

"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

Integrative Biology 200

University of California, Berkeley

Spring 2014

D.D. Ackerly

March 5, 2014

Reading:

Felsenstein J. 1985. Phylogenies and the comparative method. *Amer. Nat.* 125:1-15

Oakley TH, Cunningham CW. 2000. Independent contrasts succeed where ancestral reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54:397-405

I. Continuous characters – ancestral states

Many traits of interest are measured on **continuous or metric scales** – size and shape, physiological rates, etc. Continuous traits are often useful for species identification and taxonomic descriptions; historically, they were also used in phylogenetic analysis through the use of clustering algorithms that can group taxa based on multivariate phenetic similarity. With the advent of cladistics, reconstruction has shifted entirely to discrete trait analysis.

In principal, ordinal traits take on discrete states while continuous traits are real numbers. In reality, there is a continuum from ordered discrete traits with only a few states (1,2,3,4,5 petals) to those with enough states that we may treat them as continuous (dozens to hundreds). It is also possible to 'discretize' a continuous trait at real or arbitrary breaks in the distributions across species.

Distributions of continuous traits may take on any arbitrary shape, but certain distributions occur repeatedly, possibly reflecting underlying 'natural' processes.

Normal distribution: sum of many small additive effects

Exponential distribution: product of many small multiplicative effects

Poisson distribution: frequencies of rare events in discrete intervals

etc.

This becomes important when we consider whether ancestral reconstruction of continuous traits should reflect an underlying evolutionary model of the process that describes or dictates trait evolution. Traits may be transformed to better meet an appropriate distribution (e.g. log-transform).

I.B. Parsimony methods for ancestral states

1. Linear or Wagner Parsimony: minimize the sum of absolute or linear changes along each branch (analogous to normal parsimony for discrete traits). The ancestral value at each node will be the *median* of the three values around it (two child nodes, one parent node). The root is a special case, where it will be the median of the two child nodes, as there is no parent node.

2. Squared Change Parsimony: minimize the sum of squared changes along each branch. The ancestral value at each node will be the *mean* of the three values around it (two child nodes, one parent node). Weighted SCP can also be calculated, where the change along each branch is

divided by its branch length before summing – a given change on a long branch is penalized less, and values at any given node will be most similar to adjacent nodes connected by short branches.

I.C. Brownian motion and maximum likelihood ancestral states

Brownian motion (BM) is the term for a random walk in a continuous valued variable. If a trait was determined by multiple, independent additive factors of small effect, and if each factor was mutating or changing at random (e.g., by drift), then the character change would constitute BM. Brownian motion is the starting point for discussions of continuous character evolution, for its simplicity and its close ties to parametric statistics based on normal distributions.

In Brownian motion the size of each step can be drawn from any distribution, and by the Central Limit Theorem the outcome over large number of steps will be a normal distribution. For analytic purposes, it is useful to also use a normal distribution for the individual steps, with mean = 0 (no trend) and variance s^2 (= standard deviation s), where each step is a unit of time. When we consider Brownian motion as a process, this variance is viewed as a rate parameter, β . One of the fundamental principles of probability theory is that the variance of the sum of two random processes is the sum of their variances. In other words, if the variance of a brownian motion process is β after one time step, it will be $\beta + \beta = 2\beta$ after two time steps and $t\beta$ after t times steps. So the variance increases linearly with time.

If you apply BM to a large number of independent random walks, with time = t along each walk, then you can probably see that the variance of the resulting values at the tips of the walks will be $t\beta$. What is less intuitive for most of us (if you are not used to statistical thinking) is that a single value resulting from a random walk also has a variance that refers to the underlying (and unobserved) distribution from which that value has been drawn.

Ancestral states: Now we apply the same principles to solve for ancestral states under ML and Brownian motion, treating each ancestral state as the ML solution for a local BM process derived from that node, and finding the set of ancestral states that maximizes the likelihood over the entire tree. The branch lengths are key now, as the overall s^2 value at each node is proportional to the BM rate parameter times the branch lengths.

