Cladistics (1991) 7:81–91

EXCESS HOMOPLASY RATIOS

James S. Farris[1]

[1] *Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794, U.S.A. and Department of Entomology, American Museum of Natural History,, Central Park West at 79th Street, New York, NY 10024, U.S.A. and Department of Biology, City College, City University of New York, U.S.A.*

Received for publication 31 October 1990; accepted 2 November 1990

Archie (1990) prefers his "homoplasy excess ratio" HER to Farris' (1989)[1] ensemble retention index R. HER, he writes, lacks R's defects: R's minimum is not zero, and varies with number of terminals or characters.

HER has those defects. Archie has misunderstood the permutation method on which HER rests, and mistaken the properties of R as well. After explaining those points, I describe other permutation-based indices and briefly consider their utility. These provide relative measures of congruence among the characters of the data; one corresponds to a significance test.

Information

I begin by reviewing calculation of R and HER. I have extended my 1989 notation to accommodate permutation methods. Formulae have been restated to reduce algebraic terms.

Data matrix $\mathbf{X}$ has t columns (terminals) and n rows (characters). T is any tree for the terminals of $\mathbf{X}$, and U is the entirely unresolved tree with those terminals. Character k requires amount $h(k, T)$ of homoplasy (extra steps) on tree T; this cannot be negative. $H(\mathbf{X}, T)$ is $\sum h(k, T)$ over the characters of $\mathbf{X}$. Define $i(k)$ as $h(k, U)$, and note that $h(k, T) \leq i(k)$. Likewise, $H(\mathbf{X}, T) \leq I(\mathbf{X})$, which is $H(\mathbf{X}, U) = \sum h(k, U) = \sum i(k)$.

A character that is phylogenetically uninformative (comprises only autapomorphies and symplesiomorphies) for the terminals of $\mathbf{X}$ requires no homoplasy on any tree for those terminals, and conversely. Character k is informative just when $i(k) > 0$; $i(k)$ is called the amount of informative variation in k. $I(\mathbf{X})$ is the total informative variation in $\mathbf{X}$, and $N(\mathbf{X})$ is the number of informative characters in $\mathbf{X}$. The data $\mathbf{X}$ are called uninformative when $I(\mathbf{X}) = N(\mathbf{X}) = 0$.

Two characters of $\mathbf{X}$ are incongruent just when no tree T can be free of homoplasy in both. Consequently, an uninformative character cannot be incongruent with any character. Define $B(\mathbf{X})$ as the minimum of $H(\mathbf{X}, .)$ among possible trees. If tree V is most parsimonious for $\mathbf{X}$, then $B(\mathbf{X}) = H(\mathbf{X}, V)$. The data $\mathbf{X}$ show incongruence just when $B(\mathbf{X}) > 0$, and this can happen only if $N(\mathbf{X}) > 1$.

[1] I had introduced the retention index at the 1988 meeting of this Society (cf. Seberg, 1989), having already implemented that measure in Hennig86. I thank W. Day for his enthusiastic interest on that occasion.

0748-3007/91/010081 + 11$03.00/0

© 1991 The Willi Hennig Society

Retention

$I(\mathbf{X}) - H(\mathbf{X}, T)$ is the amount of variation in data $\mathbf{X}$ that is applicable as synapomorphy on tree T. The ensemble retention index $R(\mathbf{X}, T)$ expresses this as a fraction of the total $I(\mathbf{X})$ of informative variation in $\mathbf{X}$. The ensemble is the suite of characters. That fraction is viewed as retained in that it comprises similarity not dismissed as homoplasy on the tree. $R(\mathbf{X}, U) = 0$ because no synapomorphy can be applied on the unresolved tree; applying a synapomorphy resolves a group.

For informative data, $R(\mathbf{X}, T) = 1 - H(\mathbf{X}, T)/I(\mathbf{X})$. $R(\mathbf{X}, U) = 1 - I(\mathbf{X})/I(\mathbf{X}) = 0$. For uninformative data, R is $1 - 0/0$ for any tree. Distinguish R for a most parsimonious tree as $R^*(\mathbf{X}) = 1 - B(\mathbf{X})/I(\mathbf{X})$. The "homoplasy excess ratio maximum" HERM as described by Archie (1989: 258–260) is equivalent to R^*.

For examples use data matrices ONE and TWO.

Row	ONE terminals					TWO terminals				
	1	2	3	4	5	1	2	3	4	5
1	1	1	0	0	0	1	0	0	0	0
2	1	0	1	0	0	0	0	1	1	0
3	1	0	0	1	0	0	1	1	0	0
4	1	0	0	0	1	1	1	0	0	0

Each of the four rows is a character. Terminal 5 is the outgroup. All characters are informative in both matrices: $N(\text{ONE}) = N(\text{TWO}) = 4$. Each character of either matrix requires one extra step on the unresolved tree, so that $I(\text{ONE}) = I(\text{TWO}) = 4$. In general $I(\mathbf{X})$ need not equal $N(\mathbf{X})$.

