Exact distribution of divergence times under the birth-death-sampling model

Gilles Didier

Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

gilles.didier@univ-amu.fr

July 25, 2018

Abstract

Divergence times estimation is an essential component of evolutionary studies. It generally involves several types of data (molecular, morphological, fossils. . .) and relies on stochastic models of evolution. An important part of these models is the prior distribution of divergence times that they use. The birth-death-sampling model, which is certainly the simplest realistic model of phylogenetic trees, is a natural choice for deriving such prior distributions. It has been well studied and is still widely used in this context.

The main result provided here is a method for computing the exact distribution of the divergence times of any phylogenetic tree under a birth-death-sampling model. This computation has a cubic time-complexity, allowing us to deal with phylogenies of hundreds of tips on standard computers. The approach can be used for dating phylogenetic trees from their topologies only, for visualizing effects of diversification parameters etc.

An additional result shows how to directly sample all the divergence times of a phylogenetic tree with linear complexity under the same model. This sampling procedure can be integrated into phylogenetic inference methods, e.g., for proposing accurate MCMC moves.

1   Introduction

Estimating divergence times is an essential and difficult stage of phylogenetic inference [18, 19, 12, 3, 16, 6]. 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 [25, 20, 7, 9]. Both in a Bayesian and a maximum likelihood context, an important component of these stochastic models is the "prior" probability distribution of divergence times (i.e., that which does not take into account information about genotype or phenotype of species [26, 11, 9]), which is often derived from diversification models [4, 10, 26, 13, 14]. Among these, the birth-death-sampling model is arguably the simplest realistic model since it includes three important features shaping phylogenetic trees [27, 28]. 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. Birth-death-sampling model has been further studied and is currently used for phylogenetic inference [22, 24, 10, 4].

The main result of this work is a method to compute the exact distributions of divergence times of a given tree topology from the parameters of a birth-death-sampling model. Namely, for any internal node of the phylogeny and any time t, we provide an algorithm which computes the exact probability for the divergence time associated with this node to be anterior to t. The complexity of this algorithm is polynomial with the size of the phylogeny, namely cubic in time and quadratic in memory space. In practice, it can deal with phylogenetic trees with up to hundreds of tips on standard desktop computers.

The computation of divergence time distributions 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* [8, 17]. Second, it can provide prior distributions in phylogenetic inference frameworks. 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.

An additional result of this work is a method allowing to directly draw samples of the set all the divergence times of a phylogenetic tree according to the birth-death-sampling model. Though based on the same ideas, this sampling procedure is independent of the computation of distributions referred to just above. The sampling procedure is very fast and can easily be integrated into phylogenetic inference software [6, 21], e.g., for proposing accurate MCMC moves.

The approaches presented here can be extended in several directions. In particular, it is quite straightforward to deal with heterogeneity in time and to model massive extinctions as in [23]. Taking into account heterogeneity in the speciation-extinction rates and/or the sampling probability between clades seems feasible but is a more difficult
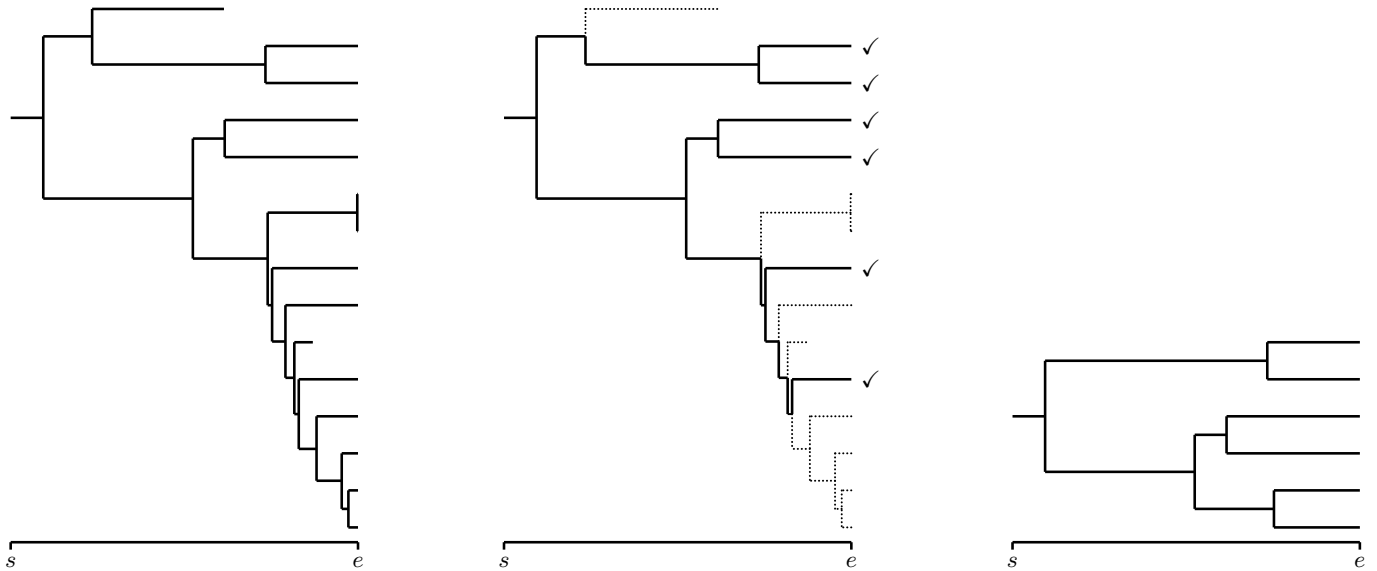
1

Figure 1: Left: the whole diversification process (e.g., from a simple birth-death 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.

question which I plan to investigate soon. In collaboration with a co-author of [2], we are currently integrating fossil finds in the 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 birth-death-sampling model is available at https://github.com/gilles-didier/DateBDS.

