J Mol Evol (1989) 28:460–464

Journal of Molecular Evolution

On the Maximum Likelihood Method for Estimating Molecular Trees: Uniqueness of the Likelihood Point

Kaoru Fukami[1] and Yoshio Tateno[2]

[1] Ochanomizu University, Ohtsuka, Tokyo 112, Japan
[2] National Institute of Genetics, Mishima 411, Japan

Summary. Studies are carried out on the uniqueness of the stationary point on the likelihood function for estimating molecular phylogenetic trees, yielding proof that there exists at most one stationary point, i.e., the maximum point, in the parameter range for the one parameter model of nucleotide substitution. The proof is simple yet applicable to any type of tree topology with an arbitrary number of operational taxonomic units (OTUs). The proof ensures that any valid approximation algorithm be able to reach the unique maximum point under the conditions mentioned above. An algorithm developed incorporating Newton's approximation method is then compared with the conventional one by means of computer simulation. The results show that the newly developed algorithm always requires less CPU time than the conventional one, whereas both algorithms lead to identical molecular phylogenetic trees in accordance with the proof.

Key words: Maximum likelihood method — Likelihood point — Molecular taxonomy — Molecular evolution — Computer simulation

Introduction

The maximum likelihood method for estimating molecular phylogenetic trees (or molecular trees, for short) from nucleic acid sequence data was developed by Felsenstein (1981). Because this method has a sound statistical basis, it has attracted the attention of serious molecular evolutionists who are aware of the fact that nucleotide substitution in evolution is a stochastic process. In particular, Hasegawa and his colleagues (Hasegawa and Yano 1984; Hasegawa et al. 1985) have used this method extensively in making statistical inferences about the taxonomic relationships of several eucaryotic species. The computer program of the method developed and revised several times by the originator has been widely distributed among concerned scientists.

In the method, Felsenstein introduced an iteration algorithm to search for the maximum point on a likelihood surface for a given range of a parameter, as a specific case of the general EM algorithm of Dempster et al. (1977). The algorithm, however, guarantees only that an iteration pass never goes downhill on the likelihood surface, and thus is effective only in determining a stationary point. The question then is how many stationary points exist in that range: if there is one, the stationary point is certainly the maximum point (see below); if there are two or more, the stationary point reached may not be the maximum point. Although this problem is implicit in the theoretical background of his method, Felsenstein (1981) did not pay much attention to it.

In this paper, we prove that there is at most one stationary point on the likelihood surface under a certain condition for the case of one parameter model of nucleotide substitution (the case of no distinction between transition and transversion). It is noteworthy that the proof is simple yet independent of tree topology. This ensures that any algorithm to search for a stationary point should lead to the max-

Offprint requests to: Y. Tateno
Contribution No. 1780 from the National Institute of Genetics, Mishima 411, Japan

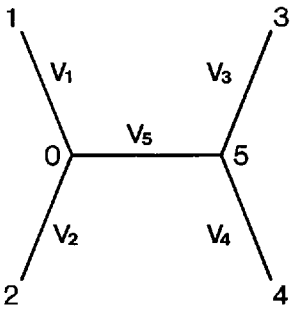

Fig. 1. Molecular tree of four OTUs. Numbers 1 through 4 refer to the four OTUs and 0 and 5 to interior nodes. The branch lengths, v_i (i = 1, 2, 3, 4, 5), are given in number of nucleotide substitutions. The tree is an unrooted one, in which the place of the common ancestor of the OTUs involved is left undetermined.

imum point on a likelihood surface in those circumstances.

