

Points of View

Syst. Biol. 43(4):560-564, 1994

The Maximum Likelihood Point for a Phylogenetic Tree Is Not Unique

MIKE STEEL

*Mathematics Department, University of Canterbury,
Christchurch, New Zealand¹*

A simple example was constructed to counter a proposed claim that under simple models of sequence evolution there exists at most one point where the likelihood function for a given phylogenetic tree is maximized. This claim suggests that for real data standard methods may not truly be maximizing the likelihood for each tree. Although the method of maximum likelihood has much to recommend it, caution should be exercised in its implementation.

The method of maximum likelihood has, justifiably, become an increasingly important method for reconstructing phylogenetic trees and for testing phylogenetic hypotheses from genetic sequences (Felsenstein, 1981; Barry and Hartigan, 1987; Goldman, 1990; Saitou, 1990; Swofford and Olsen, 1990). The method starts with a suitable stochastic model for describing the probability of "evolving" the observed sequences and adjusts the parameters of this model to maximize this probability. Generally, the parameters consist of the underlying evolutionary tree, T , and values that define the transition matrices assigned to the edges of T ; we call these the edge parameters (sometimes additional parameters may be involved, for instance in describing a distribution of substitution

rates across sequence sites, however we do not need to consider these here). Different models impose different constraints on the types of transition matrices that can be assigned to edges of the tree.

In applying the maximum-likelihood method to evolutionary tree reconstruction, the optimization procedure is generally a two-step process: for each tree, the edge parameters are chosen to maximize the likelihood (the probability of evolving the observed sequences) and then the tree(s) that maximizes these optimal likelihood values is selected.

A difficulty with this approach, however, is that it is computationally intensive (Fukami and Tateno, 1989; Swofford and Olsen, 1990). In particular, the first level of optimization (on the edge parameters) has no simple analytical solution and requires iterative numerical methods, unlike some of the other tree-building methods that are also based on a double-level optimization (i.e., first across edge parameters and then across trees), such as (1) maximum parsimony, (2) minimal evolution, and (3) closest tree. For these three methods, the first level of optimization is explicitly solved by, respectively, Fitch's algorithm (Fitch, 1971), Rzhetsky and Nei's method (Rzhetsky and Nei, 1993), and the closest tree algorithms (Hendy, 1991; Steel et al., 1992).

¹ E-mail: mas@math.canterbury.ac.nz.

For maximum likelihood there is no comparable result. Nevertheless, if it could be shown that the likelihood landscape was always maximized at a unique point in the edge-parameter space, then, as pointed out by Fukami and Tateno (1989), simple hill-climbing methods (which stop whenever a local maximum value is found) would efficiently handle this first level of optimization. With this in mind, Fukami and Tateno (1989) presented an argument aimed at showing that for the simple stochastic model described by Felsenstein (1981) such a unique result holds. The reassuring implication, therefore, was that the optimization algorithm suggested by Felsenstein (1981) and the apparently faster one devised by Fukami and Tateno (1989) would always guarantee to find the maximum-likelihood value for a given tree.

However as shown here, the underlying uniqueness claim is in error. Consequently, local optimization techniques (including those described by Felsenstein [1981], Barry and Hartigan [1987], and Fukami and Tateno [1989]) may yield erroneous maximum-likelihood values and thereby select trees with a less-than-maximum likelihood, even when the search is extended over all trees. Caution should be exercised in implementing maximum likelihood to ensure that the edge-parameter space is properly searched during optimization.

Before describing a simple counterexample to the uniqueness claim, I outline the principal error in the argument, which is perhaps best understood by way of an analogy. Suppose one is standing on an island and that wherever one is standing on this island the north-south line passes through at most one high point and the east-west line also encounters at most one high point. Then it might be concluded that there can be at most one high point on the entire island. However, upon reflection, it is clear that this need not be so; for instance, there may be a small hill in the northwest corner and another in the southeast corner, with the remainder of the island being flat.

Mathematically, the situation is as follows. Suppose F is a smooth function of n

real variables x_1, \dots, x_n , each of which ranges over a fixed, bounded interval $[a, b]$ (in our previous description, $n = 2$; $F(x_1, x_2)$ is the height of the point having north-south/east-west coordinates (x_1, x_2)). Suppose that whenever $n - 1$ of the variables are given fixed, arbitrary values, then F assumes at most one local maximal value (either at an endpoint of $[a, b]$ or at an interior point) as the remaining variable ranges over $[a, b]$. Then it is still possible for F to have many isolated local maxima.

