R+\delta v_R}\,x_R}{\frac{1}{v_L+\delta v_L} \;+\; \frac{1}{v_R+\delta v_R}}, \tag{12}$$

the variance of that weighted average is:

$$\delta v \;=\; 1 \Big/ \left( \frac{1}{v_L \,+\, \delta v_L} \;+\; \frac{1}{v_R \,+\, \delta v_R} \right), \tag{13}$$

the sum of squares of the coefficients of the individual measurements above nodes $L$ and $R$ in the weighted average is then

$$s \;=\; f_L^2\, s_L \;+\; f_R^2\, s_R, \tag{14}$$

where the coefficients $f_L$ and $f_R$ are:

$$f_L \;=\; \frac{\frac{1}{v_L+\delta v_L}}{\frac{1}{v_L+\delta v_L} + \frac{1}{v_R+\delta v_R}} \tag{15}$$

and

$$f_R = \frac{\frac{1}{v_R+\delta v_R}}{\frac{1}{v_L+\delta v_L} \;+\; \frac{1}{v_R+\delta v_R}}. \tag{16}$$

The equations (12) and (13) are the same as in the usual contrasts method. The third equation (14) computes for each node the sum of squares of the coefficients of individual measurements in the weighted average at that node. This will be needed when the orthonormal contrasts are computed.

As we compute the value for a character at each interior node, we also compute a contrast for that character at that node. The set of contrasts we compute will be

orthonormal, which implies that they will be independent and each will have the sum of squares of its coefficients be 1. The contrast at this node will be of the form

$$c \;=\; K(x_L \;-\; x_R) \tag{17}$$

for some value of $K$. The value of $K$ must be chosen so that the sum of squares of the coefficients of the contrasts is 1. this implies that

$$K^2\, s_L \;+\; K^2\, s_R \;=\; 1, \tag{18}$$

so that

$$K \;=\; \sqrt{\frac{1}{s_L \;+\; s_R}} \tag{19}$$

The quantities $w_{ii}$, which are the variances of the values of $K(x_R - x_L)$, are from (14) and (19)

$$w_{ii} \;=\; K^2\,(v_L + \delta v_L + v_R + \delta v_R) \;=\; \frac{v_L + \delta v_L + v_R + \delta v_R}{s_L + s_R} \tag{20}$$

As we go down the tree, successively considering nodes, we update these three quantities using equations 12, 13, and 14 and compute the contrast values using equations 19 and 17. We use equation (17) separately for each character. For $n$ species this will compute $n - 1$ contrasts for each character. It is rather easy to see that the coefficients and sums of squares of a contrast will be the same for all characters, so that although we need to compute the weighted averages and the contrast values for each character, we do not need to recompute $\delta v$, the $f_i$, $K$ and $s$ for each character. As we carry out this algorithm, we are in effect defining the

entries in a matrix of coefficients of individual measurements in the contrasts, $\mathbf{C}$. The matrix contains one row for each contrast, including all of the within-species contrasts. The $n-1$ contrasts being orthonormal, it follows that

$$\mathbf{C}\,\mathbf{C}^T \;=\; \mathbf{I}_{n-1} \tag{21}$$

where $\mathbf{I}_{n-1}$ is the $(n-1)\times(n-1)$ identity matrix.

The matrix $\mathbf{C}$ has in effect been applied to all characters, so the transformation of the vector $\mathbf{y}$ makes it into the $(n-1)p$ contrasts

$$\mathbf{z} \;=\; (\mathbf{C}\otimes\mathbf{I})\;\mathbf{y}. \tag{22}$$

## An example of the contrasts

The tree in Figure 1 can be used to show how this works. Let us assume that the branch lengths are as given. The first between-species contrast for a character is between species A and B. The sample sizes for the species A, B, C, D, and E are respectively 3, 4, 4, 4, and 2. The sum of squares of coefficients when we finish doing the contrasts within a species with $n_i$ samples is $s = 1/n_i$. The contrast between species means of A and B has, from (19),

$$K \;=\; \sqrt{\frac{1}{\frac{1}{3}+\frac{1}{4}}} \;=\; 1.309307 \tag{23}$$

so that the contrast is

$$1.309307\,\bar{x}_A \;-\; 1.309307\,\bar{x}_B. \tag{24}$$

The weighted average for the ancestor of A and B is, from (12)

$$\bar{x}_{AB} \;=\; \frac{\frac{1}{1.2}\,\bar{x}_A + \frac{1}{0.8}\,\bar{x}_B}{\frac{1}{1.2}+\frac{1}{0.8}} \;=\; 0.4\,\bar{x}_A + 0.6\,\bar{x}_B \tag{25}$$

and the extra branch length added above that ancestor is, from (13)

$$\delta v_{AB} \;=\; \frac{1}{\frac{1}{1.2}+\frac{1}{0.8}} \;=\; 0.48. \tag{26}$$

