Probabilities of tree topologies with temporal constraints and diversification shifts

Gilles Didier
IMAG, Univ Montpellier, CNRS, Montpellier, France
gilles.didier@umontpellier.fr

September 4, 2019

Abstract

Dating the tree of life is a task far more complicated than only determining the evolutionary relationships between species. It is therefore of interest to develop approaches apt to deal with undated phylogenetic trees.

The main result of this work is a method to compute probabilities of undated phylogenetic trees under Markovian diversification models by constraining some of the divergence times to belong to given time intervals and by allowing diversification shifts on certain clades. If the diversification models considered are lineage-homogeneous, the time complexity of this computation is quadratic with the number of species of the phylogenetic tree and linear with the number of temporal constraints.

The interest of this computation method is illustrated with three applications, namely,

- to compute the distribution of the divergence times of a tree topology with temporal constraints,
- to directly sample the divergence times of a tree topology, and
- to test for a diversification shift at a given clade.

Keywords: Phylogenetics, Datation, Shift Detection, Diversification, Birth-death process

1 Introduction

Estimating divergence times (i.e., the times of the speciation events corresponding to the internal nodes of a phylogenetic tree) is an essential and difficult stage of phylogenetic inference [23, 24, 18, 5, 21]. In order to perform this estimation, current approaches use stochastic models for combining different types of information: molecular and/or morphological data, fossil calibrations, evolutionary assumptions etc [34, 25, 9, 14]. An important point here is that dating speciation events is far more complicated and requires stronger assumptions on the evolutionary process than just determining the evolutionary relationships between species, not to mention the uncertainty with which divergence times can be estimated. It is therefore preferable to use, as much as possible, methods that do not require the exact knowledge of the divergence times. This is in particular true for studying questions related to the diversification process since the diversification process and divergence times are intricately linked. Diversification models are used in order to provide "prior" probability distributions of divergence times (i.e., which does not take into account information about genotype or phenotype of species [36, 16, 14]) [6, 15, 36]. Conversely, estimating parameters of diversification models requires temporal information about phylogenies.

We shall consider Markovian diversification models (i.e., with independence and memoryless properties) among which the birth-death-sampling model is arguably the simplest realistic model since it includes three important features shaping phylogenetic trees [37, 38]. Namely, it models cladogenesis and extinction of species by a birth-death process and takes account of the incompleteness of data by assuming a uniform sampling of extant taxa. The birth-death-sampling model has been further studied and is currently used for phylogenetic inference [30, 32, 15, 6]. Since assuming constant diversification rates along time is sometimes unrealistic, the birth-death-sampling model has been extended in various ways. The generalized birth-death process proposed in [17] allows to consider time-varying rates. In the model presented in [31], the diversification rates are piecewise-constant and the model allows to sample lineages not only at the present time but also at given past times in order to model mass extinction events. We combined the features of this two models to devise the *sampled-generalized-birth-death model* which allows both times-varying rates and past and extant samplings (Appendix B). The main goal of this work is to present methods to compute probabilities of undated phylogenies under certain assumptions about divergence times and about the diversification process under general models. Though this study focuses on methodological and computational aspects, three applications illustrating its practical interest are provided.

The first result is a method to compute the probability, under a Markovian diversification model, of a tree topology in which the divergence times are not exactly known but can be "constrained" to belong to given time intervals. This

computation is performed by splitting the tree topology into small parts involving the times of the temporal constraints, referred to as *patterns*, and by combining their probabilities in order to get the probability of the whole tree topology. If the diversification model is lineage-homogeneous, the total time complexity of this computation is quadratic with the size of the phylogeny (i.e., its total number of nodes) and linear with the total number of constraints. Its memory space complexity is quadratic with the size of the phylogeny. In practice, it can deal with phylogenetic trees with hundreds of tips on standard desktop computers.

This computation can be used to obtain the divergence time distributions of a given undated phylogeny with temporal constraints, which can be applied to various questions. First, it can be used for dating phylogenetic trees from their topology only, like the method implemented in the function *compute.brlen* of the R-package *APE* [12, 22]. It also allows to visualize the effects of the model parameters on the prior divergence times distributions, to investigate consequences of evolutionary assumptions etc. Last, it can provide prior distributions in phylogenetic inference frameworks. Note that the ability to take into account temporal constraints on the divergence times is particularly interesting in this context since in the calibration process, fossil ages are generally used for bracketing some of the divergence times [19]. The computation of the divergence time distribution is illustrated with a contrived example in order to show the influence of the temporal constraints and on a real phylogenetic tree in order to show the influence of the parameters of a simple birth-death-sampling model on the divergence time distributions. A previous method for computing divergence time distributions under the birth-death model [10] is briefly recalled in Section 7.1. It is based on a different idea and it seems difficult to extend it in order to take into account temporal constraints.

The computation of the probability of a tree topology under a given model allows us to sample all its divergence times under this model. In particular, this sampling procedure can easily be integrated into phylogenetic inference software [8, 26], e.g., for proposing accurate MCMC moves.

A second result shows how to calculate the probability of a tree topology in which a given clade is assumed to diversify following a diversification model different from that of the rest of the phylogeny. A natural application of this computation is to test diversification shift in undated phylogenies. It is used to define a likelihood ratio test for diversification shift which is compared with three previous approaches studied in [35].

Last, the approach presented here can be extended in order to take into account fossils. In [4], we started to work in this direction by determining divergence time distributions from tree topologies and fossil ages under the fossilized-birth-death model in order to obtain better node-calibrations for phylogenetic inference.

C-source code of the software performing the computation of divergence time distributions and their sampling under (piecewise-constant-)birth-death-sampling model and the shift detection test is available at `https://github.com/gilles-didier/DateBDS`.

The rest of the paper is organized as follows. Diversification models and birth-death-sampling models are formally introduced in Section 2.1. Section 3 presents definitions and some results about tree topologies. The standard and special patterns, i.e., the subparts of the diversification process from which are computed our probabilities, are introduced in Section 4. Sections 5 and 6 describe the computation of the probabilities of tree topologies with temporal constraints and diversification shifts, and show that this computation is quadratic with the size of the tree topology. Divergence time distributions obtained on two examples are displayed and discussed in Section 7. The method for directly sampling the divergence times is described in Section 8. Last, Section 9 presents a likelihood ratio test derived from the computation devised here, for determining if a diversification shift occurred in a tree topology. Its accuracy is assessed and compared with three previous tests of diversification shift. Appendices start with a table of notations, followed with the presentation of the sampled-generalized-birth-death model then with the proofs of theorems.

2 Diversification models

The methods presented below apply to general diversification models. Namely, a diversification model Θ provides the parameters of a stochastic process which starts with a single lineage at time s and ends at time e, where e is usually the present time (both s and e are parameters of Θ). At any time t between s and e, two type of event may occur on a lineage alive at t: a *speciation* event, which gives rise to a new lineage and an *extinction* event which basically kills the lineage. We also assume that the lineages alive at the end time e are sampled in a certain way (Fig. 1). A lineage alive at the end time which is not sampled has to be interpreted as a taxa which is not included in the study, for instance because it is unknown.

A important point is to distinguish between the part of the process that actually happened, which will be referred to as the *complete process* (Fig. 1-Left) and the part that can be observed from the available information at the present time (i.e., from the sampled extant taxa), which will be referred to as the *reconstructed process* (Fig. 1-Right). More formally, for all times $t \in [s, e]$, a lineage alive at time t is *observable* if itself or at least one of its descendants are both alive and sampled at the end time e. We assume that the reconstructed process, which encompasses all the observable parts of the diversification, is the only information available.

Let Θ be a diversification model starting at s and ending at e, $\mathcal{T}$ be a tree topology and t and t' be two times such that $s \leq t \leq t' \leq e$. We assume that we are able to compute under Θ:

- $\mathbf{T}_\Theta(\mathcal{T})$, the probability that the reconstructed tree topology is $\mathcal{T}$ conditionally on the number of tips of $\mathcal{T}$,

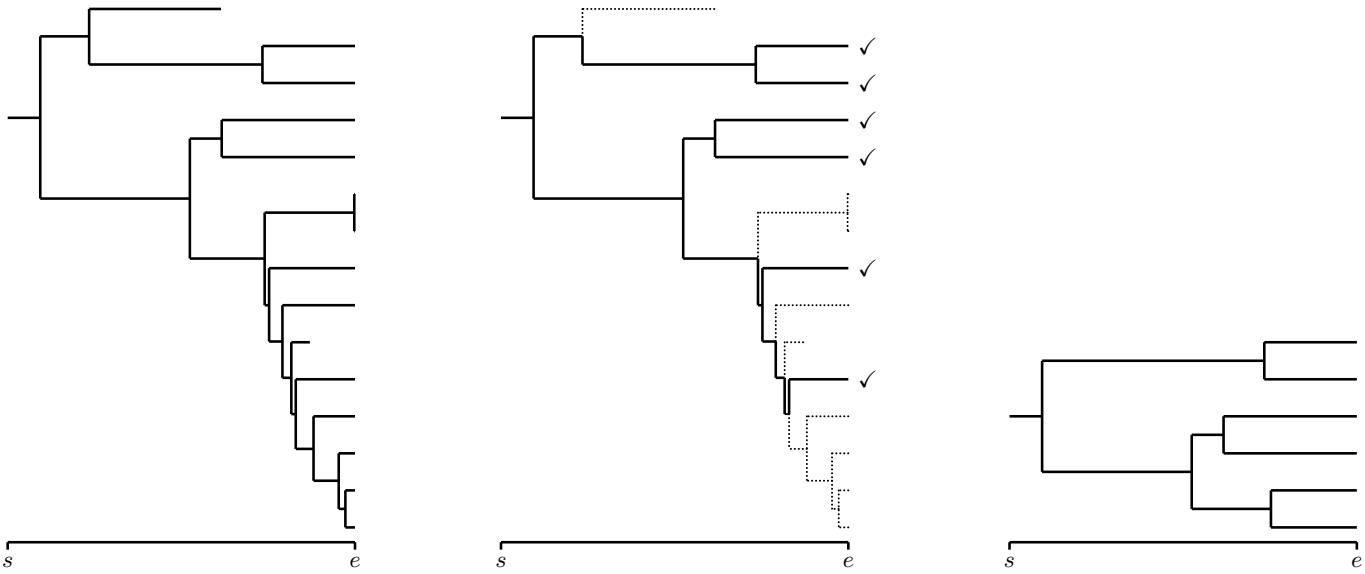

Figure 1: Left: the whole diversification process; Center: the part of the process that can be reconstructed is represented in plain – the dotted parts are lost (sampled extant species are those with '✓'); Right: the resulting phylogenetic tree.

- $\mathbf{Q}_\Theta(t, t', N)$, the probability that a lineage alive at time t has N descendants at time t' if $t' < e$ and N descendants *sampled* at time e otherwise,
- $\mathbf{O}_\Theta(t)$, the probability that a lineage alive at time t has at least a sampled descendant at the end time e.

For a diversification model starting at s and ending at e and a time $t \in [s, e]$, we put $\Theta_{[t]}$ for the model Θ restricted to the time interval $[t, e]$. Namely, $\Theta_{[t]}$ models the evolution of a lineage alive at t until e under Θ.

A diversification model Θ is *Markovian* if conditionally of being alive at a time t between s and e, the evolution of a lineage from t is independent of its evolution before t and of that of the other lineages. A diversification model Θ is *lineage-homogeneous* if conditionally at occurring at a time t, any event of the process occurs on all the lineages alive at t with equal probabilities.

2.1 Birth-death-sampling models

Under a birth-death-sampling model, the dynamics of speciation and extinction of species follows a birth-death process with constant rates λ and μ both through time and lineage, starting at origin time s and ending at time e which is generally the present time [20]. Following [37], each extant species is assumed to be independently sampled at the end time e with probability ρ. The whole model will be referred to as the *birth-death-sampling model* and has thus five parameters:

- s: the origin time of the diversification process,
- e: the end/present time,
- λ: the speciation rate,
- μ: the extinction rate and
- ρ: the probability for an ending/extant taxa to be sampled.

Let us start by recalling some already derived probabilities of interest. By assuming that the diversification follows a simple birth-death process (i.e., with $\rho = 1$) with speciation rate λ and extinction rate μ, the probability $p_N(t)$ that a single lineage at time 0 has exactly N descendants at time t was given in [20]. We have that

$$p_0(t) = \frac{\mu\left(1 - \mathrm{e}^{-(\lambda-\mu)t}\right)}{\lambda - \mu\mathrm{e}^{-(\lambda-\mu)t}}, \quad \text{and for all } N > 0,$$

$$p_N(t) = (\lambda - \mu)^2 \mathrm{e}^{-(\lambda-\mu)t} \frac{\left(\lambda(1 - \mathrm{e}^{-(\lambda-\mu)t})\right)^{N-1}}{\left(\lambda - \mu\mathrm{e}^{-(\lambda-\mu)t}\right)^{N+1}}.$$

If one assumes that the diversification follows a birth-death-sampling process with speciation rate λ and extinction rate μ, the probability $\dot{p}_N(t)$ that a single lineage at time 0 has exactly N descendants sampled with probability ρ at time t was given in [37]. We have that

$$\dot{p}_0(t) = \frac{\mu(1-\mathrm{e}^{-(\lambda-\mu)t}) + (1-\rho)\left(\lambda\mathrm{e}^{-(\lambda-\mu)t} - \mu\right)}{\rho\lambda + (\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)t}}, \quad \text{and for all } N > 0,$$

$$\dot{p}_N(t) = \rho^N(\lambda-\mu)^2\mathrm{e}^{-(\lambda-\mu)t}\frac{\left(\lambda(1-\mathrm{e}^{-(\lambda-\mu)t})\right)^{N-1}}{\left(\rho\lambda + (\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)t}\right)^{N+1}}.$$

Let $\Theta = (s, e, \lambda, \mu, \rho)$ be a birth-death-sampling model. For all pair of times t and t' with $s \leq t \leq t'$ and all number N, we define $\mathbf{Q}_\Theta(t, t', 0)$ as the probability under the model Θ that a lineage alive at time t has N descendants alive at time t' if $t' < e$ and N descendants alive and sampled if $t' = e$. We have that

$$\mathbf{Q}_\Theta(t,t',0) = \begin{cases} \frac{\mu\left(1-\mathrm{e}^{-(\lambda-\mu)(t'-t)}\right)}{\lambda-\mu\mathrm{e}^{-(\lambda-\mu)(t'-t)}} & \text{if } t' < e, \\ \frac{\rho\mu+(\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)(e-t)}}{\rho\lambda+(\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)(e-t)}} & \text{if } t' = e \text{ and for all } N > 0, \end{cases}$$

$$\mathbf{Q}_\Theta(t,t',N) = \begin{cases} (\lambda-\mu)^2\mathrm{e}^{-(\lambda-\mu)(t'-t)}\frac{\left(\lambda(1-\mathrm{e}^{-(\lambda-\mu)(t'-t)})\right)^{N-1}}{\left(\lambda-\mu\mathrm{e}^{-(\lambda-\mu)(t'-t)}\right)^{N+1}} & \text{if } t' < e, \\ \rho^N(\lambda-\mu)^2\mathrm{e}^{-(\lambda-\mu)(e-t)}\frac{\left(\lambda(1-\mathrm{e}^{-(\lambda-\mu)(e-t)})\right)^{N-1}}{\left(\rho\lambda+(\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)(e-t)}\right)^{N+1}} & \text{if } t' = e. \end{cases}$$

The probability $\mathbf{O}_\Theta(t)$ for a lineage living at time t in the complete diversification process (as in Figure 1-Left) to be observable (i.e., to be part of the reconstructed process) is the complementary probability of having no descendant sampled at time e. We have that

$$\begin{aligned} \mathbf{O}_\Theta(t) &= 1 - \mathbf{Q}_\Theta(0,t,e) \\ &= \frac{\rho(\lambda-\mu)}{\rho\lambda + (\lambda(1-\rho)-\mu)\mathrm{e}^{-(\lambda-\mu)(e-t)}}. \end{aligned}$$

By construction, birth-death-sampling models are both Markovian and lineage-homogeneous.

3 Tree topologies

Tree topologies arising from diversification processes are binary (since a speciation event gives rise to a single new lineage under the models considered here) and rooted thus so are all the tree topologies considered here. Moreover, all the tree topologies considered below will be *labeled*, which means their tips, and consequently all their nodes, are unambiguously identified. From now on, "tree topology" has to be understood as "labeled-rooted-binary tree topology".

