The tree length of an evolving coalescent

by P. Pfaffelhuber[•,*,†], A. Wakolbinger[◦,*] and H. Weisshaupt[•,†]

Albert-Ludwigs University, Freiburg[•] and Goethe University, Frankfurt[◦]

August 11, 2009

Abstract

A well-established model for the genealogy of a large population in equilibrium is Kingman's coalescent. For the population together with its genealogy evolving in time, this gives rise to a time-stationary tree-valued process. We study the sum of the branch lengths, briefly denoted as tree length, and prove that the (suitably compensated) sequence of tree length processes converges, as the population size tends to infinity, to a limit process with càdlàg paths, infinite infinitesimal variance, and a Gumbel distribution as its equilibrium.

1 Introduction

Kingman's coalescent [Kin82a, Kin82b] is a widely used model for the single-locus genealogy in a population, see [Wak08] and references therein. It arises in a suitable rescaling of time under the assumptions of a neutral evolution and an exchangeable reproduction dynamics with short-tailed offspring distribution. An intuitive way to think of Kingman's coalescent is to imagine a random tree with infinitely many leaves at time t, where backwards in time any two lineages independently coalesce at rate 1. Taking N instead of infinitely many leaves gives Kingman's N-coalescent. The latter figures as the genealogy of an N-sample taken from a large population, and also as the genealogy of the total population in a standard Moran model with population size N.

Two functionals of coalescent trees are of particular interest: the distance from the root to the leaves, or *depth*, and the sum of branch lengths, or *tree length*. It is well known that the expected depth of Kingman's N-coalescent equals $2(1 - \frac{1}{N})$, whereas its expected tree length is $\sim 2 \log N$ as $N \to \infty$. More can be said: when compensated by $2 \log N$, half the tree length of Kingman's N-coalescent converges in law to a Gumbel distributed random variable (having the cumulative distribution function $x \mapsto e^{-e^{-x}}$). This result can be read off from [Tav84, p. 153, first equation]; see also [WH99], [Tav04], [DIMR07] and [Wak08].

With individual offspring distributions that are not short-tailed, coalescents different from Kingman's appear as the genealogies of large populations. In the so-called Λ-coalescents [Pit99], more than two lines can coalesce, giving rise to *multiple mergers*, and asymptotic tree length distributions arise that are different from Gumbel distributions. For special classes including those of Beta-coalescents, results on the asymptotic tree length were obtained in [Möh06], [DIMR07], [BBS08] and [DDSJ09a].

With a population evolving in time, its genealogical relationships evolve as well. Their evolution is described by a *tree-valued process* ([GPW09]) similarly as the change of allele frequencies is

captured by measure-valued diffusions ([Daw93, Eth01]). Jumps of the tree depth correspond to the loss of one of the currently two oldest families from the population, and hence to the establishment of a new most recent common ancestor (MRCA) of the population. The resulting tree depth process in the case of Kingman's coalescent was analyzed in [PW06] and [DDSJ09b].

In the present paper we focus on the (compensated) tree length in Kingman's coalescent and describe its evolution in an infinite population. Our main result is that this process has càdlàg paths and infinite infinitesimal variance (Theorem 1). As already stated, the one-dimensional projections of this process are Gumbel distributed.

We construct the process of compensated tree length, denoted by $\mathscr{L}$, as a limit using tools from weak convergence of processes. In addition, we also provide a strong convergence result, i.e. a version of Theorem 1 in terms of convergence towards $\mathscr{L}$ in probability. For this we use the lookdown process introduced in [DK99], which provides genealogies of Moran models of any population size on one and the same probability space. Our Proposition 3.2 shows that on this space the compensated Kingman tree lengths lead to a càdlàg path-valued limit in probability.

The length of a coalescent is of relevance in empirical population genetics ([Wak08]). In the infinite sites model the number of mutations seen in a population of size N at time t is Poisson distributed with parameter proportional to tree length and to the mutation rate. This leads to an estimator of the mutation rate in an equilibrium population based on the expected tree length, known as Watterson's estimator ([Wat75]).

The process of tree lengths has attracted interest in the study of diversity in real populations [RBY04, Fig. 2c]. There, sudden losses of diversity in a population are related to jumps of the tree length process. Such jumps occur at any resampling event and correspond to the length of an external branch breaking off the tree. The asymptotics of external branch lengths are investigated in detail in [CNKR07], see also Remark 2.2 and Section 4.2.

Our results concern convergence of the tree length processes of evolving Kingman N-coalescents as $N \to \infty$, and properties of the limiting tree length process $\mathscr{L}$. The process $\mathscr{L}$ can be defined directly in terms of a sequence Moran models (or in terms of the lookdown graph), and as such is a natural object to study. Some challenging questions remain, e.g. a) Is the limit robust in the sense that $\mathscr{L}$ describes also the limiting tree length process for (a large class of) Cannings models with short-tailed offspring distributions? b) Is there an intrinsic characterization of $\mathscr{L}$ in terms of a stochastic dynamics? In particular, is $\mathscr{L}$ a semimartingale?

Our paper is organized as follows. After specifying the model we present our results on weak (Section 2) and strong (Section 3) convergence of tree lengths and tree length processes. In Section 4 we provide some auxiliary results on Kingman's coalescent for fixed times. Section 5 completes the proof of Theorem 1, and Section 6 contains the proofs of the 'strong convergence' results, Propositions 3.1 and 3.2.

2 Convergence of tree length distributions

Consider a *Moran model* with constant population size N, started at time $-\infty$.[1] Each (unordered) pair of individuals *resamples* at rate 1; in any such resampling event, one of the two individuals reproduces and the other one dies. See Figure 1(A) for an illustration.

At any time $t \in \mathbb{R}$, the common ancestry of all individuals in the population is described by a random genealogical tree, which is Kingman's N-coalescent [Kin82a]. With time t varying, we obtain a tree-valued process denoted by $\mathscr{T}^{\mathrm{N}} = (\mathscr{T}^{\mathrm{N}}_t)_{t \in \mathbb{R}}$, whose random path we can read off from the graphical representation, see Figure 1(B) and (C).

Let $\widetilde{\ell}$ be the map that sends a (finite) tree to its length, i.e. to the sum of the lengths of all branches. Back from a fixed time t, each (unordered) pair of ancestral lines coalesces at rate 1, therefore the length of the time interval during which the genealogical tree $\mathscr{T}^{\mathrm{N}}_t$ has k

---

[1]We use roman upper case letters K, N, ... for real (non-random) numbers in order to distinguish them from random variables denoted by K, N, ...
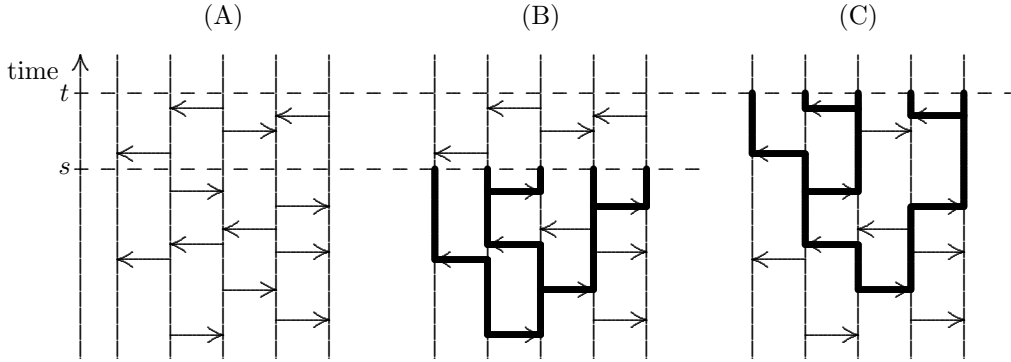
Figure 1: (A) The graphical representation of a Moran model of size N = 5. Arrows between lines indicate resampling events. By resampling the genealogical relationships between individuals change. The individual at the tip dies and the other one reproduces. (B) The genealogical relationships at time s can be read off from the graphical representation. The vertical bold lines constitute the genealogical tree $\mathscr{T}_s^{\mathrm{N}}$. (C) By time t the genealogical tree has changed to $\mathscr{T}_t^{\mathrm{N}}$. However, some parts of the tree $\mathscr{T}_t^{\mathrm{N}}$ were already present in the tree $\mathscr{T}_s^{\mathrm{N}}$.

