Review of *Probabilities of tree topologies with temporal constraints and diversification shifts*

Dominik Schrempf

March 18, 2019

1 Introduction

In the manuscript "Probabilities of tree topologies with temporal constraints and diversification shifts", the author Gilles Didier presents recursive formulae to calculate the probability of a time-like, labeled, rooted, bifurcating trees (simply called topologies hereafter) assuming the piecewise-constant, birth, death, and sampling diversification model. Additionally, for a set of temporal constraints on node ages, a recursive method is presented to calculate the joint probability of a given topology with nodes following temporal constraints. Furthermore, the probability of a model shift with different birth, death, and sampling rates on one arbitrary interior node of a given topology is derived. This probability can be used to test for model shifts using the maximum likelihood ratio test. The complexity of the calculations is quadratic with the number of leaves and linear with the number of time coefficients — allowing for fast computations.

2 Opinion

In my opinion, the manuscript is an impressive demonstration of the power of combinatorics and algebra, and presents several new findings to the phylogenetic community. Especially, the possibility to obtain model based priors for time constrained phylogeny nodes will be of high interest to phylogeneticists working on dating with fossils, or fast evolving species with time series data such as pathogens.

Nevertheless, I still feel that improvements to the readability and possibility to understand the results can be made, and I think that the results are a bit hidden behind a variety of mathematical symbols and definitions. This is also, why I was not able to completely follow the proofs of the theorems, although I understood the examples.

I would also appreciate a more detailed discussion of the results and implications. Especially, the interpretation of the presented applications could be elongated. With respect to this, one could describe in more detail, how your results can used on data sets involving sequence data. To me it is not completely clear, if one should use your model to replace other prior distributions, or to assess the likelihood of a specific sample during Bayesian (or maximum likleihood) analysis.

Minor comments can be found as digital annotations on the PDF; comments related to the English language are recommendations, but I am also not a native. Below I list other suggestions in the order they appear in the main text.

3 Suggestions

- Abstract, page 1, page 4: *size of phylogeny.* Could you please be more specific and move the definition of *size* from page 4 before the first use of this term?
- Abstract: *divergence time.* When reading the manuscript for the first time, the exact meaning of this term was not clear to me. It could refer to the divergence time between 2, or any number of leaves on a tree (possibly also the height of a tree). It could also refer to the divergence time between inner nodes on a tree. It may be good to be more specific. Would it be precise to just state that the divergence times are the branch lengths of the tree?
- Abstract and Page 2: *exact divergence time distributions.* Exact sounds very strong in this context. Do you mean *exact* when assuming the piecewise-constant, birth, death, and sampling model?
- Chapter 2, first paragraph: I am confused about the meaning of ρ_i. According to Tanja Stadlers paper [29], it is the *survival probability* when not at the present, and the *sampling probability* when at the present. Could you please explain the meaning of ρ_i in more detail here? Especially, the sentence: "The samplings of ancestral lineages .. are interpreted as extinction events .." confuses me. How can ancestral lineages be extinct?
- Chapter 3, first sentence: I do not understand this sentence. Trees obtained from the piecewise constant birth, death and sampling process are rooted and binary, but I can think of diversification processes that do not yield binary trees.
- Section 3.1.: Why does the reconstructed birth-death-sampling process not have extinction? Do I mis-understand something here?
- Chapter 4; Figure 2: Could you please explain what a special lineage is before referencing Figure 2?
- Section 5.1: To me it is not clear what the events $\tau_{n_i} < u_i$ are. As far as I can see, we restrict the node (event?) n_i, to be younger than u_i. Isn't it clearer to write: node ages τ_{n_i}, instead of events?

- Section 5.1: At this point, I also got confused about the nomenclature (which is very consistent but involves many different, new symbols). Let me summarize:
  - Bold symbols denote probabilities. Especially, the symbol $\mathbf{P}$ denotes "probability of".
  - The symbol $\mathbf{T}$ for instance, is just $\mathbf{P}(\mathcal{T}|n)$, where n are the number of tips of the topology $\mathcal{T}$.
  - Subscripts denote model parameters (mostly Θ). Why are the times u_i, and u_i' not part of Θ?
- Theorem 2: So the time constraints are not part of the model?
- Theorem 2: "$\Omega_{T,n}$ if $s1 = o$". What is $\Omega_{T,n}$? It does not form part of the definitions from before. I guess you mean Ω_T? What if $s_1 \neq o$?
- Theorem 2: "$s'_{k'+1} = s_k$". Why introduce a new symbol, when it is just the same as a symbol that was already introduced before?
- Theorem 2: "Θ'". Let the earliest time constraint be at time u. Is it correct, that if $u < s_1$, just another slice is introduced at u? Can you please describe in words what is being done here?
- Proof of Theorem 2: "being a divergence time assignation of $\mathcal{T}$": I do not understand this sentence.
- Proof of Theorem 3: "The set of subsets of internal nodes with divergence time anterior to t, and consistent with the assumptions of the Theorem is thus exactly $\Omega^{\times}_{\mathcal{T},m}$".

  $\Omega^{\times}_{\mathcal{T},m}$ is the set of all start-sets A of $\mathcal{T}$ such that m is a tip of $\Gamma_{\mathcal{T},A}$. I argue that $\Omega^{\times}_{\mathcal{T},m}$ is the "set of subsets of internal nodes with divergence time anterior to t, and consistent with the assumptions of the Theorem" together with start-sets including nodes in $\mathcal{T}_m$.
- Figure 7 and paragraph afterwards: The abbreviation receiver operating characteristic (ROC) was not defined.

Probabilities of tree topologies with temporal constraints and diversification shifts

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

January 29, 2019

Abstract

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

The main result of this work is a method to compute probabilities of undated phylogenetic trees under piecewise-constant-birth-death-sampling models by constraining some of the divergence times to belong to given time intervals and by allowing diversification shifts on certain clades. The computation is quite fast since its time complexity is quadratic with the size of the tree topology and linear with the number of time constraints and of "pieces" in the model.

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

- to compute the exact 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 is an essential and difficult stage of phylogenetic inference [22, 23, 17, 4, 20]. 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 [31, 24, 8, 13]. 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 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 [33, 15, 13]) [5, 14, 33]. Conversely, estimating parameters of diversification models requires temporal information about phylogenies. The birth-death-sampling model is arguably the simplest realistic diversification model since it includes three important features shaping phylogenetic trees [34, 35]. Namely, it models cladogenesis and extinction of species by a birth-death process and takes account of the incompleteness of data by assuming ~~an~~ uniform sampling of extant taxa. The birth-death-sampling model has been further studied and is currently used for phylogenetic inference [28, 30, 14, 5]. Since assuming constant diversification rates along time is sometimes unrealistic, the birth-death-sampling model has been extended in [29] in order to allow a finite number of shifts in diversification rates through time, ~~i.e.,~~ the diversification time is split into time intervals over which the diversification rates are constant (they may differ between intervals). This "piecewise-constant-birth-death-sampling model" also allows to model past extinction events. The main goal of this work is to devise methods to compute probabilities of undated phylogenies under certain assumptions about divergence times and about the diversification process under piecewise-constant-birth-death-sampling models from [29]. 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 piecewise-constant-birth-death-sampling 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 and of the shifts of the model, called patterns, and by combining their probabilities in order to get that of whole tree topology. The total time complexity of this computation is quadratic with the size of the phylogeny and linear with the total number of constraints and shifts of the model. 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 exact 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, as the method implemented in the function *compute.brlen* of the R-package *APE* [11, 21]. It also allows to visualize the effects of the birth-death-sampling 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 [18]. The computation of the divergence time distribution is illustrated with a contrived example in order to show the influence of the temporal constraints and of the model shifts 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 [9] 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 piecewise-constant-birth-death-sampling 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 [7, 25], 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 birth-death-sampling 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 [32].

Last, the approach presented here can be extended in order to take into account fossils. In [3], 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 a 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. Piecewise-constant-birth-death-sampling models are formally introduced in Section 2. 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.

2 Piecewise-constant-birth-death-sampling models

The dynamics of speciation and extinction of species is assumed following a *piecewise-constant-birth-death-sampling model* $((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$ where $s_0 < s_1 < \ldots < s_k$ are times, $\lambda_0, \ldots, \lambda_{k-1}$ and $\mu_0, \ldots, \mu_{k-1}$ are speciation and extinction rates and $\rho_0, \ldots, \rho_{k-1}$ are sampling probabilities. This model was introduced in [29] for modeling a diversification process starting with a single lineage at s_0 and ending at s_k, which is usually the present time. Under model $((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$, the diversification time $[s_0, s_k]$ is sliced in periods $[s_i, s_{i+1})$ during which the speciation and extinction rates are constant and equal to λ_i and μ_i respectively. At each time s_i with $i = 1, \ldots, k$ the lineages alive are uniformly sampled with probability ρ_i. The samplings of ancestral lineages at times s_i with $i < k$ (i.e., anterior to the the ending time s_k) are interpreted as extinction events while the last one, if s_k is the present time, accounts for our incomplete knowledge of extant species.

