The Bayesian "Star Paradox" Persists for Long Finite Sequences

Mike Steel[1] *and Frederick A. Matsen*

Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

The "star paradox" in phylogenetics is the tendency for a particular resolved tree to be sometimes strongly supported even when the data is generated by an unresolved ("star") tree. There have been contrary claims as to whether this phenomenon persists when very long sequences are considered. This note settles one aspect of this debate by proving mathematically that the chance that a resolved tree could be strongly supported stays above some strictly positive number, even as the length of the sequences becomes very large.

Introduction

Two recent papers (Yang and Rannala 2005; Lewis et al. 2005) highlighted a phenomenon that occurs when sequences evolve on a tree that contains a polytomy—in particular a 3-taxon unresolved rooted tree. As longer sequences are analyzed using a Bayesian approach, the posterior probability of the trees that give the different resolutions of the polytomy do not converge on relatively equal probabilities—rather a given resolution can sometimes have a posterior probability close to one. This has been called the "star paradox" because the data evolved on an unresolved tree, and thus a high posterior for a particular resolved tree must be artifactual. In response Kolaczkowski and Thornton (2006) investigated this phenomenon further, providing some interesting simulation results and offering an argument that seems to suggest that for very long sequences the tendency to sometimes infer strongly supported resolutions suggested by the earlier papers would disappear with sufficiently long sequences. As part of their case, the authors use the expected site frequency patterns to simulate the case of infinite length sequences, concluding that "with infinite length data, posterior probabilities give equal support for all resolved trees, and the rate of false inferences falls to zero." Of course these findings concern sequences that are effectively infinite, and, as is well known in statistics, the limit of a function of random variables (in this case site pattern frequencies for the first n sites) does not necessarily equate with the function of the limit of the random variables. Accordingly Kolaczkowski and Thornton offer this appropriate cautionary qualification of their findings:

"Analysis of ideal data sets does not indicate what will happen when very large data sets with some stochastic error are analyzed, but it does show that when infinite data are generated on a star tree, posterior probabilities are predictable, equally supporting each possible resolved tree."

Yang and Rannala (2005) had attempted to simulate the large sample posterior distribution but ran into numerical problems and commented that it was "unclear" what the limiting distribution on posterior probabilities was as n became large.

In particular, all of the aforementioned papers have left open an interesting statistical question, which this short note formally answers—namely, does the Bayesian posterior probability of the 3 resolutions of a star tree on 3 taxa converge to 1/3 as the sequence length tends to infinity? That is, does the distribution on posterior probabilities for "very long sequences" converge on the distribution for infinite length sequences? We show that for most reasonable priors it does not. Thus the star paradox does not disappear as the sequences get longer.

As noted by Yang and Rannala (2005) and Lewis et al. (2005), one can demonstrate such phenomena more easily for related simpler processes such as coin tossing (particularly if one imposes a particular prior). Here, we avoid this simplification as such results do not rigorously establish corresponding phenomena in the phylogenetic setting, which in contrast to coin tossing involves considering a parameter space of dimension greater than 1 (moreover, as we will see there is a complication that arises in the phylogenetic problem that is entirely absent from the coin-tossing problem). We also frame our main result so that it applies to a fairly general class of priors. Whether or not the star paradox is a practical concern for biologists is likely to depend heavily on the data, the priors, and the methods used to establish Bayesian posterior probabilities. The purpose of this paper is merely to indicate that from a mathematical perspective there is no reason to think that the star paradox will automatically vanish given long enough sequences. Some further comments and earlier references on the phenomenon have been described in the recent review paper by Alfaro and Holder (2006, p. 35–36).

