Probabilistic models for the (sub)tree(s) of life

Amaury Lambert

UPMC Univ Paris 06

Abstract. The goal of these lectures is to review some mathematical aspects of random tree models used in evolutionary biology to model species trees.

We start with stochastic models of tree shapes (finite trees without edge lengths), culminating in the β-family of Aldous' branching models.

We next introduce real trees (trees as metric spaces) and show how to study them through their contour, provided they are properly measured and ordered.

We then focus on the reduced tree, or coalescent tree, which is the tree spanned by species alive at the same fixed time. We show how reduced trees, like any compact ultrametric space, can be represented in a simple way via the so-called comb metric. Beautiful examples of random combs include the Kingman coalescent and coalescent point processes.

We end up displaying some recent biological applications of coalescent point processes to the inference of species diversification, to conservation biology and to epidemiology.

Contents

1 Tree shapes

Standard references on the topic of this section include: Barthélémy and Guénoche (1991), Stanley (1999) and Semple and Steel (2003).

1.1 Definitions

In these lecture notes, we call *tree shape* any rooted, finite tree (no edge lengths, no plane embedding). A finite tree τ is an acyclic graph $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \mathcal{V}(\tau)$ is a finite set whose elements are called the *vertices* or *nodes* of τ, and $\mathcal{E} = \mathcal{E}(\tau)$ is a subset of $\mathcal{V} \times \mathcal{V}$ whose elements are (not ordered and) called the *edges* of τ. The root of τ is a distinguished vertex of τ.

   We will also use the following terminology.

- Degree. The degree of a node u of the tree shape τ is the number of its neighbors, where a neighbor of u is an element v of τ such that (u, v) is an edge of τ. We will always assume that the root has degree 1.
- Partial order. If u and v are two nodes of τ, we say that v is descending from u and we write $u \preceq v$, if v is not in the same connected component of $\tau \setminus \{u\}$ as the root. The relation $\preceq$ is a (partial) order on the vertex set of τ.
- Tips/leaves. All nodes with degree 1 but the root, are called *tips* or *leaves*. All other nodes (including the root) are called *internal nodes* (or internal vertices).

- A tree shape is said *binary* when each of its internal nodes (but the root) has degree 3.

In the next section, we will introduce a framework not needed at this stage, known as Ulam–Harris–Neveu coding, in which finite trees can be embedded into the set $\mathscr{U}$ of finite words, where a word v descends from a word u if u is a prefix of v.

We also mention that in combinatorial phylogenetics, a tree whose tips are labelled by some finite set X is usually identified as a so-called X-*hierarchy* (Semple and Steel (2003)). An X-hierarchy is a collection $\mathcal{H}$ of subsets of X containing all singletons and such that for any $A, B \in \mathcal{H}$, $A \cap B \in \{\varnothing, A, B\}$.

Exercise 1.1.1. Display the labelled tree coded by the $\{1, 2, 3, 4, 5\}$-hierarchy $\mathcal{H} = \{\{1\}, \{1, 4\}, \{1, 4, 5\}, \{2\}, \{2, 3\}, \{3\}, \{4\}, \{5\}\}$.

1.2 Combinatorics warm up

1.2.1 *Counting trees*.

Definition 1.2.1. We let $\mathscr{T}_n$ denote the set of all binary tree shapes with n tips, and $\mathscr{T}_n^{\ell}$ the set of binary tree shapes with n tips labelled by $\{1, \ldots, n\}$. The elements of $\mathscr{T}_n^{\ell}$ are called *cladograms* or *labelled tree shapes* (with n tips).

Definition 1.2.2. For any (labelled or not) tree shape τ, we call *radial order* any total order $\sqsubset$ on the internal nodes of τ respecting the genealogical order. In other words, $\sqsubset$ is a radial order if for any internal nodes u and v of τ,

$$u \preceq v \quad \Rightarrow \quad u \sqsubset v,$$

where we recall that $u \preceq v$ means that v is descending from u.

Note that the minimal element in a radial order is always the root.

Example 1.2.3. The archetypal example of radial order is the order in which splits occur through continuous time in a tree produced by a birth–death process. Note that this is specific to continuous time, since in discrete time there are always several nodes with the same generation (i.e., the same graph distance to the root).

Definition 1.2.4. We let $\mathscr{R}_n$ denote the set of all binary tree shapes with n tips endowed with a radial order. The elements of $\mathscr{R}_n$ are called *ranked tree shapes* (with n tips).

Also, we let $\mathscr{R}_n^{\ell}$ denote the set of ranked tree shapes with n tips labelled by $\{1, \ldots, n\}$. The elements of $\mathscr{R}_n^{\ell}$ are called *labelled, ranked tree shapes* (with n tips).

There is a canonical surjection $\ell : \mathcal{T}_n^\ell \to \mathcal{T}_n$ mapping each labelled tree shape to the same tree shape without labels. We use the same notation for the surjection $\ell : \mathcal{R}_n^\ell \to \mathcal{R}_n$.

Similarly, there is a canonical surjection $r : \mathcal{R}_n \to \mathcal{T}_n$ mapping each ranked tree shape to the same tree shape without radial order. We use the same notation for the surjection $r : \mathcal{R}_n^\ell \to \mathcal{T}_n^\ell$.

The mappings ℓ and r are sometimes called the *forgetful maps*, because they consist in "forgetting" the labels or the ranks, respectively. It is obvious that $\ell \circ r = r \circ \ell$, so that the following diagram commutes.

$$
\begin{array}{ccc}
\mathcal{R}_n^\ell & \xrightarrow{\;r\;} & \mathcal{T}_n^\ell \\
\downarrow{\scriptstyle \ell} & & \downarrow{\scriptstyle \ell} \\
\mathcal{R}_n & \xrightarrow{\;r\;} & \mathcal{T}_n
\end{array}
$$

We now give explicit expressions for the cardinal numbers of $\mathcal{T}_n^\ell$ and $\mathcal{R}_n^\ell$.

Proposition 1.2.5 (Murtagh (1984)). *For each $n \geq 2$, set $t_n := \#\mathcal{T}_n^\ell$ and $r_n := \#\mathcal{R}_n^\ell$. Then*

$$
t_n = (2n - 3)!! := (2n - 3)(2n - 5)\cdots(3)(1),
$$
$$
r_n = \frac{n!(n-1)!}{2^{n-1}}.
$$

Remark 1.2.6. No explicit expression is known for $\#\mathcal{T}_n$ (sometimes called Wedderburn–Etherington number) or $\#\mathcal{R}_n$.

Proof of Proposition 1.2.5. We reason by induction. Let $p_n : \mathcal{T}_{n+1}^\ell \to \mathcal{T}_n^\ell$ denote the projection which maps each τ with $n + 1$ labelled tips to the tree spanned by the tips of τ carrying labels in $\{1, \ldots, n\}$. Note that

$$
t_{n+1} = \sum_{\tau \in \mathcal{T}_n^\ell} \#p_n^{-1}(\{\tau\}).
$$

Now let us compute $\#p_n^{-1}(\{\tau\})$. It is immediate that each tree shape with n labelled tips $\tau \in \mathcal{T}_n^\ell$ has n external edges, $n - 2$ internal edges and a root edge. This gives $2n - 1$ distinct locations where to grow a new external edge carrying the label $n + 1$, which means that $\#p_n^{-1}(\{\tau\}) = 2n - 1$. In conclusion, $t_{n+1} = (2n - 1)t_n$ and the first result follows, since $t_2 = 1$.

For ranked tree shapes, we can similarly consider the projection (still denoted) $p_n : \mathcal{R}_{n+1}^\ell \to \mathcal{R}_n^\ell$ and use the similar equality

$$
r_{n+1} = \sum_{\tau \in \mathcal{R}_n^\ell} \#p_n^{-1}(\{\tau\}).
$$

Now let us compute $\#p_n^{-1}(\{\tau\})$. Each ranked tree shape with n labelled tips $\tau \in \mathcal{R}_n^\ell$ has $n-1$ ordered internal nodes. Growing a new edge with label $n+1$ requires to add a new internal node. There are n distinct locations where to insert it in the radial order, say between the $(k-1)$st internal node and the kth internal node for $1 \le k \le n$. For a given k, there are k distinct edges where to grow the new edge, which means $\#p_n^{-1}(\{\tau\}) = 1 + \cdots + n = n(n+1)/2$. In conclusion, $r_{n+1} = n(n+1)r_n/2$ and the second result follows, since $r_2 = 1$. $\square$

Exercise 1.2.7. Check that $t(x) := \sum_{n \ge 1} t_n \frac{x^n}{n!}$ is well defined for $|x| < \frac{1}{2}$. By combinatorial arguments, prove that for $n \ge 2$

$$t_n = \frac{1}{2} \sum_{i=1}^{n-1} \binom{n}{i} t_i t_{n-i}$$

and deduce that $t(x) = x + \frac{1}{2} t(x)^2$. Conclude that $t(x) = 1 - \sqrt{1 - 2x}$.

Exercise 1.2.8. A tree shape is said *oriented* when each internal vertex has a *left* and a *right* descending subtree. Each orientation gives a tree a unique plane embedding. Prove that the number of ranked oriented trees with n tips is $(n-1)!$ and that the number of ranked, oriented trees with n labelled tips is $n!(n-1)!$. Also if o denotes the map forgetting orientation, check that for any $\tau \in \mathcal{R}_n^\ell$, $o^{-1}(\{\tau\}) = 2^{n-1}$. This gives another proof of the formula $r_n = \frac{n!(n-1)!}{2^{n-1}}$.

1.2.2 *Counting rankings and labellings.*

Definition 1.2.9. For any $\tau \in \mathcal{T}_n$, an internal node u of τ is said *symmetric* if the two subtrees descending from u (i.e., the two connected components of $\tau \setminus \{u\}$ not containing the root) are identical. A particular case of symmetric node is when u subtends (i.e., is the most recent common ancestor of) a *cherry*, that is u only subtends (two) tips.

We denote by $s(\tau)$ the number of symmetric nodes of τ, and by $c(\tau)$ the number of cherries of τ.

Assume we were to extend the notion of symmetric node to labelled (resp. ranked) tree shapes, in the sense that the two descending subtrees of a symmetric node should have not only the same shape but also the same tip labels (resp. the same internal node ranks). Then in a labelled or ranked tree shape, only cherries would be symmetric. This explains the following convention.

Definition 1.2.10. For any $\tau \in \mathcal{T}_n^\ell \cup \mathcal{R}_n \cup \mathcal{R}_n^\ell$, we denote invariably by $s(\tau)$ or $c(\tau)$ the number of cherries of τ.

Exercise 1.2.11. Prove that for any unlabelled tree shape τ with n tips ($\tau \in \mathscr{T}_n \cup \mathscr{R}_n$), the number of distinct labellings of τ is

$$\#\ell^{-1}(\{\tau\}) = 2^{-s(\tau)}n!, \tag{1.1}$$

where we recall that if τ is ranked ($\tau \in \mathscr{R}_n$), $s(\tau)$ is the number of cherries of τ (see previous discussion).

Definition 1.2.12. For any tree shape τ (labelled or not, ranked or not), for any vertex $v \in \mathcal{V}(\tau)$, if τ' denotes the subtree descending from v, we denote invariably by $\lambda(v)$ or by $\lambda(\tau')$ the number of leaves subtended by v, which is also the number of leaves of τ'.

Note that $\lambda(v) = 1$ iff v is a leaf.

Equation (1.1) gives an explicit expression for the number of distinct labellings of a given tree shape τ. The next statement gives the number of distinct rankings of any given (either labelled or not labelled) tree shape.

Proposition 1.2.13 (Knuth (1997)). *For any unranked tree shape τ with n tips ($\tau \in \mathscr{T}_n \cup \mathscr{T}_n^\ell$), the number of distinct rankings of τ is*

$$\#r^{-1}(\{\tau\}) = 2^{c(\tau)-s(\tau)} \frac{(n-1)!}{\prod_{v\in\mathring{\mathcal{V}}(\tau)}(\lambda(v)-1)}, \tag{1.2}$$

where $\mathring{\mathcal{V}}(\tau)$ denotes the set of internal vertices of τ. Recall that when τ is labelled ($\tau \in \mathscr{T}_n^\ell$), $s(\tau) = c(\tau)$.

Proof. First, note that the total number of internal vertices of τ is $\#\mathring{\mathcal{V}}(\tau) = \lambda(\tau) - 1$.

Now assume that $\tau \in \mathscr{T}_n^\ell$ and let τ' and τ'' denote the two labelled subtrees descending from the root of τ (say, e.g., that τ' is the subtree containing the tip with label 1). Let k denote the number of internal vertices of τ'. Assuming that the k internal nodes of τ' are ordered and that the $n - k - 2$ internal nodes of τ'' are ordered, the number of ways of ordering the internal nodes of τ' and τ'' with respect to each other is $\binom{n-2}{k}$. In conclusion,

$$\#r^{-1}(\{\tau\}) = \binom{n-2}{k}\#r^{-1}(\{\tau'\})\#r^{-1}(\{\tau''\}),$$

which also reads

$$\frac{\#r^{-1}(\{\tau\})}{(n-1)!} = \frac{1}{n-1}\frac{\#r^{-1}(\{\tau'\})}{k!}\frac{\#r^{-1}(\{\tau''\})}{(n-2-k)!}.$$

An immediate induction yields

$$\frac{\#r^{-1}(\{\tau\})}{(n-1)!} = \prod_{v\in\mathring{\mathcal{V}}(\tau)}\frac{1}{\lambda(v)-1},$$

which is the expected result for labelled tree shapes.

Now assume that $\tau \in \mathscr{T}_n$. The following proof relies on the equality

$$\#\ell^{-1}\big(r^{-1}(\{\tau\})\big) = \#r^{-1}\big(\ell^{-1}(\{\tau\})\big),$$

due to the fact that ℓ and r commute.

First, let $\tilde{\tau} \in \mathscr{R}_n$ such that $r(\tilde{\tau}) = \tau$. Recall from equation (1.1) that the number of distinct labellings of $\tilde{\tau}$ is $2^{-s(\tilde{\tau})}n! = 2^{-c(\tau)}n!$, which yields

$$\#\ell^{-1}\big(r^{-1}(\{\tau\})\big) = \#r^{-1}(\{\tau\})2^{-c(\tau)}n!.$$

Second, let $\bar{\tau} \in \mathscr{T}_n^\ell$ such that $\ell(\bar{\tau}) = \tau$. From what precedes, we know that

$$\#r^{-1}(\{\bar{\tau}\}) = (n-1)! \prod_{v \in \mathring{\mathcal{V}}(\bar{\tau})} \frac{1}{\lambda(v)-1} = (n-1)! \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v)-1},$$

which yields

$$\#r^{-1}\big(\ell^{-1}(\{\tau\})\big) = \#\ell^{-1}(\{\tau\})(n-1)! \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v)-1}$$

$$= 2^{-s(\tau)}n!(n-1)! \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v)-1},$$

thanks again to equation (1.1). Equalling the expression for $\#\ell^{-1}(r^{-1}(\{\tau\}))$ and the expression for $\#r^{-1}(\ell^{-1}(\{\tau\}))$ provides the final result. □

1.3 Random tree shapes

1.3.1 *Uniform distributions.*

Definition 1.3.1. Let $\mathtt{Unif}_{\mathscr{T}_n^\ell}$ and $\mathtt{Unif}_{\mathscr{R}_n^\ell}$ denote the uniform distributions on $\mathscr{T}_n^\ell$ and $\mathscr{R}_n^\ell$, respectively.

We will adopt the following notation. The distributions $P_n^{\mathtt{pda}}$ and $P_n^{\mathtt{erm}}$ are the probabilities on $\mathscr{T}_n^\ell$ defined as

$$P_n^{\mathtt{pda}} := \mathtt{Unif}_{\mathscr{T}_n^\ell} \quad \text{and} \quad P_n^{\mathtt{erm}} = \mathtt{Unif}_{\mathscr{R}_n^\ell} \circ r^{-1}.$$

The distribution $P_n^{\mathtt{urt}}$ is the probability on $\mathscr{R}_n$ defined as

$$P_n^{\mathtt{urt}} := \mathtt{Unif}_{\mathscr{R}_n^\ell} \circ \ell^{-1}.$$

Note that $P_n^{\mathtt{urt}} \circ r^{-1} = P_n^{\mathtt{erm}} \circ \ell^{-1}$ is the push forward of $\mathtt{Unif}_{\mathscr{R}_n^\ell}$ by $\ell \circ r$.

Remark 1.3.2. The preceding denominations come from the terminology used in the phylogenetics literature (Aldous (1996, 2001), Blum and François (2006), Brown (1994), Lambert and Stadler (2013), Semple and Steel (2003)), where these three TLAs[1] have the following meanings.

---

[1]Three-Letter Acronym.

- PDA stands for "proportional to distinguishable arrangements",
- ERM stands for "equal rates Markov",
- URT stands for "uniform on ranked (labelled) trees".

Thanks to Proposition 1.2.5, and to equations (1.1) and (1.2), here are the probabilities of any given tree τ under various of the previously defined distributions.

For any $\tau \in \mathscr{T}_n^\ell$,

$$P_n^{\mathrm{pda}}(\tau) = \frac{1}{t_n} = \frac{1}{(2n-3)!!}$$

and

$$P_n^{\mathrm{erm}}(\tau) = \frac{\#r^{-1}(\{\tau\})}{r_n} = \frac{2^{n-1}}{n!} \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v)-1}.$$

For any $\tau \in \mathscr{R}_n$,

$$P_n^{\mathrm{urt}}(\tau) = \frac{\#\ell^{-1}(\{\tau\})}{r_n} = \frac{2^{n-1-c(\tau)}}{(n-1)!}.$$

For any $\tau \in \mathscr{T}_n$,

$$P_n^{\mathrm{pda}} \circ \ell^{-1}(\tau) = \frac{\#\ell^{-1}(\{\tau\})}{t_n} = \frac{n!2^{-s(\tau)}}{t_n} = \frac{2^{n-1-s(\tau)}}{c_{n-1}},$$

where $c_k := \frac{1}{k+1}\binom{2k}{k}$ is the kth Catalan number, and

$$P_n^{\mathrm{erm}} \circ \ell^{-1}(\tau) = \#\ell^{-1}(\{\tau\})\frac{2^{n-1}}{n!} \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v)-1}$$

$$= \frac{2^{n-1-s(\tau)}}{\prod_{v \in \mathring{\mathcal{V}}(\tau)}(\lambda(v)-1)}.$$

Exercise 1.3.3. We will see that the tree generated by a Yule (pure-birth) process stopped upon reaching n particles, has the law P_n^{urt} of the uniform labelled, ranked tree shape after ignoring labels. Check the computations shown for the Yule tree with n tips ($n = 4, 5, 6$) in Figures 1, 2 and 3.