A important point is to distinguish between the part of the process that actually happened, which will be referred to as the *whole* or 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 *observed* or the *reconstructed process* (Fig. 1-Right). More formally, for all times $t \in [s_0, s_k]$, a lineage alive at time t is *observable* if itself or one of its descendants is both alive and sampled at the ending time s_k. The probability for a lineage alive at a time $t \in [s_i, s_{i+1}]$ to have no sampled descendant at the ending time s_k under the piecewise-constant-birth-death-sampling model $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$ was provided in [29]. It is

$$p_0^i(t) = \frac{\mu_i(c_i - 1)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-s_{i+1})} + (\mu_i - c_i\lambda_i)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-t)}}{\lambda_i(c_i - 1)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-s_{i+1})} + (\mu_i - c_i\lambda_i)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-t)}}$$

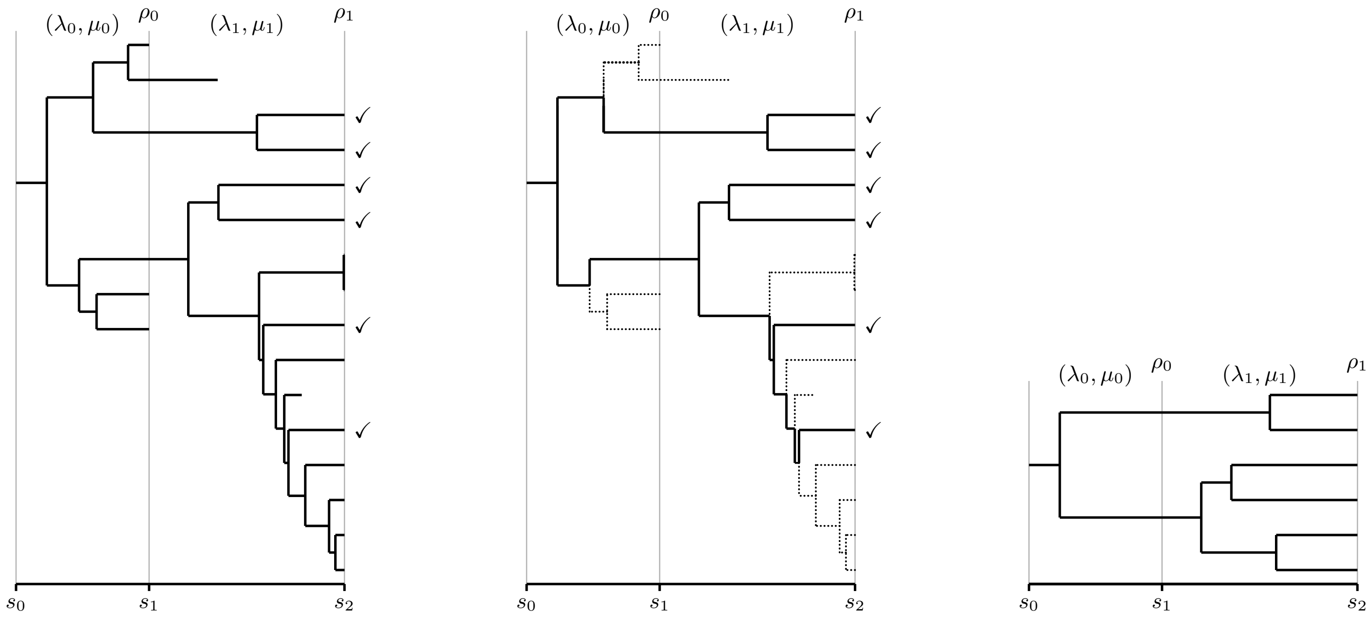


Figure 1: Left: the whole diversification process (sampled extant species are those with '✓'); Center: the part of the process that can be reconstructed is represented in plain – the dotted parts are lost; Right: the resulting phylogenetic tree.

where $c_i = (1-\rho_i) + \rho_i p_0^{i+1}(s_{i+1})$ if $i < k-1$ and $c_{k-1} = 1 - \rho_{k-1}$. The formula was slightly adapted since we consider here the "from past to present" direction of time, opposite to that of [29].

Basically, the probability $\mathbf{O}_\Theta(t)$ for a lineage living at time $t \in [s_i, s_{i+1}]$ in the complete diversification process (as in Figure 1-Left) to be observable at the ending time s_k is the complementary probability of having no descendant sampled at time s_k. We have that

$$\begin{aligned}\mathbf{O}_\Theta(t) &= 1 - p_0^i(t)\\ &= \frac{(\lambda_i-\mu_i)(c_i-1)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-s_{i+1})}}{\lambda_i(c_i-1)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-s_{i+1})} + (\mu_i - c_i\lambda_i)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-t)}}.\end{aligned}$$

The probability that a lineage alive at time $t \in [s_i, s_{i+1})$ has exactly one sampled descendant at time s_k was provided in [29]. The probability $\mathbf{I}_\Theta(t, t')$ that a lineage alive at time $t \in [s_i, s_{i+1})$ has a single descendant at a posterior time $t' \in (s_j, s_{j+1}]$ can be derived in the very same way to get that

$$\mathbf{I}_\Theta(t,t') = \frac{\widehat{g}_i(t)\prod_{\ell=i}^{j-1}\rho_\ell(\lambda_\ell-\mu_\ell)^2\widehat{g}_{\ell+1}(s_{\ell+1})}{\widehat{g}_j(t')}\delta,$$

where $\widehat{g}_i(t) = \dfrac{\mathrm{e}^{-(\lambda_i-\mu_i)(2s_k-(s_{i+1}+t))}}{\left(\lambda_i(c_i-1)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-s_{i+1})} + (\mu_i-c_i\lambda_i)\mathrm{e}^{-(\lambda_i-\mu_i)(s_k-t)}\right)^2}$ and $\delta = \begin{cases}\rho_j & \text{if } t' = s_{j+1},\\ 1 & \text{otherwise.}\end{cases}$

Birth-death-sampling models studied in [34], are basically piecewise-birth-death-sampling models with a single "time-slice", i.e., of the form $((s_0, \lambda_0, \mu_0, \rho_0), s_1)$. In the case where the ending lineages are all sampled, one talks about *birth-death models*. Under the simple birth-death model with birth rate λ and death rate μ, the probability $\mathbf{Q}_{(\lambda,\mu)}(n,t)$ that a single lineage at time 0 has exactly n descendants at time t is [16, 19],

$$\mathbf{Q}_{(\lambda,\mu)}(0,t) = \frac{\mu\left(1-\mathrm{e}^{-(\lambda-\mu)t}\right)}{\lambda-\mu\mathrm{e}^{-(\lambda-\mu)t}},$$

and, for all $n > 0$,

$$\mathbf{Q}_{(\lambda,\mu)}(n,t) = (\lambda-\mu)^2\mathrm{e}^{-(\lambda-\mu)t}\frac{\left(\lambda(1-\mathrm{e}^{-(\lambda-\mu)t})\right)^{n-1}}{\left(\lambda-\mu\mathrm{e}^{-(\lambda-\mu)t}\right)^{n+1}}.$$

3 Tree topologies

Tree topologies arising from diversification processes are rooted and binary 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 $\mathcal{T}$ and, for all nodes n of $\mathcal{T}$, $\mathcal{T}_n$ for 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 $\mathbf{T}(\mathcal{T})$ as the probability of a tree topology $\mathcal{T}$ given its number of tips under a lineage-homogeneous process with no extinction, such as the reconstructed birth-death-sampling process.

Theorem 1 ([12]). *Given its number of tips, a tree topology $\mathcal{T}$ resulting from a pure-birth realization of a lineage-homogeneous 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 [12] are slightly different from those of Theorem 1 but its arguments still hold~~s~~. The probability provided in [2, Supp. Mat., Appendix 2] is actually the same as that just above though it was derived in a different way from [12] and expressed in a slightly different form (see [3, 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 $\mathcal{T}$.

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 $\Gamma_{\mathcal{T},A}$.