Local likelihood solution: The ML reconstruction of ancestral states can be calculated as a local ML solution, based only on the trait data of tips descended from a node. This amounts to a recursive averaging process down the tree, except that at each node you calculate the weighted average of the two daughter nodes, weighted by the inverse of the square root of the branch length (more on that later). (The everyday mean that we calculate for data sets is also a maximum likelihood solution for that value that minimizes the squared changes between the mean and the data points, i.e. minimizes the variance around the mean).

Global likelihood solution: The global ML solution uses information from the entire tree, including descendent taxa and all sister clades at and above (towards the root) a given node. Since ML is minimizing the sum of squared changes, the ancestral states found under the global likelihood solution are equivalent to the results under squared change parsimony. When a solution is found, a BM rate parameter is also calculated, based on the variance of the normal distribution per unit branch length (see Schluter al. 1997, top right of p. 1701). *The big difference*

between squared change parsimony and ML is that ML techniques use branch lengths and can provide confidence intervals on the ancestors! Given the BM rate parameter, we can calculate a distribution of ancestral states (i.e. support limits) that are consistent with the observed data. The rather troubling result of much work in this area is that sometimes the error bars exceed the range of values observed in the terminal taxa (in other words the ancestor could be anywhere in the range of present-day trait values, or even outside it!). See Schluter et al. 1997 Fig. 7 and Fig. 8 (error bars in Fig. 8 don't seem to show up in the pdf - check the printed journal).

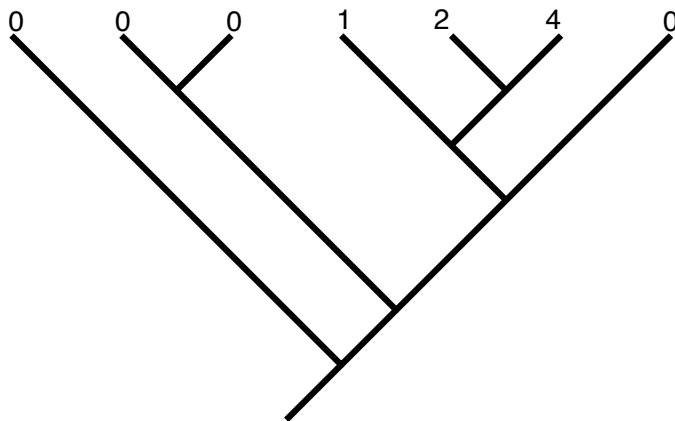
I.D. Do ancestral state reconstructions work?

Three attempts to test methods: Oakley and Cunningham 2000, Polly 2001, Webster and Purvis 2002. First used experimental bacteriophage lineages, directly examining properties of ancestral populations. Other two used comparisons with fossils. Polly found that fossil values were quite close to ancestral estimates, and closer than might be expected based on the confidence limits. Oakley and Webster papers both conclude that ancestral state methods perform very poorly—in both cases, there were significant evolutionary trends across the entire clade that caused the problems.

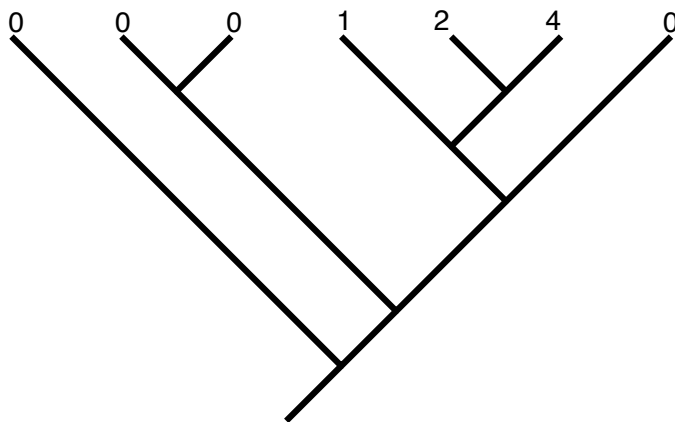
I.E. Citations on ancestral states for continuous characters