Number of Stationary Points

Let us illustrate our argument using a molecular tree of 4 operational taxonomic units (OTUs) (Sokal and Sneath 1963) with two interior nodes 0 and 5 as shown in Fig. 1. In the figure, v_k (k = 1, 2, 3, 4, 5) given alongside a branch is its length in number of nucleotide substitutions. Let us suppose that a nucleotide sequence n sites long is available for each OTU in the tree. Let us also arbitrarily designate node 0 as the pivotal point, from which the likelihood computation begins. Then the likelihood of the tree at one of the n sites, say r, is given by,

$$L_r = \sum_{s_{0r}} \sum_{s_{5r}} \pi(s_{0r})P(s_{0r}, s_{1r}; v_1)P(s_{0r}, s_{2r}; v_2) \times P(s_{0r}, s_{5r}; v_5)P(s_{5r}, s_{3r}; v_3) \times P(s_{5r}, s_{4r}; v_4), \quad (1)$$

in which s_{ir} is the state (base) of site r at OTU (or node) i (Felsenstein 1981). Because the states of interior nodes 0 and 5 are unknown, L_r is defined by taking the sum over all possible base assignments to those interior nodes as shown in formula (1). In the formula, $\pi(s_{0r})$ is the prior probability of taking state s_{0r} at node 0, and $P(s_{ir}, s_{jr}; v_k)$ is the transition probability that state s_{ir} will be replaced by state s_{jr} after v_k substitutions occur. Note that v_k is the parameter that we are to estimate by maximizing the likelihood.

We here make no distinction between transition and transversion, and assume that the substitution follows a Poisson process with parameter v_k. Then $P(s_{ir}, s_{jr}; v_k)$ is given by,

$$P(s_{ir}, s_{jr}; v_k) = e^{-v_k}\delta(s_{ir}, s_{jr}) + (1 - e^{-v_k})\pi(s_{jr}), \quad (2)$$

where $\delta(s_{ir}, s_{jr})$ is Kronecher's delta function, which is 0 for $s_{ir} \neq s_{jr}$ and 1 for $s_{ir} = s_{jr}$, and $\pi(s_{jr})$ for $j \neq 0$ is the probability of replacing s_{ir} by s_{jr} on the condition that s_{ir} be replaced. The conditional probability $\pi(s_{jr})$ is set to the constant at any OTU or node on the assumption that the base composition at any site at any OTU or node be in evolutionary steady state.

If we focus our argument on branch v_5 in the tree, and substitute formula (2) into formula (1), we obtain a simplified expression for L_r:

$$L_r = A_r q_5 + B_r p_5, \quad (3)$$

where $q_5 = e^{-v_5}$, $p_5 = 1 - e^{-v_5}$, and

$$A_r = \sum_{s_{0r}} \sum_{s_{5r}} \pi(s_{0r})P(s_{0r}, s_{1r}; v_1) \times P(s_{0r}, s_{2r}; v_2)P(s_{5r}, s_{3r}; v_3) \times P(s_{5r}, s_{4r}; v_4)\delta(s_{0r}, s_{5r}), \quad (4)$$

and

$$B_r = \sum_{s_{0r}} \sum_{s_{5r}} \pi(s_{0r})\pi(s_{5r})P(s_{0r}, s_{1r}; v_1) \times P(s_{0r}, s_{2r}; v_2) \times P(s_{5r}, s_{3r}; v_3)P(s_{5r}, s_{4r}; v_4), \quad (5)$$

(Felsenstein 1981). If we assume here that nucleotide substitution at each site occurs independently of the rest, we can obtain easily the total likelihood, L, as the product of L_r's for all n sites given by,

$$L = \prod_{r=1}^{n} (A_r q_5 + B_r p_5). \quad (6)$$

Parameter p_5 or q_5 $(=1 - p_5)$ is then estimated as that which maximizes L, and v_5 is obtained accordingly. Technically, p_5 is computed as the particular point at which the first derivative of ln L with respect to p_5 is zero, and the iteration algorithm mentioned earlier is used for this purpose.