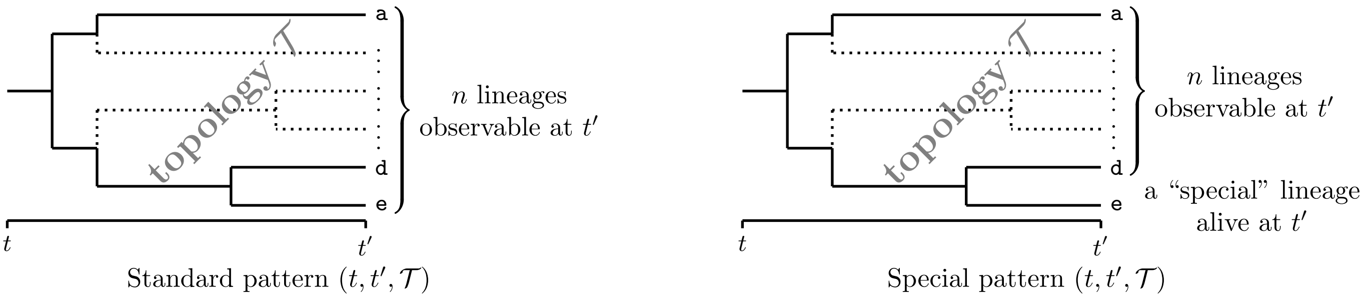
4 Patterns

In this section, we shall consider diversification processes starting at origin time s_0 and ending at time s_k ~~by evolving~~ following a piecewise-constant-birth-death-sampling model $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$. 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. It consists of a 3-tuple $(t, t', \mathcal{T})$ where t and t' are the start and ending times of the pattern and $\mathcal{T}$ is the resulting tree topology. We shall consider two types of patterns: standard and special patterns (Fig. 2). Standard and special patterns are very similar to patterns defined in [2] for the fossilized-birth-death process. Proofs of Lemmas 1 and 2 are essentially the same as those of the corresponding claims in [2].

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. 2-left).*

Let us compute the probability $\mathbf{X}_\Theta(t, t', n)$ that a single lineage at time $t \in [s_0, s_1)$ has n descendants observable from s_k at time $t' \in (t, s_1]$ under the piecewise-constant-birth-death-sampling model $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$. 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 (i.e., without sampling, which is equal to $\mathbf{Q}_{(\lambda_0,\mu_0)}(j+n, t'-t)$), among which exactly n ones are observable (i.e., $\binom{j+n}{n}\mathbf{O}_\Theta(t')^n\,(1-\mathbf{O}_\Theta(t'))^j$). We thus have


a
topology $\mathcal{T}$
d
e
n lineages observable at t'
t
t'
a
topology $\mathcal{T}$
d
e
n lineages observable at t'
a "special" lineage alive at t'
t
t'

Standard pattern $(t, t', \mathcal{T})$

Special pattern $(t, t', \mathcal{T})$

Figure 2: The two types of patterns used to compute probability distributions.

$$\begin{aligned}\mathbf{X}_\Theta(t,t',n) &= \sum_{j=0}^{\infty} \mathbf{Q}_{(\lambda_0,\mu_0)}(j+n, t'-t)\binom{j+n}{n}\mathbf{O}_\Theta(t')^n\left(1-\mathbf{O}_\Theta(t')\right)^j \\ &= \frac{(\lambda_0-\mu_0)^2 \mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)}\left(\lambda_0(1-\mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)})\right)^{n-1}\mathbf{O}_\Theta(t')^n}{\left(\lambda_0\mathbf{O}_\Theta(t') + (\lambda_0(1-\mathbf{O}_\Theta(t'))-\mu_0)\mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)}\right)^{n+1}}\end{aligned}$$

Lemma 1. *Under the piecewise-constant-birth-death-sampling model $\Theta = [(s_i, \lambda_i, \mu_i, \rho_i)]_{0\leq i<k}$, the probability of the standard pattern $(t, t', \mathcal{T})$ with $s_0 \leq t < t' \leq s_1$ is*

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

Proof. 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 $\mathbf{T}(\mathcal{T})$ from Theorem 1 and since a piecewise-constant-birth-death-sampling model is lineage homogeneous, 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, which is $\mathbf{X}_\Theta(t, t', |\mathrm{L}_\mathcal{T}|)$. □

4.2 Special patterns

Definition 2. *A special pattern $(t, t', \mathcal{T})$ starts with a single lineage at time $t \in (s_0, s_1]$ and ends with the tree topology $\mathcal{T}$ at t', 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. 2-right).*

Let us now compute the probability $\mathbf{Y}_\Theta(t, t', n+1)$ that a single lineage at time $t \in [s_0, s_1)$ has one special descendant and n descendants observable from s_k at time $t' \in (t, s_1]$. 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 (i.e., without sampling, which is equal to $\mathbf{Q}_{(\lambda_0,\mu_0)}(j+n+1, t'-t)$), 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. We have that

$$\begin{aligned}\mathbf{Y}_\Theta(t,t',n+1) &= \sum_{j=0}^{\infty} \mathbf{Q}_{(\lambda_0,\mu_0)}(j+n+1, t'-t)(n+1)\binom{j+n+1}{n+1}\mathbf{O}_\Theta(t')^n\left(1-\mathbf{O}_\Theta(t')\right)^j \\ &= \frac{(n+1)(\lambda_0-\mu_0)^2 \mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)}\left(\lambda_0(1-\mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)})\mathbf{O}_\Theta(t')\right)^{n}}{\left(\lambda_0\mathbf{O}_\Theta(t') + (\lambda_0(1-\mathbf{O}_\Theta(t'))-\mu_0)\mathrm{e}^{-(\lambda_0-\mu_0)(t'-t)}\right)^{n+2}}\end{aligned}$$

Lemma 2. *Under the the piecewise-constant-birth-death-sampling model $\Theta = [(s_i, \lambda_i, \mu_i, \rho_i)]_{0\leq i<k}$, the probability of the special pattern $(t, t', \mathcal{T})$ with $s_0 \leq t < t' \leq s_1$ is*

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

Proof. 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 $\mathbf{T}(\mathcal{T})$ from Theorem 1 and since a piecewise-constant-birth-death-sampling model is lineage homogeneous, multiplied by the probability of observing this ending configuration in a special pattern, which is $\mathbf{Y}_\Theta(t, t', |\mathrm{L}_\mathcal{T}|)$. □

5 Probability densities of topologies with temporal constraints and shifts

5.1 Temporal constraints

We shall see how to compute the probability density of a tree topology $\mathcal{T}$ with time~~s~~ constraints under a piecewise-constant-birth-death-sampling model $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0<i<k}, s_k)$. Namely, given internal nodes $n_1, \ldots, n_\ell$, $n'_1, \ldots, n'_{\ell'}$

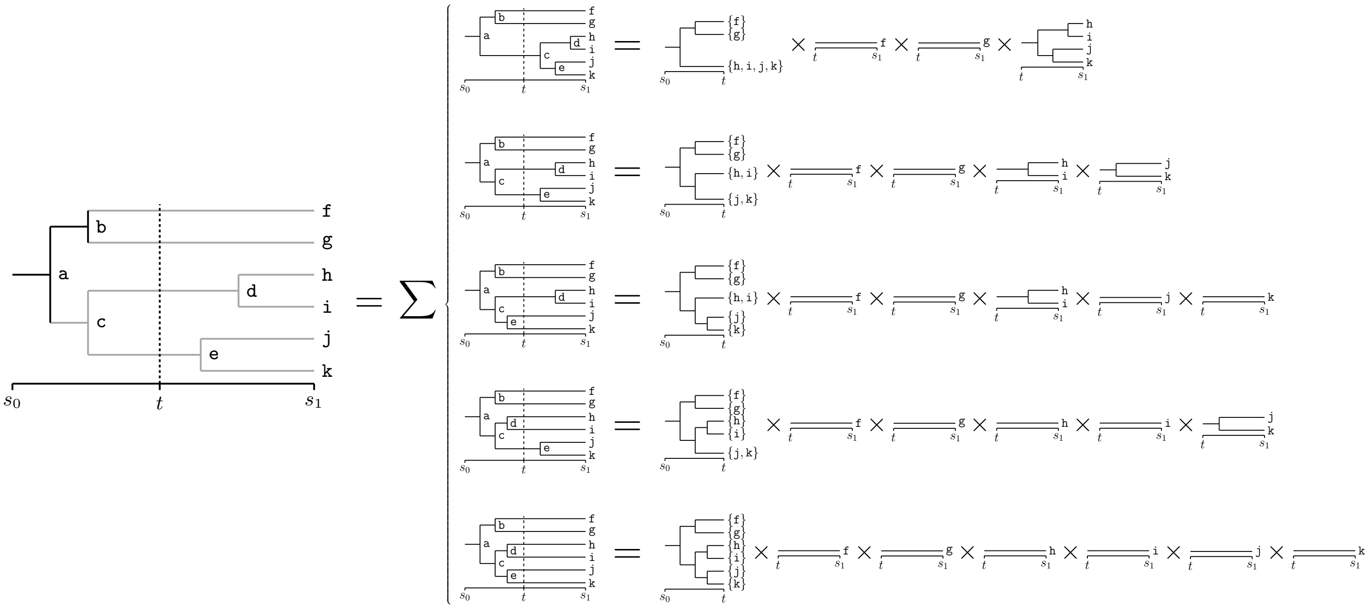


Figure 3: Schematic of the computation of the probability $\mathbf{D}_{((s_0,\lambda_0,\mu_0,\rho_0),s_1)}(\mathcal{T},\{(\mathtt{b},t)\},\emptyset)$, i.e., that the divergence time associated with node b is strictly anterior to t. Under the notation of Theorem 2, we have that $\mathcal{S} = \Omega^{\bullet}_{\mathcal{T},\mathtt{b}}$. 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 whether there are anterior or posterior to t.

