Journal of Classification 4:57-71 (1987)

Journal of
Classification

Invariants of Phylogenies in a Simple Case with Discrete States

James A. Cavender

Martin Marietta Denver Aerospace

Joseph Felsenstein

University of Washington

Abstract: Under a simple model of transition between two states, we can work out the probabilities of different data outcomes in four species with any given phylogeny. For a given tree topology, if all characters are evolving under the same probabilistic model, there are two quadratic forms in the frequencies of outcomes that must be zero. It may be possible to test the null hypothesis that the tree is of a particular topology by testing whether these quadratic forms are zero. One of the tests is a test for independence in a simple 2 × 2 contingency table. If there are differences of evolutionary rate among characters, these quadratic forms will no longer necessarily be zero.

Keywords: Phylogenies; Statistical tests.

1. Introduction

How to infer phylogenetic relations among a group of species by examining discrete-state data such as molecular sequences is a problem of statistical inference that remains quite poorly understood. For a discussion of some of the problems, see reviews by one of us (Felsenstein 1983a, 1983b). For simple models of nucleic acid evolution, maximum likelihood estimates can be made under simple models of nucleotide change (Neyman 1971, Felsenstein 1981), but little is known about the exact behavior of these estimation methods. In particular, there has been little examination of methods of testing whether the data support one topology of the evolutionary tree or another. Some progress has been made (Cavender 1978, Cavender 1981,

---

The authors wish to thank Thomas Haigh and an anonymous reviewer for helpful comments on the manuscript. Work by J. Felsenstein was supported in part by task agreement number DE-AT06-76EV71005 of contract number DE-AM06-76RL02225 between the U.S. Department of Energy and the University of Washington.

Authors' Addresses: James A. Cavender, Martin Marietta Denver Aerospace, Box 179, Denver, Colorado 80201 and Joseph Felsenstein, Department of Genetics SK-50, University of Washington, Seattle, Washington 98195, USA (to whom requests for reprints are to be addressed).

Felsenstein 1985a) by studying simplified versions of the problem, with 2 states instead of the 20 that would be needed at minimum to represent amino acids, or the 4 (or 64) that would be needed to cope with nucleic acid sequences. These studies have assumed that there are only 3 or 4 species, and have assumed symmetry of substitution probabilities. They have also restricted examination to test statistics based on parsimony. These studies have revealed that even in these idealized cases, this popular method of inference has surprising and unwelcome statistical properties. It is also possible to use "bootstrap" data resampling techniques to infer a confidence limit (Felsenstein 1985b), although this is necessarily an approximate technique.

The objective of this paper is to take a broader look at the possible statistics that may be used to test differences between tree topologies. In particular, we point out that in a simple model there are polynomial constraints on the expected frequencies of various data configurations. Although this does not solve the problem of how to test differences between tree topologies, it does suggest that there may be a way of reducing problems of phylogenetic inference to questions of whether the class frequencies in multinomial distributions satisfy certain algebraic constraints.

We consider four species, A, B, C, and D. These are related by one of the three unrooted phylogenetic trees of Figure 1. Our problem is to distinguish which of these three topologies is the correct one. There are assumed to be N characters, each one of which may have any one of S states. If a character represented a site in a nucleic acid sequence, S would be 4. Evolution of a nucleic acid sequence is particularly likely to depart from this model if it codes for a protein, so that many of the base mutations would be rejected by natural selection; noncoding sequences are more likely to have their evolution well-approximated by it. For protein sequences, we could take S to be 20, although to do so might be to ignore the complexity of the mapping from the 64 codons to the 20 amino acids plus stop signal. Dayhoff and Eck (1968, pp. 33-41) have adopted just such a simple Markov chain model of amino acid change, tabulating the state transition probabilities empirically. Let f_{ijkl} be the probability of a character being in states i, j, k, l in species A, B, C, D, respectively, given some particular phylogeny. In an example with nucleic acid sequences, perhaps $f_{A,C,C,G} = 0.003$. (Note that in this expression the symbols A, C, and G stand for three nucleic acids, and are not species names). We assume that the evolutionary processes, and hence these probabilities, are the same for all characters and that the evolutionary processes in different characters are independent. We give the name *pattern* to these chance assignments of states to species; that is, (A, C, C, G) is an example of a pattern.