- Ackerly DD, Reich PB. 1999. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. *Amer. J. Bot.* 86:1272-1281
- Cunningham CW, Omland K, Oakley TH. 1998. Reconstructing ancestral character states: a critical reappraisal. *TREE* 13:361-366
- Losos J. 1999. Commentaries : Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. *Animal Behaviour* 58:1319-1324
- Maddison W. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Syst. Zool.* 40:304-314
- Martins EP. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. *Syst. Biol.* 48:642-650
- Oakley TH, Cunningham CW. 2000. Independent contrasts succeed where ancestral reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54:397-405
- Polly PD. 2001. Paleontology and the comparative method: ancestral node reconstructions versus observed node values. *Amer. Nat.* 157:596-609
- Schluter D, Price T, Mooers A, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699-1711
- Schultz T, Cocroft R, Churchill G. 1996. The reconstruction of ancestral character states. *Evolution* 50:504-511
- Swofford DL, Maddison WP. 1987. Reconstructing ancestral states under Wagner parsimony. *Math. Biosci.* 87:199-229
- Webster AJ, Purvis A. 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proc. Roy. Soc. London Ser. B* 269:143-149

Linear Parsimony



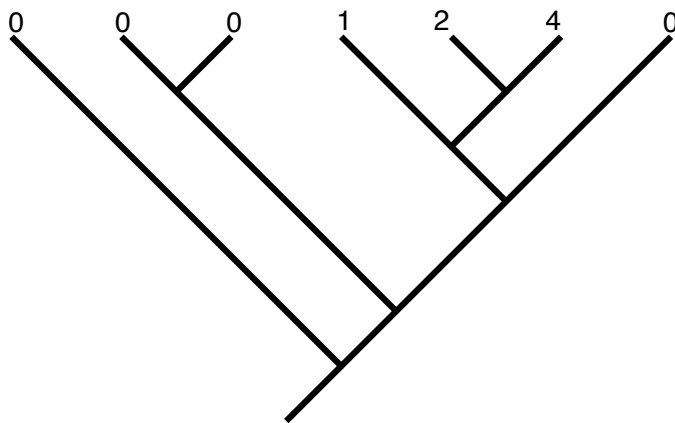
Squared change parsimony (no BL, or equivalently all BL = 1)



```
## 7 taxon example, ace solution
require(ape)
tree =
"(t1:5,((t2:1,t3:1):3,((t4:2,(t5:1
,t6:1):1):1,t7:3):1):1);"

phy = read.tree(text=tree)
plot(phy)
x = c(0,0,0,1,2,4,0)
ace(x,phy)
```

ML under Brownian motion (with BL as drawn)



TEST YOURSELF: Under ML ancestral state reconstructions, do you think the root value will be lower on tree 2 (all BL = 1) or on tree 3 (BL as drawn).

II. Continuous characters – Independent contrasts

II.A. One of the most common questions in comparative biology is whether two traits are correlated or allometrically proportional to one another. Is tree height correlated with seed size? Does brain size increase allometrically (disproportionately) with body size? Questions of trait correlations usually address an underlying adaptive hypothesis. E.g., If larger animals require more food, and the rate of energy supply to ecosystems is relatively constant (e.g. solar energy inputs), then larger animals should require larger home range and have lower population densities.

As an evolutionary question, these hypotheses can be addressed at many levels. For a population geneticist, it would be ideal to find individual genotypes that exhibit variation in the different traits, and test for genetic correlations and selection on different trait combinations. This is usually extremely difficult and requires very large sample sizes. It could also be addressed by comparing different populations within a species, which is a little easier if multiple populations occupying different conditions can be identified.