Since the context will avoid any confusion, we still write $\mathcal{T}$ for the set of nodes of any tree topology $\mathcal{T}$. For all tree topologies $\mathcal{T}$, we put $\mathrm{L}_\mathcal{T}$ for the set of tips of $\mathcal{T}$. For all nodes n of $\mathcal{T}$, we note $\mathcal{T}_n$ the subtree of $\mathcal{T}$ rooted at n.

For all sets S, $|S|$ denotes the cardinality of S. In particular, $|\mathcal{T}|$ denotes the size of the tree topology $\mathcal{T}$ (i.e., its total number of nodes, internal or tips) and $|\mathrm{L}_\mathcal{T}|$ its number of tips.

3.1 Probability

Let us define $\mathbf{T}_\Theta(\mathcal{T})$ as the probability of a reconstructed tree topology $\mathcal{T}$ given its number of tips under a lineage-homogeneous diversification process.

Theorem 1 ([13])**.** *Given its number of tips, the reconstructed tree topology $\mathcal{T}$ of a realization of a lineage-homogeneous diversification process has probability $\mathbf{T}(\mathcal{T}) = 1$ if $|\mathcal{T}| = 1$, i.e., $\mathcal{T}$ is a single lineage. Otherwise, by putting a and b for the two direct descendants of the root of $\mathcal{T}$, the probability of the tree topology $\mathcal{T}$ is*

$$\mathbf{T}(\mathcal{T}) = \frac{2|\mathrm{L}_{\mathcal{T}_a}|!\,|\mathrm{L}_{\mathcal{T}_b}|!}{(|\mathrm{L}_\mathcal{T}|-1)|\mathrm{L}_\mathcal{T}|!}\mathbf{T}(\mathcal{T}_a)\mathbf{T}(\mathcal{T}_b).$$

Assumptions of [13] are slightly different from those of Theorem 1 but its arguments still hold. The probability provided in [3, Supp. Mat., Appendix 2] is actually the same as that just above though it was derived in a different way from [13] and expressed in a slightly different form (see [4, Appendix 1]).

Theorem 1 implies in particular that $\mathbf{T}(\mathcal{T})$ can be computed in linear time through a post-order traversal of the tree topology $\mathcal{T}$.

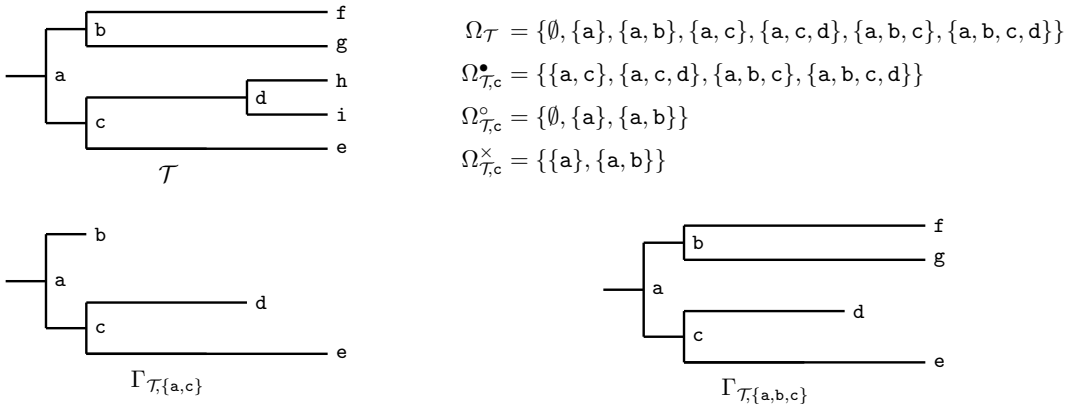

Figure 2: A tree $\mathcal{T}$ (top-left), examples of sets of start-sets of $\mathcal{T}$ of various types (top-right) and the start-trees of $\mathcal{T}$ associated to the start-sets $\{\mathtt{a},\mathtt{c}\}$ and $\{\mathtt{a},\mathtt{b},\mathtt{c}\}$ (bottom).

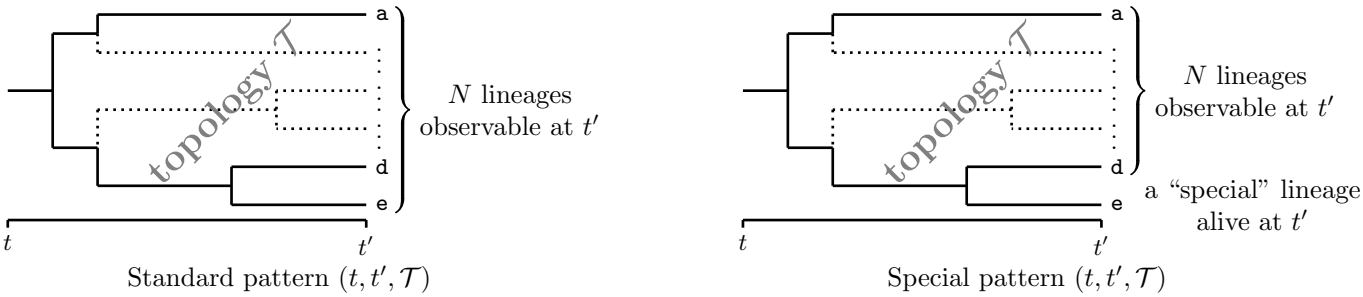

Figure 3: The two types of patterns used to compute probability distributions.

3.2 Start-sets

A *start-set* of a tree topology $\mathcal{T}$ is a possibly empty subset A of internal nodes of $\mathcal{T}$ which is such that if an internal node of $\mathcal{T}$ belongs to A then so do all its ancestors. Remark that, basically, the empty set $\emptyset$ is start-set of any tree topology and that if A and A' are two start-sets of $\mathcal{T}$ then both $A \cup A'$ and $A \cap A'$ are start-sets of $\mathcal{T}$.

Being given a tree topology $\mathcal{T}$ and a non-empty start-set A, we define the *start-tree* $\Gamma_{\mathcal{T},A}$ as the subtree topology of $\mathcal{T}$ made of all nodes in A and their direct descendants. By convention, $\Gamma_{\mathcal{T},\emptyset}$, the start-tree associated to the empty start-set, is the subtree topology made only of the root of $\mathcal{T}$.

For all tree topologies $\mathcal{T}$, we define

- $\Omega_{\mathcal{T}}$ as the set of all start-sets of $\mathcal{T}$, and for all internal nodes n,
- $\Omega^{\bullet}_{\mathcal{T},n}$ as the set of all start-sets A of $\mathcal{T}$ such that $n \in A$,
- $\Omega^{\circ}_{\mathcal{T},n}$ as the set of all start-sets A of $\mathcal{T}$ such that $n \notin A$, and
- $\Omega^{\times}_{\mathcal{T},n}$ as the set of all start-sets A of $\mathcal{T}$ such that n is a tip of $\Gamma_{\mathcal{T},A}$.

Figure 2 displays examples of start-trees and of sets of start-sets of the types above.

4 Patterns

In this section, we shall consider diversification processes starting at origin time s and ending at time e following a birth-death-sampling model $\Theta = (s, e, \lambda, \mu, \rho)$. A *pattern* is a part of the observed diversification process starting from a single lineage at a given time and ending with a certain number of lineages at another given time, these ending lineages being either observable or *special*, where "special" means "only known to be alive at the end time and distinguished for some reason". It consists of a 3-tuple $(t, t', \mathcal{T})$ where t and t' are the start and end times of the pattern and $\mathcal{T}$ is the resulting tree topology. We shall consider two types of patterns: standard and special patterns. Standard patterns ends with only observable lineages. All the ending lineages of a special pattern are observable except one which is special (Fig. 3). Standard and special patterns are very similar to patterns defined in [3] for the fossilized-birth-death process.

4.1 Standard patterns

Definition 1. *A standard pattern $(t, t', \mathcal{T})$ starts with a single lineage at time t and ends with a tree topology $\mathcal{T}$ and $|\mathrm{L}_{\mathcal{T}}|$ observable lineages at time t' (Fig. 3-left).*

Let us compute the probability $\mathbf{X}_\Theta(t, t', n)$ that a single lineage at time $t \in [s, e)$ has N descendants observable at time $t' \in (t, e]$ under the diversification model Θ. This probability is the sum over all numbers $j \geq 0$, of the probability that the lineage at t has $j + N$ descendants at t' in the whole process, which is equal to $\mathbf{Q}_\Theta(t, t', j + N)$, among which exactly n ones are observable (i.e., $\binom{j+N}{N}\mathbf{O}_\Theta(t')^N (1 - \mathbf{O}_\Theta(t'))^j$). Under the diversification model Θ, we thus have

$$\mathbf{X}_\Theta(t, t', N) = \sum_{j=0}^{\infty} \mathbf{Q}_\Theta(t, t', j + N)\binom{j + N}{N}\mathbf{O}_\Theta(t')^N (1 - \mathbf{O}_\Theta(t'))^j . \tag{1}$$

If Θ is the birth-death-sampling model $(s, e, \lambda, \mu, \rho)$, Equation 1 becomes

$$\mathbf{X}_\Theta(t, t', N) = \frac{(\lambda - \mu)^2 \mathrm{e}^{-(\lambda-\mu)(t'-t)} \left(\lambda(1 - \mathrm{e}^{-(\lambda-\mu)(t'-t)})\right)^{N-1} \mathbf{O}_\Theta(t')^N}{\left(\lambda\mathbf{O}_\Theta(t') + (\lambda(1 - \mathbf{O}_\Theta(t')) - \mu)\mathrm{e}^{-(\lambda-\mu)(t'-t)}\right)^{N+1}}.$$

The probability of the standard pattern $(t, t', \mathcal{T})$ is the probability of the tree topology $\mathcal{T}$ conditioned on its number of tips, which is $\mathbf{T}_\Theta(\mathcal{T})$ multiplied by the probability of observing this number of tips in a standard pattern, which is that of getting $|\mathrm{L}_{\mathcal{T}}|$ observable lineages at t' from a single lineage at t, i.e., $\mathbf{X}_\Theta(t, t', |\mathrm{L}_{\mathcal{T}}|)$.

Claim 1. *Under the diversification model Θ, the probability of the standard pattern $(t, t', \mathcal{T})$ with $s \leq t < t' \leq e$ is*

$$\mathbf{T}_\Theta(\mathcal{T})\mathbf{X}_\Theta(t, t', |\mathrm{L}_{\mathcal{T}}|),$$

where $\mathbf{T}_\Theta = \mathbf{T}$ if Θ is lineage-homogeneous.

4.2 Special patterns

Definition 2. *A special pattern $(t, t', \mathcal{T})$ starts with a single lineage at time $t \in [s, e)$ and ends with the tree topology $\mathcal{T}$ at $t' \in (t, e]$, thus with $|\mathrm{L}_{\mathcal{T}}|$ descendants at t' among which $|\mathrm{L}_{\mathcal{T}}|-1$ are observable and one is a distinguished "special" lineage of fate a priori unknown after t' (Fig. 3-right).*

Let us now compute the probability $\mathbf{Y}_\Theta(t, t', N + 1)$ that a single lineage at time $t \in [s, e)$ has one special descendant and N descendants observable from e at time $t' \in (t, e]$. This probability is the sum over all numbers j, of the probability that the lineage at t has $j + N + 1$ descendants at t' in the whole process, which is equal to $\mathbf{Q}_\Theta(t, t', j+N+1)$, among which the special one is picked, exactly n ones are observable and j ones are not observable, which leads to $(j + N + 1)\binom{j+N}{N} = (N + 1)\binom{j+N+1}{N+1}$ possibilities. Under the diversification model Θ, we have that

$$\mathbf{Y}_\Theta(t, t', N + 1) = \sum_{j=0}^{\infty} \mathbf{Q}_\Theta(t, t', j + N + 1)(N + 1)\binom{j + N + 1}{N + 1}\mathbf{O}_\Theta(t')^N (1 - \mathbf{O}_\Theta(t'))^j . \tag{2}$$

If Θ is the birth-death-sampling model $(s, e, \lambda, \mu, \rho)$, Equation 2 gives us that

$$\mathbf{Y}_\Theta(t, t', N + 1) = \frac{(N + 1)(\lambda - \mu)^2 \mathrm{e}^{-(\lambda-\mu)(t'-t)} \left(\lambda(1 - \mathrm{e}^{-(\lambda-\mu)(t'-t)})\mathbf{O}_\Theta(t')\right)^N}{\left(\lambda\mathbf{O}_\Theta(t') + (\lambda(1 - \mathbf{O}_\Theta(t')) - \mu)\mathrm{e}^{-(\lambda-\mu)(t'-t)}\right)^{N+2}}.$$

The probability of the special pattern $(t, t', \mathcal{T})$ is the probability of the tree topology $\mathcal{T}$ conditioned on its number of tips, which is $\mathbf{T}_\Theta(\mathcal{T})$ multiplied by the probability of observing this ending configuration in a special pattern, i.e., $\mathbf{Y}_\Theta(t, t', |\mathrm{L}_{\mathcal{T}}|)$.

Claim 2. *Under the diversification model Θ, the probability of the special pattern $(t, t', \mathcal{T})$ with $s \leq t < t' \leq e$ is*

$$\mathbf{T}_\Theta(\mathcal{T})\mathbf{Y}_\Theta(t, t', |\mathrm{L}_{\mathcal{T}}|),$$

where $\mathbf{T}_\Theta = \mathbf{T}$ if Θ is lineage-homogeneous.

5 Probability densities of topologies with temporal constraints and shifts

The probability $\mathbf{P}_\Theta(\mathcal{T})$ of observing a tree topology $\mathcal{T}$ under a diversification model Θ with origin and end times s and e is that of the corresponding standard pattern, i.e., we have that

$$\mathbf{P}_\Theta(\mathcal{T}) = \mathbf{T}_\Theta(\mathcal{T})\mathbf{X}_\Theta(s, e, |\mathrm{L}_{\mathcal{T}}|).$$

We shall see in this section how to compute the probability of a tree topology under the constraint that some of its divergence times are known to be anterior or posterior to given times.

5.1 Temporal constraints

Let us put τ_n for the (random variable associated to the) divergence time corresponding to the node n of $\mathcal{T}$. Being given internal nodes $n_1, \ldots, n_\ell$, $n'_1, \ldots, n'_{\ell'}$ of $\mathcal{T}$ and times $u_1, \ldots, u_\ell$, $u'_1, \ldots, u'_{\ell'}$ between s and e (both not included), we aim to compute the joint probability of $\mathcal{T}$ and of observing $\tau_{n_1} < u_1, \ldots, \tau_{n_\ell} < u_\ell$, $\tau_{n'_1} > u'_1, \ldots, \tau_{n'_{\ell'}} > u'_{\ell'}$ under the model Θ, i.e.,

$$\mathbf{P}_\Theta(\mathcal{T}, \mathcal{U}, \mathcal{L}) = \mathbf{P}_\Theta(\mathcal{T}, \tau_{n_1} < u_1, \ldots, \tau_{n_\ell} < u_\ell, \tau_{n'_1} > u'_1, \ldots, \tau_{n'_{\ell'}} > u'_{\ell'}).$$

The temporal constraints induced by the tree topology, i.e., that we have necessarily $\tau_n \leq \tau_m$ if n is an ancestor of m are implicitly assumed granted in the probability above. The constraints $\tau_{n_1} \leq u_1, \ldots, \tau_{n_\ell} \leq u_\ell$ will be referred to as *upper temporal constraints* and summarized as the set of pairs "node-time" $\mathcal{U} = \{(n_1, u_1), \ldots, (n_\ell, u_\ell)\}$, and the constraints $\tau_{n'_1} \geq u'_1, \ldots, \tau_{n'_{\ell'}} \geq u'_{\ell'}$, will be referred to as *lower temporal constraints* and summarized as the set of pairs $\mathcal{L} = \{(n'_1, u'_1), \ldots, (n'_{\ell'}, u'_{\ell'})\}$. We assume that the temporal constraints are consistent with another (otherwise they would basically lead to a null probability). For all subsets of internal nodes $\mathcal{S}$ of $\mathcal{T}$, we write $\mathcal{U}_{[\mathcal{S}]}$ (resp. $\mathcal{L}_{[\mathcal{S}]}$) for the set of upper (resp. lower) temporal constraints of $\mathcal{U}$ (resp. of $\mathcal{L}$) involving nodes in $\mathcal{S}$, namely $\mathcal{U}_{[\mathcal{S}]} = \{(n_j, u_j) \mid (n_j, u_j) \in \mathcal{U} \text{ and } n_j \in \mathcal{S}\}$ (resp. $\mathcal{L}_{[\mathcal{S}]} = \{(n'_j, u'_j) \mid (n'_j, u'_j) \in \mathcal{L} \text{ and } n'_j \in \mathcal{S}\}$). For all times t, we define $\mathcal{U}^{=t}$ (resp. $\mathcal{L}^{=t}$) as the set of temporal constraints of $\mathcal{U}$ (resp. $\mathcal{L}$) involving t, namely, $\mathcal{U}^{=t} = \{(n_j, u_j) \mid (n_j, u_j) \in \mathcal{U} \text{ and } u_j = t\}$ (resp. $\mathcal{L}^{=t} = \{(n'_j, u'_j) \mid (n'_j, u'_j) \in \mathcal{L} \text{ and } u'_j = t\}$). In the same way, we define $\mathcal{U}^{>t}$ and $\mathcal{L}^{>t}$ as the subsets of temporal constraints of $\mathcal{U}$ and $\mathcal{L}$ respectively, which involved times strictly posterior to t, i.e., $\mathcal{U}^{>t} = \{(n_j, u_j) \mid (n_j, u_j) \in \mathcal{U} \text{ and } u_j > t\}$ and $\mathcal{L}^{>t} = \{(n'_j, u'_j) \mid (n'_j, u'_j) \in \mathcal{L} \text{ and } u'_j > t\}$.

