Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Asymptotic Guarantees for Bayesian Phylogenetic Tree Reconstruction

Alisa Kirichenko, Luke J. Kelly & Jere Koskela

To cite this article: Alisa Kirichenko, Luke J. Kelly & Jere Koskela (23 Jul 2025): Asymptotic Guarantees for Bayesian Phylogenetic Tree Reconstruction, Journal of the American Statistical Association, DOI: 10.1080/01621459.2025.2485359

To link to this article: https://doi.org/10.1080/01621459.2025.2485359

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2025.2485359

Asymptotic Guarantees for Bayesian Phylogenetic Tree Reconstruction

Alisa Kirichenko[a], Luke J. Kelly[b], and Jere Koskela[a,*]

[a]Department of Statistics, University of Warwick, Coventry, UK; [b]School of Mathematical Sciences, University College Cork, Cork, Ireland

ABSTRACT

We derive tractable criteria for the consistency of Bayesian tree reconstruction procedures, which constitute a central class of algorithms for inferring common ancestry among DNA sequence samples in phylogenetics. Our results encompass several Bayesian algorithms in widespread use, such as `BEAST`, `MrBayes`, and `RevBayes`. Unlike essentially all existing asymptotic guarantees for tree reconstruction, we require no discretization or boundedness assumptions on branch lengths. Our results are also very flexible, and easy to adapt to variations of the underlying inference problem. We demonstrate the practicality of our criteria on two examples: a Kingman coalescent prior on rooted, ultrametric trees, and an independence prior on unconstrained binary trees, though we emphasize that our result also applies to nonbinary tree models. In both cases, the convergence rate we obtain matches known, frequentist results obtained using stronger boundedness assumptions, up to logarithmic factors. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received October 2022
Accepted March 2025

KEYWORDS
Bayesian methods; Coalescent process; Phylogenetic tree; Posterior consistency

1. Introduction

Reconstructing the genealogical tree of a sample of DNA sequences is a central task in phylogenetics. Genealogies cannot usually be observed directly, but they are informative about the evolutionary history of the population and encode the positive correlations among the samples brought about by common ancestry. There are many well-established methods for constructing a genealogy from a sample of sequences. Some recent contributions include parsimony- or consensus-based methods (Sayyari and Mirarab 2016; Zhang et al. 2018), maximum likelihood methods which seek to maximize the likelihood of the observed sample under a given generative model of DNA sequence diversity, possibly subject to regularization (Stamatakis 2006; Zhang, Dinh, and Matsen IV 2021), and Bayesian methods which seek to compute (or, more practically, sample from) the posterior distribution of genealogical trees given sequences observed at their leaves (Borges and Kosiol 2020; Ronquist et al. 2012; Suchard et al. 2018). For introductions to mathematical and computational phylogenetics, see Semple and Steel (2013) and Warnow (2017).

Our focus is consistency of Bayesian tree reconstruction procedures. Informally, a procedure is consistent if, in the absence of model error, it recovers the true, data-generating genealogical tree with certainty in an idealized, infinite-data limit, which we take to be the length of the observed sequences. Consistency is widely regarded as a minimal reliability criterion for Bayesian inference, and general conditions for it have been established: see for example Schwartz (1965). When consistency fails, pathological results such as posterior distributions which express ever-increasing confidence in an incorrect model can arise (Freedman and Diaconis 1983).

Consistency of tree reconstruction has been established for many consensus methods (Markin and Eulenstein 2021, and references therein), as well as maximum likelihood methods with (Dinh et al. 2018; Zhang, Dinh, and Matsen IV 2021) and without (Roch and Sly 2017) regularization. These results rest on a variety of strong assumptions: Roch and Sly (2017) assume that branch lengths take values in a discrete, finite set which is known a priori, Dinh et al. (2018) require branch lengths to be bounded above and away from zero, while Zhang, Dinh, and Matsen IV (2021) assume branch lengths are only bounded from above. Consistency in the Bayesian setting has also been established, but only for the tree topology, treating branch lengths as nuisance parameters (Steel 2013).

Our contribution is to establish Bayesian posterior consistency jointly for the tree topology and its branch lengths, regarding branch lengths as quantities of interest rather than nuisance parameters. Our approach requires no boundedness or discretization assumptions. Moreover, we demonstrate that phylogenetic tree models are amenable to standard tools of Bayesian asymptotic theory (Ghosal, Ghosh, and van der Vaart 2000), whereas the results of Steel (2013) were established using bespoke arguments for genealogical trees. Our method is highly flexible, and can be easily applied to for example tree models with nonbinary branching and hence a variable number of edges, and to constrained models such as ultrametric trees. Our results apply directly to several prominent Bayesian tree inference packages: `MrBayes` (Ronquist et al. 2012), `BEAST 2` (Bouckaert et al. 2019), and `RevBayes` (Höhna et al. 2016).

We demonstrate the utility of our result on two examples: unrooted, binary trees with independent branch lengths, and rooted, ultrametric binary trees with a Kingman coalescent prior

CONTACT Alisa Kirichenko *alicekirichenko@optiver.com* Optiver, Strawinskylaan 3095, 1077 ZX Amsterdam, The Netherlands.
*Current affiliation: School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, UK..

(Kingman 1982). In both cases we obtain a convergence rate of $(\log k)^{3/4+\beta/2}/\sqrt{k}$ for any $\beta > 1$, where k is the length of the observed DNA sequences, matching known rates for maximum likelihood procedures up to logarithmic corrections (Dinh et al. 2018, Theorem 5.2).

The rest of the article is structured as follows. In Section 2 we recall the general Bayesian posterior consistency theorem of Ghosal, Ghosh, and van der Vaart (2000, Theorem 2.1) which will serve as our main mathematical tool. We will also introduce the state space and likelihood model for our genealogical trees. In Section 3 we state and prove our main results by providing tractable consistency criteria on Bayesian tree priors. Section 4 concludes by demonstrating that two examples of standard priors satisfy our criteria, and Section 5 concludes with a discussion. Proofs of results are postponed to Sections 6 and 7.

2. Preliminaries

2.1. Bayesian Posterior Consistency

Let $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ be a collection of probability distributions with common support $\mathcal{Z}$ and corresponding probability densities $\{p_\theta\}_{\theta\in\Theta}$ relative to a given dominating measure. Let $\mathbf{Z}^{(k)} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ be a vector consisting of independent and identically distributed random variables with common distribution P_{θ_0} for $\theta_0 \in \Theta$. Finally, let $\mathbf{z}^{(k)} = (\mathbf{z}_1, \dots, \mathbf{z}_k)$ denote a vector arising as the observed value of $\mathbf{Z}^{(k)}$, and let Π be a prior distribution on $\mathcal{P}$. Under appropriate identifiability conditions, we can always define a prior Π on $\mathcal{P}$ using a prior $\tilde{\Pi}$ on the set Θ, and push it forward to $\mathcal{P}$ via the mapping $\theta \mapsto P_\theta$. Throughout the article we will slightly abuse the notation by using Π to denote both priors.

The central object in Bayesian inference is the posterior distribution, defined for $A \in \mathcal{B}(\Theta)$ as

$$\Pi(A|\mathbf{z}^{(k)}) := \frac{\int_A p_\theta(\mathbf{z}^{(k)})\Pi(\mathrm{d}\theta)}{\int_\Theta p_\theta(\mathbf{z}^{(k)})\Pi(\mathrm{d}\theta)}.$$

For a sequence $\varepsilon_k \to 0$, the posterior is said to contract around the truth P_{θ_0} at a rate ε_k if

$$\Pi(d_H(P_\theta, P_{\theta_0}) > M_k\varepsilon_k \,|\, \mathbf{Z}^{(k)}) \to 0 \tag{1}$$

as $k \to \infty$ in P_{θ_0}-probability for every sequence $M_k \to \infty$, where d_H is the Hellinger distance,

$$d_H(P_\theta, P_{\theta_0}) := \frac{1}{2}\int_{\mathcal{Z}} \left(\sqrt{p_\theta(\mathbf{z})} - \sqrt{p_{\theta_0}(\mathbf{z})}\right)^2 \mathrm{d}\mathbf{z}.$$

In order to establish posterior consistency criteria and derive contraction rates for phylogenetic tree priors, we are going to employ the general machinery first introduced by Ghosal, Ghosh, and van der Vaart (2000). More specifically, we use Theorem 8.9 from Ghosal and van der Vaart (2017). Consider the following statistical distances: for $P_{\theta_0}, P_\theta \in \mathcal{P}$, let

$$\mathrm{KL}(P_{\theta_0}, P_\theta) := \int_{\mathcal{Z}} \log\left(\frac{p_{\theta_0}(\mathbf{z})}{p_\theta(\mathbf{z})}\right) P_{\theta_0}(\mathrm{d}\mathbf{z}),$$

$$V_0(P_{\theta_0}, P_\theta) := \int_{\mathcal{Z}} \log^2\left(\frac{p_{\theta_0}(\mathbf{z})}{p_\theta(\mathbf{z})}\right) P_{\theta_0}(\mathrm{d}\mathbf{z}) - [\mathrm{KL}(P_{\theta_0}, P_\theta)]^2$$

be, respectively, the Kullback–Leibler divergence and the centered second Kullback–Leibler variation. Additionally, for a set $A \subseteq \mathcal{P}$ and $\varepsilon > 0$ let $N(\varepsilon, A, d_H)$ be the ε-covering number of A with respect to d_H. Then the following result holds.

Theorem 1 (Theorem 8.9 of Ghosal and van der Vaart 2017). Suppose there exist sets $\mathcal{P}_k \subseteq \mathcal{P}$, sequences $\varepsilon_k \to 0$ and $\zeta_k \to 0$ such that $k\zeta_k^2 \to \infty$ and $\zeta_k \leq \varepsilon_k$, as well as a constant $C > 0$ such that

$$\Pi(\mathcal{P}\backslash\mathcal{P}_k) \leq \exp(-(C+4)k\zeta_k^2), \tag{2}$$

$$\log N(\varepsilon_k/2, \mathcal{P}_k, d_H) \leq k\varepsilon_k^2, \tag{3}$$

$$\Pi(P_\theta : \mathrm{KL}(P_{\theta_0}, P_\theta) \leq \zeta_k^2, V_0(P_{\theta_0}, P_\theta) \leq \zeta_k^2) \geq \exp(-Ck\zeta_k^2). \tag{4}$$

Then the posterior $\Pi(\cdot|\mathbf{z}^{(k)})$ converges to the data-generating distribution P_{θ_0} at rate ε_k.

