Letter to Editor

Consistency of Bayesian inference of resolved phylogenetic trees

ABSTRACT

Bayesian inference is now a leading technique for reconstructing phylogenetic trees from aligned sequence data. In this short note, we formally show that the maximum posterior tree topology provides a statistically consistent estimate of a fully resolved evolutionary tree under a wide variety of conditions. This includes the inference of gene trees from aligned sequence data across the entire parameter range of branch lengths, and under general conditions on priors in models where the usual 'identifiability' conditions hold. We extend this to the inference of species trees from sequence data, where the gene trees constitute 'nuisance parameters', as in the program *BEAST. This note also addresses earlier concerns raised in the literature questioning the extent to which statistical consistency for Bayesian methods might hold in general.

1. Introduction

Bayesian inference has become a mainstream approach for inferring phylogenetic tree topology from aligned DNA sequence data (Lemey et al., 2009). The approach has a number of desirable features, and there exist powerful software packages for analysing genetic sequence data in this way. At the same time, some potential theoretical limitations of Bayesian phylogenetics have been identified and studied. These include potential problems with the convergence of MCMC-based Bayesian methods (Mossel and Vigoda, 2005), and properties that appear to be surprising at first, such as the Bayesian star 'paradox' (Steel and Matsen, 2007; Susko, 2008; Yang, 2007).

A further property of Bayesian phylogentic inference was raised in a simulation study of Kolackzkowski and Thornton (2009), suggesting that Bayesian methods applied to unresolved four-leaf trees (with a zero-length interior edge) with certain combinations of long/short pendant branches tended to show increasing bias towards one of the three particular resolved trees as the sequence length increased. By contrast, maximum likelihood was found to favour each of the three resolutions equally. Kolackzkowski and Thornton (2009) initially suggested the possibility that for data generated by a resolved four-leaf tree with a certain combination of short and long edges, Bayesian inference might even be statistically inconsistent (i.e. the tree with the highest posterior probability for the data being different from the tree that generated the data, with a probability that does not tend to zero as the sequence length grows) even for models for which maximum likelihood is known to be statistically consistent (Chang, 1996). While Kolackzkowski and Thornton (2009) stepped back from this suggestion in a subsequent correction to their original paper, the issue drew attention to a lack of a formal proof of the statistical consistency of Bayesian inference for in molecular phylogenetics. We provide this here by establishing a more general result that includes the phylogenetic setting as a particular case.

This enhanced generality serves a further purpose, as it allows us to establish formally the statistical consistency of Bayesian species tree estimation directly from sequence data where the gene trees (and their branch lengths) are treated as further 'nuisance parameters' (as in the program *BEAST, Heled and Drummond, 2010).

While it might be possible that these results could be derived from other theoretical results in Bayesian statistics, we provide here a self-contained and essentially elementary proof that is tailored towards easy application in the phylogenetic setting. This follows the spirit of Joseph Chang's tailored version of Wald's theorem that provided a convenient tool to check and establish the consistency of maximum likelihood in phylogenetics (Chang, 1996), and which curtailed an unproductive debate in the literature about whether the detailed theoretical assumptions of Wald's original theorem applied.

2. A general result

Consider the general problem of identifying a discrete parameter lying in an arbitrary finite set A from a sequence of independent and identically distributed (i.i.d.) observations that take values in an arbitrary finite set U. Suppose further that the probability distribution on U is determined not just by the discrete parameter $a \in A$ but also by some additional (nuisance) parameters. In this paper, we will assume that these additional parameters are continuous, and we denote the parameter space associated with each discrete parameter $a \in A$ by $\Theta(a)$. We assume throughout that $\Theta(a)$ is an open subset of some Euclidean space.

In the usual phylogenetic setting, A is the set of fully resolved (binary) phylogenetic tree topologies on a given leaf set, U is the set of possible site patterns, and the parameter set $\Theta(a)$ specifies, for the tree topology a the branch lengths of the tree each of which lies in the range $(0, \infty)$, and possibly other parameters relevant to the model. Thus, if we are only concerned with branch lengths, and trees are unrooted, then $\Theta(a) = (0, \infty)^{2n-3}$ where n is the number of leaves of tree a. The trees in A may be either rooted or unrooted, and for reconstruction we estimate the same type of

tree (thus in the rooted case, the branch lengths are assumed to be ultrametric).