Under our assumptions of independence, the sequence of patterns is exchangeable, and hence the observed frequencies of the patterns constitute all of the data. Any method of inference must proceed by comparing these

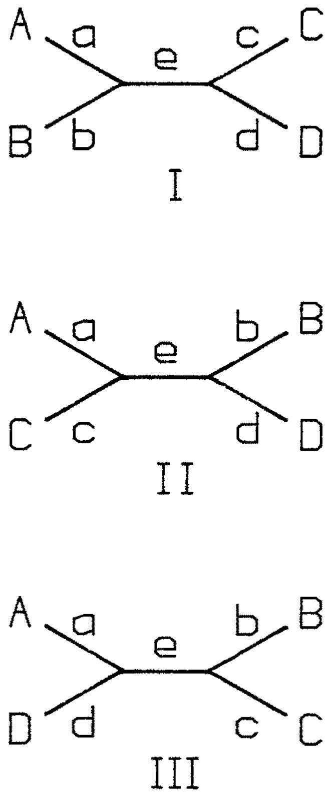

Figure 1. The three unrooted tree topologies, with their branch lengths indicated.

observed frequencies with pattern probabilities f_{ijkl} that may be computed given the tree topology and branch lengths. We regard the set of possible frequencies as a vector or point in a space of S^4-tuples. A particular tree with a particular topology and substitution probabilities leads to a set of pattern probabilities which may be regarded as a vector in the same space. A set of trees leads to a set of such points. For example, the set of all trees of the second topology leads to a surface in the space, as we shall show. If the sample of characters is large, then the vector of frequencies will, with high probability, lie close to the vector of probabilities of the true tree. In fact, the vector of frequencies is a random vector with a multinomial distribution and the expected class frequencies of that distribution make up the vector of probabilities.

Thus the hypothesis that the tree has the second topology, for example, is precisely the hypothesis that the probability vector lies in a certain point set, in this case the surface mentioned above. (In the degenerate case where the two interior vertices coincide, the tree belongs to all three topologies). This hypothesis may be rejected if the observed frequency vector lies far enough from that point set. In this paper, we do not attempt to discover how far is "far enough" or how the distance is to be measured. Instead, we algebraically characterize the point set, under the assumption that this is an essential step toward the invention of such tests.

2. Substitution Probabilities

Consider two species, X and Y. These may be chosen from among both internal and terminal nodes of the tree. Define p_{ij} to be the conditional probability of a character having state j in Y given that it has state i in X. Then the matrix

$$\mathbf{P}_{XY} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1S} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ p_{S1} & p_{S2} & \cdots & p_{SS} \end{bmatrix}$$

which we may also denote as $[p_{ij}]$ is the Markov transition matrix from X to Y. If X evolves into Y which evolves into Z, then elementary Markov chain theory (Feller 1968, section XVII.9, p. 424) gives

$$\mathbf{P}_{XZ} = \mathbf{P}_{XY}\,\mathbf{P}_{YZ} \quad .$$

This depends on an additional assumption: that changes in different edges of the tree are independent. Letting T be the determinant of $\mathbf{P}$ we have

$$T_{XZ} = T_{XY}\,T_{YZ}$$

by virtue of the general identity for determinants, $det\,(\mathbf{AB}) = (det\,\mathbf{A})\,(det\,\mathbf{B})$. We assume that the determinants are strictly positive. This will be the case if the transition matrices vary continuously with branch length. We let $W = -\ln T$ so that

$$W_{XZ} = W_{XY} + W_{YZ} \quad .$$

Because it adds in this simple way, W is a natural measure of branch length. A general form of the evolutionary clock hypothesis is that W is proportional to time along a branch. (In this paper, however, we make no assumption as to the truth of this.)

We will also assume that a Markov equilibrium exists throughout our evolutionary tree. That is, we assume there is one probability distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_S)$ such that

$$\pi \, \mathbf{P}_{XY} = \pi$$

for every pair of species X and Y, and also assume that the probability of a character being in state i is actually π_i, whichever species you are looking at. This assumption of a common equilibrium distribution at all vertices of the tree is not inevitable; it would be of interest to investigate models in which it was not true, but we will not do this here.