The sum of squares of coefficients at the ancestor of A and B is, from (14),

$$s_{AB} \;=\; 0.4^2\,s_A \;+\; 0.6^2\,s_B \;=\; 0.143333. \tag{27}$$

The next step is to do the same calculations for the adjacent tips E and C. This yields again

$$K \;=\; \sqrt{\frac{1}{\frac{1}{2}+\frac{1}{4}}} \;=\; 1.154701, \tag{28}$$

the contrast is then

$$1.154701\,\bar{x}_E \;-\; 1.154701\,\bar{x}_C. \tag{29}$$

The weighted average at their ancestor is:

$$\bar{x}_{EC} \;=\; \frac{\frac{1}{1.1}\,\bar{x}_E + \frac{1}{0.7}\,\bar{x}_C}{\frac{1}{1.1}+\frac{1}{0.7}} \;=\; 0.388888\,\bar{x}_E + 0.611111\,\bar{x}_C \tag{30}$$

and the extra branch length added above that ancestor is:

$$\delta v_{EC} \;=\; \frac{1}{\frac{1}{1.1}+\frac{1}{0.7}} \;=\; 0.427777. \tag{31}$$

The sum of squares of the coefficients is, from (19)

$$s_{EC} \;=\; (0.388888)^2\,s_E + (0.611111)^2\,s_C \;=\; 0.1689815. \tag{32}$$

The next stage involves the contrast between D and the ancestor of E and C. The contrast coefficient $K$ is

$$K \;=\; \frac{1}{\sqrt{\frac{1}{4} + 0.1689815}} \;=\; 1.544908 \tag{33}$$

so that the contrast is

$$1.544908\,\bar{x}_D \;-\; 1.544908\,\bar{x}_{EC}. \tag{34}$$

The weighted average for the character is

$$\begin{aligned} \bar{x}_{DEC} \;&=\; \frac{\frac{1}{0.7}}{\frac{1}{0.7} + \frac{1}{0.9 + 0.427777}}\,\bar{x}_D + \frac{\frac{1}{0.9 + 0.427777}}{\frac{1}{0.7} + \frac{1}{0.9 + 0.427777}}\,\bar{x}_{EC} \\ &=\; 0.6547945\,\bar{x}_D \;+\; 0.3452055\,\bar{x}_{EC} \end{aligned} \tag{35}$$

The sum of squares of coefficients is

$$s_{DEC} \;=\; (0.6547945)^2\,s_D \;+\; (0.3452055)^2\,s_{EC} \;=\; 0.1273260 \tag{36}$$

This gives us what we need to compute the last contrast, between AB and DEC. The coefficients are calculated as

$$K \;=\; \frac{1}{\sqrt{0.143333 + 0.1273260}} \;=\; 1.922157 \tag{37}$$

so that the contrast is

$$1.922157\,\bar{x}_{AB} \;-\; 1.922157\,\bar{x}_{DEC}. \tag{38}$$

The further calculation of values of $\bar{x}$, $s$, and $\delta v$ is unnecessary because we have already obtained all of the contrasts.

I have shown the calculation for a single character. In practice, as their coefficients are obtained, the contrasts can be taken separately for each character, using the same coefficients for each character.

*The covariances of the new variables*

The contrasts are $(n-1)p$ new variables, which of course have a multivariate normal distribution, as they are linear combinations of variables from a multivariate normal distribution. Each has expectation zero. The resulting variables are shown in Appendix I to consist of $n-1$ sets of $p$ variables, each set independent of all the others. The $p$ variables in the $i$th set have expectation zero and covariance

$$\mathrm{Cov}[\mathbf{z}^{(i)}] \;=\; w_{ii}\,\mathbf{A} \;+\; \mathbf{P} \tag{39}$$

where the $w_{ii}$ are the obtained in the process of computing the contrast coefficients.

We may compare this with sets of contrasts calculated in the original contrasts method, which have no within-species error term $\mathbf{P}$ and thus can be made to have a common variance by dividing each by the square root of $w_{ii}$.