The rest of the paper is organized as follows. Birth-death-sampling models are formally introduced in Section 2. Section 3 presents definitions and some results about tree topologies. The "start- and end-patterns", i.e., the subparts of the diversification process from which are computed divergence times probabilities, are introduced in Section 4. Sections 5 and 6 present the computation of the divergence time distributions and a polynomial algorithm to perform it respectively. The method for directly sampling the divergence times is described in Section 7. The computation of divergence time distributions is illustrated with the phylogenetic tree of Hominoidae from [5] in Section 8.

2 Birth-death-sampling model

The dynamics of speciation and extinction of species is modelled as a birth-death process with constant rates λ and μ both through time and lineage [15]. Following [27], each extant species is assumed to be independently sampled (i.e. included in the study) with probability ρ. The whole model will be referred to as the *birth-death-sampling model* and has thus three parameters:

- λ: the speciation rate,
- μ: the extinction rate and
- ρ: the probability for an extant taxa to be sampled.

In all what follows, we make the technical assumption that $\lambda > \mu$ and $\rho > 0$.

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* (Figure 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* (Figure 1-Right). It can be shown that the reconstructed process of a birth-death-sampling process is a pure-birth process with time-inhomogeneous birth rate.

Let us start by recalling some already derived probabilities of interest under this model. Under the simple birth-death model (i.e. with $\rho = 1$) with parameters (λ, μ), the probability $\overline{\mathbf{Q}}(n, t)$ that a single lineage at time 0 has exactly n descendants at time t was given in [15]. We have that

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

and, for all $n > 0$,

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

2

Under the birth-death-sampling model (λ, μ, ρ), the probability $\mathbf{Q}(n, t)$ that a single lineage at time 0 has exactly n descendants sampled at time t was given in [27]. We have that

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

and, for all $n > 0$,

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

In the case where all the tips are sampled, we obtain the same equations as those of [15] just above, i.e., if $\rho = 1$ then $\mathbf{Q}(n, t) = \overline{\mathbf{Q}}(n, t)$ for all λ, μ, n and t.

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 *labelled*, which means their tips, and consequently all their nodes, are unambiguously identified. From now on, "tree topology" has to be understood as "labelled-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 $\mathbf{L}_\mathcal{T}$ 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 $|\mathbf{L}_\mathcal{T}|$ 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 [2, Supp. Mat., Appendix 2].

Theorem 1 ([2]). *A tree topology $\mathcal{T}$ resulting from a pure-birth realization of a lineage-homogeneous process has probability*

$$\mathbf{T}(\mathcal{T}) = \frac{\mathbf{R}_\mathcal{T}}{(|\mathbf{L}_\mathcal{T}|-1)!\, |\mathbf{L}_\mathcal{T}|!},$$

conditioned on having $|\mathbf{L}_\mathcal{T}|$ tips, where $\mathbf{R}_\mathcal{T}$ denotes the number of divergence time rankings consistent with $\mathcal{T}$.

If $|\mathcal{T}| = 1$, i.e. $\mathcal{T}$ is a single tip, we have $\mathbf{R}_\mathcal{T} = 1$.

Otherwise, by putting a and b for the two direct descendants of the root of $\mathcal{T}$, we have

$$\mathbf{R}_\mathcal{T} = 2\binom{|\mathbf{L}_{\mathcal{T}_a}|+|\mathbf{L}_{\mathcal{T}_b}|-2}{|\mathbf{L}_{\mathcal{T}_a}|-1}\mathbf{R}_{\mathcal{T}_a}\mathbf{R}_{\mathcal{T}_b}.$$

Theorem 1 implies that both $\mathbf{R}_\mathcal{T}$ and $\mathbf{T}(\mathcal{T})$ 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 the empty set $\emptyset$ is start-set of any topology.

Being given a tree topology $\mathcal{T}$ and a non-empty start-set A, we define

- the *start-tree* $\alpha_{\mathcal{T},A}$ as the subtree topology of $\mathcal{T}$ made of all nodes in A and their direct descendants (Figure 2-Center);

- the *set of end-trees* $\Omega_{\mathcal{T},A}$ as the set of all the subtrees of $\mathcal{T}$ rooted at the tips of $\alpha_{\mathcal{T},A}$, i.e. $\Omega_{\mathcal{T},A} = \{\mathcal{T}_m \mid m \in \mathbf{L}_{\alpha_{\mathcal{T},A}}\}$ (Figure 2-Right).

By convention, $\alpha_{\mathcal{T},\emptyset}$, the start-tree associated to the empty start-set, is the subtree topology made only of the root of $\mathcal{T}$. There is then only a single end-tree that is the whole tree topology $\mathcal{T}$, i.e., $\Omega_{\mathcal{T},\emptyset} = \{\mathcal{T}\}$.

In all cases, there are as many trees in $\Omega_{\mathcal{T},A}$ as tips in the tree $\alpha_{\mathcal{T},A}$.