lines is exponentially distributed with the number of pairs, $\binom{k}{2}$, as parameter. Consequently, the expectation and the variance of the tree length are

$$\mathbb{E}[\widetilde{\ell}(\mathscr{T}_t^{\mathrm{N}})] = \sum_{i=2}^{\mathrm{N}} i \frac{1}{\binom{i}{2}} = 2 \sum_{i=1}^{\mathrm{N}-1} \frac{1}{i}, \qquad \mathbb{V}[\widetilde{\ell}(\mathscr{T}_t^{\mathrm{N}})] = \sum_{i=2}^{\mathrm{N}} i^2 \frac{1}{\binom{i}{2}^2} = 4 \sum_{i=1}^{\mathrm{N}-1} \frac{1}{i^2}.$$

We are going to study the *compensated tree length process*

$$\mathscr{L}^{\mathrm{N}} := \left(\widetilde{\ell}(\mathscr{T}_t^{\mathrm{N}}) - 2 \log \mathrm{N}\right)_{t \in \mathbb{R}} \tag{2.1}$$

in the limit $\mathrm{N} \to \infty$.

One realization of the process $\mathscr{L}^{100}$ can be seen in Figure 2. Here, several large jumps of the tree length can be observed. Particularly large jumps in the tree length arise when the MRCA of the total population changes.

2.1 Tree lengths at fixed times

We recall a basic fact about the asymptotics of the law of $\mathscr{L}_t^{\mathrm{N}}$ as $\mathrm{N} \to \infty$.

Proposition 2.1 (Tree lengths for fixed times)**.** *For $t \in \mathbb{R}$, the law of $\frac{1}{2}\mathscr{L}_t^{\mathrm{N}}$ converges as $\mathrm{N} \to \infty$ weakly to the standard Gumbel distribution with cumulative distribution function $x \mapsto e^{-e^{-x}}$.*

Proof. We briefly repeat the argument from [WH99, p. 255]. Let $X_2, X_3, \dots$ be independent random variables such that X_j has an exponential distribution with rate $\binom{j}{2}$. In addition, let $Y_1, Y_2, \dots$ be independent such that Y_j has an exponential distribution with rate j and $Z_1, Z_2, \dots$ be independent exponential, each with parameter 1. Then

$$\tfrac{1}{2}\widetilde{\ell}(\mathscr{T}_t^{\mathrm{N}}) \stackrel{d}{=} \tfrac{1}{2} \sum_{j=2}^{\mathrm{N}} j X_j \stackrel{d}{=} \sum_{j=1}^{\mathrm{N}-1} Y_j \stackrel{d}{=} \max_{1 \le j \le \mathrm{N}-1} Z_j,$$

which when shifted by $\log \mathrm{N}$ has the asserted limit in distribution as $\mathrm{N} \to \infty$. $\qquad\square$
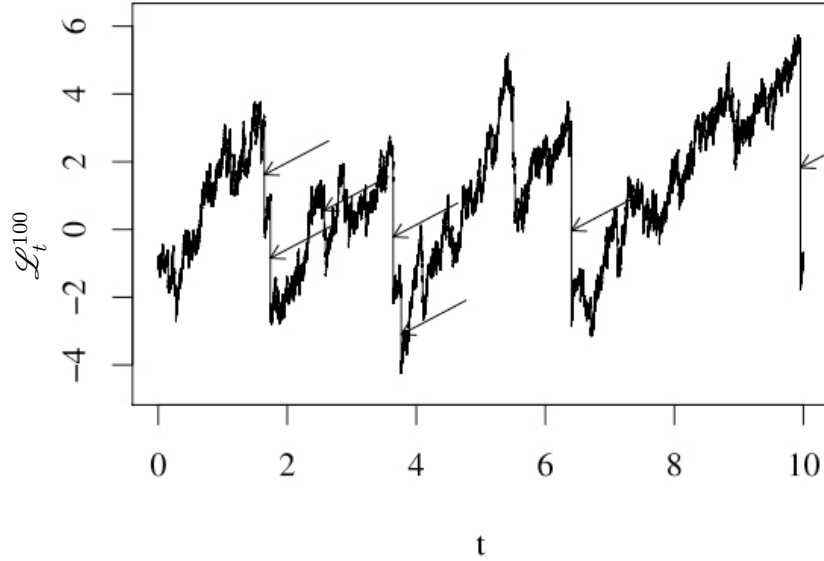
Figure 2: One realization of the tree length process $\mathscr{L}^{100}$ started at time $t = 0$ in equilibrium. The arrows point to those jumps of the tree length at which the MRCA of the total population changes.

2.2 The evolution of tree lengths

Next, we come to our main result on the limit of the compensated tree length processes $\mathscr{L}^{N}$. We denote by $\mathbb{D}$ the space of real-valued càdlàg functions on the time axis $(-\infty, \infty)$, equipped with the Skorokhod topology.

Theorem 1. *There is a process $\mathscr{L} = (\mathscr{L}_t)_{t \in \mathbb{R}}$ with sample paths in $\mathbb{D}$ such that*

$$\mathscr{L}^{N} \Longrightarrow \mathscr{L} \ as \ N \to \infty.$$

The distribution of $\frac{1}{2}\mathscr{L}_t$ is Gumbel for all $t \in \mathbb{R}$. The process $\mathscr{L}$ has infinite infinitesimal variance, with

$$\frac{1}{t|\log t|}\mathbb{E}[(\mathscr{L}_t - \mathscr{L}_0)^2] \xrightarrow{t \to 0} 4. \tag{2.2}$$

Remark 2.2 (Connection to external branch lengths)**.** An important characteristics of the tree length process is its jump size statistics. For this, let F be a randomly chosen jump time for the equilibrium process $\mathscr{L}$. By the independence properties of the Poisson processes which generate $\mathscr{T}^{N}$, we have $\mathscr{T}_{F-}^{N} \overset{d}{=} \mathscr{T}_0^{N}$ and consequently $\mathscr{L}_{F-}^{N} \overset{d}{=} \mathscr{L}_0^{N}$. Moreover, the jump removes a randomly chosen *external branch* from $\mathscr{T}_{F-}^{N}$. Since the N-coalescent, restricted to $N - 1$ randomly chosen individuals, is in distribution identical to the $(N-1)$-coalescent, it follows that $\mathscr{T}_F^{N} \overset{d}{=} \mathscr{T}_0^{N-1}$ and consequently $\mathscr{L}_F^{N} \overset{d}{=} \mathscr{L}_0^{N-1} + 2\log\left(1 - \frac{1}{N}\right)$. Moreover, the jump size, given by $\mathscr{L}_{F-}^{N} - \mathscr{L}_F^{N}$, is in distribution identical to a randomly chosen external branch of a N-coalescent in equilibrium. Properties of the external branch length distribution are recalled in Section 4.2 and were studied in more detail in [CNKR07]. For our setup, these results imply