In TWO, character 3 is incongruent with any other, but the rest are mutually congruent. The most parsimonious tree is ((1 2)(3 4)5), which requires one extra step. Three of the four characters distinguish monophyletic groups, while the apomorphies in character 3 are parallelisms. $B(\text{TWO})$ is 1, and $R^*(\text{TWO}) = 1 - 1/4 = 0.75$. Seventy-five per cent of the informative variation in the data is applicable as synapomorphy on the best tree; 25% is dismissed as homoplasy. For the tree T = (1(2 3)4 5)—picked to match character 3—$H(\text{TWO}, T)$ is 3, and $R(\text{TWO}, T) = 0.25$.

In ONE any two distinct characters are incongruent. This naturally produces homoplasy on a most parsimonious tree; the best a tree can do is to match one of the characters, so that $B(\text{ONE}) = 3$. Incongruence may well also increase the number of most parsimonious trees; here there are 12 (I used Hennig86). These fall into four types, corresponding to the four ways of placing terminal 1 with another.

$R^*(\text{ONE}) = 1 - 3/4 = 0.25$. Twenty-five per cent of informative variation is indeed retained on any one of the 12 trees, but that is too charitable to the trees collectively. In practice, conclusions are often based on a consensus tree, and then one finds R for the consensus. In this case the consensus is unresolved, and requires four extra steps. $R(\text{ONE}, U)$ is $1 - 4/4 = 0$, indicating that none of the informative variation is retained as synapomorphy.

Permutation

HER is based on permutations of the data matrix, and I describe these first. Permutation is a well-known way of obtaining the distribution of test statistics, and my comments throughout reflect common statistical ideas.

A permutation of a row (character) of a data matrix is just one of the $t!$ ways of arranging the t existing entries of the row. A *matrix permutation* of $\mathbf{X}$ is determined by selecting an ordinary permutation separately for each row of $\mathbf{X}$. ONE and TWO are matrix permutations of each other. $P(\mathbf{X})$, the *permutation class* of $\mathbf{X}$, is the suite of matrices that can be obtained by matrix permutations of $\mathbf{X}$; this includes $\mathbf{X}$ itself.

A *randomization* of $\mathbf{X}$ is a matrix $\mathbf{Y}$ selected at random from $P(\mathbf{X})$; all members of $P(\mathbf{X})$ have the same probability $p(\mathbf{X})$ of being chosen. If another data matrix $\mathbf{X}'$ belongs to $P(\mathbf{X})$, then $P(\mathbf{X}') = P(\mathbf{X})$, and any randomization of $\mathbf{X}$ is equally well a randomization of $\mathbf{X}'$. $\mathbf{Y}$ would be more precisely called a random (i.e., randomly chosen) matrix permutation of $\mathbf{X}$; "randomization" is used for brevity.

Considered simply as a matrix, $\mathbf{X}$ or $\mathbf{X}'$ is identical to some $\mathbf{Y}$. But $\mathbf{Y}$s, being randomizations, have a particular probability distribution, uniform on $P(\mathbf{X})$. $\mathbf{X}$ or $\mathbf{X}'$ is a data matrix. Neither of their distributions—if they have distributions—need be uniform on $P(\mathbf{X})$.

The $t!$ ways of arranging the entries of a row need not all yield visibly distinct rows. The number of such rows is needed to find the number of matrices in a permutation class.

Use $f(k,z)$ for the frequency of condition z in character k, that is, the number of terminals showing that condition. A condition is usually just a state, but might be a missing value or a polymorphism. Write $e(k)$ for the product of $f(k,z)!$ over possible conditions of k. The number of distinct character state (condition) distributions that have the given frequencies in character k is then $d(k) = t!/e(k)$, the mutinomial coefficient for those frequencies.

In character 1 of ONE $f(1,0) = 3$, while $f(1,1) = 2$. The number of visibly distinct permutations of that character is $d(1) = 5!/(3!2!) = 10$, the number of ways of choosing two objects from among five. The other characters of ONE happen to have the same state frequencies, so that $d(k)$ is the same for each. In general $d(k)$ can vary among characters. It can also increase rapidly with number of terminals. If some matrix has $t = 20, f(1,0) = f(1,1) = 10$, then $d(1)$ is $20!/(10!10!)$, or 184 756.

A permutation is selected separately for each character, so that the number $D(\mathbf{X})$ of distinct matrix permutations in $P(\mathbf{X})$ is the product of $d(k)$ over the characters of $\mathbf{X}$. ONE has four characters with 10 distinct permutations apiece, so that $D(\text{ONE}) = 10^4$. When randomly selected, each matrix in $P(\mathbf{X})$ has probability $p(\mathbf{X}) = 1/D(\mathbf{X})$; $p(\text{ONE}) = 10^{-4}$. $P(\text{TWO})$ and $p(\text{TWO})$ are just the same, ONE and TWO being matrix permutations of each other.

