Chapter 25

Comparative methods

Comparative methods use the distribution of traits across species to make inferences about the effect on their evolution of other traits or of environments. Gallons of ink have been spilled over the years by biologists writing about the importance of the comparative method, but only in the last 25 years have they understood that use of the comparative method requires phylogenies and statistical methods that use them. Figures 25.1 and 25.3 show how, when we do not use phylogenies, a seemingly straightforward analysis of individual species can create an artifactual signal.

An example with discrete states

Figure 25.1 shows a phylogeny with 10 species and two 0/1 characters. On the phylogeny both characters happen to change once, in the same branch. When the species are taken as individual sample points and a contingency table is created, it appears to support a strong correlation between the states of the character. Yet the phylogeny makes it clear that this is an illusion. If all branches on the phylogeny were of equal length, it would be safer to make a contingency table of branches. If the branch where both characters changed was longer than the other branches, the coincidence would be even less surprising.

The contingency table on the left side of the figure shows the apparent tight correlation between the two characters. If we accepted the species as being independently distributed, we could use a Fisher's exact test and we would find that the probability of a correlation this close or tighter is very small, $1/210 = 0.0047619$. This would seem to establish that the two characters have evolved in a closely correlated manner. But if we made the calculation this way, we would be deceiving ourselves. The species cannot be regarded as independent outcomes of evolution. They are related, having evolved on a phylogeny.

A more relevant way of looking at the evolution of these characters is to consider on which branches of the phylogeny the characters changed. The figure re-

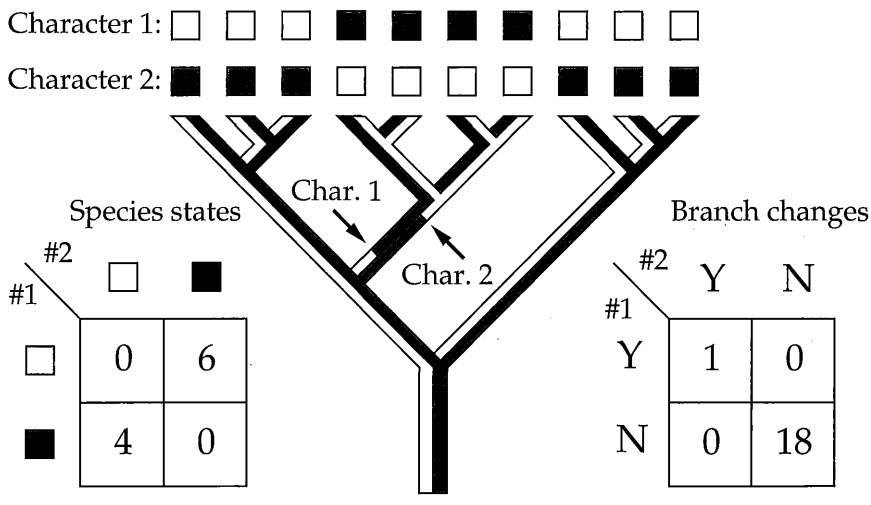


Figure 25.1: An example of two discrete two-state characters evolving along a phylogeny so that their distributions among the species are correlated. Each branch of the phylogeny consists of two regions side by side, for characters 1 and 2. Shading shows which state each is in. The states of the two characters are shown at the tips of the tree by empty or shaded boxes.

constructs where each character changed. Both changes happened on the same branch of the phylogeny. Not counting the root, there are 19 branches on this phylogeny (18 if we remove the root and unite the two most basal branches). The probability that both characters would change on the same branch would be 1/18 = 0.05555. This fails to be significant; it is far less extreme than the previous probability.

An example with continuous characters

Figure 25.2 shows a tree with 22 species, which has two clades of 11 species each. The vertical scale indicates the branch length. There is no structure within each of the clades, but they have been separated for some time. Figure 25.3 shows two characters that have evolved by Brownian motion along this phylogeny, with no correlation between the evolution of the characters. The 22 species are plotted as points, and in the left plot there seems to be a correlation between the two characters — a larger value of one goes with a larger value of the other. But on the right side we have shaded in the points according to which clade they come from (the shadings are the same as in Figure 25.2).