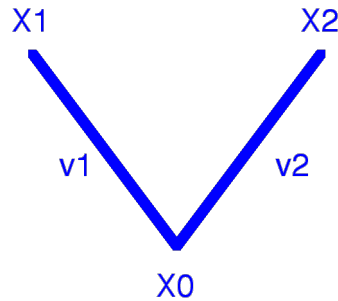
However, most comparative tests are conducted using species as the units of analysis, representing the outcomes of the evolutionary process. The underlying logic is that if selection has acted on two traits together, such that evolutionary changes in one trait are correlated with changes in another trait, and if the selective history is different for different species, leading to different outcomes, then we can use correlations of species trait values to test the hypothesis of correlated evolution between the two traits.

Numerous studies have conducted such tests over the past decades, many of them based on simple correlation statistics using species values as data points. This is sometimes called the cross-species, ahistorical, or TIP correlation (based on the species values at the tips of the phylogeny). Since Darwin (see Ridley 1992), many researchers recognized that comparisons of closely related species were particularly valuable. For example, Salisbury (1942) compared the seed size of congeneric species found in open vs. closed habitats, to test the hypothesis that shaded habitats favored larger seeds. Why should it be important to use congeneric species pairs? Presumably there are some other factors shaping seed size evolution, such as breeding system, dispersal mode, or overall plant stature. If these factors are shared by species within a genus, then the use of congeneric pairs provides a way of holding these factors constant within each comparison, while testing the adaptive hypothesis based on the divergence within each pair. Note the underlying assumption: there are some, possibly unmeasured, factors which exhibit strong phylogenetic signal such that they are shared among relatives, and these can be held constant while testing the role of another factor that exhibits less signal, and has diverged between close relatives. Statistically, this is essentially a paired t-test.

II.B. Independent contrasts: In 1985, with the rise of phylogenetics and comparative biology, Felsenstein introduced a generalization of this paired comparisons method, based on the divergences that have occurred at each bifurcation in a phylogenetic tree. These divergences represent contrasts (differences) between the trait values. If we assume that traits evolve independently in each lineage, following speciation, then the trait divergences that occur at one node are independent of the divergences at other nodes. Hence the name, phylogenetic

independent contrasts (often abbreviated PICs).

The independent contrasts method is derived from the Brownian Motion model. Let's start with a single divergence:



Assume a trait of interest starts with value X_0 at the ancestor. After speciation, the trait evolves independently along each branch, with a Brownian Motion parameter s^2 (the variance of expected character change per unit branch length). If the length of the subtending branches are v_1 and v_2 , then the changes that occur on each branch will be drawn from normal distributions with mean = 0 and variance = $s^2 \cdot v_1$ and $s^2 \cdot v_2$, respectively. The resulting trait values will be:

$$X1 = X0 + N\{0, s^2 \cdot v1\}$$

$$X2 = X0 + N\{0, s^2 \cdot v2\}$$

Now we want to know the expected distribution of the difference or contrast in the resulting trait values. We can take advantage of the fact that the variance of the sum or difference of two random variates is the sum of the variances. Therefore,

$$C_U = (X1 - X2) = N\{0, s^2 (v1+v2)\}$$

In other words, the mean value of the contrast will be zero, on average. And the variance around this mean will be the rate parameter s^2 , times the sum of the branch lengths connecting the two taxa.

This value is termed the unstandardized contrast, because it is not standardized for the branch lengths. If two taxa are connected by long branches, we expect larger contrasts, and if they share a more recent common ancestor, we expect smaller contrasts. Statistical theory is developed on the assumption that the variates are independent, and are drawn from normal distributions with equal variance or standard deviation. To achieve this, we now standardize the contrast, by dividing by the standard deviation of the expected distribution (since the rate parameter β is constant over the tree, and only relative values of the contrasts are important, β or s^2 is usually dropped at this step and the contrast is just standardized by the summed branch lengths):

$$C_S = \frac{C_U}{\sqrt{v_1 + v_2}}$$

In theory, if the trait has evolved following Brownian Motion, the values of C_S now represent a set of independent, normal variates drawn from distributions of equal variance, and these meet the assumptions for use in standard statistical tests such as correlations, regressions, and analysis of variance.