Theorem 2. *Let $\mathcal{T}$ be a tree topology, Θ be a Markovian diversification model from origin time s to end time e and $\mathcal{U} = \{(n_1, u_1), \ldots, (n_\ell, u_\ell)\}$ and $\mathcal{L} = \{(n'_1, u'_1), \ldots, (n'_{\ell'}, u'_{\ell'})\}$ be two sets of upper and lower temporal constraints respectively. Let o be the oldest time involved in a temporal constraint or the end time if there are none, namely,*

$$o = \min\{e, \min\{t \mid \exists n \in \mathcal{T} \text{ such that } (n,t) \in \mathcal{U}\}, \min\{t \mid \exists n \in \mathcal{T} \text{ such that } (n,t) \in \mathcal{L}\}\}.$$

Let us define the set $\mathcal{S}$ of internal node subsets of $\mathcal{T}$ as the intersection of

$$\begin{cases} \bigcap_{(n,o) \in \mathcal{U}^{=o}} \Omega^{\bullet}_{\mathcal{T},n} & \text{if } \mathcal{U}^{=o} \neq \emptyset, \\ \Omega_{\mathcal{T}} & \text{otherwise,} \end{cases} \quad \text{and} \quad \begin{cases} \bigcap_{(n,o) \in \mathcal{L}^{=o}} \Omega^{\circ}_{\mathcal{T},n} & \text{if } \mathcal{L}^{=o} \neq \emptyset, \\ \Omega_{\mathcal{T}} & \text{otherwise.} \end{cases}$$

The joint probability $\mathbf{P}_\Theta(\mathcal{T}, \mathcal{U}, \mathcal{L})$ of observing the tree topology $\mathcal{T}$ with the temporal constraints $\mathcal{U}$ and $\mathcal{L}$ under Θ verifies

$$\mathbf{P}_\Theta(\mathcal{T}, \mathcal{U}, \mathcal{L}) = \begin{cases} \dfrac{1}{|\mathrm{L}_{\mathcal{T}}|!} \displaystyle\sum_{A \in \mathcal{S}} |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|! \, \mathbf{T}_\Theta(\Gamma_{\mathcal{T},A}) \mathbf{X}_\Theta(s, o, |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|) \prod_{n \in \mathrm{L}_{\Gamma_{\mathcal{T},A}}} \dfrac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n, \mathcal{U}^{>o}_{[\mathcal{T}_n]}, \mathcal{L}^{>o}_{[\mathcal{T}_n]}) |\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)} & \text{if } o < e, \\ \mathbf{T}_\Theta(\mathcal{T}) \mathbf{X}_\Theta(s, e, |\mathrm{L}_{\mathcal{T}}|) & \text{otherwise,} \end{cases}$$

where $\Theta_{[o]}$ is the model Θ restricted to the time interval $[o, e]$.

Proof. Appendix C.1. □

The general idea of the computation presented in Theorem 2 is first to consider the oldest time involved in a temporal constraint (referred to as o in Theorem 2 and which is t in Figure 4), then to consider the part of the diversification process which occurred before this oldest time and the part(s) which occurred after the oldest time.

An issue here is that these parts are a priori unknown since, by construction, they are determined by the relative positions of the divergence times with regard to the oldest time. If some of the divergence times are directly or indirectly (from their ancestors or their descendants) constrained, thus known, to be posterior or anterior to the oldest time, some of them may be not and we have to consider all the possibilities consistent with the temporal constraints as displayed in the left-hand side of the second column of Figure 4. Remark that since all these possibilities are mutually exclusive, their respective probabilities can be summed in order to obtained the probability of the initial tree topology with temporal constraints.

In order to compute the probability of each possibility (i.e., in which all the relative positions of the divergence with regard to the oldest time are fixed), we use the Markov property to write the probability of the tree topology with the temporal constraints as the product of the probability of the part of the diversification which happened before the oldest time (which by construction contains no temporal constraint and therefore is a standard pattern) to the probabilities of the parts which happened after the oldest which is a set of independent tree topologies starting from the oldest time with possibly some time constraints, which all involved times posterior to the oldest time. This part of the computation is illustrated by the five lines of Column 2 in Figure 4, in which the right-hand sides start with standard patterns corresponding to the diversification anterior to the oldest time which are followed by a certain number of tree topologies starting from the oldest time (in this case without time constraints since the initial contained only one constraint) corresponding to the diversification posterior to the oldest time. For all the possibilities, Claim 1

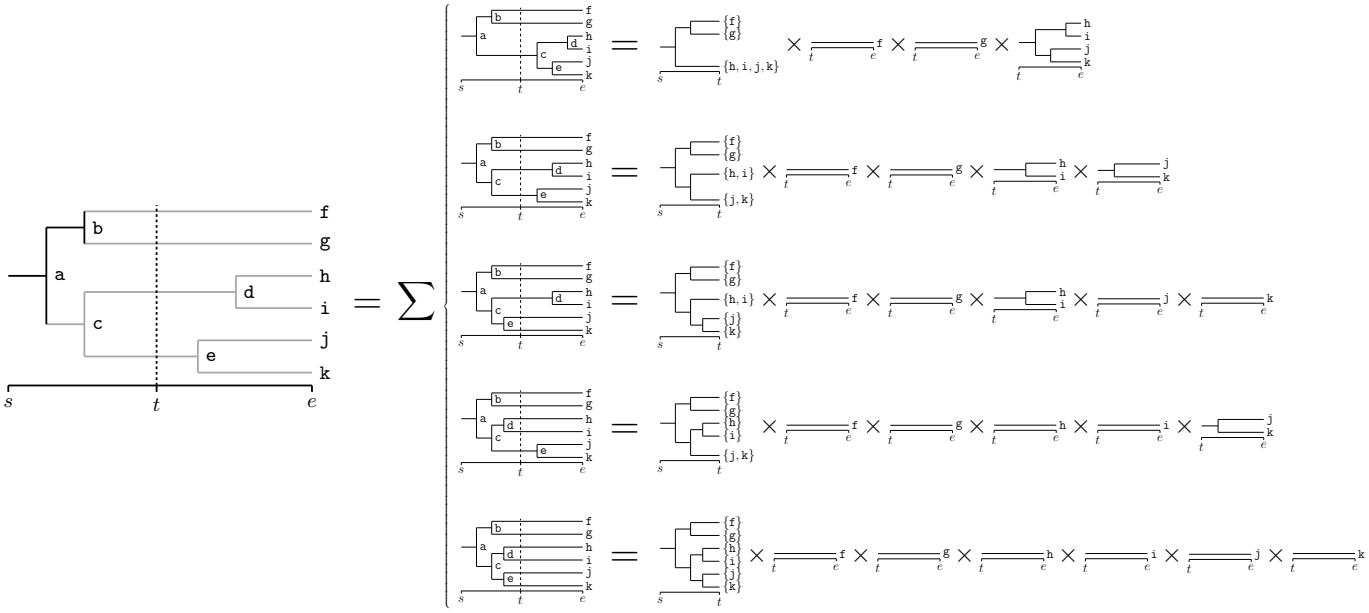

Figure 4: Schematic of the computation of the probability $\mathbf{P}_\Theta(\mathcal{T}, \{(\mathtt{b}, t)\}, \emptyset)$, i.e., that the divergence time associated with node b is strictly anterior to t. Under the notations of Theorem 2, we have that $\mathcal{S} = \Omega^\bullet_{\mathcal{T},\mathtt{b}}$. Nothing is known about divergence times in the gray part of the tree at the left. The only information about divergence times in black parts of all trees is their relative position with regard to t.

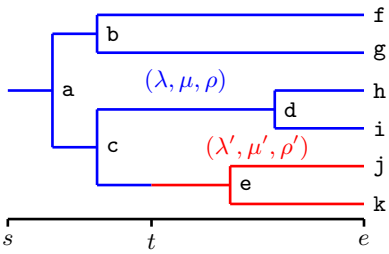

Figure 5: A tree topology with a shift at time t for the clade $\{\mathtt{e}, \mathtt{j}, \mathtt{k}\}$.

allows us to directly compute the part of diversification anterior to the oldest time. The parts posterior to the oldest time are tree topologies starting from the oldest time simpler than the initial one which can be themselves handled by Theorem 2.

Theorem 2 states that $\mathbf{P}_\Theta(\mathcal{T}, \mathcal{U}, \mathcal{L})$ can be either calculated directly (if $o = e$) or expressed as a sum-product of probabilities of tree topologies with temporal constraints under birth-death-sampling models whose starting time is strictly posterior to the starting time of Θ, on which Theorem 2 can be applied and so on. Since each time that Theorem 2 is applied, we get tree topologies under models and temporal constraints in which the starting time has been discarded, we eventually end up in the case where the oldest time is the end time of the diversification for which the probability can be calculated directly. To summarize, the probability $\mathbf{P}_\Theta(\mathcal{T}, \mathcal{U}, \mathcal{L})$ can be computed by recursively applying Theorem 2.

5.2 Shifts

We shall see how to compute the probability of a tree topology $\mathcal{T}$ under a diversification model Θ starting at s and ending at e by assuming that one of its clades follows another diversification model $\widetilde{\Theta}$ from a given time $t \in [s, e]$ to the end time e. Note that this implicitly assumes that the lineage originating this particular clade was alive at t (Fig. 5).

Theorem 3. *Let $\mathcal{T}$ be a tree topology, $s \leq t \leq e$ be three times, Θ and $\widetilde{\Theta}$ be two Markovian diversification models from origin times s and t respectively and both to end time e, and m be a node of $\mathcal{T}$. By putting $\Theta_{[t]}$ for the model Θ restricted to $[t, e]$, the probability $\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T}, m, t)$ of observing the tree topology $\mathcal{T}$ assuming that evolution follows Θ on $\mathcal{T}$ except on $\mathcal{T}_m$ on which it follows $\widetilde{\Theta}$ from time t verifies*

$$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T}, m, t) = \frac{1}{|\mathrm{L}_{\mathcal{T}}|!} \sum_{A \in \Omega^\times_{\mathcal{T},m}} (|\mathrm{L}_{\Gamma_{\mathcal{T},A}}| - 1)!\, \mathbf{T}_\Theta(\Gamma_{\mathcal{T},A}) \mathbf{Y}_\Theta(s, t, |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|) \mathbf{P}_{\widetilde{\Theta}}(\mathcal{T}_m) |\mathrm{L}_{\mathcal{T}_m}|! \prod_{n \in \mathrm{L}_{\Gamma_{\mathcal{T},A}} \setminus \{m\}} \frac{\mathbf{P}_{\Theta_{[t]}}(\mathcal{T}_n) |\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(t)}.$$

Proof. Appendix C.2. □

The idea of the proof is essentially the same as that of Theorem 2.

Let us remark that the trees starting from t are standard patterns. It follows that $\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t)$ can be equivalently written as

$$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t) = \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{A\in\Omega^{\times}_{\mathcal{T},m}}(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\mathbf{T}_{\Theta}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\mathbf{T}_{\Theta}(\mathcal{T}_m)\mathbf{X}_{\widetilde{\Theta}}(t,e,|\mathrm{L}_{\mathcal{T}_m}|)|\mathrm{L}_{\mathcal{T}_m}|!\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{T}_{\Theta}(\mathcal{T}_n)\mathbf{X}_{\Theta_{[t]}}(t,e,|\mathrm{L}_{\mathcal{T}_n}|)|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(t)}.$$

6 A quadratic computation

Since the number of start-sets may be exponential with the size of the tree, notably for balanced trees, Equations of Theorems 2 and 3 do not directly provide a polynomial algorithm for computing the probabilities considered in these theorems. In the case where the diversification is lineage-homogeneous, the form of the probability of the tree topology conditioned on its number of tips provided by Theorem 1 allows us to factorize the computation of Theorem 2 (and of Theorem 3) in order to obtain a polynomial algorithm. Let us sketch the general idea of this computation.

In the case where the diversification model Θ is lineage-homogeneous, Theorem 1 implies that $\mathbf{T}_{\Theta}=\mathbf{T}$. Let us assume that the temporal constraints are such that $\mathcal{U}\cup\mathcal{L}\neq\emptyset$ and let a and b be the two direct descendants of the root of $\mathcal{T}$. For all start sets A of $\mathcal{T}$, we define Λ^a_A and Λ^b_A as the subtrees of $\Gamma_{\mathcal{T},A}$ rooted at a and b respectively, namely $\Lambda^a_A=\Gamma_{\mathcal{T}_a,A\cap\mathcal{T}_a}$ and $\Lambda^b_A=\Gamma_{\mathcal{T}_b,A\cap\mathcal{T}_b}$. From Theorems 2 and 1, we have that

$$\begin{aligned}\mathbf{P}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L}) &= \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{A\in\mathcal{S}}|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_{\Theta}(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(o)}\\ &= \frac{2}{(|\mathrm{L}_{\mathcal{T}}|-1)(|\mathrm{L}_{\mathcal{T}}|!)^2}\sum_{A\in\mathcal{S}}\frac{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{X}_{\Theta}(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)}{\mathbf{O}_{\Theta}(o)^{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|}}\times|\mathrm{L}_{\Lambda^a_A}|!\,|\mathrm{L}_{\Lambda^b_A}|!\,\mathbf{T}(\Lambda^a_A)\mathbf{T}(\Lambda^b_A)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\end{aligned}$$

Since by construction a tip of $\Gamma_{\mathcal{T},A}$ is either a tip of Λ^a_A or a tip of Λ^b_A, we have basically that

$$\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|! = \left(\prod_{n\in\mathrm{L}_{\Lambda^a_A}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\right)\times\left(\prod_{n\in\mathrm{L}_{\Lambda^b_A}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\right).$$

It follows that

$$\mathbf{P}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L}) = \frac{2}{(|\mathrm{L}_{\mathcal{T}}|-1)(|\mathrm{L}_{\mathcal{T}}|!)^2}\sum_{A\in\mathcal{S}}f(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\times g(\Lambda^a_A)\times g(\Lambda^b_A),$$

where

$$\begin{aligned}f(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|) &= \frac{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{X}_{\Theta}(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)}{\mathbf{O}_{\Theta}(o)^{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|}},\\ g(\Lambda^a_A) &= |\mathrm{L}_{\Lambda^a_A}|!\,\mathbf{T}(\Lambda^a_A)\prod_{n\in\mathrm{L}_{\Lambda^a_A}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!,\quad\text{and,}\\ g(\Lambda^b_A) &= |\mathrm{L}_{\Lambda^b_A}|!\,\mathbf{T}(\Lambda^b_A)\prod_{n\in\mathrm{L}_{\Lambda^b_A}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!.\end{aligned}$$

In plain English, computing $\mathbf{P}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L})$ requires to sum over all start sets A in $\mathcal{S}$, $f(|\mathrm{L}|)\times g(\Lambda^a_A)\times g(\Lambda^b_A)$, a product of three factors where the first one depends only on the number of tips of $\Gamma_{\mathcal{T},A}$ and the two following ones depend on the subtrees Λ^a_A and Λ^b_A respectively (i.e., on the start set A and on the subtrees $\mathcal{T}_a$ and $\mathcal{T}_b$).