Analysis of the Star Tree Paradox for 3 Taxa

On tree T_1 (in fig. 1), let $p_i = p_i(t_0, t_1)$, $i = 0, 1, 2, 3$, denote the probabilities of the 4 site patterns (xxx, xxy, yxx, xyx, respectively) under the simple 2-state symmetric Markov process (the argument extends to more general models, but it suffices to demonstrate the phenomena for this simple model). From equation (2) of Yang and Rannala (2005) we have

$$p_0(t_0, t_1) = \frac{1}{4}(1 + \mathrm{e}^{-4t_1} + 2\mathrm{e}^{-4(t_0+t_1)}),$$

$$p_1(t_0, t_1) = \frac{1}{4}(1 + \mathrm{e}^{-4t_1} - 2\mathrm{e}^{-4(t_0+t_1)}),$$

and

$$p_2(t_0, t_1) = p_3(t_0, t_1) = \frac{1}{4}(1 - \mathrm{e}^{-4t_1}).$$

It follows by elementary algebra that for $i = 2, 3$,

[1] Present address: Biomathematics Research Centre, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.

Key words: phylogenetic trees, Bayesian statistics, star trees.

E-mail: m.steel@math.canterbury.ac.nz.

Mol. Biol. Evol. 24(4):1075–1079. 2007
doi:10.1093/molbev/msm028
Advance Access publication February 13, 2007

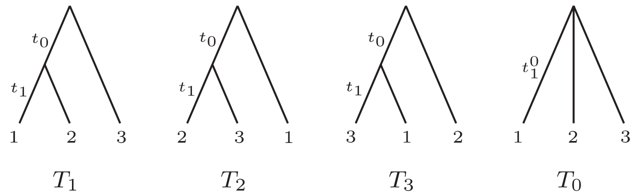

FIG. 1.—The 3 resolved rooted phylogenetic trees on 3 taxa T_1, T_2, T_3, and the unresolved 'star' tree on which the sequences are generated T_0.

$$\frac{p_1(t_0,t_1)}{p_i(t_0,t_1)} \geq 1 + 2e^{-4t_1}(1-e^{-4t_0}), \quad (1)$$

and thus $p_1(t_0, t_1) \geq p_i(t_0, t_1)$ with strict inequality unless $t_0 = 0$ (or in the limit as t_1 tends to infinity).

To allow maximal generality, we make only minimal assumptions about the prior distribution on trees and branch lengths. Namely, we assume that the 3 resolved trees on 3 leaves (trees T_1, T_2, T_3 in fig. 1) have equal prior probability $\frac{1}{3}$ and that the prior distribution on branch lengths t_0, t_1 is the same for each tree with a smooth joint probability density function that is bounded and everywhere nonzero. This condition applies, for example, to the exponential prior discussed by Yang and Rannala (2005). Any prior that satisfies these properties we call "tame."

Let $\mathbf{n} = (n_0, n_1, n_2, n_3)$ be the counts of the different types of site patterns (corresponding to the same patterns as for the P_i's). Thus $n = \sum_{i=0}^{3} n_i$ is the total number of sites (i.e., the length of the sequences). Given a prior distribution on (t_0, t_1) for the branch lengths of T_i (for $i = 1, 2, 3$) let $\mathbb{P}[T_i|\mathbf{n}]$ be the posterior probability of tree T_i given the site pattern counts $\mathbf{n}$. Now suppose the n sites are generated on a star tree T_0 with positive branch lengths. We are interested in whether the posterior probability $\mathbb{P}[T_i|\mathbf{n}]$ could be close to 1 or whether the chance of generating data with this property goes to zero as the sequence length gets very large. We show that in fact the latter possibility is ruled out by our main result, namely, the following:

Theorem 2.1

Consider sequences of length n generated by a star tree T_0 on 3 taxa with strictly positive edge length t_1^0 and let $\mathbf{n}$ be the resulting data (in terms of site pattern counts). Consider any prior on the 3 resolved trees (T_1, T_2, T_3) and their branch lengths that is tame (as defined above). For any $\epsilon > 0$, and each resolved tree T_i ($i = 1, 2, 3$), the probability that $\mathbf{n}$ has the property that

$$\mathbb{P}(T_i|\mathbf{n}) > 1 - \epsilon$$

does not converge to 0 as n tends to infinity.

Proof of Theorem 2.1

Because of the considerable details involved in the proof, we present a brief, intuitive outline of the argument.