Remark 1. Assumption (8.2) of Ghosal and van der Vaart (2017) is omitted in the theorem statement above. That is because it is satisfied by the Hellinger distance d_H whenever the underlying collection of models $\mathcal{P}$ is identifiable (Ghosal and van der Vaart 2017, Proposition D.8). We will work in the Hellinger metric and assume identifiability throughout; see (10).

Remark 2. The premise of Theorem 1 is that the observations $\mathbf{z}^{(k)}$ are iid draws from P_{θ_0}, but the prior Π can depend on k. This fact will be essential in later sections, where we apply Theorem 1 to phylogenetic trees in settings where the number of leaves in the latent tree increases with the amount of observed data.

The main (and the strongest) condition of the theorem is the so-called prior mass condition (4), which ensures that we put sufficient prior probability mass near the true density p_{θ_0}. Here the proximity to the truth is measured through Kullback–Leibler divergence and the second Kullback–Leibler variation. The condition is usually satisfied when the prior has exponential tails. The entropy condition (3) is the other main determinant of the contraction rate. It bounds the complexity of the sieves $\mathcal{P}_k$ which contain the majority of the prior probability mass. The latter is ensured by the remaining mass condition (2).

2.2. A State Space of Trees

Consider a tree with a fixed number $n \in \mathbb{N}$ of leaves, and let $\mathcal{T}_n$ denote a finite set of possible topologies. For each $T \in \mathcal{T}_n$, let $m(T) \in \mathbb{N}$ be the number of degrees of freedom required to specify the branch lengths of T, and let

$$(\psi_T(x_1, \dots, x_{m(T)})_{(u,v)})_{(u,v)\in T} =: (\psi_T(\mathbf{x}_{1:m(T)})_{(u,v)})_{(u,v)\in T}$$

be the mapping from variables specifying those degrees of freedom, $\mathbf{x}_{1:m(T)} \in (0,\infty)^{m(T)}$, to the lengths of branches $((u,v))_{(u,v)\in T}$, where (u,v) denotes a branch between two nodes u, v of the tree T. When the length of the vector $\mathbf{x}_{1:m(T)}$ is clear from context, we will often omit the subscript and write $\mathbf{x}$ to lighten notation. We will also denote by $I(T)$ the set of internal or non-leaf nodes of T. For a given topology T, a branch $(u,v) \in T$ will be called external if either u or v is a leaf, and the set of external branches will be denoted by $Ex(T)$. Finally, let $m_n^* := \max_{T\in\mathcal{T}_n}\{m(T)\}$.

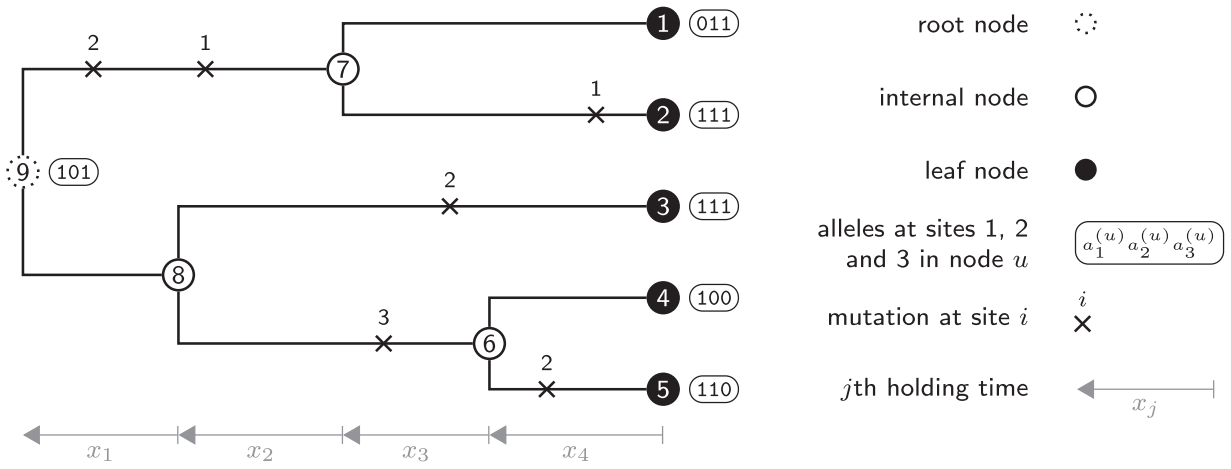


Figure 1. A binary, ultrametric tree T with $n = 5$ leaves. Time runs forward from the root $u_0 = 9$ toward the leaves. We require $m(T) = 4$ holding times, $x_1, \ldots, x_4$, to specify the branch lengths; for example, the branch from node 8 to 6 has length $\psi_T(\mathbf{x})_{(8,6)} = x_2 + x_3$. The alleles at $k = 3$ sites taking values in $H = \{0, 1\}$ are drawn independently from distribution $r(\cdot)$ at the root and mutate according to a Markov jump process with rate matrix $\mathbf{Q}$ along the branches of T. We only observe an allele at the leaves so integrate over possible states at the internal nodes $I(T) = \{6, \ldots, 9\}$ in calculating the likelihood (5).

This somewhat abstract description encompasses a broad range of tree models, such as those where the degree of tree nodes is not fixed and the number of edges varies by topology. It also encompasses both rooted and unrooted trees, and can easily incorporate constraints, such as a requirement that trees be ultrametric. Figure 1 illustrates our notation on an ultrametric tree, and examples of the state space construction are presented in Sections 4.1 and 4.2. These examples consist of binary trees, but multifurcating trees could also be considered by taking $\mathcal{T}_n$ to be, say, the set of all unrooted trees with n leaves. Then $m(T)$ can be taken to be the number of edges in $T \in \mathcal{T}_n$, which depends on the number and size of multifurcations. Similarly, ψ_T can be taken to be the identity function by imposing and ordering of edges $(u, v) \in T$ matching that of the vector $\mathbf{x}$. In this way, the model selection problem of whether a tree contains multifurcations, as studied by Zhang, Dinh, and Matsen IV (2021), can be recast as a single Bayesian inference problem with a prior distribution supporting both binary and multifurcating trees.

We assume the set $\mathcal{T}_n$, as well as the maps $m(\cdot)$ and $\psi_T(\cdot)$, are known a priori, and that ψ_T is invertible for each T. Then, our state space of trees is

$$\Theta^{(n)} := \{(T, \mathbf{x}) : T \in \mathcal{T}_n, \mathbf{x} \in (0, \infty)^{m(T)}\}.$$

2.3. A Model of Evolution on Trees

Consider a fixed tree $(T, \mathbf{x}) \in \Theta^{(n)}$. We assume that the state, or allele of each leaf is observed at $k \in \mathbb{N}$ sites. All sites evolve conditionally independently along the branches of the common genealogical tree. At a given site, the root of the tree is assigned a random allele a among a finite set of alleles H from a distribution $r(\cdot)$. Then, alleles are propagated toward leaves along branches according to the law of a Markov jump process with irreducible rate matrix $\mathbf{Q} = (Q_{\alpha,\beta} : \alpha, \beta \in H)$. These dynamics give rise to the single-site likelihood

$$P_{(T,\mathbf{x})}(\mathbf{a}) := \sum_{\mathbf{a}^* \in H^{|I(T)|}} r(a^{(u_0)}) \prod_{(u,v)\in T} \exp(\psi_T(\mathbf{x})_{(u,v)}\mathbf{Q})_{a^{(u)}a^{(v)}}, \tag{5}$$

where $\psi_T(\mathbf{x})_{(u,v)}$ is the length of the branch connecting nodes u and v, $\mathbf{a}$ is the vector of observed alleles at the n leaves, and the sum is over all non-leaf alleles in the extension, denoted $\mathbf{a}^*$, which tracks alleles at the internal nodes of the tree as well. The allele at node u is denoted by $a^{(u)}$, with u_0 corresponding to the root. Figure 1 illustrates this mutation process on a rooted tree.

The likelihood (5) depends on the position of the root, but can be extended to unrooted trees provided that the Markov jump process with rate matrix $\mathbf{Q}$ is r-reversible. In that case, Felsenstein (1981) showed that the position of the root does not affect the value of (5), and hence can be set arbitrarily. He also presented an efficient algorithm for evaluating (5), which applies to both the reversible and rooted, nonreversible cases. Likelihoods for multiple sites are products of terms of the form (5):

$$P_{(T,\mathbf{x})}(\mathbf{a}_1, \ldots, \mathbf{a}_k) := \prod_{i=1}^{k} \sum_{\mathbf{a}_i^* \in H^{|I(T)|}} r(a_i^{(u_0)}) \times \prod_{(u,v)\in T} \exp(\psi_T(\mathbf{x})_{(u,v)}\mathbf{Q})_{a_i^{(u)}a_i^{(v)}}. \tag{6}$$

In order to use Theorem 1, we need upper bounds on the Kullback–Leibler divergence and centered second Kullback–Leibler variation corresponding to (6). To that end, we show that both distances can be bounded by a Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_0$ under suitable conditions on the lengths of tree branches, when the topology is fixed. We first introduce some new notation. Let $|\mathbf{Q}|$ be the matrix whose entries are the absolute values of those of $\mathbf{Q}$. Let

$$\eta := \min_{\alpha,\beta\in H}\{Q_{\alpha,\beta} : Q_{\alpha,\beta} > 0\},$$
$$\gamma := \max_{\alpha,\beta\in H}\{|Q_{\alpha,\beta}|\},$$
$$w := \min\{j \in \mathbb{N} : \min_{\alpha,\beta}\{(|\mathbf{Q}|^j)_{\alpha,\beta}\} > 0\},$$

be the least positive entry of $\mathbf{Q}$, the entry of $\mathbf{Q}$ with maximal magnitude, and the least number of steps in which the mutation Markov chain with generator $\mathbf{Q}$ can reach any state from any other with positive probability, respectively. The last is guaranteed to exist since $\mathbf{Q}$ is a finite, irreducible matrix (Norris 1997,

chap. 1, Theorem 1.2.1). For $f > 0$, $g > 0$, and a tree topology $T \in \mathcal{T}_n$, let

$$L_f(T) := \big\{\mathbf{x} \in (0,\infty)^{m(T)} : \min_{(u,v)\in Ex(T)} \{\psi_T(\mathbf{x})_{(u,v)}\} > f\big\},$$
$$U_g(T) := \big\{\mathbf{x} \in (0,\infty)^{m(T)} : \max_{(u,v)\in T} \{\psi_T(\mathbf{x})_{(u,v)}\} < g\big\},$$

be the respective events that lengths of external branches are bounded below by f, and that the lengths of all branches are bounded above by g. The complements of these events are denoted as $L_f^c(T)$ and $U_g^c(T)$, respectively.