Let us start by regrouping the terms of the sum with respect to the number of tips of the corresponding start trees:

$$\sum_{A\in\mathcal{S}}f(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\times g(\Lambda^a_A)\times g(\Lambda^b_A) = \sum_k f(k)\times\sum_{A\in\mathcal{S},|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|=k}g(\Lambda^a_A)\times g(\Lambda^b_A).$$

Next, we put $\Delta^a_{\mathcal{S}}$ (resp. $\Delta^b_{\mathcal{S}}$) for the set of subtrees rooted at a (resp. at b) of the start trees $\Gamma_{\mathcal{T},A}$ for all $A\in\mathcal{S}$, namely $\Delta^a_{\mathcal{S}}=\{\Lambda^a_A\mid A\in\mathcal{S}\}$ and $\Delta^b_{\mathcal{S}}=\{\Lambda^b_A\mid A\in\mathcal{S}\}$. Moreover, for all $A\in\mathcal{S}$ and since Λ^a_A and Λ^b_A are the two subtrees pending from the root of $\Gamma_{\mathcal{T},A}$, we have that $|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|=k$ if and only if $|\mathrm{L}_{\Lambda^a_A}|+|\mathrm{L}_{\Lambda^b_A}|=k$. For all k, a set A belongs to $\mathcal{S}$ and is such that $|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|=k$ if and only if there exist a tree $\mathcal{T}'\in\Delta^a_{\mathcal{S}}$ and a tree $\mathcal{T}''\in\Delta^b_{\mathcal{S}}$ such

that A is the union of the root of $\mathcal{T}$ and of nodes of $\mathcal{T}'$ and $\mathcal{T}''$ and $|\mathrm{L}_{\mathcal{T}'}|+|\mathrm{L}_{\mathcal{T}''}|= k$. It follows that the terms of $\sum_{A\in\mathcal{S},|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|=k} g(\Lambda_A^a)\times g(\Lambda_A^b)$ can be regrouped and factorized in order to get that

$$\sum_{A\in\mathcal{S},|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|=k} g(\Lambda_A^a)\times g(\Lambda_A^b) = \sum_{i=1}^{k-1}\left(\sum_{\mathcal{T}'\in\Delta_{\mathcal{S}}^a,|\mathrm{L}_{\mathcal{T}'}|=i} g(\mathcal{T}')\right)\times\left(\sum_{\mathcal{T}''\in\Delta_{\mathcal{S}}^b,|\mathrm{L}_{\mathcal{T}''}|=k-i} g(\mathcal{T}'')\right).$$

To summarize, if one assume that the quantities $\sum_{\mathcal{T}'\in\Delta_{\mathcal{S}}^a,|\mathrm{L}_{\mathcal{T}'}|=i} g(\mathcal{T}')$ and $\sum_{\mathcal{T}''\in\Delta_{\mathcal{S}}^b,|\mathrm{L}_{\mathcal{T}''}|=k-i} g(\mathcal{T}'')$ are known, we have written $\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ as a sum of a quadratic number of terms, namely

$$\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \frac{2}{(|\mathrm{L}_{\mathcal{T}}|-1)(|\mathrm{L}_{\mathcal{T}}|!)^2}\sum_k f(k)\sum_{i=1}^{k-1}\left(\sum_{\mathcal{T}'\in\Delta_{\mathcal{S}}^a,|\mathrm{L}_{\mathcal{T}'}|=i} g(\mathcal{T}')\right)\times\left(\sum_{\mathcal{T}''\in\Delta_{\mathcal{S}}^b,|\mathrm{L}_{\mathcal{T}''}|=k-i} g(\mathcal{T}'')\right).$$

Appendix C.3 shows how the quantities $\sum_{\mathcal{T}'\in\Delta_{\mathcal{S}}^a,|\mathrm{L}_{\mathcal{T}'}|=i} g(\mathcal{T}')$ and $\sum_{\mathcal{T}''\in\Delta_{\mathcal{S}}^b,|\mathrm{L}_{\mathcal{T}''}|=k-i} g(\mathcal{T}'')$, which are referred to as $\mathbf{W}_{a,i}$ and $\mathbf{W}_{b,k-i}$ respectively in the appendix, can be recursively computed in order to obtain a computation with total complexity quadratic with the size of $\mathcal{T}$. We eventually obtain the following theorem.

Theorem 4. *Let $\mathcal{T}$ be a tree topology, Θ be a lineage-homogeneous Markovian diversification model and $\mathcal{U}$ and $\mathcal{L}$ be two sets of upper and lower temporal constraints respectively. If the probability of the ending configuration of any standard pattern can be computed in constant time then the probability $\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ can be computed with time complexity $O(|\mathcal{U}\cup\mathcal{L}|\times|\mathcal{T}|^2)$ and memory space complexity $O(|\mathcal{T}|^2)$.*

Proof. Appendix C.3 □

Theorem 4 holds in particular for birth-death-sampling models and sampled-generalized-birth-death models presented in Appendix B.

It can be proved in the same way that the shift probability $\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t)$ of Theorem 3 can be computed with time and memory space complexity $O(|\mathcal{T}|^2)$ if the probability of the ending configuration of any standard or special pattern can be computed in constant time and if the diversification process is lineage-homogeneous.

7 Divergence time distributions

We shall apply Theorem 2 to compute divergence time distributions of tree topologies with temporal constraints.

Claim 3. *Let $\mathcal{T}$ be a tree topology, Θ be a diversification model from origin time s to end time e, $\mathcal{U}=\{(n_1,u_1),\dots,(n_\ell,u_\ell)\}$ and $\mathcal{L}=\{(n'_1,u'_1),\dots,(n'_{\ell'},u'_{\ell'})\}$ be two sets of upper and lower temporal constraints respectively and m be an internal node of $\mathcal{T}$. The probability that the divergence time τ_m associated with m is anterior to a time $t\in[s,e]$ conditioned on observing the tree topology $\mathcal{T}$ with the temporal constraints $\mathcal{U}$ and $\mathcal{L}$ under Θ is*

$$\mathbf{P}_\Theta(\mathcal{T},\tau_m<t,\tau_{n_1}<u_1,\dots,\tau_{n'_1}>u'_1,\dots\mid\mathcal{T},\tau_{n_1}<u_1,\dots,\tau_{n'_1}>u'_1,\dots) = \frac{\mathbf{P}_\Theta(\mathcal{T},\mathcal{U}\cup\{(m,t)\},\mathcal{L})}{\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})}.$$

The computation of the divergence time distributions was performed on a contrived tree topology and on the Hominoidea subtree. Results are displayed in Figures 6 and 7 where the probability densities are computed from the corresponding distributions by finite difference approximations.

Figure 6 shows how considering models which are not time-homogeneous such as the piecewise-constant-birth-death models and adding temporal constraints on some of the divergence times influences the shapes of the divergence times distributions of all the nodes of the tree topology. In particular, divergence time distributions may become multimodal, thus hard to sample. Let us remark that a temporal constraint on the divergence time of a node influences the divergence time distributions of the other nodes of the tree topology, even if they are not among its ascendants or descendants.

In order to illustrate the computation of the divergence time distributions on a real topology, let us consider the Hominoidea subtree from the Primates tree of [7]. The approach can actually compute the divergence time distributions of the whole Primates tree of [7] but they cannot be displayed legibly because of its size.

The divergence time distributions were computed under several (simple) birth-death-sampling models, namely all parameter combinations with $\lambda=0.1$ or 1, $\mu=\lambda-0.09$ or $\lambda-0.01$ and $\rho=0.1$ or 0.9. Since the difference $\lambda-\mu$ appears in the probability formulas, several sets of parameters are chosen in such a way that they have the same difference between their birth and death rates.

Divergence time distributions obtained in this way are displayed in Figure 7 around their internal nodes (literally, since nodes are positioned at the median of their divergence times). Each distribution is plotted at its own scale in order to be optimally displayed. This representation allows to visualize the effects of each parameter on the shape and

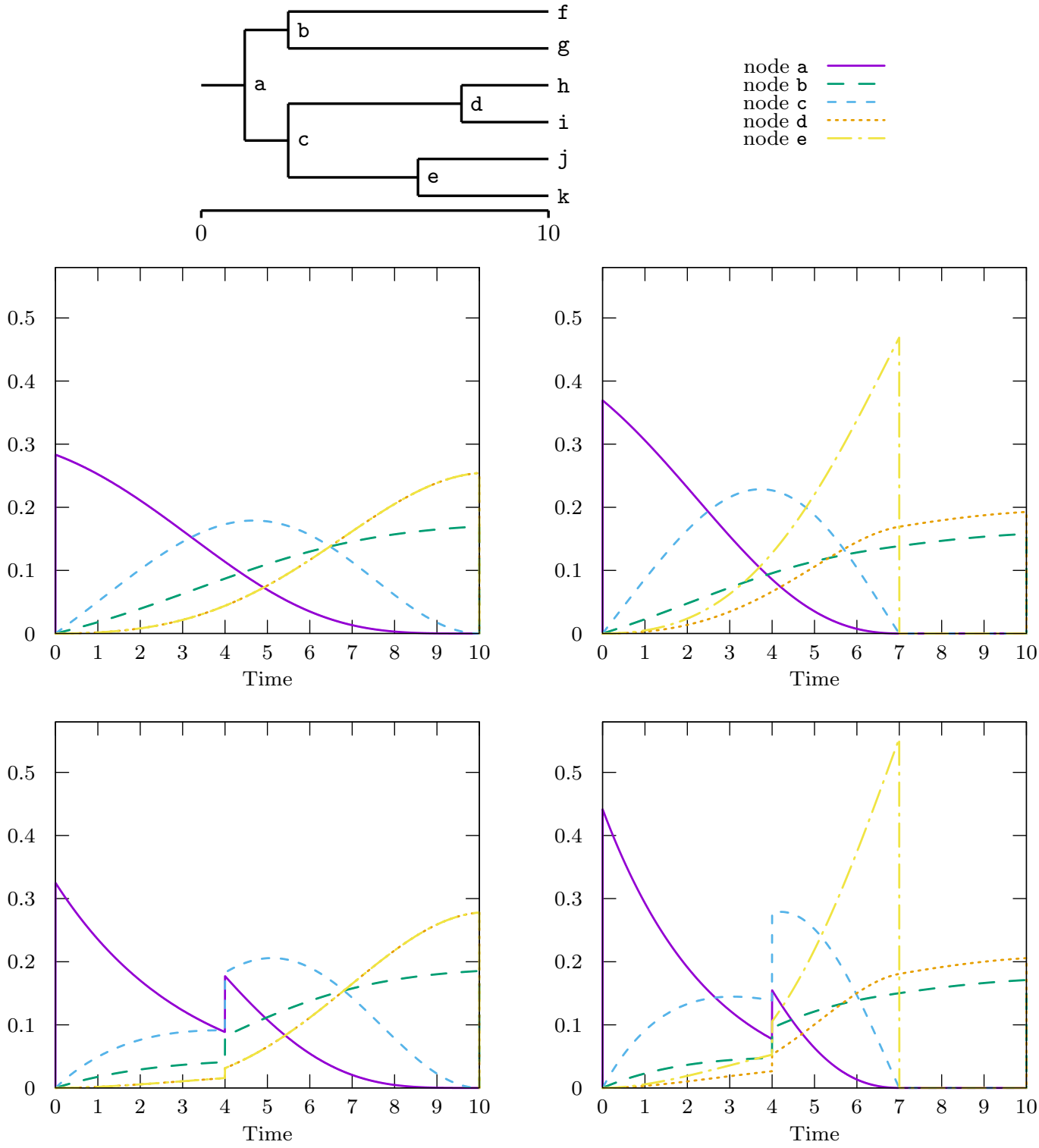

Figure 6: Divergence time probability densities of the tree displayed at the top, in the first row of plots by assuming a diversification process running from time 0 to 10 under a birth-death-sampling model with parameters $\lambda = 0.2$, $\mu = 0.02$ and $\rho = 0.5$ between times 0 and 10 and in the second row of plots by assuming a piecewise constant birth-death-sampling model with parameters $\lambda_0 = 0.1$, $\mu_0 = 0.02$ and $\rho_0 = 0.1$ between times 0 and 4 (only 10% of the lineages survives to time 4) and parameters $\lambda_1 = 0.2$, $\mu_1 = 0.02$ and $\rho_1 = 0.5$ between times 4 and 10. Plots of the first column are computed with no constraint on the divergence times and those of the second column by constraining the divergence time associated to node `e` to be anterior to 7. Densities of nodes `d` and `e` are confounded in the plots of the first column. Densities was obtained from the corresponding distributions by finite difference approximations.

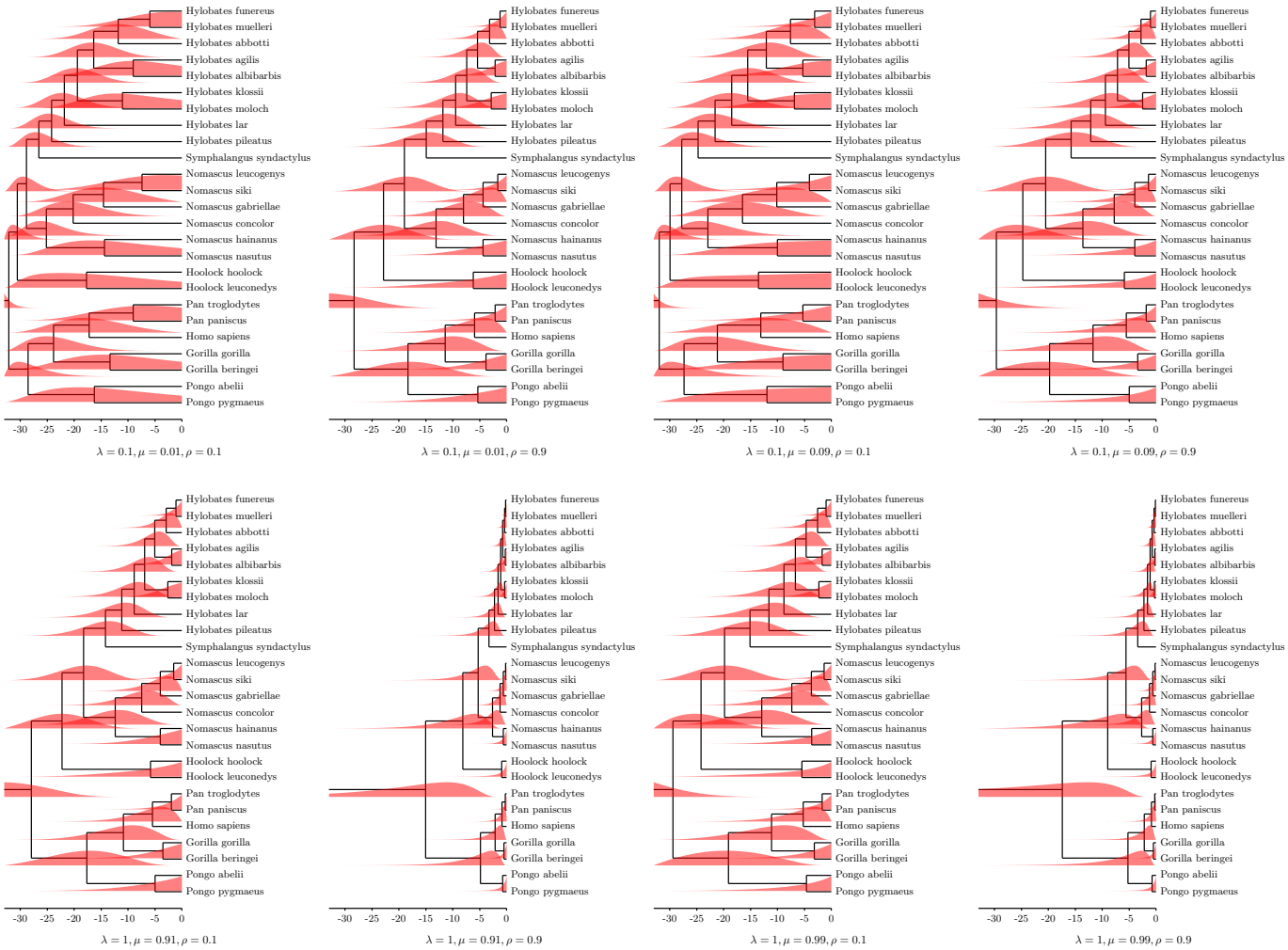

Figure 7: Divergence time probability densities of the Hominoidea tree from [7] under birth-death-sampling models with parameters $\lambda = 0.1$ or 1, $\mu = 0.01$ or 0.09 and $\rho = 0.1$ or 0.9. Internal nodes are positioned at their median divergence time. Densities was obtained from the corresponding distributions by finite difference approximations.

the position of distributions, to investigate which parameter values are consistent with a given evolutionary assumption etc.