For all internal nodes n of the tree topology $\mathcal{T}$, we define $\Gamma_{\mathcal{T},n}$ as the set of all start-sets of $\mathcal{T}$ which contain n.
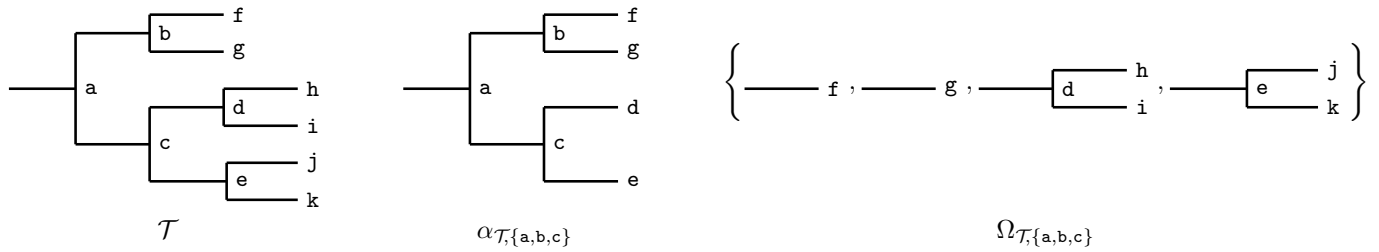
Figure 2: Left: A tree topology $\mathcal{T}$ of which $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ is start-set. Center: The start-tree $\alpha_{\mathcal{T}, \{\mathtt{a},\mathtt{b},\mathtt{c}\}}$. Right: The set of end-trees $\Omega_{\mathcal{T}, \{\mathtt{a},\mathtt{b},\mathtt{c}\}}$. Root nodes are represented with the branch that they end.
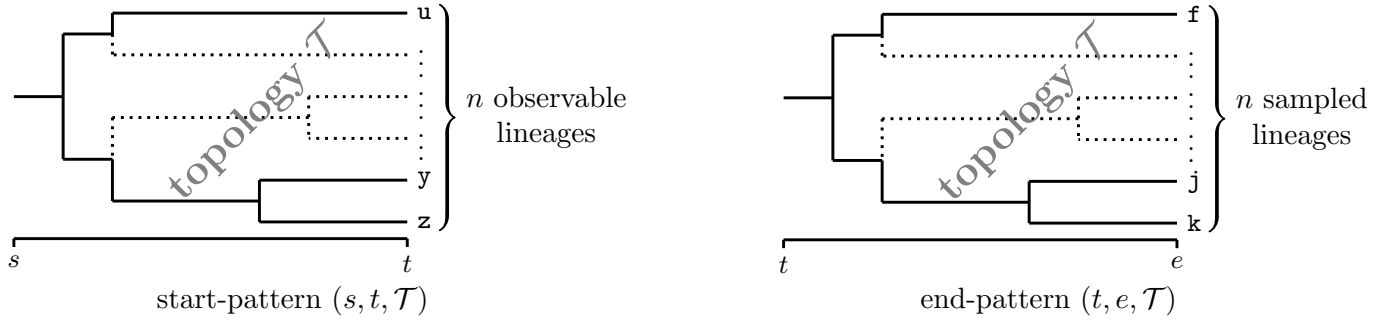

Figure 3: The two types of patterns which are the basis of our computation.

4 Patterns

From now on, we shall consider diversification processes starting at origin time s and ending at time e by evolving following a birth-death-sampling model (λ, μ, ρ). In practice, the ending time e is usually the present time. 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 the resulting tree topology and of the bounding times (Figure 3). We shall consider two types of patterns. *Start-patterns* start from the origin o of the diversification process and end at a given time $t \in [s, e]$. *End-patterns* go from a given time $t \in [s, e]$ to the ending time e.

Start- and end-patterns are very similar to patterns of types **c** and **a** defined in [2], respectively. Proofs of Lemmas 1 and 2 are essentially the same as those of the corresponding claims in [2].

4.1 Start-patterns

Start-patterns encompass the *observed* part of the diversification process running from the origin time s until a given time $t \in [s, e]$. We recall that, by *observed* part, we mean the part that can be reconstructed from the ending time. More formally, for all times $t \in [s, e]$, a lineage alive at time t is *observable* if itself or one of its descendants is sampled at e.

Definition 1. *A start-pattern $(s, t, \mathcal{T})$ starts with a single lineage at the time origin s and ends with n observable lineages and a tree topology $\mathcal{T}$ at t (Figure 3-left).*

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

$$\mathbf{O}(t) = 1 - \mathbf{Q}(0, e - t)$$
$$= \frac{\rho(\lambda - \mu)}{\rho\lambda + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)(e-t)}}.$$

Let us now compute the probability $\mathbf{X}(k, s, t)$ that a single lineage at time s has k descendants observable from e at time $t \in [s, e]$. This probability is the sum over all numbers $n \geq k$, of the probability that the lineage at s has n descendants at t in the whole process (i.e., without sampling, which is equal to $\overline{\mathbf{Q}}(n, t - s)$), among which exactly k ones are observable (i.e., $\binom{n}{k}\mathbf{O}(t)^k(1 - \mathbf{O}(t))^{n-k}$). By setting $n = j + k$, we thus have

$$\mathbf{X}(k, s, t) = \sum_{j=0}^{\infty} \overline{\mathbf{Q}}(j + k, t - s)\binom{j + k}{k}\mathbf{O}(t)^k(1 - \mathbf{O}(t))^j$$
$$= \frac{\rho^k(\lambda - \mu)e^{-(\lambda-\mu)(t-s)}\left[\lambda(1 - e^{-(\lambda-\mu)(t-s)})\right]^{k-1}\left[\rho\lambda + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)(e-t)}\right]}{\left[\rho\lambda + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)(e-s)}\right]^{k+1}}.$$

4

Figure 4: Schematic of the computation of the probability that the divergence time associated with node b is strictly anterior to t. 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.

Lemma 1. *Under the birth-death-sampling model (λ, μ, ρ), the probability of the start-pattern $(s, t, \mathcal{T})$ is*

$$\mathbf{S}(s, t, \mathcal{T}) = \mathbf{T}(\mathcal{T})\mathbf{X}(|\mathbf{L}_\mathcal{T}|, s, t).$$

Proof. The probability of the start-pattern $(s, t, \mathcal{T})$ is the probability of the tree topology $\mathcal{T}$ conditioned on its number of tips, which is $\mathbf{T}(\mathcal{T})$ from Theorem 1, multiplied by the probability of observing this number of tips in a start-pattern, which is that of getting $|\mathbf{L}_\mathcal{T}|$ observable lineages at t from a single lineage at s, i.e., $\mathbf{X}(|\mathbf{L}_\mathcal{T}|, s, t)$. $\square$

4.2 End-patterns

End-patterns encompass all the observed parts of the diversification process which start from a single lineage at a given time $t \in [s, e]$ until the ending time of the process.