1.3.2 *The law of your favourite random tree.*

The Bienaymé–Galton–Watson tree shape. Let $\mathcal{B}$ stand for the law of a binary (unlabelled) Galton–Watson tree with parameter $p \in (0, 1)$, where p is the probability of begetting two offspring, and $\mathcal{B}_n$ the law of the same tree conditioned to have n leaves. Set σ_n the probability of having n leaves under $\mathcal{B}$
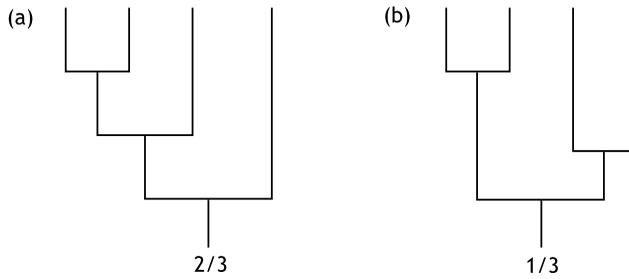
$$\sigma_n := \mathcal{B}(\lambda = n),$$

Figure 1 *Probabilities of trees shapes with n = 4 tips under P_n^{urt}. (a) The caterpillar tree with 4 tips has only one conforming ranked shape, with probability 2/3; (b) The symmetric tree with 4 tips also has one conforming ranked shape, with probability 1/3.*

Figure 2 *Probabilities of trees shapes with n = 5 tips under P_n^{urt}. (a) Is the caterpillar tree with 5 tips, which has only one conforming ranked shape, with probability 1/3; (b) Has one conforming ranked shape with probability 1/6; (c) Has 3 conforming ranked shapes, each with probability 1/6.*

so that for each $\tau \in \mathcal{T}_n$,

$$\mathcal{B}_n(\tau) = \frac{\mathcal{B}(\tau)}{\sigma_n}.$$

Now I will leave the reader convince herself (by drawing an example, say the two different binary trees with 4 leaves; or more rigorously, proving that the total number of plane orientations of an unlabelled, unranked tree shape τ is $2^{n-1-s(\tau)}$) that

$$\mathcal{B}(\tau) = 2^{n-1-s(\tau)} \left[\prod_{v \in \mathring{\mathcal{V}}(\tau)} p \right] \left[\prod_{u \text{ tip of } \tau} (1-p) \right]$$

$$= 2^{n-1-s(\tau)} p^{n-1} (1-p)^n.$$

(1.3)

Now thanks to equation (1.1), we know that

$$t_n = \sum_{\tau \in \mathcal{T}_n} \ell^{-1}(\{\tau\}) = \sum_{\tau \in \mathcal{T}_n} 2^{-s(\tau)} n!$$
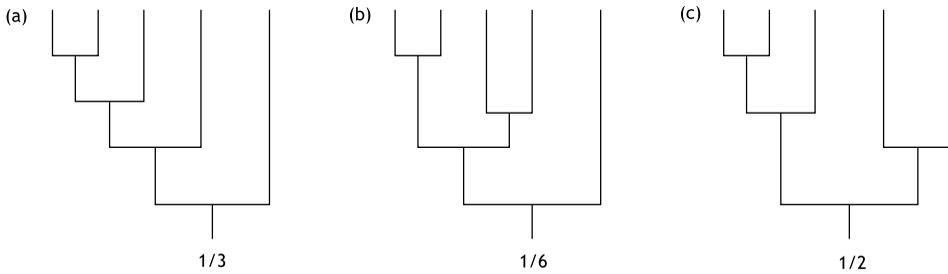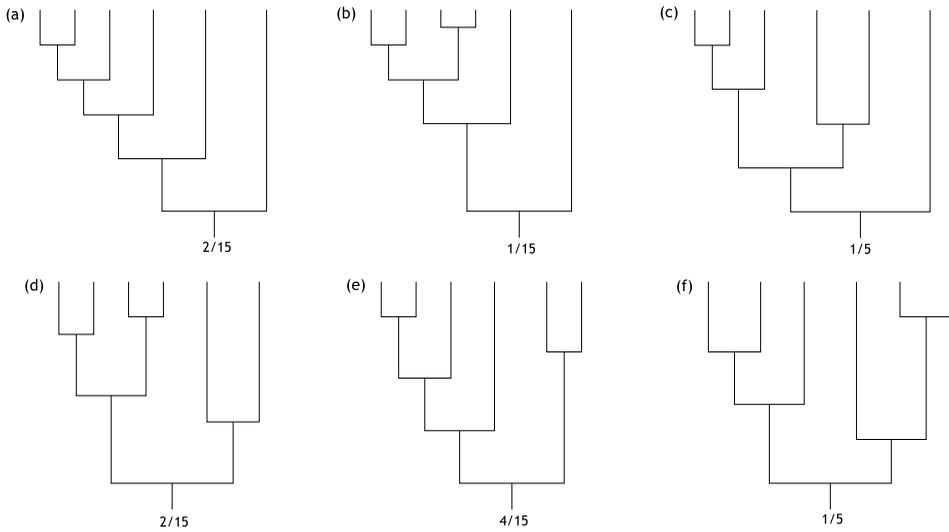
Figure 3 *Probabilities of trees shapes with $n = 6$ tips under $P_n^{\mathtt{urt}}$. (a) Is the caterpillar tree with 6 tips, which has only one conforming ranked shape, with probability 2/15; (b) Has one conforming ranked shape with probability 1/15; (c) Has 3 conforming ranked shapes, each with probability 1/15; (d) Has 4 conforming ranked shapes, each with probability 1/30; (e) Has 4 conforming ranked shapes, each with probability 1/15; (f) Has 3 conforming ranked shapes, each with probability 1/15.*

so that

$$\sigma_n = \sum_{\tau \in \mathscr{T}_n} 2^{n-1-s(\tau)} p^{n-1}(1-p)^n = \frac{2^{n-1} t_n}{n!} p^{n-1}(1-p)^n. \qquad (1.4)$$

In conclusion, for each $\tau \in \mathscr{T}_n$,

$$\mathcal{B}_n(\tau) = \frac{\mathcal{B}(\tau)}{\sigma_n} = \frac{n! 2^{-s(\tau)}}{t_n} = \frac{\#\ell^{-1}(\{\tau\})}{t_n} = P_n^{\mathrm{pda}} \circ \ell^{-1}(\tau),$$

which can be recorded in the following statement.

Proposition 1.3.4. *For each integer $n \geq 2$, $\mathcal{B}_n = P_n^{\mathrm{pda}} \circ \ell^{-1}$.*

If the reader is not convinced that (1.3) does hold, another proof will be given in the context of Markov branching models page 429.

Notice that $\mathcal{B}_n$ does not depend on p. Also note for the record that the probability of a given labelled tree shape under the Galton–Watson model with uniform labelling is $\frac{2^{n-1}}{n!} p^{n-1}(1-p)^n$ (where n is its number of tips).

The Yule tree shape. The reason why $\mathrm{Unif}_{\mathscr{R}_n^\ell} \circ r^{-1}$ is denoted P_n^{erm} is that it is the law of the (uniformly labelled) tree shape given by the genealogy of a population where all particles split independently and at the same rate b, called birth rate, into

two new particles (more details to come in Section 3). The process counting the size of the population is a Markov process jumping from k to $k+1$ at rate bk and is usually called pure birth process, or Yule (sometimes Yule–Furry) process (see Lambert (2008) for an introduction to stochastic models of population dynamics and genealogies).

More specifically, let $\mathcal{Y}_n$ denote the probability on $\mathcal{R}_n$ defined as the law of the *ranked* tree shape generated by a pure-birth process started at 1 and stopped upon reaching n, where the radial order is the chronological order of node splits. It is clear that this probability does not depend on the birth rate b.

Proposition 1.3.5. *For each integer $n \geq 2$, $\mathcal{Y}_n = P_n^{\mathrm{urt}}$. In particular, the law $\mathcal{Y}_n \circ r^{-1}$ of a Yule tree whose node ranks are ignored is $P_n^{\mathrm{erm}} \circ \ell^{-1}$.*

Proof. Let us prove the proposition by induction on n (the proposition obviously holds for $n = 2$). Let $n \geq 2$ and assume the proposition holds for this n. Now let τ be a ranked tree shape with $n + 1$ tips. Let v denote the maximal interior node in the radial order and let v' denote the maximal interior node in the genealogical order of the path from the root to v. Define $\hat{\tau}$ as the ranked tree shape with n tips obtained from τ by collapsing the cherry subtended by v into a single terminal edge. By definition of the Yule process, if v' subtends a cherry in $\hat{\tau}$, then splitting any of the two tips of this cherry of $\hat{\tau}$ into a new cherry yields τ. Otherwise, only one tip of $\hat{\tau}$ can be split to yield τ. This can be expressed as

$$\mathcal{Y}_{n+1}(\tau) = \frac{2^{1_{\lambda(v')=3}}}{n} \mathcal{Y}_n(\hat{\tau}).$$

By the induction hypothesis,

$$\mathcal{Y}_{n+1}(\hat{\tau}) = P_n^{\mathrm{urt}}(\hat{\tau}) = \frac{2^{n-1-c(\hat{\tau})}}{(n-1)!},$$

so that

$$\mathcal{Y}_{n+1}(\tau) = \frac{2^{1_{\lambda(v')=3}} 2^{n-1-c(\hat{\tau})}}{n!}.$$

But now check that $c(\tau) = c(\hat{\tau}) + 1_{\lambda(v')\neq 3}$, which yields the result.   □

The Kingman tree shape.   Conversely, consider a population where each pair of particles independently merges at the same rate c, called coalescence rate (or competition rate) into one single new particle. The process counting the size of the population is a Markov pure-death process jumping from k to $k-1$ at rate $ck(k-1)/2$.

More specifically, let $\mathcal{K}_n$ denote the probability on $\mathcal{R}_n^{\ell}$ of the labelled, ranked tree shape generated by this process started from n labelled particles and naturally stopped when it reaches 1, usually called Kingman n-coalescent tree (Kingman (1982)). Obviously, $\mathcal{K}_n$ does not depend on c.

Proposition 1.3.6. *For each integer $n \geq 2$, $\mathcal{K}_n = \mathtt{Unif}_{\mathscr{R}_n^\ell}$. In particular, the law $\mathcal{K}_n \circ \ell^{-1}$ of a Kingman tree whose labels are ignored is the law $\mathtt{Unif}_{\mathscr{R}_n^\ell} \circ \ell^{-1} = \mathcal{Y}_n$ of a Yule tree.*

Proof. Let us prove the proposition by induction on n (the proposition obviously holds for $n = 2$). Let $n \geq 2$ and assume the proposition holds for this n. Let τ be a ranked tree shape with $n + 1$ labelled tips. Let i and j be the labels of the cherry subtended by the maximal interior node in the radial order, and as in the previous proof, let $\hat{\tau}$ be the ranked tree shape with n labelled tips obtained from τ by collapsing this cherry into one single terminal edge and relabelling the new tip. By induction $\mathcal{K}_n(\hat{\tau}) = 1/r_n$, so that

$$\mathcal{K}_{n+1}(\tau) = \frac{2}{n(n+1)}\mathcal{K}_n(\hat{\tau}) = \frac{2}{n(n+1)}\frac{1}{r_n} = \frac{1}{r_{n+1}},$$

which terminates the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Exercise 1.3.7. Explain why the probability distribution P_n^{erm} puts more weight on balanced trees than P_n^{pda}. If a_n denotes the most imbalanced tree with n tips, known as the caterpillar tree (see panel (a) in each of Figures 1–3), prove that

$$P_n^{\mathrm{pda}} \circ \ell^{-1}(a_n) = \frac{2^{n-2}}{c_{n-1}} \quad \text{and} \quad P_n^{\mathrm{erm}} \circ \ell^{-1}(a_n) = \frac{2^{n-2}}{(n-1)!}.$$

One can easily see that the probability of the caterpillar tree with n tips, under either of these two distributions, vanishes as $n \to \infty$. Nevertheless, the probability that one of the two subtrees incident to the root subtends exactly one tip (i.e., a long external edge) converges to $\frac{1}{4}$ under P_n^{pda}, as will be seen in Exercise 1.4.5 (see also Table 1).

It is generally observed that none of the random tree shapes generated by P_n^{erm} or by P_n^{pda} statistically give a good fit to empirical species trees. More specifically, real phylogenies are less balanced than random trees under P_n^{erm} but more balanced than random trees under P_n^{pda}. In the following section, we introduce more general models of random tree shapes, as well as a one-parameter family of tree shape distributions interpolating in particular P_n^{erm} and P_n^{pda}.

Table 1 *Numerical values of the probability of the caterpillar tree under P_n^{erm} and P_n^{pda}. Observe the larger weight put on this tree under P_n^{pda} than under P_n^{erm}*

n	3	4	5	6	7
$P_n^{\mathrm{pda}} \circ \ell^{-1}(a_n)$	1	$\frac{4}{5}$	$\frac{4}{7}$	$\frac{8}{21}$	$\frac{8}{33}$
$P_n^{\mathrm{erm}} \circ \ell^{-1}(a_n)$	1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{15}$	$\frac{2}{45}$

1.4 Markov branching models

1.4.1 *Definitions and interval splitting.*   In what follows, for each $\tau \in \mathscr{T}_n \cup \mathscr{T}_n^{\ell}$, we will write $\tau = \tau' \oplus \tau''$ to denote the fact that τ' and τ'' are the two subtrees descending from the root of τ. It will be convenient to assume that τ' is chosen uniformly at random among the two possible subtrees (except in the case when τ is not labelled and the root is symmetric, since then $\tau' = \tau''$). Note that it would also have been possible to select τ' deterministically as the subtree containing the tip with label 1 say, or the subtree with the smaller number of tips.

Definition 1.4.1 (Aldous (1996)). A family of distributions (P_n) on $\mathscr{T}_n^{\ell}$ is a Markov branching model if there is a family of laws (q_n) on $\{1, \ldots, n-1\}$ such that $q_n(i) = q_n(n-i)$ for all i, and for any $\tau \in \mathscr{T}_n^{\ell}$,

$$P_n(\tau) = \frac{2q_n(i)}{\binom{n}{i}} P_i(\tau') P_{n-i}(\tau''),$$

where $\tau = \tau' \oplus \tau''$ and $i = \lambda(\tau')$ (when $n = 3$, and $i = 2$, $P_2(\tau') = 1 = P_1(\tau'')$).

Note the abuse of notation, since τ' is not in general labelled by $\{1, \ldots, i\}$. It is implicit that P_n is invariant by permutations of labels (P_n is said exchangeable or equivariant) and defined independently from the chosen label set.

Remark 1.4.2. If τ' was chosen to be the subtree containing the tip with label 1 (instead of being chosen at random), the previous display would become

$$P_n(\tau) = \frac{q_n(i)}{\binom{n-1}{i-1}} P_i(\tau') P_{n-i}(\tau'').$$

A third possibility would be to choose τ' as the subtree with the smaller number of tips. The details are left to the reader (see also Aldous (2001)).

To generate a labelled tree shape with law P_n, proceed recursively.

1. Draw a random variable $K_n \in \{1, \ldots, n-1\}$ with law q_n.
2. Conditional on $K_n = i$, select a subset I of $\{1, \ldots, n\}$ with cardinality i, uniformly at random.
3. Create two edges joining the root of τ to τ' and τ'' respectively, where τ' and τ'' are two independent tree shapes labelled respectively by I and its complement, with respective laws P_i and P_{n-i}.

Exercise 1.4.3. If $\tau \in \mathscr{T}_n$, recall that $\tau = \tau' \oplus \tau''$ still makes sense, and check that $s(\tau) = s(\tau') + s(\tau'')$ except when $\tau' = \tau''$, where $s(\tau) = 1 + s(\tau') + s(\tau'')$. Deduce that

$$P_n \circ \ell^{-1}(\tau) = 2^{\mathbf{1}_{\tau' \neq \tau''}} q_n(i) P_i \circ \ell^{-1}(\tau') P_{n-i} \circ \ell^{-1}(\tau''),$$

where again $i = \lambda(\tau')$. Explain why $(P_n \circ \ell^{-1})$ can also be seen as a Markov branching model (on unlabelled tree shapes).

Now we introduce a specific way of designing Markov branching models. Let f be a non-negative function on $(0, 1)$ such that $f(x) = f(1 - x)$ and $\int_0^1 x(1 - x) f(x) \, dx < \infty$. Set

$$\alpha_n := \sum_{i=1}^{n-1} \binom{n}{i} \int_0^1 x^i (1 - x)^{n-i} f(x) \, dx = \int_0^1 \left(1 - x^n - (1 - x)^n\right) f(x) \, dx,$$

which is finite by assumption. Then we define

$$q_n^f(i) := \alpha_n^{-1} \binom{n}{i} \int_0^1 x^i (1 - x)^{n-i} f(x) \, dx,$$

as well as (P_n^f) the associated Markov branching model.

In the case when f is integrable, one can generate a labelled tree shape with law P_n^f by throwing uniformly n points in $(0, 1)$ and performing a recursive interval splitting procedure, down until each interval contains at most two of the initial points. Let $U_1, \ldots, U_n$ be i.i.d. random variables uniform in $(0, 1)$. Assume that $\int_0^1 f(x) \, dx = 1$.

1. Draw an independent r.v. X with density f in $(0, 1)$.
2. Let I be the subset of $\{1, \ldots, n\}$ defined by: $i \in I \Leftrightarrow U_i < X$. Then conditional on $X = x$, proceed as in the general case, by putting the labels of I in τ' and those of its complement in τ''.
3. Apply recursively the same procedure to the intervals $(0, x)$ and $(x, 1)$ independently.

It is intuitive from this description that the more f puts weight close to the boundaries of $(0, 1)$, the more the associated random tree shape is imbalanced.

1.4.2 *ERM, PDA and Aldous' β-splitting model.*

Theorem 1.4.4 (Harding (1971), Slowinski (1990), Brown (1994)). *Both (P_n^{erm}) and (P_n^{pda}) are Markov branching models with*

$$q_n^{\mathrm{pda}}(i) = \frac{1}{2} \binom{n}{i} \frac{t_i t_{n-i}}{t_n} \quad \text{and} \quad q_n^{\mathrm{erm}}(i) = \frac{1}{n-1}.$$

Proof. We know that for any $\tau \in \mathscr{T}_n^\ell$ written $\tau = \tau' \oplus \tau''$, with $i = \lambda(\tau')$, $P_n^{\mathrm{pda}}(\tau) = \frac{1}{t_n}$, $P_i^{\mathrm{pda}}(\tau') = \frac{1}{t_i}$ and $P_{n-i}^{\mathrm{pda}}(\tau'') = \frac{1}{t_{n-i}}$, so that

$$P_n^{\mathrm{pda}}(\tau) = \frac{1}{t_n} = \frac{t_i t_{n-i}}{t_n} P_i^{\mathrm{pda}}(\tau') P_{n-i}^{\mathrm{pda}}(\tau''),$$

Table 2 *The limiting value, as $n \to \infty$, of a basal split i vs. $n - i$, under P_n^{pda}*

i	1	2	3	4
$\lim_n q_n^{\text{pda}}(i)$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{5}{128}$

which agrees with the characterization of Markov branching models choosing q_n^{pda} as in the theorem (see Table 2).

We now focus on the distribution of $\lambda(\tau') =: J_n$ under P_n^{erm}. Recall that $P_n^{\text{erm}} \circ \ell^{-1}$ is the law of the unranked and unlabelled Yule tree. Consider two particles, a red one and a black one. Give these two particles independent Yule descendances and stop the process when the total number of particles equals n. By the strong Markov property of the Yule process (applied at the time it reaches 2), J_n has the same law as the number of red (say) particles when the process stops. By the strong Markov property applied at the first time when there are $n - 1$ particles, it is easy to see that

$$\mathbb{P}(J_n = i) = \frac{i - 1}{n - 1}\mathbb{P}(J_{n-1} = i - 1) + \frac{n - 1 - i}{n - 1}\mathbb{P}(J_{n-1} = i),$$

An immediate induction shows that $\mathbb{P}(J_n = i) = \frac{1}{n-1}$.

Now recall that P_n^{erm} is also the law of the unranked Kingman coalescent tree. Fixing a subset I of $\{1, \ldots, n\}$ with cardinality i and conditioning the Kingman coalescent tree to have on the one hand all lineages initially labelled by I and on the other hand all lineages initially labelled by the complement of I, coalesce within each other before coalescing between each other, yields two independent Kingman coalescent trees (forbidding a restricted subset of the pairwise exponential clocks to ring does not alter the independence of the other clocks). Now these two unranked, labelled subtrees follow respectively, P_i^{erm} and P_{n-i}^{erm}, which yields the result. $\square$

Exercise 1.4.5. Show that $\lim_{n \to \infty} q_n^{\text{pda}}(i) = c_{i-1}4^{-i}$, which 'contrasts sharply with the flat distribution q_n^{erm}' (Semple and Steel (2003)).

Back to the Galton–Watson tree. Here, we want to give a proof of Proposition 1.3.4 via Markov branching models, without using (1.3). Let $\tau \in \mathscr{T}_n$ and write $\tau = \tau' \oplus \tau''$, $i = \lambda(\tau')$. Then by the branching property,

$$\mathcal{B}(\tau) = p 2^{1_{\tau' \neq \tau''}} \mathcal{B}(\tau')\mathcal{B}(\tau''),$$

which becomes

$$\mathcal{B}_n(\tau) = p 2^{1_{\tau' \neq \tau''}} \frac{\sigma_i \sigma_{n-i}}{\sigma_n} \mathcal{B}_i(\tau')\mathcal{B}_{n-i}(\tau'').$$

Now recalling Exercise 1.4.3, this shows that $\mathcal{B}_n = P_n \circ \ell^{-1}$, where (P_n) is the Markov branching model associated with

$$q_n(i) = p \frac{\sigma_i \sigma_{n-i}}{\sigma_n}.$$

Then it remains to show that

$$\sigma_n = \frac{2^{n-1} t_n}{n!} p^{n-1} (1-p)^n, \tag{1.5}$$

for this will ensure that

$$q_n(i) = \frac{1}{2} \binom{n}{i} \frac{t_i t_{n-i}}{t_n},$$

which indeed is the splitting probability of the PDA model. Actually, (1.5) was already obtained as equation (1.4), but this equation was derived from (1.3), so we have to prove (1.5) by other means. This can actually be done in multiple ways, using for example the Lukasiewicz path associated to the tree and Dwass identity (see, e.g., Pitman (2006)).

Aldous' β-splitting model.   The β-splitting model of Aldous (1996, 2001) is a one-parameter family of interval splitting branching models. Specifically, for any $\beta \in (-2, +\infty)$, consider

$$f_\beta(x) = x^\beta (1-x)^\beta, \qquad x \in (0, 1).$$

The law $q_n^{f_\beta}$ associated with f_β will be denoted q_n^β.

Exercise 1.4.6. Prove that

$$q_n^\beta(i) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + n - i + 1)}{\Gamma(i+1)\Gamma(n - i + 1)}, \tag{1.6}$$

where Γ is the usual Gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$, for $x > 0$, and

$$a_n(\beta) := \frac{\Gamma(2\beta + n + 2)}{\Gamma(n+1)} \int_0^1 (1 - x^n - (1-x)^n) x^\beta (1-x)^\beta\, dx.$$

This family has some interesting special cases. As noticed earlier, the balance of the tree increases with β. As $\beta \to \infty$, the intervals are split deterministically in their middle, while as $\beta \to -2$, the splitting procedure converges to pure erosion, that is P_n^β puts weight converging to 1 on the caterpillar tree. The three other cases of interest are $\beta = 0$, $\beta = -3/2$ and $\beta = -1$.