Birth-death-sampling models are not identifiable, since several set of parameters leads to the same probability distributions. Namely, ρ and ρ' being two sampling probabilities, if one sets $\lambda' = \rho\lambda/\rho'$ and $\mu' = \mu - \lambda(1 - \rho/\rho')$, the probability densities of any phylogenetic tree (by considering or not considering its divergence times) is the same under the model $(s, e, \lambda, \mu, \rho)$ as under the model $(s, e, \lambda', \mu', \rho')$ [30].

We observe on Figure 7 that, all other parameters being fixed, the greater the speciation/birth rate λ (resp. the sampling probability ρ), the closer are the divergence time distributions to the end time.

Influence of the extinction/death rate on the divergence time distributions is more subtle and ambiguous, at least for this set of parameters. All other parameters being fixed, it seems that an increase of the extinction rate tends to push distributions of nodes close to the root towards the starting time and, conversely, those of nodes close to the tips towards the end time.

The divergence time distributions obtained for $\lambda = 0.1$, $\mu = 0.01$ and $\rho = 0.9$ (Fig. 7, column 2, top) and for $\lambda' = 1$, $\mu' = 0.91$ and $\rho' = 0.1$ (Fig. 7, column 1, bottom) are very close one to another. The same remark holds for $\lambda = 0.1$, $\mu = 0.09$ and $\rho = 0.9$ (Fig. 7, column 4, top) and for $\lambda' = 1$, $\mu' = 0.99$ and $\rho' = 0.1$ (Fig. 7, column 3, bottom). This certainly comes from identifiability issue of the birth-death sampling model since in both cases we have that $\lambda' \sim \rho\lambda/\rho'$ and $\mu' \sim \mu - \lambda(1 - \rho/\rho')$.

The variety of shapes of divergence times probability densities observed in Figures 6 and 7 exceeds that of standard prior distributions used in phylogenetic inference, e.g., uniform, lognormal, gamma, exponential [16, 14].

7.1 A previous approach

A previous approach for computing the probability density of a given divergence time is provided in [10]. It is based on the explicit computation of the probability density $f_{\mathcal{A}^k_{n,t}}$ of the kth divergence time of a tree topology with n tips

starting at t from the present, provided in [10], and the computation of the probability $\mathbb{P}(r(v) = k)$ for the rank $r(v)$ of the divergence time associated to the vertex v to be the kth which was given in [11]. The probability density f_v of the divergence time associated to a vertex v of a tree topology with n tips is then given for all times s by

$$f_v(s) = \sum_{k=1}^{n-1} \mathbb{P}(r(v) = k) f_{\mathcal{A}^k_{n,t}}(s).$$

The probability density $f_{\mathcal{A}^k_{n,t}}$ is computed in constant time and the probabilities $\mathbb{P}(r(v) = k)$ for all nodes v are computed in a time quadratic with the size of the tree.

The computation of the probability density of the kth divergence time of tree relies on the fact that, under some homogeneity assumption, the divergence times are independent and identically distributed random variables. Approach provided in [10] was described in the case of birth-death models. It can be easily adapted to deal with piecewise-constant-birth-death-sampling models but extending this approach in order to compute divergence times distribution with temporal constraints seems not straightforward.

8 Direct sampling of divergence times

Theorems 2 and 4 and Claim 3 show how to compute the marginal (with regard to the other divergence times) of the divergence time distribution of any internal node of a phylogenetic tree from a given birth-death-sampling model. It allows in particular to sample any divergence time of the phylogenetic tree disregarding the other divergence times. We shall see in this section how to draw a sample of all the divergence times of any tree topology from a given birth-death-sampling model.

Lemma 1. *Let $\mathcal{T}$ be a tree topology of root r, Θ be a Markovian diversification model from origin time s to end time e and t be a time in $(x_i, x_{i+1}]$. The probability that the root divergence time τ_r is anterior to a time $t \in [s, e]$ conditioned on observing the tree topology $\mathcal{T}$ under Θ is*

$$\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T}) = 1 - \frac{\mathbf{Q}_\Theta(s, t, 1)\mathbf{P}_{\Theta_{[t]}}(\mathcal{T})}{\mathbf{P}_\Theta(\mathcal{T})}.$$

Proof. The probability that the divergence time τ_r associated with r is anterior to a time $t \in [s, e]$ is the complementary probability that $\tau_r > t$. Observing $\tau_r > t$ means that the starting lineage at s has a single descendant observable at t from which descends the tree topology $\mathcal{T}$ sampled at e. It follows that

$$\begin{aligned}\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T}) &= 1 - \mathbf{P}_\Theta(\mathcal{T}, \tau_r > t \mid \mathcal{T}) \\ &= 1 - \frac{\mathbf{Q}_\Theta(s, t, 1)\mathbf{P}_{\Theta_{[t]}}(\mathcal{T})}{\mathbf{P}_\Theta(\mathcal{T})}.\end{aligned}$$

□

The probability $\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T})$ can be directly written as $\mathbf{P}_\Theta(\mathcal{T}, \{(r,t)\}, \emptyset)/\mathbf{P}_\Theta(\mathcal{T})$. Lemma 1 allows to avoid considering a temporal constraint, which is particularly interesting in the birth-death-sampling case.

Remark 1. *Under the birth-death-sampling model $\Theta = (s, e, \lambda, \mu, \rho)$, we have that*

$$\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T}) = 1 - \left[\frac{(1 - \mathrm{e}^{-(\lambda-\mu)(e-t)})(\rho\lambda + (\lambda(1-\rho) - \mu)\mathrm{e}^{-(\lambda-\mu)(e-s)})}{(1 - \mathrm{e}^{-(\lambda-\mu)(e-s)})(\rho\lambda + (\lambda(1-\rho) - \mu)\mathrm{e}^{-(\lambda-\mu)(e-t)})}\right]^{|\mathrm{L}_\mathcal{T}|-1},$$

which can be computed in constant time.

Let us first show how to sample the divergence time of the root of a tree topology. The marginal, with regard to the other divergence times, of the distribution of the root-divergence time conditioned on the tree topology $\mathcal{T}$ is the cumulative distribution function (CDF) $F_r : t \to \mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T})$. In order to sample τ_r under this distribution, we shall use *inverse transform sampling* which is based on the fact that if a random variable U is uniform over $[0, 1]$ then $F_r^{-1}(U)$ has distribution function F_r (e.g., [2, chapter 2]). Since finding an explicit formula for F_r^{-1} is not straightforward, we have to rely on numerical inversion at a given precision level in order to get a sample of the distribution F_r from an uniform sample on $[0, 1]$. The current implementation uses the *bisection method*, which computes an approximate inverse with a number of F_r-computations smaller than minus the logarithm of the required precision [2, p 32].

In order to sample the other divergence times, let us remark that by putting a and b for the two direct descendants of the root of $\mathcal{T}$ and t for the time sampled for the root-divergence, we have two independent diversification processes both starting at t and giving the two subtree topologies $\mathcal{T}_a$ and $\mathcal{T}_b$ at e. By applying Lemma 1 to $\mathcal{T}_a$ and $\mathcal{T}_b$ between t and e, the divergence times of the roots of these subtrees, i.e., a and b, can thus be sampled in the same way as above.

The very same steps can then be performed recursively in order to sample all the divergence times of $\mathcal{T}$. The time complexity of each sampling of a divergence time of $\mathcal{T}$ is obtained by multiplying the complexity of computing the probability of Lemma 1 with minus the logarithm of the precision required for the samples. From Remark 1, under the birth-death-sampling model $\Theta = (s, e, \lambda, \mu, \rho)$, the computation of $\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T})$ requires only the number of tips of $\mathcal{T}$ (in particular, the shape of $\mathcal{T}$ does not matter). In this case, the CDF F_r can be computed at any time t with complexity $O(1)$ and a pre-order traversal of $\mathcal{T}$ allows to sample all its divergence times in a time linear in $|\mathcal{T}|$ with a multiplicative factor proportional to minus the logarithm of the precision required for the samples.

The same approach can be applied in order to sample divergence times with temporal constraints and/or shifts.

9 Testing diversification shifts

Theorem 3 yields the computation of the probability density of a tree topology in which a given clade diversifies from a given "shift time" according a (simple) birth-death-sampling model different from that of the rest of the topology. This allows us to estimate the likelihood-ratio test for comparing the null model assuming a unique diversification model for the whole topology with the alternative model including a shift as displayed in Figure 5. Since the alternative model requires the implicit assumption that the lineage originating the "shifted" clade was alive at the shift time, we make the same assumption for the null model, i.e., the divergence time associated to the crown-node of the clade (resp. to the direct ancestor of the crown node) is assumed to be posterior (resp. anterior) to the shift-time. Basically, being given a tree topology, one of its clade and the shift time, we compute the ratio Λ_N of the maximum likelihoods of this topology with to without shift at the clade and shift time from Theorems 3 and 2 by using numerical optimization whenever a direct determination is not possible. Namely, in order to test a diversification shift at time t on the clade originating at node m of the tree topology $\mathcal{T}$, we consider the ratio

$$\Lambda_N = \frac{\mathbf{S}_{\Theta_1, \widetilde{\Theta}_1}(\mathcal{T}, m, t)}{\mathbf{P}_{\Theta_0}(\mathcal{T}, \tau_{\mathrm{f}(m)} \leq t, \tau_m \geq t)},$$

where $\mathrm{f}(m)$ is the direct ancestor of m, Θ_0, Θ_1, $\widetilde{\Theta}_1$ are diversification models with

$$\Theta_0 = \arg\max_{\Theta} \mathbf{P}_\Theta(\mathcal{T}, \tau_{\mathrm{f}(m)} \leq t, \tau_m \geq t) \text{ and } (\Theta_1, \widetilde{\Theta}_1) = \arg\max_{(\Theta, \widetilde{\Theta})} \mathbf{S}_{\Theta, \widetilde{\Theta}}(\mathcal{T}, m, t).$$

In order to assess the accuracy of Λ_N, we compare it to three sister-group diversity tests considered in [35]. Namely, for two sister groups originating at shift time t with $N_1 > N_2$ terminal taxa and total sums of branch lengths B_1 and B_2 respectively, we have that

- the probability of observing this or greater difference between sister group diversities from [29] is $P = \dfrac{2N_2}{N_1 + N_2 - 1}$,
- the likelihood ratio alternative provided in [27] is
  $\Lambda_A = 1.629 \times [h(N_1 - 1) - h(N_1) + h(N_2 - 1) - h(N_2) - h(2) - h(N_1 + N_2 - 2) + h(N_1 + N_2)]$,
  where $h(x) = \begin{cases} x \log(x) & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases}$
- the likelihood ratio from perfect-information given in [35] is $\Lambda_P = 2 \times \left(\dfrac{\hat{\lambda}_1^+}{\hat{\lambda}^+}\right)^{N_1 - 1} \left(\dfrac{\hat{\lambda}_2^+}{\hat{\lambda}^+}\right)^{N_2 - 1}$,
  where $\hat{\lambda}^+ = \dfrac{N_1 + N_2 - 2}{B_1 + B_2}$, $\hat{\lambda}_1^+ = \dfrac{N_1 - 1}{B_1}$ and $\hat{\lambda}_2^+ = \dfrac{N_2 - 1}{B_2}$.

I simulated topologies with and without shift according to pure-birth models, a.k.a. Yule models which are special cases of birth-death-sampling models with null death rate and full sampling, in the following way. Being given a general birth rate, a shift birth rate and the shift time, I first simulated topologies without shift from the general birth rate. Next, I filtered the simulated topologies by discarding those with less than 10 or more than 50000 nodes and those with a single lineage alive at the shift time. For each remaining simulation, I randomly picked a lineage alive at the shift time and replaced the clade originating from this lineage with a clade simulated with the shift rate from the shift to the end times in order to eventually obtain a topology with shift.

The quantities Λ_N, the likelihood ratio obtained from Theorem 3, P, Λ_A and Λ_P are then evaluated with regard to their ability to discriminate between tree topologies with or without shift. Figure 8 displays the Receiver Operating Characteristic (ROC) plots obtained for all these quantities. We first observe that Λ_N significantly outperforms measures P and Λ_A. In particular, in the case where the difference between the general and the shift birth rates is small (e.g., 0.6 and 1.0 in Fig. 8-left), performances of P and Λ_A are close to that of a random guess while Λ_N is still accurate. This was expected to at least some extent since Λ_N takes into account both the shift time and the whole tree topology while P and Λ_A are computed from the clade with the shift and its sister group. More surprisingly, Λ_N

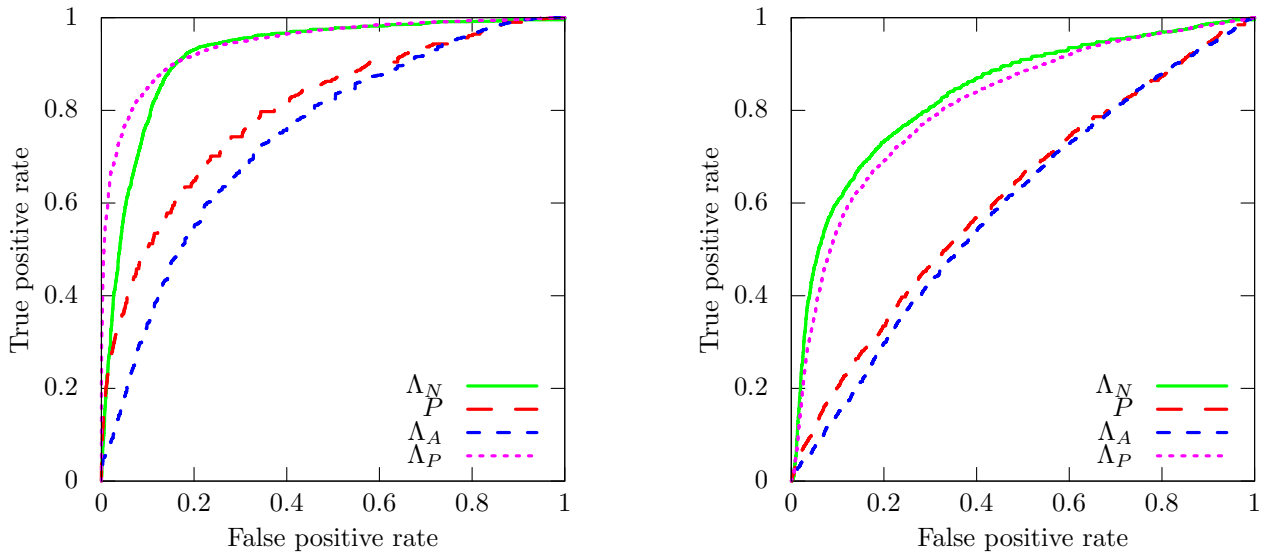

Figure 8: ROC plots of different measures for shift detection at left (resp. at right) are obtained by simulated 10000 Yule topologies with birth rate 0.4 (resp. 0.6) from times 0 to 10 and birth rate 1.0 from the shift time 5 to 10 for one of the clades present at time 5.

is only partially outperformed by Λ_P, which is obtained from all the divergence times and the shift time. In the case where the general birth rate is 0.4 and the the shifted one is 1, the ability to distinguishes between phylogenies with or without diversification shift is almost as good with our likelihood ratio as with that of the perfect information. In the case where the general birth rate is 0.6, the likelihood ratio test Λ_N obtained from Theorem 3 outperforms the other tests for all positive discovery rates lower than 40%.

In order to illustrate the diversification tests on a biological dataset, let us consider the calibrated phylogeny of Cetacea from [28], which displayed in Figure 9. In [28], authors detected a diversification rate increase in Delphinidae using MEDUSA, a detection method developed in [1]. The general idea of MEDUSA is to fit birth and death models with increasing numbers of diversification shifts by stopping when the improvement in the Akaike Information Criterion (AIC) is smaller than a fixed threshold. Note that the MEDUSA method requires all the divergence times in order to fit the models [1].

We computed the quantities Λ_N, P, Λ_A and Λ_P for all clades of the phylogeny of Cetacea, each time by setting the shift time to the time corresponding to the middle of the branch supporting the clade. Figure 9 displays the phylogenetic positions of the maxima observed for all these quantities. The maximal/most significant with regard to the likelihood ratio Λ_N was achieved at the position where the diversification rate increase was detected by MEDUSA [28, Fig. 1]. None of the other quantities P, Λ_A and Λ_P were maximal for this branch (Fig. 9).

References

[1] M. E. Alfaro, F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, 106(32):13410–13414, 2009.

[2] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag New York, 1986.