Firstly, we will show that the probability that a star tree generates a "moderate" excess of site patterns favoring any 1 of the 3 resolved trees stays bounded above zero as n (the sequence length) goes to infinity. Here, by moderate we mean that the excess is in the order of the square root of n (the standard deviation for site pattern counts). Conditioning on this event (and some related moderate events, collectively called F_c below), we would like to show that a given tree T has unusually high posterior probability for this data; this can be simplified to showing that the ratio of the posterior probability of the given resolved tree to either of the other 2 resolved trees is large. These ratios can in turn be conveniently expressed as a ratio of expectations of 2 closely related random variables (X, Y—eq. 4). A crucial observation (based on symmetry considerations) is that X and Y are random variables determined just by T and the prior on its branch lengths (t_0, t_1)—that is, we have reduced a problem involving 3 resolved trees and their branch length priors to an analysis of 2 quantities involving a single resolved tree and its branch length prior. To analyze the ratio of expectations (in the hope of showing it is large), it helps to focus on the distribution of the random variables (P_0, P_1) induced by the prior on (t_0, t_1) (later it is helpful to consider a further derived pair of random variables (P_0, Z)). We do not need to calculate these distributions explicitly; indeed, working over the general class of priors is curiously helpful here, as it avoids the temptation to get bogged down in the detailed analysis of a particular distribution.

Considering the ratio of conditional expectations according to a distribution on (P_0, P_1) leads to a second helpful observation: if we condition on P_0 taking a particular value, say p_0, then p_0 cancels out of the ratio (eq. 13), reducing what began as a 2-dimensional problem to a family of 1-dimensional computations. At this point a complication arises that requires some care to resolve (and which does not arise in the much simpler [fair] coin-tossing problem). Namely, the ratio of conditional probabilities is not always large (e.g., if P_0 is conditioned to equal a value that approaches $\frac{1}{4}$, which corresponds to long branches [site saturation], then all 3 resolved trees have approximately the same posterior probability for any data). Nevertheless, we are assuming that the data is generated by a star tree with finite nonzero branch lengths and so the probability q_0 of the unvaried pattern (xxx) is a fixed number strictly between $\frac{1}{4}$ and 1. If P_0 were to differ from q_0 too much then (for long sequences that satisfy the moderate events mentioned above) the posterior probability of any resolved tree conditional on such an extreme P_0 value would be small, compared with a conditioned value of P_0 close to q_0—this is, informally, what Claim (i) in the proof below says. Moreover, if P_0 is conditioned to equal a value close to q_0, then the ratio of conditional expectations can be shown to be large. This is essentially Claim (ii) in the proof below. These 2 claims can be combined by Lemma 2.2 to handle the complication described, and thereby, establish what we required, namely, that the ratio of (unconditional) expectations is large.

We now proceed to the formal details—the proofs of Claims (i) and (ii) and Lemma 2.2 are deferred to the Appendix.

Consider the star tree T_0 with given branch lengths t_1^0 (as in fig. 1). Let (q_0, q_1, q_2, q_3) denote the probability of the 2 types of site patterns generated by T_0 with these branch lengths. Note that $q_1 = q_2 = q_3$ and so $q_0 = 1 - 3q_1$). Suppose we generate n sites on this tree, and let n_0, n_1, n_2, n_3 be the counts of the different types of site patterns (corresponding to the p_i's). Let $\Delta_0 := (n_0 - q_0 n)/\sqrt{n}$ and for $i = 1, 2, 3$, let

$$\Delta_i := \frac{n_i - \frac{1}{3}(n - n_0)}{\sqrt{n}}.$$

For a constant $c > 1$, let F_c denote the event:

$$F_c : \Delta_2, \Delta_3 \in [-2c, -c] \quad \text{and} \quad \Delta_0 \in [-c, c].$$

Notice that F_c implies $\Delta_1 \in [2c, 4c]$ because $\Delta_1 + \Delta_2 + \Delta_3 = 0$. By standard stochastic arguments (based on the asymptotic approximation of the multinomial distribution by the multinormal distribution) event F_c has probability at least some value $\delta' = \delta'(c) > 0$ for all n sufficiently large (relative to c).