**Estimating the covariances**

If we are given a tree, we can use the recursive algorithm above to compute the contrasts calculated in equation 22 to obtain $n-1$ independent contrasts for each character. We organize them into $n-1$ sets of variables, with variables in different sets being independent. The sets have unequal covariances, which depend on the as-yet-unknown phylogenetic covariances $\mathbf{A}$ and within-species phenotypic covariances

$\mathbf{P}$. Given the transformed variables $\mathbf{z}$, we could imagine estimating these covariance matrices by maximum likelihood. One algorithm that is easy to carry out, if not particularly fast, is the EM algorithm (Dempster et. al., 1977). This uses our current estimates of $\mathbf{A}$ and $\mathbf{P}$ to find the expectation of sufficient statistics for estimating those matrices. New covariance matrices then are estimated from these expectations. The EM algorithm can be proven to converge on the maximum likelihood estimates of the covariance matrices. In the present case the contrasts $\mathbf{z}$ omit the grand mean of each character, so that the ML estimation is actually REML (reduced maximum likelihood).

There are $n$ observations of each character in all, and these have been reduced to $n-1$ contrasts. Of these $n_1 + n_2 + \ldots + n_s$ are the within-species contrasts, all of which have $w_{ii} = 0$. With a total of $s$ species, the remaining $s-1$ contrasts have possibly different values of $w_{ii}$. If we look at the covariance matrix $\mathbf{W} \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{P}$, we find that it is block-diagonal with a total of $n-1$ blocks. If we break the vector of contrasts $\mathbf{z}$ into parts that correspond to the blocks, these correspond to the $n-1$ sets of contrasts. The block-diagonality reflects the independence of these sets of contrasts.

We call the $i$-th such set of contrasts $\mathbf{z}^{(i)}$. The elements of the vector $\mathbf{z}^{(i)}$ are $p$ adjacent values in the larger vector $\mathbf{z}$. In principle the vector $\mathbf{z}^{(i)}$ consists of two parts, one from the phylogenetic changes and one from the within-species phenotypic differences. We can write the random variable for the vector of characters in terms

of its unknown phylogenetic and phenotypic components

$$\mathbf{z}^{(i)} \;=\; \sqrt{w_{ii}}\,\mathbf{a} \;+\; \mathbf{p}. \tag{40}$$

In practice we cannot know exactly how much of $\mathbf{z}^{(i)}$ comes from each of these sources. If we did, we could estimate the covariance matrices $\mathbf{A}$ and $\mathbf{P}$ by averaging $\mathbf{a}\mathbf{a}^T$ and $\mathbf{p}\mathbf{p}^T$ over all the contrasts. The within-species contrasts have $w_{ii} = 0$; they provide estimates of only $\mathbf{P}$, but they do not contain all of the information about $\mathbf{P}$, since it also appears in the covariances of the between-species contrasts.

The EM algorithm uses our current estimates of $\mathbf{A}$ and $\mathbf{P}$ to compute the expectations of their sufficient statistics, which are sums over the sets of contrasts of $\mathbf{a}\mathbf{a}^T$ and of $\mathbf{p}\mathbf{p}^T$ given the $\mathbf{z}^{(i)}$. These are used in place of the unknown actual values to make new estimates of $\mathbf{A}$ and $\mathbf{P}$. Those are then used to compute new expectations, and so on. When the process converges, the results are the (reduced) maximum likelihood estimates. Appendix II derives the formulas for the expectations of $\mathbf{a}\mathbf{a}^T$ and $\mathbf{p}\mathbf{p}^T$ from the $\mathbf{z}^{(i)}$ and the formulas for the current estimates of the covariances, $\mathbf{A}(t)$ and $\mathbf{P}(t)$.

## Tests of covariation

The estimates of the phylogenetic and phenotypic covariance matrices $\mathbf{A}$ and $\mathbf{P}$ allow us to test hypotheses about the covariation between characters. If we wish to test whether the phylogenetic covariance is nonzero, we can maximize the likelihood under the assumption that all elements of $\mathbf{A}$ are zero, by carrying out the EM

algorithm while holding all of the $a_{ij}$ zero. Another hypothesis of interest would be that two sets of characters have no phylogenetic covariation. If these are (say) the first $q$ characters and the remaining $p-q$ characters, we would instead hold the $q \times (p-q)$ block of covariances between them, and also its transpose, zero in $\mathbf{A}$. These restricted hypotheses would be compared to the results of the full EM algorithm in which all elements of $\mathbf{A}$ and $\mathbf{P}$ are estimated.