$$N(\mathscr{L}_{F-}^{N} - \mathscr{L}_F^{N}) \xrightarrow{N \to \infty} J$$

for some random variable J with expectation 2 and density $x \mapsto 8/(2 + x)^3$. This power law with exponent 3 was already guessed in [RBY04] based on simulations. The random variable J
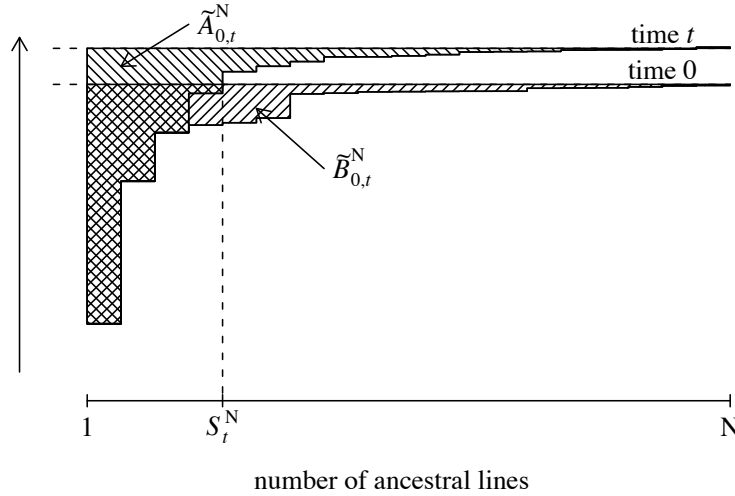
Figure 3: Schematic picture of tree change between two times 0 and t for a population of size N. Tree topology is ignored in the figure and only the number of ancestral lines is given. The population at time t has S_t^N ancestors at time 0. The genealogical tree at time t overlaps with the tree at time 0. The subset $A_{0,t}^N$ belongs to the tree at time t but not to the tree at time 0. The subset $B_{0,t}^N$ is lost between time 0 and t. The net difference in tree length is $\mathscr{L}_t^N - \mathscr{L}_0^N \overset{d}{=} A_{0,t}^N - B_{0,t}^N$.

has unbounded variance. For the asymptotics of this variance, [FL93] already showed (see also Proposition 4.2), translated to our setup,[2]

$$\mathbb{V}[N(\mathscr{L}_{F-}^N - \mathscr{L}_F^N)] \overset{N\to\infty}{\sim} 8\log(N). \tag{2.3}$$

Remark 2.3 (Heuristics on jump sizes). The fact that the approximate size of a randomly chosen jump is of the order $2/N$ can also be seen from the dynamics of $\mathscr{L}^N$. In one time unit, tree length is gained by growth of the tree at constant speed N. Moreover, the process $\mathscr{T}^N$ makes approximately $\binom{N}{2}$ jumps. Since $\mathscr{L}^N$ is in equilibrium, the tree growth and the jumps have to compensate each other. Therefore, a single jump must be of expected size $\frac{2}{N}$; compare also Proposition 4.2.

In the light of (2.3), the fact that the limit process $\mathscr{L}$ has infinite infinitesimal variance would not be surprising if there were no dependencies between jump sizes: In a short time t, the process $\mathscr{L}^N$ makes approximately $\binom{N}{2}t$ downward jumps as $N \to \infty$. If jumps would be independent, we would get from (2.3) that the variance of $\mathscr{L}_t^N - \mathscr{L}_0^N$ is approximately $4t\log N$.

Remark 2.4 (Idea of the proof of Theorem 1). The crucial step in the proof, whose details are given in Section 5, is to establish tightness of the family $\mathscr{L}^N$. For this, it suffices to show that 'large' jumps in the tree length do not happen too often. To be more specific, we must show that during times $(t-h, t]$ and $(t, t+h]$ some moment of the smaller jump, $\mathscr{L}_t^N - \mathscr{L}_{t-h}^N$ or $\mathscr{L}_{t+h}^N - \mathscr{L}_t^N$, is bounded by Ch^θ for some constant C and $\theta > 1$. Heuristically, such a statement is true since it can be shown that the times at which one of the f oldest families of the coalescent tree dies out build a Poisson process with rate $\binom{f}{2}$, $f = 2, 3, \ldots$, see the arguments leading to (5.3). Dying out of one of the f oldest families implies larger jumps for smaller f and the proof of tightness requires bounds for f depending on the time interval h, using that loss times for one of the f oldest families in $(t-h, t]$ and $(t, t+h]$ are independent.

To obtain the form for the infinitesimal variance, it is essential to bound jumps between times 0 and t of $\mathscr{L}^N$ for small t, uniformly in N; see Figure 3 for an illustration. Note that jumps in $\mathscr{L}^N$

---

[2]For sequences $(a_N)_{N=1,2,\ldots}$ and $(b_N)_{N=1,2,\ldots}$ we write $a_N \overset{N\to\infty}{\sim} b_N$ iff $a_N/b_N \xrightarrow{N\to\infty} 1$.

consist of two sources. First, between 0 and t, additional tree length is gained by tree growth ($\tilde{A}_{0,t}^{\mathrm{N}}$ in the figure). The random variable $\tilde{A}_{0,t}^{\mathrm{N}}$ equals the tree-length of a Kingman N-coalescent gained by time t and for large N, we see from Lemma 4.10 that $\mathbb{V}[\tilde{A}_{0,t}^{\mathrm{N}}] \overset{t \to 0}{\sim} \frac{2}{3}t$. Second, a part of the tree at time t breaks off ($\tilde{B}_{0,t}^{\mathrm{N}}$ in the figure). This part is determined by the number S_t^{N} of ancestors at time 0 of the population at time t. Additionally, note that the number of ancestors at time t is approximately $\lfloor \frac{2}{t} \rfloor$ for large N; see Lemma 4.6. Therefore the random variable $B_{0,t}^{\mathrm{N}} := \tilde{B}_{0,t}^{\mathrm{N}} - \mathbb{E}[\tilde{B}_{0,t}^{\mathrm{N}}]$ converges as $N \to \infty$ to the difference of the compensated tree length of a coalescent with infinitely many lineages and the compensated length of the tree spanned by a subset of $\lfloor \frac{2}{t} \rfloor$ lineages. This random variable, denoted by $B_{\lfloor \frac{2}{t} \rfloor}$, satisfies $\mathbb{V}[B_{0,t}^{\mathrm{N}}] \overset{\mathrm{N} \to \infty}{\sim} \mathbb{V}[B_{\lfloor \frac{2}{t} \rfloor}] \overset{t \to 0}{\sim} 4t \log \frac{2}{t} \overset{t \to 0}{\sim} 4t|\log t|$ by Proposition 4.4. Combining the results for $\tilde{A}_{0,t}^{\mathrm{N}}$ and $\tilde{B}_{0,t}^{\mathrm{N}}$ we see that $\lim_{\mathrm{N} \to \infty} \mathbb{V}[\mathscr{L}_t^{\mathrm{N}} - \mathscr{L}_0^{\mathrm{N}}] \overset{t \to 0}{\sim} t|\log t|$.

Remark 2.5 (Convergence of tree-valued processes). [GPW09] specify a topology τ on the space of trees and prove that the sequence of tree-valued processes $\mathscr{T}^{\mathrm{N}}$ converges in distribution to a tree-valued process $\mathscr{T}$, whose paths are a.s. continuous with respect to the topology τ. One might be tempted to use convergence of $\mathscr{T}^{\mathrm{N}}$ to $\mathscr{T}$ in order to show that $\mathscr{L}^{\mathrm{N}}$ converges to some limit process $\mathscr{L}$ as well. Such an attempt would require that the function ℓ mapping finite trees to their (compensated) lengths is τ-continuous. However, if ℓ would be τ-continuous we would conclude that $\mathscr{L}$ has continuous paths, but $\mathscr{L}$ clearly makes jumps. Hence, ℓ is not continuous and convergence of $\mathscr{T}^{\mathrm{N}}$ cannot be used to show convergence of $\mathscr{L}^{\mathrm{N}}$.