Returning to the general set-up, let $p_{(a,\theta)}$ denote the probability distribution on some finite set U determined by the discrete-continuous parameter pair (a, θ). Suppose we have a discrete (prior) probability distribution π on A, and, for each $a \in A$, a continuous (prior) probability distribution on $\Theta(a)$ with a probability density function $f_a(\theta)$. We will suppose that the following conditions hold for all $a \in A$:

(C1) $\pi(a) > 0$;
(C2) the density $f_a(\theta)$ is continuous, bounded and nonzero on $\Theta(a)$;
(C3) the function $\theta \mapsto p_{(a,\theta)}(u)$ is continuous and nonzero on $\Theta(a)$ for each $u \in U$;
(C4) for all $\theta \in \Theta(a)$, and all $b \neq a$, we have:
$\inf_{\theta' \in \Theta(b)} d(p_{(a,\theta)}, p_{(b,\theta')}) > 0$.

In (C4) and henceforth, d denotes the L_1 metric – that is, for any two probability distributions p, q on U: $d(p, q) := \sum_{u \in U} |p(u) - q(u)|$.

In the phylogenetic setting, if π is any of the usual nonzero priors on binary phylogenetic trees (e.g. the uniform 'proportional to distinguishable arrangements' or PDA distribution, or the Yule distribution), then condition (C1) is satisfied. If we take the usual exponential prior on branch lengths then condition (C2) is satisfied. For all Markov processes on trees, condition (C3) holds (the nonzero condition holds, since in any tree with pendant edges of positive lengths all site patterns have a strictly positive probability). Finally, for all models for which identifiability holds (e.g. the general time-reversible (GTR) model or any submodel down to the highly restrictive Jukes–Cantor model), condition (C4) holds (see e.g. Steel and Székely, 2009; a specific lower bound on d for the two-state symmetric model is provided via Lemma 7.3 of Steel and Székely, 2007).

Now, suppose we are given a sequence $\mathbf{u} = (u_1, \dots, u_k) \in U^k$ generated i.i.d. by some unknown pair (a, θ) and we wish to identify the discrete parameter (a) from $\mathbf{u}$ given prior densities on A and the continuous parameters. The *maximum a-posteriori* (MAP) estimator selects the element $b \in A$ that maximizes the posterior probability of b given $\mathbf{u}$ – that is, it maximizes $\pi(b) \mathbb{E}_{\theta'}[\mathbb{P}(\mathbf{u}|b, \theta')]$, where

$$\mathbb{P}(\mathbf{u}|b, \theta') = \prod_{i=1}^{k} p_{(b,\theta')}(u_i), \tag{1}$$

which is the probability of generating the sequence of i.i.d. observations $(u_1, \dots, u_k)$ from the underlying parameters (b, θ'), and where $\mathbb{E}_{\theta'}$ refers to taking expectation with respect to the prior probability distribution on $\Theta(b)$.

Let $P(a, \theta, k)$ denote the probability that, for a sequence $u_1, \dots, u_k$ generated i.i.d. by (a, θ), the MAP estimator correctly selects a. The following theorem establishes a sufficient condition for the statistical consistency of the MAP estimator in this context.

Theorem 1. *Provided conditions* (C1)–(C4) *hold for all $a \in A$, then*

$$\lim_{k \to \infty} P(a, \theta, k) = 1$$

for all $a \in A$, and $\theta \in \Theta(a)$.

Proof. Our proof relies on a general but technical lemma, the proof of which we defer to the Appendix. □

Lemma 2. *For any $\epsilon_1, \epsilon_2 > 0$ there exists a value $\delta > 0$ for which the following holds: for any finite set U, and any four probability distributions p, q, r, s on U that satisfy the three conditions:*

(i) $d(p, q) \geq \epsilon_1$;
(ii) *for all $u \in U$ with $r(u) > 0$, $p(u) \geq \epsilon_2$ and $q(u) > 0$;*
(iii) $d(p, r) < \delta$ *and* $d(p, s) < \delta$;

the quantity $Q = \sum_{u \in U: r(u) > 0} r(u) \log(s(u)/q(u))$ is well defined (i.e. logarithms are applied to positive quantities) and $Q \geq \frac{1}{3}\epsilon_1^2$.