Given the data $\mathbf{n} = (n_0, n_1, n_2, n_3)$ write $\mathbb{P}(n_0, n_1, n_2, n_3 | T_i, t_0, t_1)$ for the probability of $\mathbf{n}$ assuming the data was generated on tree T_i with branch lengths t_0, t_1. The assumption of equality of priors across T_1, T_2, and T_3 implies that

$$\mathbb{P}(n_0, n_1, n_2, n_3 | T_2, t_0, t_1) = \mathbb{P}(n_0, n_2, n_3, n_1 | T_1, t_0, t_1), \quad (2)$$

and

$$\mathbb{P}(n_0, n_1, n_2, n_3 | T_3, t_0, t_1) = \mathbb{P}(n_0, n_3, n_1, n_2 | T_1, t_0, t_1). \quad (3)$$

Now, as (t_0, t_1) are random variables with some prior density, when we view p_0, p_1, p_2, p_3 as random variables by virtue of their dependence on (t_0, t_1), we will write them as P_0, P_1, P_2, P_3 (note that Yang and Rannala [2005] use P_i differently). With this notation, the posterior probability of T_1 conditional on $\mathbf{n}$ can be written as

$$\mathbb{P}(T_1 | \mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_2} P_3^{n_3}],$$

where $p(\mathbf{n})$ is the posterior probability of $\mathbf{n}$ (assuming that data is generated on one of the resolved trees chosen with equal probability) and $\mathbb{E}_1$ denotes expectation with respect to the prior for t_0, t_1 on T_1. Moreover, because $P_2 = P_3$, we can write this as $\mathbb{P}(T_1|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_2+n_3}]$. By equation (2) and equation (3), we have

$$\mathbb{P}(T_2|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_2} P_2^{n_1+n_3}] \quad \text{and}$$
$$\mathbb{P}(T_3|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_3} P_2^{n_1+n_2}],$$

where again expectation is taken with respect to the prior for t_0, t_1 on T_1. Consequently,

$$\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_2|\mathbf{n})} = \frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]}, \quad (4)$$

where

$$X = P_0^{n_0} P_1^{n_1} P_2^{n_2+n_3} \quad \text{and} \quad Y = P_0^{n_0} P_1^{n_2} P_2^{n_1+n_3}.$$

As will be shown later, it suffices to demonstrate that the ratio in equation (4) can be large with nonvanishing probability in order to obtain the conclusion of the theorem. In order to do so, we use the following lemma, whose proof is provided in the Appendix.

Lemma 2.2

Let X, Y be nonnegative continuous random variables, dependent on a third random variable Λ that takes values in an interval I = [a, b]. Suppose that for some interval I_0 strictly inside I, and $I_1 = I - I_0$ the following inequality holds:

$$\mathbb{E}[Y|\Lambda \in I_0] \geq \mathbb{E}[Y|\Lambda \in I_1] \quad (5)$$

and that for some constant $B > 0$, and all $\lambda \in I_0$,

$$\frac{\mathbb{E}[X|\Lambda = \lambda]}{\mathbb{E}[Y|\Lambda = \lambda]} \geq B. \quad (6)$$

Then, $\frac{\mathbb{E}[X]}{\mathbb{E}[Y]} \geq B \times \mathbb{P}(\Lambda \in I_0)$.

To apply this lemma, select a value $s > 0$ so that $\frac{1}{4} + s < q_0 < 1 - s$, and let $I_0 = [q_0 - s, q_0 + s]$. Then let $I = [\frac{1}{4}, 1]$ and $I_1 = I - I_0 = [\frac{1}{4}, q_0 - s) \cup (q_0 + s, 1]$.