Lemma 1. Suppose $\psi_T(\cdot)$ is differentiable for each $T \in \mathcal{T}_n$, and that

$$\max_{T\in\mathcal{T}_n} \max_{i\in\{1,\dots,m(T)\}} \left|\frac{\partial \psi_T}{\partial x_i}(\mathbf{x})\right| =: \|\partial\psi^{(n)}\|_\infty < \infty.$$

Define

$$B_n := A\|\partial\psi^{(n)}\|_\infty (n-1)B^n := \frac{4\gamma w!}{\min_{\alpha\in H}\{r(\alpha)\}} \frac{\|\partial\psi^{(n)}\|_\infty (n-1)|H|^{n-1}}{\min_{\alpha\in H, t\geq 0}\{\exp(\mathbf{Q}t)_{\alpha\alpha}\}^{n-2}\eta^{nw}}, \quad (7)$$

where $A > 0$ and $B > 0$ are constants independent of n. Then, for $0 < f < 1/(2\eta)$,

$$\mathrm{KL}(T_0,\mathbf{x}_0;T_0,\mathbf{x}) \leq B_n f^{-nw}\|\mathbf{x}-\mathbf{x}_0\|_2, \quad (8)$$
$$V_0(T_0,\mathbf{x}_0;T_0,\mathbf{x}) \leq B_n^2 f^{-2nw}\|\mathbf{x}-\mathbf{x}_0\|_2^2, \quad (9)$$

as long as $\mathbf{x}, \mathbf{x}_0 \in L_f(T_0)$.

3. A Contraction Theorem for Posteriors on Trees

Consider a prior distribution $\Pi^{(n)}$ on $\Theta^{(n)}$, and n observed sequences at k sites with respective allele configurations $(\mathbf{a}_1,\dots,\mathbf{a}_k)$. From now on we assume the identifiability condition

$$P_{\theta_0} = P_\theta \Rightarrow \theta_0 = \theta. \quad (10)$$

Remark 3. The identifiability condition (10) is known to hold for a number of standard phylogenetic tree models (Steel and Székely 2007, 2009). A typical failure mode of (10) arises when entries of the matrix $\mathbf{Q}$ are part of the inference problem, whereupon a model only distinguishes compound products $\mathbf{Q}(\psi_T(\mathbf{x}))_{(u,v)}$ of rates and branch lengths. The typical solution is to reparameterize the model with these products directly.

Let the data-generating tree be $(T_0,\mathbf{x}_0)$, and let

$$\tilde{f}_n := \min_{(u,v)\in Ex(T_0)} \{\psi_{T_0}(\mathbf{x}_0)_{(u,v)}/2\} \wedge 1/(2\eta).$$

Theorem 2. Suppose (10) holds, and that n is either bounded as $k \to \infty$, or that if $n = n_k$ with $\lim_{k\to\infty} n_k = \infty$, then there exists an infinite-leaf tree from which the sequence of data-generating trees are obtained as nested restrictions to n_k leaves, for each k. Suppose further that there exist sequences $\varepsilon_k \to 0$ and $\zeta_k \to 0$ such that $k\zeta_k^2 \to \infty$ and $\zeta_k \leq \varepsilon_k$, sequences $f_k \to 0$ and $g_k \to \infty$ such that

$$\left(\frac{g_k}{\varepsilon_k^2 f_k^{nw}}\right)^{m_n^*} \exp(-k\varepsilon_k^2) \leq \frac{1}{|\mathcal{T}_n|(4B_n)^{m_n^*}}, \quad (11)$$

and a constant $C > 0$ such that

$$\Pi^{(n)}(\cup_{T\in\mathcal{T}_n}\{L_{f_k}^c(T) \cup U_{g_k}^c(T)\}) \leq \exp(-(C+4)k\zeta_k^2), \quad (12)$$
$$\Pi^{(n)}(T_0)\Pi^{(n)}(\mathbf{x} \in L_{\tilde{f}_n}(T_0) : \|\mathbf{x}-\mathbf{x}_0\|_2 \leq B_n^{-1}\tilde{f}_n^{nw}\zeta_k^2 \mid T_0) \geq \exp(-Ck\zeta_k^2). \quad (13)$$

Then the posterior $\Pi^{(n)}(\cdot \mid \mathbf{a}_1,\dots,\mathbf{a}_k)$ converges to $(T_0,\mathbf{x}_0)$ at rate ε_k as $k \to \infty$.

Theorem 2 gives sufficient conditions for posterior consistency and a general tool for deriving posterior contraction rates. Notice that conditions (11)–(13) can be matched one-to-one to the conditions in Theorem 1 by defining the sieves

$$\Theta_k^{(n)} := \{(T,\mathbf{x}) : T \in \mathcal{T}_n, \mathbf{x} \in L_{f_k}(T) \cap U_{g_k}(T)\}, \quad (14)$$

which corresponds to selecting the trees that have a lower bound on the lengths of the external branches and an upper bound on all the branch lengths. Then (11) ensures that the entropy condition (3) is satisfied, while (12) and (13) correspond to the remaining and the prior mass conditions, respectively.

The requirement of a nested sequence of data-generating trees in Theorem 2 ensures that there is a well-defined limit object around which the posterior can converge as $n, k \to \infty$. In order to prove convergence, we require k to grow rapidly enough with n so that (11)–(13) can be shown to hold. The examples in Section 4 illustrate this requirement.

4. Examples

In this section we show that the conditions of Theorem 2 hold for two popular priors: the Kingman coalescent prior for ultrametric trees, and a uniform topology prior with independent exponential branch lengths on unconstrained trees. Both settings will use the following elementary lemma.

Lemma 2. Let $(X_1,\dots,X_n)$ be random variables with marginal laws $X_i \sim \mathrm{Exp}(\lambda_i)$. For any $f, g > 0$,

$$\mathbb{P}(X_i < f) \leq \lambda_i f\left[1 + \frac{\lambda_i}{2}f\right],$$
$$\mathbb{P}(\max\{X_1,\dots,X_n\} > g) \leq n\exp(-\min\{\lambda_1,\dots,\lambda_n\}g).$$

4.1. Kingman Coalescent Prior

Let $\mathcal{T}_n^K$ be the set of *ranked topologies* of rooted binary trees with n labeled leaves, of which there are $\binom{n}{2}\binom{n-1}{2}\cdots\binom{2}{2} = 2^{-n+1}(n!)^2/n \sim 4\pi[n/(\sqrt{2}e)]^{2n}$. A ranked topology specifies both the binary mergers which take place, and the order in which they occur. To ensure the prior places full mass on ultrametric trees, it is convenient to parameterize branch lengths by the $n-1$ holding times between mergers, $\mathbf{x} \in (0,\infty)^{n-1}$, so $m(T) \equiv n-1$ for all T, and

$$\psi_T(\mathbf{x})_{(u,v)} = \sum_{j\in\{1,\dots,n-1\}: x_j \in (u,v)} x_j, \quad (15)$$

where $x_j \in (u,v)$ means that holding time x_j contributes to the length of branch (u,v). See Figure 1 for an illustration. Let

$$\Pi^{K,n}(T,\mathbf{x}) := \exp\left(-\sum_{i=1}^{n-1}\binom{n-i+1}{2}x_i\right)$$

denote the Kingman coalescent prior (Kingman 1982) on $\Theta^{K,n} := \mathcal{T}_n^K \times (0,\infty)^{n-1}$. It is obtained by initializing n disconnected leaf lineages from time zero and merging each pair at rate 1 until their most recent common ancestor is reached. The resulting distribution on trees is uniform over ranked topologies, and has independent holding times $X_i \sim \text{Exp}(\binom{i+1}{2})$ for $i \in \{1,\dots,n-1\}$ governing its branch lengths.

Corollary 1 considers a sequence of trees tending to an infinite-leaf limit. The fact that such a limit exists is a well-known consistency property of the Kingman coalescent (Kingman 1982, Theorem 3), and we denote its law by Π^K. Indeed, a Kingman coalescent tree with n leaves can be extended to one with $n+1$ leaves by splitting a randomly chosen leaf into two, and extending all external branches (including the two zero-length ones just created by the split) by an independent $X_n \sim \text{Exp}(\binom{n+1}{2})$. This ordering of holding times, where x_{n-1} is the time until the first merger when there are n leaves, and x_1 is the holding time between the last two mergers, is depicted in Figure 1. For an infinite-leaf tree $(T,\mathbf{x}) \in \mathcal{T}_{\mathbb{N}}^K \times (0,\infty)^{\mathbb{N}}$, let $(T,\mathbf{x})|_n = (T|_n, \mathbf{x}|_n)$ be the restriction of the ranked topology to the first n leaves with branch lengths determined by holding times $\mathbf{x}|_n = (x_1,\dots,x_{n-1})$.

Corollary 1. Suppose $(T_0,\mathbf{x}_0)$ is an infinite-leaf tree in the support of Π^K, and that the data-generating sequence of trees is given by the restrictions $(T_0,\mathbf{x}_0)|_n$. Let the observed sites $(\mathbf{a}_1,\dots\mathbf{a}_k)$ be coupled such that for any n, trees $(T_0,\mathbf{x}_0)|_n$ and $(T_0,\mathbf{x}_0)|_{n+1}$ use the same Poisson process on $(T_0,\mathbf{x}_0)|_n$ and $((T_0,\mathbf{x}_0)|_{n+1})|_n$, and a further independent Poisson process on the difference $(T_0,\mathbf{x}_0)|_{n+1} \setminus (T_0,\mathbf{x}_0)|_n$. Suppose $n \equiv n_k$ increases slowly enough that for any sufficiently large k, and a constant $C > 0$ independent of k and n, we have $\log k \geq n^2/C^2$. Then posterior distributions $\Pi^{K,n}(\cdot \mid \mathbf{a}_1,\dots,\mathbf{a}_k)$ obtained from priors $\Pi^{K,n}$ concentrate around $(T_0,\mathbf{x}_0)$ at rate $\varepsilon_k \asymp k^{-1/2}(\log k)^{3/4+\beta/2}$ for any $\beta > 1$ as $k \to \infty$.