HER

HER compares homoplasy for data matrix $\mathbf{X}$ to that for randomizations $\mathbf{Y}$ of $\mathbf{X}$. As is usually done, homoplasy for any one matrix is assessed as $B(\mathbf{X})$ or $B(\mathbf{Y})$. The $B(\mathbf{Y})$s are summarized by their mean.

$B(\mathbf{Y})$s can vary because permutation can change congruence among characters, as is evident on comparing ONE and TWO. Uninformative characters can be ignored for purposes of permutation; they cannot affect B. Notice that permutation cannot change the amount of informative variation in any character. Thus for any $\mathbf{Y}$, $M(\mathbf{Y}) = M(\mathbf{X})$, and $I(\mathbf{Y}) = I(\mathbf{X})$.

Define $A(\mathbf{X})$ as $\sum p(\mathbf{X})B(\mathbf{Y})$ over $\mathbf{Y}$ in $P(\mathbf{X})$. This is the expectation (population mean) of $B(\mathbf{Y})$ under equiprobability of randomizations. Then $\text{Her}(\mathbf{X}) = 1 - B(\mathbf{X})/A(\mathbf{X})$ (compare Archie, 1989: 258). Within a permutation class, HER is a fixed linear function of B, so that $\text{HER}(\mathbf{Y})$ has expectation $1 - A(\mathbf{X})/A(\mathbf{X}) = 0$.

If $A(\mathbf{X}) = 0$, HER($\mathbf{X}$) is $1 - 0/0$. As $A(\mathbf{X})$ is the mean of $B(\mathbf{Y})$s, $A(\mathbf{X}) > 0$ only if some $B(\mathbf{Y}) > 0$. Data satisfying the latter condition are called *usefully permutable*. If $\mathcal{N}(\mathbf{X}) < 2$, every $B(\mathbf{Y}) = 0$, so useful permutability requires at least two informative characters.

One must beware HER's name. The "homoplasy excess ratio" increases as the amount $B(\mathbf{X})$ of homoplasy in the data decreases, reaching 100% when the data imply no homoplasy. Archie (1990: 173) clouds interpretation further:

> "The homoplasy excess ratio (HER) appears to be the most appropriate statistic for measuring average levels of homoplasy in an entire data set."

HER($\mathbf{X}$) is certainly not the level of homoplasy in $\mathbf{X}$, although $B(\mathbf{X})/A(\mathbf{X}) = 1 - \text{HER}(\mathbf{X})$ might be. $A(\mathbf{X})$ is averaged over $\mathbf{Y}$s, not over the data $\mathbf{X}$, and $B(\mathbf{X})$ is not an average.

Taking $\mathbf{X}$ = ONE as an example, the exact (population) distribution of $B(\mathbf{Y})$ is found by computing a most parsimonious tree for each of the 10^4 possible $\mathbf{Y}$s, counting the instances of each B value.

$B(\mathbf{Y})$	0	1	2	3
Prob ($B(\mathbf{Y})$)	0.022	0.276	0.690	0.012
HER($\mathbf{Y}$)	1	0.409	−0.182	−0.773

The mean of $B(\mathbf{Y})$ is 1.692, and this is $A(\mathbf{X} = \text{ONE})$. The distribution of $\text{HER}(\mathbf{Y}) = 1 - B(\mathbf{Y})/1.692$ is determined by that of $B(\mathbf{Y})$. HER($\mathbf{Y}$) is naturally negative whenever $B(\mathbf{Y})$ exceeds $A(\mathbf{X})$. The mean of HER($\mathbf{Y}$) is 0, as always (these HERs were rounded to three places).

ONE itself is at the extreme in this case: $B(\text{ONE}) = 3$, $\text{HER}(\text{ONE}) = -0.773$. If $\mathbf{X}$ is instead taken as TWO, the same distribution of $B(\mathbf{Y})$ is obtained. Since $P(\text{TWO}) = P(\text{ONE})$, necessarily $A(\text{TWO}) = A(\text{ONE})$. $B(\text{TWO}) = 1$, and $\text{Her}(\text{TWO}) = 0.409$.

Archie (1989) calculated HER only for data with several characters, but he now suggests (Archie, 1990: 170)

> "Of course, the limiting case would have only a single character used in the analysis for a set of taxa. In such a case HER (discussed below) could be calculated directly and the HERM terminology would be inappropriate."

The last refers to his 1989 use of HERM in a method for approximating HER, but he does not say how the terminology should be changed. In any case, if there is just one character, the data cannot be usefully permutable: $\text{HER} = 1 - 0/0$.