Claim:

Let $c > 1$. Then conditional on the data $\mathbf{n} = (n_0, n_1, n_2, n_3)$ satisfying F_c, and for n sufficiently large, the following two inequalities hold:

(i) $\mathbb{E}_1[Y|P_0 \in I_0] \geq \mathbb{E}_1[Y|P_0 \in I_1]$
(ii) For all $p_0 \in I_0$, $\mathbb{E}_1[X|P_0 = p_0]/\mathbb{E}_1[Y|P_0 = p_0] \geq 6c^2$.

The proofs of these claims are given in the Appendix.

To apply these claims, select $c > 1/(\sqrt{3\epsilon\mathbb{P}(P_0 \in I_0)})$ (this is finite by the assumption that the prior on (t_0, t_1) is smooth and everywhere nonzero). Then, $6c^2 \times \mathbb{P}(P_0 \in I_0) > \frac{2}{\epsilon}$.

Combining Lemma 2.2 with Claims (i) and (ii) for this value of c we deduce that conditional on $\mathbf{n}$ satisfying F_c and for n sufficiently large,

$$\frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]} \geq 6c^2 \times \mathbb{P}(P_0 \in I_0) > \frac{2}{\varepsilon}. \quad (7)$$

As stated before, the probability that $\mathbf{n}$ satisfies F_c is at least $\delta' = \delta'(c) > 0$ for n sufficiently large, and so, by equation (4) and equation (7), we have $\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_2|\mathbf{n})} = \frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]} > \frac{2}{\epsilon}$. Similarly, $\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_3|\mathbf{n})} > \frac{2}{\epsilon}$. Now, because $\mathbb{P}(T_1|\mathbf{n}) + \mathbb{P}(T_2|\mathbf{n}) + \mathbb{P}(T_3|\mathbf{n}) = 1$ it follows that, for n sufficiently large, and conditional an event of probability at least $\delta' > 0$, that $\mathbb{P}(T_1|\mathbf{n}) > 1 - \varepsilon$ as claimed. This completes the proof. □

Concluding Remarks

One feature of the argument we have provided is that it does not require stipulating in advance a particular prior on the branch lengths—that is, the result is somewhat generic as it imposes relatively few conditions. Moreover, it seems possible to weaken these even further. For example, the requirement that the prior on (t_0, t_1) be everywhere nonzero could be weakened to simply being nonzero in a neighborhood of $(0, t_1^0)$ (thereby allowing, e.g., a uniform distribution on bounded range).

An interesting open question in the spirit of this paper is to explicitly calculate the limit of the posterior density $f(P_1, P_2, P_3)$ described in (Yang and Rannala 2005). It may also be of interest to study posterior support for resolved trees when one weakens the molecular clock assumption on the star tree that generates the data. For example, one could imagine combinations of (non-clocklike) branch lengths that may lead to more frequent and/or stronger support for particular resolved trees.

Appendix: Proof of Lemma 2.2 and Claims (i) and (ii)

Proof of Lemma 2.2:

For W = X, Y we have

$$\mathbb{E}[W] = \mathbb{E}[W|\Lambda \in I_0]\mathbb{P}(\Lambda \in I_0) + \mathbb{E}[W|\Lambda \in I_1]\mathbb{P}(\Lambda \in I_1). \tag{8}$$

In particular, for $W = X$ we have: $\mathbb{E}[X] \geq \mathbb{E}[X|\Lambda \in I_0]$ $\mathbb{P}(\Lambda \in I_0)$. Note that equation (6) implies that $\mathbb{E}[X|\Lambda \in I_0] \geq B \times \mathbb{E}[Y|\Lambda \in I_0]$, so

$$\mathbb{E}[X] \geq B \times \mathbb{E}[Y|\Lambda \in I_0]\mathbb{P}(\Lambda \in I_0). \tag{9}$$

Taking $W = Y$ in equation (8) and applying equation (5) gives us

$$\mathbb{E}[Y] \leq \mathbb{E}[Y|\Lambda \in I_0](\mathbb{P}(\Lambda \in I_0) + \mathbb{P}(\Lambda \in I_1)) = \mathbb{E}[Y|\Lambda \in I_0],$$

which combined with equation (9) gives the result. □

Proof of Claim (i), $\mathbb{E}_1[Y|P_0 \in I_0] \geq \mathbb{E}_1[Y|P_0 \in I_1]$:

We will first bound $\mathbb{E}_1[Y|P_0 \in I_1]$ above. Let $\mu(n) = (q_0^{q_0} q_1^{q_1} q_2^{q_2} q_3^{q_3})^n$. Now, conditional on $\mathbf{n}$ satisfying F_c we have

$$n^{-1}\log(\mu(n)/Y(t_0, t_1)) = d_{KL}(q, p) + o(1), \tag{10}$$

where $p = (p_0, p_1, p_2, p_3)$ and $q = (q_0, q_1, q_2, q_3)$, and d_{KL} denotes Kullback–Leibler distance. Now, $d_{\mathrm{KL}}(q, p) \geq \frac{1}{2}\|q - p\|_1^2 \geq \frac{1}{2}|q_0 - p_0|^2$ (the first inequality is a standard one in probability theory). In particular, if $p_0 \in I_1$, then $|q_0 - p_0| > s > 0$. Moreover,

$$\mathbb{E}_1[Y|P_0 \in I_1] \leq \max\{Y(t_0, t_1) : p_0(t_0, t_1) \in I_1\}.$$

The right hand side can then be bounded above by rearranging equation (10) and using the lower bound on the Kullback–Leibler distance, giving

$$\begin{aligned}\mathbb{E}_1[Y|P_0 \in I_1] &\leq \max\{Y(t_0, t_1) : p_0(t_0, t_1) \in I_1\} \\ &< \mu(n)\mathrm{e}^{-\frac{1}{2}s^2 n + o(n)}.\end{aligned} \tag{11}$$

In the reverse direction, we have

$$\mathbb{E}_1[Y|P_0 \in I_0] \geq A(n)B(n),$$

where

$$A(n) = \min\{Y(t_0, t_1) : (t_0, t_1) \in [0, n^{-1}] \times [t_1^0, t_1^0 + n^{-1}]\}$$

and

$$B(n) = \mathbb{P}((t_0, t_1) \in [0, n^{-1}] \times [t_1^0, t_1^0 + n^{-1}]).$$

Now,

$$A(n)/\mu(n) = \left(\frac{p_0^{q_0} p_1^{q_1} p_2^{2q_1}}{q_0^{q_0} q_1^{3q_1}}\right)^n \times \left(p_0^{\Delta_0} p_1^{\Delta_2 - \frac{1}{3}\Delta_0} p_2^{\Delta_1 + \Delta_3 - \frac{2}{3}\Delta_0}\right)^{\sqrt{n}} \tag{12}$$

for (p_0, p_1, p_2) determined by some $(t_0, t_1) \in [0, n^{-1}] \times [t_1^0, t_1^0 + n^{-1}]$. Now, the first term of the product in equation (12) converges to a constant as n grows (because $p_0 - q_0, p_1 - q_1$, and $p_2 - q_1$ are each of order n^{-1}), whereas the condition F_c ensures that the second term decays no faster than $\mathrm{e}^{-C_1\sqrt{n}}$ for a constant C_1. Thus, $A(n) \geq C_2\mu(n)\mathrm{e}^{-C_1\sqrt{n}}$ for a positive constant C_2. The term $B(n)$ is asymptotically proportional to n^{-2}. Summarizing, for a constant $C_3 > 0$ (dependent just on t_1^0)

$$\mathbb{E}_1[Y|P_0 \in I_0] \geq C_3\mu(n)n^{-2}\mathrm{e}^{-C_1\sqrt{n}},$$

which combined with equation (11) establishes Claim (i) for n sufficiently large. □

In order to prove Claim (ii), we need some preliminary results.

Lemma 3.1

Let $\eta < 1$. Then for each $x > 0$ there exists a value $K = K(x) < \infty$ that depends continuously on x so that the following holds. For any continuous random variable Z on [0, 1] with a smooth density function f that satisfies $f(1) \neq 0$ and $|f'(z)| < B$ for all $z \in (\eta, 1]$, we have

$$k \times \frac{(\mathbb{E}[Z^k] - \mathbb{E}[Z^{k+1}])}{\mathbb{E}[Z^k]} \geq \frac{1}{2}$$

for all $k \geq K(\frac{B}{f(1)})$.