Looking at the shaded points, we can see that within each of the clades there is no particular sign that the two characters are correlated. All of the correlation

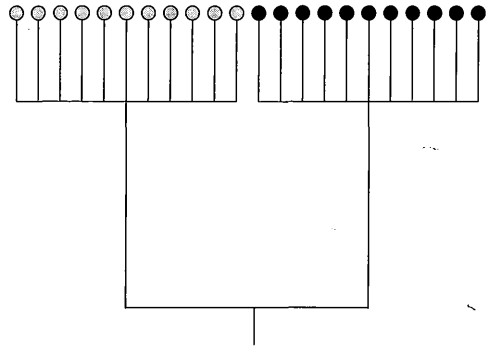


Figure 25.2: A tree with 22 species and two clades, which was used to evolve the characters shown in Figure 25.3.

arises from the difference between the clades — from the changes of the two characters along the two large internal branches of the tree. Those two branches are effectively only one branch, as we have no outgroup to give us separate information about the root of the tree. So the correlation in a plot of 22 points comes from 11 of them being displaced upwards and to the right, compared to the other 11. This is due to a single event, a burst of change in the branch separating the two

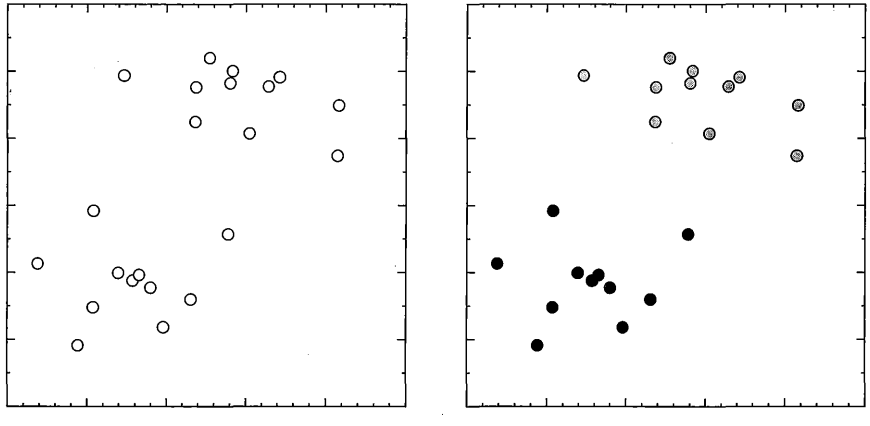


Figure 25.3: Two characters that evolved in an uncorrelated fashion on the tree of Figure 25.2. There appears to be a correlation, but when the points are shaded to show which clade they came from, it becomes apparent that this is an artifact of the phylogeny, as within each of the two clades there is no correlation. The scales on the axes are omitted, as they are arbitrary.

clades. The number of points (22) is a misleading indication of the number of independent sources of variation.

The whole point of phylogenies is that species did not evolve independently, but that historical events affected large groups of species. Only by untangling the correlation of the species can we untangle the correlations of the characters.

The contrasts method

That there was a problem in the statistical analysis of the comparative method was recognized by a number of evolutionary biologists in the late 1970s (cf. Clutton-Brock and Harvey, 1977, 1979; Harvey and Mace, 1982; Ridley, 1983). Methods based on analysis of variance were attempted (Harvey and Mace, 1982), and Ridley (1983) discussed how to analyze data on discrete characters. We return to Ridley's method and other discrete characters methods later; for the moment we concentrate on the analysis of continuous character data, for which the answer is found in the contrasts method.

The *contrasts method* was introduced by me (Felsenstein, 1985a), using techniques developed for computational efficiency when computing likelihoods on phylogenies where the characters have evolved by Brownian motion. In Chapter 23 we have seen that, for any character evolving by Brownian motion along a phylogeny, we can find a series of contrasts between the character values at the tips that are statistically independent. Figure 25.4 shows a phylogeny, together with the contrasts it implies. Below the figure are the contrasts that are derived from this phylogeny, together with their variances.

The contrasts are expressed by their coefficients, shown in the figure. When these are used, the resulting numbers may be called contrast scores. The contrast scores may be taken as new variables. They express all the variation between species, leaving out only the grand mean of the character. We can also divide each contrast score by the square root of its variance, so that (in our example):

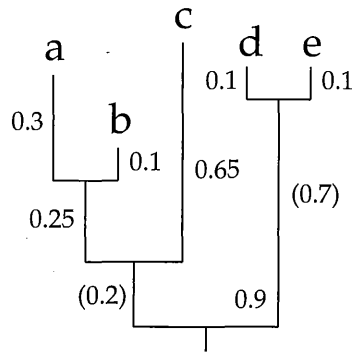
$$\begin{aligned} Y_1 &= y_1/\sqrt{0.4} \\ Y_2 &= y_2/\sqrt{0.975} \\ Y_3 &= y_3/\sqrt{0.2} \\ Y_4 &= y_4/\sqrt{1.11666} \end{aligned} \tag{25.1}$$

and we obtain standardized contrast scores.