Definition 2. *An end-pattern $(t, e, \mathcal{T})$ starts with a single lineage at time t and ends with $|\mathbf{L}_\mathcal{T}|$ lineages and a tree topology $\mathcal{T}$ at the ending time e (Figure 3-right).*

Lemma 2. *Under the birth-death-sampling model (λ, μ, ρ), the probability of the end-pattern $(t, e, \mathcal{T})$ is*

$$\mathbf{E}(t, e, \mathcal{T}) = \mathbf{T}(\mathcal{T})\mathbf{Q}(|\mathbf{L}_\mathcal{T}|, e - t).$$

Proof. The probability of the end-pattern $(t, e, \mathcal{T})$ is the probability of the tree topology $\mathcal{T}$ conditioned on its number of tips, which is $\mathbf{T}(\mathcal{T})$ from Theorem 1, multiplied by the probability of observing this number of tips in an end-pattern, which is the probability that a single lineage at time t has $|\mathbf{L}_\mathcal{T}|$ descendants sampled at time e, i.e., $\mathbf{Q}(|\mathbf{L}_\mathcal{T}|, e - t)$ [27]. $\square$

5 Distribution of divergence times

Theorem 2. *Let $\mathcal{T}$ be a tree topology, $s < e$ be two times and n be an internal node of $\mathcal{T}$. The joint probability of observing the tree topology $\mathcal{T}$ and that the divergence time τ_n associated with n is anterior to a time $t \in [s, e]$ from a diversification process starting at s and ending at e following the birth-death-sampling model (λ, μ, ρ) is*

$$\mathbf{P}(\mathcal{T}, \tau_n < t) = \frac{1}{|\mathbf{L}_\mathcal{T}|!} \sum_{A \in \Gamma_{\mathcal{T},n}} \mathbf{S}(s, t, \alpha_{\mathcal{T},A}) \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \frac{\mathbf{E}(t, e, \mathcal{T}')|\mathbf{L}_{\mathcal{T}'}|!}{\mathbf{O}(t)}.$$

Proof. Under the notations of the theorem, let us define $\mathcal{A}_n$ as the set of nodes of $\mathcal{T}$ whose divergence times are anterior to t (i.e. $\mathcal{A}_n = \{m \in \mathcal{T} \mid \tau_m < t\}$). Since divergence times corresponding to ancestors of a given node are always posterior to its own divergence time, all sets $\mathcal{A}_n$ are start-sets. Moreover, any event including "$\tau_n < t$" implies that $n \in \mathcal{A}_n$, thus that $\mathcal{A}_n \subseteq \Gamma_{\mathcal{T},n}$. By construction, the set $\Gamma_{\mathcal{T},n}$ contains all the possible configurations of nodes of

5

$\mathcal{T}$ with divergence times anterior to t such that $\tau_n < t$. Since all these possibilities are mutually exclusive, the law of total probabilities gives us that

$$\mathbf{P}(\mathcal{T}, \tau_n < t) = \sum_{A \in \Gamma_{\mathcal{T},n}} \mathbf{P}(\mathcal{T}, \tau_n < t, \mathcal{A}_n = A). \tag{1}$$

The entries of the second column of Figure 4 (just after the sign sum) represent all the start-sets A in $\Gamma_{\mathcal{T},b}$.

In order to compute the probability $\mathbf{P}(\mathcal{T}, \tau_n < t, \mathcal{A}_n = A)$ for a start-set $A \in \Gamma_{\mathcal{T},n}$, we remark that

- the part of the diversification process anterior to t is the start-pattern $(s, t, \alpha_{\mathcal{T},A})$ and

- the part of the diversification process posterior to t consists of all end patterns $(t, e, \mathcal{T}')$ with $\mathcal{T}' \in \Omega_{\mathcal{T},A}$.

Since the birth-death-sampling process is a Markov process, evolution of all the end patterns are independent one to another and with regard to the part of the process anterior to t, conditional upon starting with an observable lineage at time t. By assuming that its set of tip labels is known and conditional upon starting with an observable lineage, the probability of the end pattern $(t, e, \mathcal{T}')$ is

$$\frac{\mathbf{E}(t, e, \mathcal{T}')}{\mathbf{O}(t)}, \quad \text{for all } \mathcal{T}' \in \Omega_{\mathcal{T},A}.$$

From Lemma 1, the probability of the start-pattern $(s, t, \alpha_{\mathcal{T},A})$ is $\mathbf{S}(s, t, \alpha_{\mathcal{T},A})$ under the assumption that $\alpha_{\mathcal{T},A}$ is labelled. We have not a direct labeling of $\alpha_{\mathcal{T},A}$ here : tips of $\alpha_{\mathcal{T},A}$ are identified though the labels of their tip descendants in $\mathcal{T}$, i.e., the tips of the subtrees in $\Omega_{\mathcal{T},A}$. Since all labellings of $\mathcal{T}$ are equiprobable, the probability of a labeling of $\alpha_{\mathcal{T},A}$ is the inverse of the number of ways of choosing a subset of $|\mathbf{L}_{\mathcal{T}_m}|$ labels from $|\mathbf{L}_{\mathcal{T}}|$ ones for all tips m of $\alpha_{\mathcal{T},A}$ without replacement, i.e. the inverse of corresponding multinomial coefficient, which is

$$\frac{\prod_{m \in \mathbf{L}_{\alpha_{\mathcal{T},A}}} |\mathbf{L}_{\mathcal{T}_m}|!}{|\mathbf{L}_{\mathcal{T}}|!} = \frac{\prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} |\mathbf{L}_{\mathcal{T}'}|!}{|\mathbf{L}_{\mathcal{T}}|!}.$$