- $\beta = 0$. Thanks to equation (1.6), $q_n^0(i)$ does not depend on i, which implies $q_n^0(i) = \frac{1}{n-1} = q_n^{\mathrm{erm}}(i)$, so that

$$\beta = 0 \quad \Longrightarrow \quad P_n^\beta = P_n^{\mathrm{erm}}.$$

- $\beta = -3/2$. We are going to prove that

$$\beta = -3/2 \quad \implies \quad P_n^\beta = P_n^{\text{pda}}.$$

First notice that we can write

$$t_n = \frac{(2n-3)!}{2^{n-2}(n-2)!} = \frac{\Gamma(2n-2)}{2^{n-2}\Gamma(n-1)} = 2^{n-1}\frac{\Gamma(n-\frac{1}{2})}{\Gamma(\frac{1}{2})},$$

where we used the identities $\Gamma(x)\Gamma(x+\frac{1}{2}) = 2^{1-2x}\sqrt{\pi}\Gamma(2x)$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Now tedious calculations (note the missing '4' in Aldous (1996), equation following equation 5) show that

$$a_n\left(-\frac{3}{2}\right) = \frac{4\Gamma(n-\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(n+1)},$$

so that

$$q_n^{-\frac{3}{2}}(i) = \frac{\Gamma(n+1)}{4\Gamma(n-\frac{1}{2})\Gamma(\frac{1}{2})} \frac{\Gamma(i-\frac{1}{2})\Gamma(n-i-\frac{1}{2})}{\Gamma(i+1)\Gamma(n-i+1)}$$

$$= \frac{1}{2}\frac{\Gamma(n+1)}{\Gamma(i+1)\Gamma(n-i+1)}\frac{t_i t_{n-i}}{t_n} = q_n^{\text{pda}}(i).$$

- $\beta = -1$. This model is sometimes called AB model for "Aldous branching". First, check thanks to (1.6) that

$$q_n^{-1}(i) = \frac{1}{a_n(-1)}\frac{\Gamma(i)\Gamma(n-i)}{\Gamma(i+1)\Gamma(n-i+1)} = \frac{1}{a_n(-1)}\frac{1}{i(n-i)}.$$

Then

$$a_n(-1) = \sum_{i=1}^{n-1}\frac{1}{i(n-i)} = \frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{1}{i} + \frac{1}{n-i}\right) = \frac{2h_{n-1}}{n},$$

with the usual notation h_n for the harmonic series

$$h_n := 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

In the end, we get

$$q_n^{-1}(i) = \frac{n}{2h_{n-1}}\frac{1}{i(n-i)}.$$

The remarkable feature of the β-splitting family is that it interpolates between maximally imbalanced (caterpillar) trees and random (maximally) balanced trees passing through P_n^{pda} and P_n^{erm}. Table 3 is taken from Aldous (2001), Table 3, and provides the median size of the smaller daughter clade at the basal split under P_n^β.

Table 3 *Median value of the smaller clade size at the basal split as $n \to \infty$, under P_n^β (taken from Aldous (2001))*

β	-2	$-\frac{3}{2}$	-1	0	∞
Smaller clade size	1	1.5	$\sqrt{n}$	$\frac{n}{4}$	$\frac{n}{2}$

Remark 1.4.7. The reason why the name of Aldous is tied to the special case $\beta = -1$ is due to the empirical observation that the trees generated by P_n^{-1} give the best fit to real phylogenies. In Aldous (2001), this was shown by a visual fit to a linear dependence with slope $1/2$ in the log–log scale of the size of the median split vs. n (see Table 3). In Blum and François (2006), the MLE of β for species trees is remarkably centered around -1. The biological reason for this pattern is still very much debated (Hagen et al. (2015), Manceau, Lambert and Morlon (2015)).

For the record, we finally provide a closed-form expression for the probability of a given labelled tree shape $\tau \in \mathcal{T}_n^\ell$ under P_n^β.

Proposition 1.4.8. *For any $\beta > -2$ and $\tau \in \mathcal{T}_n^\ell$,*

$$P_n^\beta(\tau) = \frac{\Gamma(\beta+2)^n 2^{n-1}}{\Gamma(\beta+n+1)} \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{\Gamma(\beta+\lambda(v)+1)}{\Gamma(\lambda(v)+1)a_{\lambda(v)}(\beta)}. \tag{1.7}$$

Proof. Thanks to Definition 1.4.1 and equation (1.6), writing $\tau = \tau' \oplus \tau''$ and $i = \lambda(\tau')$, we get by an immediate induction

$$P_n^\beta(\tau) = \frac{2q_n(i)}{\binom{n}{i}} P_i^\beta(\tau') P_{n-i}^\beta(\tau'')$$

$$= \frac{2}{a_n(\beta)n!}\Gamma(\beta+\lambda(\tau')+1)\Gamma(\beta+\lambda(\tau'')+1)P_i^\beta(\tau')P_{n-i}^\beta(\tau'')$$

$$= \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{2}{a_{\lambda(v)}(\beta)\lambda(v)!}\Gamma(\beta+\lambda(v_1)+1)\Gamma(\beta+\lambda(v_2)+1),$$

where for each internal node v, we have denoted by v_1 and v_2 its two offspring vertices. Therefore,

$$P_n^\beta(\tau) = \Gamma(\beta+n+1)^{-1}\left[\prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{2\Gamma(\beta+\lambda(v)+1)}{a_{\lambda(v)}(\beta)\lambda(v)!}\right]$$

$$\times \left[\prod_{u \text{ tip of } \tau} \Gamma(\beta+\lambda(u)+1)\right],$$

which yields (1.7). $\square$

Exercise 1.4.9. By giving the value 0 or $-3/2$ to β, recover from (1.7) the explicit expressions

$$P_n^{\mathrm{erm}}(\tau) = \frac{2^{n-1}}{n!} \prod_{v \in \mathring{\mathcal{V}}(\tau)} \frac{1}{\lambda(v) - 1} \quad \text{and} \quad P_n^{\mathrm{pda}}(\tau) = \frac{1}{t_n}.$$

For $\beta = 0$ you will need to remember that $a_n(0) = n - 1$, and for $\beta = -3/2$ that

$$t_n = 2^{n-1} \frac{\Gamma(n - \frac{1}{2})}{\Gamma(\frac{1}{2})} \quad \text{and} \quad a_n\left(-\frac{3}{2}\right) = \frac{4\Gamma(n - \frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(n + 1)}.$$

1.4.3 *Sampling consistency.* Following Aldous (1996), we will say that the Markov branching model (P_n) is *sampling consistent* if for each $n \geq 2$, the random tree shape $\hat{\tau}$ obtained from the tree τ with law P_{n+1}, after removing its edge subtending the label $n + 1$, has law P_n.

Exercise 1.4.10. Prove that a Markov branching model where $q_4(2, 2) = 1$ cannot be sampling consistent.

It is obvious that the interval splitting branching models (P_n^f) are sampling consistent. The converse statement is given in the following theorem.

Theorem 1.4.11. *Let (P_n) be a sampling consistent Markov branching model. Then there is a measure μ on $[0, 1]$ invariant by $x \mapsto 1 - x$ such that $\int_{(0,1)} x(1 - x)\mu(dx) < \infty$ and*

$$q_n(i) = \alpha_n^{-1}\left\{\binom{n}{i} \int_{(0,1)} \mu(dx)x^i(1 - x)^{n-i} \right.$$

$$\left. + n\mu(\{0\})1_{i=1} + n\mu(\{1\})1_{i=n-1}\right\}, \tag{1.8}$$

where

$$\alpha_n = \int_{(0,1)} \mu(dx)\left(1 - x^n - (1 - x)^n\right) + n\mu(\{0, 1\}).$$

Remark 1.4.12. If μ has a density f w.r.t. Lebesgue measure, then we are left with the interval splitting models of Aldous. The terms due to atoms at 0 and 1 correspond to single labels being taken away from the rest, a phenomenon called *erosion*. The theorem states that the only Markov branching models that are sampling-consistent combine interval splitting with erosion. In the case of pure erosion (i.e., when μ charges only $\{0, 1\}$), all trees are caterpillar trees a.s.

Remark 1.4.13. A more general version of the previous statement (i.e., not restricted to binary trees) is shown in Haas et al. (2008) by identifying the splitting

rules with the transitions of a general fragmentation process (Bertoin (2006)). This also allows the authors to study scaling limits of these random tree shapes. In particular, when $\beta \in (-2, -1)$, the trees with n tips generated by the β-splitting branching model, converge as $n \to \infty$ when their edges are given properly scaled lengths, to some closed set called real tree (weakly in the Gromov–Hausdorff topology, see next section).

Proof of Theorem 1.4.11. Assume that (P_n) form a Markov branching model, by definition exchangeable, and by assumption sampling consistent. Let $1 \le i \le n \le m$. Consider the tree generated by P_m and let $\sigma_{n,m}$ be the most recent common ancestor (mrca, i.e., the ancestor with maximal distance to the root) of all its tips carrying a label in $\{1, \ldots, n\}$, let $\tau_{n,m}$ denote the descending subtree of $\sigma_{n,m}$, and write $\tau_{n,m} = \tau'_{n,m} \oplus \tau''_{n,m}$ (each subtree being equally likely to be chosen as $\tau'_{n,m}$). Now let $J_{n,m}$ (resp. $J'_{n,m}$, $J''_{n,m}$) denote the set of labels carried by the tips of $\tau_{n,m}$ (resp. $\tau'_{n,m}$, $\tau''_{n,m}$). By construction, $\{1, \ldots, n\}$ is entirely contained in $J_{n,m}$, and intersects both $J'_{n,m}$ and $J''_{n,m}$, which form a partition of $J_{n,m}$.

Next, by sampling consistency, the triple $(J_{n,m+1}, J'_{n,m+1}, J''_{n,m+1})$ restricted to $\{1, \ldots, m\}$ has the same law as $(J_{n,m}, J'_{n,m}, J''_{n,m})$. By Kolmogorov's extension theorem, all these triples can be coupled on a same probability space, that is, there exists a random triple (J_n, J'_n, J''_n) of random subsets of $\mathbb{N}$, such that J_n contains $\{1, \ldots, n\}$, (J'_n, J''_n) form a partition of J_n, both intersect $\{1, \ldots, n\}$, and the restriction of (J_n, J'_n, J''_n) to $\{1, \ldots, m\}$ has the same law as $(J_{n,m}, J'_{n,m}, J''_{n,m})$.

From here on, we denote $\{1, \ldots, n\}$ by $[n]$. We define I_n as the set $(J_n \setminus [n]) - n$, that is

$$I_n := s_n \big(J_n \cap \{n+1, n+2, \ldots\} \big),$$

where $s_n(x) = x - n$. We define similarly I'_n and I''_n. Now the triple (I_n, I'_n, I''_n) is exchangeable, so by de Finetti's theorem, they all have an asymptotic frequency, say Y_n, Y'_n and Y''_n, respectively. Note that conditional on $Y_n = y$, the random variables $(\mathbb{1}_{\{k \in I_n\}})_k$ are i.i.d. Bernoulli random variables with parameter y. In particular, I_n is empty iff $Y_n = 0$ and if $Y_n \ne 0$, I_n is infinite with positive asymptotic frequency Y_n.

Let us prove that $\mathbb{P}(I_n = \varnothing) = 0$ so that $Y_n > 0$ a.s. Set $p_{n,m} := \mathbb{P}(J_{n,m} = [n])$. In particular, $p_{2,m}$ is the probability that in the tree T_m, the tips labelled 1 and 2 form a cherry. Then by exchangeability,

$$1 \ge \mathbb{P}(\text{ the tip labelled 1 belongs to a cherry in } T_m) = (m-1) p_{2,m},$$

so that $p_{2,m}$ vanishes as $m \to \infty$. This shows that

$$\mathbb{P}(I_2 = \varnothing) = \mathbb{P}\big(J_2 = [2]\big) = \lim_{m \to \infty} p_{2,m} = 0.$$

Following the preceding discussion, I_2 is a.s. infinite. Now it can be seen that $\tau_{2,m}$ is a subtree of $\tau_{n,m}$, so that $J_{2,m} \subseteq J_{n,m}$. As a consequence, $\#J_2$ is stochastically

smaller than $\#J_n$, which shows that J_n (and hence I_n) is infinite a.s. We have proved that $Y_n > 0$ a.s.

Now let X'_n (resp. X''_n) be the asymptotic fraction of I'_n (resp. I''_n) in I_n, that is

$$X'_n := \lim_{k \to \infty} \frac{\#I'_n \cap [k]}{\#I_n \cap [k]} = \frac{Y'_n}{Y_n} \quad \text{and} \quad X''_n := \lim_{k \to \infty} \frac{\#I''_n \cap [k]}{\#I_n \cap [k]} = \frac{Y''_n}{Y_n}.$$

Note that $X'_n + X''_n = 1$. Since X'_n and X''_n have the same law, we can record that X'_n and $1 - X'_n$ have the same law. The Markov property of the branching model and the exchangeability imply that X'_n is independent of I_n and in particular of Y_n. In addition, conditional on J_n and X'_n, the indicator variables $(\mathbb{1}_{\{a_i \in I'_n\}})$, where a_i is the ith element of I_n, are independent copies of a Bernoulli r.v. with success parameter X'_n.

From now on, we write $K'_n := J'_n \cap [n]$ and $K''_n = J''_n \cap [n]$, so that $K'_n \cup K''_n = [n]$. Let A be a fixed subset of $[n]$ such that $1 \in A$ but $2 \notin A$. Now observe that on the event $K'_n = A$, the triple (J_2, J'_2, J''_2) is either equal to (J_n, J'_n, J''_n) or to (J_n, J''_n, J'_n) with probabilities equal to $1/2$. So for any $x \in (0, 1)$, writing A^c for $[n] \setminus A$,

$$\mathbb{P}(X'_n \in dx, K'_n = A) = \frac{1}{2}\mathbb{P}(X'_2 \in dx, J'_2 \cap [n] = A, J''_2 \cap [n] = A^c).$$

Now recall that J_2 is independent of X'_2 and conditional on J_2 and $X'_2 = x$, the indicator variables $(\mathbb{1}_{\{a_i \in I'_2\}})$, where a_i is the ith element of J_2, are independent copies of a Bernoulli r.v. with success parameter x. So we get

$$\begin{aligned}
\mathbb{P}(X'_2 &\in dx, J'_2 \cap [n] = A, J''_2 \cap [n] = A^c) \\
&= \mathbb{P}([n] \subset J_2)\mathbb{P}(X'_2 \in dx)\mathbb{P}(J'_2 \cap [n] = A, J''_2 \cap [n] = A^c | [n] \subset J_2, X'_2 = x) \\
&= \mathbb{P}([n] \subset J_2)\mathbb{P}(X'_2 \in dx)x^{i-1}(1 - x)^{n-i-1},
\end{aligned}$$

where we let i denote the cardinality of A. We can rewrite the next-before-last equality as

$$\mathbb{P}(X'_n \in dx, K'_n = A) = \frac{1}{2}\mathbb{P}([n] \subset J_2)\mathbb{P}(X'_2 \in dx)x^{i-1}(1 - x)^{n-i-1}.$$

Summing all these equalities over all possible A's with cardinality $1 \le i \le n - 1$, gives

$$\begin{aligned}
\mathbb{P}(X'_n &\in dx, \#K'_n = i, 1 \in K'_n, 2 \notin K'_n) \\
&= \frac{1}{2}\binom{n-2}{i-1}\mathbb{P}([n] \subset J_2)\mathbb{P}(X'_2 \in dx)x^{i-1}(1 - x)^{n-i-1}.
\end{aligned}$$

By exchangeability again, the left-hand side equals

$$\mathbb{P}(X'_n \in dx, \#K'_n = i, 1 \in K'_n, 2 \notin K'_n) = \mathbb{P}(X'_n \in dx, \#K'_n = i)\frac{i(n - i)}{n(n - 1)},$$

so that

$$\mathbb{P}(X_n' \in dx, \#K_n' = i) = \binom{n}{i}\mathbb{P}([n] \subset J_2)\frac{\mathbb{P}(X_2' \in dx)}{2x(1-x)}x^i(1-x)^{n-i}.$$

Now let us treat the case when $X_n' = 0$, that is I_n' is empty and J_n' is reduced to a singleton. By exactly the same reasoning as above,

$$\mathbb{P}(X_n' = 0, J_n' = \{1\}) = \frac{1}{2}\mathbb{P}([n] \subset J_2)\mathbb{P}(X_2' = 0),$$

so again by exchangeability

$$\mathbb{P}(X_n' = 0, \#K_n' = 1) = \frac{n}{2}\mathbb{P}([n] \subset J_2)\mathbb{P}(X_2' = 0).$$

Reasoning symmetrically with X_n'', we finally get for all $x \in [0, 1]$ and $i \in [n-1]$,

$$\mathbb{P}(X_n' \in dx, \#K_n' = i)$$

$$= \mathbb{P}([n] \subset J_2)\binom{n}{i}\left(x^i(1-x)^{n-i}1_{x \in (0,1)} + 1_{(x,i)=(0,1) \text{ or } (1,n-1)}\right)\mu(dx),$$

where μ is the positive measure on $[0, 1]$ defined by

$$\mu(dx) := \frac{1}{2x(1-x)}\mathbb{P}(X_2' \in dx)1_{x \in (0,1)}$$

$$+ \frac{1}{2}\mathbb{P}(X_2' = 0)\delta_0(dx) + \frac{1}{2}\mathbb{P}(X_2' = 1)\delta_1(dx).$$

Summing on $i \in [n-1]$ yields

$$\mathbb{P}(X_n' \in dx) = \mathbb{P}([n] \subset J_2)\left((1 - x^n - (1-x)^n)1_{x \in (0,1)} + n1_{x \in \{0,1\}}\right)\mu(dx).$$

Integrating w.r.t. x, we have

$$\mathbb{P}([n] \subset J_2) = \left(\int_{x \in (0,1)}(1 - x^n - (1-x)^n)\mu(dx) + n\mu(\{0,1\})\right)^{-1}$$

$$=: \frac{1}{\alpha_n},$$

so that

$$\mathbb{P}(X_n' \in dx, \#K_n' = i)$$

$$= \alpha_n^{-1}\left(\binom{n}{i}x^i(1-x)^{n-i}1_{x \in (0,1)} + 1_{(x,i)=(0,1) \text{ or } (1,n-1)}\right)\mu(dx).$$

Integrating w.r.t. x gives the result, because $q_n(i) = \mathbb{P}(\#K_n' = i)$. $\qquad\qquad\square$

2 Real trees

Textbooks and surveys available on the topic of this section include: Duquesne and Le Gall (2002), Evans (2008) and Le Gall (2005).

2.1 Preliminaries

2.1.1 *Scaling limits*.　As seen in the previous section, it is tempting to investigate the limiting behaviour of some marginals of our random tree shapes with n tips, as $n \to \infty$. Typically interesting marginals include the maximal leaf height (the generation at which the population becomes extinct), the maximal width (the maximal population size), the coalescence time (number of generations back to the most common recent ancestor of a given subpopulation). There are also higher dimensional, natural marginals like the leaf-height process (the sequence of heights of tip i, $i = 1, \ldots, n$, for some plane embedding of the tree), or the width process (the process counting the number of individuals at each successive generation).

If for the same proper rescaling several of these marginals converge in distribution, it is relevant to ask whether the trees themselves converge in some sense to some continuous object. Such a limit theorem would have several important implications.

First, if our marginals of interest can be obtained from the tree by a continuous mapping, then by the continuous mapping theorem, they should converge to the image of the limiting object by the same mapping (in practice however, it can be easier to prove the convergence directly than to prove the mapping's continuity...).

Second, some difficult computations in the finite case can be smoothened out in the limiting case, just as solving differential equations can be simpler than solving difference equations. The limit theorem already provides the scaling, now computations can provide the constant in front of the scaling. For example, the maximal distance achieved by a random walker in n time units scales like $\sqrt{n}$ thanks to Donsker's theorem, and when rescaled by $\sqrt{n}$, it converges in distribution to the maximum of the reflected Brownian motion on $[0, 1]$.

Third, exactly as in the case of Donsker's theorem, we could hope that the limit theorem is an invariance principle, in the sense that the law of the limiting object is the same for a wide class of converging random sequences. In practice, this has the very important consequence that the patterns predicted by the model do not depend on the details of the model.

It is beyond the scope of these notes to give more details about limit theorems for random trees, see Le Gall and Miermont (2012) or Haas (2016). In this section, we want to directly pounce to the continuous objects, only mentioning in passing how they arise as limits of discrete objects. We end this section recalling some well-known definitions in this area. We will then introduce the general framework of real trees and explain how they can be usefully coded by a real function called the (jumping) contour process.

2.1.2 *Local time.* If A is a closed subset of $[0, \infty)$, a *local time* associated to A is a nondecreasing mapping $L : [0, \infty) \to [0, \infty)$ such that $L(0) = 0$ and whose points of increase coincide with A. If A is discrete, L can be defined simply, for example as the counting process $L_t = \#[0, t] \cap A$. If A is not discrete, the counting process will blow up at the first accumulation point of A, so a different strategy is needed.

Assume that A is perfect, that is, it has empty interior and no isolated point. For any compact interval, say $[0, M]$, we can construct a continuous mapping $L : [0, M] \to [0, 1]$ such that $L(0) = 0$, $L(M) = 1$ and for any $0 \leq s < t \leq M$

$$L_t > L_s \quad \nLeftrightarrow \quad (s, t) \cap A \neq \varnothing.$$

The construction is recursive, exactly as for the Cantor–Lebesgue function, also called "devil's staircase". The reader who already knows this construction may skip the next paragraph.

First, recall that the open set $B := (0, M) \setminus A$ can be written as a countable union of open intervals, say $(I_n)_{n \geq 1}$. Note that for any $\varepsilon > 0$, there can be only finitely many of these intervals which have length larger than ε. Therefore, we can assume that the intervals (I_n) are ranked by decreasing order of their lengths (in case of equality, in their order of appearance, say). Finally for each $n \geq 1$, write $I_n = (g_n, d_n)$.