To illustrate Corollary 1, Figure 2(a) depicts the posterior support for the true tree T_0 as the number of leaves n and sites k increase in simple experiments with binary alleles and symmetric mutation rates $\mu = Q_{0,1} = Q_{1,0}$. Figure 2(b) displays the intervals on which the curves in Figure 2(a) first exceed 0.5.

4.2. Uniform Prior

Let $\mathcal{T}_n^U$ be the set of unrooted, binary trees with n labeled leaves. There are

$$|\mathcal{T}_n^U| = (2n-5)!! := 2^{3-n}(2n-5)!/(n-3)! \sim 2^{3-n}\sqrt{\frac{2n-5}{n-3}}\frac{(2n-5)^{2n-5}}{e^{n-8}(n-3)^{n-3}} \tag{16}$$

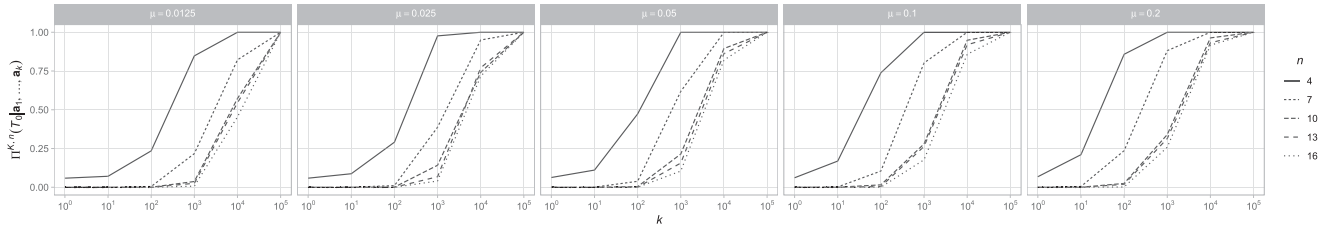
labeled topologies for such trees (Felsenstein 1978), each requiring the specification of $2n-3$ branch lengths. Thus, we set $m(T) \equiv 2n-3$, take the variables $(x_1,\dots,x_{2n-3}) \in (0,\infty)^{2n-3}$ to parameterize branch lengths directly, and set $\psi_T(\mathbf{x}) = \mathbf{x}$.

For $\lambda > 0$, let

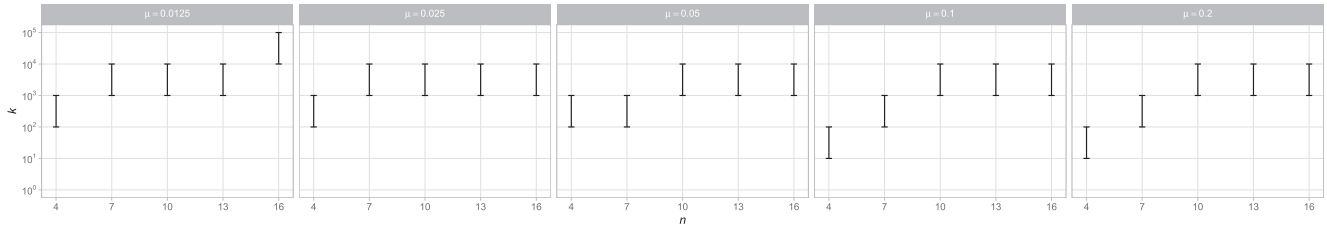
$$\Pi^{U,n}(T,\mathbf{x}) := \frac{\lambda^{2n-3}}{|\mathcal{T}_n^U|}\exp\left(-\lambda\sum_{i=1}^{2n-3}x_i\right)$$

be the prior on $\Theta^{U,n} := \mathcal{T}_n^U \times (0,\infty)^{2n-3}$ which is uniform over topologies and assigns independent exponentially distributed lengths with rate λ to branches. This prior was used for example by Whidden and Matsen IV (2015) to benchmark MCMC algorithms on phylogenetic trees.

Corollary 2 below considers a sequence of trees tending to an infinite-leaf limit. A consistent sequence of finite restrictions



(a) Posterior support for the true tree topology $T_0|_n$ averaged across 100 replicate data sets for each n, μ and k A figure depicting the posterior support from each data set is contained in the supplementary materials.



(b) Intervals for k on which the curves in Figure 2a first exceed 0.5.

Figure 2. Posterior support on the true rooted tree topology $T_0|_n$ under a Kingman coalescent prior $\Pi^{K,n}$ as the number of leaves n, number of sites k and mutation rate μ vary. The sequence of trees $(T_0,\mathbf{x}_0)|_n \sim \Pi^{K,n}$ was generated according to the algorithm in Section 4.1 and binary datasets were sampled according to a Markov model with mutation rate μ. See Section 4.3 for details of the inference.

$(T, \mathbf{x})|_n$ of an infinite-leaf tree can be constructed via the *sequential construction* of Aldous (1993):

1. To extend $(T, \mathbf{x}) \in \Theta^{U,n}$ to $\Theta^{U,n+1}$, choose uniformly at random a branch b on T. Attach two new branches at a uniformly chosen end of b: one of length $x_{2n-2} \sim \text{Exp}(\lambda)$ connecting b to its previous endpoint, and one of independent length $x_{2n-1} \sim \text{Exp}(\lambda)$ connecting the resulting node to a new leaf.
2. To restrict $(T, \mathbf{x}) \in \Theta^{U,n+1}$ to $\Theta^{U,n}$, choose a leaf uniformly at random and remove it. Remove the resulting degree-2 internal node by deleting the adjacent edge whose length has a higher index in $\mathbf{x}$ and attaching its endpoints to each other directly.

Let Π^U be the law of the infinite-leaf projective limit obtained from the above procedure, which exists by the Kolmogorov extension theorem. For an infinite-leaf tree $(T, \mathbf{x}) \in \mathcal{T}_{\mathbb{N}}^U \times (0,\infty)^{\mathbb{N}}$, let $(T, \mathbf{x})|_n = (T|_n, \mathbf{x}|_n)$ be the restriction of the unrooted topology to the first n leaves with branch lengths $\mathbf{x}|_n = (x_1, \ldots, x_{2n-3})$.

Corollary 2. Suppose $(T_0, \mathbf{x}_0)$ be an infinite-leaf tree in the support of Π^U, and that the data-generating sequence of trees is given by the restrictions $(T_0, \mathbf{x}_0)|_n$. Let the observed sites $(\mathbf{a}_1, \ldots \mathbf{a}_k)$ be coupled such that for any n, trees $(T_0, \mathbf{x}_0)|_n$ and $(T_0, \mathbf{x}_0)|_{n+1}$ use the same Poisson process on $(T_0, \mathbf{x}_0)|_n$ and $((T_0, \mathbf{x}_0)|_{n+1})|_n$, and a further independent Poisson process on the difference $(T_0, \mathbf{x}_0)|_{n+1} \setminus (T_0, \mathbf{x}_0)|_n$. Suppose $n \equiv n_k$ increases slowly enough that for any sufficiently large k, and a constant $C > 0$ independent of k and n, we have $\log k \geq n^2/C^2$. Then posterior distributions $\Pi^{U,n}(\cdot \mid \mathbf{a}_1, \ldots, \mathbf{a}_k)$ obtained from priors $\Pi^{U,n}$ concentrate around $(T_0, \mathbf{x}_0)$ at rate $\varepsilon_k \asymp k^{-1/2}(\log k)^{3/4+\beta/2}$ for any $\beta > 1$ as $k \to \infty$, provided the rate matrix $\mathbf{Q}$ is r-reversible.

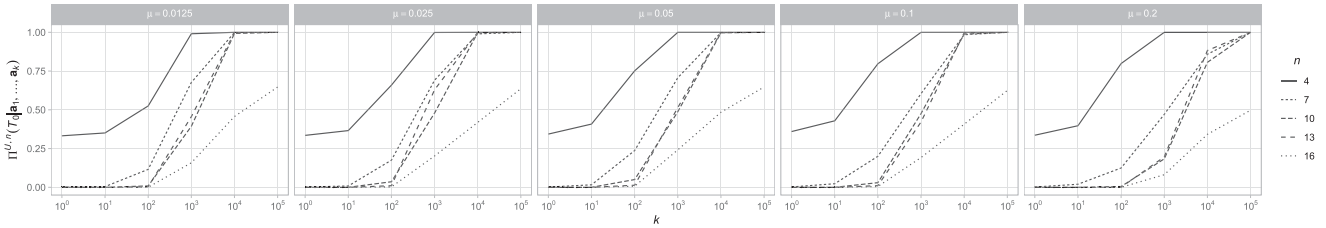
To illustrate Corollary 2, Figure 3(a) shows the posterior distributions on unrooted tree topologies concentrating in a simple experiment with binary alleles and symmetric mutation rates $\mu = Q_{0,1} = Q_{1,0}$. Figure 3(b) displays the intervals for which the posterior support on the true topology first exceeds 0.5.

4.3. Further Details on Experiments

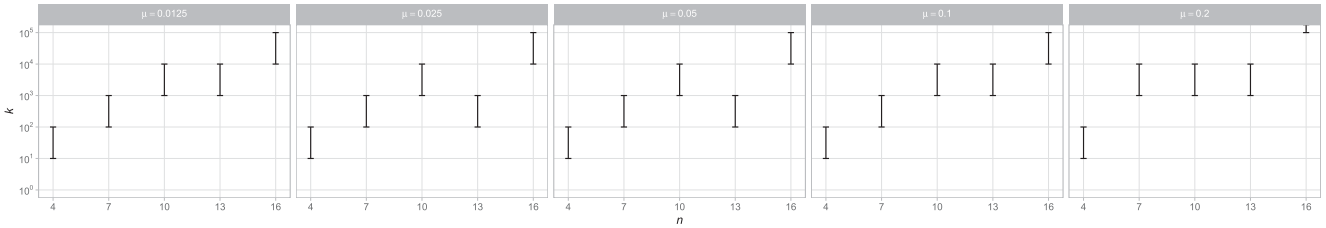
`RevBayes` (Höhna et al. 2016) was used to perform inference via Markov chain Monte Carlo. Each chain was initialized randomly and the first 10^6 iterations discarded as warm up, then 10^3 samples were drawn at intervals of 10^3 iterations and used to estimate the posterior support on the true tree topology. Mixing and convergence was assessed visually via trace plots of the log-likelihood at each iteration and computing the overall acceptance rate of proposals. Figures 2 and 3 were created using `ggplot2` (Wickham 2016) in R (R Core Team 2024). Code to recreate the experiments is included with the reproducibility materials and is also available at `github.com/lukejkelly/TreeConsistency`. The routines to generate trees and allele data make use of `ape` (Paradis and Schliep 2019) and `phangorn` (Schliep 2011). The supplementary material contains additional figures and details of the experiments.