2.1. Application to the proof of Theorem 1

To apply Lemma 2 we need to define the quantities mentioned by it, and we will do this in the order p, s then q, r followed by ϵ_1 and ϵ_2. Notice first that the statement of Lemma 2 is sufficiently general to allow (but not require) for q, r and s to depend on the data (i.e. to be random variables), as will be the case in our application of the lemma. This causes no problem for the argument, as we remark at the end of the proof.

We suppose throughout that the sequence $\mathbf{u} = u_1, \dots, u_k$ is generated i.i.d. by (a, θ_0) where θ_0 is any particular element of $\Theta(a)$. Then the MAP estimator will correctly select a from $\mathbf{u}$ if and only if the *Bayes Factor* defined by

$$BF_{a/b} = \frac{\pi(a)\mathbb{E}_\theta[\mathbb{P}(\mathbf{u}|a, \theta)]}{\pi(b)\mathbb{E}_{\theta'}[\mathbb{P}(\mathbf{u}|b, \theta')]}$$

is strictly greater than 1 for all $b \neq a$. By the Bonferroni inequality, it suffices to show that for each $b \neq a$ the probability that $\mathbf{u}$ is such that $BF_{a/b} > 1$ tends to 1 as k grows. To achieve this we first observe that $BF_{a/b} = (\pi(a)/\pi(b)) \cdot R_{a/b}$ where

$$R_{a/b} := \frac{\mathbb{E}_\theta[\mathbb{P}(\mathbf{u}|a, \theta)]}{\mathbb{E}_{\theta'}[\mathbb{P}(\mathbf{u}|b, \theta')]}, \tag{2}$$

and where $\pi(a)/\pi(b)$ is finite and strictly positive by (C1). Thus, it suffices to show that, for each $b \neq a$ and for every finite constant M, the inequality $R_{a/b} > M$ holds with a probability that tends to 1 as $k \to \infty$. We will establish this inequality by providing an explicit lower bound to the numerator of $R_{a/b}$ and an explicit upper bound to the denominator of $R_{a/b}$, and showing that, with probability tending to 1 as k grows, their ratio exceeds M.

Before describing the lower bound, observe that we can re-write Eq. (1) as follows:

$$\mathbb{P}(\mathbf{u}|b, \theta) = \prod_{u \in U} p_{(b,\theta)}(u)^{n_u}, \tag{3}$$

where, for each $u \in U$,

$$n_u := |\{i : u_i = u\}|.$$

For the lower bound on the numerator of $R_{a/b}$, consider the subset N_τ of $\Theta(a)$ consisting of a closed ball centered on θ_0 and of radius $\tau > 0$. Note that we can always select a sufficiently small value of $\tau > 0$ for which $N_\tau \subset \Theta(a)$ by the assumption that $\Theta(a)$ is an open subset of some Euclidean space. Letting $\mu(N_\tau) = \int_{N_\tau} f_a(\theta) \, d\theta > 0$ we have

$$\mathbb{E}_\theta[\mathbb{P}(\mathbf{u}|a, \theta)] = \int_{\Theta(a)} \mathbb{P}(\mathbf{u}|a, \theta) f_a(\theta) \, d\theta \geq \int_{N_\tau} \mathbb{P}(\mathbf{u}|a, \theta) f_a(\theta) \, d\theta,$$

and so

$$\mathbb{E}_\theta[\mathbb{P}(\mathbf{u}|a, \theta)] \geq \mu(N_\tau) \cdot \inf_{\theta \in N_\tau} \{\mathbb{P}(\mathbf{u}|a, \theta)\}. \tag{4}$$

2.2. Lower bound and the distributions p and s

Let $p = p_{(a,\theta_0)}$ (the generating probability distribution on the true parameters) and let s be the probability distribution of the form $p_{(a,\theta)}$ that minimizes $\mathbb{P}(\mathbf{u}|a, \theta)$ when θ is restricted to N_τ; such a distribution s exists from the compactness of N_τ and the continuity condition of (C3). Then, from (3) we have: $\inf_{\theta \in N_\tau} \{\mathbb{P}(\mathbf{u}|a, \theta)\} = \prod_{u \in U} s(u)^{n_u}$. Applying this to (4) gives

$$\mathbb{E}_\theta[\mathbb{P}(\mathbf{u}|a, \theta)] \geq \mu(N_\tau) \cdot \prod_{u \in U} s(u)^{n_u}. \tag{5}$$