By construction, a labeling of $\alpha_{\mathcal{T},A}$ fully determines the set of tip-labels of all subtrees $\mathcal{T}_m$ with $m \in \mathbf{L}_{\alpha_{\mathcal{T},A}}$. Putting all together, we eventually get that

$$\mathbf{P}(\mathcal{T}, \tau_n < t, \mathcal{A}_n = A) = \mathbf{S}(s, t, \alpha_{\mathcal{T},A}) \frac{\prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} |\mathbf{L}_{\mathcal{T}'}|!}{|\mathbf{L}_{\mathcal{T}}|!} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \frac{\mathbf{E}(t, e, \mathcal{T}')}{\mathbf{O}(t)}$$

$$= \frac{1}{|\mathbf{L}_{\mathcal{T}}|!} \mathbf{S}(s, t, \alpha_{\mathcal{T},A}) \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \frac{\mathbf{E}(t, e, \mathcal{T}') |\mathbf{L}_{\mathcal{T}'}|!}{\mathbf{O}(t)},$$

which, with Equation 1, ends the proof. The whole computation of a toy example is schematized in Figure 4. $\qquad \square$

Corollary 1. *Let $\mathcal{T}$ be a tree topology, s and e be the origin and end times of the diversification process and n be an internal node of $\mathcal{T}$. The probability that the divergence time τ_n associated with n is anterior to a time $t \in [s, e]$ conditioned on observing the tree topology $\mathcal{T}$ under the birth-death-sampling model (λ, μ, ρ) is*

$$\mathbf{P}(\tau_n < t \mid \mathcal{T}) = \frac{\mathbf{P}(\mathcal{T}, \tau_n < t)}{\mathbf{E}(s, e, \mathcal{T})}$$

Proof. It is enough to remark that the probability of observing a tree topology $\mathcal{T}$ (without constraint on its divergence times other than belonging to $[s, e]$) for a diversification process starting from time s with a single lineage and ending at time e is, by definition, exactly that of the end-pattern $(s, e, \mathcal{T})$ given in Lemma 2. $\qquad \square$

We put F_n for the cumulative distribution function (CDF) of the divergence time associated with node n, namely,

$$F_n : t \to \mathbf{P}(\tau_n < t \mid \mathcal{T}).$$

Figure 5 displays the probability densities of all the divergence times of the tree at its left. In Figures 5 and 6, the densities are computed from the corresponding distributions by finite difference approximations.
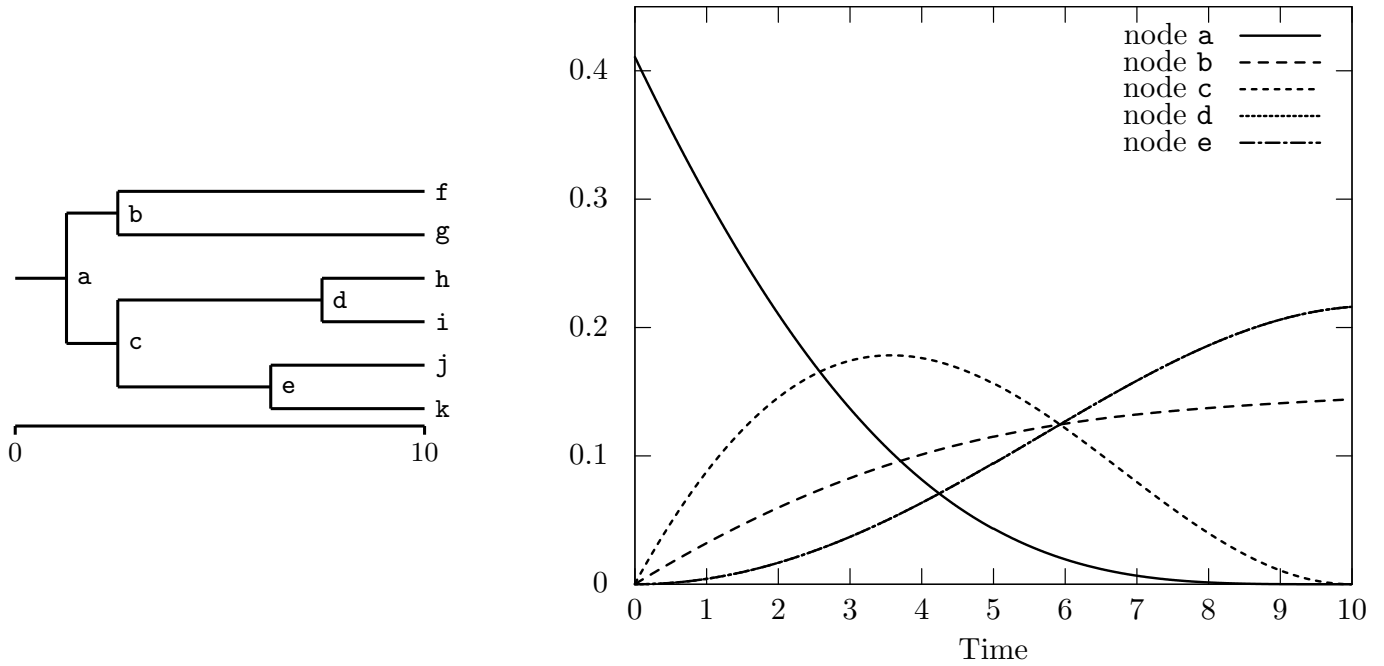
6

Figure 5: Divergence time probability densities of the tree at the left, by assuming a diversification process running from time 0 to 10 under a birth-death-sampling model with parameters $\lambda = 0.1$, $\mu = 0.02$ and $\rho = 0.5$. Densities for nodes d and e are confounded.

6 A polynomial computation

Since the number of start-sets may be exponential with the size of the tree, notably for balanced trees, Theorem 2 does not directly provide a polynomial algorithm for computing the divergence time distributions. 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.