5. Discussion

We have demonstrated the consistency of Bayesian tree reconstruction procedures in a wide range of settings, and under mild technical conditions. In particular, we do not require any discretization or boundedness assumptions on branch lengths, in contrast with earlier results. Our results apply to rooted and unrooted trees, unconstrained and ultrametric (or otherwise constrained) trees, and to both binary and more flexible tree models.



(a) Posterior support for the true tree topology $T_0|_n$ averaged across 100 replicate data sets for each n, μ and k A figure depicting the posterior support from each data set is contained in the supplementary materials.



(b) Intervals for k on which the curves in Figure 3a first exceed 0.5.

Figure 3. Posterior support on the true unrooted tree topology $T_0|_n$ under a prior $\Pi^{U,n}$ uniform across topologies and exponential on branch lengths as the number of leaves n, number of sites k and mutation rate μ vary. A sequence of trees $(T_0, \mathbf{x}_0)|_n \sim \Pi^{U,n}$ was generated according to the algorithm in Section 4.2 and binary datasets were sampled according to a Markov model with mutation rate μ. See Section 4.3 for details of the inference.

We presented several prototypical examples of tree reconstruction problems, and obtained explicit convergence rates for them which match known frequentist rates, obtained under stronger assumptions.

While our results apply to the setting in which the number of leaves, and hence the size of the tree, diverge with the number of replicate sites, we require the number of sites to be super-exponential in the number of leaves. Under stronger assumptions on branch lengths (and sometimes the mutation model), a polynomial number of sites suffices for accurate maximum likelihood reconstruction (Roch and Sly 2017; Dinh et al. 2018). Since maximum likelihood reconstruction corresponds to maximum a posteriori estimation under a uniform prior when branch lengths are bounded, it is clear that similar refinements can be obtained for particular models in the Bayesian setting. Our numerical results in Figures 2(b) and 3(b) are consistent with sub-exponential growth of k in n, strongly suggesting that the super-exponential growth rate which we require for our method of proof can be improved, at least in particular cases.

Our rates pertain to convergence on the space of distributions, $P_\theta \to P_{\theta_0}$. In addition, it is also of interest to obtain rates for convergence of the corresponding parameter estimation problem, $\theta \to \theta_0$. In the context of genetics, the latter corresponds to learning the ancestral tree, while the former corresponds to learning the distribution of alleles at the n leaves of the tree at an arbitrary site. The identifiability assumption (10) guarantees that $P_\theta \to P_{\theta_0} \Rightarrow \theta \to \theta_0$, but the rate of the latter depends on the fluctuations of the $\theta \mapsto P_\theta$ mapping and can, in principle, be arbitrarily slow. When they are available, model-specific results bounding $\|\theta - \theta_0\|$ by $\|P_\theta - P_{\theta_0}\|$ in suitable norms can be used to obtain rates for the $\theta \to \theta_0$ convergence as well; see Zhang, Dinh, and Matsen IV (2021, eq. (3)) for an example.

6. Proofs of Main Results

6.1. Proof of Theorem 2

We are going to match conditions (11)–(13) to the conditions (2)–(4) in Theorem 1. Let the subsets $\Theta_k^{(n)} \subset \Theta^{(n)}$ be given by (14), and define $\mathcal{P}_k^{(n)} = \{P_\theta : \theta \in \Theta_k^{(n)}\}$. Then the remaining mass condition (2) is clearly satisfied when (12) holds.

For the entropy condition (3), recall that $d_H^2(p,q) \le \mathrm{KL}(p;q)$ for any pair of probability densities p, q (Ghosal and van der Vaart 2017, Lemma B.1(iii)). Therefore, we can use (8) and the identifiability condition (10) to bound the covering number as

$$\begin{aligned} N(\varepsilon_k/2, \mathcal{P}_k^{(n)}, d_H) &\le N(\varepsilon_k^2/4, \mathcal{P}_k^{(n)}, \mathrm{KL}) \\ &\le |\mathcal{T}_n| N\left(\frac{\varepsilon_k^2 f_k^{nw}}{4B_n}, [0, g_k]^{m_n^*}, \|\cdot\|_2\right) \\ &\le |\mathcal{T}_n| \left(\frac{4B_n g_k}{\varepsilon_k^2 f_k^{nw}}\right)^{m_n^*}. \end{aligned}$$

Therefore, for (3) to hold we require the following, which is equivalent to (11):

$$\frac{g_k^{m_n^*}}{\varepsilon_k^{2m_n^*} f_k^{nwm_n^*}} \le \frac{\exp(k\varepsilon_k^2)}{|\mathcal{T}_n|(4B_n)^{m_n^*}}.$$

It remains to obtain (4). Let $\theta_0 = (T_0, \mathbf{x}_0)$ be the true parameter. Then for $\mathbf{x} \in L_{\tilde{f}_n}(T_0)$ we can use (8) and (9) to obtain

$$\begin{aligned} &\Pi^{(n)}((T,\mathbf{x}) \in \Theta : \mathrm{KL}(P_{\theta_0}; P_\theta) \le \zeta_k^2,\ V_0(P_{\theta_0}; P_\theta) \le \zeta_k^2) \\ &\quad \ge \Pi^{(n)}(T_0)\Pi^{(n)}(\mathbf{x} \in (0,\infty)^{m(T_0)} : \\ &\qquad \mathrm{KL}(P_{(T_0,\mathbf{x}_0)}; P_{(T_0,\mathbf{x})}) \le \zeta_k^2, \\ &\qquad V_0(P_{(T_0,\mathbf{x}_0)}; P_{(T_0,\mathbf{x})}) \le \zeta_k^2 \mid T_0) \\ &\quad \ge \Pi^{(n)}(T_0)\Pi^{(n)}\left(\mathbf{x} \in L_{\tilde{f}_n}(T_0) : \frac{B_n}{\tilde{f}_n^{nw}}\|\mathbf{x}-\mathbf{x}_0\|_2 \le \zeta_k^2, \right. \\ &\qquad \left. \frac{B_n^2}{\tilde{f}_n^{2nw}}\|\mathbf{x}-\mathbf{x}_0\|_2^2 \le \zeta_k^2 \mid T_0\right). \end{aligned}$$

For k large enough that $\zeta_k \le 1$, the condition $B_n^2\|\mathbf{x}-\mathbf{x}_0\|_2^2 \le \tilde{f}_n^{2nw}\zeta_k^2$ is automatically satisfied once $B_n\|\mathbf{x}-\mathbf{x}_0\|_2 \le \tilde{f}_n^{nw}\zeta_k^2$. Therefore,

$$\begin{aligned} &\Pi^{(n)}((T,\mathbf{x}) \in \Theta : \mathrm{KL}(P_{\theta_0}; P_\theta) \le \zeta_k^2,\ V_0(P_{\theta_0}; P_\theta) \le \zeta_k^2) \\ &\ge \Pi^{(n)}(T_0)\Pi^{(n)}(\mathbf{x} \in L_{\tilde{f}_n}(T_0) : \|\mathbf{x}-\mathbf{x}_0\|_2 \le B_n^{-1}\tilde{f}_n^{nw}\zeta_k^2 \mid T_0), \end{aligned}$$

at which point (13) implies (4).

6.2. Proof of Corollary 1

Before proving Corollary 1, we collect some useful properties of the infinite-leaf coalescent tree $(T,\mathbf{x}) \sim \Pi^K$ in a preparatory lemma.

Lemma 3. With Π^K-probability one, $\|\mathbf{x}\|_1 < \infty$, and only finitely many of the events $\{x_{n-1} < n^{-4}\}_{n\ge 2}$ happen.

Proof. Recall that $(x_1, x_2, \ldots)$ are independent and $x_i \sim \mathrm{Exp}\left(\binom{i+1}{2}\right)$. For the first claim,

$$\mathbb{E}[\|\mathbf{x}\|_1] = \mathbb{E}\left[\sum_{i=1}^\infty x_i\right] = \sum_{i=1}^\infty \frac{2}{i(i+1)} = 2\sum_{i=1}^\infty \left(\frac{1}{i} - \frac{1}{i+1}\right) = 2,$$

where the interchange of sum and expectation can be justified by Tonelli's theorem. Hence, $\Pi^K(\|\mathbf{x}\|_1 = \infty) = 0$. For the second claim, Lemma 2 yields

$$\begin{aligned} \sum_{n=2}^\infty \Pi^K(x_{n-1} < n^{-4}) &\le \sum_{n=2}^\infty \frac{n^2}{2} n^{-4}\left(1 + \frac{n^2}{4}n^{-4}\right) \\ &= \sum_{n=2}^\infty \frac{5}{8n^2} < \infty. \end{aligned}$$

The claim follows by the Borel–Cantelli lemma. □

For $\beta > 1, \delta > 0$, and the constant $C > 0$ from the statement of Corollary 1, define

$$\begin{aligned} \zeta_k^2 &= \frac{C(\log k)^{3/2}}{k}, \\ \varepsilon_k^2 &= \frac{C(\log k)^{3/2+\beta}}{k}, \\ f_k &= \frac{1}{n(n-1)}\exp(-(C+4)k\zeta_k^2), \\ g_k &= (C+4+\delta)k\zeta_k^2. \end{aligned}$$

Then, using $\log k \geq n^2/C^2$ yields

$$\begin{aligned}
&\frac{g_k^{n-1}}{\varepsilon_k^{2(n-1)} f_k^{nw(n-1)}} \exp(-k\varepsilon_k^2) \\
&\leq \frac{(C+4+\delta)^n n^{2wn^2}}{(\log k)^{\beta(n-1)}} \exp([n + C(C+4)wn^2(\log k)^{1/2} \\
&\qquad - C(\log k)^{1/2+\beta}]\log k) \\
&\leq \frac{(C+4+\delta)^n n^{2wn^2}}{(\log k)^{\beta(n-1)}} \exp([n + \{C(C+4)wn^2 \\
&\qquad - C(\log k)^{\beta}\}(\log k)^{1/2}]\log k) \\
&\leq \frac{(C+4+\delta)^n n^{2wn^2}}{(\log k)^{\beta(n-1)}} \exp([n + \{C(C+4)wn^2 \\
&\qquad - C^{1-2\beta} n^{2\beta}\}(\log k)^{1/2}]\log k).
\end{aligned}$$

Since $\beta > 1$, the factor inside the {}-braces is negative for large enough n. Hence, for such a large enough n,

$$\begin{aligned}
&\frac{g_k^{n-1}}{\varepsilon_k^{2(n-1)} f_k^{nw(n-1)}} \exp(-k\varepsilon_k^2) \\
&\leq \frac{(C+4+\delta)^n n^{2wn^2}}{(\log k)^{\beta(n-1)}} \exp([n + \{C(C+4)wn^2 \\
&\qquad - C^{1-2\beta} n^{2\beta}\}(n/C)]\log k) \\
&\leq \frac{(C+4+\delta)^n}{(n^2/C^2)^{\beta(n-1)}} \exp(2wn^2 \log n + [n + \{C(C+4)wn^2 \\
&\qquad - C^{1-2\beta} n^{2\beta}\}(n/C)]\log k) \\
&\leq \frac{(C+4+\delta)^n}{(n^2/C^2)^{\beta(n-1)}} \exp([2w\log n + n/C^2 + C^{-2}(C+4)wn^3 \\
&\qquad - C^{-2(1+\beta)} n^{1+2\beta}]n^2), \qquad (17)
\end{aligned}$$

where the last line follows because the negative sign dominates for large enough n since $\beta > 1$.