We are going to construct recursively, for each $n \geq 1$, a continuous mapping $L^n : [0, M] \to [0, 1]$ which is piecewise affine and constant exactly on $\bigcup_{k=1}^{n} I_k$. First, L^1 is the function equal to $1/2$ on I_1, affine on $[0, g_1]$ and on $[d_1, 1]$, such that $L^1(0) = 0$ and $L^1(M) = 1$. Now assume that we are given a continuous function $L^n : [0, M] \to [0, 1]$ which is piecewise affine and constant exactly on $\bigcup_{k=1}^{n} I_k$. Writing $d_0 = 0$ and $g_0 = 1$, there is a unique pair $0 \leq k, j \leq n$ such that $d_j < g_{n+1} < d_{n+1} < g_k$ minimizing $g_k - d_j$. Then we can define L^{n+1} as the continuous function equal to L^n outside (d_k, g_j), constant to $\frac{1}{2}(L^n(g_j) + L^n(d_k))$ on $[g_{n+1}, d_{n+1}]$, affine on $[d_j, g_{n+1}]$ and on $[d_{n+1}, g_k]$. It is easy to see that L^{n+1} satisfies the announced properties.

Now for any $p \in \mathbb{N}$, let $\mathscr{D}_p$ denote the set of dyadic numbers of $(0, 1)$ whose dyadic expansion has length smaller than p, that is, $\mathscr{D}_p = \{x \in (0, 1) : \exists (x_1, \ldots, x_p) \in \{0, 1\} : x = \sum_{k=1}^{p} x_k 2^{-k}\}$. Also for $n \in \mathbb{N}$, let $\mathscr{J}_n$ denote the set of values taken by L^n on its constancy intervals. Because A has no isolated point, for each $p \geq 1$ there is an integer N such that for all $n \geq N$, $\mathscr{D}_p \subset \mathscr{J}_n$, so that for any $n' \geq N$, $\|L^n - L^{n'}\| \leq 2^{-p}$, where $\|\cdot\|$ is the supremum norm on $[0, 1]$. This shows that (L^n) is a Cauchy sequence for the supremum norm, and so converges uniformly on $[0, 1]$ to a continuous function L. It is not difficult to see, using the stationarity of (L^n) on B, that L increases exactly on A.

In probability theory, local times are most often used to "count" the visits to a point or set by a stochastic process, for example, the visit times of zero by Brownian motion. In this case, there are alternative ways of constructing the local time

which ensure that it is at the same time adapted and unique up to a multiplicative constant. This contrasts with the recursive construction given above, where the local time is measurable, but is certainly neither adapted nor unique.

In the case of the standard Brownian motion B (but also of many other Markov processes or semi-martingales), there exist several possible such constructions. For example, if $N_\varepsilon(t)$ denotes the number of positive excursions of B with height larger than ε, then a.s. for all t, $2\varepsilon N_\varepsilon(t)$ converges as $\varepsilon \downarrow 0$ to L_t^0, which is a local time at 0 for B. Also,

$$L_t^0 = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{|B_s| < \varepsilon}\, ds,$$

where the limit again holds a.s. for all t. Even more interestingly, the local times of B at all levels (not only 0) can be simultaneously constructed in a consistent manner. Namely, let ν_t denote the so-called occupation measure of B, that is for any non-negative Borel function f

$$\int_{\mathbb{R}} f\, d\nu_t = \int_0^t f(B_s)\, ds.$$

Then a.s. for all t, ν_t has a density $(L_t^a; a \in \mathbb{R})$ w.r.t. Lebesgue measure. In addition, the doubly indexed process $(L_t^a; a \in \mathbb{R})$ is bicontinuous and for each a, $(L_t^a; t \geq 0)$ is a local time for B at level a.

Not the least utility of an adapted local time L (at 0 say) for a stochastic process X, is that it provides a way of indexing the excursions of X away from 0 in their order of appearance. Indeed, if J denotes the inverse J of L, then to each jump ΔJ_s of J corresponds an excursion of X away from 0 with length ΔJ_s, say e_s. In particular, if X is a strong Markov process, $((s, \Delta J_s); s \geq 0)$ are the atoms of a Poisson point process in $[0, \infty)^2$. Furthermore, $((s, e_s); s \geq 0)$ are the atoms of a Poisson point process in $[0, \infty) \times \mathscr{E}$, where $\mathscr{E}$ is the space of paths with finite lifetime V visiting 0 at most at 0 and at V. The intensity measure of this Poisson point process is called *Itô's excursion measure* and is the analogue to the common probability distribution of excursions when the visit times of 0 by X form a discrete set.

2.2 Definitions and examples

2.2.1 *The real tree*.

Definition 2.2.1. A real tree, or $\mathbb{R}$-tree, is a complete metric space $(\mathfrak{t}, d)$ satisfying:

(A) Uniqueness of geodesics. For any $x, y \in \mathfrak{t}$, there is a unique isometric map $\phi_{x,y} : [0, d(x, y)] \to \mathfrak{t}$ such that $\phi_{x,y}(0) = 0$ and $\phi_{x,y}(d(x, y)) = y$.
The geodesic $\phi_{x,y}([0, d(x, y)])$, also called *arc*, is denoted $[\![x, y]\!]$.

(B) No loop. For any continuous, injective map $\psi : [0, 1] \to \mathfrak{t}$, $\psi([0, 1]) = [\![\psi(0), \psi(1)]\!]$.

The root of an $\mathbb{R}$-tree $\mathfrak{t}$ is a distinguished element of $\mathfrak{t}$ denoted ρ.

Theorem 2.2.2 (Four points condition). *The metric space $(\mathfrak{t}, d)$ is a real tree if it is complete, path-connected and satisfies for any $x_1, x_2, x_3, x_4 \in \mathfrak{t}$*

$$d(x_1, x_2) + d(x_3, x_4) \leq \max\big[d(x_1, x_3) + d(x_2, x_4), d(x_1, x_4) + d(x_2, x_3)\big].$$

For references on real trees and the paternity of the last theorem, see Dress, Moulton and Terhalle (1996) and Duquesne (2006), page 2.

Definition 2.2.3. For any $x \in \mathfrak{t}$, the *multiplicity*, or *degree* of x denotes the number of connected components of $\mathfrak{t} \setminus \{x\}$.

- $m(x) = 1$: x is called a *leaf*.
- $m(x) = 2$: x is an *internal vertex*.
- $m(x) \geq 3$: x is a *branching point*.

The set of leaves of $\mathfrak{t}$ is denoted $\mathrm{Lf}(\mathfrak{t})$ and the set of branching points $\mathrm{Br}(\mathfrak{t})$. The *skeleton* of $\mathfrak{t}$ is $\mathrm{Sk}(\mathfrak{t}) := \mathfrak{t} \setminus \mathrm{Lf}(\mathfrak{t})$.

Exercise 2.2.4. Prove that for any sequence (x_n) dense in the $\mathbb{R}$-tree $\mathfrak{t}$,

$$\mathrm{Sk}(\mathfrak{t}) = \bigcup_n [\![\rho, x_n[\![.$$

From now on, we will assume that $\mathfrak{t}$ *denotes a* binary $\mathbb{R}$-*tree, that is, $m(x) \leq 3$ for all $x \in \mathfrak{t}$. We will also assume that $m(\rho) = 1$. We will further need the following notation and terminology.*

- Mrca. For any $x, y \in \mathfrak{t}$ the *most recent common ancestor* (in short mrca) of x and y, denoted $x \wedge y$, is the unique $z \in \mathfrak{t}$ such that $[\![\rho, x]\!] \cap [\![\rho, y]\!] = [\![\rho, z]\!]$.
- Partial order. For any $x, y \in \mathfrak{t}$, y is said to *descend* from x, and then x is called an *ancestor* of y if $x \in [\![\rho, y]\!]$, and this is denoted $x \preceq y$.
- Lebesgue measure. Whenever $\mathfrak{t}$ is locally compact, there is a unique measure ℓ on the Borel σ-field of $\mathfrak{t}$, called Lebesgue measure or *length measure*, such that for any $x, y \in \mathfrak{t}$, $\ell([\![x, y]\!]) = d(x, y)$ (see Section 4.3.5 in Evans (2008)).
- Orientation. For any $x \in \mathrm{Br}(\mathfrak{t})$, $\mathfrak{t} \setminus \{x\}$ has 3 connected components: the one containing ρ and two others, which are assumed to be labelled as the *left subtree* L_x and the *right subtree* R_x.

2.2.2 *First constructive example*: *Connecting segments*. Fix some infinite-dimensional, complete vector space $\mathcal{X}$ and let (γ_n) be a linearly independent, countable family of $\mathcal{X}$. Then construct a real tree by recursively connecting segments as follows.

1. Start with a segment colinear to γ_1.
2. Given a set consisting of the first n segments properly connected, draw a uniform point (according to Lebesgue measure) in this set and glue one of the extremities of a segment colinear to γ_{n+1} to this point.
3. When all segments are connected, take the closure $\mathfrak{t}$ of the resulting set.

The (random) set $\mathfrak{t}$ is obviously path connected and satisfies the four points condition. So it is a real tree iff it is complete, which holds iff the sequence $(\|\gamma_n\|)$ vanishes (see Evans (2008), Lemma 4.33). In this case $\mathfrak{t}$ is even compact.

2.2.3 *Second constructive example*: *Chronological trees.* We start with a well-known coding of discrete, rooted (plane) trees, sometimes denoted UHN, for Ulam–Harris–Neveu.

Definition 2.2.5. A (rooted) *discrete tree* $\mathcal{T}$ is a subset of $\mathscr{U} := \bigcup_{n \in \mathbb{Z}_+} \mathbb{N}^n$ (finite words), with the convention $\mathbb{N}^0 = \{\varnothing\}$, whose elements are called *vertices*, satisfying

  (i) $\varnothing \in \mathcal{T}$,
 (ii) if $v = uj \in \mathcal{T}$, then $u \in \mathcal{T}$,
(iii) for any $u \in \mathcal{T}$, there is $K_u \in \mathbb{Z}_+ \cup \{+\infty\}$ such that

$$uj \in \mathcal{T} \quad \Leftrightarrow \quad 1 \le j < K_u.$$

The vertex $\varnothing$ is called the root of $\mathcal{T}$.

  Let us give some further terminology and notation.

- Edge. An *edge* is any (non-ordered) pair $\{u, uj\}$ such that $u \in \mathcal{T}$, $uj \in \mathcal{T}$.
- Partial order. The vertex v is said to *descend* from u, and then u is called an *ancestor* of v if there is a finite word w such that $v = uw$, and this is denoted $u \preceq v$.
- Generation. The number of letters in the word u is called its *length*, or *generation*, and denoted $|u|$.
- Ancestor. The ancestor of u at generation k is denoted $u|k$.
- Mrca. The *most common recent ancestor* (mrca) of u and v is

$$u \wedge v := \arg\max_{w \in \mathcal{T}} \{|w| : w \preceq u, w \preceq v\}.$$

- Distance. The graph distance d in $\mathcal{T}$ can be written as

$$d(u, v) = |u| + |v| - 2|u \wedge v|, \qquad u, v \in \mathcal{T}.$$

- Boundary. We denote by $\partial \mathcal{T}$ the *boundary* of $\mathcal{T}$ defined as

$$\partial \mathcal{T} := \{u \in \mathbb{N}^{\mathbb{N}} : \forall n \in \mathbb{N}, u|n \in \mathcal{T}\}.$$

Definition 2.2.6. Let $\mathcal{T}$ be a discrete, rooted tree. Assume that each vertex $u \in \mathcal{T}$ is endowed with a date of birth $\alpha(u) \in [0, +\infty)$, a date of death $\omega(u) \in (\alpha(u), +\infty]$ and a lifetime duration $\zeta(u) := \omega(u) - \alpha(u)$, satisfying

  (i) $\alpha(\varnothing) = 0$,
 (ii) for any $u \in \mathcal{T}$, for any $j \in \mathbb{N}$,

$$uj \in \mathcal{T} \quad \Rightarrow \quad \alpha(u) < \alpha(uj) \leq \omega(u),$$

(iii) for any $u \in \partial\mathcal{T}$,

$$\alpha(u) := \lim_{n \uparrow \infty} \uparrow \alpha(u|n) < \infty \quad \Rightarrow \quad \lim_{n \to \infty} \zeta(u|n) = 0.$$

Then we can define the *chronological tree* $\mathbb{t}$ as the subset of $\mathbb{U} := \mathscr{U} \times [0, \infty]$ defined by

$$\mathbb{t} := \bigcup_{u \in \mathcal{T}} \{u\} \times \big(\alpha(u), \omega(u)\big] \cup \bigcup_{u \in \partial\mathcal{T}: \alpha(u) < \infty} \{u\} \times \{\alpha(u)\}.$$

The chronological tree is naturally rooted at $\rho := (\varnothing, 0)$. If in addition

$$\alpha(ui) \neq \alpha(uj), \qquad u \in \mathcal{T}, i \neq j,$$

then the tree $\mathcal{T}$ is said *binary*.

We will further need the following notation and terminology. We use (here only) the notation $p_2 : \mathbb{U} \to [0, \infty]$ for the canonical projection $p_2((u, t)) = t$.

- Partial order. The point $y = (v, t) \in \mathbb{t}$ is said to *descend* from $x = (u, s) \in \mathbb{t}$, and then x is called an *ancestor* of y, if either $u = v$ and $s \leq t$, or $u \neq v$, $u \preceq v$ and $s \leq \alpha(v)$. This is denoted $x \preceq y$.
- Mrca. The *most common recent ancestor* of $x \in \mathbb{t}$ and $y \in \mathbb{t}$ is

$$x \wedge y := \arg\max_{z \in \mathbb{t}} \{p_2(z) : z \preceq x, z \preceq y\}.$$

- Distance. We still denote by d the graph distance in $\mathbb{t}$, defined by

$$d(x, y) = p_2(x) + p_2(y) - 2p_2(x \wedge y), \qquad x, y \in \mathbb{t}.$$

Exercise 2.2.7. Prove the following statement.

Theorem 2.2.8. *The metric space* $(\mathbb{t}, d)$ *is a locally compact* $\mathbb{R}$-*tree.*

Exercise 2.2.9. Check that the notions of partial order and mrca in the chronological tree $\mathbb{t}$ coincide with the corresponding notions in $\mathbb{R}$-trees. Characterize the leaves and branching points of $\mathbb{t}$.
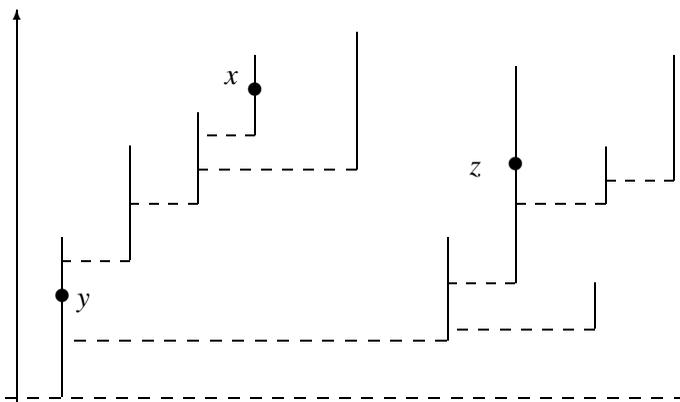
Figure 4 *A binary chronological tree, where edges are vertical, time flows upward, dashed lines represent filiation and daughters sprout to the right of their mother, conferring a natural orientation to the tree. The three points x, y, z satisfy $y \preceq x$ and $x \leq y \leq z$.*

Remark 2.2.10. For chronological trees represented in the plane with vertical edges, there is a natural orientation stemming from the rule that "daughters sprout to the right of their mother" (see Figure 4 and see Lambert (2010) for a rigorous definition). We will always assume that chronological trees are endowed with this specific orientation.

2.3 The contour process

2.3.1 *From the $\mathbb{R}$-tree to its contour process.*   Let $\mathfrak{t}$ be a binary, oriented $\mathbb{R}$-tree.

Definition 2.3.1. The relation $\leq$ on $\mathfrak{t}$ is defined as follows. For any $x, y \in \mathfrak{t}$,

$$x \preceq y \quad \Rightarrow \quad y \leq x,$$

otherwise $x \wedge y \in \mathrm{Br}(\mathfrak{t})$ and

$$\begin{cases} x \in L_{x \wedge y} & \Rightarrow \quad x \leq y, \\ x \in R_{x \wedge y} & \Rightarrow \quad y \leq x. \end{cases}$$

Exercise 2.3.2. Prove that $\leq$ is a total order on $\mathfrak{t}$, that $\rho = \max \mathfrak{t}$, and find $\min \mathfrak{t}$ on the example shown in Figure 4.

Exercise 2.3.3. Prove that for any $x \in \mathfrak{t}$,

$$\pi(x) := \{y \in \mathfrak{t} : y \leq x\}$$

is a Borel set.

Now we assume that we are given a finite measure μ on the Borel σ-field of $\mathfrak{t}$, called *mass measure*, satisfying

Mes 1 for any $x \leq y \in \mathfrak{t}$,

$$x \neq y \quad \Rightarrow \quad \mu\big(\pi(x)\big) < \mu\big(\pi(y)\big).$$

Mes 2 μ is diffuse (no atom).

Remark 2.3.4. Whenever $\ell(\mathfrak{t}) < \infty$, the length measure ℓ is a natural example of finite measure which satisfies both **Mes 1** and **Mes 2**.

Now define $\varphi : (\mathfrak{t}, \leq, \mu) \to ([0, \mu(\mathfrak{t})], \leq, \text{Leb})$ by

$$\varphi(x) := \mu\big(\pi(x)\big), \qquad x \in \mathfrak{t},$$

which always makes sense, since $\pi(x)$ is a Borel set of $\mathfrak{t}$. It is clear that φ is one-to-one (by **Mes 1**), preserves the order and the measure. But it is not clear whether it is onto.

Exercise 2.3.5. Display an example of a real tree that has no minimal element, and so for which $0 \notin \varphi(\mathfrak{t})$.

Lemma 2.3.6. *The set $D := \varphi(\mathfrak{t})$ is dense in $[0, \mu(\mathfrak{t})]$.*

Proof. Let $t \in (0, \mu(\mathfrak{t}))$. Set $G_t := \{x \in \mathfrak{t} : \varphi(x) < t\}$ and $D_t := \{y \in \mathfrak{t} : \varphi(y) \geq t\}$. Also set $s_t := \sup\{\varphi(x) : x \in G_t\}$ and $i_t := \inf\{\varphi(y) : y \in D_t\}$, so that in particular $s_t \leq t \leq i_t$.

First notice that for any $x \in G_t$, $\pi(x) \subset G_t$, so that G_t is necessarily of the form $\pi(x)$ or $\pi(x) \setminus \{x\}$, which yields $\mu(G_t) = s_t$.

Now by definition of i_t, there is some $\leq$-decreasing sequence (y_n) of elements of D_t such that $\lim_n \downarrow \varphi(y_n) = i_t$. Since (y_n) is decreasing, the sequence $\pi(y_n)$ is also decreasing, let L denote its limit. If there were two elements in $L \setminus G_t$, say $z_1 < z_2$, we would have $t \leq \varphi(z_1) < \varphi(z_2) \leq i_t$ by **Mes 1**. Now this contradicts the definition of i_t, since $L \setminus G_t \subset D_t$, so $L \setminus G_t$ contains at most one element. By **Mes 2**, this shows that $\mu(L) = \mu(G_t)$. Now recall that $\mu(G_t) = s_t$, so that

$$i_t = \lim_n \downarrow \varphi(y_n) = \lim_n \downarrow \mu\big(\pi(y_n)\big) = \mu(L) = \mu(G_t) = s_t,$$

which shows that $i_t = s_t = t$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

In light of the previous lemma, we can define $\phi : [0, \mu(\mathfrak{t})] \to \mathfrak{t}$ as

$$\phi(t) := \lim_{s \downarrow t, s \in D} \varphi^{-1}(s),$$

which we call the *exploration process*. The existence and the uniqueness of this limit come from the fact that all monotonic sequences of $\mathfrak{t}$ do converge (see Lambert and Uribe Bravo (2016b)). Of course, ϕ does not preserve the order any longer.
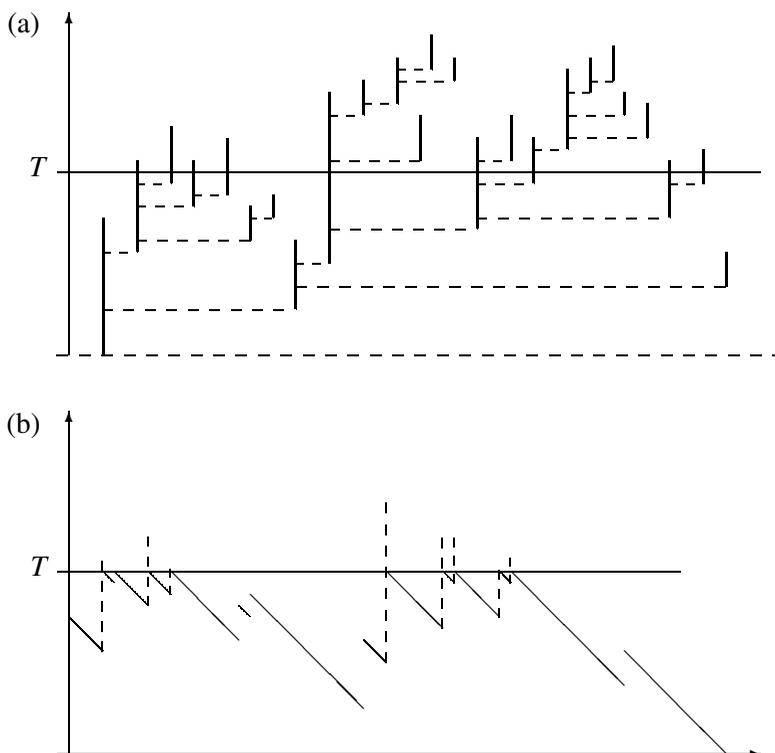
Figure 5 *A chronological tree* (a) *and* (b) *the jumping contour process of its truncation below T, where μ is chosen equal to ℓ.*

Theorem 2.3.7 (Lambert and Uribe Bravo (2016b)). *The exploration process is the only càdlàg extension to φ^{-1}. The mapping $h : [0, \mu(\mathfrak{t})] \to [0, \infty)$ defined by $h(s) := d(\rho, \phi(s))$ is called the* jumping contour process *of $\mathfrak{t}$. It is càdlàg and has no negative jumps.*

Figure 5 shows an example of a real tree and of the jumping contour process of its truncation below T, when μ is chosen equal to ℓ. Figure 6 shows how to recover a chronological tree from its contour.