2.3. Upper bound and the distributions q and r

Regarding the upper bound on the denominator of $R_{a/b}$, we have

$$\mathbb{E}_{\theta'}[\mathbb{P}(\mathbf{u}|b,\theta')] \leq \sup_{\theta' \in \Theta(b)} \{\mathbb{P}(\mathbf{u}|b,\theta')\}. \tag{6}$$

Given $\mathbf{u}$, let θ_i be a sequence of elements of $\Theta(b)$ for which $\lim_{i\to\infty}\mathbb{P}(\mathbf{u}|b,\theta_i) = \sup_{\theta' \in \Theta(b)}\{\mathbb{P}(\mathbf{u}|b,\theta')\}$.

Notice that $p_{(b,\theta_i)}$, $i \geq 1$, is a sequence in a bounded subset of Euclidean space (the probability simplex) and so, by the Bolzano–Weierstrass theorem, it has a convergent subsequence, with limit q (a probability distribution on U).

It remains to specify the fourth distribution r, which is determined purely by the data, and records the proportion of occurrences of the various outcomes. That is, for each $u \in U$ let $r(u) := (1/k)n_u$. Notice that $r = (r(u) : u \in U)$ is a (empirical) probability distribution on U (i.e. its entries are nonzero and sum to 1). In the phylogenetic setting r describes the frequency of site patterns in the data.

2.4. Combining the two bounds

Eqs. (2), (5) and (6) give:

$$R_{a/b} \geq \frac{\mu(N_\tau) \cdot \prod_{u \in U} s(u)^{n_u}}{\prod_{u \in U} q(u)^{n_u}}. \tag{7}$$

By (C3), $p(u) > 0$ for all $u \in U$, and by the continuity condition (C4), we can select $\tau > 0$ sufficiently small so that $s(u) > 0$ for all $u \in U$. Suppose there exists some $u_0 \in U$ with $r(u_0) > 0$ (i.e. $n_{u_0} \geq 1$) and with $q(u_0) = 0$. Then Eq. (7) implies that $R_{a/b} = +\infty$ and so Bayesian inference will select a over b. Otherwise we may assume that $q(u) > 0$ for all $u \in U$ for which $r(u) > 0$, in which case we can take logarithms of both sides of Eq. (7) and so obtain the fundamental inequality

$$\log(R_{a/b}) \geq \log(\mu(N_\tau)) + k \sum_{u \in U: r(u) > 0} r(u) \log\left(\frac{s(u)}{q(u)}\right). \tag{8}$$

2.5. Definitions of ϵ_1 and ϵ_2

Fix $b \in A - \{a\}$ and let $\epsilon_1 := \inf_{\theta' \in \Theta(b)}\{d(p, p_{(b,\theta')})\}$ and let $\epsilon_2 := \min\{p(u) : u \in U, r(u) > 0\}$. Notice that $\epsilon_1 > 0$ by (C4) and $\epsilon_2 > 0$ by (C3).

2.6. Completing the argument

Returning to the proof of Theorem 1, we are now in a position to apply Lemma 2. First observe that parts (i) and (ii) of Lemma 2 hold by definition of ϵ_1 and ϵ_2, respectively.

Next, observe that the event that $d(p, r) \leq \delta$ has probability converging to 1 as k grows, by the law of large numbers. Thus, with probability converging to 1 as $k \to \infty$ the first half of part (iii) of Lemma 2 holds (i.e. $d(p, r) < \delta$). Moreover, by the continuity condition in (C3), we can select $\tau > 0$ sufficiently small so that $d(p, s) < \delta$, and so the second half of part (iii) of Lemma 2 also holds.

In summary, with probability converging to 1 as k grows, the conditions of Lemma 2 are satisfied, in which case (by (8))

$$\log(R_{a/b}) \geq \log(\mu(N_\tau)) + k \cdot \tfrac{1}{3}\epsilon_1^2.$$

Thus, with probability converging to 1 as k grows, for any finite value M, $R_{a/b} > M$. By the comments following Eq. (2), this completes the proof. $\square$

Remark. In the proof, notice that only the probability distribution $p = p_{(a,\theta_0)}$ is fixed; the other three distributions r, s, q depend on

the data $\mathbf{u}$ that is generated by p. However, Lemma 2 applies over all choices of r, s, q once the ϵ_1 and ϵ_2 values have been specified, and these two ϵ_i values depend ultimately just on p by definition. $\square$