[3] G. Didier, M. Fau, and M. Laurin. Likelihood of Tree Topologies with Fossils and Diversification Rate Estimation. *Systematic Biology*, 66(6):964–987, 2017.

[4] G. Didier and M. Laurin. Exact distribution of divergence times from fossil ages and tree topologies. *bioRxiv*, 2018.

[5] P. C. J. Donoghue and Z. Yang. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1699), 2016.

[6] M. dos Reis. Notes on the birth-death prior with fossil calibrations for Bayesian estimation of species divergence times. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1699), 2016.

[7] M. dos Reis, G. F. Gunnell, J. Barba-Montoya, A. Wilkins, Z. Yang, and A. D. Yoder. Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Systematic Biology*, to appear, 2018.

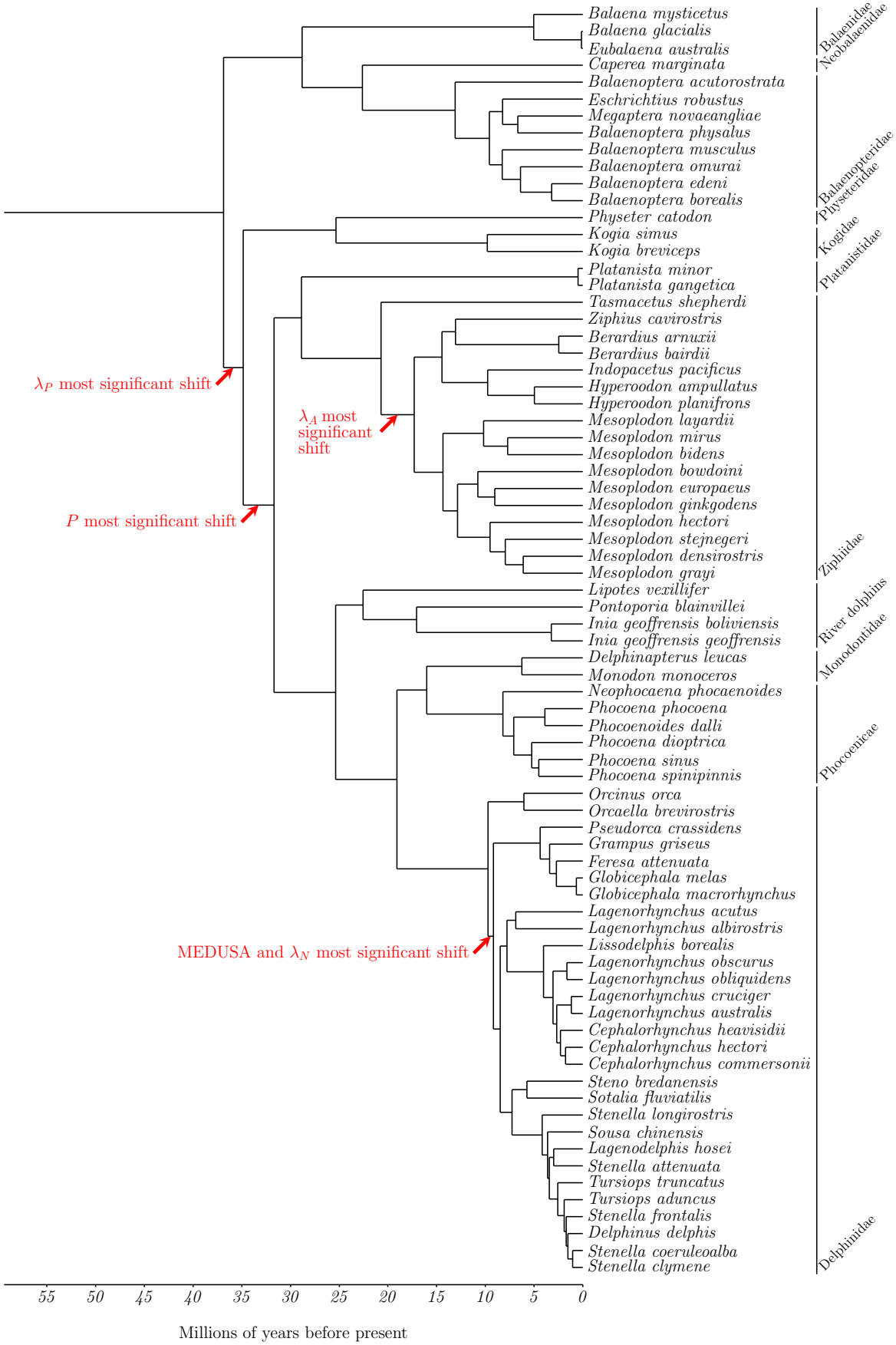

Figure 9: Time calibrated phylogeny of Cetacea [28].

[8] A. J. Drummond, M. A. Suchard, D. Xie, and A. Rambaut. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973, 2012.

[9] A. Gavryushkina, T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond. Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology*, 66(1):57–73, 2017.

[10] T. Gernhard. The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4):769–778, 2008.

[11] T. Gernhard, D. Ford, R. Vos, and M. Steel. Estimating the Relative Order of Speciation or Coalescence Events on a Given Phylogeny. *Evolutionary Bioinformatics*, 2:117693430600200012, 2006.

[12] A. Grafen. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 326(1233):119–157, 1989.

[13] E. F. Harding. The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability*, 3(1):44–77, 1971.

[14] T. A. Heath. A Hierarchical Bayesian Model for Calibrating Estimates of Species Divergence Times. *Systematic Biology*, 61(5):793–809, 2012.

[15] J. Heled and A. J. Drummond. Calibrated Birth-Death Phylogenetic Time-Tree Priors for Bayesian Inference. *Systematic Biology*, 64(3):369–383, 2015.

[16] S. Y. W. Ho and M. J. Phillips. Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology*, 58(3):367–380, 2009.

[17] D. G. Kendall. On the Generalized "Birth-and-Death" Process. *The Annals of Mathematical Statistics*, 19(1):1–15, 1948.

[18] H. Kishino, J. L. Thorne, and W. J. Bruno. Performance of a Divergence Time Estimation Method under a Probabilistic Model of Rate Evolution. *Molecular Biology and Evolution*, 18(3):352–361, 2001.

[19] C. Marshall. A Simple Method for Bracketing Absolute Divergence Times on Molecular Phylogenies Using Multiple Fossil Calibration Points. *The American Naturalist*, 171(6):726–742, 2008. PMID: 18462127.

[20] S. Nee, E. Holmes, R. May, and P. Harvey. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344(1307):77–82, 1994.

[21] J. E. O'Reilly, M. dos Reis, and P. C. Donoghue. Dating Tips for Divergence-Time Estimation. *Trends in Genetics*, 31(11):637–650, 2015.

[22] E. Paradis, J. Claude, and K. Strimmer. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2):289–290, 2004.

[23] B. Rannala and Z. Yang. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution*, 43(3):304–311, Sep 1996.

[24] B. Rannala and Z. Yang. Inferring Speciation Times under an Episodic Molecular Clock. *Systematic Biology*, 56(3):453–466, 2007.

[25] F. Ronquist, S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology*, 61(6):973–999, 2012.

[26] F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3):539–542, 2012.

[27] H. J. Sims and K. J. McConway. Nonstochastic variation of species-level diversification rates within angiosperms. *Evolution*, 57(3):460–479, 2003.

[28] G. J. Slater, S. A. Price, F. Santini, and M. E. Alfaro. Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*, 2010.

[29] J. B. Slowinski and C. Guyer. Testing the Stochasticity of Patterns of Organismal Diversity: An Improved Null Model. *The American Naturalist*, 134(6):907–921, 1989.

[30] T. Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58–66, 2009.

[31] T. Stadler. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences*, 108(15):6187–6192, 2011.

[32] T. Stadler and Z. Yang. Dating Phylogenies with Sequentially Sampled Tips. *Systematic Biology*, 62(5):674–688, 2013.

[33] E. A. Thompson. *Human Evolutionary Trees.* Cambridge University Press, Cambridge, MA, 1975.

[34] J. L. Thorne and H. Kishino. Estimation of divergence times from molecular sequence data. In *Statistical methods in molecular evolution*, pages 233–256. Springer, 2005.

[35] J. O. Wertheim and M. J. Sanderson. Estimating diversification rates: How useful are divergence times? *Evolution*, 65(2):309–320, 2010.

[36] Z. Yang. Empirical evaluation of a prior for Bayesian phylogenetic inference. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1512):4031–4039, 2008.

[37] Z. Yang and B. Rannala. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Molecular Biology and Evolution*, 14(7):717–724, 1997.

[38] Z. Yang and B. Rannala. Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. *Molecular Biology and Evolution*, 23(1):212–226, 2006.

A Table of the notations

s	origin time of the diversification process
e	end time of the diversification process/present time
λ	speciation rate
μ	extinction rate
ρ	sampling probability of extant taxa
Θ	(parameters of) a diversification model; in the case of a birth-death-sampling model, $\Theta = (s, e, \lambda, \mu, \rho)$
$\Theta_{[t]}$	diversification model Θ restricted to the time interval $[t, e]$
$\mathbf{P}_\Theta(.)$	probability of an event, a tree topology etc. under Θ
$\mathbf{O}_\Theta(t)$	probability for a lineage alive at t to be observable under Θ
$\mathbf{Q}_\Theta(t, t', N)$	probability that a single lineage alive at time t has exactly N descendants at time t' under Θ
$\mathbf{X}_\Theta(t, t', N)$	probability of the ending configuration of a standard pattern $(t, t', \mathcal{T})$ with $\lvert \mathrm{L}_\mathcal{T} \rvert = N$ under Θ
$\mathbf{Y}_\Theta(t, t', N)$	probability of the ending configuration of a special pattern $(t, t', \mathcal{T})$ with $\lvert \mathrm{L}_\mathcal{T} \rvert = N$ under Θ
$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T}, m, t)$	probability of the tree topology $\mathcal{T}$ by assuming a shift from model Θ to model $\widetilde{\Theta}$ at time t on the branch ending at node m
$\mathbf{W}_{m,k}$	quantity associated to a time o, a node m and a number of tips k used to compute the probability of a tree topology with temporal constraints
$\mathbf{T}_\Theta(\mathcal{T})$	probability of the tree topology $\mathcal{T}$ conditioned on its number of tips under Θ
$\mathbf{T}(\mathcal{T})$	probability of the tree topology $\mathcal{T}$ conditioned on its number of tips under any lineage-homogeneous model (i.e., in the Yule-Harding distribution)
$\lvert . \rvert$	cardinality of a set or size of a tree topology
$\mathcal{T}_n$	subtree of $\mathcal{T}$ rooted at node n
$\mathrm{L}_\mathcal{T}$	set of tips of $\mathcal{T}$
$\Gamma_{\mathcal{T},A}$	subtree topology of $\mathcal{T}$ made of all nodes in A and their direct descendants if $A \neq \emptyset$ and $\mathcal{T}$ if $A = \emptyset$
$\Omega_\mathcal{T}$	set of all start-sets of $\mathcal{T}$
$\Omega^\bullet_{\mathcal{T},n}$	set of start-sets of $\mathcal{T}$ containing node n
$\Omega^\circ_{\mathcal{T},n}$	set of start-sets of $\mathcal{T}$ not containing node n
$\Omega^\times_{\mathcal{T},n}$	set of start-sets A of $\mathcal{T}$ such that n is a tip of $\Gamma_{\mathcal{T},A}$
$\Upsilon^{(k)}_\mathcal{S}$	set of start-sets $A \in \mathcal{S}$ which are such that $\Gamma_{\mathcal{T},A}$ has exactly k tips
$\Upsilon_{\mathcal{S},m}$	set of start-sets of $\mathcal{S}$ restricted to $\mathcal{T}_m$ (i.e., $\bigcup_{A\in\mathcal{S}}\{A \cap \mathcal{T}_m\}$)
$\Upsilon^{(k)}_{\mathcal{S},m}$	set of start-sets $A \in \Upsilon_{\mathcal{S},m}$ which are such that $\Gamma_{\mathcal{T},A}$ has exactly k tips

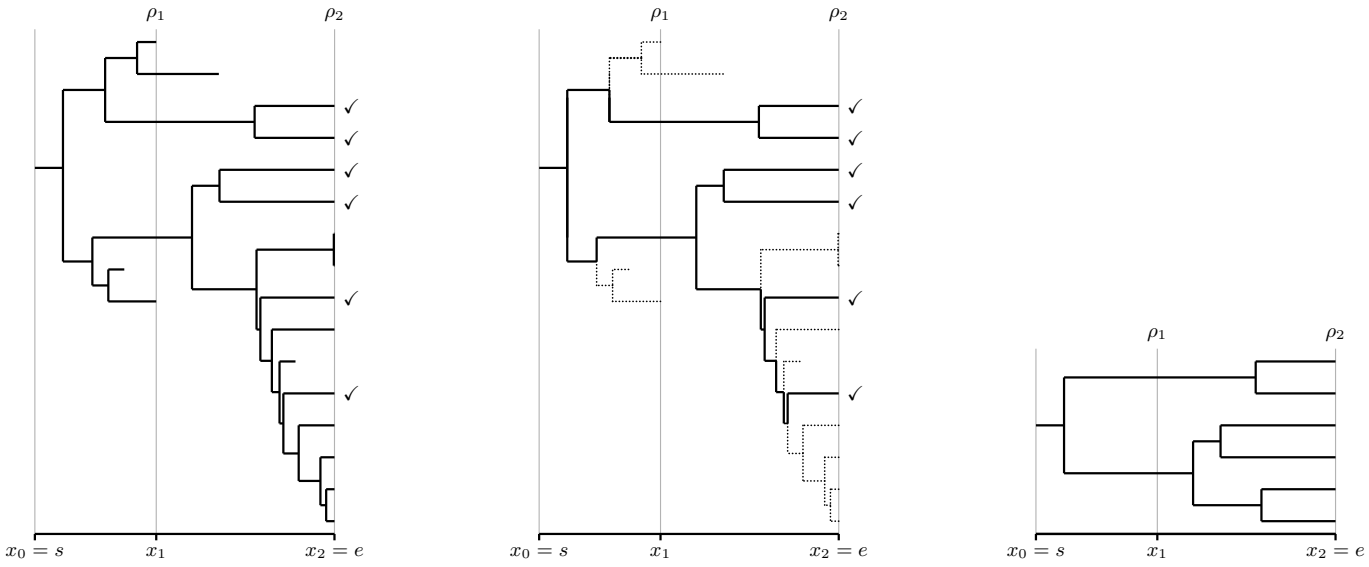

Figure 10: Left: the whole diversification process under the model $(s, e, \lambda, \mu, (x_1, \rho_1), (x_2, \rho_2))$ (sampled extant species are those with '✓'); Center: the part of the process that can be reconstructed is represented in plain – the dotted parts are lost; Right: the resulting phylogenetic tree.

B The generalized birth-death model with mass extinction events and extant sampling

The generalized birth-death process was introduced and studied in [17]. In this model, the speciation and extinction rates are allowed to change through time and are therefore given as two functions of the time, $\lambda : t \to \lambda(t)$ and $\mu : t \to \mu(t)$ (in this section, λ and μ denotes two functions of the time and are not real numbers like in Section 2.1). The probability $\ddot{p}_N(t, t')$ that a single lineage at time t has exactly N descendants at time t' by following the generalized birth-death (λ, μ) was given in [17]. We have that

$$\ddot{p}_0(t, t') = 1 - \frac{\mathrm{e}^{\delta_{t,t'}}}{\omega_{t,t'}}, \quad \text{and for all } N > 0,$$

$$\begin{aligned}\ddot{p}_N(t, t') &= (1 - \ddot{p}_0(t, t'))\, \frac{1}{\omega_{t,t'}} \left(1 - \frac{1}{\omega_{t,t'}}\right)^{N-1} \\ &= \frac{\mathrm{e}^{\delta_{t,t'}}\, (\omega_{t,t'} - 1)^{N-1}}{\omega_{t,t'}^{N+1}},\end{aligned}$$

where $\quad \delta_{t,t'} = \int_t^{t'} (\mu(z) - \lambda(z))dz \quad$ and $\quad \omega_{t,t'} = \mathrm{e}^{-\delta_{t,t'}} \left(1 + \int_t^{t'} \mathrm{e}^{-\delta_{t,z}} \mu(z)dz\right).$

Following the idea of [31], we shall consider a more general model by allowing to uniformly sample lineages at a given set of times $x_1, \ldots, x_k$ with respective probabilities $\rho_1, \ldots, \rho_k$. Namely, under the *sampled-generalized-birth-death* model $\Theta = (s, e, \lambda, \mu, (x_i, \rho_i)_{1 \le i \le k})$, lineages evolve following the generalized-birth-death model (λ, μ) between s and e, the origin and end times of the diversification process, and are uniformly sampled with probability ρ_i at each time x_i for $1 \le i \le k$ (Fig 10). In practice, sampling lineages at a time x_i anterior to the present time has to be interpreted as a mass extinction event (a lineage not sampled at x_i is assumed to have become extinct exactly at x_i) while sampling at the present time accounts for our incomplete knowledge of extant species (a species not sampled at the present time is assumed unknown). From now on, we assume without loss of generality that the last sampling time is the end/present time, i.e., $x_k = e$, and we set $x_0 = s$, the origin time of the diversification process. Like in Section 2.1, we are interested in the reconstructed process, i.e., the part of the process which is observable from the present/end time (Fig 10).