Recalling the definition of B_n in (7) and the fact that $\|\partial\psi\|_\infty \leq n-1$ in (15), the right-hand side of (11) is bounded below as

$$\begin{aligned}
\frac{1}{|\mathcal{T}_n^K|(4B_n)^{n-1}} &\sim \frac{1}{4\pi}\left(\frac{\sqrt{2}e}{n}\right)^{2n} \frac{1}{A\|\partial\psi^{(n)}\|_\infty (n-1)B^n} \\
&\geq \frac{1}{4\pi A n^2}\left(\frac{2e^2}{n^2 B}\right)^n. \qquad (18)
\end{aligned}$$

When $\beta > 1$, (17) decays faster than (18) as $n \to \infty$. Hence, (11) holds for any sufficiently large n.

In order to show (12) we use Lemma 2 twice. First, the holding time $x_{n-1} \sim \mathrm{Exp}\left(\binom{n}{2}\right)$ contributes to the lengths of all external branches in $(T,\mathbf{x})|_n$. Hence, $\cup_{T\in\mathcal{T}_n^K} L_f^c(T) = \{x_{n-1} < f\}$. Thus, by Lemma 2 and the definition of f_k,

$$\begin{aligned}
\Pi^{K,n}(\cup_{T\in\mathcal{T}_n^K} L_{f_k}^c(T)) &\leq \frac{1}{2}\exp(-(C+4)k\zeta_k^2)\left(1+\frac{1}{4}\right) \\
&= \frac{5}{8}\exp(-(C+4)k\zeta_k^2).
\end{aligned}$$

On the other hand, $\cup_{T\in\mathcal{T}_n^K} U_g^c(T) \subseteq \{\max_{i\in\{1,\dots,n-1\}}\{x_i\} > g\}$. Again by Lemma 2,

$$\begin{aligned}
\Pi^{K,n}(\cup_{T\in\mathcal{T}_n^K} U_{g_k}^c(T)) &\leq (n-1)\exp(-g_k) \\
&= (n-1)\exp(-(C+4+\delta)k\zeta_k^2) \\
&\leq \frac{3}{8}\exp(-(C+4)k\zeta_k^2),
\end{aligned}$$

when k is large enough that $-\delta C(\log k)^{3/2} + \log(n-1) \leq \log(3/8)$, which exists because k grows superlinearly in n. Therefore, (12) is satisfied.

The left-hand side of (13) does not depend on T_0 because ranked topologies and branch lengths are independent under Π^K and $\Pi^{K,n}$, and the marginal distribution over ranked topologies is uniform. Let $B_n' := \tilde{f}_n^{nw}/(\sqrt{n}B_n)$. We have that $\|\partial\psi^{(n)}\|_\infty \geq 1$, and by definition $\tilde{f}_n \leq 1/(2\eta)$. Hence

$$\begin{aligned}
B_n'\zeta_k^2 = \frac{\tilde{f}_n^{nw}}{\sqrt{n}B_n}\frac{C(\log k)^{3/2}}{k} &\leq \frac{\tilde{f}_n^{nw}}{n^{3/2}A\|\partial\psi^{(n)}\|_\infty B^n}\frac{n^3}{C^2\exp(n^2/C^2)} \\
&\leq \frac{n^{3/2}}{AC^2B^n(2\eta)^{wn}\exp(n^2/C^2)} \to 0. \qquad (19)
\end{aligned}$$

Again by Lemma 3, $\min\{x_1,\dots,x_{n-1}\} \to 0$ no faster than n^{-4}, which is slower than the decay in (19). Hence, we can take a large enough n and k, so that $x_{0,i} - B_n'\zeta_k^2 > 0$ for each $i \in \{1,\dots,n-1\}$, and $|x_{n-1} - x_{0,n-1}| \leq B_n'\zeta_k^2 \Rightarrow x_{n-1} \geq \tilde{f}_n$, whereupon

$$\begin{aligned}
&\Pi^{K,n}(\mathbf{x} : x_1 \geq \tilde{f},\ \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \leq B_n^{-1}\tilde{f}^{nw}\zeta_k^2) \\
&= \int_{\mathbf{x}\in\mathbb{R}_+^{n-1}:\, x_1\geq\tilde{f},\, \|\mathbf{x}-\mathbf{x}_0|_n\|_2\leq\sqrt{n}B_n'\zeta_k^2} \left[\prod_{i=1}^{n-1}\binom{i}{2}\exp\left(-\binom{i}{2}x_i\right)\right]\mathrm{d}\mathbf{x} \\
&\geq \int_{\mathbf{x}\in\mathbb{R}_+^{n-1}:\, |x_i-x_{0,i}|\leq B_n'\zeta_k^2} \left[\prod_{i=1}^{n-1}\binom{i}{2}\exp\left(-\binom{i}{2}x_i\right)\right]\mathrm{d}\mathbf{x} \\
&= \prod_{i=1}^{n-1}\left[\exp\left(-\binom{i}{2}(x_{0,i}-B_n'\zeta_k^2)\right) - \exp\left(-\binom{i}{2}(x_{0,i}+B_n'\zeta_k^2)\right)\right].
\end{aligned}$$

We use a Taylor expansion to show that for any $i \in \{1,\dots,n-1\}$,

$$\begin{aligned}
&\exp\left(-\binom{i}{2}(x_{0,i}-B_n'\zeta_k^2)\right) - \exp\left(-\binom{i}{2}(x_{0,i}+B_n'\zeta_k^2)\right) \\
&\geq 2\binom{i}{2}B_n'\zeta_k^2\exp\left(-\binom{i}{2}x_{0,i}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Pi^{K,n}(\mathbf{x}\in(0,\infty)^{n-1} : x_1 \geq \tilde{f},\ \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \leq \sqrt{n}B_n'\zeta_k^2) \\
&\geq (B_n')^{n-1}\frac{(n!)^2}{n}\zeta_k^{2(n-1)}\exp\left(-\binom{n}{2}\|\mathbf{x}_0\|_1\right).
\end{aligned}$$

By Stirling's formula and $\log k \geq n^2/C^2$,

$$\begin{aligned}
&\Pi^{K,n}(\mathbf{x}\in(0,\infty)^{n-1} : x_1 \geq \tilde{f}_n, \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \leq B_n^{-1}\tilde{f}_n^{nw}\zeta_k^2) \\
&\geq 2\pi(\log k)^{3(n-1)/2} \\
&\qquad \times \exp\left(2n\log(n/e) + (n-1)(\log B_n' + \log C)\right. \\
&\qquad \left. - \binom{n}{2}\|\mathbf{x}_0\|_1 - (n-1)\log k\right) \\
&\sim 2\pi\exp\left(2n\log(n/e) + (n-1)(\log B_n' + \log C)\right.
\end{aligned}$$

$$- \binom{n}{2}\|\mathbf{x}_0\|_1 - (n-1)\frac{n^2}{C^2} + 3(n-1)\log(n/C)\Big).$$

By Lemma 3 we have that $\|\mathbf{x}_0\|_1$ is bounded, and that $\log B_n'$ satisfies

$$\begin{aligned} wn\log n \geq \log B_n' &\geq wn\log\tilde{f}_n - \log A - \log\|\partial\psi^{(n)}\|_\infty \\ &\quad - \frac{3}{2}\log n - n\log B \\ &\gtrsim -4wn\log n - \log A - \frac{5}{2}\log n - n\log B \end{aligned}$$

because $\tilde{f}_n \geq \min_{i\in\{1,\ldots,n-1\}}\{x_{0,i}\}/2$ and $\|\partial\psi^{(n)}\|_\infty \leq n$. Hence, for sufficiently large n,

$$\begin{aligned} &\Pi^{K,n}(\mathbf{x}\in(0,\infty)^{n-1} : x_1 \geq \tilde{f}_n, \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \leq B_n^{-1}\tilde{f}_n^{nw}\zeta_k^2) \\ &\geq \exp\Big(-(1+o(n))\frac{n^3}{C^2}\Big). \end{aligned}$$

Condition (13) follows because

$$\frac{n^3}{C^2} = C(\log k)^{3/2} = k\zeta_k^2.$$

Hence, Theorem 2 holds, which shows that the sequence $\Pi^{K,n}$ is consistent at the claimed rate.

6.3. Proof of Corollary 2

Before proving Corollary 2, we collect some useful properties of the infinite-leaf uniform tree $(T,\mathbf{x}) \sim \Pi^U$ in a preparatory lemma.

Lemma 4. With Π^U-probability one, $\lambda\|\mathbf{x}|_n\|_1/n \to 1$ as $n\to\infty$, and only finitely many of the events $\{x_{n-1} < n^{-2}\}_{n\geq 2}$ happen.

Proof. The first claim is just the strong law of large numbers for independent, Exp(λ)-distributed random variables.