2.3.2 *From the contour to the tree.* Let $h : [0, \infty) \to [0, \infty)$ be càdlàg with no negative jumps and compact support. Set $\sigma_h := \sup\{t > 0 : h(t) \neq 0\}$, as well as

$$m_h(s, t) := \inf_{[s \wedge t, s \vee t]} h, \qquad s, t \geq 0,$$

and

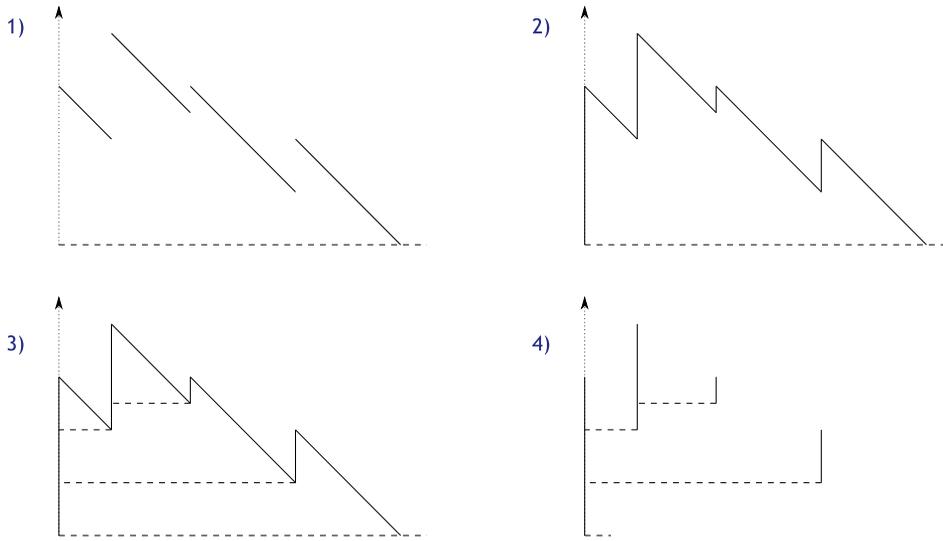$$d_h(s, t) := h(s) + h(t) - 2m_h(s, t).$$

Figure 6 *The jumping contour process of a chronological tree with finite length, where μ is chosen equal to ℓ: how to recover the tree from the contour.* (1) *Start with a càdlàg map with compact support*; (2) *Draw vertical solid lines in the place of jumps*; (3) *Report horizontal dashed lines from each edge bottom left to the rightmost solid point*; (4) *Erase diagonal lines.*

It is clear that d_h is a pseudo-distance on $[0, \infty)$. Further let $\sim_h$ denote the equivalence relation on $[0, \infty)$

$$s \sim_h t \quad \Leftrightarrow \quad d_h(s, t) = 0 \quad \Leftrightarrow \quad h(s) = h(t) = m_h(s, t).$$

Theorem 2.3.8. *Denote by $\mathbb{t}_h$ the quotient space $[0, \sigma_h]|_{\sim_h}$. Then $(\mathbb{t}_h, d_h)$ is a compact $\mathbb{R}$-tree.*

Exercise 2.3.9. Prove the last statement using the four points condition.

From now on, let $p_h : [0, \sigma_h] \to \mathbb{t}_h$ map any element of $[0, \sigma_h]$ to its equivalence class relative to $\sim_h$. We can also endow $\mathbb{t}_h$ with a total order and a mass measure, as follows.

- Total order. We define $\leq_h$ as the order of first visits, that is for any $x, y \in \mathbb{t}_h$,

$$x \leq_h y \quad \Leftrightarrow \quad \inf p_h^{-1}(\{x\}) \leq \inf p_h^{-1}(\{y\}).$$

- Mass measure. The measure μ_h is defined as the push forward of Lebesgue measure by p_h.

Theorem 2.3.10 (Duquesne (2006), Lambert and Uribe Bravo (2016b)). *Let $(\mathbb{t}, d)$ be a compact, binary $\mathbb{R}$-tree endowed with an orientation inducing a total order $\leq$ (as in Definition 2.3.1) and with a (finite) mass measure μ (satisfying*

Mes 1 *and* **Mes 2**). *Let h denote the jumping contour process associated with* $\leq$ *and* μ. *Then h is the unique càdlàg map such that the tree* $(\mathbb{t}_h, d_h, \leq_h, \mu_h)$ *is isomorphic to* $(\mathbb{t}, d, \leq, \mu)$.

2.3.3 *A few words on topology.* Real trees are metric spaces. The "space" of real trees only makes sense if one can imbed all trees into the same metric space, say $(\mathcal{X}, \delta)$, and if two compact real trees, seen as closed subsets of $\mathcal{X}$, are identified when there is a root-preserving isometry mapping one tree onto the other.

Rigorously, the space of real trees is then the set of isometry classes of trees which are closed subsets of $\mathcal{X}$, and it can then be endowed with the usual Hausdorff metric δ_H associated with δ (i.e., $\delta_H(F_1, F_2)$ is the smallest ε such that the ε-enlargement of F_i is contained in F_j, $i \neq j$), called on this occasion the *Gromov–Hausdorff distance* and denoted d_{GH}, see, for example, Burago, Burago and Ivanov (2001) and Paulin (1989).

In other words, this distance is defined for any two $\mathbb{R}$-trees $\mathbb{t}_1$ and $\mathbb{t}_2$ as

$$d_{GH}(\mathbb{t}_1, \mathbb{t}_2) = \inf_{f_i : \mathbb{t}_i \to \mathcal{X}} \delta_H\big(f_1(\mathbb{t}_1), f_2(\mathbb{t}_2)\big) \vee \delta\big(f_1(\rho_1), f_2(\rho_2)\big),$$

where the infimum is taken over all isometries imbedding $\mathbb{t}_1$ and $\mathbb{t}_2$ into $\mathcal{X}$.

Theorem 2.3.11 (Evans, Pitman and Winter (2005)). *The Gromov–Hausdorff distance makes the space of compact real trees a complete, separable space.*

Actually, one can avoid resorting to the abstract space $\mathcal{X}$ and directly deal with correspondences between $\mathbb{t}_1$ and $\mathbb{t}_2$.

Definition 2.3.12. A correspondence between $(\mathbb{t}_1, d_1)$ and $(\mathbb{t}_2, d_2)$ is a subset $\mathcal{R}$ of $\mathbb{t}_1 \times \mathbb{t}_2$ such that

$$\forall x_1 \in \mathbb{t}_1 \, \exists x_2 \in \mathbb{t}_2 \qquad (x_1, x_2) \in \mathcal{R},$$

$$\forall y_2 \in \mathbb{t}_2 \, \exists y_1 \in \mathbb{t}_1 \qquad (y_1, y_2) \in \mathcal{R}.$$

The distortion $\text{dis}(\mathcal{R})$ of the correspondence $\mathcal{R}$ is defined as

$$\text{dis}(\mathcal{R}) = \sup\big\{\big|d_1(x_1, y_1) - d_2(x_2, y_2)\big| : (x_1, x_2) \in \mathcal{R}, (y_1, y_2) \in \mathcal{R}\big\}.$$

Then we have the following useful equality (Burago, Burago and Ivanov (2001))

$$d_{GH}(\mathbb{t}_1, \mathbb{t}_2) = \frac{1}{2} \inf_{\mathcal{R}} \text{dis}(\mathcal{R}),$$

where the infimum is taken over all distortions $\mathcal{R}$ between $\mathbb{t}_1$ and $\mathbb{t}_2$. This equality has the following consequence (which is a slight improvement of Lemma 2.4 in Le Gall (2005)).

Proposition 2.3.13. *Let $h_1, h_2 : \mathbb{R}_+ \to \mathbb{R}_+$ be two càdlàg functions with compact support, and let $\mathfrak{t}_1 := \mathfrak{t}_{h_1}$ and $\mathfrak{t}_2 := \mathfrak{t}_{h_2}$ denote the real trees associated with h_1 and h_2, respectively. Then*

$$d_{\mathrm{GH}}(\mathfrak{t}_1, \mathfrak{t}_2) \le 2d_S(h_1, h_2),$$

where d_S denotes the Skorokhod distance.

Remark 2.3.14. A very important consequence of this proposition is that whenever a sequence (X_n) of càdlàg, non-negative stochastic processes with compact support converges weakly in Skorokhod space to X, the trees coded by X_n converge weakly in the Gromov–Hausdorff sense to the tree coded by X.

Also note that the separability of the Gromov–Hausdorff tree space stems from the separability of the Skorokhod space, thanks again to the last statement.

Proof of Proposition 2.3.13. Let $\varepsilon > 0$ and let λ be a perturbation such that $\|h_1 \circ \lambda - h_2\| \le d_S(h_1, h_2) + \varepsilon$, where $\|\cdot\|$ denotes the supremum norm. Then let $\mathscr{R}$ be the correspondence defined by

$$\mathscr{R} = \big\{(x_1, x_2) \in \mathfrak{t}_1 \times \mathfrak{t}_2 : \exists t \ge 0, \, p_{h_1}(\lambda(t)) = x_1, \, p_{h_2}(t) = x_2\big\}.$$

Then for any $(x_1, x_2) \in \mathscr{R}$ and $(y_1, y_2) \in \mathscr{R}$, there are $s, t \ge 0$ such that

$$x_1 = p_{h_1}(\lambda(s)), x_2 = p_{h_2}(s) \quad \text{and} \quad y_1 = p_{h_1}(\lambda(t)), y_2 = p_{h_2}(t).$$

Now

$$d_1(x_1, y_1) = h_1(\lambda(s)) + h_1(\lambda(t)) - 2m_{h_1}(\lambda(s), \lambda(t))$$

and

$$d_2(x_2, y_2) = h_2(s) + h_2(t) - 2m_{h_2}(s, t),$$

so that

$$\big|d_1(x_1, y_1) - d_2(x_2, y_2)\big|$$
$$\le |h_1 \circ \lambda - h_2|(s) + |h_1 \circ \lambda - h_2|(t) + 2\left|\inf_{[s,t]} h_1 \circ \lambda - \inf_{[s,t]} h_2\right|$$
$$\le 4\|h_1 \circ \lambda - h_2\|.$$

Then by definition of the distortion, $\mathrm{dis}(\mathscr{R}) \le 4\|h_1 \circ \lambda - h_2\|$, so the inequality stated before the proposition yields

$$d_{\mathrm{GH}}(\mathfrak{t}_1, \mathfrak{t}_2) = \frac{1}{2} \inf_{\mathscr{R}} \mathrm{dis}(\mathscr{R}) \le 2\|h_1 \circ \lambda - h_2\| \le 2d_S(h_1, h_2) + 2\varepsilon,$$

which yields the result. Note that we have not needed to control the difference between λ and the identity. $\qquad\square$

2.4 Random $\mathbb{R}$-trees

2.4.1 *Splitting trees*. Splitting trees (Geiger (1996), Geiger and Kersting (1997), Lambert (2010)) are random chronological trees satisfying the branching property. More specifically, let Λ be a positive measure on $(0, \infty]$, called the *lifespan measure*, such that $\int_{(0,\infty]} (r \wedge 1) \Lambda(dr) < \infty$. A splitting tree is a random chronological tree, where individuals live and reproduce independently and conditional on the life span $(\alpha(u), \omega(u)]$ of a given individual u, pairs of birth times and lifetimes of the newborns of u form a Poisson point process on $(\alpha(u), \omega(u)] \times (0, \infty]$ with intensity Leb $\otimes \Lambda$. If Λ is finite with mass $b = \Lambda((0, \infty])$, then individuals give birth at rate b to individuals with lifetime distribution $b^{-1}\Lambda$.

Observe that the width process of a splitting tree is not necessarily Markovian. When Λ is finite, it is a binary, homogeneous Crump–Mode–Jagers (CMJ) process, and it is not Markovian unless the lifetime distribution is exponential (or a Dirac mass at $\{\infty\}$ in the pure-birth case). For modeling purposes, note that splitting trees with absolutely continuous lifetimes can equivalently be defined via a "death rate" that can be age-dependent.

Remark 2.4.1. There are two branching processes hidden in a splitting tree other than its width process. The first one is the process tracking the number of individuals alive at each given (discrete) generation, and the second one is the process tracking the total sum of lifetimes of individuals of each given generation. Both are Markovian branching processes in discrete time, the first one with integer values (a Galton–Watson process), and the second one with non-negative real values (a Jirina process). Note that the Jirina process can take finite values even when Λ is not finite, which is not the case of the Galton–Watson process.

Recall that chronological trees are naturally endowed with the orientation associated with the rule that "daughters sprout to the right of their mother", so they are given a natural order $\leq$ associated with this orientation.

In addition, thanks to our assumption on Λ, the length measure is a.s. locally finite. So it is possible to use the length measure to define the exploration process and the jumping contour process for the tree truncated under some fixed, finite height. The law of the jumping contour process is particularly appealing in this setting (see also Figure 6).

Theorem 2.4.2 (Lambert (2010)). *Let $X_t = Y_t - t$, where Y is the subordinator with Lévy measure Λ. Conditional on the lifetime x of the root individual, the jumping contour process of the splitting tree with lifespan measure Λ truncated below height a is distributed like the process X started at x, reflected below a and killed upon hitting 0.*

2.4.2 *The continuum random tree.* Recall how in Section 2.2.2 we have constructed a (random) real tree $\mathfrak{t}$ by connecting together segments colinear to the elements (γ_n) of a linearly independent family of a complete vector space $\mathcal{X}$. To define the *Continuum Random Tree* (CRT) discovered by Aldous (1991, 1993), we need the lengths of the segments to be random ($\mathcal{X} = \ell_1$ in the original paper).

Definition 2.4.3. The CRT is the tree $\mathfrak{t}$ obtained by connecting segments whose lengths $(\|\gamma_n\|)$ are distributed as the successive distances between consecutive atoms of a Poisson point process on the half line with inhomogeneous intensity $t \, dt$.

Theorem 2.4.4 (Aldous (1991, 1993), Le Gall (1993)). *Let* **e** *stand for the normalized Brownian excursion, i.e., the positive Brownian excursion conditioned to have lifetime* 1. *The* $\mathbb{R}$-*tree* $\mathfrak{t}_{\mathbf{e}}$ *coded by* **e** *is isometric in law to the CRT.*

It is standard that the contour process (in some appropriate meaning) of a Galton–Watson tree with finite offspring variance conditioned to have n vertices, rescaled by a factor $\sqrt{n}$, converges weakly in the Skorokhod space to the normalized excursion. The results of the last section then imply that the CRT is the scaling limit in the Gromov–Hausdorff topology of conditioned Galton–Watson trees with finite offspring variance (see, e.g., Aldous (1993) and Le Gall (1993)).

Also, since binary trees with n tips have $2n - 1$ vertices, binary Galton–Watson trees conditioned to have n tips, which follow P_n^{pda}, converge to the tree coded by the Brownian excursion with length normalized to 2.

This conditioning (on number of vertices or tips) can sometimes be awkward in some practical situations since it is evanescent in the limit. It can be more convenient to consider a forest of n independent critical Galton–Watson trees, whose contour process (in the same appropriate meaning) is the concatenation of n independent contour processes (Le Gall (2005)). In the limit, we should get a contour process which is the concatenation of a fixed amount of excursions. This fixed amount is measured by the local time at 0 of this process.

Actually, the contour processes of Galton–Watson processes are not in general Markovian (Duquesne and Le Gall (2002), Le Gall (2005)), so it will be more convenient to display the same kind of result starting with splitting trees, whose contour process is Markovian, thanks to Theorem 2.4.2. For the sake of generality, we will also allow the trees to be subcritical ($\alpha > 0$ in the following statement).

Let $\zeta > 0$. Consider a forest of $[A\zeta]$ i.i.d. splitting trees, where $A > 0$ is a scaling parameter, characterized by a finite birth rate b_A and a lifetime distribution given by a random variable V_A. We make three assumptions.

(H1) $b_A \mathbb{E}(V_A) = 1 - \frac{\alpha}{A} + o(\frac{1}{A})$.

(H2) $\lim_{A \to \infty} \frac{A \mathbb{E}(V_A^2)}{2 \mathbb{E}(V_A)} = \beta > 0$.

(H3) $\lim_{A \to \infty} A b_A \mathbb{E}(V_A^3 \wedge 1) = 0$.

Remark 2.4.5. Note that if V_A is exponentially distributed with parameter d_A, then the assumptions (H1)–(H3) hold as soon as $d_A = A/\beta = b_A + \alpha/\beta$.

Theorem 2.4.6 (Lambert, Simatos and Zwart (2013)). *Let X^A denote the jumping contour process of this forest and Z_t^A denote the total number of individuals alive at time t, rescaled by A. Then the pair (X^A, Z^A) converges weakly in Skorokhod space.*

First, the limit of the sequence (X^A) is the process $X - \underline{X}$ killed when $\underline{X}$ hits $-\zeta$, where $\underline{X}_t = \inf_{s \in [0,t]} X_s$, and

$$X_t = -\alpha t + \sqrt{2\beta} B_t,$$

where B is the standard Brownian motion started at 0. Second, the limit of (Z^A) is a diffusion process Z started at $Z_0 = \zeta$ and solution to an SDE of the form

$$dZ_t = -\frac{\alpha}{\beta} Z_t \, dt + \sqrt{\frac{2Z_t}{\beta}} \, dW_t.$$

Note that $-\underline{X}$ is a local time at 0 for $X - \underline{X}$, so the limiting contour process $X - \underline{X}$ is indeed killed when its local time hits ζ.

The fact that the width processes Z^A converge cannot be deduced from the convergences of the contours X^A, since the local time functional (mapping the tree to its width process) is not continuous. In this direction, notice that the theorem does not specify that Z is the local time process of $X - \underline{X}$.

Last, the convergence in Skorokhod space of the contours ensures that the splitting trees themselves converge in the Gromov–Hausdorff sense to what we could call Brownian tree, in a wider sense than the CRT (no normalization of the contour excursion interval, possible subcriticality).

Remark 2.4.7. Actually, the previous theorem only holds if the lifetimes of the $[A\zeta]$ progenitors are (i.i.d. and) distributed as the forward recurrence time $V_A^\star$ of V_A, otherwise the width process is not continuous at 0

$$\mathbb{P}(V_A^\star \in dx) = \frac{\mathbb{P}(V_A \geq x)}{\mathbb{E}(V_A)} \, dx, \qquad x > 0.$$

Note that $V_A^\star$ is distributed like V_A if (and only if) V_A is exponentially distributed.

3 Reduced trees

For a real tree $\mathfrak{t}$ and a fixed real number $T > 0$, the so-called *reduced tree* at height T is the tree spanned by points at distance T from the root

$$\{y \in \mathfrak{t} : \exists x \in \mathfrak{t}, y \preceq x, d(\rho, x) = T\}.$$

It is usually called *reconstructed tree* in phylogenetics and *coalescent tree* in population genetics. Its topology can be understood from the topology of the sphere of $\mathtt{t}$ with center ρ and radius $T > 0$

$$\mathtt{t}^{\{T\}} := \{x \in \mathtt{t} : d(\rho, x) = T\},$$

which will thus be the focus of the present section.

3.1 The comb metric

Most of this section is taken from Lambert and Uribe Bravo (2016a).

3.1.1 *Definition and examples.* Let I be a compact interval and $f : I \to \mathbb{R}_+$ such that for any $\varepsilon > 0$, $\{f \geq \varepsilon\}$ is finite.

For any $s, t \in I$, define $\bar{d}_f$ by

$$\bar{d}_f(s, t) = 2 \sup_{(s \wedge t, s \vee t)} f.$$

It is clear that $\bar{d}_f$ is a pseudo-distance on $\{f = 0\}$ and that it is ultrametric, in the sense that

$$\bar{d}_f(r, t) \leq \max(\bar{d}_f(r, s), \bar{d}_f(s, t)), \qquad r, s, t \in I.$$

It is a distance on $\{f = 0\}$ whenever $\{f \neq 0\}$ is dense in I for the usual distance. This may not be the case in general, so we need to consider $\dot{I}$ the quotient space $\{f = 0\}|_\sim$ where $\sim$ is the equivalence relation

$$s \sim t \quad \Leftrightarrow \quad \bar{d}_f(s, t) = 0 \quad \Leftrightarrow \quad f = 0 \qquad \text{on } [s \wedge t, s \vee t].$$

Definition 3.1.1. We call f a *comb-like function* or *comb*, and $\bar{d}_f$ the *comb metric* on $\dot{I}$.