If $\mathbf{P}_{XY} = [p_{ij}]$ and we employ the notation $[q_{ij}]$ for the components of the matrix $\mathbf{P}_{YX}$ from Y to X, then

$$\pi_i \, p_{ij} = \pi_j \, q_{ji} \; , \tag{1}$$

both expressions being the probability of a character being in state i at X but in state j at Y (Feller 1968, section XV.11, p. 373). By noting that (1) can be rewritten as

$$\mathbf{D} \, \mathbf{P}_{XY} \, \mathbf{D}^{-1} = \mathbf{P}_{YX}$$

where

$$\mathbf{D} = diag\,(\pi_1, \pi_2, \ldots, \pi_S)$$

and using the product rule for determinants of products of matrices, we easily establish that

$$\begin{gathered} det\,\mathbf{P}_{XY} = det\,\mathbf{P}_{YX} \\ T_{XY} = T_{YX} \\ W_{XY} = W_{YX} \; . \end{gathered} \tag{2}$$

We note that $T \leqslant 1$ and hence that every W is non-negative.

3. Inequalities for Different Topologies

The transition probabilities between two tip species, such as A and D, can easily be related to the expected pattern frequencies. Say for example that

$$\mathbf{P}_{AD} = [p_{il}] \ .$$

Then each side of

$$\pi_i \, p_{il} = \sum_j \sum_k f_{ijkl}$$

is the probability of a character having state i at A but state l at D. Hence

$$p_{il} = \frac{1}{\pi_i} \sum_j \sum_k f_{ijkl} \tag{3}$$

Since p_{il} is expressible in terms of the f's, so are T_{AD} and W_{AD}.

For a tree of the first topology,

$$W_{AC} + W_{BD} = W_{AD} + W_{BC}$$

as a glance at Figure 1 shows, equating the W's with branch length. For the second topology, however, the same two quantities are not equal,

$$W_{AC} + W_{BD} \leqslant W_{AD} + W_{BC} \ .$$

This is also evident from Figure 1, but a careful proof of it must actually employ (2) and the fact that no W is negative. Since each distance W in these relations is observable, this provides a way of distinguishing topologies. Corresponding relations in the determinants T are

$$T_{AC} \, T_{BD} = T_{AD} \, T_{BC}$$

and

$$T_{AC} \, T_{BD} \geqslant T_{AD} \, T_{BC}$$

depending on the underlying topology. We summarize the relations of this type in Table 1.

4. An Example

To see this idea in action, we study the case of $S = 2$ with the simplest possible transition matrices. We assume that all transition matrices have the form

TABLE 1

Signs of inequalities in determinants and branch lengths under the three possible topologies.

	Topology I	Topology II	Topology III
$T_{AC}\ T_{BD} - T_{AD}\ T_{BC}$ and $W_{AD} + W_{BC} - W_{AC} - W_{BD}$	0	+	-
$T_{AD}\ T_{BC} - T_{AB}\ T_{CD}$ and $W_{AB} + W_{CD} - W_{AD} - W_{BC}$	-	0	+
$T_{AB}\ T_{CD} - T_{AC}\ T_{BD}$ and $W_{AC} + W_{BD} - W_{AB} - W_{CD}$	+	-	0

$$\mathbf{P} = \begin{bmatrix} 1-z & z \\ z & 1-z \end{bmatrix}$$

where z varies from branch to branch and $0 \leqslant z \leqslant 1/2$. Then

$$\begin{aligned} T &= (1-z)^2 - z^2 \\ &= 1 - 2z \end{aligned} \tag{4}$$

That is, T is the probability $1 - z$ that the two species match at this character minus the probability z that they do not. The equilibrium distribution is $\pi = (1/2, 1/2)$.

We introduce a more compact notation related to that of Cavender (1978) but with a different system of numbering. Define

$$\begin{aligned} f_0 &= f_{0000} + f_{1111} \\ f_1 &= f_{0001} + f_{1110} \\ f_2 &= f_{0010} + f_{1101} \\ f_3 &= f_{0011} + f_{1100} \end{aligned}$$

$$\begin{aligned}
f_4 &= f_{0100} + f_{1011} \;, \\
f_5 &= f_{0101} + f_{1010} \\
f_6 &= f_{0110} + f_{1001} \\
f_7 &= f_{0111} + f_{1000}
\end{aligned}$$

We can write the determinants as functions of these.