For the second claim, Lemma 2 yields

$$\begin{aligned} \sum_{n=2}^{\infty}\Pi^U(x_{n-1} < n^{-2}) &\leq \sum_{n=2}^{\infty}\frac{\lambda}{n^2}\Big(1+\frac{\lambda}{2n^2}\Big) \\ &= \sum_{n=1}^{\infty}\frac{\lambda(2n^2+\lambda)}{2n^4} < \infty. \end{aligned}$$

The claim follows by the Borel–Cantelli lemma. □

We will follow the same argument as the proof of Corollary 1 by checking (11)–(13), and concluding the claimed convergence using Theorem 2. For $\beta > 1, \delta > 0$, and $C > 0$ as in Corollary 2, define

$$\begin{aligned} \zeta_k^2 &= \frac{2C(\log k)^{3/2}}{k}, \\ \varepsilon_k^2 &= \frac{2C(\log k)^{3/2+\beta}}{k}, \\ f_k &= \frac{1}{2n\lambda}\exp(-(C+4)k\zeta_k^2), \\ g_k &= \frac{(C+4+\delta)}{\lambda}k\zeta_k^2. \end{aligned}$$

Then, using $\log k \geq n^2/C^2$,

$$\begin{aligned} &\frac{g_k^{2n-3}}{\varepsilon_k^{2(2n-3)}f_k^{wn(2n-3)}}\exp(-k\varepsilon_k^2) \\ &= \frac{(C+4+\delta)^{2n-3}(2n)^{wn(2n-3)}\lambda^{(2n-3)(wn-1)}}{(\log k)^{\beta(2n-3)}} \\ &\quad\times\exp\Big([2n-3+2C(C+4)wn(2n-3)(\log k)^{1/2} \\ &\qquad - 2C(\log k)^{1/2+\beta}]\log k\Big) \\ &\leq \exp\Big((2n-3)[\log(C+4+\delta)+(wn-1)\log\lambda \\ &\qquad - \beta\log(n^2/C^2)] \\ &\qquad + \Big[w\frac{(2n-3)}{n}\log(2n) + \frac{2n-3}{C^2} + \frac{2(C+4)}{C^2} \\ &\qquad\times w(2n-3)n^2 - \frac{2n^{1+2\beta}}{C^{2+2\beta}}\Big]n^2\Big). \end{aligned} \tag{20}$$

Recalling (16), the definition of B_n in (7), and the fact that $\|\partial\psi\|_\infty \equiv 1$, the right-hand side of (11) is bounded below as

$$\begin{aligned} \frac{1}{|\mathcal{T}_n^U|(4B_n)^{n-1}} &\sim 2^{n-3}\sqrt{\frac{n-3}{2n-5}}\frac{e^{n-8}(n-3)^{n-3}}{(2n-5)^{2n-5}} \\ &\quad\times \frac{1}{A\|\partial\psi\|_\infty(n-1)B^n} \\ &\geq \frac{(1+o(n))}{\sqrt{2}An}\exp((n-3)[\log 2+\log(n-3)] \\ &\qquad + n - 8 - (2n-5)\log(2n-5) \\ &\qquad - n\log B - \log n). \end{aligned} \tag{21}$$

When $\beta > 1$, (20) decays faster than (21) as $n\to\infty$. Hence, (11) holds for any sufficiently large n.

In order to show (12), we use Lemma 2. When $f_k \leq 1$

$$\Pi^{U,n}(\cup_{T\in\mathcal{T}_n^U}L_{f_k}^c(T)) \leq n\lambda f_k\Big(1+\frac{n\lambda}{2}f_k\Big) \leq \frac{5}{8}\exp(-(C+4)k\zeta_k^2).$$

On the other hand,

$$\begin{aligned} \Pi^{U,n}(\cup_{T\in\mathcal{T}_n^U}U_g^c(T)) &\leq (2n-3)\exp(-\lambda g_k) \\ &= (2n-3)\exp(-(C+4+\delta)k\zeta_k^2) \\ &\leq \frac{3}{8}\exp(-(C+4)k\zeta_k^2) \end{aligned}$$

when k is large enough that $-\delta C(\log k)^{3/2} + \log(2n-3) \leq \log(3/8)$, which exists because k grows faster than linearly with n. Therefore, (12) is satisfied.

The left-hand side of (13) does not depend on T_0 in this case because topologies and branch lengths are independent under $\Pi^{U,n}$ and Π^U, and the marginal distribution over topologies is uniform. Let $B_n' := \tilde{f}_n^{nw}/(\sqrt{n}B_n)$. We have by definition that $\|\partial\psi^{(n)}\|_\infty \equiv 1$ and $\tilde{f}_n \leq 1/(2\eta)$. Hence

$$\begin{aligned} B_n'\zeta_k^2 = \frac{\tilde{f}_n^{nw}}{\sqrt{n}B_n}\frac{2C(\log k)^{3/2}}{k} &\leq \frac{1}{(2\eta)^{wn}n^{3/2}AB^n}\frac{2n^3}{C^2\exp(n^2/C^2)} \\ &\sim \frac{n^{3/2}}{AC^2(2\eta)^{nw}B^n\exp(n^2/C^2)} \to 0. \end{aligned} \tag{22}$$

Again by Lemma 4, $\min\{x_1,\dots,x_{n-1}\} \to 0$ no faster than n^{-2}, which is slower than the decay in (22). Hence, we can take a large enough n and k so that $x_{0,i} - B'_n\zeta_k^2 > 0$ for each $i \in \{1,\dots,2n-3\}$, and $|x_i - x_{0,i}| \le B'_n\zeta_k^2 \Rightarrow x_i \ge \tilde{f}_n$ for all $i \in \{1,\dots,n\}$, whereupon

$$\begin{aligned}
&\Pi^{U,n}(\mathbf{x} \in (0,\infty)^{2n-3} : \mathbf{x} \in L_{\tilde{f}_n}(T_0), \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \le B_n^{-1}\tilde{f}_n^{nw}\zeta_k^2)\\
&= \lambda^{2n-3}\int_{\mathbf{x}\in\mathbb{R}_+^{2n-3}:\,\mathbf{x}\in L_{\tilde{f}_n}(T_0),\,\|\mathbf{x}-\mathbf{x}_0|_n\|_2\le\sqrt{n}B'_n\zeta_k^2} \exp\left(-\lambda\sum_{i=1}^{2n-3}x_i\right)\mathrm{d}\mathbf{x}\\
&\ge \lambda^{2n-3}\int_{\mathbf{x}\in\mathbb{R}_+^{2n-3}:\,|x_i-x_{0,i}|\le B'_n\zeta_k^2} \exp\left(-\lambda\sum_{i=1}^{2n-3}x_i\right)\mathrm{d}\mathbf{x}\\
&= \prod_{i=1}^{2n-3}\left(\exp(-\lambda(x_{0,i}-B'_n\zeta_k^2)) - \exp(-\lambda(x_{0,i}+B'_n\zeta_k^2))\right).
\end{aligned}$$

We use a Taylor expansion to show that for any $i \in \{1,\dots,2n-3\}$,

$$\begin{aligned}
&\exp(-\lambda(x_{0,i}-B'_n\zeta_k^2)) - \exp(-\lambda(x_{0,i}+B'_n\zeta_k^2))\\
&\qquad \ge 2\lambda B'_n\zeta_k^2\exp(-\lambda x_{0,i}).
\end{aligned}$$

Therefore, using $\log k \ge n^2/C^2$ and $\|\mathbf{x}_0|_n\|_1 \sim n/\lambda$ from Lemma 4,

$$\begin{aligned}
&\Pi^{U,n}(\mathbf{x} \in (0,\infty)^{2n-3} : \mathbf{x} \in L_{\tilde{f}_n}(T_0), \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \le \sqrt{n}B'_n\zeta_k^2)\\
&\ge (2\lambda B'_n)^{2n-3}\zeta_k^{2(2n-3)}\exp(-\lambda\|\mathbf{x}_0|_n\|_1)\\
&\ge \frac{\tilde{f}_n^{wn(2n-3)}(2C)^{2n-3}(\log k)^{3(2n-3)/2}}{A^{2n-3}n^{3(2n-3)/2}B^{n(2n-3)}k^{2n-3}}\exp(-\lambda\|\mathbf{x}_0|_n\|_1)\\
&\ge \exp\Bigg((2n-3)\Bigg[wn\log\tilde{f}_n + \log(2C) + 3\log(n/C)\\
&\qquad - \log A - \frac{3}{2}\log n - n\log B - \frac{n^2}{C^2}\Bigg] - (1+o(n))n\Bigg).
\end{aligned}$$

By definition $\tilde{f}_n \le 1/(2\eta)$. To obtain a lower bound, note $\tilde{f}_n \ge \min\{x_{0,1},\dots,x_{0,2n-3}\}/2$. By Lemma 4, $\min\{x_{0,1},\dots,x_{0,2n-3}\}/2 < n^{-2}/2$ only finitely often as $n\to\infty$. These particularly short branches will all be internal eventually Π^U-almost surely by the second Borel–Cantelli lemma, because any given branch gets chosen as the attachment point of a new leaf with probability $1/(2n-3)$ when there are n leaves, and successive attachment points are independent. Hence, $\tilde{f}_n > n^{-2}/2$ for any sufficiently large n, and

$$\begin{aligned}
&\Pi^{U,n}(\mathbf{x} \in (0,\infty)^{2n-3} : \mathbf{x} \in L_{\tilde{f}_n}(T_0), \|\mathbf{x}-\mathbf{x}_0|_n\|_2 \le \sqrt{n}B'_n\zeta_k^2)\\
&\ge \exp\Bigg((2n-3)\Bigg[\log(2C) + 3\log(n/C) - 2wn\log(n/\sqrt{2})\\
&\qquad - \log A - \frac{3}{2}\log n - n\log B\Bigg]\\
&\qquad - (1+o(n))n - \frac{2n-3}{2n}\frac{2n^3}{C^2}\Bigg)\\
&= \exp\left(-(1+o(n))\frac{2n^3}{C^2}\right).
\end{aligned}$$

Condition (13) follows because

$$\frac{2n^3}{C^2} = 2C(\log k)^{3/2} = k\zeta_k^2.$$

Hence, Theorem 2 holds, which shows that the sequence $\Pi^{U,n}$ is consistent at the claimed rate.