The calculation of contrasts, as illustrated above, require trait values at each pair of nodes sharing a common ancestor. Starting at the tips, these are the species values provided by the data set. Moving down the tree, however, we need to calculate a set of estimated internal values to

obtain the appropriate contrasts. These are based on the local (top-down) maximum likelihood algorithm—essentially a weighted average of the nodes above, using the inverse of the branch lengths as weighting factors:

$$X_0 = \frac{\left(\frac{1}{v_1}\right)X_1 + \left(\frac{1}{v_2}\right)X_2}{\frac{1}{v_1} + \frac{1}{v_2}}$$

As a final step, the branch length below node 0 is lengthened, reflecting the increased uncertainty associated with deeper nodes (since a longer branch results in higher variance):

$$v'_0 = v_0 + v_1 v_2 (v_1 + v_2)$$

With the new internal node values calculated, the calculation of contrasts proceeds down the tree. These internal values are essentially the local (not global) maximum likelihood ancestral states based only on information from descendents, not the totality of information obtained when considering both descendents and other taxa in the tree, as in the full ancestral state algorithms discussed above.

II.C. Correlations of independent contrasts: Before we conduct a statistical test with independent contrasts, note that they have one unusual property. Each contrast is based on subtraction of one value from another. Clearly, the direction of subtraction is arbitrary, as long as it is kept the same for all traits at each node. As a result, each contrast has a mirror image of the opposite sign. So clearly the average value of all contrasts must be zero, since each one could be flipped around. As a result, all correlations and regression analyses must be calculated through the origin, and anovas would have to be conducted without a grand mean term. The formula for the correlation coefficient through the origin is a bit different than the familiar one in a stats textbook:

$$r_{xy} = \frac{\sum C_x C_y}{\left\{ \sum C_x^2 \sum C_y^2 \right\}^{1/2}}$$

where C_x and C_y are the standardized contrasts for traits x and y . See Garland et al. 1992 for additional discussion.

Citations below provide examples of independent contrast correlations and the interpretations that can be drawn from comparisons of TIPs and PICs analyses.

Addendum – the mean as a ML solution:

Whether you know it or not, we all solve a maximum likelihood (ML) problem on a daily basis when we calculate the mean for a set of numbers. The mean of X (a set of numbers) is the sum of X divided by N, the number of values in X, right? Yes. Alternatively, the mean of X is the ML solution for the starting point of N random walks that end with values X. This can be solved from the following steps:

- 1) From the central limit theorem, we know that random walks generate values drawn from a normal distribution, with mean μ and variance σ^2 .
- 2) The probability of each value of X, under a normal distribution is:

$$P(x) = \frac{\exp\left[-(x - \mu)^2 / 2\sigma^2\right]}{\sigma\sqrt{2\pi}}$$

- 3) The ML solution for μ and σ^2 are the values that maximize their cumulative probability over all values of x, and the cumulative probability is the product of the individual probabilities. A product of a series of values for P(x) looks pretty nasty, so instead let's take the sum of the log of P(x) (because the log of a product is the sum of the logs):

$$\log(P(x)) = \frac{-(x - \mu)^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})$$

- 4) To maximize this, we can ignore the denominator ($2\sigma^2$) and the second term, since they will be constants. And if we are maximizing the sum of negative terms, we can instead minimize the sum of the positive terms. So the mean of X is that value which minimizes:

$$\sum (x - \mu)^2$$

Look familiar!? It's the sum of squares of X. And now there's some magic, and the sum of squares is also how we calculate σ^2 , but we won't try and derive that as a ML problem here.