$$\begin{aligned}
T_{AB} &= f_0 + f_1 + f_2 + f_3 - f_4 - f_5 - f_6 - f_7 \\
T_{AC} &= f_0 + f_1 + f_4 + f_5 - f_2 - f_3 - f_6 - f_7 \\
T_{AD} &= f_0 + f_2 + f_4 + f_6 - f_1 - f_3 - f_5 - f_7 \\
T_{BC} &= f_0 + f_1 + f_6 + f_7 - f_2 - f_3 - f_4 - f_5 \\
T_{BD} &= f_0 + f_2 + f_5 + f_7 - f_1 - f_3 - f_4 - f_6 \\
T_{CD} &= f_0 + f_3 + f_4 + f_7 - f_1 - f_2 - f_5 - f_6
\end{aligned}$$

using (4) or (3). For example, $f_0 + f_1 + f_2 + f_3$ is the probability that A and B match because it is the sum of the probabilities of those patterns in which A matches B; similarly, $f_4 + f_5 + f_6 + f_7$ is the probability that A and B do not match. Thus

$$\begin{aligned}
T_{AB} &= (1 - z) - z \\
&= (f_0 + f_1 + f_2 + f_3) - (f_4 + f_5 + f_6 + f_7)
\end{aligned}$$

from (4). To derive the same formula from (3) is just a very laborious substitution. Upon substituting these expressions into the right side of $K_1 = (T_{AC}\ T_{BD} - T_{AD}\ T_{BC})/4$, we obtain

$$K_1 = (f_4 - f_7)(f_2 - f_1) - (f_6 - f_5)(f_0 - f_3) \;.$$

(In this calculation, 120 terms cancel.) We have included the factor of 1/4 in our definition of K_1 merely to improve the appearance of this equation and its relatives in (5). We can define K_2 and K_3 in an analogous fashion so that

$$\begin{aligned}
K_1 &= (f_4 - f_7)(f_2 - f_1) - (f_6 - f_5)(f_0 - f_3) \\
K_2 &= (f_2 - f_7)(f_1 - f_4) - (f_3 - f_6)(f_0 - f_5) \qquad (5) \\
K_3 &= (f_1 - f_7)(f_4 - f_2) - (f_5 - f_3)(f_0 - f_6) \;.
\end{aligned}$$

These respectively equal

$$(T_{AC}\,T_{BD} - T_{AD}\,T_{BC})/4$$

$$(T_{AD}\,T_{BC} - T_{AB}\,T_{CD})/4$$

$$(T_{AB}\,T_{CD} - T_{AC}\,T_{BD})/4.$$

We have thus shown that

$$K_1 = 0,\ K_2 \leqslant 0,\ K_3 \geqslant 0 \ \text{ under topology I} \tag{6}$$

$$K_1 \geqslant 0,\ K_2 = 0,\ K_3 \leqslant 0 \ \text{ under topology II} \tag{7}$$

$$K_1 \leqslant 0,\ K_2 \geqslant 0,\ K_3 = 0 \ \text{ under topology III} \tag{8}$$

5. Statistical Properties

We will not attempt to develop statistical tests in this paper, but we should make a few observations. The quantities K_i are quadratic forms in the expected pattern frequencies. Suppose that we have sampled N characters. Let x_i be the number observed in pattern i. Then x_i/N is an estimate of f_i and may be substituted into (5) to give estimates $\hat{K}_1$ of K_1, $\hat{K}_2$ of K_2, etc. By the law of large numbers, $\hat{K}_i$ will converge to K_i as N increases. Consequently, the following method of choosing a phylogeny is consistent:

Prefer Topology II	over Topology III	if $\hat{K}_1>0$
Prefer Topology III	over Topology II	if $\hat{K}_1<0$
Prefer Topology I	over Topology III	if $\hat{K}_2<0$
Prefer Topology III	over Topology I	if $\hat{K}_2>0$
Prefer Topology I	over Topology II	if $\hat{K}_3>0$
Prefer Topology II	over Topology I	if $\hat{K}_3<0$

By "consistent" we mean that the probability of this rule leading to a wrong answer vanishes as N grows very large.

In principle, a statistical test could be built on this. One rejects topology I as a hypothesis if the data deviates with statistical significance from any of the relationships (6), and so on. Rejecting all three topologies means rejecting the model itself, an interesting and important possibility. We will not pursue this problem of hypothesis testing here. The question is how to test the hypothesis that the parameters of a multinomial distribution satisfy the quadratic equations and inequalities (6) through (8). It is complicated by the existence of further constraints on the parameters, which we take up next.