If the null hypothesis has covariance estimates $\mathbf{A}_0$ and $\mathbf{P}_0$, and the alternative has $\mathbf{A}_1$ and $\mathbf{P}_1$, the likelihood ratio test uses the densities of the multivariate normal distribution of the contrasts to compute

$$\begin{aligned} 2\ln(L_1/L_0) &= \sum_i \ln\left(\frac{|w_{ii}\mathbf{A}_0+\mathbf{P}_0|}{|w_{ii}\mathbf{A}_1+\mathbf{P}_1|}\right) \\ &\quad - \sum_i \mathbf{z}^{(i)\,T}(w_{ii}\mathbf{A}_0+\mathbf{P}_0)^{-1}\mathbf{z}^{(i)} + \sum_i \mathbf{z}^{(i)\,T}(w_{ii}\mathbf{A}_1+\mathbf{P}_1)^{-1}\mathbf{z}^{(i)} \end{aligned} \tag{41}$$

This will have an approximate chi-square distribution with $q(p-q)$ degrees of freedom in the case where we test whether one set of characters has no phylogenetic covariation with the rest.

One might think that it would be useful to infer whether one phylogenetic covariance (say $v_{kl}$) was zero, but in view of the possibility that these characters could still covary through their mutual covariation with a third character, one would need to be cautious in interpreting this hypothesis. It would be possible to fit a full Gaussian graphical model for the characters by constraining the appropriate set of partial correlations to zero, which amounts to constraining some entries in the

inverse of the covariance matrix, to zero (Lauritzen, 1996, section 5.1.3).

Another test of interest is whether we can assume that the phylogenetic covariance $\mathbf{A}$ is proportional to the phenotypic covariance matrix $\mathbf{P}$. Appendix II includes iteration equations $(55) - (57)$ which can maintain proportionality between these two covariance matrices (and estimate the constant of proportionality). These two matrices can be used as $\mathbf{A}_0$ and $\mathbf{P}_0$ in the likelihood ratio test. In this test the degrees of freedom are $p(p+1)/2 - 1$ which is $(p+2)(p-1)/2$.

## Simulations

Table 1 and Figures 2 and 3 show the results of simulation tests of the method. Three sets of runs were done. 10 trees of 40 species were generated by simulating a birth process with birth rate 1, stopping just as the 41st species was about to be born. This is expected to generate a tree of average depth (from the tips to the bottommost fork)

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots + \frac{1}{40} \;=\; 3.278543 \tag{42}$$

This same set of 10 trees was used as the true trees for all three sets of simulation runs.

In the first set of runs two characters were simulated as evolving up the tree by Brownian motion, both with a variance of change 1.0 per unit time, and with no correlation of change. For each species four individuals were measured, with the within-species phenotypic variance being 1.0 in both characters and no within-

species correlation of the characters. For each of the 10 trees 1000 such data sets were simulated. These were each analyzed in two ways: by using the species means with the original contrasts method, and by taking within-species phenotypic variation and covariation into account using the methods discussed above.

(Insert Table 1 about here)

For each of these two analyses a likelihood ratio test was carried out of the (true) hypothesis that there was no phylogenetic correlation among the characters. Table 1 shows, for each of the 10 trees, how many times out of 1000 data sets the null hypothesis was rejected. It should be immediately apparent that when there is no correction for within-species variance, the null hypothesis is rejected far too often (an average of 20% of the time when the nominal level of rejection is set to 5%). This is consistent with the results of similar simulations carried out by Harmon and Losos (2005), which found the same effect. When the analysis allowed for within-species phenotypic variance, the null hypothesis is rejected an average of 5.67% of the time, significantly more often than 5%. One possible explanation for the discrepancy is that it is reflects the use of a likelihood ratio test that is only asymptotically valid.

(Insert Figure 2 about here)

In Figures 2 and 3 histograms of estimates of the phylogenetic correlation between two characters are shown. In each column, the top histogram is for analysis of species means by the original contrasts method, and the bottom histogram is for a

full within- and between-species analysis as described above. Each column is for one true tree and 1000 simulated data sets, for each of which the phylogenetic correlation is estimated. The four columns of histograms are for the first four of the 10 trees used for Table 1.

(Insert Figure 3 about here)

Each histogram shows the distribution of estimates of the phylogenetic correlation on a scale from -1 to 1 (only for the leftmost column is the label -1 visible). The value of 0 is indicated by a vertical line of thin dashes. The two tables are both for cases with a true phylogenetic correlation of 0.4 (shown on each histogram as the darker vertical dashed lines).