Because formula (6) is at most an n-th order polynomial in p_5, equation $L(p_5) = 0$ has at most n real roots, $A_r/(A_r - B_r)$ (r = 1, 2, . . . , n). Accordingly, equation $dL(p_5)/dp_5 = 0$ harbors at most $(n - 1)$ real roots, each of which exists between two consecutive solutions of $L(p_5) = 0$ in descending (or ascending) order of magnitude. It is noted that this also holds true for the case of multiple roots of $L(p_5) = 0$, in which the root merges into that of $dL(p_5)/dp_5 = 0$. Thus, the problem of the number of the stationary points can be reduced to the relationship between the root $A_r/(A_r - B_r)$ and a pair of A_r and B_r in the parameter range at each of the n sites. Depending on the range taken by A_r and B_r, the root falls into four separate cases that are mutually exclusive and comprehensive: (1) $A_r > 0$, $B_r > 0$, and $A_r \neq B_r$; (2) $A_r > 0$, $B_r > 0$, and $A_r = B_r$; (3) $A_r =$

0 and $B_r > 0$; and (4) $A_r = B_r = 0$. It is needless to mention that $L(p_5)$ is always nonnegative in the parameter range for each of the four cases.

In the first case, which occurs most frequently, the root can be rewritten as $1/[1 - (B_r/A_r)]$. Thus, the root always lies outside the parameter range [0, 1] as shown in Fig. 2a. In the second case, $L_r(p_5)$ becomes independent of p_5, and the root does not exist for site r. The orders of $L(p_5)$ and $dL(p_5)/dp_5$ are consequently reduced. In the third case, $p_5 = 0$ is a root of $L(p_5) = 0$. If the root is a multiple root, it is also a root of $dL(p_5)/dp_5 = 0$. In the fourth case, the root is indefinite, making it impossible to obtain the unique stationary point. Although the last three cases are extreme and rare, they cannot be neglected in the theoretical consideration of the maximum likelihood method, in particular when implementing the iteration algorithm.

From the above characterization of the root of $L(p_5) = 0$, we see that there are no roots in the parameter range [0, 1] except for $p_5 = 0$. Thus, there can be at most one stationary point excepting $p_5 = 0$, and this must be the maximum point in the range. Even if $p_5 = 0$ is a stationary point as in the third case, it is not the maximum point, because $L(p_5)$ must be nonnegative in the parameter range. As can be deduced from the argument in the first case, there is at most one stationary point besides $p_5 = 0$, which is the maximum point in the range (see Fig. 2b). If $L(p_5)$ is a monotonically increasing or decreasing function in the range, there is, strictly speaking, no stationary point there. If this is the case, $p_5 = 1$ or $p_5 = 0$ is taken as the substitute for the maximum point, according to whether $L(p_5)$ is monotonically increasing or decreasing (see Fig. 2a).

The aforementioned discussion is shown to hold for each of the other branches, v_1, v_2, v_3, and v_4. Moreover, it can be extended readily to any type of tree topology with an arbitrary number of OTUs, because we can always focus our attention on one branch at a time and carry out the same analysis on it.

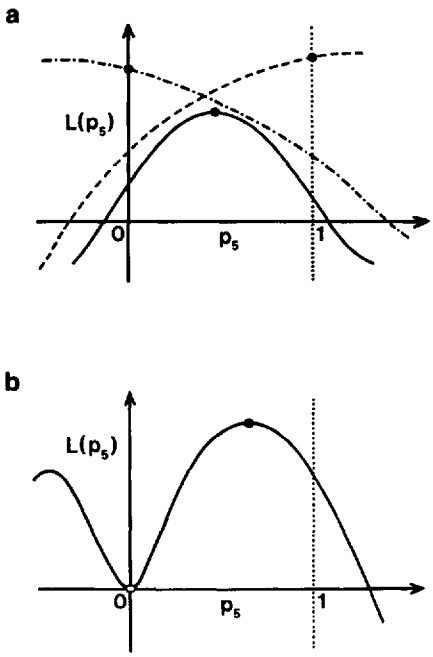

Fig. 2. Likelihood functions and the location of the maximum likelihood point in the parameter range. **a** Three likelihood functions are shown for the case that there exists at most one stationary point in the range. The maximum likelihood point or its substitute is shown by a dot. The broken line and chain line are monotonically increasing and decreasing functions, respectively, in which the maximum likelihood point is substituted by one for the former and by zero for the latter. **b** Likelihood function with two stationary points is shown for the case of the multiple root. Because the function is nonnegative in the range, $p_5 = 0$ is the minimal point as shown by an open circle, and the other is the maximum point as marked with a dot.