Let us give the canonical example of a comb. Let $\mathtt{t}$ be an oriented $\mathbb{R}$-tree with finite length, total order $\leq$ and jumping contour process h associated (e.g.) with its length measure (i.e., the mass measure μ is taken equal to the length measure ℓ). Let $T > 0$ such that the sphere

$$\mathtt{t}^{\{T\}} = \{x \in \mathtt{t} : d(\rho, x) = T\}$$

has finite cardinality $N_T \geq 2$. Let $x_1 \leq \cdots \leq x_{N_T}$ denote its elements labelled in the order $\leq$. Then for any $1 \leq i < j \leq N_T$, writing $s_i := \inf p_h^{-1}(\{x_i\})$ (each set $p_h^{-1}(\{x\})$ is actually a singleton in the case when the sphere is finite),

$$d(x_i, x_j) = h(s_i) + h(s_j) - 2 \inf_{[s_i, s_j]} h = 2\left(T - \inf_{[s_i, s_j]} h\right) = 2 \max(h_i, \ldots, h_{j-1}),$$
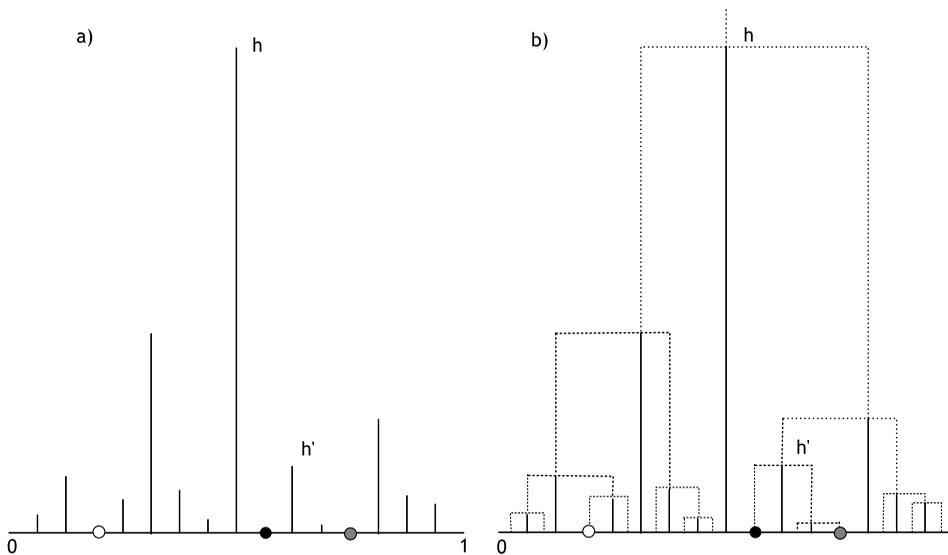
where

$$h_i := T - \inf_{[s_i, s_{i+1}]} h.$$

Figure 7   (a) *A comb-like function with finite support on* [0, 1]. *The distance between the black dot and the grey dot is h′, whereas h is the distance from either of these dots to the white dot*; (b) *In dashed lines, the ultrametric tree associated to the comb shown in* (a).

In conclusion, the metric on $\mathfrak{t}^{\{T\}}$ is isomorphic to the comb metric $\bar{d}_f$ on $\dot{I}$, where $I = [1, N_T]$ and $f := \sum_{i=1}^{N_T-1} h_i \mathbb{1}_{\{i\}}$. We will extend this description to all locally compact $\mathbb{R}$-trees in Section 3.1.2.

Figure 7 shows a comb and how an ultrametric tree can be embedded into it. As in Figure 8 the same metric space can be represented by identifying points of $\dot{I}$ to their left-neighbour in $I \setminus \dot{I}$ (or to their right-neighbour) and reporting the distances accordingly.

Remark 3.1.2. The space $(\dot{I}, \bar{d}_f)$ is not complete in general. To make it complete, one has to distinguish for each point $t \in I$ between its *left face* (t, l) and its *right face* (t, r). The distance $\bar{d}_f$ is extended to the space $I \times \{l, r\}$ by the following definitions for $s < t \in I$

$$\bar{d}_f\big((s, r), (t, l)\big) = 2 \sup_{(s,t)} f, \qquad \bar{d}_f\big((s, l), (t, l)\big) = 2 \sup_{[s,t)} f,$$

$$\bar{d}_f\big((s, r), (t, r)\big) = 2 \sup_{(s,t]} f, \qquad \bar{d}_f\big((s, l), (t, r)\big) = 2 \sup_{[s,t]} f,$$

and the symmetrized definitions for $s > t$. If $s = t$, the four last quantities are respectively, defined as $f(t), 0, 0, f(t)$. This extension of $\bar{d}_f$ is a pseudo-distance and it can be shown (Lambert and Uribe Bravo (2016a)) that the associated quotient space $\bar{I}$ is a compact, ultrametric space called *comb metric space*. Actually the converse is also true, as states the next theorem.
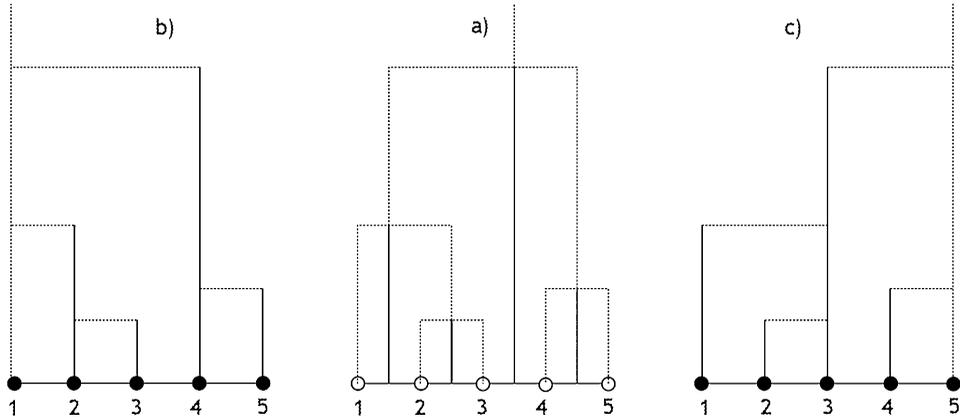
Figure 8   (a) *A comb with finite support, and the associated ultrametric tree in dashed lines; equivalent representations of this ultrametric space can be obtained by reporting all tips of the comb to the left* (b) *or to the right* (c).

Theorem 3.1.3 (Lambert and Uribe Bravo (2016a)). *Any compact ultrametric space is isometric to a comb metric space.*

3.1.2 *Spheres of $\mathbb{R}$-trees.*   Here, we consider a locally compact $\mathbb{R}$-tree $\mathfrak{t}$ and we assume that the sphere

$$\mathfrak{t}^{\{T\}} = \{x \in \mathfrak{t} : d(\rho, x) = T\}$$

is not empty. Note that by the four-points condition, for any $x, y, z \in \mathfrak{t}^{\{T\}}$,

$$
\begin{aligned}
T + d(x, z) &= d(\rho, y) + d(x, z) \\
&\leq \max\big[d(\rho, x) + d(y, z), d(\rho, z) + d(y, x)\big] \\
&= \max\big[T + d(y, z), T + d(y, x)\big],
\end{aligned}
$$

which yields $d(x, z) \leq \max[d(y, z), d(y, x)]$, that is the metric induced by d on $\mathfrak{t}^{\{T\}}$ is ultrametric.

Since $\mathfrak{t}$ is locally compact, $\mathfrak{t}^{\{T\}}$ is a compact ultrametric space and Theorem 3.1.3 ensures that provided it has no isolated point, it is isometric to a comb metric space. There is actually an isometry between $\mathfrak{t}^{\{T\}}$ and a comb metric space preserving the order on $\mathfrak{t}^{\{T\}}$ inherited from the order $\leq$ on $\mathfrak{t}$. Let h denote the jumping contour process of the tree truncated at height T, which is the closed ball with center ρ and radius T.

Exercise 3.1.4. Assume that $\mathfrak{t}^{\{T\}}$ has no isolated point. Prove that $\{h = T\}$ has no isolated point and empty interior. Also prove that $\mathfrak{t}^{\{T\}}$ has empty interior.

Recall from the paragraph page 438 on local time, that since $\{h = T\}$ is perfect, we can construct a local time at level T for h, that is a nondecreasing, continuous map $L^T : [0, \infty) \to [0, \infty)$ such that $L^T(0) = 0$ and for any $s < t$

$$L_t^T > L_s^T \quad \Leftrightarrow \quad (s, t) \cap \{h = T\} \neq \varnothing.$$

Let $I = [0, L_{\sigma_h}^T]$, and set

$$\dot{\mathfrak{t}}^{\{T\}} := \Big\{ x \in \mathfrak{t}^{\{T\}} : \exists (x_n^+) \uparrow, (x_n^-) \downarrow \in \mathfrak{t}^{\{T\}}, \lim_n \uparrow x_n^+ = \lim_n \downarrow x_n^- = x \Big\},$$

where the sequences in the previous definition are requested to be strictly monotonic.

Theorem 3.1.5 (Lambert and Uribe Bravo (2016a)). *Assume as previously that $\mathfrak{t}$ is a locally compact $\mathbb{R}$-tree such that $\mathfrak{t}^{\{T\}}$ is not empty and has no isolated point. Then there is a comb-like function f on I and two global isometries $\dot{\theta} : (\dot{I}, \bar{d}_f) \to (\dot{\mathfrak{t}}^{\{T\}}, d)$ and $\bar{\theta} : (\bar{I}, \bar{d}_f) \to (\mathfrak{t}^{\{T\}}, d)$ preserving the order and mapping the Lebesgue measure to the push forward μ^T of the measure dL^T by p_h.*

3.2 Coalescent point processes

3.2.1 *The reduced tree of splitting trees, of the Brownian tree.*   Consider a splitting tree $\mathfrak{t}$ with lifespan measure Λ satisfying $\int_{(0,\infty]} (r \wedge 1) \Lambda(dr) < \infty$. We have seen in the last section that $\mathfrak{t}$ is an oriented $\mathbb{R}$-tree naturally endowed with the associated total order $\leq$, and that it has locally finite length. Taking the mass measure equal to the length measure ℓ, the jumping contour process X of the tree $\mathfrak{t}$ truncated below T is well defined and thanks to Theorem 2.4.2, it has the law of the process $(Y_t - t; t \geq 0)$, where Y is the subordinator with Lévy measure Λ, reflected below T and killed upon hitting 0.

So $\mathfrak{t}$ falls into the category of canonical examples given after Definition 3.1.1 and conditional on $N_T \geq 1$ (where $N_T = \#\mathfrak{t}^{\{T\}}$), it is isometric to the comb metric space $(\dot{I}, \bar{d}_f)$, where $I = [1, N_T]$ and $f := \sum_{i=1}^{N_T - 1} H_i \mathbb{1}_{\{i\}}$, with

$$H_i := T - \inf_{[\sigma_i, \sigma_{i+1}]} X,$$

and σ_i is the ith visit of T by X.

Note that $\mathbb{P}_x(N_T \neq 0) = \mathbb{P}_x(\tau_T^+ < \tau_0)$, where the subscript x records the lifetime x of the progenitor, which is also the starting point of the contour process, and τ_z^+ (resp. τ_y) denotes the first hitting time by X of $[z, +\infty)$ (resp. of $\{y\}$).

Also note, thanks to the strong Markov property of X, that conditional on $N_T \geq 1$, N_T is geometric with failure probability $\mathbb{P}_T(\tau_T^+ > \tau_0)$. Furthermore, conditional on $N_T = n$, the H_i's are n i.i.d. random variables, all distributed as the depth of the excursion of X away from T conditioned to be smaller than T. It is then elementary to get the following result.
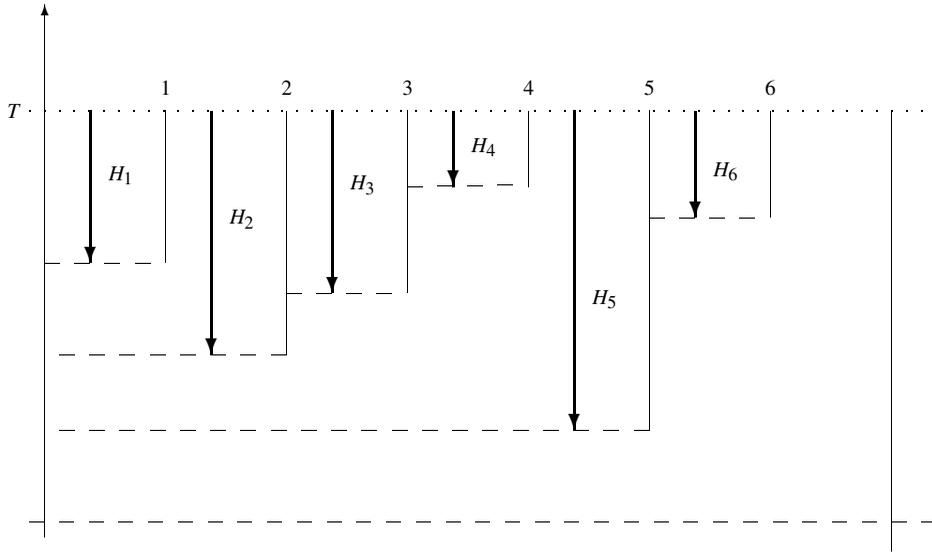
Figure 9 *A coalescent point process* (*upside down compared to previous pictures of combs*) *with* 6 *nonzero values* (*the* 7*th one is the first one larger than T*). *To recover an oriented ultrametric tree with* 7 *tips*, *draw horizontal lines from each tip left to the rightmost vertical line.*

Proposition 3.2.1. *The sphere* $\mathfrak{t}^T$ *of the splitting tree* $\mathfrak{t}$ *is non empty with probability* $\mathbb{P}_x(\tau_T^+ < \tau_0)$. *Conditional on being non-empty, it is isometric to the random comb* $\sum_{i \geq 1} H_i \mathbb{1}_{\{i\}}$, *where the* H_i's *form a sequence of i.i.d. random variables killed at its first value larger than T, and whose common distribution is given by*

$$\mathbb{P}(H_1 > t) = \mathbb{P}_T(\tau_{T-t} < \tau_T^+). \tag{3.1}$$

We say that $\mathfrak{t}^{\{T\}}$ *is* (*isometric to*) *a* coalescent point process (*CPP*).

Figure 9 shows a coalescent point process and how it codes for an ultrametric tree. By extension, we make the following definition.

Definition 3.2.2. Let ν be a σ-finite measure on $(0, \infty)$ such that $\nu([\varepsilon, \infty)) < \infty$ for all $\varepsilon > 0$. Let $\mathcal{M}$ be a Poisson point process on $(0, \infty)^2$ with intensity Leb $\otimes \nu$ and denote by $(S_i, H_i)_i$ its atoms. Finally, let (D, H) denote the first (i.e., smallest in the first dimension) atom such that $H > T$. We will say that the random comb metric space associated with the comb $\sum_{i:S_i < D} H_i \mathbb{1}_{\{S_i\}}$ is a *coalescent point process* with height T and intensity measure ν.

The term "coalescent point process" has first been coined in the setting of the Brownian tree by Popovic (2004). Recall that we called Brownian tree the tree coded by a positive Brownian excursion (CRT when the excursion length is normalized).

The Brownian excursion has a local time at level T, which allows one to construct as in Theorem 3.1.5 the comb giving the metric of the reduced tree at level T. This comb is a "list", in the plane order, of the depths of excursions of the contour away from T.

Theorem 3.2.3 (Popovic (2004), Aldous and Popovic (2005)). *Conditional on being non-empty, the sphere $\mathfrak{t}^{\{T\}}$ of the Brownian tree $\mathfrak{t}$ is a coalescent point process with height T and intensity measure ν, where ν is the push forward of the Brownian excursion measure by the depth mapping, that is*

$$\nu_0(dx) = \frac{dx}{2x^2}. \qquad (3.2)$$

More generally, as a follow up to Theorem 2.4.6, it can be shown that under suitable scaling, the coalescent point processes of (sub)critical splitting trees conditioned on reaching height T converge to the Poisson point process of excursion depths of a Brownian motion (with negative drift).

More specifically, under the assumptions (H1)–(H3) of Theorem 2.4.6, if $H_1^A, H_2^A, \ldots$ denote the coalescence times of the splitting tree conditioned on reaching height T, then the point processes $\sum_{i \geq 1} \delta_{(\frac{i}{A}, H_i^A)}$ converge as $A \to \infty$ to a CPP with height T and intensity measure ν, where

$$\nu_\alpha\big((x, \infty)\big) = \frac{\alpha}{1 - e^{-\alpha x / \beta}}. \qquad (3.3)$$

Formulae (3.1), (3.2) and (3.3) can actually all be rephrased in terms of the scale function W of the relevant contour process, as we will now see. We refer the reader to Bertoin (1996) and Kyprianou (2006) for more information about what follows.

Recall that a Lévy process X with no negative jumps is characterized by the Laplace transform of its one-dimensional marginals

$$\mathbb{E}\big(\exp(-\lambda X_t)\big) = \exp(t\psi(\lambda)), \qquad t, \lambda \geq 0,$$

where ψ is called the *Laplace exponent* of X. If $X_t = Y_t - t$ (where Y is the subordinator with Lévy measure Λ), the Lévy–Khinchin formula gives

$$\psi(\lambda) = \lambda - \int_{(0,\infty]} \big(1 - e^{-\lambda r}\big)\Lambda(dr), \qquad \lambda \geq 0.$$

If $X_t = -\alpha t + \sqrt{2\beta}B_t$, then $\psi(\lambda) = \alpha\lambda + \beta\lambda^2$.

Notice that $\psi(0) = 0$, that ψ is convex and has at most one positive root, here denoted η. Then there is a unique non-negative, increasing function W called the *scale function* such that

$$\int_0^\infty W(x)e^{-\lambda x}\,dx = \frac{1}{\psi(\lambda)}, \qquad \lambda \geq \eta. \qquad (3.4)$$

In the case when $X_t = -\alpha t + \sqrt{2\beta}\, B_t$, it is not difficult to compute $W(x) = x/\beta$ when $\alpha = 0$ and when $\alpha \neq 0$,

$$W(x) = \frac{1 - e^{-\alpha x/\beta}}{\alpha}, \qquad x \geq 0. \tag{3.5}$$

In general, one can prove (see Bertoin (1996)) that

$$W(x) = \exp\left\{ \int_0^x \underline{N}(\sup > s)\, ds \right\}, \qquad x \geq 0,$$

where $\underline{N}$ is the excursion measure of $X - \underline{X}$ away from 0. When X has finite variation, like when $X_t = Y_t - t$, $\underline{N}$ is merely the birth rate $b = \Lambda((0, \infty))$ times the law of X started from a random jump with law $b^{-1}\Lambda$.

In all cases, it is a consequence of the exponential formula for the Poisson point process of excursions of $X - \underline{X}$ away from 0 that

$$\mathbb{P}_x\big(\tau_0 < \tau_a^+\big) = \frac{W(a - x)}{W(a)}, \qquad 0 \leq x \leq a. \tag{3.6}$$

In the cases with infinite variation, such as Brownian motion (with or without drift), $W(0) = 0$ and

$$N(-\inf > x) = \frac{1}{W(x)}, \qquad x \geq 0, \tag{3.7}$$

where N is the excursion measure of X away from 0, so that equations (3.2) and (3.3) are a consequence of (3.5) and (3.7). In the case when $X_t = Y_t - t$, $W(0) = 1$, and we can be more accurate in equation (3.1) using equation (3.6)

$$\mathbb{P}(H_1 > t) = \mathbb{P}_T\big(\tau_{T-t} < \tau_T^+\big) = \frac{1}{W(t)},$$

where W can be identified inverting the Laplace transform (3.4).

Remark 3.2.4. In Lambert and Popovic (2013), we have expressed the distribution of the coalescent point process of non binary branching trees, including Galton–Watson processes and continuous-state branching processes. One of the main difficulties is to cope with the existence of points with arbitrarily large degree in the tree.

3.2.2 *A more general class of models.*   In this section, we seek to investigate population processes which generate trees whose spheres are (isometric to) CPPs.

Consider a population where all individuals live and reproduce independently, and each individual is endowed with a trait (some random character living in $\mathbb{R}$ for simplicity) that evolves through time according to independent copies of the same, possibly time-inhomogeneous, Markov process K with generator $L_t = L(t, \cdot)$. Further assume what follows.

- This trait is non-heritable, in the sense that any individual born at time t draws the value of her trait at birth from the same distribution ν_t, independently of her mother's history;
- All individuals give birth at the same, possibly time-inhomogeneous rate $b(t)$;
- An individual holding trait x at time t dies at rate $d(t, x)$.

Theorem 3.2.5 (Lambert and Stadler (2013)). *Under the previously defined population model, starting with one individual at time* 0 *and conditional on having at least one alive individual at time* T, *the reduced genealogical tree at level* T *is given by a coalescent point process with typical depth* H *whose inverse tail distribution is given by*

$$W(t) := \frac{1}{\mathbb{P}(H > t)} = \exp\left(\int_{T-t}^{T} b(s)\big(1 - q(s)\big)\,ds\right), \qquad t \in [0, T],$$

where $q(t)$ *denotes the probability that an individual born at time* t *has no descendants alive by time* T.

We will see later why the function W defined in the previous statement becomes the scale function of the last section when there is no time-inhomogeneity and the inheritable trait is the age.

In addition, W can be computed from the knowledge of g, where $g(t, \cdot)$ denotes the density of the death time of an individual born at time t.

Proposition 3.2.6 (Lambert and Stadler (2013)). *The function* W *is solution to the following integro-differential equation*

$$W'(t) = b(T - t)\left(W(t) - \int_0^t W(s)g(T - t, T - s)\,ds\right), \qquad t \geq 0, \qquad (3.8)$$

with initial condition $W(0) = 1$.