In Figure 2, the within-species phenotypic variation has a correlation of 0.8 between characters. In the analysis using only means, this correlation affects the estimates of the phylogenetic correlation, making it visibly biased upwards. In Figure 3, the within-species phenotypic correlation is -0.2, and the result for the analysis of means is a downward bias of the estimated phylogenetic correlation. In both cases the full analysis removes this bias, resulting in estimates whose mean is close to the truth. The distribution of estimates of the phylogenetic correlation is somewhat asymmetrical, particularly in the case of high estimates. As with ordinary correlation coefficients, the distributions become more symmetrical if we plot instead Fisher's $z$ transformation.

## Genetic experiments

The evolutionary covariation of characters between species, **A**, reflects either genetic drift or varying natural selection. A lineage can have the mean of the characters change by either of these. It has been known since the work of Sewall Wright that the covariance of means among lineages due to genetic drift is expected to be a multiple of the within-species additive genetic covariance. With natural selection the between-species evolutionary covariance will reflect a complicated compromise of the additive genetic covariances and the selective covariance, which is the covariance of selection pressures (see discussions in Felsenstein, 1988, pp. 446-454 and Felsenstein, 2004, pp. 415-426).

### *Mating designs*

In principle, if we had estimates of the additive genetic covariance, we could tease these apart and infer the selective covariance. Additive genetic covariances are often inferred by mating designs (such as half-sib families or diallel crosses). A design of sets of half-sub families that share male parents but have different female parents, has two levels of nesting. The design has a series of males, each mated to a set of females (the same female is not mated to more than one of these males). A large genetic experiment is desirable, but not always possible.

### *Contrasts*

REML analysis of variance in such a design strongly resembles the contrasts analysis that we have been doing. For example, in a conventional half-sib design, a contrast between half sibs will have expectation zero and covariances proportional to $\frac{3}{4}\mathbf{G} + \mathbf{D} + \mathbf{E}$, where $\mathbf{G}$ is the additive genetic covariances, $\mathbf{D}$ the dominance covariances, and $\mathbf{E}$ the environmental covariances. A contrast between means of half-sib families will have expectation zero and covariances proportional to $\frac{1}{4}\mathbf{G}$. In this design, it would be possible to incorporate the analysis of the genetic experiment with the comparative method analysis. For more complex designs of the genetic experiment it might be necessary to do the inferences of the covariances separately in the genetics experiment and the comparative method analysis.

*Estimating selective covariances*

We can estimate the covariance matrices for the genetic components from a mating design, and then compare the results to the phylogenetic covariances $\mathbf{A}$ and the within-species phenotypic covariance (here called $\mathbf{P}$). In this notation, we expect the phylogenetic covariances to be

$$\mathbf{A} \;=\; \mathbf{G}\,\mathbf{P}^{-1}\,\mathbf{S}\,\mathbf{P}^{-1}\,\mathbf{G} \tag{43}$$

(Felsenstein, 1988, p. 451) where $\mathbf{S}$ is the covariances of the selection differentials along lineages.

Given $\mathbf{A}$ and $\mathbf{P}$ inferred from the within- and between-species contrasts, and $\mathbf{G}$ from the genetic experiments, we can use this to estimate the covariance of selection

differentials as

$$\mathbf{S} \;=\; \mathbf{P}\,\mathbf{G}^{-1}\,\mathbf{A}\,\mathbf{G}^{-1}\,\mathbf{P} \tag{44}$$

It is to be expected that these selective covariances (covariances of selection differentials) will be hard to estimate accurately. One approach to their error would be to resample individuals in the genetic experiments, and to resample individuals in the population samples. The variation of $\mathbf{S}$ in this resampling could then be observed. However, it is not clear whether it is also desirable to resample whole species in the phylogeny. Another approach would be to use the likelihood surfaces for the estimates of the matrices $\mathbf{G}$, $\mathbf{P}$, and $\mathbf{A}$ together with equation (44) to infer the likelihood surface for $\mathbf{S}$. If the two analyses (genetic and comparative) could be combined, then the inference of $\mathbf{S}$ or tests of hypotheses about it would be particularly straightforward.