Let us first introduce an additional notation. For all tree topologies $\mathcal{T}$, all internal nodes n of $\mathcal{T}$ and all numbers k between 1 and the number of tips of $\mathcal{T}$, we put $\Gamma_{\mathcal{T},n}^{(k)}$ for the set of start-sets A of $\mathcal{T}$ containing n and such that the corresponding start-tree $\alpha_{\mathcal{T},A}$ has exactly $|\mathbf{L}_{\alpha_{\mathcal{T},A}}| = k$ tips. By construction, a start-tree of $\mathcal{T}$ has at least one tip and at most $|\mathbf{L}_{\mathcal{T}}|$ tips. We have:

$$\mathbf{P}(\tau_n < t) = \frac{1}{|\mathbf{L}_{\mathcal{T}}|!} \sum_{k=1}^{|\mathbf{L}_{\mathcal{T}}|} \sum_{A \in \Gamma_{\mathcal{T},n}^{(k)}} \mathbf{S}(s, t, \alpha_{\mathcal{T},A}) \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \frac{\mathbf{E}(t, e, \mathcal{T}')|\mathbf{L}_{\mathcal{T}'}|!}{\mathbf{O}(t)}$$

$$= \frac{1}{|\mathbf{L}_{\mathcal{T}}|!} \sum_{k=1}^{|\mathbf{L}_{\mathcal{T}}|} \sum_{A \in \Gamma_{\mathcal{T},n}^{(k)}} \frac{\mathbf{R}_{\alpha_{\mathcal{T},A}} \mathbf{X}(|\mathbf{L}_{\alpha_{\mathcal{T},A}}|, s, t)}{(|\mathbf{L}_{\alpha_{\mathcal{T},A}}| - 1)! |\mathbf{L}_{\alpha_{\mathcal{T},A}}|!} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \frac{\mathbf{E}(t, e, \mathcal{T}')|\mathbf{L}_{\mathcal{T}'}|!}{\mathbf{O}(t)}$$

$$= \frac{1}{|\mathbf{L}_{\mathcal{T}}|!} \sum_{k=1}^{|\mathbf{L}_{\mathcal{T}}|} \frac{\mathbf{X}(k, s, t)}{(k-1)! \, k! \, \mathbf{O}(t)^k} \sum_{A \in \Gamma_{\mathcal{T},n}^{(k)}} \mathbf{R}_{\alpha_{\mathcal{T},A}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T},A}} \mathbf{E}(t, e, \mathcal{T}')|\mathbf{L}_{\mathcal{T}'}|!$$

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

$$\Upsilon_{\mathcal{T},n}^{(m)} = \bigcup_{A \in \Gamma_{\mathcal{T},n}} A \cap \mathcal{T}_m,$$

where $\mathcal{T}_m$ stands here for the set of nodes of the subtree topology rooted at m. Since, by construction, the elements of $\Upsilon_{\mathcal{T},n}^{(m)}$ are start-sets of the tree topology $\mathcal{T}_m$, the start-tree $\alpha_{\mathcal{T}_m,A}$ and the set of end-trees $\Omega_{\mathcal{T}_m,A}$ are well-defined for all $A \in \Upsilon_{\mathcal{T},n}^{(m)}$. For all numbers $1 \le k \le |\mathbf{L}_{\mathcal{T}_m}|$, we put $\Upsilon_{\mathcal{T},n}^{(m,k)}$ for the set of start-sets $A \in \Upsilon_{\mathcal{T},n}^{(m)}$ such that the corresponding start-tree $\alpha_{\mathcal{T}_m,A}$ has exactly k tips.

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

$$\mathbf{W}_{m,k} = \sum_{A \in \Upsilon_{\mathcal{T},n}^{(m,k)}} \mathbf{R}_{\alpha_{\mathcal{T}_m,A}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_m,A}} \left(\mathbf{E}(t, e, \mathcal{T}')|\mathbf{L}_{\mathcal{T}'}|! \right).$$

7

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

$$\mathbf{P}(\tau_n < t) = \frac{1}{|\mathbf{L}_\mathcal{T}|!} \sum_{k=1}^{|\mathbf{L}_\mathcal{T}|} \frac{\mathbf{X}(k,s,t)}{(k-1)!\,k!\,\mathbf{O}(t)^k} \mathbf{W}_{r,k}. \tag{2}$$

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

Let us first consider the case where $k = 1$. Since $\Upsilon_{\mathcal{T},n}^{(m,1)} = \{\emptyset\}$ for all nodes m of $\mathcal{T}$, we have that

$$\mathbf{W}_{m,1} = \mathbf{R}_{\mathcal{T}_m} \mathbf{E}(t,e,\mathcal{T}_m). \tag{3}$$

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

From Theorem 1, we have that

$$\mathbf{R}_{\alpha_{\mathcal{T}_m,A}} = 2\binom{|\mathbf{L}_{\alpha_{\mathcal{T}_a,I}}| + |\mathbf{L}_{\alpha_{\mathcal{T}_b,J}}| - 2}{|\mathbf{L}_{\alpha_{\mathcal{T}_a,I}}| - 1} \mathbf{R}_{\alpha_{\mathcal{T}_a,I}} \mathbf{R}_{\alpha_{\mathcal{T}_b,J}}.$$

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

$$\mathbf{R}_{\alpha_{\mathcal{T}_m,A}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_m,A}} \mathbf{E}(t,e,\mathcal{T}') = 2\binom{|\mathbf{L}_{\alpha_{\mathcal{T}_a,I}}| + |\mathbf{L}_{\alpha_{\mathcal{T}_b,J}}| - 2}{|\mathbf{L}_{\alpha_{\mathcal{T}_a,I}}| - 1} [\mathbf{R}_{\alpha_{\mathcal{T}_a,I}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_a,I}} \mathbf{E}(t,e,\mathcal{T}')][\mathbf{R}_{\alpha_{\mathcal{T}_b,J}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_b,J}} \mathbf{E}(t,e,\mathcal{T}')].$$