3. Inferring species trees directly from sequences with gene trees treated as 'nuisance parameters'

Consider a fully resolved species tree with branch parameters corresponding to inter-speciation times, and ancestral population sizes. Such a model induces a probability distribution on gene trees under a process of incomplete lineage sorting that is modelled by the multi-species coalescent model (Degnan and Rosenberg, 2009). Suppose we generate k independent gene trees under this process, and on each gene tree, we evolve sequence sites under a time-reversible site substitution model in which the branch lengths on the gene tree are (in expectation) an i.i.d. scalar multiple of the branch lengths in the species tree (i.e. we allow different genes to evolve at different rates, but assume that these rates are chosen independently from a given distribution).

Now, for any fully resolved species tree T and any tree T' on the same leaf set that has a different topology from T, there exists at least one triplet of taxa x, y, z, say, for which $T|\{x, y, z\} = xy|z$ and $T'|\{x, y, z\} \neq xy|z$. Under the multispecies coalescent, if T is the generating species tree, then the probability that the induced gene tree has the topology $xy|z$ is strictly greater than the probability it has one of the other two topologies (which have equal probability) (Degnan and Rosenberg, 2009). Moreover, for any time-reversible site substitution process, the probability that two taxa are both in a given state (say 0) is a continuous and strictly monotone decreasing function of the temporal separation between them (Aldous and Fill, 2010).

Consequently, if T is the generating species tree and if we let P_1 (respectively P_2) denote the probability that any given sequence site (generated by the gene tree sampled from the species tree under the multispecies coalescent) has the same state (0) for taxa x and y (respectively, for taxa x and z), then $P_1 > P_2$. However for the tree T' (for which $T'|\{x, y, z\} \neq xy|z$) we have $P_1 \leq P_2$, regardless of the branch lengths of T'. Thus (C4) holds (with $a = T$ and $b = T'$).

If we now take A to be the finite set of species tree topologies and U to be site patterns, then the conditions for Theorem 1 apply and so the posterior probability of the generating species tree converges to 1 as $k \to \infty$.

4. Concluding comments

In certain Bayesian implementations, the output tree is not the tree that is most frequently found; rather, a score is assigned to each cluster (subset of taxa) according to its frequency as a clade in the posterior distribution of trees, and a consensus tree is constructed on the clusters with the highest posterior support (Heled and Drummond, 2010). There are various options here as to how this can be implemented, but it is clear that, in general, such a tree could differ from the MAP tree on a given set of data. This raises an obvious question: Is this consensus tree constructed from the clusters with highest posterior support as clades a consistent estimator of the true species tree? In the limit, any such tree will converge on the true tree (and the MAP tree) as k (the sequence length) grows, for the following reason: Since we are assuming that the species tree T is fully resolved and that the posterior probability of T converges to 1 (with increasing k), the only clusters that will have a posterior probability greater than any positive value $\epsilon > 0$ for all k will be clades in T (and each clade in T will have posterior support approaching 1 as k grows); otherwise, if some cluster C not in T had this property, then another tree T' would exist for which the posterior probability of T' would be at

least $\epsilon' > 0$ for all k, contradicting the assumption that the posterior probability of T converges to 1 as $k \to \infty$ (we can take ϵ' to be ϵ divided by the number of fully resolved trees that contain the cluster C as a clade).

For future work, the consistency of phylogenetic questions on nonresolved trees could be of interest, as in that case condition (C4) does not hold. Condition (C4) can also fail on resolved trees, under site substitution models that allow too much flexibility in rate variation across sites, or in certain phylogenetic 'mixture models' (where various classes of sites evolve under different branch lengths).

Acknowledgments

Appendix A. Proof of Lemma 2

Proof. We will require at the outset that $\delta < \min\{\epsilon_1, \frac{1}{2}\epsilon_2\}$; later we place a third upper bound on δ. Applying the triangle inequality to conditions (ii) and (iii), with $\delta < \frac{1}{2}\epsilon_2$ implies that $r(u)$ and $s(u)$ are both at least $\frac{1}{2}\epsilon_2$ for all $u \in U$, and so Q is well-defined (i.e. logarithms are only applied to positive entries). Let $\eta = \min\{q(u) : r(u) > 0\}$. By condition (ii), $\eta > 0$. For each $u \in U$, let $\Delta_u := r(u) - s(u)$. Then

$$\sum_{u \in U} r(u) \log\left(\frac{s(u)}{q(u)}\right) = \sum_{u \in U} s(u) \log\left(\frac{s(u)}{q(u)}\right) + \sum_{u \in U} \Delta_u \log\left(\frac{s(u)}{q(u)}\right). \quad (9)$$