Fukami and Tateno (1989) correctly showed that when F is a likelihood function (the probability of evolving the observed sequences under a simple stochastic model) and the x_i 's are the edge parameters for that model, the above conditions are met. However, they appear to have tacitly, but incorrectly, assumed that this implies that the domain of F must possess a unique point that realizes a local maximum for F . Actually, the authors do allow for a certain exceptional case in which the likelihood function is locally flat; however, I will now present a much stronger counterexample in which the local maxima are only realized at widely separated points and so any small perturbation of this example would also have multiple local maxima. This example shows that the initial conclusion is incorrect. For real data, the maximum value may be realized at a unique point, at least on some trees (particularly on the maximum likelihood tree), although this now requires further investigation.

Consider first the tree with four taxa, as in Fukami and Tateno (1989: fig. 1), which has taxa 1 and 2 separated from taxa 3 and 4 by an internal edge (see Fig. 1).

For simplicity, consider just two nucleotide states, denoted by + and -, and the data will consist of four sequences, each of length 2 (i.e., consisting of just two sites) (Table 1).

More realistic examples, such as with four-state sequences and using a variety of different patterns for the sites including constant sites, can also be considered; however, it is sufficient to find one counterexample to disprove a claim, and so we deal with the simplest.

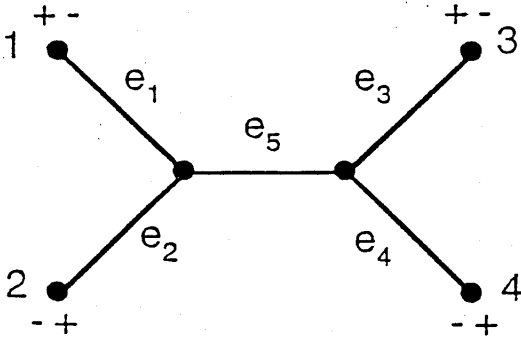


FIGURE 1. A resolved (binary) phylogenetic tree of taxa 1, 2, 3, and 4, and the labeling of the five edges (e_1, \dots, e_5) of this tree. The corresponding state (+ or -) of each taxon for the two-site counterexample is as shown.

For each taxon, the frequency of state + and of - in the above sequences is 0.5, and so the simple Poisson process model considered by Fukami and Tateno (1989; see also Felsenstein, 1981) reduces, for two states, to the Cavender-Farris model (Cavender, 1978), which has symmetric 2×2 transition matrices on all edges of the tree.

Thus, the probability p_i that the ends of edge e_i of T are in different states is given by

$$p_i = \frac{1}{2}[1 - \exp(-2v_i)],$$

where v_i is the expected number of substitutions along edge e_i . Given values for the p_i , the probability of generating the above two-site sequences under the usual independence assumption (between sites and across edges of the tree) is simply

$$L = L_1^2,$$

TABLE 1. Nucleotide states (-, +) for a sequence with two sites for four taxa (1-4).

Taxon	Sites	
	1	2
1	+	-
2	-	+
3	+	-
4	-	+

where

$$L_1 = \frac{1}{2}(q_1 p_2 q_5 q_3 p_4 + q_1 p_2 p_5 p_3 q_4 + p_1 q_2 q_5 p_3 q_4 + p_1 q_2 p_5 q_3 p_4) \quad (1)$$

and where $q_i = 1 - p_i$, the probability that the ends of edge i are in the same state (the edges are ordered as in Fig. 1). The quantity L is called the likelihood of generating the sequences, and we wish to maximize L subject to the constraints that $0 \leq p_i < 0.5$ (which is equivalent to the condition that $v_i \geq 0$). We will show that L assumes strictly maximal values at two separate points on the closed set $C = \{(p_1, \dots, p_5): 0 \leq p_i \leq 0.5\}$, namely,

$$P1: p_1 = p_3 = 0.5; p_2 = p_4 = p_5 = 0,$$

and

$$P2: p_2 = p_4 = 0.5; p_1 = p_3 = p_5 = 0,$$

where, as is easily checked from Equation 1, $L = 1/64$ in both cases.

These two points are local maxima, that is, deviating from them by any small amount must strictly lower L ; this suffices to establish a counterexample to the claimed uniqueness. Actually, I will do this by establishing a stronger (though less trivial to derive) result, namely that P1 and P2 maximize L on C and that no other point in C gives such a large L value.