6. Invariants

There are further functions of the expected frequencies f_i that are constrained by the tree topology, and these can be found by a simple argument, although we do not have a general form for them analogous to the relations in Table 1. Consider tree I of Figure 1. In the symmetric two-state case that we are considering the probability of a state transition in a segment of the tree does not depend on which state is present at the moment of the transition. We can treat the state transitions as events that occur on segments of the tree. Whether there is a transition in a given segment is then independent of whether there is one in another segment. Note that whether A and B are both in the same state depends only on events in branches a and b, and whether C and D are in the same state depends only on events in branches c and d. It is then clear that these two events are independent. The probability that A and B are identical is $f_0 + f_1 + f_2 + f_3$, and the probability that C and D are identical is $f_0 + f_3 + f_4 + f_7$. More generally, we have the following probabilities

	$C = D$	$C \neq D$
$A = B$	$f_0 + f_3$	$f_1 + f_2$
$A \neq B$	$f_4 + f_7$	$f_5 + f_6$

The independence of the events $A = B$ and $C = D$ implies that the probabilities in this table are equal to the product of the row and column totals. This in turn implies that product of the upper-left and lower-right elements equals the product of the other two elements. Therefore under topology I we have

$$(f_0 + f_3)(f_5 + f_6) = (f_4 + f_7)(f_1 + f_2) \ .$$

By an exactly analogous argument we can find similar equalities for the other two topologies. If we define L_1, L_2, and L_3 by

$$L_1 = (f_4 + f_7)(f_1 + f_2) - (f_5 + f_6)(f_0 + f_3)$$

$$L_2 = (f_2 + f_7)(f_1 + f_4) - (f_3 + f_6)(f_0 + f_5)$$

$$L_3 = (f_1 + f_7)(f_2 + f_4) - (f_3 + f_5)(f_0 + f_6)$$

then we can easily show that L_1, L_2, and L_3 are zero under tree topologies I, II, and III respectively.

As an aside we may comment that the test of whether a set of observed frequencies could have been generated by expected frequencies f_i satisfying (say) $L_1 = 0$ is particularly simple: we construct a table of observed numbers of characters having $A = B$ versus $A \neq B$ and $C = D$ versus $C \neq D$. Regarding this as a 2 × 2 contingency table we immediately see that independence of $A = B$ from $C = D$ can be tested by the standard tests of independence in a 2 × 2 table, Fisher's Exact Test and the chi-square test with one degree of freedom.

Before dropping this algebraic study, we point out the identities

$$L_1 - L_2 = K_3$$

$$L_2 - L_3 = K_1$$

$$L_3 - L_1 = K_2$$

$$K_1 + K_2 + K_3 = 0 \ .$$

which imply that among the three K_i and the three L_i there are only three independent quantities.

7. Some Geometry

If the model is expanded so as to permit each character to have its own set $\{a,b,c,d,e\}$ of edge lengths, then the inequalities (6) through (8) are no longer a reliable guide to the topology. An example will show this.

Say A, B, C, and D are species joined by a tree of the first topology. We expect based on the above that

$$K_2 = (f_1 - f_4)(f_2 - f_7) - (f_3 - f_6)(f_0 - f_5) < 0 \ . \tag{9}$$

Now say that for each character there is a probability one-half that $a = d = e = 0.02$ and $b = c = 0.2$. Then, for these characters only, $f =$ (0.602, 0.016, 0.150, 0.018, 0.150, 0.006, 0.038, 0.016).

Say that for the remaining characters, $a = d = e = 0.057632$ and $b = c = 0.392$. For these characters, the W-lengths of the edges are exactly triple those that apply to the other characters. Then for these characters, $f =$ (0.311, 0.039, 0.201, 0.043, 0.201, 0.032, 0.131, 0.039). The value of f_i for the whole set of characters is just the average of the two values for the two subsets. That is,