Discussion

The problem of the number of stationary points in the parameter range possibly could be solved by Strum's theorem in algebra, because equation, $d\ln L(p_i)/dp_i = 0$ for branch i, could be considered equivalent to an algebraic equation of the $(n - 1)$th order or less. In practice, however, we had difficulty in applying the theorem to our problem. At first glance, $L(p_i) = 0$ looks like an n-th order algebraic equation with coefficients of A_r's and B_r's. Unfortunately, that is not the case; those "coefficients" are dependent upon p_i, as formulas (4) and (5) show. Namely, A_r's and B_r's are changed each time in the iteration pass and settled as the pass converges to the maximum point. It is still possible to use the theorem for examination of the number of stationary points, because the range of coefficients can be determined as was done in the present study. Computation of the coefficients is, however, so complicated that it prevents us from gaining insight into the relationship between the coefficients and the number of stationary points by the use of Strum's theorem. This approach is, in any event, more complicated than the present proof in the one-parameter case.

Because the proof ensures that any valid approximation algorithm be able to reach the unique maximum point, it is worthwhile to develop an algorithm that works better than Felsenstein's. With this intention in mind, we incorporated Newton's popular approximation algorithm to obtain the maximum point, because this algorithm is believed to quickly converge to the solution, in particular, when the initial value given is close to the solution. Then, to examine the performance of our algorithm in comparison with Felsenstein's, a computer simu-

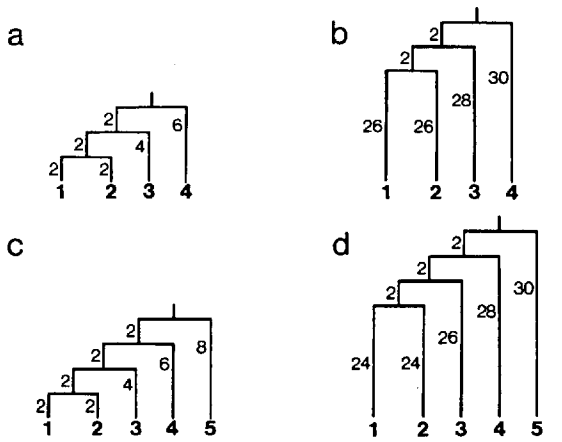

Fig. 3. Four model molecular trees used in the computer simulation. Trees **a** and **b** are of four OTUs, and **c** and **d** of five OTUs. The value given alongside a branch is its length in number of nucleotide substitutions. The value was used as the parameter of a Poisson distribution from which the actual number of nucleotide substitutions was sampled.

lation was conducted. The model and method of the simulation were the same as those of Tateno et al. (1982) except for the model tree and the number of replications. The model trees used in the present study are shown in Fig. 3, whereas the number of replications was set to be 10 for each model tree. The results show that, as expected from the proof, Felsenstein's and our algorithms constructed the identical tree in every replication for each model tree. Our algorithm, however, always took less CPU time than Felsenstein's to reach the maximum point, as shown in Table 1. In particular, when the branch lengths of the model tree were long as in model trees b and d, the time required by ours was about half that by Felsenstein's.

Though the main objective of the present paper is to discuss the number of stationary points, it is also of interest to see how the maximum likelihood method performs in the construction of molecular trees. General discussion on this subject is evidently difficult, but we can make inference into the subject to some extent, using the results of the aforementioned simulation study.