of $\mathcal{T}$ and times $u_1, \dots, u_\ell, u'_1, \dots, u'_{\ell'}$ between s_0 and s_k (both not included), we aim to compute the joint probability density of $\mathcal{T}$ and events $\tau_{n_1} < u_1, \dots, \tau_{n_\ell} < u_\ell, \tau_{n'_1} > u'_1, \dots, \tau_{n'_{\ell'}} > u'_{\ell'}$ under the model Θ, i.e.,

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

The events $\tau_{n_1} \leq u_1, \dots, \tau_{n_\ell} \leq u_\ell$ will be referred to as *upper temporal constraints* and resumed as the set of pairs "node-time" $\mathcal{U} = \{(n_1,u_1),\dots,(n_\ell,u_\ell)\}$, and the events $\tau_{n'_1} \geq u'_1, \dots, \tau_{n'_{\ell'}} \geq u'_{\ell'}$, will be referred to as *lower temporal constraints* and resumed as the set of pairs $\mathcal{L} = \{(n'_1,u'_1),\dots,(n'_{\ell'},u'_{\ell'})\}$. We assume that the temporal constraints are consistent ~~one~~ 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 time 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. $\mathcal{L}^{(t)} = \{(n'_j,u'_j) \mid (n'_j,u'_j) \in \mathcal{L} \text{ and } u'_j = t\}$).

Theorem 2. *Let $\mathcal{T}$ be a tree topology, $\Theta = ((s_i,\lambda_i,\mu_i,\rho_i)_{0\leq i<k}, s_k)$ a piecewise-constant-birth-death-sampling model from origin time s_0 to end time s_k and $\mathcal{U} = \{(n_1,u_1),\dots,(n_\ell,u_\ell)\}$ and $\mathcal{L} = \{(n'_1,u'_1),\dots,(n'_{\ell'},u'_{\ell'})\}$ be two sets of upper and lower temporal constraints respectively. Let ~~us put~~ o for the oldest time involved in the model or in a time constraint~~s~~, s_0 excluded, namely,*

$$o = \min\{s_1, \min\{t \mid \exists n \in \mathcal{T} \text{ such that } (n,t) \in \mathcal{U}\}, \min\{t \mid \exists n \in \mathcal{T} \text{ such that } (n,t) \in \mathcal{L}\}\}.$$

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

- $\Omega_{\mathcal{T},n}$ *if* $s_1 = o$,
- $\bigcap_{(n,o)\in\mathcal{U}^{(o)}} \Omega^{\bullet}_{\mathcal{T},n}$ *if* $\mathcal{U}^{(o)} \neq \emptyset$,
- $\bigcap_{(n,o)\in\mathcal{L}^{(o)}} \Omega^{\circ}_{\mathcal{T},n}$ *if* $\mathcal{L}^{(o)} \neq \emptyset$,

and let us set $\Theta' = ((s'_i,\lambda'_i,\mu'_i,\rho'_i)_{0\leq i<k'}, s'_{k'+1})$ where $s'_{k'+1} = s_k$ and

- $k' = k-1$ *and* $(s'_i,\lambda'_i,\mu'_i,\rho'_i) = (s_{i+1},\lambda_{i+1},\mu_{i+1},\rho_{i+1})$ *for all* $0 \leq i \leq k'$ *if* $s_1 = o$,
- $k' = k$, $(s'_0,\lambda'_0,\mu'_0,\rho'_0) = (o,\lambda_0,\mu_0,\rho_0)$ *and* $(s'_i,\lambda'_i,\mu'_i,\rho'_i) = (s_i,\lambda_i,\mu_i,\rho_i)$ *for all* $1 \leq i \leq k'$ *otherwise,*

$\mathcal{U}' = \mathcal{U} \setminus \mathcal{U}^{(o)}$ *and* $\mathcal{L}' = \mathcal{L} \setminus \mathcal{L}^{(o)}$.

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

$$\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \begin{cases} \dfrac{1}{|\mathrm{L}_\mathcal{T}|!}\displaystyle\sum_{A\in\mathcal{S}} |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}} \dfrac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)} & \text{if } o < s_k, \\ \mathbf{T}(\mathcal{T})\mathbf{X}_\Theta(s_0,s_1,|\mathrm{L}_\mathcal{T}|) & \text{otherwise.} \end{cases}$$

Proof. Let us start with the case where $o = s_k$, i.e., the case where the oldest time is the ending time of the diversification. By construction, we then have necessarily that $k = 1$ and that $\mathcal{U}$ and $\mathcal{L}$ are both empty. It follows that $(s_0, s_1, \mathcal{T})$ is a standard pattern of probability $\mathbf{T}(\mathcal{T})\mathbf{X}_\Theta(s_0,s_1,|\mathrm{L}_\mathcal{T}|)$ from Lemma 1.

Let us now assume that $o < s_k$. Under the notations of the theorem and being a divergence time assignation of $\mathcal{T}$ 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 time constraints $\mathcal{U}$ and $\mathcal{L}$. Since all these configurations are mutually exclusive, by putting $\mathbf{D}_{\Theta,A}(\mathcal{T},\mathcal{U},\mathcal{L})$ for the probability of observing the topology $\mathcal{T}$ with $\mathcal{A}_o = A$ and the time constraints $\mathcal{U}$ and $\mathcal{L}$, the law of total probabilities gives us that

$$\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L}) = \sum_{A\in\mathcal{S}} \mathbf{D}_{\Theta,A}(\mathcal{T},\mathcal{U},\mathcal{L}). \tag{1}$$

For instance, the entries of the second column of Figure 3 (just after the sign sum) represent all the start-sets A in $\Omega^\bullet_{\mathcal{T}_\mathfrak{b}}$.

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

- the part of the diversification process anterior to o is the standard pattern $(s_0, o, \Gamma_{\mathcal{T},A})$ and that
- the part of the diversification process posterior to o consists of all the tree topologies $\mathcal{T}_n$ with times constraints $\mathcal{U}_{[\mathcal{T}_n]}, \mathcal{L}_{[\mathcal{T}_n]}$ with $n \in \mathrm{L}_{\Gamma_{\mathcal{T},A}}$ under the model Θ' (i.e., the model Θ restricted to the interval of times $[o, s_k]$), which have probability ${\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}_{[\mathcal{T}_n]},\mathcal{L}_{[\mathcal{T}_n]})}/{\mathbf{O}_\Theta(o)}$ conditioned on the observability of their starting lineages.

Since piecewise-constant-birth-death-sampling models are Markovian, evolution of all the tree topologies $\mathcal{T}_n$ are independent one to another and with regard to the part of the process anterior to o, conditional upon starting with an observable lineage at time o.

From Lemma 1, the probability of the standard pattern $(s_0, o, \Gamma_{\mathcal{T},A})$ is $\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)$ 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 though 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 Lemma 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

$$\begin{aligned} \mathbf{D}_{\Theta,A}(\mathcal{T},\mathcal{U},\mathcal{L}) &= |\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\frac{\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}|\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_\mathcal{T}|!}\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}_{[\mathcal{T}_n]},\mathcal{L}_{[\mathcal{T}_n]})}{\mathbf{O}_\Theta(o)} \\ &= \frac{|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_\Theta(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)}{|\mathrm{L}_\mathcal{T}|!}\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}_{[\mathcal{T}_n]},\mathcal{L}_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_\Theta(o)}, \end{aligned}$$

which, with Equation 1, ends the proof. The whole computation of a toy example is schematized in Figure 3. □

Theorem 2 states that $\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ can be either calculated directly (if $o = s_k$) or expressed as a sum-product of probabilities of tree topologies with temporal constraints under piecewise-constant-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 ending time of the diversification for which the probability can be calculated directly. To summarize, the probability density $\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ can be computed by recursively applying Theorem 2.