By construction, sampled-generalized-birth-death models are both Markovian and lineage-homogeneous. Extending the approaches which are presented in the main text in order to deal with the sampled-generalized-birth-death model $\Theta = (s, e, \lambda, \mu, (x_i, \rho_i)_{1 \le i \le k})$ only requires to compute the probabilities of the ending probabilities of standard and special patterns, which can be obtained from the probabilities $\mathbf{O}_\Theta(t)$ and $\mathbf{Q}_\Theta(t, t', N)$ for all positive numbers N and all times $s \le t \le t' \le e$. Let us see how to compute these last two probabilities.

In order to avoid ambiguity, we put t^+ (resp. t^-) for "an infinitesimal time after (resp. before) the time t". In particular, time x_i^+ (resp. x_i^-) is immediately after (resp. before) the ith sampling. By convention, we set

$\mathbf{Q}_\Theta(t,t',N) = \mathbf{Q}_\Theta(t,(t')^+,N)$, i.e., $\mathbf{Q}_\Theta(t,t',N)$ is the probability that a single lineage at time t has N descendants immediately after t'.

In order to compute $\mathbf{Q}_\Theta(t,e,0)$, the probability that a lineage alive at t has no sampled descendant at the end/present time $e = x_k$ under the sampled-generalized-birth-death model $\Theta = (s,e,\lambda,\mu,(x_i,\rho_i)_{1\le i\le k})$, let us remark that for all $1 \le i \le k$, a lineage alive at x_i^- has no sampled extant descendant with probability $\mathbf{Q}_\Theta(x_i^-,e,0) = (1-\rho_i) + \rho_i \mathbf{Q}_\Theta(x_i^+,e,0)$ if $i < k$ (since it is either not sampled at x_i or sampled with no sampled extant descendant) and with probability $\mathbf{Q}_\Theta(x_k^-,e,0) = 1-\rho_k$ otherwise [31]. It follows that for all $t \in [x_{i-1},x_i]$, we have that

$$\begin{aligned}\mathbf{Q}_\Theta(t,e,0) &= \ddot{p}_0(t,x_i) + \sum_{j=1}^{\infty} \ddot{p}_j(t,x_i)\mathbf{Q}_\Theta(x_i^-,e,0)^j \\ &= 1 - \frac{\mathrm{e}^{\delta_{t,x_i}}(1-\mathbf{Q}_\Theta(x_i^-,e,0)}{\omega_{t,x_i}(1-\mathbf{Q}_\Theta(x_i^-,e,0)) + \mathbf{Q}_\Theta(x_i^-,e,0)}.\end{aligned}$$

If $t = x_{i-1}$, the formula above gives $\mathbf{Q}_\Theta(x_{i-1}^+,e,0)$, thus allows to compute $\mathbf{Q}_\Theta(x_{i-1}^-,e,0)$ and iteratively any $\mathbf{Q}_\Theta(t,e,0)$ from $\mathbf{Q}_\Theta(x_k^-,e,0) = 1-\rho_k$. The probability $\mathbf{O}_\Theta(t)$ that a lineage alive at time t has at least a sampled extant descendant is basically $1-\mathbf{Q}_\Theta(t,e,0)$ and can therefore be computed under the sampled-generalized-birth-death model $\Theta = (s,e,\lambda,\mu,(x_i,\rho_i)_{1\le i\le k})$.

Let us compute $\mathbf{Q}_\Theta(t,e,1)$, the probability that a lineage alive at t has exactly one sampled descendant at the end/present time e. For all $1 \le i \le k$ and all $t \in [x_{i-1},x_i]$, the probability that a lineage alive at t has a single lineage sampled at e is the sum over all $j \ge 1$ of the probabilities that this lineage has j descendants at time x_i^- among which

- one is both sampled at x_i and with a single descendant sampled at e and,
- $j-1$ ones have no sampled descendants at e.

Namely, we have that

$$\begin{aligned}\mathbf{Q}_\Theta(t,e,1) &= \sum_{j=1}^{\infty} \binom{j}{1} \ddot{p}_j(t,x_i)\mathbf{Q}_\Theta(x_i^-,e,0)^{j-1}\rho_j\mathbf{Q}_\Theta(x_i^+,e,1) \\ &= \frac{\mathrm{e}^{\delta_{t,x_i}}\rho_j\mathbf{Q}_\Theta(x_i^+,e,1)}{\left(\omega_{t,x_i}(1-\mathbf{Q}_\Theta(x_i^-,e,0)) + \mathbf{Q}_\Theta(x_i^-,e,0)\right)^2},\end{aligned}$$

Since $\mathbf{Q}_\Theta(x_k^+,e,1) = 1$, the formula above allows to compute $\mathbf{Q}_\Theta(t,e,1)$ for all $t \in [s,e]$.

In order to show how to compute $\mathbf{Q}_\Theta(s,e,N)$, we shall first show that the probability density of observing the set of divergence times $(\tau_n)_{n\in\mathcal{T}\setminus \mathrm{L}_\mathcal{T}}$ of any tree topology $\mathcal{T}$ with N tips (i.e., with $|\mathcal{T}\setminus \mathrm{L}_\mathcal{T}| = N-1$ inner nodes and divergence times) under a sampled-generalized-birth-death model and disregarding the tree topology $\mathcal{T}$ is

$$\mathbf{P}_\Theta((\tau_n)_{n\in\mathcal{T}\setminus\mathrm{L}_\mathcal{T}}) = \mathbf{Q}_\Theta(s,e,1)\prod_{n\in\mathcal{T}\setminus\mathrm{L}_\mathcal{T}} \lambda(\tau_n)\mathbf{Q}_\Theta(\tau_n,e,1).$$

This result was actually already proved in [33, pp. 57-58]. But, since it was under the simple birth-death process and without including the first divergence time, let us sketch its proof.

The author of [33] first remarks that for all times $t < t'$, we have that

$$\mathbf{Q}_\Theta(t,e,1) = \sum_{j=1}^{\infty} \binom{j}{1} \ddot{p}_j(t,t')\mathbf{Q}_\Theta(t',e,0)^{j-1}\mathbf{Q}_\Theta(t',e,1).$$

The equation above implies that the probability of observing no speciation event which gives rise to a lineage sampled at e between times t and t' on a lineage of the reconstructed tree is $\mathbf{Q}_\Theta(t,e,1)/\mathbf{Q}_\Theta(t',e,1)$.

For all internal nodes n, the branch ending by n ends at time τ_n and starts at time $\widehat{\tau}_n$ where $\widehat{\tau}_n = \tau_{\mathrm{f}(n)}$ if n is not the root and $\mathrm{f}(n)$ is its direct ancestor, and where $\widehat{\tau}_n = s$ if n is the root. Considering the diversification process only on this branch, we observe a single lineage alive at $\widehat{\tau}_n$ which goes to time τ_n without (observable) speciation, which has probability $\mathbf{Q}_\Theta(\widehat{\tau}_n,e,1)/\mathbf{Q}_\Theta(\tau_n,e,1)$. If n is a tip (i.e., $n \in \mathrm{L}_\mathcal{T}$) then we set $\tau_n = e$ and we have $\mathbf{Q}_\Theta(\tau_n,e,1) = 1$. If n is an inner node (i.e., $n \in \mathcal{T}\setminus\mathrm{L}_\mathcal{T}$) then we observe a speciation event at τ_n, which occurs with probability $\lambda(\tau_n)$. From the Markov property, all the branchs evolves independently and we get that the probability density of observing the divergence times $(\tau_n)_{n\in\mathcal{T}\setminus\mathrm{L}_\mathcal{T}}$ is

$$\mathbf{P}_\Theta((\tau_n)_{n\in\mathcal{T}\setminus\mathrm{L}_\mathcal{T}}) = \left(\prod_{n\in\mathcal{T}\setminus\mathrm{L}_\mathcal{T}} \lambda(\tau_n)\frac{\mathbf{Q}_\Theta(\widehat{\tau}_n,e,1)}{\mathbf{Q}_\Theta(\tau_n,e,1)}\right) \times \left(\prod_{n\in\mathrm{L}_\mathcal{T}} \mathbf{Q}_\Theta(\widehat{\tau}_n,e,1)\right).$$

In the product above, each divergence time τ_n occurs twice in numerators as $\widehat{\tau}_{\mathrm{a}(n)}$ and $\widehat{\tau}_{\mathrm{b}(n)}$, where $\mathrm{a}(n)$ and $\mathrm{b}(n)$ are the direct descendants of n, and once in denominators. The time origin occurs only once in the numerator associated to the root. By simplifying the product above, we eventually get that

$$\mathbf{P}_\Theta((\tau_n)_{n\in\mathcal{T}\setminus \mathrm{L}_\mathcal{T}}) = \mathbf{Q}_\Theta(s,e,1)\prod_{n\in\mathcal{T}\setminus \mathrm{L}_\mathcal{T}}\lambda(\tau_n)\mathbf{Q}_\Theta(\tau_n,e,1),$$

which does not depend on the tree topology $\mathcal{T}$ (except on its size) and which is the probability density of observing the divergence times $(\tau_n)_{n\in\mathcal{T}\setminus \mathrm{L}_\mathcal{T}}$ on any tree topology with N tips.

The probability of observing N lineages at e by starting with a single lineage at s (in any tree topology) is then obtained by integrating the probability density of $N-1$ divergence times $(\tau_j)_{1\le j\le N-1}$ between s and e:

$$\begin{aligned}\mathbf{Q}_\Theta(s,e,N) &= \int_s^e \dots \int_s^e \mathbf{Q}_\Theta(s,e,1)\prod_{1\le j\le N-1}\lambda(\tau_j)\mathbf{Q}_\Theta(\tau_j,e,1)d\tau_1\dots d\tau_{N-1}\\ &= \mathbf{Q}_\Theta(s,e,1)\left(\int_s^e \lambda(\tau)\mathbf{Q}_\Theta(\tau,e,1)d\tau\right)^{N-1}.\end{aligned}$$

For all times t and t' with $s\le t\le t'\le e$, the probability $\mathbf{Q}_\Theta(t,t',N)$ can be computed in the same way by considering the restriction of the model Θ to the time interval $[t,t']$ (with full sampling at t' if $t'\neq x_i$ for all $1\le i\le k$).

The probabilities of ending configurations of standard and special patterns under the sampled-generalized-birth-death model $\Theta=(s,e,\lambda,\mu,(x_i,\rho_i)_{1\le i\le k})$ can then be computed from the probabilities $\mathbf{Q}_\Theta(t,t',N)$ and $\mathbf{O}_\Theta(t)$ (Eq. 1 and 2). By putting $\Theta_{[t,t']}$ for the model Θ restricted to the time interval $[t,t']$, we have that

$$\mathbf{X}_\Theta(t,t',N) = \frac{\mathbf{Q}_{\Theta_{[t,t']}}(t,t',1)\mathbf{O}_\Theta(t')^N\left(\int_t^{t'}\lambda(\tau)\mathbf{Q}_{\Theta_{[t,t']}}(\tau,t',1)d\tau\right)^{N-1}}{\left(1-(1-\mathbf{O}_\Theta(t'))\int_t^{t'}\lambda(\tau)\mathbf{Q}_{\Theta_{[t,t']}}(\tau,t',1)d\tau\right)^{N+1}},\quad \text{and,}$$

$$\mathbf{Y}_\Theta(t,t',N) = \frac{N\mathbf{Q}_{\Theta_{[t,t']}}(t,t',1)\left(\mathbf{O}_\Theta(t')\int_t^{t'}\lambda(\tau)\mathbf{Q}_{\Theta_{[t,t']}}(\tau,t',1)d\tau\right)^{N-1}}{\left(1-(1-\mathbf{O}_\Theta(t'))\int_t^{t'}\lambda(\tau)\mathbf{Q}_{\Theta_{[t,t']}}(\tau,t',1)d\tau\right)^{N+1}}.$$

C Proofs of Theorems

C.1 Proof of Theorem 2

Let us start with the case where the oldest time is the end time of the diversification process, i.e., the case where $o=e$. By construction, we then have that $\mathcal{U}$ and $\mathcal{L}$ are both empty. It follows that $(s,e,\mathcal{T})$ is a standard pattern of probability $\mathbf{T}_\Theta(\mathcal{T})\mathbf{X}_\Theta(s,e,|\mathrm{L}_\mathcal{T}|)$ from Claim 1.

Let us now assume that $o<e$. Under the notations of the theorem and by assuming that the divergence times of $\mathcal{T}$ are consistent with the temporal constraints, let us define $\mathcal{A}_o$ as the set of nodes of $\mathcal{T}$ whose divergence times are anterior to o (i.e. $\mathcal{A}_o=\{m\in\mathcal{T}\mid \tau_m<o\}$). Since divergence times corresponding to ancestors of a given node are always posterior to its own divergence time, all sets $\mathcal{A}_o$ are start-sets. By construction, the set $\mathcal{S}$ contains all the possible configurations of nodes of $\mathcal{T}$ with divergence times anterior to o which are consistent with the temporal constraints $\mathcal{U}$ and $\mathcal{L}$. Since all these configurations are mutually exclusive, by putting $\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L},\mathcal{A}_o=A)$ for the probability of observing the topology $\mathcal{T}$ with $\mathcal{A}_o=A$ and the temporal constraints $\mathcal{U}$ and $\mathcal{L}$, the law of total probabilities gives us that

$$\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \sum_{A\in\mathcal{S}}\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L},\mathcal{A}_o=A). \tag{3}$$

For instance, the entries of the second column of Figure 4 (just after the sum sign) represent all the start-sets of $\Omega^\bullet_{\mathcal{T},\mathsf{b}}$.

In order to compute the probability $\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L},\mathcal{A}_o=A)$ for a start-set $A\in\mathcal{S}$, we remark that

- the part of the diversification process anterior to o is the standard pattern $(s,o,\Gamma_{\mathcal{T},A})$ and that
- the part of the diversification process posterior to o consists of all the tree topologies $\mathcal{T}_n$ with temporal constraints $\mathcal{U}_{[\mathcal{T}_n]},\mathcal{L}_{[\mathcal{T}_n]}$ with $n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}$ under the model $\Theta_{[o]}$ (i.e., the model Θ restricted to the interval of times $[o,e]$), which have probability $\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})/\mathbf{O}_\Theta(o)$ conditioned on the observability of their starting lineages.

Since the diversification model Θ is Markovian, evolution of all the tree topologies $\mathcal{T}_n$ is independent of each other and with regard to the part of the process anterior to o, conditional upon starting with an observable lineage at time o.

From Claim 1, the probability of the standard pattern $(s, o, \Gamma_{\mathcal{T},A})$ is $\mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s, o, |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)$ under the assumption that $\Gamma_{\mathcal{T},A}$ is labeled. This part is a little tricky since we don't have a direct labeling of $\Gamma_{\mathcal{T},A}$ here (the tips of $\Gamma_{\mathcal{T},A}$ are identified through the labels of their tip descendants in $\mathcal{T}$, i.e., the tips of the subtrees pending from the tips of $\Gamma_{\mathcal{T},A}$). Since it assumes that $\Gamma_{\mathcal{T},A}$ is (exactly) labeled, we have to multiply the probability obtained from Claim 1 with the number of ways of connecting the tips/labels of $\Gamma_{\mathcal{T},A}$ to the subtrees starting from o, which is $|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!$, and with the probability of observing the groups of labels corresponding to the subtrees starting from o. Since all labelings of $\mathcal{T}$ are equiprobable, the probability of the groups of labels corresponding to the subtrees starting from o is the inverse of the number of ways of choosing a subset of $|\mathrm{L}_{\mathcal{T}_n}|$ labels from $|\mathrm{L}_{\mathcal{T}}|$ ones for all tips n of $\Gamma_{\mathcal{T},A}$ without replacement, i.e., the inverse of corresponding multinomial coefficient, which is

$$\frac{\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}} |\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_{\mathcal{T}}|!}.$$

Putting all together, we eventually get that

$$\begin{aligned}\mathbf{P}_{\Theta,A}(\mathcal{T},\mathcal{U},\mathcal{L}) &= |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\frac{\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}|\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_{\mathcal{T}}|!}\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})}{\mathbf{O}_\Theta(o)}\\ &= \frac{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)}{|\mathrm{L}_{\mathcal{T}}|!}\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)},\end{aligned}$$

which, with Equation 3, ends the proof. The whole computation of a toy example is schematized in Figure 4.