Remark 2.6 (Connection to empirical population genetics). Coalescent trees are of particular importance in empirical population genetics and in the analysis of sequence diversity data. In the infinite sites model, mutations leading to segregating sites fall on the genealogical tree at constant rate. As a consequence, the number of segregating sites is Poisson distributed with a parameter proportional to the tree length. As illustrated by Figure 2, the tree length process makes jumps. Particularly large jumps occur at a time F when the most recent common ancestor of the total population changes. At time F, one of the two oldest families in the population dies out and a long external branch breaks off the genealogical tree (see also [Taj90]). At time $F-$ there are several segregating sites which are carried by all individuals which belong to the family which does not die out. Such segregating sites become fixed in the population at time F. In particular, fixation of segregating sites (also denoted by *substitutions*) come in bursts as time evolves, an observation already made by [Wat82]. In addition, segregating sites which are present only in the oldest family which dies out at time F, are lost.

Observations concerned with the substitutions of segregating sites are special properties of the *mutation-drift balance*. This dynamic equilibrium is between the introduction of new segregating sites due to mutation and loss of present ones due to genetic drift. Considered between times 0 and t, the introduction of new mutations in the population are due to mutation events falling on the part of the genealogical tree gained between times 0 and t while the loss of existing mutations is due to some part of the genealogical tree at time 0 breaking off by time t. Most interestingly, the number of segregating sites in the total population, unlike many processes in population genetics, is 'super-diffusive' in that it has infinite infinitesimal variance, as stated in Theorem 1.

3 Strong convergence of tree lengths

Proposition 2.1 and Theorem 1 establish convergence in distribution for the real-valued random variables $\mathscr{L}_t^{\mathrm{N}}$ (t fixed) and the $\mathbb{D}$-valued random variables $\mathscr{L}^{\mathrm{N}}$. We extend these results by stronger notions of convergence, i.e. convergence in probability, almost sure convergence and convergence in L^2. We start with extensions for fixed times (Proposition 3.1) and then come to the extension involving the processes $\mathscr{L}^{\mathrm{N}}$ (Proposition 3.2). The proofs are given in Section 6.
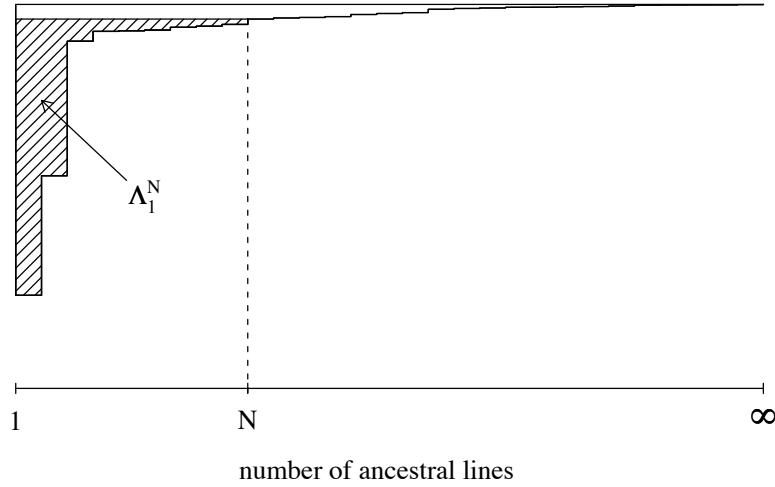
Figure 4: Kingman's temporal coupling: In the tree $\mathcal{T}$ coming down from infinitely many leaves, a tree with N leaves can be embedded by considering the part of $\mathcal{T}$ which is below the time at which $\mathcal{T}$ comes down to N lines. The resulting tree lengths is Λ_1^N.

3.1 Tree lengths at fixed times

For the extension of Proposition 2.1, fix $t \in \mathbb{R}$. An elegant way to encode a random coalescent tree $\mathcal{T} := \mathscr{T}_t$ is in terms of (the completion of) a random metric on $\mathbb{N}$, as proposed by Evans [Eva00]. To visualize this, consider a sequence of lineages indexed by $\mathbb{N}$, where lineage i starts at time 0 in leaf i. Any pair of lineages coalesces independently at rate 1, and a random (ultra)metric R is defined by

$$R(i,j) := 2 \cdot \text{time to the most recent common ancestor of leaves } i, j. \qquad (3.1)$$

The completion of $(\mathbb{N}, R)$ is a.s. a compact ultra-metric space that represents the (uncountable set of) leaves of the coalescent tree.

There are two canonical ways to approach the compensated length of $\mathcal{T}$ by a sequence of lengths of finite trees.

1. *From the root to the leaves:* Let $R_{(1)}, R_{(2)}, \ldots$ be a listing of the set $\{R(i,j) : i, j \in \mathbb{N}, i \neq j\}$ in decreasing order. The random variable

$$X_k := R_{(k-1)} - R_{(k)} \qquad (3.2)$$

then gives the time the tree $\mathcal{T}$ spends with k lines "in parallel". We set

$$\Lambda_1^N := \sum_{k=2}^{N} k X_k - 2 \log N, \quad N = 2, 3, \ldots$$

The random sequence $(\Lambda_1^N)_{N=2,3,\ldots}$ is called *temporal coupling* in [Kin82b]; see also Figure 4.

2. *Across lineages:* For $N = 2, 3, \ldots$, consider the finite subtree

$$\mathcal{T}^N \text{ encoded by } \left(\{1, \ldots, N\}, R\big|_{\{1,\ldots,N\}^2} \right), \qquad N = 1, 2 \ldots. \qquad (3.3)$$

Define

$$\Lambda_2^N := \widetilde{\ell}(\mathcal{T}^N) - 2 \log N, \quad N = 2, 3, \ldots$$

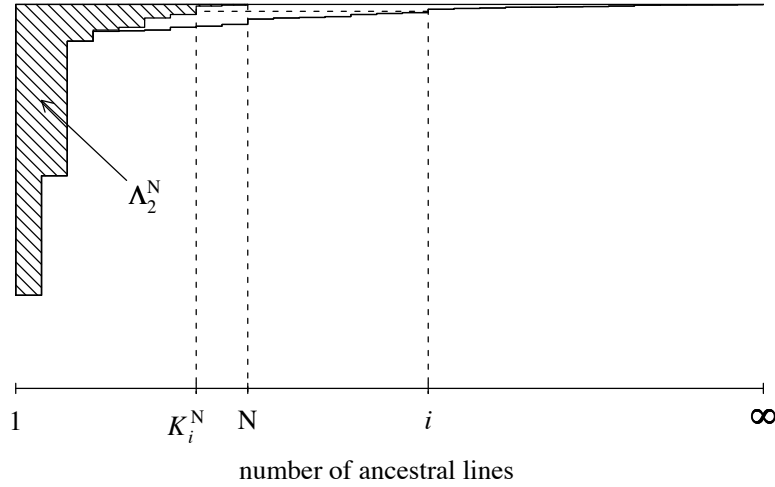This random sequence is called *natural coupling* in [Kin82b]; see also Figure 5.

Figure 5: Kingman's natural coupling: In the tree $\mathcal{T}$ coming down from infinitely many leaves, a tree with N leaves can be embedded by working down from the first N leaves (obtained by sampling) from the infinite tree. The resulting tree length is Λ_2^N. In the natural coupling, K_i^N is the number of lines in the small tree while the full tree has i lines.

Proposition 3.1. *There is random variable* $\Lambda = \Lambda(R)$ *such that* $\frac{1}{2}\Lambda$ *is Gumbel distributed,*

$$\Lambda_1^N \xrightarrow{\text{N}\to\infty} \Lambda \tag{3.4}$$

almost surely and in L^2 *and*

$$\Lambda_2^N \xrightarrow{\text{N}\to\infty} \Lambda \tag{3.5}$$

in L^2.