Proof.

Let $t_k = 1 - \frac{1}{\sqrt{k}}$. Then,

$$\mathbb{E}[Z^k] = \int_0^{t_k} t^k f(t)\mathrm{d}t + \int_{t_k}^1 t^k f(t)\mathrm{d}t.$$

Now,

$$0 \leq \int_0^{t_k} t^k f(t)\mathrm{d}t \leq t_k^k \sim \mathrm{e}^{-\sqrt{k} - 1/2},$$

where ~ denotes asymptotic equivalence (i.e., $f(k) \sim g(k)$ if $\lim_{k\to\infty} f(k)/g(k) = 1$). Using integration by parts,

$$\int_{t_k}^1 t^k f(t)\mathrm{d}t = \frac{1}{k+1}t^{k+1}f(t)\Big|_{t_k}^1 - \frac{1}{k+1}\int_{t_k}^1 t^{k+1}f'(t)\mathrm{d}t.$$

Now, provided $k > (1 - \eta)^{-2}$, we have $t_k > \eta$ and so the absolute value of the second term on the right is at most $\frac{B}{k+1}\int_{t_k}^1 t^{k+1}\mathrm{d}t = \frac{B}{(k+1)(k+2)}(1 - t_k^{k+2})$. Consequently, $|\mathbb{E}[Z^k] - (f(1)/k+1)|$ is bounded above by B times a term of order k^{-2}. A similar argument, again using integration by parts, shows that $|k(\mathbb{E}[Z^k] - \mathbb{E}[Z^{k+1}]) - (f(1)/k+1)|$ is bounded above by B times a term of order k^{-2}, and the lemma now follows by some routine analysis. □

Lemma 3.2

Let $y = (1+2x)(1-x)^2$. Then, for $x \in [0,1)$ and $m \geq 3$ we have

$$\left(\frac{1+2x}{1-x}\right)^m \geq m^2(1-y).$$

Proof.

$$\left(\frac{1+2x}{1-x}\right)^m = \left(1+\frac{3x}{1-x}\right)^m \geq \frac{m(m-1)}{2}\left(\frac{3x}{1-x}\right)^2 \geq \frac{9m(m-1)x^2}{2},$$

and $m^2(1-y) = m^2(3x^2-2x^3) \leq 3m^2x^2$, and for $m \geq 3$ this upper bound is less or equal to the lower bound in the previous expression. □

Proof of Claim (ii), for all $p_0 \in I_0$, $\mathbb{E}_1[X|P_0{=}p_0]/\mathbb{E}_1[Y|P_0{=}p_0] \geq 6c^2$:

Write $\mathbb{E}_1[W|p_0]$ as shorthand for $\mathbb{E}[W|P_0{=}p_0]$. Note that, for any r, $s > 0$, $\mathbb{E}_1[P_0^{n_0}P_1^rP_2^s|p_0]{=}p_0^{n_0}\mathbb{E}_1[P_1^rP_2^s|p_0]$. Consequently, if we let $k{=}k(n){=}\frac{1}{3}(n-n_0)$ then, by definition of the Δ_i's,

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} = \frac{\mathbb{E}_1[(P_1P_2^2)^k \times (P_1^{\Delta_1}P_2^{\Delta_2+\Delta_3})^{\sqrt{n}}|p_0]}{\mathbb{E}_1[(P_1P_2^2)^k \times (P_1^{\Delta_2}P_2^{\Delta_1+\Delta_3})^{\sqrt{n}}|p_0]}. \qquad (13)$$