Assuming that the Brownian motion model is correct and so is the tree, the contrasts all have the same expected mean and variance. Note that

1. The contrasts all have expectation zero.

2. They assume that we know the mean phenotype of the populations at the tips of the trees precisely.



Contrast												Variance proportional to
y_1	$=$	x_a	$-$	x_b								0.4
y_2	$=$	$\frac{1}{4}\,x_a$	$+$	$\frac{3}{4}\,x_b$	$-$	x_c						0.975
y_3	$=$							x_d	$-$	x_e		0.2
y_4	$=$	$\frac{1}{6}\,x_a$	$+$	$\frac{1}{2}\,x_b$	$+$	$\frac{1}{3}\,x_c$	$-$	$\frac{1}{2}\,x_d$	$-$	$\frac{1}{2}\,x_e$		1.11666

Figure 25.4: An example phylogeny and the independent contrasts that it implies under a model of evolution by Brownian motion. The branch passing through the bottommost node has total length 0.9, as two branches of length 0.2 and 0.7 separate these two clades.

3. If the tree's branch lengths are all multiplied by the same constant (say, 2.347) the contrasts will still all be independent, have zero means, and all have the same variance; their variances will all be divided by 2.347.

Figure 25.5 shows the contrasts for the numerical example of Figures 25.2 and 25.3. The points for the contrasts within the two clades have the corresponding shadings, and the single contrast between the clades is unshaded.

Correlations between characters

If we have two or more characters, we can apply the contrasts method to each simultaneously. The contrast formulas will be the same in each character. We have seen in equation 23.39 that when there are multiple characters undergoing correlated Brownian motion, a set of contrasts can be found that are independent. The covariances of the characters in these contrasts will be proportional to the covariances ($\mathbf{A}$) of evolutionary change among the characters.

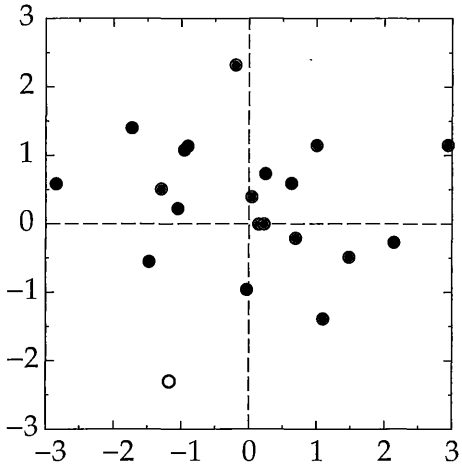


Figure 25.5: The contrasts for the two characters in the example of Figures 25.2 and 25.3. The 10 contrasts within each of the two clades are shaded correspondingly, and the single contrast between the two clades is unshaded. The scales are in standard deviation units.

As an example, consider the tree of Figure 25.4 with three characters. We have five species and three characters, so that we have 15 species means. If we take the four contrasts in each of the characters, for contrast y_1 we will have three values, one for each of the characters. The covariance matrix of changes in the three characters along the tree (**A**) will then also be the covariance we expect to see among the three values of y_1. There will be a similar covariance among the three values of y_2, with the same covariance matrix. But the y_1 values and the y_2 values will be independent.

Thus we can simply take the four contrast scores for each of the three characters, and feed them into a standard multivariate statistics package. They give us four samples with which to estimate the covariances among the characters. We can do the usual kinds of multivariate statistical analysis on them, such as using principal components to find the combination of characters that shows the most change along the tree.

When the tree is not completely known

The contrasts method requires that we know the phylogeny, including its branch lengths. Frequently, some regions of the tree have unresolved structure, as they are considered to be poorly known. These are often described as "soft polytomies," as they are not positively known to be polytomous, but describe our ignorance instead. A number of suggestions have been made as to how to deal with these:

- I suggested (1988a) that we bootstrap-sample the original data that inferred the tree, inferring a tree from each bootstrap sample. The comparative method would be performed on each bootstrap sample. To place a confidence limit on an inferred parameter such as a slope, one would make a histogram of its values from all these analyses, and reject a value such as 0 only if it fell in the appropriate tail of the histogram.

- Grafen (1989), Harvey and Pagel (1991), and Pagel (1992) proposed methods of dealing with polyfurcations in the tree, methods that go beyond the contrasts computation. Purvis and Garland (1993) suggested reducing the degrees of freedom to count for the polyfurcation. Garland and Díaz-Uriarte (1999) presented simulation results arguing that this worked well. Rohlf (2001) pointed out that in their method if the tree were a complete multifurcation, there would be zero degrees of freedom in spite of the presence of data.

- Martins (1996a) proposed that when a tree topology is completely unknown, one simulate trees at random for these species, then analyze the data on each and take the mean of the inferred quantity as the best estimate. Housworth and Martins (2001) suggested ways of simulating random parts of trees when only a portion of the tree is unknown. Abouheif (1998) suggested that randomly generated bifurcating trees will give the same result as having only a single multifurcation. In this he was supported by the simulations of Symonds (2002). It would seem possible to treat this analytically. The approach of Martins and Housworth (2002) would seem relevant.