3.2   The evolution of tree lengths

For the extension of Theorem 1 we briefly review the lookdown construction of [DK99]; see also [PW06] for a detailed description and Figure 6 for an illustration.

Consider the set of vertices $\mathcal{V} := \mathbb{R} \times \mathbb{N}$. A vertex (t, i) is referred to as *the individual at time t at level i*. The source of randomness in the lookdown construction is a family of rate one Poisson processes $(\mathcal{P}_{ij})_{1 \leq i < j}$. At times $t \in \mathcal{P}_{ij}$, the individual at level j *looks down* to level i. As the illustration in Figure 6 shows, at a lookdown event in $\mathcal{P}_{ij}$, all individuals at levels $k \geq j$ are pushed one level up, and a new *line of ascent* is born at level j. Note that the individual at level k is pushed to level $k+1$ at rate $\binom{k}{2}$.

We define the partition $\mathcal{G}$ of $\mathbb{R} \times \{2, 3, \ldots\}$ into lines of ascent as follows. Every $s_0 \in \mathcal{P}_{ij}$ gives rise to a partition element G of the form

$$([s_0, s_1) \times \{j\} \cup ([s_1, s_2) \times \{j+1\}) \cup ([s_2, s_3) \times \{j+2\}) \cup \ldots$$

with $s_{k+1} > s_k$ for all $k = 0, 1, 2, \ldots$. Here, s_{k+1} is the smallest element in $[s_k, \infty) \cap \bigcup_{1 \leq i' < j' \leq k+j} \mathcal{P}_{i'j'}$. We say that G is born by (s_0, i) and pushed one level up at times $s_1, s_2, \ldots$. If $(s_0, i) \in G'$, we say that G descends from G'. Since the individual at level k is pushed up by one at rate $\binom{k}{2}$ and hence, for G as above, $\lim_{k \to \infty} s_k$ is finite.

For $s \leq t$, $i \leq j$ let $G, G' \in \mathcal{G}$ be such that $(s, i) \in G', (t, j) \in G$. We say that (s, i) is ancestor of (t, j) if either $G = G'$ or there are $G_1, \ldots, G_n$ such that G descends from G_n, G_k descends from
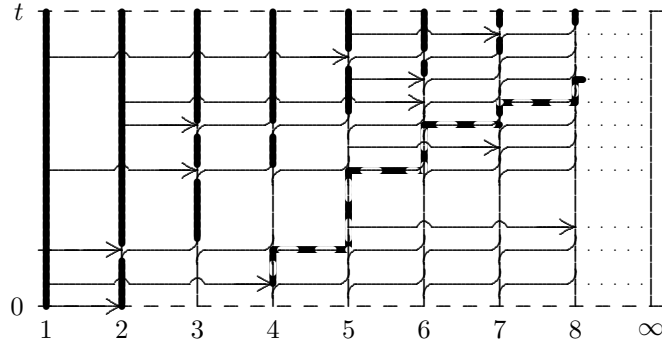
Figure 6: Detail of a look-down graph. Time is running upwards; all lines at the first 8 levels are drawn between times 0 and t. At times in $\mathcal{P}_{ij}$ an arrow is drawn from i to j, and all lines at levels $\geq j$ are pushed upwards as indicated by bent lines. The dashed line initiates one partition element of the partition $\mathcal{G}$ of $\mathbb{R} \times \{2, 3, \ldots\}$. The solid lines constitute the tree $\mathcal{T}_t^{\mathrm{ld},8}$ of the first 8 ancestral lineages back from time t. The sum of the lengths of the solid lines is the tree length of the first 8 levels at time t.

G_{k-1}, $k = 2, \ldots, n$ and G_1 descends from G'. In this case, we define $A_s(t, j) := i$. In addition, for $t \in \mathbb{R}$, define the random metric R_t^{ld} (compare (3.1)) on $\mathbb{N}$ by

$$R_t^{\mathrm{ld}}(i, j) = 2 \cdot \inf\{t - s : A_s(t, i) = A_s(t, j)\}.$$

We define the random trees

$$\mathcal{T}_t^{\mathrm{ld},\mathrm{N}} := \left(\{1, \ldots, \mathrm{N}\}, R_t^{\mathrm{ld}}|_{\{1, \ldots, \mathrm{N}\}^2}\right), \qquad \mathrm{N} = 2, 3, \ldots$$

and

$$\mathcal{L}_t^{\mathrm{ld},\mathrm{N}} := \widetilde{\ell}(\mathcal{T}_t^{\mathrm{ld},\mathrm{N}}) - 2 \log \mathrm{N}.$$

Recall the complete Skorokhod metric d_{Sk} from [EK86, Section 3.5]. We are now ready to state a result extending Theorem 1 to convergence in probability, proved in Section 6.

Proposition 3.2. *There is a process $\mathcal{L}^{\mathrm{ld}}$, having the same distribution as $\mathcal{L}$ from Theorem 1, such that*

$$d_{\mathrm{Sk}}(\mathcal{L}^{\mathrm{N},\mathrm{ld}}, \mathcal{L}^{\mathrm{ld}}) \xrightarrow{\mathrm{N} \to \infty} 0$$

in probability.

4   Auxiliary results on Kingman's coalescent

In this section we collect some facts on Kingman's coalescent which will be required for the proof of Theorem 1. For this, we take the tree $\mathcal{T}$ as introduced in Section 3. Recall the subtrees $\mathcal{T}^{\mathrm{N}}, \mathrm{N} = 2, 3, \ldots$ as defined in (3.3) and the inter-coalescence times $X_2, X_3, \ldots$ from (3.2). Recall that X_k is exponentially distributed with parameter $\binom{k}{2}$. We define

$$K_i^{\mathrm{N}} := \text{ number of lines in } \mathcal{T}^{\mathrm{N}} \text{ while } \mathcal{T} \text{ has } i \text{ lines} \tag{4.1}$$

and $K^{\mathrm{N}} := (K_i^{\mathrm{N}})_{i=1,2,\ldots}$; see also Figure 5.

4.1   The Markov Chain K^{N}

The connection between the trees $\mathcal{T}^{\mathrm{N}}$ and $\mathcal{T}$ has been described e.g. by [Tav84], [STW84] and [EPW06, Section 4.4]. Lemma 4.8 of [EPW06] states that K^{N} builds a Markov chain with one-

and two-dimensional distributions

$$\mathbb{P}[K_i^N = k] = \frac{\binom{N-1}{N-k}\binom{i}{k}}{\binom{N+i-1}{N}}, \qquad i \geq 1, 1 \leq k \leq N \tag{4.2}$$

$$\mathbb{P}[K_j^N = \ell | K_i^N = k] = \frac{\binom{N-k}{N-\ell}\binom{j+k-1}{i+l-1}}{\binom{N+j-1}{N+i-1}}, \qquad 1 \leq i \leq j, 1 \leq k \leq \ell \leq N. \tag{4.3}$$

We will need some moment properties of this Markov chain.