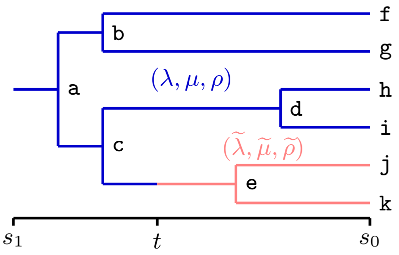


Figure 4: A tree topology with a shift at time t for the clade $\{\mathtt{e},\mathtt{j},\mathtt{k}\}$.

5.2 Shifts

We shall see how to compute the probability of a tree topology $\mathcal{T}$ under a simple birth-death-sampling model $((s_0,\lambda,\mu,\rho),s_1)$ by assuming that one of its clades follows another birth-death-sampling model $((t,\widetilde{\lambda},\widetilde{\mu},\widetilde{\rho}),s_1)$ from a given time $t\in[s_0,s_1]$ to the ending time s_1. Note that this implicitly assumes that the lineage originating this particular clade was alive at t (Fig. 4). For the sake of simplicity, we consider simple birth-death-sampling models in this section. Computing probabilities under the general piecewise-constant-birth-death-sampling is possible but requires some extra work since the probability for a clade to be observable depends on whether it contains the clade following the other model.

Theorem 3. *Let $\mathcal{T}$ be a tree topology, $s_0\leq t\leq s_1$ be three times, $\Theta=((s_0,\lambda,\mu,\rho),s_1)$ and $\widetilde{\Theta}=((t,\widetilde{\lambda},\widetilde{\mu},\widetilde{\rho}),s_1)$ be two birth-death-sampling models from origin times s_0 and t respectively and both to end time s_1, and m be an internal node of $\mathcal{T}$. By setting $\Theta'=((t,\lambda,\mu,\rho),s_1)$, 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*

$$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t)=\frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{A\in\Omega^{\times}_{\mathcal{T},m}}(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s_0,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\mathbf{D}_{\widetilde{\Theta}}(\mathcal{T}_m,\mathcal{U},\mathcal{L})|\mathrm{L}_{\mathcal{T}_m}|!\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U},\mathcal{L})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(t)}$$

Proof. Assuming that a diversification shift of the clade originating at m occurs at time t implies that all the divergence time 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. ~~Conversely~~ any divergence time assignation verifying these two conditions is consistent with this assumption. 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~~s~~ 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

$$\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t)=\sum_{A\in\Omega^{\times}_{\mathcal{T},m}}\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t). \qquad (2)$$

From the Markov property, we have that the $\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t)$ can be written as the product of the part of the diversification anterior to t, which is the special pattern $(s_0,t,\Gamma_{\mathcal{T},A})$ 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 s_1 by following model Θ' 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 ${}^{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U},\mathcal{L})}/_{\mathbf{O}_{\Theta}(t)}$ while the special one is not conditioned and has probability $\mathbf{D}_{\widetilde{\Theta}}(\mathcal{T}_m,\mathcal{U},\mathcal{L})$.

From Lemma 2, the probability of the special pattern $(s_0,t,\Gamma_{\mathcal{T},A})$ is $\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s_0,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)$ under the assumption that $\Gamma_{\mathcal{T},A}$ is labeled. The situation slightly differs with 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 to multiply the probability provided by Lemma 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

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

We get that

$$\begin{aligned}\mathbf{S}_{\Theta,\widetilde{\Theta},A}(\mathcal{T},m,t) &= \mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s_0,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\frac{\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}|\mathrm{L}_{\mathcal{T}_n}|!}{|\mathrm{L}_{\mathcal{T}}|!}\mathbf{D}_{\widetilde{\Theta}}(\mathcal{T}_m,\mathcal{U},\mathcal{L})\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U},\mathcal{L})}{\mathbf{O}_{\Theta}(o)}\\ &= \frac{(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s_0,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\mathbf{D}_{\widetilde{\Theta}}(\mathcal{T}_m,\mathcal{U},\mathcal{L})|\mathrm{L}_{\mathcal{T}_m}|!}{|\mathrm{L}_{\mathcal{T}}|!}\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U},\mathcal{L})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(o)},\end{aligned}$$

which with Equation 2 ends the proof.

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

$$\begin{aligned}&\mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T},m,t)\\ &= \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{A\in\Omega^{\times}_{\mathcal{T},m}}(|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|-1)!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{Y}_{\Theta}(s_0,t,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\mathbf{T}(\mathcal{T}_m)\mathbf{X}_{\widetilde{\Theta}}(t,s_1,|\mathrm{L}_{\mathcal{T}_m}|)|\mathrm{L}_{\mathcal{T}_m}|!\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}\setminus\{m\}}\frac{\mathbf{T}(\mathcal{T}_n)\mathbf{X}_{\Theta'}(t,s_1,|\mathrm{L}_{\mathcal{T}_n}|)|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(t)}.\end{aligned}$$

□

6 A quadratic computation

Since the number of start-sets may be exponential with the size of the tree, notably for balanced trees, Theorems 2 and 3 do not directly provide a polynomial algorithm for computing the probabilities. We shall show in this section that the left-side of the equation of Theorem 2 can be factorized in order to obtain a polynomial computation. Under the assumptions and notations of Theorem 2, we have that

$$\mathbf{D}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L}) = \begin{cases}\dfrac{1}{|\mathrm{L}_{\mathcal{T}}|!}\displaystyle\sum_{A\in\mathcal{S}}|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_{\Theta}(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}\dfrac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(o)} & \text{if } o<s_k,\\ \mathbf{T}(\mathcal{T})\mathbf{X}_{\Theta}(s_0,s_1,|\mathrm{L}_{\mathcal{T}}|) & \text{otherwise.}\end{cases}$$

Since in the case where $o = s_k$, the computation of $\mathbf{D}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L})$ is performed in constant time, we focus on the case where $o < s_k$. 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:

$$\begin{aligned}\mathbf{D}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L}) &= \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{A\in\mathcal{S}}|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_{\Theta}(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(o)}\\ &= \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{k=1}^{|\mathrm{L}_{\mathcal{T}}|}\sum_{A\in\Upsilon^{(k)}_{\mathcal{S}}}|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|!\,\mathbf{T}(\Gamma_{\mathcal{T},A})\mathbf{X}_{\Theta}(s_0,o,|\mathrm{L}_{\Gamma_{\mathcal{T},A}}|)\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}\frac{\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!}{\mathbf{O}_{\Theta}(o)}\\ &= \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{k=1}^{|\mathrm{L}_{\mathcal{T}}|}\frac{\mathbf{X}_{\Theta}(s_0,o,k)k!}{\mathbf{O}_{\Theta}(o)^k}\sum_{A\in\Upsilon^{(k)}_{\mathcal{S}}}\mathbf{T}(\Gamma_{\mathcal{T},A})\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,.\end{aligned}$$

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

$$\mathbf{W}_{m,k} = \sum_{A\in\Upsilon^{(k)}_{\mathcal{S},m}}\mathbf{T}(\Gamma_{\mathcal{T},A})\prod_{n\in \mathrm{L}_{\Gamma_{\mathcal{T},A}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,.$$

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

$$\mathbf{D}_{\Theta}(\mathcal{T},\mathcal{U},\mathcal{L}) = \frac{1}{|\mathrm{L}_{\mathcal{T}}|!}\sum_{k=1}^{|\mathrm{L}_{\mathcal{T}}|}\frac{\mathbf{X}_{\Theta}(s_0,o,k)k!}{\mathbf{O}_{\Theta}(o)^k}\mathbf{W}_{r,k}. \tag{3}$$

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

$$\mathbf{W}_{m,1} = \mathbf{D}_{\Theta'}(\mathcal{T}_m, \mathcal{U}'_{[\mathcal{T}_m]}, \mathcal{L}'_{[\mathcal{T}_m]})|\mathrm{L}_{\mathcal{T}_m}|!\,. \tag{4}$$

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

$$\mathbf{T}(\Gamma_{\mathcal{T}_m A})\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_m A}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|! =$$
$$\frac{2|\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}|!\,|\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}|!}{(k-1)k!}\mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J})\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big)\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big).$$

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

$$\mathbf{W}_{m,k} = \sum_{\substack{i,j\\ i+j=k}}\ \sum_{\substack{(I,J)\in\\ \Upsilon^{(i)}_{\mathcal{S},a}\times\Upsilon^{(j)}_{\mathcal{S},b}}}\frac{2i!\,j!}{(k-1)k!}\mathbf{T}(\Gamma_{\mathcal{T}_a,I})\mathbf{T}(\Gamma_{\mathcal{T}_b,J})\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_a,I}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big)\Big(\prod_{n\in\mathrm{L}_{\Gamma_{\mathcal{T}_b,J}}}\mathbf{D}_{\Theta'}(\mathcal{T}_n,\mathcal{U}'_{[\mathcal{T}_n]},\mathcal{L}'_{[\mathcal{T}_n]})|\mathrm{L}_{\mathcal{T}_n}|!\,\Big).$$

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

$$\mathbf{W}_{m,k} = \sum_{\substack{i,j\\ i+j=k}}\frac{2i!\,j!}{(k-1)k!}\mathbf{W}_{a,i}\mathbf{W}_{b,j}. \tag{5}$$

The following remark is straightforward to prove by induction.