- 5) Try this R script to solve for the mean of a set of numbers by finding the minimum of the sum of squares:

```
## enter a set of numbers in xx
xx = c(1,2,4,5)

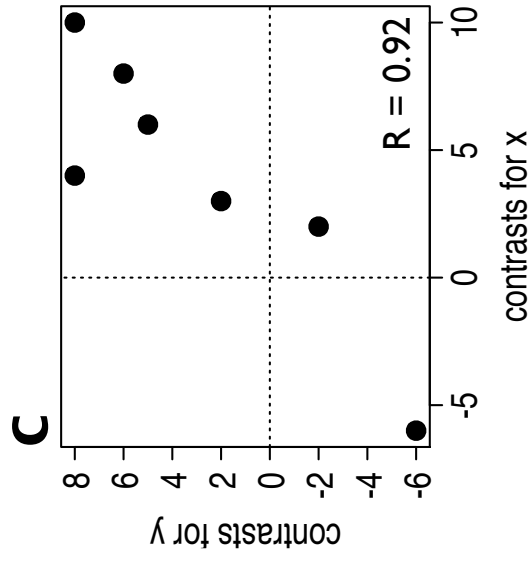
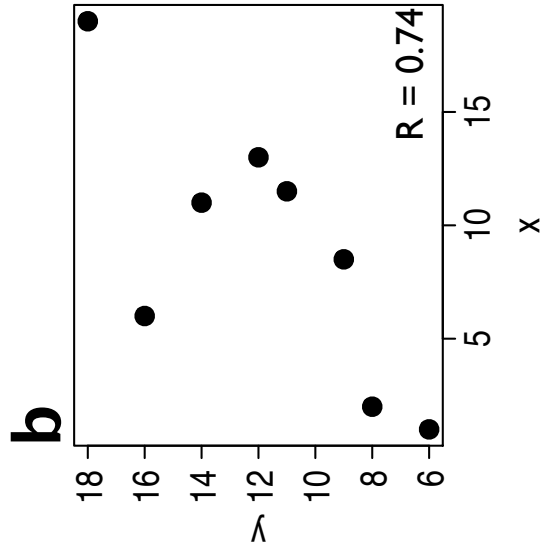
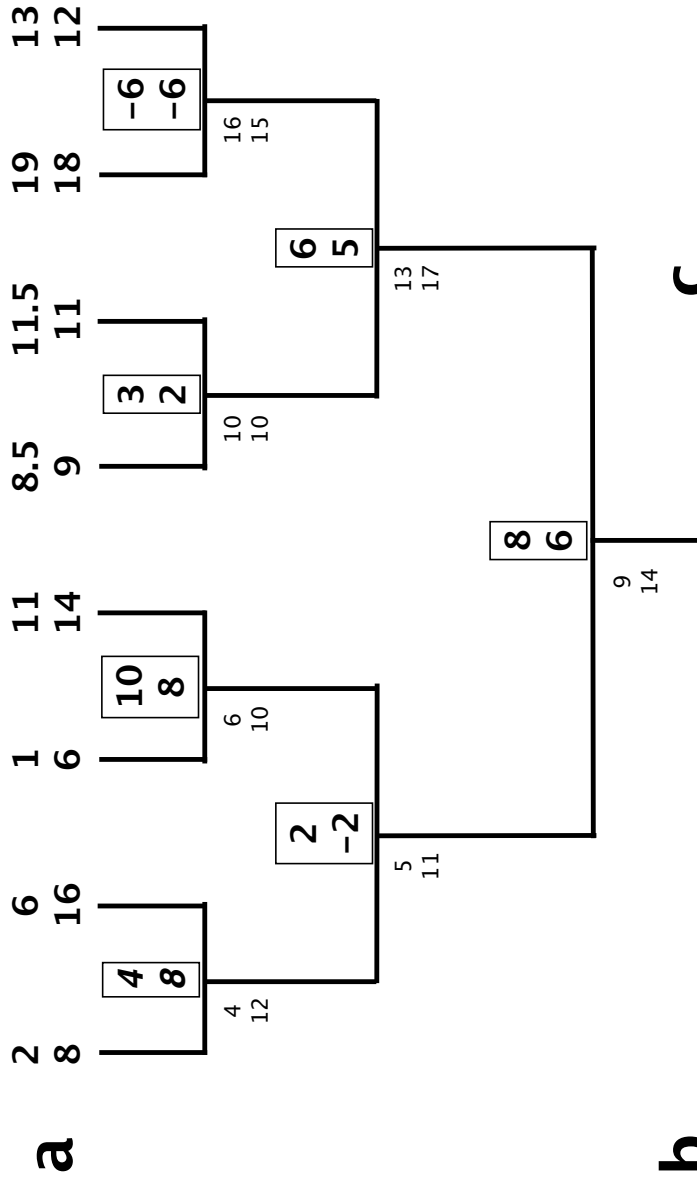
## create a sequence of candidate values for the mean of xx
xu = seq(1,5,by=0.1)

## create a variable to hold the sum of squares
lxu = rep(NA,length(xu))