## Program

The EM algorithm approach to estimating the covariances $\mathbf{A}$ and $\mathbf{P}$ from a bifurcating phylogeny (with branch lengths) on $n$ species, and samples of $p$ phenotypes from $n_i$ individuals in the $i$th species has been available as the `W` menu option in program Contrast of the PHYLIP package, versions 3.6a and later, since July, 2000. However, the regression coefficients in that program were calculated incorrectly (owing to a mistaken form of equations 50 and 51) in all versions prior to version 3.67. That version was released in July, 2007. PHYLIP is available free from `http://evolution.gs.washington.edu/phylip.html` (or you can simply

type PHYLIP into any search engine).

The program estimates the covariances, and from them the regressions of the variables on each other, and the correlations between the variables. Another option is the analysis of multiple trees, so that bootstrap samples of trees can be analysed. Likelihood ratio tests are available of the hypothesis that a set of $q$ characters have no phylogenetic covariation with the remaining $p - q$ characters. A likelihood ratio test of the assertion that all characters have no phylogenetic covariation is also available. The program does not yet include the ability to make inferences from genetic experiments.

## Acknowledgments

I wish to thank Emília Martins and Michael Lynch for most helpful discussions and encouragement, and for sharing with me and allowing me to describe the results of their investigations of the behavior of Lynch's method. Thanks also to Luke Harmon for useful comments. I also wish to thank Ted Garland and Tony Ives for showing me their manuscript before publication. The associate editor and her reviewers made comments that were of great help and are much appreciated. This research was funded by National Science Foundation grant DEB-9815650 and National Institutes of Health grant R01 GM071639.

## APPENDIX A

### Distribution of the contrasts

If $\mathbf{z}$ is the vector of contrasts produced using the $(n-1) \times (n-1)$ matrix of contrast coefficients $\mathbf{C}$, their covariance matrix will be

$$\begin{aligned} \mathbf{V} &= \mathbb{E}\,(\mathbf{z}\mathbf{z}^T) = \mathbb{E}[\,(\mathbf{C} \otimes \mathbf{I})\,(\mathbf{y}\mathbf{y}^T)\,(\mathbf{C} \otimes \mathbf{I})^T\,] \\ &= (\mathbf{C} \otimes \mathbf{I})\,\mathbb{E}[\,\mathbf{y}\mathbf{y}^T\,]\,(\mathbf{C} \otimes \mathbf{I})^T \qquad (45) \\ &= (\mathbf{C} \otimes \mathbf{I})\,(\mathbf{T} \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{P})\,(\mathbf{C} \otimes \mathbf{I})^T \end{aligned}$$

Applying the multiplication rules for Kronecker products this becomes

$$\mathbf{V} = (\mathbf{C}\,\mathbf{T}\,\mathbf{C}^T) \otimes \mathbf{A} + (\mathbf{C}\mathbf{I}\mathbf{C}^T) \otimes \mathbf{P} \qquad (46)$$

which using orthonormality (equation 21) easily becomes

$$\mathbf{V} = (\mathbf{C}\,\mathbf{T}\,\mathbf{C}^T) \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{P}, \qquad (47)$$

For each character the contrasts $\mathbf{C}$ are mutually independent. Thus the matrix $\mathbf{C}\,\mathbf{T}\,\mathbf{C}^T$ is diagonal. The diagonal elements corresponding to within-species contrasts are all 0, and for the between-species contrasts the diagonal elements have nonzero values. Let us call this diagonal matrix $\mathbf{W}$. Then

$$\mathbf{V} = \mathbf{W} \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{P}. \qquad (48)$$

This covariance matrix is of size $(n-1)p \times (n-1)p$, consisting of $n-1$ diagonal blocks, each $p \times p$ in size.

We will use the diagonal elements of $\mathbf{W}$, the $w_{ii}$ in the estimation of the covariances. It is not necessary to form the product $\mathbf{C}\,\mathbf{T}\,\mathbf{C}^T$ explicitly. The variance of the unnormalized contrast $\bar{x}_L - \bar{x}_R$ is $v_L + \delta v_L + v_R + \delta v_R$ (for a character with unit variance of evolutionary change). The variance of the orthonormal contrast $K(\bar{x}_L - \bar{x}_R)$ is $w_{ii}$, so that from (19) we can obtain equation (20), and the relevant one of the $w_{ii}$ can easily be computed at each step of the recursion.

By applying the orthonormal contrasts within and between species, we have found sets of variables in the vector $\mathbf{z}$ that are independent. The variables within each set covary, and they have heterogeneous variances (as they have different $w_{ii}$).