Remark 1. *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 5 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 1 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 $O(|\mathcal{T}|^2)$.

Theorem 4. *Let $\mathcal{T}$ be a tree topology, $\Theta=((s_i,\lambda_i,\mu_i,\rho_i)_{0\leq i<k},s_k)$ be a piecewise-constant-birth-death-sampling model and $\mathcal{U}$ and $\mathcal{L}$ be two sets of upper and lower temporal constraints respectively. By setting $\Delta=k+|\mathcal{U}|+|\mathcal{L}|$, the probability $\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ can be computed with time complexity $O(\Delta\times|\mathcal{T}|^2)$ and memory space complexity $O(|\mathcal{T}|^2)$.*

Proof. We shall proceed by induction on Δ, i.e., the total number of times involved in the model and the temporal constraints, by proving that at each stage, all the probabilities $\mathbf{D}_\Theta(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$ can be calculated with a total time complexity $O(|\mathcal{T}|^2)$.

In the base case where $\Delta=1$, we have necessarily that Θ is a simple birth-death-sampling model $((s_0,\lambda_0,\mu_0,\rho_0),s_1)$ and that both $\mathcal{U}$ and $\mathcal{L}$ are empty. From Theorem 2, the probability $\mathbf{D}_\Theta(\mathcal{T},\mathcal{U},\mathcal{L})$ can then be calculated in constant time since, under the notations of the theorem, we have $o=s_1$. In the same way, the probabilities $\mathbf{D}_\Theta(\mathcal{T}_m,\mathcal{U}_{[\mathcal{T}_m]},\mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$ can be calculated with a total time complexity $O(|\mathcal{T}|)$.

Let us now assume that $\Delta>1$ and that we have already computed the probabilities $\mathbf{D}_{\Theta'}(\mathcal{T}_m,\mathcal{U}'_{[\mathcal{T}_m]},\mathcal{L}'_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$ (under the notations of Theorem 2). From Equation 4, the quantities $\mathbf{W}_{m,1}$ for all internal nodes m are calculated directly from the probabilities $\mathbf{D}_{\Theta'}(\mathcal{T}_m,\mathcal{U}'_{[\mathcal{T}_m]},\mathcal{L}'_{[\mathcal{T}_m]})$, thus in $O(|\mathcal{T}|)$. From Remark 1, 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 3 can then be applied to all subtrees of $\mathcal{T}$ in order to compute the probabilities $\mathbf{D}_\Theta(\mathcal{T}_m, \mathcal{U}_{[\mathcal{T}_m]}, \mathcal{L}_{[\mathcal{T}_m]})$ from the quantities $\mathbf{W}_{m,k}$ for all internal nodes m of $\mathcal{T}$. Since computing each $\mathbf{D}_\Theta(\mathcal{T}_m, \mathcal{U}_{[\mathcal{T}_m]}, \mathcal{L}_{[\mathcal{T}_m]})$ requires to sum $|\mathrm{L}_{\mathcal{T}_m}|$ terms, computing all the $\mathbf{D}_\Theta(\mathcal{T}_m, \mathcal{U}_{[\mathcal{T}_m]}, \mathcal{L}_{[\mathcal{T}_m]})$ has total time complexity $O(|\mathcal{T}|^2)$.

To sum up, being given the probabilities $\mathbf{D}_{\Theta'}(\mathcal{T}_m, \mathcal{U}'_{[\mathcal{T}_m]}, \mathcal{L}'_{[\mathcal{T}_m]})$, computing the probabilities $\mathbf{D}_\Theta(\mathcal{T}_m, \mathcal{U}_{[\mathcal{T}_m]}, \mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$ has total time complexity $O(|\mathcal{T}|^2)$. Since we have that $k' + |\mathcal{U}'| + |\mathcal{L}'| < k + |\mathcal{U}| + |\mathcal{L}| = \Delta$, it requires at most $\Delta - 1$ stages to end up with the base case which has time complexity $O(|\mathcal{T}|)$. The total time complexity is thus $O(\Delta \times |\mathcal{T}|^2)$.

Last, since, at each stage, we have to store only the quantities $(\mathbf{W}_{m,k})_{m\in\mathcal{T},k=1,\dots,|\mathrm{L}_{\mathcal{T}_m}|}$ and the probabilities $\mathbf{D}_{\Theta'}(\mathcal{T}_m, \mathcal{U}'_{[\mathcal{T}_m]}, \mathcal{L}'_{[\mathcal{T}_m]})$ and $\mathbf{D}_\Theta(\mathcal{T}_m, \mathcal{U}_{[\mathcal{T}_m]}, \mathcal{L}_{[\mathcal{T}_m]})$ for all internal nodes m of $\mathcal{T}$, the total memory space complexity is $O(|\mathcal{T}|^2)$. □

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 $O(|\mathcal{T}|^2)$.

7 Divergence time distributions

We shall apply Theorem 2 to compute divergence time distributions of tree topologies with time constraints under piecewise-constant-birth-death-sampling models.

Corollary 1. *Let $\mathcal{T}$ be a tree topology, $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$ a piecewise-constant-birth-death-sampling model from origin time s_0 to end time s_k, $\mathcal{U} = \{(n_1, u_1), \dots, (n_\ell, u_\ell)\}$ 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 $t \in [s_0, s_k]$ conditioned on observing the tree topology $\mathcal{T}$ with the temporal constraints $\mathcal{U}$ and $\mathcal{L}$ under Θ is*

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

The computation of the divergence time distributions was performed on a contrived tree topology and on the Hominoidea subtree. Results are displayed in Figures 5 and 6 where the probability densities are computed from the corresponding distributions by finite difference approximations.

Figure 5 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 [6]. The approach can actually compute the divergence time distributions of the whole Primates tree of [6] 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 6 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 the position of distributions, to investigate which parameter values are consistent with a given evolutionary assumption etc.

We observe on Figure 6 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 ending 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 ending time.

The divergence time distributions obtained for $\lambda = 0.1$, $\mu = 0.01$ and $\rho = 0.9$ (Fig. 6, column 2, top) and for $\lambda = 1$, $\mu = 0.91$ and $\rho = 0.1$ (Fig. 6, 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. 6, column 4, top) and for $\lambda = 1$, $\mu = 0.99$ and $\rho = 0.1$ (Fig. 6, column 3, bottom). This point suggests that estimating the birth-death-sampling parameters from the divergence times might be difficult, even if the divergence times are accurately determined.

The variety of shapes of divergence times probability densities observed in Figures 5 and 6 exceeds that of standard prior distributions used in phylogenetic inference, e.g., uniform, lognormal, gamma, exponential [15, 13].