Lemma 4.1. *For $i \geq 1$,*

$$\mathbb{E}[i - K_i^N] = \frac{i(i-1)}{N+i-1}, \tag{4.4}$$

and for $1 \leq i \leq j$

$$\mathbb{E}[(i - K_i^N)(j - K_j^N)] = \frac{i(i-1)j(j-1)}{(N+i-1)(N+j-1)} + \frac{i(i-1)N(N-1)}{(N+j-1)(N+i-1)(N+i-2)}. \tag{4.5}$$

Proof. In the proof we omit the superscript N. It is a consequence of (4.2) that

$$\mathbb{E}[i - K_i] = \sum_{k=1}^{\infty}(i-k)\frac{\binom{N-1}{N-k}\binom{i}{k}}{\binom{N+i-1}{N}} = i\sum_{k=1}^{\infty}\frac{\binom{N-1}{N-k}\binom{i-1}{k}}{\binom{N+i-1}{N}} = i\frac{\binom{N+i-2}{N}}{\binom{N+i-1}{N}} = \frac{i(i-1)}{N+i-1} \tag{4.6}$$

which shows (4.4). Moreover,

$$
\begin{aligned}
\mathbb{E}[(i - K_i)^2] &= \sum_{k=1}^{\infty}(i-k)(i-k-1)\frac{\binom{N-1}{N-k}\binom{i}{k}}{\binom{N+i-1}{N}} + \mathbb{E}[i - K_i] \\
&= i(i-1)\sum_{k=1}^{\infty}\frac{\binom{N-1}{N-k}\binom{i-2}{k}}{\binom{N+i-1}{N}} + \mathbb{E}[i - K_i] \\
&= \frac{i(i-1)\binom{N+i-3}{N}}{\binom{N+i-1}{N}} + \frac{i(i-1)}{N+i-1} \\
&= \frac{i(i-1)^2(i-2)}{(N+i-1)(N+i-2)} + \frac{i(i-1)}{N+i-1}
\end{aligned}
$$

and, using (4.3) for $1 \leq i \leq j$,

$$
\begin{aligned}
\mathbb{E}[j - K_j | K_i] &= \sum_{\ell=1}^{\infty}(j-\ell-i+K_i)\frac{\binom{N-K_i}{N-\ell}\binom{j+K_i-1}{i+\ell-1}}{\binom{N+j-1}{N+i-1}} + i - K_i \\
&= (j+K_i-1)\sum_{\ell=1}^{\infty}\frac{\binom{N-K_i}{N-\ell}\binom{j+K_i-2}{i+\ell-1}}{\binom{N+j-1}{N+i-1}} + i - K_i \\
&= \frac{(j+K_i-1)\binom{N+j-2}{N+i-1}}{\binom{N+j-1}{N+i-1}} + i - K_i \\
&= \frac{(j+K_i-1)(j-i)}{N+j-1} + i - K_i \\
&= \frac{(i-K_i)(N+i-1)}{N+j-1} + \frac{(j+i-1)(j-i)}{N+j-1}.
\end{aligned}
$$

Combining the last two equations we obtain

$$
\begin{aligned}
\mathbb{E}[(i - K_i)(j - K_j)] &= \mathbb{E}[(i - K_i) \cdot \mathbb{E}[j - K_j | K_i]] \\
&= \mathbb{E}[(i - K_i)^2] \frac{N + i - 1}{N + j - 1} + \mathbb{E}[i - K_i] \frac{(j + i - 1)(j - i)}{N + j - 1} \\
&= \frac{i(i-1)^2(i-2)}{(N + j - 1)(N + i - 2)} + \frac{i(i-1)}{N + j - 1} + \frac{i(i-1)\big(j(j-1) - i(i-1)\big)}{(N + i - 1)(N + j - 1)} \\
&= \frac{i(i-1)j(j-1)}{(N + i - 1)(N + j - 1)} + \frac{i(i-1)N(N-1)}{(N + j - 1)(N + i - 1)(N + i - 2)}
\end{aligned}
$$

i.e., we have shown (4.5). □

4.2   The length of an external branch

We will recall several facts of the length of a randomly chosen external branch in an N-coalescent $\mathcal{T}^N$. In this setting we take the inter-coalescence times $X_2^N, ..., X_N^N$ are such that X_i^N is exponentially distributed with rate $\binom{i}{2}$, $i = 2, ..., N$. We denote by J^N the length of a randomly chosen external branch. The results we describe in this section are collected from [FL93], [Dur08], [BF05] and [CNKR07] and stated here for completeness.

We define

$$
F^N := f \qquad \text{iff} \qquad J^N = \sum_{k=f+1}^{N} X_k,
$$

i.e. F^N denotes the number of lines extant in the N-coalescent at the time at which the external branch connects to the tree. We give a basic fact about F^N and properties of J^N.

Proposition 4.2 (External branches). *For $f = 1, ..., N - 1$,*

$$
\mathbb{P}[F^N < f] = \frac{f(f-1)}{N(N-1)} \qquad i.e. \qquad \mathbb{P}[F^N = f] = \frac{2f}{N(N-1)}
$$

The first two moments of J^N are given by

$$
\mathbb{E}[J^N] = \frac{2}{N},
$$

$$
\mathbb{V}[J^N] = \frac{8 \sum_{k=1}^{N} \frac{1}{k} - 12 + \frac{4}{N}}{N(N-1)}.
$$

Proof. We obtain the distribution of F^N as follows: With probability $\frac{\binom{N-1}{2}}{\binom{N}{2}}$ the randomly chosen external branch is not involved in the first coalescence event (bringing the number of lines from N down to N − 1). Iterating this argument, we immediately see that the probability that the randomly chosen line did not take part in the first N − f + 1 coalescence events is

$$
\mathbb{P}[F^N < f] = \frac{\binom{N-1}{2}}{\binom{N}{2}} \cdot \frac{\binom{N-2}{2}}{\binom{N-1}{2}} \cdots \frac{\binom{f}{2}}{\binom{f+1}{2}} = \frac{\binom{f}{2}}{\binom{N}{2}} = \frac{f(f-1)}{N(N-1)}.
$$

To compute moments of J^N, we use the representation

$$
J^N = \sum_{k=F^N}^{N} X_k^N,
$$

where $F^N, X_2^N, ..., X_N^N$ are independent. The first two moments of J^N are now obtained by

$$\mathbb{E}[J^N] = \mathbb{E}\big[\mathbb{E}[J^N|F^N]\big] = \mathbb{E}\Big[\frac{2}{F^N} - \frac{2}{N}\Big] = \Big(2 \sum_{f=1}^{N-1} \frac{f}{\binom{N}{2}} \frac{1}{f}\Big) - \frac{2}{N} = \frac{2}{N},$$

$$\mathbb{V}[J^N] = \mathbb{E}\big[\mathbb{V}[J^N|F^N]\big] + \mathbb{V}\big[\mathbb{E}[J^N|F^N]\big] = \mathbb{E}\Big[\sum_{k=F^N+1}^{N} \frac{1}{\binom{k}{2}^2}\Big] + \mathbb{V}\Big[\frac{2}{F^N}\Big]$$

$$= \frac{1}{\binom{N}{2}}\Big(\sum_{k=2}^{N}\sum_{f=1}^{k-1} f\frac{1}{\binom{k}{2}^2} + 4\sum_{f=1}^{N-1} f\frac{1}{f^2}\Big) - \frac{16}{N^2}$$

$$= \frac{8}{N(N-1)}\Big(\sum_{k=1}^{N-1} \frac{1}{k}\Big) + \frac{4}{N^2} - \frac{16}{N^2}$$

$$= \frac{8\sum_{k=1}^{N} \frac{1}{k} - 12 + \frac{4}{N}}{N(N-1)}.$$

$\square$

4.3   Subtrees of coalescents and their lengths

The aim of this section is to analyze the difference of the tree lengths of $\mathcal{T}$ and of $\mathcal{T}^N$. Since this difference is infinite, we have to carry out a limiting procedure, compensating by the mean. Recall that the inter-coalescence times $X_2, X_3, ...$ are independent of the tree topology of $\mathcal{T}$ in general and of $K_2^N, K_3^N, ...$ in particular. We study the random variable

$$B_N := \sum_{i=2}^{\infty} \Big((i - K_i^N)X_i - \mathbb{E}[i - K_i^N] \cdot \mathbb{E}[X_i]\Big) \tag{4.7}$$

which is the compensated difference of the tree lengths of $\mathcal{T}$ and $\mathcal{T}^N$.