## APPENDIX B

### The EM algorithm for inferring the covariances

For one of the sets of contrasts $\mathbf{z}^{(i)}$, and given the current estimates $\mathbf{A}(t)$ and $\mathbf{P}(t)$ of the covariance matrices, we want to obtain the expectation of $\mathbf{a}\mathbf{a}^T$ and $\mathbf{p}\mathbf{p}^T$. For this we will need the matrix of regression coefficients of $\mathbf{a}$ on $\mathbf{z}^{(i)}$. The covariance matrix of the contrasts $\mathbf{z}^{(i)}$ we know to be $w_{ii}\mathbf{A} + \mathbf{P}$. From 40 we can show that the covariance between $\mathbf{a}$ and $\mathbf{z}^{(i)}$ is

$$\mathbb{E}(\mathbf{z}^{(i)}\mathbf{a}^T) \;=\; \sqrt{w_{ii}}\,\mathbf{A}(t) \tag{49}$$

which through the usual expression for regression coefficients leads to

$$\mathbf{B_{a.z}} \;=\; \sqrt{w_{ii}}\;\mathbf{A}(t)\,(w_{ii}\mathbf{A}(t)+\mathbf{P}(t))^{-1}. \tag{50}$$

An exactly analogous derivation for the regression of $\mathbf{p}$ on $\mathbf{z}^{(i)}$ gives

$$\mathbf{B_{p.z}} \;=\; \mathbf{P}(t)\,(w_{ii}\mathbf{A}(t)+\mathbf{P}(t))^{-1}. \tag{51}$$

Having the regression coefficients, we want the expectations of the sufficient statistics. The expectation of $\mathbf{aa}^T$ given $\mathbf{z}^{(i)}$ can be computed by finding $\mathbb{E}[\mathbf{a}\,|\,\mathbf{z}^{(i)}]\;\mathbb{E}[\mathbf{a}\,|\,\mathbf{z}^{(i)}]^T$ and adding to it the residual variance around the regression line of $\mathbf{a}$ on $\mathbf{z}$. The result is

$$\mathbb{E}[\mathbf{aa}^T] \;=\; \mathbf{B_{a.z}}\,\mathbf{zz}^T\,\mathbf{B_{a.z}}^T \;+\; \mathbf{A}(t) \;-\; \mathbf{B_{a.z}}(w_{ii}\mathbf{A}(t)+\mathbf{P}(t))\mathbf{B_{a.z}}^T \tag{52}$$

which when rearranged gives the expectation of $\mathbf{aa}^T$ given $\mathbf{z}^{(i)}$ as

$$\mathbf{A}(t) \;+\; \mathbf{B_{a.z}}\,(\mathbf{zz}^T-(w_{ii}\mathbf{A}(t)+\mathbf{P}(t))\,)\,\mathbf{B_{a.z}}^T \tag{53}$$

A very similar derivation for the expectation of $\mathbf{pp}^T$ given $\mathbf{z}^{(i)}$ gives

$$\mathbf{P}(t) \;+\; \mathbf{B_{p.z}}\,(\mathbf{zz}^T-(w_{ii}\mathbf{P}(t)+\mathbf{P}(t))\,)\,\mathbf{B_{p.z}}^T \tag{54}$$

One of these computations is done for each set of contrasts; the new estimates of $\mathbf{A}$ and $\mathbf{P}$ are then computed as the averages of the expectations of $\mathbf{aa}^T$ and $\mathbf{pp}^T$ for all contrasts. For the contrasts within species the regressions $\mathbf{B_{a.z}}$ are zero; for those equation (53) shows that the expectations of $\mathbf{aa}^T$ are to be taken as given by $\mathbf{A}(t)$.

If we have a constraint on the covariances, such as that there is no phylogenetic covariance between two sets of characters, this is easily maintained by setting those elements of $\mathbf{A}$ to zero at each stage of the EM iteration.

If we are testing the proportionality of $\mathbf{A}$ to $\mathbf{P}$, we can use a model in which $\mathbf{A} = \alpha\mathbf{P}$. Iteration equations can be derived from the likelihood functions to obtain estimates of $\alpha$ and $\mathbf{P}$. Given current estimates of $\alpha$, the new estimate of $\mathbf{P}$ is

$$\mathbf{P}' = \frac{1}{n-1} \sum_i \frac{\mathbf{z}^{(i)}\mathbf{z}^{(i)\,T}}{w_{ii}\alpha + 1} \tag{55}$$

and given the current estimate of $\mathbf{P}$, a new estimate of $\alpha$ is obtained from

$$\alpha' = \frac{\sum_i f_i \, \frac{1}{w_{ii}} \left( \frac{\mathbf{z}^{(i)\,T}\mathbf{P}^{-1}\mathbf{z}^{(i)}}{p} - 1 \right)}{\sum_i f_i}, \tag{56}$$

where

$$f_i = \frac{w_{ii}^2}{(\alpha\, w_{ii} + 1)^2}. \tag{57}$$

These should converge on the REML estimates of $\alpha$ and $\mathbf{P}$, and thereby on the estimate of $\mathbf{A}$ as well.