Retention index r can be calculated for a single informative character: $r(k, T) = 1 - h(k, T)/i(k)$. Incidentally, $R(\mathbf{X}, T)$ is not defined as the mean of $r(k, T)$s. This should already be clear, and I mention it only because Archie (1990: 170) devotes most of a page to expressing uncertainty on that point.[2]

Error

When $\mathbf{X}$ is large, evaluation $B(\mathbf{Y})$ for every possible $\mathbf{Y}$ in $P(\mathbf{X})$ is impractical, and $A(\mathbf{X})$ is not calculated exactly. instead, each of several (Archie tried 100)randomly chosen matrix permutations is used to obtain a $B(\mathbf{Y})$, the sample mean of those $B(\mathbf{Y})$s

[2] In that discussion Archie attributes to me a personal communication, of which, however, I have no recollection.

then providing an estimate of $A(\mathbf{X})$. The precision of that estimate is limited by sampling error, just as when any population mean is estimated by a sample mean.

Archie (1990: 171) mistakes the result of sampling error (square brackets in original):

> "HER will always be greater than or equal to 0 (except for sampling error, see Archie [1989b,c])."

In the case just illustrated, HER(ONE) = −0.773 is negative because $B(\text{ONE}) = 3$ exceeds the population mean $A(\text{ONE}) = 1.692$. A sample estimate of $A(\text{ONE})$ might differ somewhat from 1.692, but such sampling error is not the cause of the negative HER. HER(ONE) is −0.773 when $A(\text{ONE})$ is known exactly and sampling error is entirely absent.

Archie (1989: 258) supposed:

> "As the number of steps on the observed minimum-length tree approaches that for phylogenetically random data, HER will approach 0.0. For completely random data, the expected value of HER is 0.0, although HER may take on negative values due to sampling error."

By either completely or phylogenetically random data he meant randomizations, the matrices I have denoted **Y**.

Again it is not sampling error that produces negative HER. In the exact distribution given above for $P(\text{ONE})$, HER(**Y**) is negative with probability 0.702. The median $B(\mathbf{Y}) = 2$ is the mode as well, most likely by far with probability 0.69, and is also the value closest to the population mean 1.692. But when $B(\mathbf{Y}) = 2$, HER(**Y**) = −0.182. Other (and much larger) permutation classes usually show the same kind of skewness, the median $B(\mathbf{Y})$ exceeding the mean. Negative HER might well be called typical of randomizations.

In any usefully permutable case, the expectation— the mean—of HER(**Y**) is indeed 0. But for just that reason, HER(**Y**) can be positive for some **Y**s only if it is negative for others. A variable quantity, that is, cannot very well have its mean equal to its minimum.

Writing SA(**X**) for the sample estimate of $A(\mathbf{X})$, sampling error can cause sample value of $\text{SHER}(\mathbf{X}) = 1 - B(\mathbf{X})/\text{SA}(\mathbf{X})$ to differ from population quantity $\text{HER}(\mathbf{X}) = 1 - B(\mathbf{X})/A(\mathbf{X})$. SHER(**X**) can indeed be negative although HER(**X**) is positive. As $B(\text{TWO}) = 1$ and $A(\text{TWO}) = 1.692$, HER(TWO) = 0.409, but SHER(TWO) would be negative whenever SA(TWO) < 1.

But the opposite can happen, too. From the distribution of $B(\mathbf{Y})$ already given, it is seen that some matrix in the same permutation class, THREE say, has $B(\text{THREE}) = 2$, HER(THREE) = −0.182. SHER(THREE) would be positive whenever SA(THREE) > 2. That is actually somewhat more likely than SA(TWO) < 1, as the population mean 1.692 is closer to 2 than to 1.

And if SA(THREE) < 2, accurately reflecting that the population mean is less than 2, then SHER(THREE) will be negative, accurately reflecting the negative sign of HER(THREE). While sampling error can cause SHER(**X**) to differ from Her(**X**) in sign, it can change the sign either way. It does not in itself favor negative over positive values, and it is certainly not the only cause of negative ones.

Zero

In Archie's (1990: 172) estimation:

> "The most critical point, however, is that the [ensemble consistency index, C], although bounded by 1 (no homoplasy) and 0 (maximum homoplasy), will never achieve its lower bound.... In fact, Farris, new measure, [R], suffers from a similar problem."

C does, of course, have a greatest lower bound; I gave a formula for the latter in my 1989 paper. The sense of his comment is just that that minimum exceeds 0. He goes on (p. 173):

> "[Farris'] implication is that [R times C, RC] will also vary between 0 and 1. However, since [R] will not achieve a value of 0 (its minimum is approximately 0.38), [RC] cannot achieve a value of 0 either. It suffers from the same problems as [C]!"

Taking a greatest lower bound of precisely 0 as the most critical requirement for an index would provide no basis for preferring HER, unless that ratio had the desired minimum. He does hold that HER ≥ 0, "except for sampling error".[3] But HER(ONE) $= -0.773$ when sampling error is entirely eliminated. The unresolved tree has $R = 0$ for ONE, or for any informative data, and the same is consequently true of RC.