Remark 4.3. Let us first make sure that the infinite sum in the definition of B_N exists and has expectation zero. To see this, we fix $M \in \{2, 3, ...\}$ and consider the sequences $(B_{N,M})_{M=2,3,...}$ where $B_{N,M}$ is defined as B_N but with the sum ranging from $i = 2$ to $i = M$. Since, for $M < M'$, by Lemma 4.1,

$$\mathbb{E}\big[(B_{N,M'} - B_{N,M})^2\big]$$

$$= \mathbb{E}\Big[\Big(\sum_{i=M+1}^{M'} \Big((i - K_i^N)X_i - \frac{2}{N+i-1}\Big)\Big)^2\Big]$$

$$\leq 8 \sum_{i=M+1}^{M'}\sum_{j=i}^{M'} \mathbb{E}\Big[\Big(\frac{i - K_i^N}{i(i-1)} - \frac{1}{N+i-1}\Big)\Big(\frac{j - K_j^N}{j(j-1)} - \frac{1}{N+j-1}\Big)\Big]$$

$$\leq 8 \sum_{i=M+1}^{M'}\sum_{j=i+1}^{M'} \frac{1}{j(j-1)(N+j-1)}$$

$$= o(1)$$

as $M \to \infty$,, the sequence $(B^{N,M})_{M=2,3,...}$ is Cauchy in L^2 and thus converges in L^2 to the limit variable B_N defined in (4.7). From uniform integrability we conclude that $\mathbb{E}[B_N] = 0$.

Proposition 4.4 (Variance of the difference in length of an infinite and a finite coalescent in the natural coupling)**.**

$$\mathbb{E}[B_N^2] \overset{N\to\infty}{\sim} \frac{8\log N}{N}. \tag{4.8}$$

Proof. We omit the subscript N in the proof. First,

$$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] \cdot (1 + \delta_{ij}).$$

In order to obtain (4.8), we write, using Lemma 4.1,

$$\mathbb{E}[B^2] = \mathbb{E}\Big[\Big(\sum_{i=2}^{\infty}(i - K_i)X_i - \frac{2}{N+i-1}\Big)^2\Big]$$

$$= 2\sum_{j=3}^{\infty}\sum_{i=2}^{j-1}\mathbb{E}\Big[\Big((i - K_i)X_i - \frac{2}{N+i-1}\Big)\Big((j - K_j)X_j - \frac{2}{N+j-1}\Big)$$

$$+ \sum_{i=2}^{\infty}\mathbb{E}\Big[\Big((i - K_i)X_i - \frac{2}{N+i-1}\Big)^2\Big]$$

$$= 8\sum_{j=3}^{\infty}\sum_{i=2}^{j-1}\mathbb{E}\Big[\Big(\frac{i - K_i}{i(i-1)} - \frac{1}{N+i-1}\Big)\Big(\frac{j - K_j}{j(j-1)} - \frac{1}{N+j-1}\Big)\Big] \qquad (4.9)$$

$$+ \sum_{i=2}^{\infty}\Big(8\mathbb{E}\Big[\Big(\frac{i - K_i}{i(i-1)}\Big)^2\Big] - \frac{4}{(N+i-1)^2}\Big)$$

$$= 8\sum_{j=3}^{\infty}\sum_{i=2}^{j-1}\frac{N(N-1)}{j(j-1)(N+j-1)(N+i-1)(N+i-2)}$$

$$+ \sum_{i=2}^{\infty}\Big(\frac{8 \cdot N(N-1)}{i(i-1)(N+i-1)^2(N+i-2)} + \frac{4}{(N+i-1)^2}\Big).$$

It is clear that the expression in the last line go to 0 as C/N, for some $C > 0$, as $N \to \infty$. For the next to last line, we obtain

$$\sum_{j=3}^{\infty}\frac{N(N-1)}{j(j-1)(N+j-1)}\sum_{i=2}^{j-1}\Big(\frac{1}{N+i-2} - \frac{1}{N+i-1}\Big)$$

$$= \sum_{j=3}^{\infty}\frac{N(N-1)}{j(j-1)(N+j-1)}\Big(\frac{1}{N} - \frac{1}{N+j-2}\Big)$$

$$\overset{N\to\infty}{\sim} \sum_{j=1}^{\infty}\frac{N}{j(N+j)^2} \overset{N\to\infty}{\sim} \sum_{j=1}^{\infty}\frac{1}{j(N+j)} - \frac{1}{(N+j)^2}$$

$$\overset{N\to\infty}{\sim} \frac{1}{N}\sum_{j=1}^{\infty}\Big(\frac{1}{j} - \frac{1}{N+j}\Big) - \frac{1}{N}$$

$$\overset{N\to\infty}{\sim} \frac{\log N}{N}$$

which proves the Proposition. $\square$

4.4   Numbers of ancestors near the tree top

Let[3] $u > 0$. Define S_u^N to be the (random) number of ancestors at time $-u$ in $\mathcal{T}^N$ and S_u the number of ancestors by time $-u$ in $\mathcal{T}$, where $\mathcal{T}^N$ and $\mathcal{T}$ are defined as in Section 3. We give results on convergence of S_u^N as $N \to \infty$ (Lemma 4.5) and on convergence of S_u as $u \to 0$ (Lemma 4.6).

Lemma 4.5. *For $u > 0$,*

$$S_u^N \xrightarrow{N\to\infty} S_u$$

almost surely and in L^p for all $p > 0$.

---

[3]We use times s, t when going forwards in time and u, v when going backwards in time.

Proof. With probability one, the completion of R from (3.1) is compact, so $\mathcal{T}$ *comes down from infinity*, i.e. with probability one there are at most finitely many lines left by time $-u$. Since S_u^{N} is increasing with N, the almost sure convergence follows. To see the L^p-convergence, note that all moments of the distribution of S_u exist; see e.g. [Tav84, Section 5.4]. Since S_u^{N} is bounded by S_u, the families $\big((S_u^{\mathrm{N}})^p\big)_{\mathrm{N}=1,2,\ldots}$ are uniformly integrable and convergence in L^p follows.   □

Lemma 4.6. *For the number of ancestors S_u*

$$u \cdot S_u \xrightarrow{u\to 0} 2, \tag{4.10}$$

almost surely and in L^2. Moreover,

$$\frac{S_u - 2/u}{\sqrt{2/(3u)}} \xrightarrow{u\to 0} N(0,1). \tag{4.11}$$

In addition, for $u, v \to 0$, $u \le v$, $u/v \to \Gamma \le 1$

$$\Big(\frac{S_u - 2/u}{\sqrt{2/(3u)}}, \frac{S_v - 2/v}{\sqrt{2/(3v)}}\Big) \Rightarrow N(0, C) \tag{4.12}$$

with the covariance matrix C given by

$$C = \begin{pmatrix} 1 & \Gamma^{3/2} \\ \Gamma^{3/2} & 1 \end{pmatrix}. \tag{4.13}$$

Remark 4.7. As an extension of the previous lemma, we see that the finite dimensional distributions of $\big((S_{tu} - \frac{2}{tu})/\sqrt{2/(3tu)}\big)_{t\ge 0}$ converge as $u \to 0$ to those of a Gaussian process $(A_t)_{t\ge 0}$ with covariance $\mathbb{COV}[A_s, A_t] = (s/t)^{3/2}$ for $s \le t$.