More generally, the start-sets of $\Upsilon_{\mathcal{T},n}^{(m,k)}$ are in one-to-one correspondence with the set of pairs (I,J) of $\Upsilon_{\mathcal{T},n}^{(a)} \Upsilon_{\mathcal{T},n}^{(b)}$ such that $|\mathbf{L}_{\alpha_{\mathcal{T}_a,I}}| + |\mathbf{L}_{\alpha_{\mathcal{T}_b,I}}| = k$. 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 $\Upsilon_{\mathcal{T},n}^{(a,i)} \times \Upsilon_{\mathcal{T},n}^{(b,j)}$. It follows that

$$\mathbf{W}_{m,k} = \sum_{\substack{i,j \\ i+j=k}} \sum_{\substack{(I,J)\in \\ \Upsilon_{\mathcal{T},n}^{(a,i)} \times \Upsilon_{\mathcal{T},n}^{(b,j)}}} 2\binom{i+j-2}{i-1} [\mathbf{R}_{\alpha_{\mathcal{T}_a,I}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_a,I}} \mathbf{E}(t,e,\mathcal{T}')][\mathbf{R}_{\alpha_{\mathcal{T}_b,J}} \prod_{\mathcal{T}' \in \Omega_{\mathcal{T}_b,J}} \mathbf{E}(t,e,\mathcal{T}')].$$

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

$$\mathbf{W}_{m,k} = \sum_{\substack{i,j \\ i+j=k}} 2\binom{i+j-2}{i-1} \mathbf{W}_{a,i} \mathbf{W}_{b,j}. \tag{4}$$

Theorem 3. *Let $\mathcal{T}$ be a tree topology, s and e be the origin and end times of the diversification process and n be an internal node of $\mathcal{T}$. Both $\mathbf{P}(\mathcal{T}, \tau_n < t)$ and $\mathbf{P}(\tau_n < t \mid \mathcal{T})$ can be computed with time-complexity $O(|\mathcal{T}|^3)$ and memory-space-complexity $O(|\mathcal{T}|^2)$.*

Proof. For all tips m of $\mathcal{T}$, $(\mathbf{W}_{m,k})_{k=1,\dots,|\mathbf{L}_{\mathcal{T}_m}|}$ reduces to $(\mathbf{W}_{m,1})$, which is directly given by Equation 3.

For all internal nodes m of $\mathcal{T}$, $(\mathbf{W}_{m,1})$ is also given by Equation 3. Equation 4 shows that for all $2 \leq k \leq |\mathbf{L}_{\mathcal{T}_m}|$, $\mathbf{W}_{m,k}$ can be computed from $(\mathbf{W}_{a,i})_{i=1,\dots,|\mathbf{L}_{\mathcal{T}_a}|}$ and $(\mathbf{W}_{b,i})_{i=1,\dots,|\mathbf{L}_{\mathcal{T}_b}|}$, where a and b are the direct descendants of m, in $O(|\mathcal{T}_m|)$ operations.

In sum, for all nodes m of $\mathcal{T}$, the quantities $(\mathbf{W}_{m,k})_{m \in \mathcal{T}, k=1,\dots,|\mathbf{L}_{\mathcal{T}_m}|}$ require $O(|\mathbf{L}_{\mathcal{T}_m}|)$ memory space to be stored, each entry $\mathbf{W}_{m,k}$ requiring $O(|\mathbf{L}_{\mathcal{T}_m}|)$ operations to be computed. It follows that the quantities $(\mathbf{W}_{r,k})_{r \in \mathcal{T}, k=1,\dots,|\mathbf{L}_{\mathcal{T}_r}|}$ can be determined with a post-order traversal of $\mathcal{T}$ in $O(|\mathcal{T}|^3)$ time by using $O(|\mathcal{T}|^2)$ memory space.

From Equation 2, the probability $\mathbf{P}(\tau_n < t)$ can be computed with complexity $O(|\mathbf{L}_\mathcal{T}|)$ from the quantities $(\mathbf{W}_{r,k})_{k=1,\dots,|\mathbf{L}_\mathcal{T}|}$, which ends the proof. □

7 Direct sampling of divergence times

Theorems 2 and 3 and Corollary 1 show how to compute the marginal (with regard to other divergence times) of the divergence time distribution of any internal node of a phylogenetic tree from given birth-death-sampling parameters and origin and end times of the diversification. 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, still being given birth-death-sampling parameters and origin and end times of the diversification.

8

Lemma 3. *Let $\mathcal{T}$ be a tree topology, s and e be the origin and end times of the diversification process and r be the root of $\mathcal{T}$. 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 the birth-death-sampling model (λ, μ, ρ) is*

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

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. By putting $\mathcal{I}$ for the tree topology made of a single tip/lineage, it follows that

$$\begin{aligned} \mathbf{P}(\tau_r < t \mid \mathcal{T}) &= 1 - \mathbf{P}(\tau_r > t \mid \mathcal{T}) \\ &= 1 - \frac{\mathbf{S}(t, e, \mathcal{I})\mathbf{E}(t, e, \mathcal{T})}{\mathbf{E}(s, e, \mathcal{T})\mathbf{O}(t)} \\ &= 1 - \left[\frac{(1 - e^{-(\lambda-\mu)(e-t)})(\rho\lambda + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)(e-s)})}{(1 - e^{-(\lambda-\mu)(e-s)})(\rho\lambda + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)(e-t)})} \right]^{|\mathbf{L}_{\mathcal{T}}|-1}. \end{aligned}$$

$\square$

Remark that the computation of $\mathbf{P}(\tau_r < t \mid \mathcal{T})$ requires only the number of tips of $\mathcal{T}$ (in particular, the shape of $\mathcal{T}$ does not matter). Lemma 3 implies that the CDF $F_r : t \to \mathbf{P}(\tau_r < t)$ can be computed at any time t with complexity $O(1)$. Moreover, one can verify that F_r is strictly increasing under the assumptions on the birth-death-sampling parameters, i.e., $\lambda > \mu$ and $\rho > 0$.