7. Proofs of Technical Lemmas

7.1. Proof of Lemma 1

For $i \in \{1,\dots,m(T)\}$, we shorten $\partial_i\psi_T(\mathbf{x})_{(u,v)}$ for the partial derivative $[\partial\psi_T(\mathbf{x})/(\partial x_i)]_{(u,v)}$. Then

$$\begin{aligned}
\left|\frac{\partial P_{(T,\mathbf{x})}(\mathbf{a})}{\partial x_i}\right| &\le \sum_{\mathbf{a}^*\in H^{|I(T)|}} r(a^{(u_0)})\\
&\quad\times \sum_{(u,v)\in T}|(\partial_i\psi_T(\mathbf{x})_{(u,v)}\mathbf{Q}\exp(\mathbf{Q}\psi_T(\mathbf{x})_{(u,v)}))_{a^{(u)}a^{(v)}}|\\
&\quad\times \prod_{(u',v')\ne(u,v)}\exp(\mathbf{Q}\psi_T(\mathbf{x})_{(u',v')})_{a^{(u')}a^{(v')}}\\
&\le \sum_{\mathbf{a}^*\in H^{|I(T)|}}\ \sum_{(u,v)\in T:\partial_i\psi_T(\mathbf{x})\ne0}\|\partial\psi_T(\cdot)\|_\infty\\
&\quad\times|(\mathbf{Q}\exp(\mathbf{Q}\psi_T(\mathbf{x})_{(u,v)}))_{a^{(u)}a^{(v)}}|\\
&\le \gamma\|\partial\psi^{(n)}\|_\infty(2n-2)|H|^{n-1},
\end{aligned}$$

where we have bounded probabilities by 1, and used the fact that a tree with n leaves has at most $n-1$ internal nodes and $2n-2$ branches. On the event $L_f(T)$ we also have

$$P_{(T,\mathbf{x})}(\mathbf{a}) \ge \min_{\alpha\in H}\left\{r(\alpha)\min_{t\ge0}\{\exp(\mathbf{Q}t)_{\alpha\alpha}\}^{n-2}\min_{\beta\in H,t\ge f}\{\exp(\mathbf{Q}t)_{\alpha\beta}\}^n\right\} =: \lambda_f,$$

where the minima are guaranteed to be strictly positive because $\exp(\mathbf{Q}t)$ has a stationary distribution, and because all external edges have strictly positive minimum lengths. Hence

$$\left|\frac{\partial\log P_{(T,\mathbf{x})}(\mathbf{a})}{\partial x_i}\right| \le \frac{\gamma\|\partial\psi^{(n)}\|_\infty(2n-2)|H|^{n-1}}{\lambda_f},$$

and therefore, whenever $\mathbf{x}_0 \in L_f(T_0)$ and $\mathbf{x} \in L_f(T_0)$ for $f > 0$,

$$\mathrm{KL}(T_0,\mathbf{x}_0;T_0,\mathbf{x}) \le \frac{\gamma\|\partial\psi^{(n)}\|_\infty(2n-2)|H|^{n-1}}{\lambda_f}\|\mathbf{x}_0-\mathbf{x}\|_2.$$

It remains to bound λ_f from below. By definition, w mutations suffice for changing a type to any other along a lineage with positive probability. Moreover, mutations are distributed according to a Poisson process with rate at least $\eta > 0$. For a sufficiently small branch length $f > 0$, we thus have

$$\begin{aligned}
\lambda_f &\ge \min_{\alpha\in H}\left\{r(\alpha)\min_{t\ge0}\{\exp(\mathbf{Q}t)_{\alpha\alpha}\}^{n-2}\right\}\frac{(\eta f)^{nw}\exp(-nf)}{w!}\\
&\ge \min_{\alpha\in H}\left\{r(\alpha)\min_{t\ge0}\{\exp(\mathbf{Q}t)_{\alpha\alpha}\}^{n-2}\right\}\frac{(\eta f)^{nw}(1-\eta f)}{w!},
\end{aligned}$$

where the right hand side on the first line is a lower bound for the probability that the mutation process produces the observed configuration of types at the leaves. It is obtained as the probability that type α appears at each internal node (of which there are at most $n-2$ and the root), and that a sufficient number of mutations happens on each external branch (all of which have length at least f) to yield the desired type at each leaf. Hence

$$\begin{aligned}
&\mathrm{KL}(T_0,\mathbf{x}_0;T_0,\mathbf{x})\\
&\le \max_{\alpha\in H}\left\{\frac{1}{r(\alpha)\min_{t\ge0}\{\exp(\mathbf{Q}t)_{\alpha\alpha}\}^{n-2}}\right\}\\
&\quad\times\frac{\gamma\|\partial\psi^{(n)}\|_\infty(2n-2)|H|^{n-1}w!}{(\eta f)^{nw}(1-\eta f)}\|\mathbf{x}_0-\mathbf{x}\|_2,
\end{aligned}$$

whereupon using $f < 1/(2\eta)$ ensures that the right hand side is positive and yields the required bound. The calculation for V_0 is essentially identical, and omitted.

7.2. Proof of Lemma 2

By Taylor's theorem with Lagrange form remainder,

$$\mathbb{P}(X_i < f) = 1 - \exp(-\lambda_i f) \leq \lambda_i f + \frac{\lambda_i^2}{2} \exp(-\lambda_i \zeta) f^2$$

for some $\zeta \in (0, f)$. Maximizing the right hand side with $\zeta = 0$ gives the claimed bound.

On the other hand, by the union bound,

$$\begin{aligned}\mathbb{P}(\max\{X_1, \ldots, X_n\} > g) &\leq \sum_{i=1}^{n} \mathbb{P}(X_i > g) \\ &= \sum_{i=1}^{n} \exp(-\lambda_i g) \\ &\leq n \exp(-\min\{\lambda_1, \ldots, \lambda_n\} g).\end{aligned}$$

Supplementary Materials

The supplementary material contains figures illustrating how the posterior support for the true tree topology in Section 4 varies across datasets. The reproducibility materials contain the synthetic trees used in the experiments and instructions for generating data and creating figures.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

LJK received support from the French government under management of Agence Nationale de la Recherche as part of the ABSint (reference ANR-18-CE40-0034) and PRAIRIE (reference ANR-19-P3IA-0001) programs. JK was funded by UKRI through EPSRC grant EP/V049208/1.

ORCID

Luke J. Kelly http://orcid.org/0000-0003-4653-9303
Jere Koskela http://orcid.org/0000-0002-2836-8777

References

Aldous, D. (1993), "The Continuum Random Tree III," *The Annals of Probability*, 21, 248–289. [6]

Borges, R., and Kosiol, C. (2020), "Consistency and Identifiability of the Polymorphism-Aware Phylogenetic Models," *Journal of Theoretical Biology*, 486, 110074. [1]

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., de Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., and Dummond, A. J. (2019), "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis," *PLOS Computational Biology*, 15, e1006650. [1]

Dinh, V., Si Tung Ho, L., Suchard, M. A., and Matsen IV, F. A. (2018), "Consistency and Convergence Rate of Phylogenetic Inference via Regularization," *The Annals of Statistics*, 46, 1481–1512. [1,2,7]

Felsenstein, J. (1978), "The Number of Evolutionary Trees," *Systematic Biology*, 27, 27–33. [5]

——— (1981), "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach," *Journal of Molecular Evolution*, 17, 368–376. [3]

Freedman, D., and Diaconis, P. (1983), "On Inconsistent Bayes Estimates in the Discrete Case," *The Annals of Statistics*, 11, 1109–1118. [1]

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000), "Convergence Rates of Posterior Distributions," *The Annals of Statistics*, 28, 500–531. [1,2]

Ghosal, S., and van der Vaart, A. W. (2017), *Fundamentals of Nonparametric Bayesian Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press. [2,7]

Höhna, S., Landis, M., Heath, T., Boussau, B., Lartillot, N., Moore, B., Huelsenbeck, J., and Ronquist, F. (2016), "RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language," *Systematic Biology*, 65, 726–736. [1,6]

Kingman, J. F. C. (1982), "The Coalescent," *Stochastic Processes and their Applications*, 13, 235–248. [2,5]

Markin, A., and Eulenstein, O. (2021), "Quartet-based Inference is Statistically Consistent Under the Unified Duplication-Loss-Coalescence Model," *Bioinformatics*, 28, 4064–4074. [1]

Norris, J. (1997), *Markov Chains*, Cambridge: Cambridge University Press. [3]

Paradis, E., and Schliep, K. (2019), "ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R," *Bioinformatics*, 35, 526–528. [6]

R Core Team. (2024), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. *https://www.R-project.org/* [6]

Roch, S., and Sly, A. (2017), "Phase Transition in the Sample Complexity of Likelihood-based Phylogeny Inference," *Probability Theory and Related Fields*, 169, 3–62. [1,7]

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012), "Efficient Bayesian Phylogenetic Inference and Model Selection Across a Large Model Space," *Systematic Biology*, 61, 539–542. [1]

Sayyari, E., and Mirarab, S. (2016), "Anchoring Quartet-based Phylogenetic Distances and Applications to Species Tree Reconstruction," *BMC Genomics*, 17, 101–113. [1]

Schliep, K. (2011), "phangorn: Phylogenetic Analysis in r," *Bioinformatics*, 27, 592–593. [6]

Schwartz, L. (1965), "On Bayes Procedures," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 4, 10–26. [1]

Semple, C., and Steel, M. A. (2013), *Phylogenetics*, Oxford: Oxford University Press. [1]

Stamatakis, A. (2006), "RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models," *Bioinformatics*, 22, 2688–2690. [1]

Steel, M. A. (2013), "Consistency of Bayesian Inference of Resolved Phylogenetic Trees," *Journal of Theoretical Biology*, 336, 246–249. [1]

Steel, M. A., and Székely, L. A. (2007), "Teasing Apart Two Trees," *Combinatorics, Probability & Computing*, 16, 903–922. [4]

——— (2009), "Inverting Random Functions III: Discrete MLE Revisited," *Annals of Combinatorics*, 13, 365–382. [4]

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018), "Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10," *Virus Evolution*, 4, vey016. [1]

Warnow, T. (2017), *Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation*, Cambridge: Cambridge University Press. [1]

Whidden, C., and Matsen IV, F. A. (2015), "Quantifying MCMC Exploration of Phylogenetic Tree Space," *Systematic Biology*, 64, 472–491. [5]

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer-Verlag. [6]

Zhang, C., Dinh, V., and Matsen IV, F. A. (2021), "Nonbifurcating Phylogenetic Tree Inference via the Adaptive LASSO," *Journal of the American Statistical Association*, 116, 858–873. [1,3,7]

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018), "ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees," *BMC Bioinformatics*, 19, 153. [1]