If polyfurcations in the tree are taken to be real ("hard polytomies"), the correct way of dealing with them is to resolve them into bifurcations by adding zero-length branches in any reasonable way. It can be shown (Rohlf, 2001) that how this is done does not affect the values of regressions and correlations calculated from ordinary multivariate analyses.

Inferring change in a branch

McPeek (1995) has discussed using the contrasts method to infer the amount of change along a particular branch. If we take the tree as unrooted, then prune the phenotype values down to both ends of the branch, we will obtain values such as x'_1 and x'_2, with extra variances δ_1 and δ_2 added to the two ends. If the branch has original length v_3, a simple regression argument shows that the actual change of the character along that branch is distributed normally around a mean $x'_1 - x_2{}'$ multiplied by $v_3/(v_3 + \delta_1 + \delta_2)$. The variance of the distribution is $v_3(\delta_1 + \delta_2)/(v_3 + \delta_1 + \delta_2)$. McPeek points out that reconstructed changes on different branches can be correlated (in fact, all of them are). However, his formulas for the reconstructed changes differ from the ones I have given here. His formulas

give nonzero reconstructed change even when $v_3 = 0$, while the above formulas correctly infer that change is zero in such a case.

Sampling error

Ricklefs and Starck (1996, p.169), in the midst of a skeptical review of applicability of contrast methods, make an insightful point about sampling error. If we do not have the actual character mean for each species, because we have only a finite sample of individuals, this adds an extra source of error beyond the randomness of Brownian motion. This sampling error will depend on sample sizes, but not on branch lengths of the tree. If a contrast is taken between two species that are neighbors on the tree, the contrast will be divided by a small quantity, the branch length separating those two tips. This assumes that the variance of the difference between them is substantially smaller than if they had been farther apart on the tree. But if sampling error is an important source of variance, the variance of the contrast between neighbors is then underestimated. Ricklefs and Starck find that contrasts of closely related species are often outliers in their regressions. This is presumably an artifact of sampling error (some of which could be due to measurement error).

The solution to this problem is to take the sampling error explicitly into account. Riska (1991) discussed the need to include sampling error in our models. We must add a variance component due to variation between individuals within a population, and allow for its variance when inferring covariances between characters. Each individual's measurement is distributed normally, with a term added to the model for within-species variation. We have seen that if all the variation is due to evolutionary change of population means, the covariance matrix of multiple characters in multiple species is $\mathbf{T} \otimes \mathbf{A}$. When there are also sampling effects, suppose that we draw multiple individuals from species i. The covariances between characters k and ℓ will then have an extra component $\sigma_{k\ell}$ if they are measured in the same individual. This means that an extra component of variance (which we call e_{kl}) is added to covariances of characters in the same individual.

The model can be written concisely by taking the individuals as the unit. We can imagine a phylogeny like Figure 25.6 connecting all the individuals in the study. The tree that is shown is basically the same as in Figure 25.4, except that extra branches have been added for the individuals sampled for each species. In this model, the covariances added by these branches are not derived from the covariances of changes in phenotype in evolution, but from the covariances of characters within species. The effect of this is that the covariance matrix can be written as

$$\mathbf{T} \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{E} \tag{25.2}$$

where $\mathbf{T}$ is a matrix for a tree that has tip i replicated n_i times, where n_i is the sample size for species i on that tree. The extra branches that are added have length 0. The new term in this covariance matrix, $\mathbf{I} \otimes \mathbf{E}$, adds covariances of different charac-

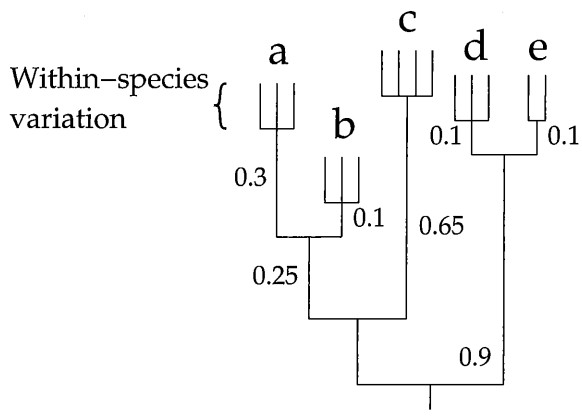


Figure 25.6: The same tree as in Figure 25.2, but with extra branches for each species, one per member of the sample for that species.

ters within each individual, which are the within-species phenotypic covariances of characters. The distribution of phenotypes in individuals is

$$\boldsymbol{x} \sim \mathcal{N}(\mathbf{1} \otimes \boldsymbol{\mu},\ \mathbf{T} \otimes \mathbf{A} + \mathbf{I} \otimes \mathbf{E}) \tag{25.3}$$

where $\mathbf{x}$ is a vector that not only stacks up the different species, but within them the members of the sample and within those their different characters, and where $\mathbf{T}$ is a tree with a branch of length 0 for each sampled individual.