R can calculated for any tree, but the figure 0.38 is based just on R values for most parsimonious trees. Archie (1989), who (in present notation) calculated HERM as R^*, argued that HERM ≥ 0.38. In his 1990 paper he simply used the same value for the minimum of R. That plainly does not follow, but 0.38 is not the minimum of R^* either.

Archie found HER and HERM for each of several data matrices with various numbers of terminals and characters. But none of those values approached the minimum of either ratio. Instead he extrapolated from a least-squares regression (p. 267):

> "HER $= -0.6 + 1.581$HERM ... When HER is predicted to be 0.0 (its lower limit), HERM will be approximately 0.38."

To find a HERM from a postulated HER, he should have regressed HERM on HER, not HER on HERM. Extrapolation from a least-squares line is notoriously unreliable. By his equation, the actual HERM(ONE) = 0.25 predicts HER(ONE) to be -0.205—off by a factor of 3.8. On the same line, the actual HER(ONE) $= -0.773$ correspones to HERM(ONE) $= -0.109$. HERM, being R^*, cannot be negative, but certainly none of this shows that $R^* \geq 0.38$.

R^* can be much less than 0.38. Consider data matrix LOW with $N(\text{LOW}) = n = (t - 1) \geq 3$. Each character k has state 1 in terminals 1 and $k + 1$, state 0 in the rest. ONE fits this description for $t = 5$. As in ONE, the characters are all mutually incongruent: $B(\text{LOW}) = t - 2$. Every character requires one extra step on U: $I(\text{LOW}) = t - 1$. Then $R^*(\text{LOW}) = 1 - (t - 2)/(t - 1) = 1/(t - 1)$. For $t = 101$, $R^*(\text{LOW}) = 0.01$. The minimum of R^* among possible data sets is arbitrarily close to 0.

Dependence

Archie (1990: 172) has other criteria:

> "[C's] lower bound is dependent on the number of taxa (t), the number of characters (N), and character state distribution (CSD) among taxa in an analysis, just as is [C] itself. In fact, Farris' new measure, [R], suffers from a similar problem."

By taxa here he means terminals, not inclusive groups. His kind of character state distribution tells how *many* terminals—not *which* terminals—possess each state.

The minimum of R is simply zero for any informative data, but his comment would

[3] In papers presented at the 1990 ICSEB conference and later at the 1990 meeting of the Hennig Society Archie claimed a 0 minimum for HER, again using this as a reason for rejecting R.

apply to R^*. These dependencies can offer grounds for preferring HER only if that ratio lacks them. I have calculated the minimum of HER for several cases.

	$t=4$	$t=5$	$t=6a$	$t=7$	$t=6b$
$\mathcal{N}=2$	−0.5	−0.67+	−0.88+	−1.1+	−0.11+
$\mathcal{N}=3$	−0.8	−0.85	−0.94+	−1.05+	−0.17
$\mathcal{N}=4$	−0.23	−0.77	−0.81	−0.88+	−0.6
$\mathcal{N}=5$	−0.35	−0.27	−0.74	−0.8	−0.45
$\mathcal{N}=6$	−0.45	−0.33	−0.36	−0.74	−0.57

For each column but the last, each informative character has state 1 in two terminals, state 0 in the rest. For the last, there are three terminals in each state. Cases marked "+" require (uninformative) autopomorphies for some terminals to make all terminals distinct. The method used has already been illustrated with ONE.

On comparing the two columns for $t=6$, it is seen that the minimum HER varies with character state frequency when t and $\mathcal{N}$ are fixed. Within any column, it varies with $\mathcal{N}$ when t and frequency are fixed. Within any row, it varies with t and frequency when $\mathcal{N}$ is fixed. The dependence of minimum HER on these factors is complex, and certainly non-linear. It is also certainly real, for all these values are free of sampling error.

Confidence

Archie's (1989: 268) reasons for preferring HER were not restricted to the behavior of minima.

> "HER (or its estimate HERM) appears to be a sensitive and useful indicator of the confidence that an investigator should have in a cladogram based on parsimony since appropriate, intrinsic properties of the data array are incorporated."

It seems that HERM is a fine idea, unless called a retention index. HERM aside, those intrinsic properties—t, $\mathcal{N}$, and the state frequencies—are just what determine a permutation class. The substance of this is simply that matrix permutations are used. By confidence he meant, as he continued (p. 268):

> "Low values of HER (or HERM) indicate that the data differ little from phylogenetically uninformative data.... In contrast, high values of HER indicate that on average the data differ substantially from random and that a significant amount of information is available for constructing phylogenetic hypotheses."

The only difference that $\mathrm{HER}(\mathbf{X}) = 1 - B(\mathbf{X})/A(\mathbf{X})$ can show is between $B(\mathbf{X})$ and $A(\mathbf{X})$. That difference is not number of informative characters; permutation cannot change that number. The difference is in homoplasy as assessed by $B(\mathbf{X})$ and the $B(\mathbf{Y})$s, $A(\mathbf{X})$ being the average of the latter.