$$f = (0.456, 0.027, 0.176, 0.030, 0.176, 0.019, 0.084, 0.027).$$

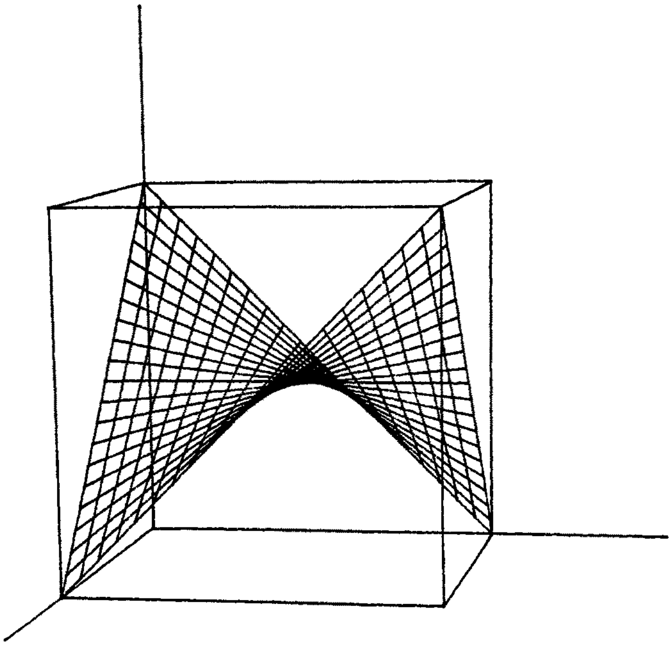

Figure 2. The hyperbolic paraboloid, an analogue in three dimensions to the seven-dimensional quadratic invariants discussed in this paper.

Substituting these values into the left side of (9) gives the value 0.0013, so the invariants have failed. (For this particular example, K_1, K_2, K_3, L_1, L_2, and L_3 are all nonzero, so with a large enough sample one has grounds to expect with a fully-developed statistical test to reject the model and not actually be misled.)

A little geometry can make this phenomenon more understandable. The graph of

$$(f_2 - f_7)(f_1 - f_4) - (f_3 - f_6)(f_0 - f_5) = 0 \ .$$

is a seven-dimensional quadratic hypersurface in the eight-dimensional Euclidean space R^8. Like the familiar, saddle-shaped hyperbolic paraboloid of solid geometry (Figure 2), it is a ruled surface, thoroughly webbed with straight lines. It divides R^8 into two parts, (one being the region where Topology I is preferred over Topology III, the other where Topology III is preferred over I) neither of which is convex. It is this non-convexity that made the above example possible. Two points were chosen on one side of the hypersurface in such a way that the midpoint between them was on the

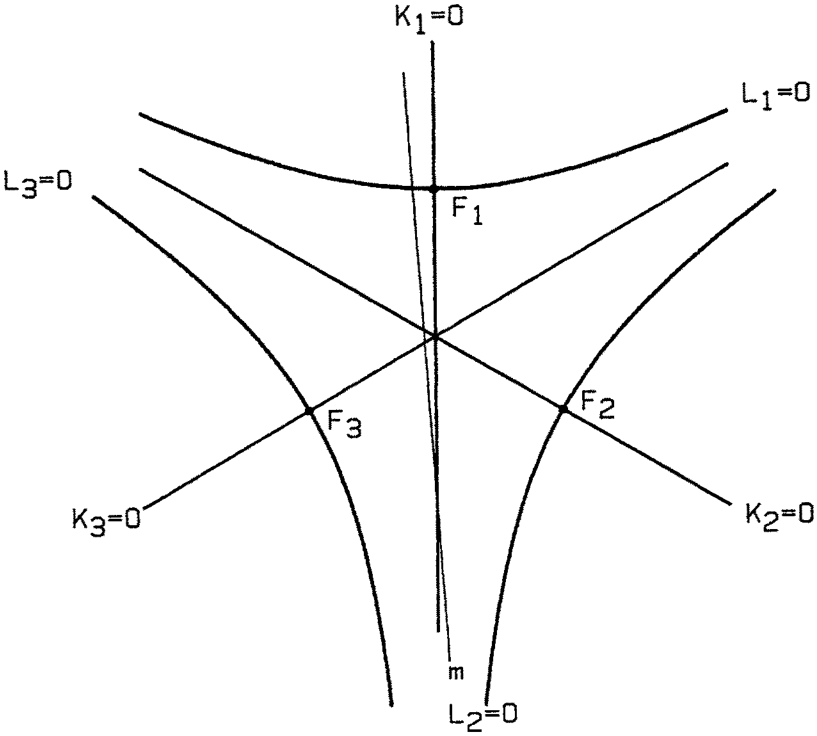

Figure 3. A low-dimensional sketch of the geometric relationships between the F_i, the K_i, and the L_i. Five dimensions have been omitted and are left to the reader's imagination.

other side. In previous papers (Cavender 1978, Cavender 1981, Felsenstein, 1985a) we investigated criteria (formulations of parsimony) that did not have this problem. There, we chose the first topology (for example) whenever the estimates x_i/N of f_i fell in a certain small corner of R^8. And that small corner was convex, with a flat boundary.