We cannot simply regard these extra branches, one for each individual in the samples, as involving the same differences as those that accumulate in evolution. It is true that contrasts among these individual branches can be taken, and they are independent of each other. If a species has n_i individuals in its sample, it is quite easy to construct $n_i - 1$ contrasts among the individuals, all of which are independent. However, the character covariances in those contrasts are the covariances $\mathbf{E}$, the within-species phenotypic covariances, and these may be quite different from the covariances $\mathbf{A}$ of between-species phenotypic change. If we take all the within-species contrasts, this leaves at each species its arithmetic mean phenotype, as seems appropriate.

You might think that there is a neat separation between one set of contrasts that measures within-species phenotypic variation and another that measures between-species changes. But when we set out to take between-species contrasts, we should remember that the species means reflect not only the random changes between species, but also our sampling error due to having only a limited number of individuals when there is within-species variation. For example, if species a has a sample size of 3, then when the within-species variance of a character is σ^2, this creates an extra variability of $\sigma^2/3$ in the species mean beyond the variability of evolutionary change.

It will not necessarily be true that the covariation of characters among species due to evolutionary change is proportional to the covariation within species due to phenotypic variation and finiteness of the samples. Two characters could covary positively within species, and yet change in a negatively correlated way between species.

One case of this model has already been treated in the literature. Lynch (1991) used a model where, in addition to its phylogenetic covariances, each species has a component due to recent change that is specific to that species and has its own covariances. This gives the same model as if we had sampled only one individual per species. Using an REML method from quantitative genetics, Lynch was able to provide iterative equations for estimating the different variance components.

I have (Felsenstein, *in prep.*) provided an alternative method of computation for the model with sampling error, using contrasts that are not standardized in the usual way, but are orthonormal contrasts (the sum of squares of their coefficients sum to 1). As we saw with the variance of a single character, the finiteness of the sample sizes implies that a small portion of the within-species phenotypic covariances becomes added to the between-species covariances. The computation must be iterative, converging on REML estimates of the within- and between-species covariances. We can no longer simply take the contrasts of all characters and feed them into a standard multivariate statistics package, but we can maximize likelihood with respect to the general model and with respect to a restrictive model (such as a model that has no phylogenetic covariance between two characters), and do a likelihood ratio test in the usual way.

The effect of allowing for sampling error and within-species phenotypic variation is to discount many of the contrasts between closely-related species. If we have a pair of species that are close relatives, ordinary contrasts analysis inflates the contrast between them, dividing it by the square root of the branch length separating them. But if there is also sampling error, most of the contrast may come from that. When the sampling error is properly taken into account, the contrast between that pair of species may contribute little to the inference of phylogenetic covariances and correlations. This makes intuitive sense.

When Lynch's method was tested (E. Martins and M. Lynch, personal communication; Felsenstein, *in prep.*) it was found that estimates of phylogenetic correlation between traits tend to be too extreme. This simply reflects the small effective sample sizes; if we have 20 species, but many fewer clusters of closely related ones, the correlation coefficients are estimated mostly from the between-cluster contrasts, which are fewer. Correlations are usually biased to be too extreme, and this bias is serious when sample size is small. For example, when sample size is 2 we know that the correlations that can be observed are almost always +1 or –1. The extreme values found in simulations reflect the fact that the between-species observations are equivalent to a rather small number of independent points.

Correction for sampling error is still rare in comparative studies; hopefully it will become more common. It should place a premium on having a sufficiently great diversity of species in the study.

The standard regression and other variations

Grafen (1989, 1992) has given a multivariate statistical framework that is an alternative to the contrasts method; he calls it the *standard regression*. In the case in which the tree is fully bifurcating, careful consideration of his method will show that it is simply an alternative computational scheme to contrasts and will always give the same results. His methods use matrix calculations instead of recursive derivation of contrasts. The computation of the contrasts is, in effect, a way of obtaining eigenvalues and eigenvectors of the covariance matrix of species.

Generalized least squares

Martins (1994) presents an alternative generalized least-squares framework that is not exactly the same as REML, but that will approximate it in most cases. Rohlf (2001) argues that these two methods are actually equivalent. Paradis and Claude (2002) suggest using generalized estimating equations (GEE) rather than the generalized least squares (GLS) framework of Martins. Their method is more general than GLS but reduces to it in the Brownian motion case. The chief utility of these methods is that when we have other evolutionary mechanisms, they can be modified to approximate them. Martins, Diniz-Filho, and Housworth (2002) have examined by computer simulation how robust some of these methods are to variations in the evolutionary model.