By phylogenetically uniformative here Archie meant randomizations, which he also called completely or phylogenetically random. It does no good to have so many names for one thing; that one of them conflicts with existing usage does not help. Those matrices are reasonably called random, but in that they are hardly unique. Calling them random matrix permutations would be much clearer.

A high value of HER does indicate that $B(\mathbf{X})$ and $A(\mathbf{X})$ differ. But then so would a large negative value, and in any case HER is hardly the only way of showing that difference. No doubt the difference is often statistically significant when HER is high, but not always.

If $t = 4$ and $N = 2$, $B(\mathbf{Y})$ and HER($\mathbf{Y}$) are distributed

$B(\mathbf{Y})$	0	1
Prob($B(\mathbf{Y})$)	1/3	2/3
HER($\mathbf{Y}$)	1	−0.5

The highest possible HER is 1. But here that value occurs with probability 1/3 in randomizations; it scarcely shows a statistically significant departure from such randomness. To indicate signifiance, use a significance test.

Congruence

HER does not meet Archie's criteria, but other indices do.

Arranging for a zero minimum is simple in itself. If measure of homoplasy $S \geq 0$ can reach maximum $\mathcal{Z} > 0$, then index $(1 - S/\mathcal{Z}) \geq 0$. The worst-case value $\mathcal{Z}$ depends on what variety of cases is considered. Suppose $H(\mathbf{X}, T)$ is to be compared among possible trees for data $\mathbf{X}$. Then the maximum is $I(\mathbf{X})$, and the index is $R(\mathbf{X}, T) = 1 - H(\mathbf{X}, T)/I(\mathbf{X})$. But rescaling by the maximum would work on anything. R is worthwhile because of its interpretation, applicable synapomorphy as a fraction of informative variation.

Homoplasy $B(\mathbf{X})$ on a most parsimonious tree for data matrix $\mathbf{X}$ reflects incongruence among the characters of $\mathbf{X}$. To compare this among possible matrices of a permutation class, maximize over that class. Define $\mathcal{J}(\mathbf{X})$ as the maximum of $B(\mathbf{Y})$ over $\mathbf{Y}$ in $P(\mathbf{X})$. $\mathcal{J}(\mathbf{X}) > 0$ just when $\mathbf{X}$ is usefully permutable. Then the *permutation congruence index* $K(\mathbf{X}) = 1 - B(\mathbf{X})/\mathcal{J}(\mathbf{X})$ has minimum 0.

$K(\mathbf{X})$ indicates the degree to which informative variation in $\mathbf{X}$ is congruent, but relative to $\mathcal{J}(\mathbf{X})$. Standards not based on permutation might well be devised. That $K(\mathbf{X})$ is well-defined only for usefully permutable matrices is no drawback in an index of congruence. Congruence is of no concern unless there are two or more informative characters.

For examples use values already calculated. $\mathcal{J}$(TWO) is 3, and B(TWO) = 1, so that K(TWO) = 1 − 1/3 = 2/3. $\mathcal{J}$(ONE) is also 3, as is B(ONE), and K(ONE) = 0. That seems to be about the evaluation that ONE deserves. Certainly ONE shows as little congruence among its characters as can be imagined.

Multiplicity of most parsimonious trees, as seen in ONE, also reflects incongruence. This seems to pose no further problem, so far as measuring the amount of incongruence is concerned. The characters that support alternative trees show up as homoplasy on any one of the solutions.

To calculate $\mathcal{J}(\mathbf{X})$ in practice, just permute characters of $\mathbf{X}$ so that resulting matrix $\mathbf{Z}$ requires as much homoplasy $B(\mathbf{Z})$ as possible. This is substantially easier than finding most parsimonious trees for 100 matrices the size of $\mathbf{X}$, as is done for HER($\mathbf{X}$). It can still take work when the number of terminals is large, but then a further saving of effort is available.

$\mathcal{J}(\mathbf{X}) < I(\mathbf{X})$ for any usefully permutable $\mathbf{X}$. As is seen in ONE and LOW, a most parsimonious tree can always retain some informative variation. But the number of possible state distributions per informative character increases rapidly with t. Then for large t, nearly all informative characters can be permuted into mutually incongruent distributions, and $\mathcal{J}(\mathbf{X})/I(\mathbf{X})$ approaches unity $\mathcal{J}$(LOW)/I(LOW), for example, is $(t-2)/(t-1)$, or 0.99 when $t = 101$.

When $\mathcal{J}(\mathbf{X})/I(\mathbf{X})$ is close to unity, $R^*(\mathbf{X}) = 1 - B(\mathbf{X})/I(\mathbf{X})$ will be close to $K(\mathbf{X}) = 1 - B(\mathbf{X})/\mathcal{J}(\mathbf{X})$. That approximation will be better still to the degree that $B(\mathbf{X})/I(\mathbf{X})$ is small; both indices approach unity as $B(\mathbf{X})$ approaches 0. Given reasonably congruent data for a reasonably large number of terminals, then, just use R^* as an index of congruence. That way the interpretation of R as an index of applicable synapomorphy is available as well.