In the simulation study we also obtained the mean and standard deviation of the distortion index d_T (Robinson and Foulds 1981; Tateno et al. 1982), which is a measure taking larger values as the topology of a constructed tree deviates further from the model tree. For a model tree with t OTUs, the value of d_T ranges from 0, for the perfect match of the two topologies, to $2(t-3)$ for the least possible match in the unrooted tree. The d_T values were 0.0 ± 0.0, 0.2 ± 0.6, 0.6 ± 1.0, and 2.6 ± 1.6 for model trees a, b, c, and d, respectively. Though 10 replications is not enough for a rigorous statistical analysis, the results show a trend that the performance

Table 1. Mean CPU time in seconds consumed by the two algorithms for estimating each of the four model trees

Model tree	Algorithm[a]	
	Felsenstein's	New
a	0.27 ± 0.05	0.26 ± 0.06
b	1.35 ± 0.51	0.55 ± 0.06
c	16.64 ± 4.28	13.71 ± 3.37
d	63.51 ± 12.35	31.17 ± 2.47

[a] Felsenstein's refers to the iteration algorithm developed by Felsenstein (1981) and New to one devised in this paper. Number of replications was 10 for each model tree

of the method declines as branch lengths are extended in both cases of four and five OTUs.

The trend can be explained first by considering the arrangement of the nucleotide sequences in question. Saitou and Nei (1986) studied phylogenetic relationships of primates using the nucleotide configuration, which is a base array at a site in a set of nucleotide sequences. In the maximum likelihood method, the nucleotide configuration also plays an important role in differentiating one topology from others. For example, such configurations as (1, 1, 1, 2, 2) and (2, 2, 1, 1, 1), in which 1 and 2 are any pair of different bases, for (OTU 1, OTU 2, OTU 3, OTU 4, OTU 5) are consistent with model trees c and d in Fig. 3, whereas configurations like (1, 1, 2, 1, 2) and (1, 1, 2, 2, 1) are not. The probability for the consistent configurations is about 50 times higher than that for the inconsistent ones in model tree c, whereas the former is only 3.4 times higher than the latter in model tree d in the one-parameter model. This clearly shows that model tree d is much more difficult to reconstruct than model tree c, though both trees have the same number of OTUs.

For the trend to have occurred, there seems to be another reason that is intrinsic to the maximum likelihood method. Let us consider an extreme situation of the tree in Fig. 1, in which all four exterior branches grow infinitely long. In this case, formula (1) takes its asymptotic form given by,

$$L_r = \pi(s_{1r})\pi(s_{2r})\pi(s_{3r})\pi(s_{4r}), \tag{10}$$

which is the likelihood at site r for the star topology (Kimura 1983) with four OTUs and is no longer dependent on tree topology, indicating that the total likelihood in formula (6) is also free from tree topology. For the star topology of t OTUs with the infinitely long exterior branches, L_r is given by,

$$L_r = \prod_{i=1}^{t} \pi(s_{ir}), \tag{11}$$

which converges to zero as t approaches infinity, as does the total likelihood. Thus, it is expected that

the difference in the likelihood value between any two tree topologies becomes less conspicuous as the exterior branches grow longer. In this situation, it is probable that by chance a tree topology more different from the model tree has a higher likelihood value than one less different.

Acknowledgments. Part of the present work was carried out by using the FACOM M-780 mainframe of the Institute of Physical and Chemical Research (RIKEN). We thank Dr. T. Yasunaga and Mr. T. Horiki of the Computer Center at RIKEN for their help. We are also grateful to Dr. K. Aoki for his valuable comments on the first draft of the paper. The present work was supported by grants from the Ministry of Education, Science, and Culture, Japan.

References

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 39:1–38

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Hasegawa M, Yano T (1984) Phylogeny and classification of Hominoidea as inferred from DNA sequence data. Proc Jpn Acad B 60:389–392

Hasegawa M, Iida Y, Yano T, Takaiwa F, Iwabuchi M (1985) Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. J Mol Evol 22:32–38

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Math Biosci 53:131–147

Saitou N, Nei M (1986) The number of nucleotides required to determine the branching order of three species, with special reference to the human–chimpanzee–gorilla divergence. J Mol Evol 24:189–204

Sokal RR, Sneath PHA (1963) Principles of numerical taxonomy. Freeman, San Francisco

Tateno Y, Nei M, Tajima F (1982) Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J Mol Evol 18:387–404

Received July 5, 1988/Revised and accepted September 19, 1988