Phylogenetic autocorrelation

Cheverud, Dow, and Leutenegger (1985) used a method developed to correct for geographic structure in data, adapting it to phylogenies (see also Gittleman and Kot, 1990). Their "phylogenetic autocorrelation" method has been tested against contrasts methods in simulations, with indifferent results (Martins, 1996b). It has been criticized by Rohlf (2001) as having some assumptions that conflict with any possible evolutionary mechanisms. Chief among these is the way error enters the model. Differences between closely related species are expected to be as variable as differences between distantly related ones; this is inconsistent with mechanisms such as Brownian motion.

Transformations of time

Gittleman and Kot (1990) added to their method a transformation of the time scale. Suspecting that the Brownian motion model might not accurately represent the evolutionary process, they allowed the time scale to be nonlinearly transformed. Actually, their scale was not so much time as it was a distance that represented

taxonomic dissimilarity. By adding a parameter α to make the covariances proportional to this distance raised to the αth power, they allow the data to dictate how taxonomic groupings are reflected in covariances. The phylogenetic autocorrelation method also has a constant ρ that controls how much phylogeny influences similarity.

Pagel (1994), in his method for discrete characters, used a different approach in which the length of each branch of the tree is raised to a power, with the power being estimated.

Should we use the phylogeny at all?

The most drastic modification of the contrast method is to discard it. As phylogenetic comparative methods have first disrupted comparative work, and then become compulsory, this has led to questioning of their celebrity status. A particularly interesting exchange in the pages of *Journal of Ecology* raises the issue of whether seeing an effect of phylogeny proves that the effect is not due to natural selection (Westoby, Leishman, and Lord, 1995a, b, c; Harvey, Read, and Nee, 1995a, b; Ackerly and Donoghue, 1995). The paper of Ricklefs and Starck (1996) is another interesting critique.

Freckleton, Harvey, and Pagel (2002) suggest that a parameter defined by Pagel (1999a) can be used to test whether there are any phylogenetic effects at all. The covariance matrix between species is taken to be λ times the matrix predicted from the phylogeny, plus $(1 - \lambda)$ times a matrix in which all species are independent. By doing likelihood ratio tests on λ one can test whether there is any sign of an effect of phylogeny. This is essentially the same as a comparable test in the work of Lynch (1991) in which the model included both a phylogenetic effect and an individual species effect.

Paired-lineage tests

An alternative to the contrasts method is to look at pairs of species, chosen from the tree in such a way that the paths between them do not overlap. Looking at pairs of related species is an old method that dates back to at least the work of Salisbury (1942). I have extended it (Felsenstein, 1985a) to allow pairs that are connected by nonintersecting paths on the tree. Figure 25.7 shows the choice of such pairs. We can make a sign test of whether two characters change in the same direction (+) or opposite directions (–) between members of the pairs. If there is no correlation in the evolution of the two characters, our expectation is that these two outcomes would be equally frequent. So a simple sign test suffices.

Figure 25.7 shows a tree with one of the ways that the species can be divided into pairs. The paths between the species are shown by the dark lines. In general there are multiple ways that species can be grouped into pairs so that the paths between the members of the pairs do not overlap. On this tree there is only one grouping that satisfies this condition and finds four pairs, and it is shown here.

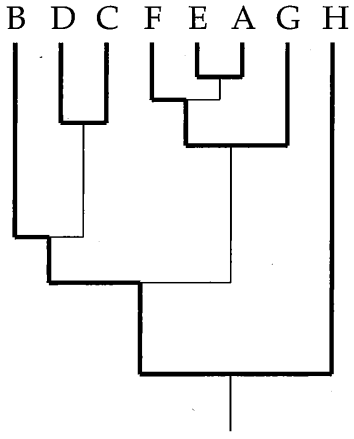


Figure 25.7: Pairs of species chosen so that the paths between the two members of each pair do not overlap. The paths between the pairs of species are indicated by the dark lines.

The paired-lineages test is appealingly simple, but it does lose some statistical information. A set of 19 species will have eight such pairs of lineages, but it will have 18 contrasts in the contrast method. An additional issue is that some pairs may have long paths between them. Depending on the exact scheme of evolution these pairs could fail to show a correlation. Of course, the same could be said for contrasts methods using a contrast with long branch lengths between the groups.

Ackerly (2000) presents simulation results on different sampling strategies used with different methods of analysis, including paired-lineage tests. He argues that choosing pairs of species so as to maximize the differences between species in one character can bias the regression of another character on that one.

Discrete characters

The contrasts method does not help us if we have discrete characters. Both parsimony and likelihood methods have been suggested to test whether characters are correlated in their evolution.