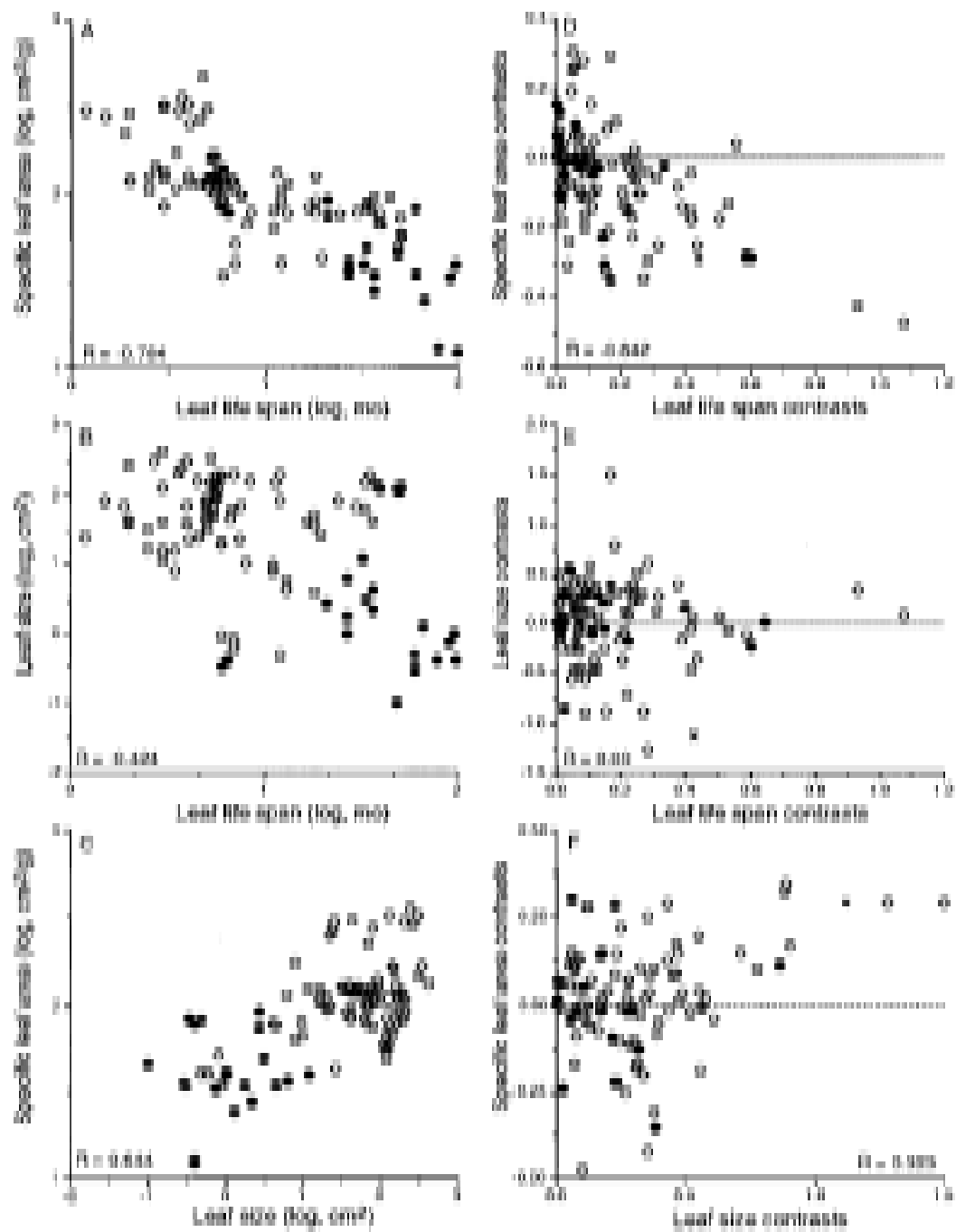
## loop through xu and calculate the likelihood score for each candidate value
## as sum of squared deviations of xx from xu
for (i in 1:length(xu)) lxu[i] = sum((xx-xu[i])^2)

## plot the likelihood score vs. xu
plot(xu,lxu)

## find the minimum; print it out and compare it to the mean of xx
minxu = xu[which(lxu==min(lxu))]
print(minxu)print(mean(xx))
```