Proof. First, observe that

$$q(t) = \int_t^T g(t, s)e^{-\int_t^s b(u)(1 - q(u))\,du}\,ds.$$

Recalling that

$$W(t) = \frac{1}{\mathbb{P}(H > t)} = \exp\left(\int_{T-t}^{T} b(s)\big(1 - q(s)\big)\,ds\right),$$

we get

$$q(t) = \int_t^T g(t, s)\frac{W(T - s)}{W(T - t)}\,ds,$$

or equivalently

$$q(T - t) = \int_0^t g(T - t, T - s) \frac{W(s)}{W(t)} ds.$$

Now check that

$$W'(t) = b(T - t)(1 - q(T - t))W(t).$$

Equation (3.8) is a consequence of the last two equations. □

Remark 3.2.7. In general, g is not given directly in terms of the model ingredients b, d and the generator L_t of the trait process K. To compute g and then W, one can proceed as follows. Recall that $g(t, \cdot)$ is the density of the death time of an individual born at time t, so that

$$g(t, s) = \int_{\mathbb{R}} v_t(dx) u_s(t, x), \qquad s \geq t,$$

where $u_s(t, x)$ is that density conditional on the value x of the trait at birth ($K_t = x$), that is,

$$u_s(t, x) := \mathbb{E}_{t,x}\big(d(s, K_s)e^{-\int_t^s d(r, K_r)\, dr}\big), \qquad s \geq t, \tag{3.9}$$

where $\mathbb{E}_{t,x}$ denotes the expectation associated to the distribution of K started at time t in state x. Now the Feynman–Kac formula ensures that u_s is solution to

$$\frac{\partial u_s}{\partial t}(t, x) + L_t u_s(t, x) = d(t, x)u_s(t, x), \tag{3.10}$$

with terminal condition $u_s(s, x) = d(s, x)$. Specifically, when K is the age, the initial trait value is $x = 0$ and the age at s of a species born at t is $K_s = s - t$ so that

$$g(t, s) = d(s, s - t)e^{-\int_t^s d(r, r - t)\, dr}, \qquad s \geq t. \tag{3.11}$$

Proof of Theorem 3.2.5. Let $n \geq 1$ be an integer, and let $h_1, \ldots, h_{n-1}$ be elements of $(0, T)$. Assume $N_T \geq n$, and condition on $H_i = h_i$ for $i = 1, \ldots, n - 1$. We are going to prove that the conditional law of H_n is given by

$$\mathbb{P}(H_n > t) = \exp\Big(-\int_{T-t}^T b(s)(1 - q(s))\, ds\Big), \qquad t \in [0, T], \tag{3.12}$$

which will show that H_n is independent of $H_1, \ldots, H_{n-1}$ and has W as inverse tail distribution. This result yields the theorem by induction. Note that conditional on $N_T \geq n$, N_T exactly equals n iff $H_n > T$.

Label by $0, 1, \ldots, n - 1$ the individuals alive at time T in the order induced by the plane orientation of the tree, where daughters sprout to the right of their mother. In particular, h_i is the coalescence time between individuals $i - 1$ and i ($1 \leq i \leq n - 1$).

We denote by k the number of generations separating individuals $n-1$ from the progenitor. We let x_k denote the time when individual $n-1$ was born, $x_{k-1} < x_k$ the time when her mother was born, and so on, until $x_0 = 0$ the birth time of the progenitor. By the orientation of the tree, there are (conditionally) deterministic indices $0 = i_0 < \cdots < i_k = n-1$, such that $x_j = T - h_{i_j}$ (and $h_v < h_{i_j}$ for all $v \in \{i_{j-1}+1, \ldots, i_j - 1\}$), so that conditional on $H_i = h_i$ for $i = 1, \ldots, n-1$, the times $x_0, \ldots, x_k$ are deterministic.

By the orientation of the tree again, apart from the individuals already labelled, individuals alive at T descend from births occurring during one of the time intervals $I_j := [x_j, x_{j+1})$, where $x_{k+1} := T$ for convenience. On each of these time intervals, births occur at rate $b(t)$, and so by thinning, successful births, that is, births with alive descendance at time T, occur at rate $b(t)(1 - q(t))$. But conditional on the (x_j), all the ancestors of individual $n-1$ (including her) independently give birth on their corresponding interval I_j. Then if A denotes any subset of $[0, T)$, the number $N(A)$ of successful births occurring during A is the sum

$$N(A) = \sum_{j=0}^{k} N(A \cap I_j),$$

where the random numbers $N(A \cap I_j)$ are independent. Now $N(A \cap I_j)$ is a Poisson random variable with parameter $\int_{A \cap I_j} b(t)(1 - q(t)) \, dt$. As a consequence, $N(A)$ is a Poisson random variable with parameter

$$\sum_{j=0}^{k} \int_{A \cap I_j} b(t)(1 - q(t)) \, dt = \int_A b(t)(1 - q(t)) \, dt.$$

The proof finishes noticing that $H_n > t$ iff $N([T - t, T)) = 0$, which occurs with the probability stated in (3.12). $\qquad \square$

Exercise 3.2.8. When there is no trait dependence of the death rate, the process is merely a time-inhomogeneous birth–death process. Prove that in this case W is given by

$$W(t) = 1 + \int_{T-t}^{T} b(s) e^{\int_s^T r(u) \, du} \, ds,$$

where $r(t) := b(t) - d(t)$. In particular, when rates do not depend on time, the process is a linear birth–death process with birth rate b and death rate d. The last formula then boils down to

$$W(t) = \begin{cases} 1 + \dfrac{b}{r}(e^{rt} - 1) & \text{if } r \neq 0, \\ 1 + bt & \text{if } r = 0. \end{cases} \tag{3.13}$$

Figure 10 shows the density W'/W^2 (recall $W(t) = 1/\mathbb{P}(H \geq t)$) of coalescence times in the pure-birth case ($d = 0$) and in the critical case ($b = d$).
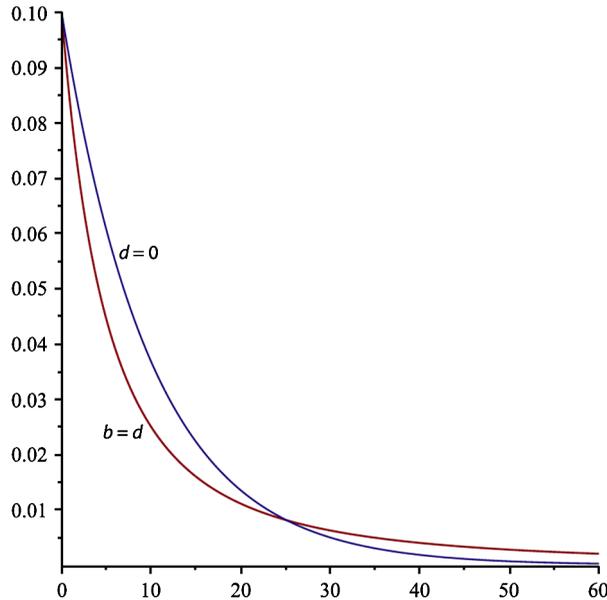
Figure 10   *The common density function $f = W'/W^2$ of the node depths of the reduced tree for a birth–death process with constant rates b and d. In blue, the pure birth case $f(t) = be^{-bt}$ ($b = 0.1$ in the figure); in red, the critical case $f(t) = bt/(1 + bt)^2$ ($b = d = 0.1$ in the figure). The critical process gives a density with a faster decay initially, but has a heavier tail than for the pure-birth model. This figure is taken from* Lambert and Steel (2013).

Exercise 3.2.9. When there is no time-dependence of the birth and death rates, and the trait is chosen to be the age, we are back to the splitting tree model. Here, $g(t, s) \equiv g(s - t)$ and $\Lambda(dr) = bg(r)\,dr$. First prove thanks to (3.8) that

$$W' = b(W - W \star g), \tag{3.14}$$

and then recover that the Laplace transform of W indeed is $1/\psi$ (as in equation (3.4)), where here ψ can be written as $\psi(\lambda) = \lambda - b \int_0^\infty (1 - e^{-\lambda r})g(r)(dr)$.

In passing, equation (3.14) offers to compute the pair (W, W') by solving numerically a 2D integro-differential equation (equation (3.14) along with $W(t) = 1 + \int_0^t W'(s)\,ds$) instead of inverting the Laplace transform of W, which can be numerically tricky.

Exercise 3.2.10. Prove that the shape of an ultrametric tree associated with a coalescent point process conditioned to have n tips is always P_n^{erm}.

3.3 Applications

In this section, we wish to give a taste of some recent applications of contour processes and coalescent point processes in evolutionary biology and in epidemiology. They rely in particular on the property that the density (or likelihood) of a given

ultrametric tree $\mathfrak{t}$ with age T and node depths (h_i), seen as the reduced tree of a tree generated under one of the models displayed in the last section, is in product form

$$\mathcal{L}(\mathfrak{t}) = p(T) \prod_i f(h_i),$$

where p and f have to be computed in terms of the model ingredients, in particular $f = W'/W^2$ is the density of a random node depth.

3.3.1 *Bottlenecks and missing tips.* Let $\mathfrak{t}$ be a rooted $\mathbb{R}$-tree interpreted as the genealogy of some population. We want to model the fact that at some fixed time point t, a macroscopic proportion say p of the population, is killed (and its entire descendance as well of course). In population genetics, such events are called *bottlenecks*, whereas in phylogenetics they model *mass extinctions*. When $t = T$ is present time, this procedure is meant to model *incomplete sampling* (or contemporary extinctions, see next section).

There are two reduced trees to consider, the reduced tree *ex ante*, spanned by individuals alive at time T in the absence of bottleneck, and the reduced tree *ex post*, spanned by individuals alive even in the presence of the bottleneck. The second one is of course included in the first one.

Note that the reduced tree *ex post* is not affected by lineages that do not even make it to the present in the absence of bottleneck, so we can equally assume that the bottleneck is only applied to the reduced tree *ex ante*. See Figure 11.

As soon as the reduced tree is compact it is a comb metric space, and if $t < T$ we can model the bottleneck by simply disconnecting each lineage of the reduced tree *ex ante* at distance t from the root, independently with probability p. The next statement ensures that this operation preserves the CPP property.

Proposition 3.3.1 (Lambert and Stadler (2013)). *Start with a CPP with inverse tail distribution W. Add bottlenecks with survival probabilities $\varepsilon_1, \ldots, \varepsilon_k$ at times $T - s_1 > \cdots > T - s_k$ (where $s_1 > 0$ and $s_k < T$). Then conditional on survival, the reduced tree* ex post *is again a coalescent point process, with inverse tail distribution W_ε given by*

$$W_\varepsilon(t) = \varepsilon_1 \cdots \varepsilon_m W(t) + \sum_{j=1}^m (1 - \varepsilon_j)\varepsilon_1 \cdots \varepsilon_{j-1} W(s_j), \tag{3.15}$$

$$t \in [s_m, s_{m+1}], 0 \le m \le k,$$

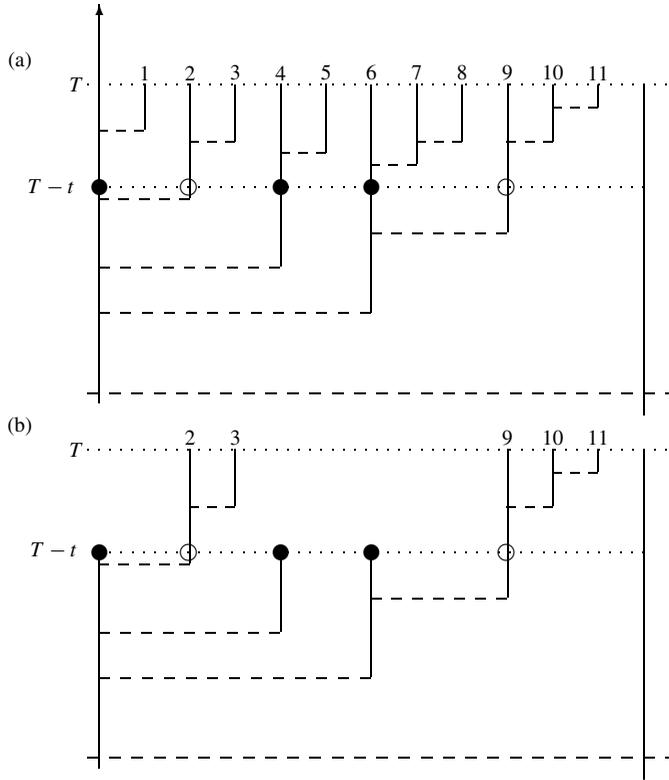*where $s_0 := 0$ and $s_{k+1} = T$ (empty sum is zero, empty product is 1).*

Figure 11 *Bottleneck at time point $T - t$, black dots disconnect lineages.* (a) *Coalescent point process* ex ante, *in the absence of bottleneck*; (b) *Reduced tree* ex post, *after passage of the bottleneck.*

In the case of a finite number of tips, this formula can also include sampling with probability p by adding a bottleneck with $s_0 = 0$ and $\varepsilon_0 = p$, resulting in

$$W_\varepsilon(t) = \varepsilon_0 \cdots \varepsilon_m W(t) + \sum_{j=0}^{m} (1 - \varepsilon_j)\varepsilon_0 \cdots \varepsilon_{j-1} W(s_j),$$

$$t \in [s_m, s_{m+1}], 0 \le m \le k,$$

which boils down to $W_\varepsilon = 1 - p + pW$ when $k = 0$ (since $W(0) = 1$).

Proof of Proposition 3.3.1. We characterize the effect of one bottleneck on a CPP with finitely many individuals at height T.

Assume $k = 1$ and $s_1 \in (0, T)$. Recall that a CPP is defined thanks to a sequence of independent, identically distributed random variables (H_i). We will see that

the tree obtained after thinning is still a coalescent point process, defined from independent random variables, say (B_i). Let (e_i) be the i.i.d. Bernoulli random variables defined by $e_i = 1$ if lineage i survives the bottleneck (this has a meaning only if $H_i \geq s_1$; it happens with probability ε_1). By the orientation of the tree, a tip terminating a pendant edge with depth smaller than s_1 is kept alive iff the rightmost pending edge on its left with depth larger than s_1 survives. As a consequence, if $H_1 < s_1$, then the first lineage is trivially still alive and its coalescence time with the left-hand ancestral lineage is $B := H_1$. Otherwise, define $1 = J_1 < J_2 < \cdots$ the indices of consecutive edges with depths larger than s_1. Then the first lineage kept alive after thinning is the least J_m such that $e_{J_m} = 1$, and its coalescence time with the ancestral lineage is $B := \max(H_{J_1}, \ldots, H_{J_m})$. By the independence property of coalescent point processes and by the independence of the Bernoulli random variables (e_i), the new genealogy is obtained by a sequence of independent random variables (B_i) all distributed as B.

Let us specify the law of B. First, with probability $P(H < s_1)$, $P(B \in \cdot) = P(H \in \cdot | H < s_1)$. Second, with probability $P(H \geq s_1)$

$$B \overset{(d)}{=} \max\{A_1, \ldots, A_M\},$$

where the A_i's are i.i.d. distributed as H conditional on $H \geq s_1$ and M is an independent (modified) geometric random variable, that is, $P(M = j) = \varepsilon_1 (1 - \varepsilon_1)^{j-1}$. Then for any $s \geq s_1$

$$\frac{1}{P(B \geq s)} = \frac{1 - \varepsilon_1}{P(H \geq s_1)} + \frac{\varepsilon_1}{P(H \geq s)}, \qquad s \geq s_1.$$

Then if W_ε denotes the inverse tail distribution of B, that is, $W_\varepsilon(s) := 1/P(B \geq s)$, we have

$$W_\varepsilon(s) = \begin{cases} W(s) & \text{if } 0 \leq s \leq s_1, \\ (1 - \varepsilon_1) W(s_1) + \varepsilon_1 W(s) & \text{if } s_1 \leq s \leq t, \end{cases}$$

where W is the inverse tail distribution of H. Iterating this procedure yields the result in Proposition 3.3.1. $\square$

3.3.2 *Loss of phylogenetic diversity.*   In the context of the contemporary crisis of biodiversity, conservation biologists have proposed to quantify the loss of evolutionary heritage by the sum of branch lengths that disappear from the Tree of Life as new extinctions occur, that is, evolutionary heritage of a clade is quantified by the sum of its branch lengths, called *phylogenetic diversity* (PD). Then a natural question to ask is the following. If a random, say 10% of species from some given clade were to disappear in the next 100 years due to current high rates of extinction, how much evolutionary heritage would be lost?

"Not so much", asserted Nee and May (1997) in a very much debated paper, where the tree of life was modeled by Kingman coalescent. "A lot more", replied Mooers et al. (2012), in a paper where the tree of life was modeled by a Yule tree.

In Lambert and Steel (2013), we generalized their calculations to the case of a splitting tree with age T and typical node depth H, where tips (contemporary species) are eliminated independently with probability $1 - p$ ("field of bullets" model). Let G denote a geometric random variable with success probability p, let (A_i) be a sequence of independent copies of H conditioned on $H \leq T$ and set

$$B := \max_{i=1,...,G} A_i,$$

that is B is the typical node depth of the tree after passage of the field of bullets (see previous paragraph on bottlenecks).

Conditional on the number of tips n of the tree *ex ante* and on the number k_n of the tree *ex post*, as $n \to \infty$ and $k_n/n \to p$, elementary SLLN-type arguments show that the ratio of the remaining PD to the old PD converges a.s. to

$$\pi_T(p) = p \frac{\mathbb{E}(B)}{\mathbb{E}(A)}.$$

The ratio of remaining to old PD is always above the identity, corresponding to the case when the tree is star-like. This obviously holds also for π_T. In addition, it is not difficult to see that π_T is always a concave, increasing function such that $\pi_T(0) = 0$ and $\pi_T(1) = 1$.

If we take $T = \infty$, we get

$$\pi_\infty(p) = p \frac{\int_0^\infty \frac{dt}{1-p+pW(t)}}{\int_0^\infty \frac{dt}{W(t)}}.$$

In the case of a birth–death tree with rates b and d, with $r := b - d > 0$, we get

$$\pi_\infty(p) = \begin{cases} \dfrac{dp}{bp-r} \dfrac{\ln(bp/r)}{\ln(b/r)} & \text{if } b > r \neq bp; \\[2ex] -\dfrac{p\ln(p)}{1-p} & \text{if } b = r > bp; \\[2ex] -\dfrac{1-p}{\ln(p)} & \text{if } b > r = bp. \end{cases}$$

Figure 12 shows the graph of π_∞ for a range of birth and death rates $b > d$. Note that the more concave the better for evolutionary heritage. Also the larger $d/b < 1$, the larger the remaining fraction of phylogenetic diversity, converging, but very slowly, to 1 as $d/b \to 1$.

3.3.3 *Do species age?*   In Lambert, Alexander and Stadler (2014a), we have developed a framework to test the assumption that the extinction rate of a species remains constant all the way through its lifetime. Specifically, we have applied a maximal likelihood procedure to the recently published bird phylogeny (Jetz et al. (2012)), to infer the lifetime distribution of bird species, assuming they are Gamma distributed. We tested the assumption that the shape parameter k of the
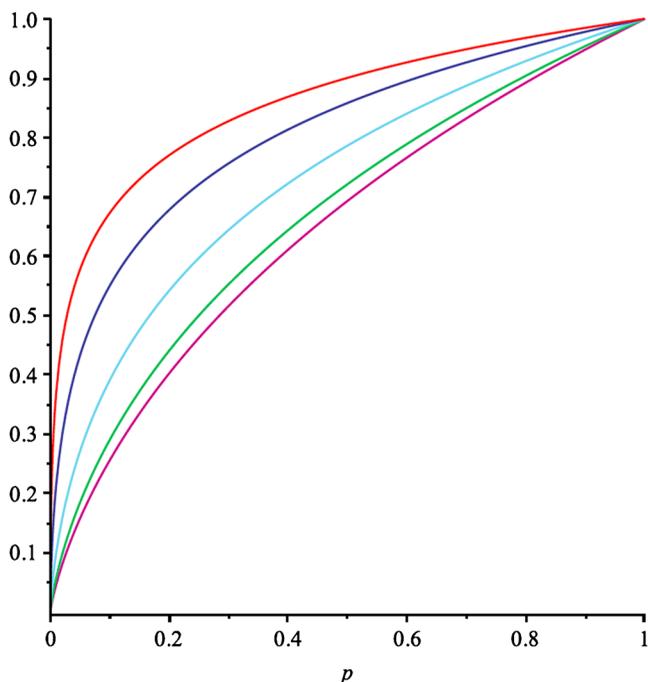
Figure 12   *Remaining fraction π_∞ of phylogenetic diversity (PD) as a function of the probability p of species survival to the mass extinction, for a constant-rate birth–death process. Observe the slow progression toward the unit step function (from pure birth to critical): $d/b = 0$ (the lowest curve) and then $d/b = 0.5, 0.9, 0.99, 0.999$. This figure is taken from* Lambert and Steel (2013).

Gamma r.v. equals 1 (exponential lifetimes, i.e., age-independent extinction rate) vs $k \neq 1$ (age-dependence). This study generalizes previous works on the inference of diversification processes from reconstructed phylogenies (e.g., Nee, May and Harvey (1994), Nee (2006), Stadler (2011)).

Our estimate of k is much larger than 1, indicating that the extinction rate is not constant but increases with age. For the record, our estimate of the speciation rate in birds is $0.11 \ My^{-1}$ and our estimate of the mean bird species lifetime is $15 \ My$.