Now, conditional on $\mathbf{n}$ satisfying F_c (and because $P_1 \geq P_2$) the following 2 inequalities hold (recalling that $\Delta_1+\Delta_2+\Delta_3 = 0$),

$$P_1^{\Delta_1}P_2^{\Delta_2+\Delta_3} = \left(\frac{P_1}{P_2}\right)^{\Delta_1} \geq \left(\frac{P_1}{P_2}\right)^{2c} \quad \text{and}$$
$$P_1^{\Delta_2}P_2^{\Delta_1+\Delta_3} = \left(\frac{P_1}{P_2}\right)^{\Delta_2} \leq 1.$$

Applying this to equation (13) gives:

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq \frac{\mathbb{E}_1\left[(P_1P_2^2)^k \times \left(\frac{P_1}{P_2}\right)^{2c\sqrt{n}}\middle|p_0\right]}{\mathbb{E}_1[(P_1P_2^2)^k|p_0]}. \qquad (14)$$

Let $U{=}(P_1-P_2)/(1-P_0)$, which takes values between 0 and 1 because $P_1 \geq P_2$. Because $P_1 + 2P_2 = 1 - P_0$, we can write $P_1 = \frac{1}{3}(1{+}2U)(1-P_0)$ and $P_2 = \frac{1}{3}(1-U)(1-P_0)$.

Thus, $P_1P_2^2 = \frac{1}{27}(1{+}2U)(1-U)^2 \quad (1-P_0)^3$ and $\frac{P_1}{P_2}{=}\frac{1+2U}{1-U}$. Substituting these into equation (14), letting $Z = (1+2U)(1-U)^2$ and noting that $\sqrt{n} \geq \sqrt{3k}$ gives

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq \frac{\mathbb{E}_1\left[Z^k \times \left(\frac{1+2U}{1-U}\right)^{2c\sqrt{3k}}\middle|p_0\right]}{\mathbb{E}_1[Z^k|p_0]}.$$

Thus, by Lemma 3.2, (taking $x{=}U, y{=}Z, m{=}2c\sqrt{3k}$) we obtain, for $m \geq 3$,

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq 12c^2k \times \frac{(\mathbb{E}_1[Z^k|p_0] - \mathbb{E}_1[Z^{k+1}|p_0])}{\mathbb{E}_1[Z^k|p_0]}. \qquad (15)$$

The mapping $(t_0, t_1) \mapsto (P_0, Z)$ is a smooth invertible mapping between $(0,\infty)^2$ and its image within $(\frac{1}{4}, 1) \times (0,1)$. Notice that Z becomes concentrated about 1 whenever P_0 approaches $\frac{1}{4}$. However, for p_0 in the interval I_0 the conditional density $f(Z|P_0 = p_0)$ is smooth, bounded away from 0, and its first derivative is also uniformly bounded above over this interval. Consequently, we may apply Lemma 3.1 (noting that the condition that $\mathbf{n}$ satisfies F_c ensures that $k(n) \geq \frac{1}{4}n - o(n)$) to show that for n sufficiently large the following inequality holds for all $p_0 \in I_0$,

$$k \times \frac{(\mathbb{E}_1[Z^k|p_0] - \mathbb{E}_1[Z^{k+1}|p_0])}{\mathbb{E}_1[Z^k|p_0]} \geq \frac{1}{2}.$$

Applying this to equation (15) gives $\mathbb{E}_1[X|p_0]/\mathbb{E}_1[Y|p_0] \geq 6c^2$ as claimed. This completes the proof of Claim (ii).

Acknowledgments

M.S. thanks Ziheng Yang for suggesting the problem of computing the limiting distribution of posterior probabilities for 3-taxon trees. We also thank Paul O. Lewis and the anonymous referee for their helpful comments. This work is funded by the Allan Wilson Centre for Molecular Ecology and Evolution.

Literature Cited

Alfaro ME, Holder MT. 2006. The posterior and prior in Bayesian phylogenetics. Annu Rev Evol Syst. 37:19–42.

Kolaczkowski B, Thornton JW. 2006. Is there a star tree paradox? Mol Biol Evol. 23:1819–1823.

Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. Syst Biol. 54(2): 241–253.

Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst Biol. 54(3): 455–470.

Herve Philippe, Associate Editor

Accepted January 31, 2007