Significance

Since $B(\mathbf{Y})$ has a probability distribution, there is another way of arranging for a zero minimum. Define $Q(\mathbf{X})$ as the probability that $B(\mathbf{Y}) > B(\mathbf{X})$ when $\mathbf{Y}$ is taken at random from $P(\mathbf{X})$. $Q(\mathbf{X})$ decreases as $B(\mathbf{X})$ increases, reaching 0 when $B(\mathbf{X})$ reaches its maximum $\mathcal{J}(\mathbf{X})$. Thus $Q(\mathbf{X})$ provides a permutation-relative index of congruence. $Q(\mathbf{X})$ is called the permutation congruence tail probability.[4]

Use P(ONE) as an example.

$B(\mathbf{X})$	0	1	2	3
Prob$(B(\mathbf{Y}) = B(\mathbf{X}))$	0.022	0.276	0.690	0.012
$Q(\mathbf{X})$	0.978	0.702	0.012	0
$Q'(\mathbf{X})$	1.658	0.526	0.005	0

Q(TWO) = 0.702. Q(ONE) = 0, as one would like. The last row will be explained shortly.

Consider testing the null hypothesis that data matrix $\mathbf{X}$ is randomly drawn from a population of matrices distributed uniformly on $P(\mathbf{X})$, that is, the population of randomizations. When $Q(\mathbf{X})$ is large enough, reject in favor of the alternative that $\mathbf{X}$ comes from a population of matrices generally showing less homoplasy than do randomizations.

$\mathbf{X}$ still belongs to permutation class $P(\mathbf{X})$; the issue is how its population is distributed on $P(\mathbf{X})$. The exact error rate for this one-tailed test is $\alpha = 1 - Q(\mathbf{X})$. For P(ONE), only $Q(\mathbf{X}) = 0.978$ is significant at the conventional 5% level, but matrices showing more homoplasy can be significant when N or t is greater.

HER is no significance test, and indeed gives no indication of the variability of $B(\mathbf{Y})$. Testing aside, if the aim is to compare congruence in $\mathbf{X}$ to that in randomizations, $Q(\mathbf{X})$ seems ideal. It shows the fraction of $\mathbf{Y}$s less congruent than $\mathbf{X}$. HER provides less information, as it depends on the distribution of $B(\mathbf{Y})$ only through the mean.

For a large permutation class, the distribution of $B(\mathbf{Y})$ may have a long left tail. There $Q(\mathbf{X})$ is close to unity, while $B(\mathbf{X})$ and $\alpha = 1 - Q(\mathbf{X})$ change substantially. Then measure congruence as $Q'(\mathbf{X}) = -\log(1 - Q(\mathbf{X}))$, which expresses $1/\alpha$ as a power of 10. If $Q(\mathbf{X}) = 0.999$, $Q'(\mathbf{X}) = 3$. As the distribution of $B(\mathbf{Y})$ is discrete, $Q(\mathbf{X})$ cannot quite reach unity, and $Q'(\mathbf{X})$ is always well-defined. The significance test can be expressed in terms of $Q'(\mathbf{X})$. Reject when $Q'(\mathbf{X}) \geq -\log(0.05) = 1.3$.

Except for small matrices, $Q(\mathbf{X})$ must be estimated from a random sample of $B(\mathbf{Y})$s. If of W $B(\mathbf{Y})$s sampled, E exceed $B(\mathbf{X})$, the point estimate of $Q(\mathbf{X})$ is E/W. Nearly all the work is in computing a most parsimonious tree for each of W randomizations. As with HER, estimating $Q(\mathbf{X})$ is about W times as much work as finding $K(\mathbf{X})$ or $R^*(\mathbf{X})$. With a large data matrix and $W = 100$, it could take weeks.

[4] In this I follow Faith and Cranston (1991), who have arrived at the same idea.

For estimation, $Q(\mathbf{X})$ is like any binomial proportion. Confidence limits can be found using Fisher's method, which is illustrated by Zar (1984: 378). The often-used normal approximation to binomial confidence limits is not adequate when the sample proportion E/W is anywhere near unity. That is the most important case, as such samples suggest that $\mathbf{X}$ shows low homoplasy.

Computing effort will make it tempting to use a small number W of randomizations. Take care in interpreting result $E/W = 1$. In that case the lower limit of a (necessarily) one-tailed confidence interval of size c for $Q(\mathbf{X})$ is $L(W, c) = (1 - c)^{1/W}$. $L(10, 0.95) = 0.74$. If 10 randomizations all have $B(\mathbf{Y}) > B(\mathbf{X})$, one can conclude with 95% confidence only that $Q(\mathbf{X}) \geq 0.74$. $L(10, 0.99)$ is only 0.63.