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 F_r. 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]). Though Lemma 3 provides an expression of $\mathbf{P}(\tau_n < t \mid \mathcal{T})$, I did not find an explicit formula for F_r^{-1}. We thus 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 e. By applying Lemma 3 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 $\mathcal{T}$.

In short, 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.

8   Example – Influence of the birth-death-sampling parameters

In order to illustrate the computation of the divergence time distributions, let us consider the Hominoidea subtree from the Primates tree of [5]. The approach can actually compute the divergence time distributions of the whole Primates tree of [5] but they cannot be displayed legibly because of its size.

The divergence time distributions were computed under several sets of birth-death-sampling parameters, namely all combinations with $\lambda = 0.1$ or 1, $\mu = \lambda - 0.09$ or $\lambda - 0.01$ and $\rho = 0.1$ or 0.9. Since $\lambda - \mu$ 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.
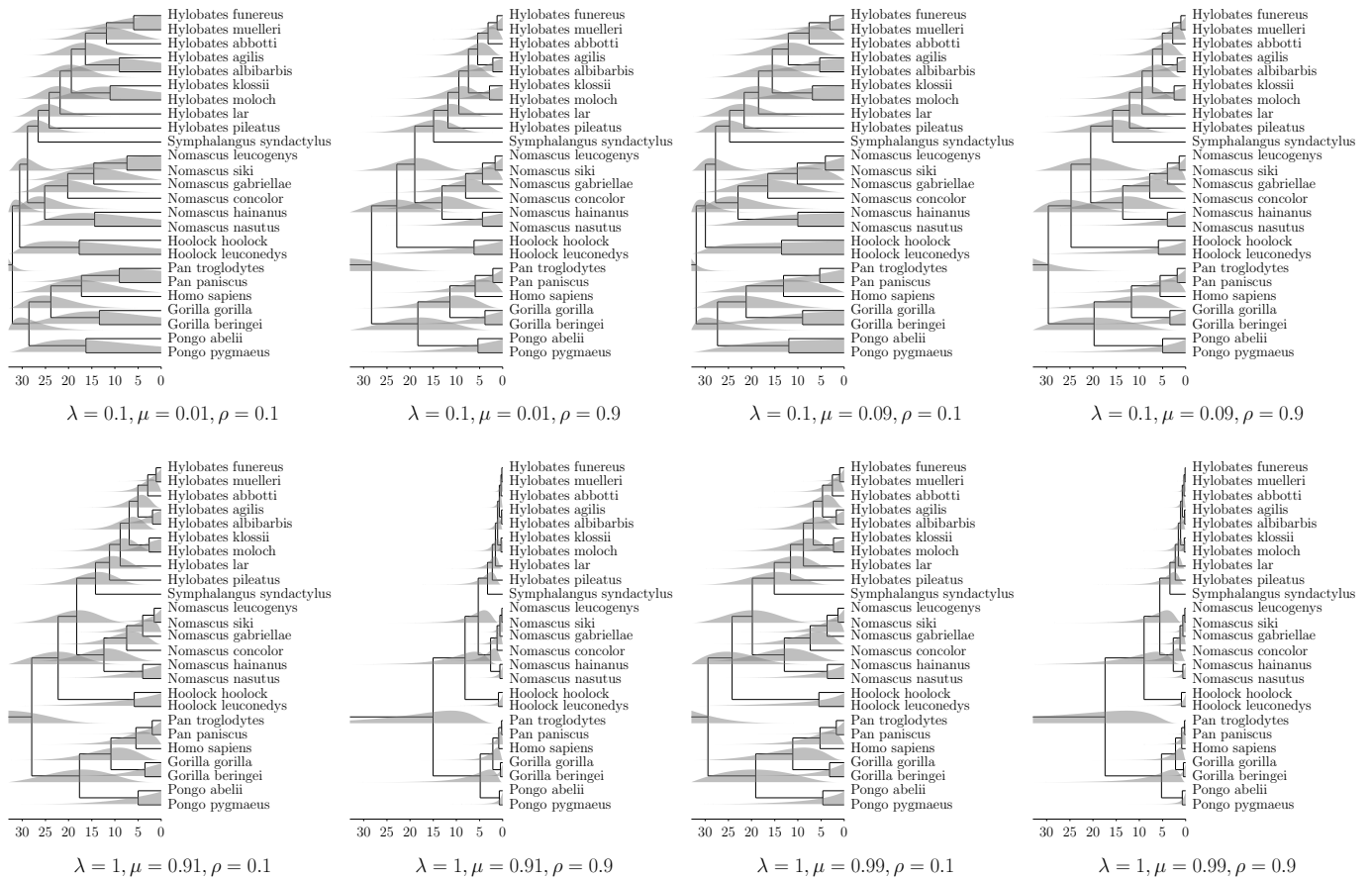
Figure 6: Divergence time probability densities of the Hominoidea tree from [5] under birth-death-sampling models with parameters $\lambda = 0.1$ or 1, $\mu = 0.01$ or 0.09 and $\rho = 0.1$ or 0.9. Internal nodes are positioned at their median divergence time.

The divergence time distributions obtained for $\lambda = 0.1$, $\mu = 0.01$ and $\rho = 0.9$ (Figure 6, column 2, top) and for $\lambda = 1$, $\mu = 0.91$ and $\rho = 0.1$ (Figure 6, column 1, bottom) are close one to another. The same remark holds for $\lambda = 0.1$, $\mu = 0.09$ and $\rho = 0.9$ (Figure 6, column 4, top) and for $\lambda = 1$, $\mu = 0.99$ and $\rho = 0.1$ (Figure 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 [11, 9].

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

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

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

[11] 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.

[12] 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.

[13] N. J. Matzke and A. Wright. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biology Letters*, 12(8), 2016.

[14] S. Möller, L. du Plessis, and T. Stadler. Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 2018.

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.