C.2 Proof of Theorem 3

Assuming that a diversification shift of the clade originating at m occurs at time t implies that the divergence times of $\mathcal{T}$ are such that both the direct ancestor of m has a divergence time strictly anterior to t and the divergence time of m is strictly posterior to t. Reciprocally, if the divergence times of m and of its direct ancestor are respectively posterior and anterior to t, then a diversification shift at time t may occur for the clade originating at m. The set of subsets of internal nodes with divergence time anterior to t consistent with the assumptions of the Theorem is thus exactly $\Omega^{\times}_{\mathcal{T},m}$.

We next follow the same outline as that of the proof of Theorem 2. For all subsets A of internal nodes of $\mathcal{T}$, let us put $\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t)$ for the probability of observing the topology $\mathcal{T}$ with a shift at time t for the clade originating at m and whose set of nodes with divergence time anterior to t is exactly A. We have that

$$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t) = \sum_{A\in\Omega^{\times}_{\mathcal{T},m}} \mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t). \tag{4}$$

From the Markov property, we have that $\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t)$ can be written as the product of the part of the diversification anterior to t, which is the special pattern $(s,t,\Gamma_{\mathcal{T},A})$ where the special lineage is the one on which the shift occurs, and the part of the diversification posterior to t which is a set of trees starting from time t and ending at time e by following model $\Theta_{[t]}$ except the special one which follows $\widetilde{\Theta}$. By construction, the non-special trees starting from t are conditioned on the observability of their starting lineage at t, thus have probability $\mathbf{P}_{\Theta_{[t]}}(\mathcal{T}_n)/\mathbf{O}_\Theta(t)$ while the special one is not conditioned and has probability $\mathbf{P}_{\widetilde{\Theta}}(\mathcal{T}_m)$.

From Claim 2, the probability of the special pattern $(s,t,\Gamma_{\mathcal{T},A})$ is $\mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{Y}_\Theta(s,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)$ under the assumption that $\Gamma_{\mathcal{T},A}$ is labeled. The situation slightly differs from the case of a standard pattern treated in the proof of Theorem 2 since the special tip of the special pattern is well identified and so is the subtree pending from it. In order to taking into account the fact that $\Gamma_{\mathcal{T},A}$ is not directly labeled, we have here to multiply the probability provided by Claim 2 with the number of ways of connecting the tips/labels of $\Gamma_{\mathcal{T},A}$ except m, the special one, to the subtrees starting from t, i.e., $(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!$, and with the probability of observing the groups of labels corresponding to the subtrees starting from t, which is

$$\frac{\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}} |\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_{\mathcal{T}}|!}.$$

Eventually, we get that

$$\begin{aligned}\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t) &= \mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{Y}_\Theta(s,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\frac{\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}|\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_{\mathcal{T}}|!}\mathbf{P}_{\widetilde{\Theta}}(\mathcal{T}_m)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{P}_{\Theta_{[t]}}(\mathcal{T}_n)}{\mathbf{O}_\Theta(o)}\\ &= \frac{(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\mathbf{T}_\Theta(\Gamma_{\mathcal{T},A})\mathbf{Y}_\Theta(s,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\mathbf{P}_{\widetilde{\Theta}}(\mathcal{T}_m)|\mathrm{L}_{\mathcal{T}_m}|!}{|\mathrm{L}_{\mathcal{T}}|!}\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{P}_{\Theta_{[t]}}(\mathcal{T}_n)|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)},\end{aligned}$$

which with Equation 4 ends the proof.

C.3 Proof of Theorem 4

If the model Θ is lineage homogeneous then, under the assumptions and notations of Theorem 2, we have that

$$\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \begin{cases} \dfrac{1}{|\mathrm{L}_\mathcal{T}|!}\displaystyle\sum_{A\in\mathcal{S}} |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}} \dfrac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)} & \text{if } o<e,\\ \mathbf{T}(\mathcal{T})\mathbf{X}_\Theta(s,e,|\mathrm{L}_\mathcal{T}|) & \text{otherwise.}\end{cases}$$

Since in the case where $o=e$, the computation of $\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ is performed in constant time under the assumptions of the theorem, we focus on the case where $o<e$. Let us first introduce an additional notation. For all sets $\mathcal{S}$ of start sets of a tree topology $\mathcal{T}$ and all numbers k between 1 and the number of tips of $\mathcal{T}$, we put $\Upsilon^{(k)}_\mathcal{S}$ for the set of start-sets $A\in\mathcal{S}$ such that the corresponding start-tree $\Gamma_{\mathcal{T},A}$ has exactly k tips. By construction, a start-tree of $\mathcal{T}$ has at least one tip and at most $|\mathrm{L}_\mathcal{T}|$ tips. We have:

$$\begin{aligned}\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) &= \frac{1}{|\mathrm{L}_\mathcal{T}|!}\sum_{A\in\mathcal{S}} |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}} \frac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)}\\ &= \frac{1}{|\mathrm{L}_\mathcal{T}|!}\sum_{k=1}^{|\mathrm{L}_\mathcal{T}|}\sum_{A\in\Upsilon^{(k)}_\mathcal{S}} |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}} \frac{\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)}\\ &= \frac{1}{|\mathrm{L}_\mathcal{T}|!}\sum_{k=1}^{|\mathrm{L}_\mathcal{T}|}\frac{\mathbf{X}_\Theta(s,o,k)}{\mathbf{O}_\Theta(o)^k}k!\sum_{A\in\Upsilon^{(k)}_\mathcal{S}}\mathbf{T}(\Gamma_{\mathcal{T},A})\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,.\end{aligned}$$

Let us set for all nodes m of $\mathcal{T}$,

$$\Upsilon_{\mathcal{S},m} = \bigcup_{A\in\mathcal{S}}\{A\cap\mathcal{T}_m\},$$

where $\mathcal{T}_m$ stands here for the set of nodes of the subtree topology rooted at m. In plain English, elements of $\Upsilon_{\mathcal{S},m}$ are elements of $\mathcal{S}$ restricted to $\mathcal{T}_m$. Since, by construction, the elements of $\Upsilon_{\mathcal{S},m}$ are start-sets of the tree topology $\mathcal{T}_m$, the start-tree $\Gamma_{\mathcal{T}_m,A}$ is well-defined for all $A\in\Upsilon_{\mathcal{S},m}$. For all numbers $1\le k\le|\mathrm{L}_{\mathcal{T}_m}|$, we put $\Upsilon^{(k)}_{\mathcal{S},m}$ for the set of start-sets $A\in\Upsilon_{\mathcal{S},m}$ such that the corresponding start-tree $\Gamma_{\mathcal{T}_m,A}$ has exactly k tips.

Let us now define for all nodes m of $\mathcal{T}$ and all $1\le k\le|\mathrm{L}_{\mathcal{T}_m}|$, the quantity

$$\mathbf{W}_{m,k} = k!\sum_{A\in\Upsilon^{(k)}_{\mathcal{S},m}}\mathbf{T}(\Gamma_{\mathcal{T},A})\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,.$$

Basically, by putting r for the root of $\mathcal{T}$, we have that

$$\mathbf{P}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \frac{1}{|\mathrm{L}_\mathcal{T}|!}\sum_{k=1}^{|\mathrm{L}_\mathcal{T}|}\frac{\mathbf{X}_\Theta(s,o,k)}{\mathbf{O}_\Theta(o)^k}\mathbf{W}_{r,k}. \tag{5}$$

We shall see how to compute $(\mathbf{W}_{m,k})_{k=1,\dots,|\mathrm{L}_{\mathcal{T}_m}|}$ for all nodes m of $\mathcal{T}$.

Let us first consider the case where $k=1$. We have that

$$\mathbf{W}_{m,1} = \mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_m,\mathcal{U}^{>o}_{[\mathcal{T}_m]},\mathcal{L}^{>o}_{[\mathcal{T}_m]})|\mathrm{L}_{\mathcal{T}_m}|!\,. \tag{6}$$

Let us now assume that $k>1$ and let a and b be the two direct descendants of m. Since we assume $k>1$, all start-sets of $\Upsilon^{(k)}_{\mathcal{S},m}$ contain m. It follows that we have $A\in\Upsilon^{(k)}_{\mathcal{S},m}$ if and only if there exist two start-sets $I\in\Upsilon_{\mathcal{S},a}$ and $J\in\Upsilon_{\mathcal{S},b}$ with $\{m\}\cup I\cup J=A$. The tree topology $\Gamma_{\mathcal{T}_m,A}$ has root m with two child-subtrees $\Gamma_{\mathcal{T}_a,I}$ and $\Gamma_{\mathcal{T}_b,J}$. In particular, we have $|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|+|\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}|=|\mathrm{L}_{\Gamma_{\mathcal{T}_m,A}}|=k$.

From Theorem 1, we have that

$$\mathbf{T}(\Gamma_{\mathcal{T}_m,A}) = \frac{2|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|!\,|\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}|!}{(|\mathrm{L}_{\Gamma_{\mathcal{T}_m,A}}|-1)|\mathrm{L}_{\Gamma_{\mathcal{T}_m,A}}|!}\mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J}) = \frac{2|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|!\,|\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}|!}{(k-1)k!}\mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J}).$$

Moreover, since by construction $\mathrm{L}_{\Gamma_{\mathcal{T}_m A}} = \mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}\cup\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}$, we get that

$$\begin{aligned}&\mathbf{T}(\Gamma_{\mathcal{T}_m A})\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_m A}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|! =\\ &\frac{2|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|!\,|\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}|!}{(k-1)k!}\mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J})\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big)\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}}\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big).\end{aligned}$$

More generally, the start-sets of $\Upsilon^{(k)}_{\mathcal{S},m}$ are in one-to-one correspondence with the set of pairs (I,J) of $\Upsilon_{\mathcal{S},a} \times \Upsilon_{\mathcal{S},b}$ such that $|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|+|\mathrm{L}_{\Gamma_{\mathcal{T}_b,I}}|= k$. This set of pairs is exactly the union over all pairs of positive numbers (i,j) such that $i+j=k$, of the product sets of $\Upsilon^{(i)}_{\mathcal{S},a} \times \Upsilon^{(j)}_{\mathcal{S},b}$. It follows that

$$\mathbf{W}_{m,k} = k! \sum_{\substack{i,j \\ i+j=k}} \sum_{\substack{(I,J)\in \\ \Upsilon^{(i)}_{\mathcal{S},a}\times\Upsilon^{(j)}_{\mathcal{S},b}}} \frac{2i!\,j!}{(k-1)k!} \mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J})(\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}} \mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!)(\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}} \mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_n,\mathcal{U}^{>o}_{[\mathcal{T}_n]},\mathcal{L}^{>o}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!).$$

After factorizing the left hand side of the equation just above, we eventually get that for all $k>1$,

$$\mathbf{W}_{m,k} = \frac{2}{k-1} \sum_{\substack{i,j \\ i+j=k}} \mathbf{W}_{a,i}\mathbf{W}_{b,j}. \tag{7}$$

The following remark is straightforward to prove by induction.

Remark 2. *Let $\mathcal{T}$ be a binary tree topology and for all internal nodes n of $\mathcal{T}$, let* $\mathrm{a}(n)$ *and* $\mathrm{b}(n)$ *denote the two direct descendants of n. We have that*

$$\sum_{n\in\mathcal{T}\setminus \mathrm{L}_{\mathcal{T}}} |\mathrm{L}_{\mathcal{T}_{\mathrm{a}(n)}}|\times|\mathrm{L}_{\mathcal{T}_{\mathrm{b}(n)}}|= \frac{|\mathrm{L}_{\mathcal{T}}|(|\mathrm{L}_{\mathcal{T}}|-1)}{2}.$$

From Equation 7 and for all internal nodes m of $\mathcal{T}$ with children a and b, computing the quantities $\mathbf{W}_{m,k}$ for all $1<k\leq|\mathrm{L}_{\mathcal{T}_m}|$ involves exactly $|\mathrm{L}_{\mathcal{T}_a}|\times|\mathrm{L}_{\mathcal{T}_b}|$ terms of the form $\mathbf{W}_{a,i}\mathbf{W}_{b,j}$. It follows that Remark 2 implies that if the quantities $\mathbf{W}_{m,1}$ are given for all nodes m of $\mathcal{T}$, the quantities $\mathbf{W}_{m,k}$ for all $m\in\mathcal{T}$ and all $1<k\leq|\mathrm{L}_{\mathcal{T}_m}|$ can be recursively computed in a time proportional to $|\mathrm{L}_{\mathcal{T}}|(|\mathrm{L}_{\mathcal{T}}|-1)/2$, thus with time complexity $O(|\mathcal{T}|^2)$.

In order to finish to prove Theorem 4, we shall proceed by induction on the total number temporal constraints by showing that the total time complexity required to compute the probabilities $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m is $O(|\mathcal{U}\cup\mathcal{L}|\times|\mathcal{T}|^2)$.

This property is basically true in the base case where $|\mathcal{U}\cup\mathcal{L}|= 0$ since from Theorem 2, the probability $\mathbf{P}_{\Theta}(\mathcal{T}_m,\emptyset,\emptyset)$ can be computed in constant time for all internal nodes m.

Let us assume that the induction assumption holds for all numbers of temporal constraints smaller than ℓ and let us consider two sets of temporal constraints $\mathcal{U}$ and $\mathcal{L}$ such that $|\mathcal{U}\cup\mathcal{L}|= \ell+1$. If o is the oldest time involved in $\mathcal{U}$ and $\mathcal{L}$, this implies that $|\mathcal{U}^{>o}\cup\mathcal{L}^{>o}|\leq\ell$. From the induction hypothesis, computing the probabilities $\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_m,\mathcal{U}^{>o}_{[\mathcal{T}_m]},\mathcal{L}^{>o}_{[\mathcal{T}_m]})$ for all internal nodes m can be performed in $O(\ell\times|\mathcal{T}|^2)$. From Equation 6, the quantities $\mathbf{W}_{m,1}$ for all internal nodes m are calculated directly from the probabilities $\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_m,\mathcal{U}^{>o}_{[\mathcal{T}_m]},\mathcal{L}^{>o}_{[\mathcal{T}_m]})$, thus in $O(|\mathcal{T}|)$. From Remark 2, all the quantities $\mathbf{W}_{m,k}$ for all internal nodes m of $\mathcal{T}$ and all $1<k\leq|\mathrm{L}_{\mathcal{T}_m}|$ can be calculated with time complexity $O(|\mathcal{T}|^2)$. Equation 5 can then be applied to all subtrees of $\mathcal{T}$ in order to compute the probabilities $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ from the quantities $\mathbf{W}_{m,k}$ for all internal nodes m of $\mathcal{T}$. Since computing each $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ requires to sum $|\mathrm{L}_{\mathcal{T}_m}|$ terms, computing all the $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ has total time complexity $O(|\mathcal{T}|^2)$.

To sum up, being given the probabilities $\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_m,\mathcal{U}^{>o}_{[\mathcal{T}_m]},\mathcal{L}^{>o}_{[\mathcal{T}_m]})$, which can be computed with complexity $O(\ell\times|\mathcal{T}|^2)$ from the induction hypothesis, computing the probabilities $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$ has time complexity $O(|\mathcal{T}|^2)$. The total time complexity required to compute all the $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ is well $O((\ell+1)\times|\mathcal{T}|^2)$, which proves the statement of the theorem about the time complexity.

Last, since at each stage of the induction, we have to store only the quantities $(\mathbf{W}_{m,k})_{m\in\mathcal{T},k=1,\ldots,|\mathrm{L}_{\mathcal{T}_m}|}$ and the probabilities $\mathbf{P}_{\Theta_{[o]}}(\mathcal{T}_m,\mathcal{U}^{>o}_{[\mathcal{T}_m]},\mathcal{L}^{>o}_{[\mathcal{T}_m]})$ and $\mathbf{P}_{\Theta}(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$, the total memory space complexity of the computation is $O(|\mathcal{T}|^2)$.