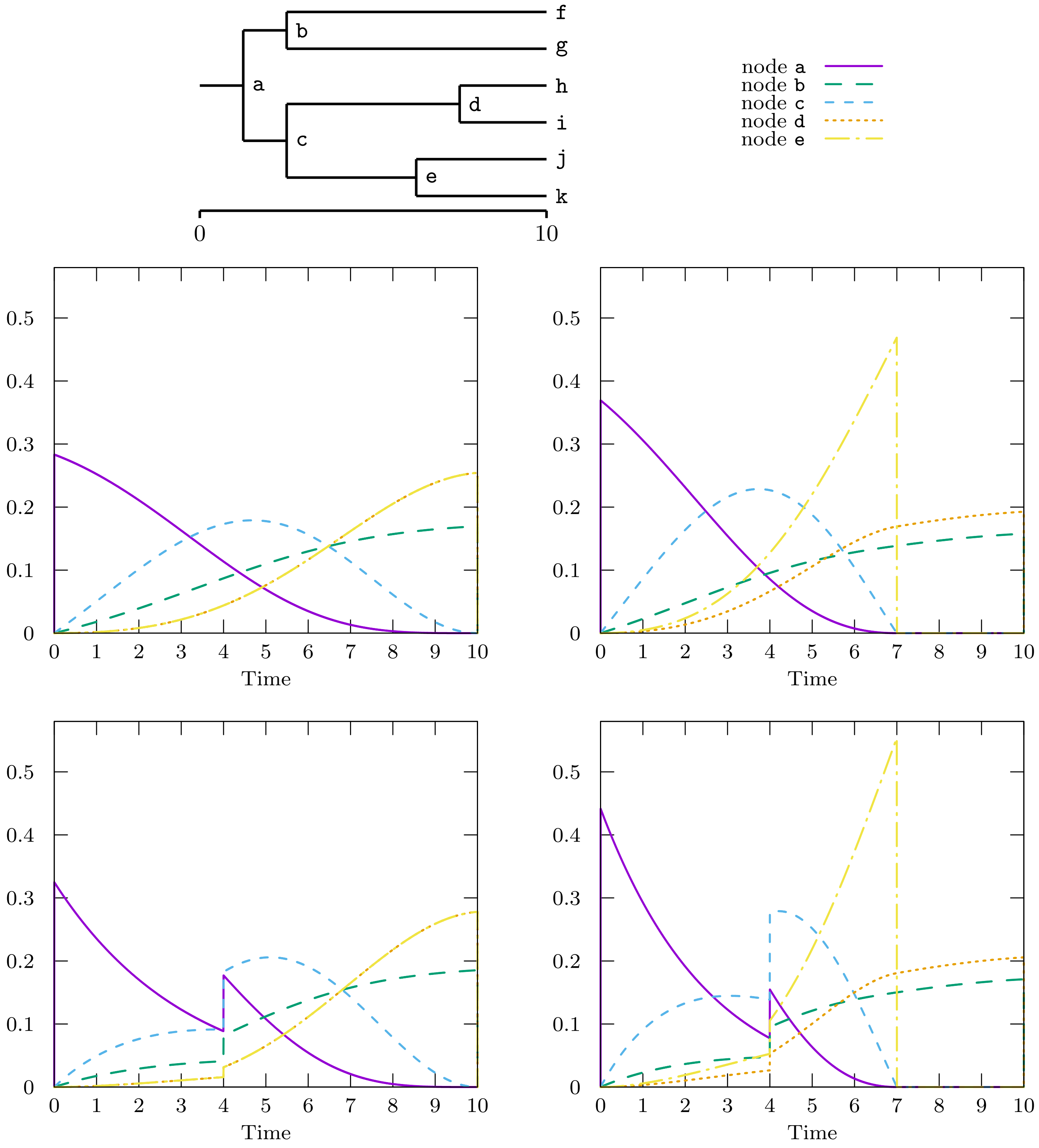
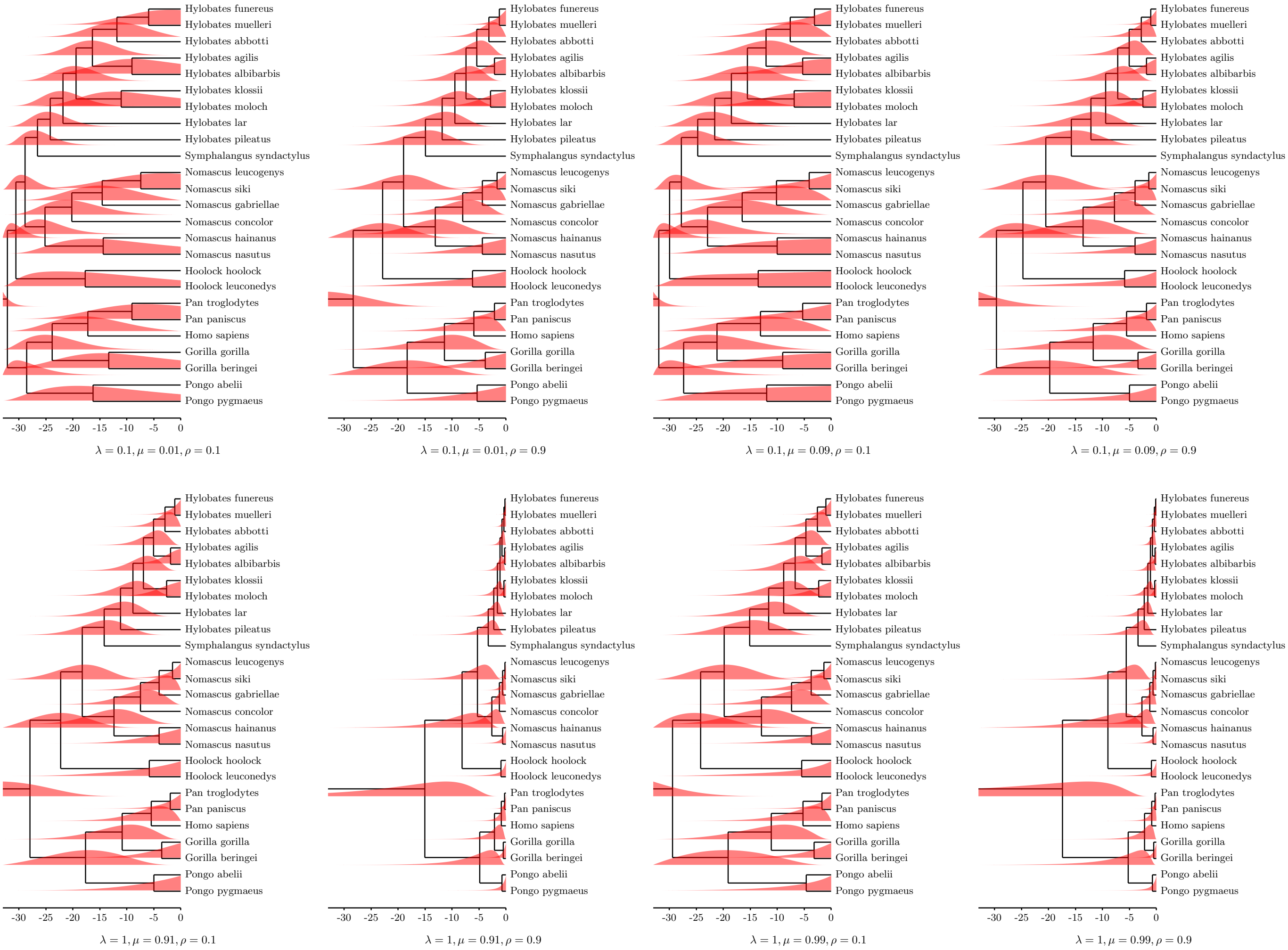


Figure 5: Divergence time probability densities of the tree displayed at the first row, in the second row by assuming a diversification process running from time 0 to 10 under a birth-death-sampling model with parameters $\lambda = 0.2$, $\mu = 0.02$ and $\rho = 0.5$ between times 0 and 10 and in the third row 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.



$\lambda = 0.1, \mu = 0.01, \rho = 0.1$

$\lambda = 0.1, \mu = 0.01, \rho = 0.9$

$\lambda = 0.1, \mu = 0.09, \rho = 0.1$

$\lambda = 0.1, \mu = 0.09, \rho = 0.9$

$\lambda = 1, \mu = 0.91, \rho = 0.1$

$\lambda = 1, \mu = 0.91, \rho = 0.9$

$\lambda = 1, \mu = 0.99, \rho = 0.1$

$\lambda = 1, \mu = 0.99, \rho = 0.9$

Figure 6: Divergence time probability densities of the Hominoidea tree from [6] 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.

7.1 A previous approach

A previous approach for computing the probability density of a given divergence time is provided in [9]. It is based on the explicit computation of the probability density $f_{\mathcal{A}^k_{n,t}}$ of the k^{th} divergence time of a tree topology with n tips starting at t from the present, provided in [9], and the computation of the probability $\mathsf{P}(r(v) = k)$ for the rank $r(v)$ of the divergence time associated to the vertex v to be the k^{th} which was given in [10]. 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

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

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

The computation of the probability density of the k^{th} 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 [9] 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 Corollary 1 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 piecewise-constant-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 piecewise-constant-birth-death-sampling model.