3.3.4 *How long does speciation take?*   In most models of diversification, like the previous one, species are seen as particles that split instantaneously into two daughter species upon speciation. It is obvious that on the contrary speciation takes time, but the last assumption would still be relevant if the time speciation takes was negligible compared to a species lifetime. There is evidence that this is not the case in nature, and some authors have recently proposed an alternative model of diversification, called *protracted speciation*, meant to take this effect into account.

In this model, a species is described as an ensemble of populations, and as time passes, these populations diverge (genetically) gradually from each other. To not have to record all (phylo)genetic distances between all populations composing
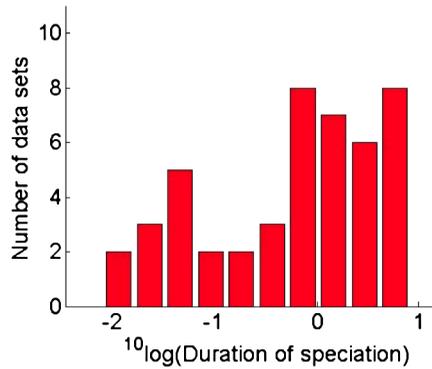
Figure 13 *Duration of speciation* (*in My*). *Distribution inferred from* 46 *bird clades*, *ranging from* 10,000 *years to* 10 *My. This figure is taken from* Etienne, Morlon and Lambert (2014).

each species, Etienne and Rosindell (2012) have proposed a model where each population passes through *k* different stages of maturation, after which it becomes a so-called good species. This model produces quite realistic patterns in terms of phylogenetic balance and branching tempo, but an inference framework was missing.

Since speciation stage is a non-heritable trait, Theorem 3.2.5 ensures that if speciation rate does not depend on speciation stage, the phylogeny produced by this model is a CPP. In Lambert, Morlon and Etienne (2014b), we have characterized the common distribution of node depths in this CPP.

In Etienne, Morlon and Lambert (2014), we have developed an inference framework that we tested against 46 bird clades. Individual parameters are difficult to infer, but the method is relatively good at inferring a composite parameter of interest, the duration of speciation. The duration of speciation is defined as the time it takes for a novel population to get a good species in its descendance. The results are shown in Figure 13.

3.3.5 *Trees with random marks*.

In this section, trees are endowed with marks to model sampling, detection or mutation.

The phylogeny of pathogens. In this paragraph, we focus on a population of patients carrying (or not) an infectious disease. We stick to the framework of splitting trees, with the interpretation that a birth is a transmission event and a death is either a real death or the end of the infectious period (exit from the infective population). The branching property assumption is justified in the case of a well-mixed population where susceptibles are always in excess.

Our setting where lifetime distributions are not exponential is mostly suited to diseases like flu or HIV, where the infectious period is known to be deterministic (flu) or heavy-tailed (HIV).

In addition, we assume that patients are detected to be infective after some random time D, at which they are tagged by a mark (see figures). Upon detection (if D is smaller than the infectious period/lifetime), the patients are assumed to exit the infective population, for example because they change their behaviour to avoid transmission or because they are treated and become non-infectious.

In Lambert, Alexander and Stadler (2014a), we have considered the case where the transmission tree spanned by detected patients has been reconstructed, similarly as Stadler (2010). Actually, the data is not directly the transmission tree but the phylogenetic tree of the pathogens carried by patients, reconstructed thanks to biological samples taken from the patients upon detection. Figure 14 shows the oriented tree of the epidemic, with black dots showing detection events, along with its reduced tree and contour process. Patients can be labelled in the plane order, so we can define S_i the detection time of patient i and R_i the coalescence time between patients $i - 1$ and i.

By considering the jumping contour process of the epidemics, we have shown that the sequence (S_i, R_i) is a killed Markov chain. The likelihood of a tree $\mathfrak{t}$ with coalescence times $(x_i)_{2 \leq i \leq n}$ and detection times $(y_i)_{1 \leq i \leq n}$ can be written in the
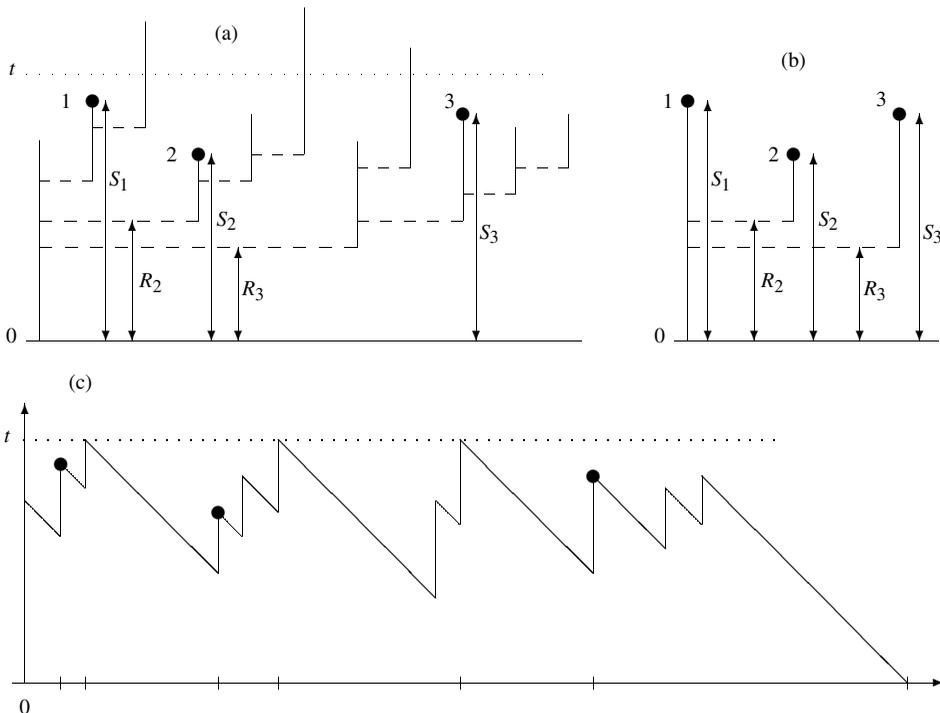

Figure 14 *The transmission tree*. (a) *Binary tree with marks* (*detection events*); (b) *its reduced tree and* (c) *its contour process.*

form

$$\mathscr{L}(\mathbb{t}) = g(y_1)k(y_n)\prod_{i=2}^{n} f(y_{i-1}; x_i, y_i),$$

where g, k and f can be semi-explicitly expressed in terms of the model ingredients (law of infectious lifetimes, transmission rate, detection rate).

The state of the epidemics at first detection. We have considered the same setting in Lambert and Trapman (2013), but for a different question, namely the structure of the epidemic at the first detection time. This question arises in the case of hospital-borne diseases due for example to bacterial antibiotic resistance. In this situation, everything can be known about all carriers of the disease, but only after the first detection of a case. At this random time, denoted T, everybody in the hospital is scanned and infected individuals are identified. Here the phylogeny of pathogens is not assumed to be known but patients' data like entrance dates or durations of stays are precisely known.

We assume that patients have i.i.d. lengths of stay in the hospital, all distributed as some r.v. K. Conditional on infection, the length of stay of a patient is supposed to be a size-biased version of K. Finally, the transmission rate is b and the detection rate per patient is denoted δ.

For individual i, set

- $U_i :=$ time elapsed from entrance of the hospital up to infection.
- $A_i :=$ time elapsed from infection up to T.
- $R_i :=$ residual lifetime in the hospital after T.

See Figure 15 for an example. Set $m := \mathbb{E}(K)$ and let ϕ denote the inverse of the convex function

$$x \mapsto x - \frac{b}{m}\int_{(0,\infty]} (1 - e^{-xy})\mathbb{P}(K > y)\,dy.$$

Using Vervaat's transform applied to the path of the contour process, we were able to show that conditional on $N_T = n$, the triples (U_i, A_i, R_i) of the n infectives at time T are i.i.d. distributed as

$$\mathbb{E}\big(f(U, A, R)\big)$$
$$= \frac{b}{m}\frac{\phi(\delta)}{\phi(\delta) - \delta}\int_{u=0}^{\infty} du \int_{a=0}^{\infty} da \int_{z=u+a}^{\infty} \mathbb{P}(K \in dz)$$
$$\times e^{-\phi(\delta)a} f(u, a, z - u - a).$$

In particular, the times $J_i = U_i + A_i$ spent in the hospital up to time T are i.i.d. distributed as the r.v. J

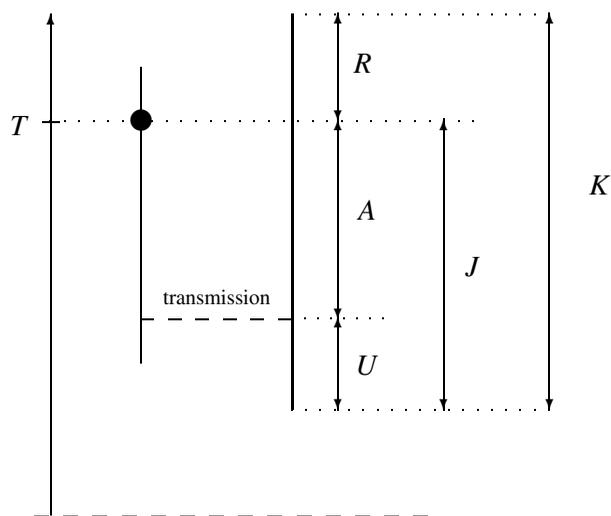$$\mathbb{P}(J \in dy) = \frac{b/m}{\phi(\delta) - \delta}\mathbb{P}(K > y)\big(1 - e^{-\phi(\delta)y}\big)\,dy.$$

Figure 15 *The structure of the stay in the hospital of an infected patient. Some* (*other*) *patient is detected at time T* (*random*). *The focal patient has total duration of stay $K = U + A + R$ (see text).*

The last formulae will allow us to infer the dynamical characteristics of hospital-borne disease epidemics from hospital data. This contrasts with the fact that inference is impossible from the sole numbers of cases found upon detection (Trapman and Bootsma (2009)), due to the geometric distribution of N_T (recall however that here T is random).

Mutations. Marks on a tree can also be used to model mutations. In population genetics, it is standard to assume that each new mutation occurs at a new site of the DNA sequence, the infinitely-many-sites assumption. The list of mutated sites of a sequence is called allele. One of the fundamental questions in population genetics is to interpret genetic data such as the number of individuals in a population carrying a specific mutation or a specific allele. Reciprocally, the number of mutations, or of alleles, carried by k individuals in the population is called frequency spectrum by sites, or by alleles.

The frequency spectrum of neutral mutations (that is, mutations with no influence on the population dynamics) has been extensively studied for random genealogies arising from models with constant population size, culminating in so-called Ewens' sampling formula (Ewens (1972)). In a series of recent papers relying heavily on contour techniques, we have studied the frequency spectrum in branching genealogies (Lambert (2009, 2011), Champagnat and Lambert (2012, 2013), Richard (2014), Delaporte, Achaz and Lambert (2016)).

3.4 Perspectives

For who has understood how to identify which forward-in-time processes generate trees whose reduced tree is a coalescent point process, and has learnt the procedure of characterizing W and the coalescent density from the model ingredients, coalescent point processes are a very convenient tool:

- They arise in a wide class of models;
- They generate robust patterns, that in particular are invariant under incomplete sampling and under the action of bottlenecks;
- The reconstructed tree has a particularly simple distribution, which is extremely fast to simulate, in contrast to the entire forward-in-time process, that may even not be Markovian;
- The inference of model parameters from the knowledge of the reconstructed tree can be done using low-tech statistical methods.

On the other hand, CPPs also have a number of shortcomings:

- The models in which CPP arise exclude some interesting features from the modeling point of view, in particular the trait/age-dependence of birth rates;
- Among the robust patterns they generate, the shape of the reduced tree is always ERM, which is certainly not the rule in empirical genealogies/phylogenies;

Currently, one of our main lines of research (not only mine but more generally that of the SMILE group—*Stochastic models for the inference of life evolution*, UPMC & Collège de France) is to produce and study models that

1. Are grounded on the microscopic description of individuals, either at the ecological scale or at the genetic scale;
2. Feature a small number of parameters, which can nevertheless be tuned so as to generate a wide range of different patterns, when the corresponding empirical patterns vary across datasets (e.g., species abundance distributions);
3. Generate robust patterns when the corresponding empirical patterns are well conserved across datasets (e.g., the MLE of β in empirical phylogenies, that revolves around -1);
4. Produce observable statistics (e.g., reconstructed trees) with computable likelihoods.

   Criterion 3 is in general difficult to satisfy, especially simultaneously with 2.

References

Aldous, D. (1991). The continuum random tree. I. *The Annals of Probability* **19**, 1–28. MR1085326

Aldous, D. (1993). The continuum random tree. III. *The Annals of Probability* **21**, 248–289. MR1207226

Aldous, D. (1996). Probability distributions on cladograms. In *Random Discrete Structures* (A. Friedman, W. Miller, D. Aldous and R. Pemantle, eds.) **76** 1–18. New York: Springer. MR1395604

Aldous, D. and Popovic, L. (2005). A critical branching process model for biodiversity. *Advances in Applied Probability* **37**, 1094–1115. MR2193998

Aldous, D. J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science* **16**, 23–34. MR1838600

Barthélémy, J.-P. and Guénoche, A. (1991). *Trees and Proximity Representations*. New York: Wiley. MR1138723

Bertoin, J. (1996). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge Univ. Press. MR1406564

Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. *Cambridge Studies in Advanced Mathematics* **102**. Cambridge: Cambridge Univ. Press. MR2253162

Blum, M. G. and François, O. (2006). Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology* **55**, 685–691.

Brown, J. K. M. (1994). Probabilities of evolutionary trees. *Systematic Biology* **43**, 78–91.

Burago, D., Burago, Y. and Ivanov, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Providence, RI: American Mathematical Society. MR1835418

Champagnat, N. and Lambert, A. (2012). Splitting trees with neutral Poissonian mutations I: Small families. In *Stochastic Processes and Their Applications* **122**, 1003–1033. MR2891445

Champagnat, N. and Lambert, A. (2013). Splitting trees with neutral Poissonian mutations II: Largest and oldest families. *Stochastic Processes and their Applications* **123**, 1368–1414. MR3016227

Delaporte, C., Achaz, G. and Lambert, A. (2016). Mutational pattern of a sample from a critical branching population. *Journal of Mathematical Biology Journal of Mathematical Biology* **73**, 627–664. MR3535416

Dress, A., Moulton, V. and Terhalle, W. (1996). T-theory: An overview. *European Journal of Combinatorics* **17**, 161–175. MR1379369

Duquesne, T. (2006). The coding of compact real trees by real valued functions. Preprint. Available at arXiv:math/0604106.

Duquesne, T. and Le Gall, J.-F. (2002). Random trees, Lévy processes and spatial branching processes. *Asterisque—Société Mathématique de France* **281**. Paris: Société Mathématique de France.

Etienne, R. S., Morlon, H. and Lambert, A. (2014). Estimating the duration of speciation from phylogenies. *Evolution* **68**, 2430–2440.

Etienne, R. S. and Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. *Systematic Biology* **61**, 204–213.

Evans, S. N. (2008). *Probability and Real Trees*. *Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Lecture Notes in Mathematics* **1920**. Berlin: Springer. MR2351587

Evans, S. N., Pitman, J. and Winter, A. (2005). Rayleigh processes, real trees, and root growth with re-grafting. *Probability Theory and Related Fields* **134**, 81–126. MR2221786

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112. Erratum *Theoretical Population Biology* **3** 240, 376. MR0325177

Geiger, J. (1996). Size-biased and conditioned random splitting trees. In *Stochastic Processes and Their Applications* **65**, 187–207. MR1425355

Geiger, J. and Kersting, G. (1997). Depth-first search of random trees, and Poisson point processes. In *Classical and Modern Branching Processes Minneapolis, MN, 1994. IMA Vol. Math. Appl.* **84**, 111–126. New York: Springer. MR1601713

Haas, B. (2016). Scaling limits of Markov-branching trees and applications. Preprint. Available at arXiv:1605.07873.

Haas, B., Miermont, G., Pitman, J. and Winkel, M. (2008). Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability* **36**, 1790–1837. MR2440924

Hagen, O., Hartmann, K., Steel, M. and Stadler, T. (2015). Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic Biology* **64**, 432–440.

Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability* **3**, 44–77. MR0282451

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444–448.

Kingman, J. (1982). The coalescent. *Stochastic Processes and Their Applications* **13**, 235–248. MR0671034

Knuth, D. E. (1997). *The Art of Computer Programming*. Reading, MA: Addison-Wesley. MR0378456

Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Berlin: Springer. MR2250061

Lambert, A. (2008). Population dynamics and random genealogies. *Stochastic Models* **24**, 45–163. MR2466449

Lambert, A. (2009). The allelic partition for coalescent point processes. *Markov Processes and Related Fields* **15**, 359–386. MR2554367

Lambert, A. (2010). The contour of splitting trees is a Lévy process. *The Annals of Probability* **38**, 348–395. MR2599603

Lambert, A. (2011). Species abundance distributions in neutral models with immigration or mutation and general lifetimes. *Journal of Mathematical Biology* **63**, 57–72. MR2806489

Lambert, A., Alexander, H. K. and Stadler, T. (2014a). Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *Journal of Theoretical Biology* **352**, 60–70. MR3198053

Lambert, A., Morlon, H. and Etienne, R. S. (2014b). The reconstructed tree in the lineage-based model of protracted speciation. *Journal of Mathematical Biology* **70**, 367–397. MR3294977

Lambert, A. and Popovic, L. (2013). The coalescent point process of branching trees. *Annals of Applied Probability* **23**, 99–144. MR3059232

Lambert, A., Simatos, F. and Zwart, B. (2013). Scaling limits via excursion theory: Interplay between Crump–Mode–Jagers branching processes and processor-sharing queues. *The Annals of Applied Probability* **23**, 2357–2381. MR3127938

Lambert, A. and Stadler, T. (2013). Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology* **90**, 113–128.

Lambert, A. and Steel, M. (2013). Predicting the loss of phylogenetic diversity under non-stationary diversification models. *Journal of Theoretical Biology* **337**, 111–124. MR3115320

Lambert, A. and Trapman, P. (2013). Splitting trees stopped when the first clock rings and Vervaat's transformation. *Journal of Applied Probability* **50**, 208–227. MR3076782

Lambert, A. and Uribe Bravo, G. (2016a). The comb representation of compact ultrametric spaces. Preprint. Available at arXiv:1602.08246.

Lambert, A. and Uribe Bravo, G. (2016b). Totally ordered, measured trees and splitting trees with infinite variation. Preprint. Available at arXiv:1607.02114.

Le Gall, J.-F. (1993). The uniform random tree in a Brownian excursion. *Probability Theory and Related Fields* **96**, 369–383. MR1231930

Le Gall, J.-F. (2005). Random trees and applications. *Probability Surveys* **2**, 245–311. MR2203728

Le Gall, J.-F. and Miermont, G. (2012). Scaling limits of random trees and planar maps. In *Probability and Statistical Physics in Two and More Dimensions* (D. Ellwood, ed.). *Clay Math. Proc.* **15**, 155–211. Providence, RI: American Mathematical Society. MR3025391

Manceau, M., Lambert, A. and Morlon, H. (2015). Phylogenies support out-of-equilibrium models of biodiversity. *Ecology Letters* **18**, 347–356.

Mooers, A., Gascuel, O., Stadler, T., Li, H. and Steel, M. (2012). Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Systematic Biology* **61**, 195–203.

Murtagh, F. (1984). Counting dendrograms: A survey. *Discrete Applied Mathematics* **7**, 191–199. MR0727923

Nee, S. (2006). Birth-death models in macroevolution. *Annual Review of Ecology, Evolution and Systematics* **37**, 1–17.

Nee, S., May, R. and Harvey, P. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **344**, 305–311.

Nee, S. and May, R. M. (1997). Extinction and the loss of evolutionary history. *Science* **278**, 692–694.

Paulin, F. (1989). The Gromov topology on *R*-trees. *Topology and its Applications* **32**, 197–221. MR1007101

Pitman, J. (2006). *Combinatorial Stochastic Processes*. In *Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics* **1875**. Berlin: Springer. MR2245368

Popovic, L. (2004). Asymptotic genealogy of a critical branching process. *Annals of Applied Probability* **14**, 2120–2148. MR2100386

Richard, M. (2014). Splitting trees with neutral mutations at birth. In *Stochastic Processes and Their Applications* **124**, 3206–3230. MR3231617

Semple, C. and Steel, M. A. (2003). *Phylogenetics*. *Oxford Lecture Series in Mathematics and its Applications* **24**. Oxford: Oxford Univ. Press. MR2060009

Slowinski, J. B. (1990). Probabilities of *n*-trees under two models: A demonstration that asymmetrical interior nodes are not improbable. *Systematic Biology* **39**, 89–94.

Stadler, T. (2010). Sampling-through-time in birth–death trees. *Journal of Theoretical Biology* **267**, 396–404. MR2974417

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

Stanley, R. P. (1999). *Enumerative Combinatorics*. *Cambridge Studies in Advanced Mathematics* **62**. Cambridge: Cambridge Univ. Press. MR1676282

Trapman, P. and Bootsma, M. C. J. (2009). A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection. *Mathematical Biosciences* **219**, 15–22. MR2518217

Stochastics & Biology Group
Laboratoire de Probabilités et Modèles Aléatoires
Box 188
UPMC Univ Paris 06
4 Place Jussieu
75252 Paris Cedex 05
France
E-mail: amaury.lambert@upmc.fr
URL: http://www.lpma-paris.fr/pageperso/amaury.lambert/