Ridley's method

Mark Ridley (1983), in probably the earliest method for treating comparative method data statistically, suggested mapping the two traits onto a tree using parsimony. Ridley's method is not entirely clear in his book. Wayne Maddison (1990) expresses some uncertainty as to what it is; Harvey and Pagel (1991) argue that it involves investigating the independence of the states in two characters at the upper ends of all branches. Sillén-Tullberg (1988) interprets Ridley's method as in-

volving examining whether the origins of the derived state in one character occur more often with one state of the other character than with the other state. However, she does not present a statistical test. Still another interpretation is that Ridley intends examining all branches and scoring each character as to whether it has changed or not in that branch. A contingency table can then be used to test association (as was done at the beginning of this chapter).

Concentrated-changes tests

Wayne Maddison (1990) has proposed a statistical test with discrete 0/1 characters. Depending on one's interpretation of Ridley's method, it may be an implementation of it. It asks whether the chance of a change in one character (call it character #1) is affected by the state of another character (call it character #2). He would have us start by reconstructing the regions of a tree that have one state or the other in character #2. Then he reconstructs the locations of changes in character #1. Suppose that there are seen to be seven occurrences of state 1 in character #1, and that these represent five changes of state in the tree, four of which are in regions with character state 1 in the other character. Is this number (4) a surprisingly large number?

Maddison uses a recursive algorithm to compute the probability of having four or more changes, out of five in total, in regions having character state 1 in character #2. He computes the number of different ways that one could have five changes of state of character #1 on the tree, and the fraction of those in which there are four or more changes in the regions that have state 1 in character #2. This fraction is taken as the probability, under the null hypothesis of no concentration of character #1's changes.

Sillén-Tullberg (1993) presents an alternative to Maddison's concentrated-changes test. Like Maddison, she makes parsimony reconstructions, including assignment of states to interior nodes in the tree. She looks at those branches that have state 0 of character #1 at their base and makes a contingency table. She classifies them according to whether or not there is a change to state 1, and whether or not the branch has state 0 or state 1 for character #2. One ambiguity in this test is when character #2 changes in a branch; it is not obvious which state of that character to assign to the branch. If multiple, equally parsimonious reconstructions are possible in either character, this can also cause trouble.

There are other possible ways to compute a probability for a concentrated-changes test. One could, for example, randomize the occurrences of state 1 in character #1 among all the tips on the tree, and count in how many of those cases there were more than four changes in the relevant regions. This would give a different result, but it would not be correct, because it would implicitly assume that even sister species were not correlated in their states. Maddison's test does a better job of allowing neighboring species on the tree to have correlated states. However, it does restrict its attention to those outcomes with a certain number of total changes in character #1. If we had a stochastic model for the evolution of that

character, we could have randomized over outcomes of evolution. However, in the absence of that kind of model, Maddison's quantity is probably as good as we can do.

In addition to this issue, this concentrated-changes test is completely dependent on the accuracy of the reconstruction of state changes by parsimony, as Wayne Maddison (1990) acknowledges. It assumes that we can know, without error, where each character's changes of state were. In doing so, it fails to take into account our full uncertainty as to how characters have changed.

A paired-lineages test

Read and Nee (1995) have proposed a test with discrete binary characters which is a natural counterpart to the continuous-characters paired-lineages test. Taking care to keep lineages from overlapping, the test takes pairs of species that differ at both characters. A simple sign test can then be done to check whether the character correlations are more in one direction than the other. A difficulty with this test is that there may be too few pairs of lineages that differ in both characters to allow for much statistical power. Wayne Maddison (2000) has discussed algorithms for choosing these pairs in as effective a way as possible.

Methods using likelihood

Pagel (1994) made the first likelihood-based comparative method for discrete characters. He assumed that two characters each have two states, 0 and 1, and that there is a simple stochastic process of change back and forth between them. There were two rates of change, so that the probability of change in a tiny interval of time of length dt would be $\alpha\, dt$ if the character were in state 0, and $\beta\, dt$ if it were in state 1. If the two characters were changing independently, for any given tree with branch lengths, we could compute the likelihoods for each character on the tree and take their product as the overall likelihood. The algorithms are essentially identical to those in Chapter 16, except that the number of states is two instead of four.

However, if the two characters are changing in a correlated fashion, the state of each can affect the probability of change of the other. In effect, we then have four states, for the four different combinations of states at the two characters. If both characters are in state 0, the combination is counted as being in state 00. There are also states 01, 10, and 11. If we assume that in any tiny interval of time only one of the characters can change, the matrix of rates of change of the character combinations is as shown in Table 25.1.