Now, the first term on the right hand-side of (9) is simply the Kullback–Leibler separation of s and q and, by Pinsker's Inequality (Cover and Thomas, 1991), this is bounded below by $\frac{1}{2}d(s,q)^2$. Moreover, by the triangle inequality, $d(s,q) \geq d(p,q) - d(p,s) \geq \epsilon_1 - \delta$ (by conditions (i) and (iii)) and since $\delta < \epsilon_1$ (so $\epsilon_1 - \delta > 0$) the first term on the right of (9) is bounded below by $\frac{1}{2}(\epsilon_1 - \delta)^2$.

Concerning the second term on the right of (9), its absolute value is bounded above by

$$\sum_{u \in U} |\Delta_u| \cdot \max_{u \in U} |\log\left(\frac{s(u)}{q(u)}\right)| = d(r,s) \cdot \max_{u \in U} |\log\left(\frac{s(u)}{q(u)}\right)|.$$

Again invoking the triangle inequality, $d(r,s) \leq d(r,p) + d(p,s) \leq 2\delta$ (by condition (iii)). Moreover, since $s(u) \geq p(u) - \delta \geq \epsilon_2 - \delta$ (by

condition (ii)) and since $\delta < \epsilon_2$ (so $\epsilon_2 - \delta > 0$) and $q(u) \geq \eta$

$$\max_{u \in U} |\log\left(\frac{s(u)}{q(u)}\right)| \leq \max_{u \in U} |\log(s(u))| + \max_{u \in U} |\log(q(u))|$$
$$\leq |\log(\epsilon_2 - \delta)| + |\log \eta|.$$

Thus we select $\delta > 0$ sufficiently small (in addition to the earlier two upper bounds on δ) so that

$$\tfrac{1}{2}(\epsilon_1 - \delta)^2 - 2\delta \cdot (|\log(\epsilon_2 - \delta)| + |\log \eta|) \geq \tfrac{1}{3}\epsilon_1^2.$$

then the bounds placed above on the terms in (9) ensure that $Q \geq \frac{1}{3}\epsilon_1^2$ as required. □

References

Aldous, D., Fill, J., 2010. Reversible Markov Chains and Random Walks on Graphs. Chapter 3, Eq. (40). ⟨http://www.stat.berkeley.edu/aldous/RWG/book.html⟩.

Chang, J., 1996. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. Mathematical Biosciences 137, 51–73.

Cover, T., Thomas, J., 1991. Elements of Information Theory. John Wiley Sons, Inc, New York.

Degnan, J., Rosenberg, N., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology and Evolution 24, 332–340.

Heled, J., Drummond, A., 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27, 570–580.

Kolackzkowski, B., Thornton, J., 2009. Long-branch attraction bias and inconsistency in Bayesian phylogenetics. PLos One 4, e7891.

Lemey, P., Salemi, M., Vandamme, A.M., 2009. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing, 2nd ed. Cambridge University Press.

Mossel, E., Vigoda, E., 2005. Phylogenetic MCMC algorithms are misleading on mixtures of trees. Science 309, 2207–2209.

Steel, M., Matsen, E., 2007. The Bayesian 'star paradox' persists for long finite sequences. Molecular Biology and Evolution 24, 1075–1079.

Steel, M., Székely, L., 2007. Teasing apart two trees. Combinatorics, Probability and Computing 16, 903–922.

Steel, M., Székely, L., 2009. Inverting random functions (iii): discrete MLE revisited. Annals of Combinatorics 13, 373–390.

Susko, E., 2008. On the distributions of bootstrap support and posterior distributions for a star tree. Systematic Biology 57, 602–612.

Yang, Z., 2007. Fair-balance paradox, star-tree paradox and Bayesian phylogenetics. Molecular Biology and Evolution 24, 1639–1655.

Mike Steel*
Allan Wilson Centre for Molecular Ecology and Evolution, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand
E-mail addresses: mathmomike@gmail.com, mike.steel@canterbury.ac.nz

* Tel.: +64 21329705.