A sample of W randomizations can be used in a significance test for the null hypothesis and alternative described before. Reject when E is a large enough fraction of $W + 1$. The error rate is $\alpha' = 1 - E/(W + 1)$. The sample value for $Q'(\mathbf{X})$ can be taken as $-\log(\alpha')$, which is always well-defined. Rejection is possible at $\alpha' \leq 0.05$ only if at least 19 randomizations are sampled.

Correlations

Significance tests as a class cannot reject unless provided with enough data, even when the null hypothesis is in fact false. And when the null hypothesis is false, adding more data changes the test statistic and usually increases the signifiance level.

$Q(\mathbf{X})$ and $Q'(\mathbf{X})$ are significance tests, and a good many real data matrices are more congruent than randomizations. Applied to real data matrices, then, $Q(\mathbf{X})$ and $Q'(\mathbf{X})$ would usually increase with number of terminals or informative characters. But those statistics are measures of congruence, too—at least relative to matrix permutations.

Archie (1989, 1990) objects to the ensemble consistency index C. One of his grounds is routine: C is influenced by autapomorphies. Another is that, among real data matrices, the observed C is negatively correlated with number of terminals[5] and number of characters. Finally, the minimum possible C has similar dependencies. For these reasons, he feels that HER should be used instead.

HER is not influenced by autopomorphies, but it does depend on the other factors. Minima can always be fixed, but correlation in real cases is another matter. In that connection a more productive question is: Used for what?

Adding terminals or informative characters to data often increases the amount of homoplasy that can be detected.[6] To study just how great the increase is, use an index that shows it. Evidently, the increase is usually faster among randomizations than in real data matrices. To show the difference between real and randomized matrices, use another index.

Do not complain if that first index shows the phenomenon of interest. Still less complain if the second index increases with the amount of data available—worry if it does not. Correlation as such is no grounds for objection.

Evidence

Failing to reject a null hypothesis does not mean that the hypothesis is true, only that the information used by the test does not suffice to distinguish between the null and the

[5] That correlation had been discussed by Donoghue at the 1988 meeting of the Society for the Study of Evolution (cf. Sanderson and Donoghue, 1989).

[6] Not that that observation is particularly original. For the effect of adding terminals, for example, see Farris (1972).

alternative. If one kind of information does not distinguish between alternative hypotheses, use other information.

Archie intended random matrix permutation as a kind of evolutionary model. It is not a description of a particular process, but rather of an outcome, what one might expect if characters are distributed independently, and randomly with respect to phylogenetic relationships.

If $\mathcal{N}(\mathbf{X}) = 1$, $Q(\mathbf{X})$—well-defined even then—is 0. Indeed, when there is one informative character, there is no congruence between informative characters. The corresponding $\alpha = 1$ does not mean that a hypothesis of random character distribution is certain, just that in this case congruence provides no evidence on the issue.

Suppose that character is holometaboly, if that feature is random with respect to phylogeny, then the seeming identity of holometaboly in different species must be considered an astonishing coincidence. But, by taking the obvious course of explaining that similarity as the result of inheritance, one rejects the hypothesis of random character distribution. Congruence is not quite all there is to phylogenetic evidence.

Rejecting the null hypothesis of random permutation may provide evidence on the amount of incongruence, but need not provide equal confidence in phylogenetic conclusions. That characters might be correlated for some other reason is commonplace, but one of its ramifications is often overlooked.

One wants confidence in some particular tree. If homoplastic characters, even though relatively abundant, are randomly and independently distributed, then at least some set of similar trees is likely to be favored. Quite distinct most parsimonious trees—as sometimes are found with real data matrices—may be produced by conflicting suites of characters, within which the characters are relatively mutually congruent.

Regardless of how they are found, such suites must reduce confidence in conclusions. At the least one is faced with accounting for the suites. Present statistical tests of congruence do not seem to address this problem. The model on which they are based includes nothing but independent distribution.

Acknowledgment

Without R. L. Shipp's distinctive brand of encouragement, I might well not have written this paper.

REFERENCES

ARCHIE, J. W. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. 38: 235 269.

ARCHIE, J. W. 1990. Homoplasy excess statistics and retention indices: A reply to Farris. Syst. Zool. 39: 169–174.

FAITH, D., AND P. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. Cladistics 7: 1–28.

FARRIS, J. S. 1972. Inferring phylogenetic trees from distance matrices. Am. Nat. 106: 645–668.

FARRIS, J. S. 1989. The retention index and the rescaled consistency index. Cladistics 5: 417–419.

SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution. 43: 1781–1795.

SEBERG, O. 1989. The seventh anmual meeting of the Willi Hennig Society. Cladistics 5: 183–191.

ZAR, J. H. 1984. Biostatistical Analysis. Prentice-Hall, Englewood Cliffs, New Jersey, 718 pp.