A sketch (Figure 3) schematically shows the parts of (5). The sketch represents the space of all octuples $f = (f_0, f_1, \ldots, f_7)$. The varieties F_1 F_2, and F_3, which are the graphs of equation (5), are reduced to points in order to draw them in the plane. (In reality, they are 5-dimensional and have a 4-dimensional set as a common intersection.) The graph of $K_1 = 0$ is represented by a line passing through F_1 and midway between F_2 and F_3. The frequencies $(x_1/N, x_2/N, \ldots, x_7/N)$, which constitute an estimate of f, can be plotted in the same sketch. For large N, such a point will fall close to F_1, F_2, or F_3, so any surface that runs between F_2 and F_3 can be the basis for a consistent choice between the second and third topologies.

Our sketch shows a line *m* representing such a surface. What distinguishes the surface $K_1 = 0$ is its symmetrical position midway between F_2 and F_3.

8. Some Questions

One may legitimately wonder whether all this has gained us anything. We started knowing that the likelihood function for a set of data had five parameters (the branch lengths) and ended with constraints on the expected frequencies of character configurations. In both cases we reduce the dimensions from 8 to 5, so it is not obvious that anything has been gained. We have converted the specification of a phylogeny from a statement in terms of branch lengths and topology to a statement in terms of quadratic functions of expected frequencies.

Of course, what we would like to have is a statistical test discriminating among the tree topologies. Although it would seem possible to carry out a likelihood ratio test by maximizing the likelihood functions under the three alternative bifurcating topologies, these hypotheses are not nested one within another, and the theoretical properties of such a test are thus unknown. We have pointed out that if a test can be found of the quadratic inequalities (6), (7), and (8), this will be equivalent to a test of the tree topology. We are still without such a test, but believe that this represents progress towards finding one.

Two questions which need investigation are (i) whether the quadratic inequalities can be generalized to larger numbers of species, and (ii) whether counterparts can be found for models with larger numbers of states or with inequalities of probabilities of change in different characters. One may doubt that things will be so simple in those models, but we believe that the investigation will yield results that cannot help being illuminating.

References

CAVENDER, J.A. (1978), "Taxonomy with Confidence," *Mathematical Biosciences, 40*, 271-280 (erratum, *44*, 308 (1979)).

CAVENDER, J.A. (1981), "Tests of Phylogenetic Hypotheses Under Generalized Models," *Mathematical Biosciences, 54*, 217-229.

DAYHOFF, M.O., and ECK, R.V. (1968), *Atlas of Protein Sequence and Structure 1967-1968*, Silver Spring, Maryland: National Biomedical Research Foundation.

FELLER, W. (1968), *An Introduction to Probability Theory and Its Applications, Vol. 1*, 3rd ed., New York: Wiley.

FELSENSTEIN, J. (1981), "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach," *Journal of Molecular Evolution, 17*, 368-376.

FELSENSTEIN, J. (1983a), "Inferring Evolutionary Trees from DNA Sequences," in *Statistical Analysis of DNA Sequence Data*, Ed. B.S. Weir, New York: Marcel Dekker, 133-150.

FELSENSTEIN, J. (1983b), "Statistical Inference of Phylogenies," *Journal of the Royal Statistical Society, A146*, 246-272.

FELSENSTEIN, J. (1985a), "Confidence Limits on Phylogenies with a Molecular Clock," *Systematic Zoology, 34*, 152-161.

FELSENSTEIN, J. (1985b), "Confidence Limits on Phylogenies: An Approach Using the Bootstrap," *Evolution, 39*, 783-791.

NEYMAN, J. (1971), "Molecular Studies of Evolution: A Source of Novel Statistical Problems," in *Statistical Decision Theory and Related Topics*, Eds. S.S. Gupta and J. Yackel, New York: Academic Press, 1-27.