First, to show that $1/64$ is the maximal value of L over the closed set C , let $z_i = 1 - 2p_i$ for $i = 1, \dots, 5$. By elementary algebra, it can be checked from Equation 1 that

$$L_1 = \frac{1}{16}(1 - z_1 z_2 + z_1 z_3 z_5 - z_2 z_3 z_5 - z_1 z_4 z_5 + z_2 z_4 z_5 - z_3 z_4 + z_1 z_2 z_3 z_4). \quad (2)$$

This equation is given by Hendy's theorem (Hendy, 1989); however, we do not need to discuss this theorem here because Equation 2 can be verified directly from Equation 1. By the independence assumption in the underlying stochastic model, it is clear that L_1 is bounded above by (the

probability that taxon 1 is in a particular state [e.g., +] and differs in state from taxon 2) \times (the probability that taxon 3 differs in state from taxon 4). These two probabilities are $(p_1q_2 + q_1p_2)/2$ and $p_3q_4 + q_3p_4$, respectively; written in terms of the z_i 's they are $(1 - z_1z_2)/4$ and $(1 - z_3z_4)/2$, respectively. Thus,

$$L_1 \leq \frac{1}{8}[(1 - z_1z_2)(1 - z_3z_4)].$$

If preferred, this inequality can be formally derived just from Equation 2. Thus, $L = L_1^2 \leq 1/64$, with equality precisely when both

$$z_1z_2 = 0 \quad (3)$$

and

$$z_3z_4 = 0. \quad (4)$$

Thus, we see immediately that L assumes maximal values at P1 and P2. I can now show that these points are the only such points by exploiting Equations 3 and 4. Either z_1 or z_2 must be 0, and either z_3 or z_4 must be 0. Consider the four possible cases, beginning with $z_1 = z_3 = 0$. Equation 2 then becomes $L_1 = (1 + z_2z_4z_5)/16$, which equals its maximum possible value, $1/8$, precisely when $z_2 = z_4 = z_5 = 1$. Then the five z_i values define point P1. Similarly, if $z_2 = z_4 = 0$, we obtain point P2. If $z_1 = z_4 = 0$, Equation 2 becomes $L_1 = (1 - z_2z_3z_5)/16 \leq 1/16$. In this case, $L \leq 1/256$ and so no maximum value can be realized. Similarly, for the remaining case, $z_2 = z_3 = 0$. Thus, P1 and P2 are the only points that maximize L_1 and thereby L , as claimed.

By slightly modifying the example, one could produce local maxima that do not have the same likelihood values. In this case, the implementations of maximum likelihood described earlier, which are based on simple hill climbing in the first level of their optimization, may select a tree with less-than-maximum likelihood (even when all trees are examined under the second level of optimization). Also, it is possible to generalize the above counterexample to more taxa, and more states, to construct examples in which the like-

lihood is maximized at a very large number of different points (Steel, in prep).

It might be objected that $p_i = 0.5$ has been allowed on edges of the tree; however, this was just a computational convenience. The result shows that the high portions of the likelihood landscape can occur in two quite different parts of the (acceptable) open interior of this domain. Also, it might be objected that this particular counterexample is overly pathological; certainly it is contrived (though surprisingly simple). The important point is that it demonstrates that even for the most simple models the likelihood function need not necessarily be optimized at a unique point for a tree (the possibility of more complex models giving rise to multiple optima had already been suggested by Golding and Felsenstein, 1990). Consequently, greedy optimization schemes, such as that described by Fukami and Tateno (1989), should be used with care until or unless it can be shown that multiple optima do not arise for real data. Computational experiments should be carried out to determine this.

REFERENCES

- BARRY, D., AND J. A. HARTIGAN. 1987. Statistical analysis of hominoid molecular evolution. *Stat. Sci.* 2:191-210.
- CAVENDER, J. A. 1978. Taxonomy with confidence. *Math. Biosci.* 40:271-280.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368-376.
- FITCH, W. M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst. Zool.* 20:406-416.
- FUKAMI, K., AND Y. TATENO. 1989. On the uniqueness of the maximum likelihood method for estimating molecular trees: Uniqueness of the likelihood point. *J. Mol. Evol.* 28:460-464.
- GOLDING, B., AND J. FELSENSTEIN. 1990. A maximum likelihood approach to the detection of selection from a phylogeny. *J. Mol. Evol.* 31:511-523.
- GOLDMAN, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. *Syst. Zool.* 39:345-361.
- HENDY, M. D. 1989. The relationship between simple evolutionary tree models and observable sequence data. *Syst. Zool.* 38:310-321.
- HENDY, M. D. 1991. A combinatorial description of the closest tree algorithm for finding evolutionary trees. *Discr. Math.* 96:51-58.

- RZHETSKY, A., AND M. NEI. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol. Biol. Evol.* 10:1073-1095.
- SAITOU, N. 1990. Maximum likelihood methods. *Methods Enzymol.* 183:584-598.
- STEEL, M. A., M. D. HENDY, L. A. SZÉKELY, AND P. L. ERDŐS. 1992. Spectral analysis and a closest tree method for genetic sequences. *Appl. Math. Lett.* 5:63-67.
- SWOFFORD, D. L., AND G. J. OLSEN. 1990. Phylogeny reconstruction. Pages 411-501 in *Molecular systematics* (D. M. Hillis and C. Moritz, eds.). Sinauer, Sunderland, Massachusetts.

Received 29 December 1993; accepted 25 May 1994