Lemma 3. *Let $\mathcal{T}$ be a tree topology of root r, $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0 \leq i < k}, s_k)$ be a piecewise-constant-birth-death-sampling model from origin time s_0 to end time s_k and t be a time in $(s_i, s_{i+1}]$. By setting $\Theta' = ((s'_i, \lambda'_i, \mu'_i, \rho'_i)_{0 \leq i < k'}, s'_{k'+1})$ where $s'_{k'+1} = s_k$, $k' = k - i + 1$ and $(s'_0, \lambda'_0, \mu'_0, \rho'_0) = (t, \lambda_0, \mu_0, \rho_0)$ and $(s'_j, \lambda'_j, \mu'_j, \rho'_j) = (s_{i+j}, \lambda_{i+j}, \mu_{i+j}, \rho_{i+j})$ for all $1 \leq j \leq k'$. The probability that the root divergence time τ_r is anterior to a time $t \in [s_0, s_k]$ conditioned on observing the tree topology $\mathcal{T}$ under the birth-death-sampling model (λ, μ, ρ) is*

$$\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T}) = 1 - \frac{\mathbf{I}_\Theta(s_0, t)\mathbf{D}_{\Theta'}(\mathcal{T}, \emptyset, \emptyset)}{\mathbf{D}_\Theta(\mathcal{T}, \emptyset, \emptyset)}.$$

Proof. The probability that the divergence time τ_r associated with r is anterior to a time $t \in [s_0, s_k]$ is the complementary probability that $\tau_r > t$. Observing $\tau_r > t$ means that the starting lineage at s_0 has a single descendant observable at t from which descends the tree topology $\mathcal{T}$ sampled at s_k. It follows that

$$\begin{aligned}\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T}) &= 1 - \mathbf{P}_\Theta(\mathcal{T}, \tau_r > t \mid \mathcal{T}) \\ &= 1 - \frac{\mathbf{I}_\Theta(s_0, t)\mathbf{D}_{\Theta'}(\mathcal{T}, \emptyset, \emptyset)}{\mathbf{D}_\Theta(\mathcal{T}, \emptyset, \emptyset)}.\end{aligned}$$

□

The probability $\mathbf{P}_\Theta(\mathcal{T}, \tau_r < t \mid \mathcal{T})$ can be directly written as ${}^{\mathbf{D}_\Theta(\mathcal{T},(r,t),\emptyset)}/_{\mathbf{D}_\Theta(\mathcal{T},\emptyset,\emptyset)}$. Lemma 3 allows to avoid considering a temporal constraint, which is particularly interesting in the simple birth-death-sampling case.

Remark 2. *Under the birth-death-sampling model $((s_0, \lambda_0, \mu_0, \rho_0), s_1)$, we have that*

$$\mathbf{P}_{((s_0,\lambda_0,\mu_0,\rho_0),s_1)}(\mathcal{T}, \tau_r < t \mid \mathcal{T}) = 1 - \left[\frac{(1 - \mathrm{e}^{-(\lambda_0-\mu_0)(s_1-t)})(\rho_0\lambda_0 + (\lambda_0(1-\rho_0) - \mu_0)\mathrm{e}^{-(\lambda_0-\mu_0)(s_1-s_0)})}{(1 - \mathrm{e}^{-(\lambda_0-\mu_0)(s_1-s_0)})(\rho_0\lambda_0 + (\lambda_0(1-\rho_0) - \mu_0)\mathrm{e}^{-(\lambda_0-\mu_0)(s_1-t)})}\right]^{|\mathrm{L}_\mathcal{T}|-1},$$

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., [1, 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 [1, 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 s_k. By applying Lemma ?? to $\mathcal{T}_a$ and $\mathcal{T}_b$ between t and s_k, 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 $\mathcal{T}$. If $\Theta = ((s_i, \lambda_i, \mu_i, \rho_i)_{0\leq i<k}, s_k)$ with $k > 1$, each sampling of a divergence time of $\mathcal{T}$ has complexity $O(-\log(\epsilon)k|\mathcal{T}|^2)$, where ϵ is the precision required on the samples. The total complexity for sampling all the divergence times is therefore $O(-\log(\epsilon)k|\mathcal{T}|^3)$.

From Remark 2, under the simple birth-death-sampling model $\Theta = ((s_0, \lambda_0, \mu_0, \rho_0), s_1)$, the computation of $\mathbf{P}_\Theta(\tau_r < t \mid \mathcal{T})$ 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.

For the sake of simplicity, we showed how to sample divergence times under a piecewise-constant-birth-death-sampling model only but the same approach can be applied in order to sample divergence times with temporal constraints and/or shifts, still under a piecewise-constant-birth-death-sampling model.

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 single birth-death-sampling model for the whole topology with the alternative model including a shift as displayed in Figure 4. 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

$$\Lambda_N = \frac{\mathbf{S}_{\Theta_1,\widetilde{\Theta}_1}(\mathcal{T}, m, t)}{\mathbf{D}_{\Theta_0}(\mathcal{T}, \emptyset, \emptyset)},$$

where Θ_0, Θ_1, $\widetilde{\Theta}_1$ are birth-death-sampling models with

$$\Theta_0 = \arg\max_{\Theta} \mathbf{D}_\Theta(\mathcal{T}, \emptyset, \emptyset) \text{ and } (\Theta_1, \widetilde{\Theta}_1) = \arg\max_{(\Theta,\widetilde{\Theta})} \mathbf{S}_{\Theta,\widetilde{\Theta}}(\mathcal{T}, m, t).$$

In order to assess the accuracy of Λ_N, we compare it to three sister-group diversity tests considered in [32]. 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 [27] is $P = \dfrac{2N_2}{N_1 + N_2 - 1}$,
- the likelihood ratio alternative provided in [26] 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 [32] 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 $\hat{\lambda}_2^+ = \dfrac{N_2 - 1}{B_2}$.

We 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, we first simulated topologies without shift from the general birth rate. Next, we 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, we 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 ending times in order to eventually obtain a topology with shift.

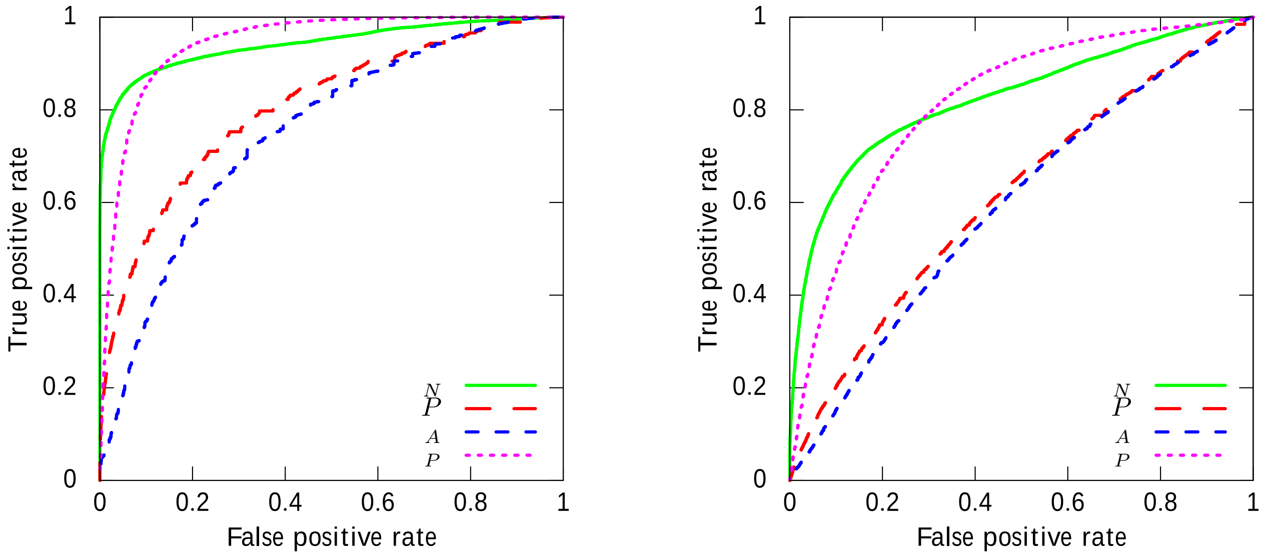


Figure 7: ROC plots of different measures for shift detection at left (resp. at right) are obtained by simulated 50000 Yule topologies with birth rate 0.4 (resp. 0.6) from times 0 to 10 and birth rate 1.0 from the shift time 5 to 10 for one of the clades present at time 5.

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 7 displays the 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. 7-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 is only partially outperformed by Λ_P, which is obtained from all the divergence times and the shift time. If one requires a false positive discovery rate below 10%, the likelihood ratio test Λ_N obtained from Theorem 3 is the most powerful.

References

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

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

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

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

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

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

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

[15] 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.

[16] D. Kendall. On some modes of population growth leading to RA Fisher's logarithmic series distribution. *Biometrika*, 35(1/2):6–15, 1948.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

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

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

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

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

[31] 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.

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

[33] 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.

[34] 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.

[35] 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.