The notation in the table, which differs from Pagel's, has rates α and β of forward and backward change at the first character, and γ and δ of change at the second character. However, they differ according to the state of the other character. When character 2 is 0, the rates of change at character 1 are α_0 and β_0. But when character 2 is 1, they are α_1 and β_1. Similarly, the rates γ and δ for character 2 are subscripted according to the state of character 1.

Table 25.1: Rates of change between combinations of states in Pagel's (1994) discrete-character comparative method.

To : / From :	00	01	10	11
00	—	γ_0	α_0	0
01	δ_0	—	0	α_1
10	β_0	0	—	γ_1
11	0	β_1	δ_1	—

I will not go into the computational details, but it is possible to compute transition probabilities and equilibrium frequencies for the four states for this model. Using them, we can compute the likelihood for a tree (perhaps one obtained by molecular methods). More tediously, one can maximize this likelihood over the values of the parameters.

Some hypotheses of interest are restrictions of the values of these parameters. If the subscripting of α, β, γ, and δ does not affect their values, then the two characters are evolving independently. This is the set of constraints

$$\begin{aligned} \alpha_0 &= \alpha_1 \\ \beta_0 &= \beta_1 \\ \gamma_0 &= \gamma_1 \\ \delta_0 &= \delta_1 \end{aligned} \tag{25.4}$$

If we maximize the likelihood while maintaining these constraints, we are restricting four of the eight parameters. It is possible to do a likelihood ratio test comparing this likelihood to the likelihood with unrestricted parameter values, and this has 4 degrees of freedom. It is also possible to test, in a similar fashion, whether character 1 is unaffected by character 2, and whether character 2 is unaffected by character 1. These each involve restricting two of the parameters (α and β or γ and δ). Pagel noted that one can also test individual ones of these four parameters for being unaffected by the state of the other character, and each of these tests has 1 degree of freedom.

Another possibility is that one character affects the rate of evolution of the other, but not the equilibrium frequency of its two states. If $\alpha_0/\beta_0 = \alpha_1/\beta_1$, then character 1 will have the same equilibrium frequency no matter what the state of character 2. There is a similar condition for character 2. Each of these restricts 1 degree of freedom. A likelihood ratio test of these assertions could be done individually or simultaneously. The latter test has 2 degrees of freedom.

Pagel's framework provides a straightforward test of the independence of change in two characters. It can be extended to more elaborate forms of dependence. If one is provided with an unrooted tree, it may be necessary to constrain the two-character model of change to be reversible, so as to avoid having to know where the root is. This can be done by adding only one constraint, which is that

$$\frac{\gamma_0\beta_0}{\delta_0\alpha_0} = \frac{\gamma_1\beta_1}{\delta_1\alpha_1} \tag{25.5}$$

Advantages of the likelihood approach

Pagel's discrete characters comparative method has some important advantages over Maddison's and Ridley's parsimony-based approaches. It takes branch lengths into account, which the methods of Maddison and Ridley cannot (as Maddison pointed out in introducing his method). When some of the branches in the tree are quite short, it will automatically adjust for this and not assume that there could be large amounts of change in those parts of the tree. When a branch is very long, it will treat the evidence from the groups connected to its opposite ends as relatively independent. Another disadvantage of the parsimony-based approaches is that they reconstruct the placement of changes and then treat those events as observations. Any time we find ourselves using something that is an estimate, and using it as a definite observation, we should be suspicious. In such cases the errors that may arise from the uncertainty of the reconstruction are not taken into account in the analysis.

All of these methods suffer from a common limitation—they use a rather naive model of character change. Populations do not make instantaneous changes from one state to another. Having arrived at state 1, they may be more likely to revert to state 0 soon, but less likely later. None of these are taken into account in parsimony or likelihood uses of a simple model of change between two states. We have seen in Chapter 24 that a model with underlying quantitative characters and a threshold can come closer to reality. It allows for polymorphism within species and for different probabilities of reversion soon after a change and later. It is also easier to generalize to covariances of change among multiple characters than the 0/1 stochastic model. It will be important to develop a comparative method that uses the threshold model.

Molecular applications

Molecular evolutionists frequently want to know whether substitutions at different sites are correlated. This might help reconstruct protein structure so as to place correlated sites near each other in space. The great difficulty with using discrete-characters comparative methods for this is that there are 20 amino acids, and many possible pairs of sites to examine, so that there are far too many possible parameters. Any successful application of these methods would necessarily involve constraining parameters and examining sets of pairs of amino acids that could all be

neighbors. Studies of correlated evolution of amino acid positions in a protein have often used nonphylogenetic measures of correlation. Wollenberg and Atchley (2000) have done so, but have used a parametric bootstrapping approach to simulate data on phylogenies to see how much of the correlations could be coming from the phylogeny. We have already seen, in Chapter 13, models of change that maintain pairing of sites in RNA structures.