## Literature Cited


Christman, M. C., R. W. Jernighan, and D. Culver. 1997. A comparison of two models for estimating phylogenetic effect on trait variation. Evolution 51:262-266.

Dempster, A. P., N. M. Laird and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39:1-38.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1-15.

Felsenstein, J. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445-471.

Felsenstein, J. 2002a. Quantitative characters, phylogenies, and morphometrics. Pages 27-44 *in* N. MacLeod and P. L. Forey, eds. Morphology, shape, and phylogenetics. Systematics Association Special Volume Series 64. Taylor and Francis, London.

Felsenstein, J. 2002b. Contrasts for a within-species comparative method. Pages 118-129 *in* M. Slatkin and M. Veuille, eds. Modern developments in theoretical population genetics. The legacy of Gustave Malécot. Oxford University Press, Oxford.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, Masschusetts.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London Series B - Biological Sciences 326:119-157.

Harmon, L. J. and J. B. Losos. 2005. The effect of intraspecific sample size on type I and type II error rates in comparative studies. Evolution 59:2705-2710.

Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. American Naturalist 163:84-96.

Ives, A. R., P. E. Midford, and T. Garland, Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Systematic Biology 56:252-270.

Lauritzen, S. L. 1996. Graphical models. Clarendon Press, Oxford.

Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065-1080.

Purvis, A., and T. Garland, Jr. 1993. Polytomies in comparative analyses of continuous characters. Systematic Biology 42:569-575.

Ricklefs, R. E. and M. J. Starck. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. Oikos 77:167-172.

**Table caption**

Number of times that the (true) null hypothesis of no correlation between two characters was rejected when the analysis used only the species means, and when the analysis took the within-species variation into account. Each row is a different 40-species tree with 4 individuals measured per species.

**Table**

| tree | $\alpha$ (at nominal 0.05) using only species means | $\alpha$ (at nominal 0.05) using within-species information as well |
|---|---|---|
| 1 | 137 / 1000 | 53 / 1000 |
| 2 | 237 / 1000 | 53 / 1000 |
| 3 | 305 / 1000 | 58 / 1000 |
| 4 | 84 / 1000 | 66 / 1000 |
| 5 | 149 / 1000 | 61 / 1000 |
| 6 | 102 / 1000 | 59 / 1000 |
| 7 | 259 / 1000 | 65 / 1000 |
| 8 | 76 / 1000 | 61 / 1000 |
| 9 | 205 / 1000 | 50 / 1000 |
| 10 | 446 / 1000 | 47 / 1000 |

## Figure Captions

Fig. 1 – The phylogeny connecting all individuals of all species in an example of the model. Five species are shown as connected by a phylogeny. One of the species has been magnified to show that it actually consists of a sample of 4 individuals, each connected to the phylogeny by its own branch. In the model, these branches are of length 0, though in the figure they are shown as having a small length so as to make the connections clear.

Fig. 2 – Each column of histograms shows the estimates of the phylogenetic correlations between two characters which have been simulated with true phylogenetic correlation 0.4 in 1000 data sets on the same tree. The top analysis uses the species means and the original contrasts method, the bottom ones use the full analysis described in this paper. The within-species variance is taken to be 1.0 with correlation of 0.8 of the within-species effects. Four individuals are sampled per species. The same 1000 data sets are analyzed in each column. The four columns are for the first four trees used in the simulation.

Fig. 3 – Each column of histograms shows the estimates of the phylogenetic correlations between two characters which have been simulated with true phylogenetic correlation 0.4 in 1000 data sets on the same tree. The top analysis uses the species means and the original contrasts method, the bottom ones use the full analysis described in this paper. The within-species variance is taken to be 1.0

with correlation of -0.2 of the within-species effects. Four individuals are sampled per species. The same 1000 data sets are analyzed in each column. The four columns are for the first four trees used in the simulation.