Ackerly and Reich, 1999, Amer J Bot



open circles: angiosperms
 closed circles: conifers
 cross: basal contrast between angios and conifers

II.D. Citations for independent contrasts

- Abouheif E. 1998. Random trees and the comparative method: a cautionary tale. *Evolution* 52:1197-1204
- Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* 1:895-909
- Ackerly DD, Reich PB. 1999. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. *Amer. J. Bot.* 86:1272-1281
- Ackerly DD. 2000. Taxon sampling, correlated evolution and independent contrasts. *Evolution* 54:1480-1492
- Armstrong D, Westoby M. 1993. Seedlings from large seeds tolerate defoliation better: a test using phylogenetically independent contrasts. *Ecology* 74:1092-1100
- Bauwens D, Garland T, Jr., Castilla A, Van Damme R. 1995. Evolution of sprint speed in Lacertid lizards: morphological, physiological and behavioral covariation. *Evolution* 49:848-863
- Burt A. 1989. Comparative methods using phylogenetically independent contrasts. *OxfSurvEvolBiol.* 6:33-53
- Diaz-Uriarte R, Garland T, Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45:27-47
- Felsenstein J. 1985. Phylogenies and the comparative method. *Amer. Nat.* 125:1-15
- Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. *Amer. Nat.* 171:713-725
- Freckleton RP. 2000. Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. *Func. Ecol.* 14:129-134
- Garland T, Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Amer. Nat.* 140:509-519
- Garland Jr, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18-32.
- McPeck, M.A. (1995) Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. *Amer Nat*, **145**, 686-703.
- Oakley TH, Cunningham CW. 2000. Independent contrasts succeed where ancestral reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54:397-405
- Pagel M. 1993. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. *JTheorBiol.* 164:191-205
- Price T. 1997. Correlated evolution and independent contrasts. *Phil. Trans. Roy. Soc. Lond. Ser. B* 352:519-529
- Purvis A, Rambaut A. 1995. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Bios.* 11:247-251
- Ricklefs RE, Starck JM. 1996. Applications of phylogenetically independent contrasts: a mixed progress report. *Oikos* 77:167-172
- Ridley M. 1992. Darwin sound on comparative method. *TREE* 7:37
- Salisbury E. 1942. The reproductive capacity of plants. Bell, London
- Starck J. 1998. Non-independence of data in biological comparisons a critical appraisal of current concepts assumptions and solutions. *Theory Biosci.* 117:109-138

R script to generate some random walks and their variance over time:

```
N = 10 #number of walks
L = 10000 #length of each walk
x = replicate(N,cumsum(rnorm(L)))
xr = range(x)
plot(x[,1],type='n',ylim=xr)
for (i in 1:10) for (j in 1:9999) lines(c(j,j+1),c(x[j,i],x[j+1,i]))

## plot variance over time of the outcome of the N walks
vx = apply(x,1,var)
plot(vx,xlab='time',ylab='variance of x',type='l')
```