Proof of Lemma 4.6. The convergence (4.11) is assertion (35) in [Ald99]. For further use below, we prove (4.11) in all detail. Define T_n as the time it takes the coalescent to come down to n lines, i.e.,

$$T_n := \sum_{i=n+1}^{\infty} X_i.$$

Note that

$$\mathbb{E}[T_n] = \sum_{i=n+1}^{\infty} \frac{2}{i(i-1)} = \frac{2}{n}, \qquad \mathbb{V}[T_n] = \sum_{i=n+1}^{\infty} \frac{4}{i^2(i-1)^2} \overset{n\to\infty}{\sim} \frac{4}{3n^3}. \tag{4.14}$$

The central ingredients in the proof are the two facts

$$\mathbb{P}[S_u \le n] = \mathbb{P}[T_n \le u],$$
$$\frac{T_n - 2/n}{\sqrt{4/(3n^3)}} \xrightarrow{n\to\infty} N(0,1), \tag{4.15}$$

where the second assertion is a consequence of (4.14) and the central limit theorem. Hence we may define

$$a_u(x) := \lfloor 2/u + x\sqrt{2/(3u)} \rfloor, \tag{4.16}$$

and write

$$\mathbb{P}\Big[\frac{S_u - 2/u}{\sqrt{2/(3u)}} \le x\Big] = \mathbb{P}[S_u \le a_u(x)]$$

$$= \mathbb{P}[T_{a_u(x)} \le u]$$

$$= \mathbb{P}\Big[\frac{T_{a_u(x)} - 2/a_u(x)}{\sqrt{4/(3a_u(x)^3)}} \le \frac{u - 2/a_u(x)}{\sqrt{4/(3a_u(x)^3)}}\Big]$$

$$\overset{u\to 0}{\sim} \mathbb{P}\Big[\frac{T_{a_u(x)} - 2/a_u(x)}{\sqrt{4/(3a_u(x)^3)}} \le x\Big],$$

since

$$u - 2/a_u(x) \overset{u \to 0}{\sim} x\sqrt{u^3/6}, \qquad \sqrt{4/3(a_u(x)^3)} \overset{u \to 0}{\sim} \sqrt{u^3/6}. \tag{4.17}$$

Now, (4.11) follows from (4.15). Since the event in (4.10) is measurable with respect to the terminal σ-algebra generated by the independent random variables $X_2, X_3, ...$, the convergence in (4.10) holds almost surely. The L^2-convergence follows from moment results for S_u given e.g. in [Tav84, Section 5.4].

Let us turn to the proof of (4.12). Since for $m \le n$

$$\mathbb{COV}[T_m, T_n] = \sum_{i=n+1}^{\infty} \mathbb{V}[X_i] \overset{n \to \infty}{\sim} \frac{4}{3n^3},$$

and using (4.14), it is an easy exercise to show that for $m, n \to \infty, m/n \to \Gamma \le 1$,

$$\left(\frac{T_n - 2/n}{\sqrt{4/(3n^3)}}, \frac{T_m - 2/m}{\sqrt{4/(3m^3)}} \right) \Rightarrow N(0, C)$$

with the covariance matrix C given in (4.13). To see (4.12) from this, note that for $m \le n$, $u \le v$, an extension of (4.15) gives

$$\mathbb{P}[T_n \le u, T_m \le v] = \mathbb{P}[S_u \le n, S_v \le m]$$

and thus for $x, y \in \mathbb{R}$, and using (4.16),

$$\begin{aligned}
\mathbb{P}\Big[\frac{S_u - 2/u}{\sqrt{2/(3u)}} \le x, \frac{S_v - 2/v}{\sqrt{2/(3v)}} \le y \Big] &= \mathbb{P}[S_u \le a_u(x), S_v \le a_v(y)] \\
&= \mathbb{P}[T_{a_u(x)} \le u, T_{a_v(y)} \le v] \\
&= \mathbb{P}\Big[\frac{T_{a_u(x)} \le 2/a_u(x)}{\sqrt{4a_u(x)^3/3}} \le \frac{u - 2/a_u(x)}{\sqrt{4a_u(x)^3/3}}, \frac{T_{a_v(y)} \le 2/a_v(y)}{\sqrt{4a_v(y)^3/3}} \le \frac{v - 2/a_v(y)}{\sqrt{4a_v(y)^3/3}} \Big] \\
&\overset{u,v \to 0}{\sim} \mathbb{P}\Big[\frac{T_{a_u(x)} - 2/a_u(x)}{\sqrt{4a_u(x)^3/3}} \le x, \frac{T_{a_v(y)} \le 2/a_v(y)}{\sqrt{4a_v(y)^3/3}} \le y \Big]
\end{aligned}$$

by (4.17) and thus, if $u, v \to 0, u/v \to \Gamma \le 1$, (4.12) follows since $a_v(x)/a_u(y) \overset{u,v \to 0}{\sim} u/v$. $\qquad\square$

4.5 The tree length near the tree top

We analyze now the contribution to the tree length that comes from a small time interval near the tree top. To this purpose we define

$$\Lambda_u^N := \int_0^u (S_v^N - \mathbb{E}[S_v^N])dv, \qquad \Lambda_u := \int_0^u (S_v - \mathbb{E}[S_v])dv. \tag{4.18}$$

Again, we give results on convergence of Λ_u^N as $N \to \infty$ (Lemma 4.9) and of Λ_u as $u \to 0$ (Lemma 4.10).

Remark 4.8 (Λ_u as an L^2-limit). Since the integrand in the definition of Λ_u is unbounded, we have to make sure that the random variable Λ_u exists. Indeed, using Lemma 4.6 it is easy to check that $\left(\int_{2^{-n}}^u (S_v - \mathbb{E}[S_v])dv \right)_{n=1,2,...}$ is a Cauchy sequence in L^2, and we define Λ_u as its L^2-limit.

Lemma 4.9. *For the random variables Λ_u^N and Λ_u,*

$$\Lambda_u^N \xrightarrow{N \to \infty} \Lambda_u \tag{4.19}$$

in L^2.

Proof. We start with proving the intuitively obvious fact that $S_v - S_v^N$ and $S_w - S_w^N$ have non-negative correlation. Using Lemma 4.1, for $w \leq v$, we write

$$\mathbb{COV}[S_v - S_v^N, S_w - S_w^N] = \mathbb{COV}\big[\mathbb{E}[S_v - S_v^N|S_v, S_w], \mathbb{E}[S_w - S_w^N|S_v, S_w]\big]$$
$$+ \mathbb{E}\big[\mathbb{COV}[S_v - S_v^N, S_w - S_w^N|S_v, S_w]\big]$$
$$= \mathbb{COV}\Big[\frac{S_v(S_v - 1)}{N + S_v - 1}, \frac{S_w(S_w - 1)}{N + S_w - 1}\Big] + \mathbb{E}\Big[\frac{S_v(S_v - 1)N(N - 1)}{(N + S_w - 1)(N + S_v - 1)(N + S_v - 2)}\Big].$$
$$(4.20)$$

The second term on the r.h.s. is nonnegative, and so is the first term, since $i \mapsto \frac{i(i-1)}{N+i-1}$ is increasing and (S_v, S_w) are associated, i.e.

$$\mathbb{COV}[f(S_v), g(S_w)] \geq 0 \tag{4.21}$$

for all increasing functions f, g. Indeed, to verify (4.21) it is enough to show this inequality for $f(S_v) = 1_{S_v \geq \ell}$ and $g(S_w) = 1_{S_w \geq k}$. This, however, is clear since

$$\mathbb{COV}[1_{\{S_w \geq \ell\}}, 1_{\{S_v \geq k\}}] = \mathbb{COV}[1_{\{S_w \geq \ell\}}, \mathbb{E}[1_{\{S_v \geq k\}}|S_w]] \geq 0$$

and $\mathbb{E}[1_{\{S_v \geq k\}}|S_w]$ is an increasing function of S_w and the single random variable S_w is associated. So we have proved that

$$\mathbb{COV}[S_v - S_v^N, S_w - S_w^N] \geq 0 \tag{4.22}$$

for all $v, w \geq 0$.

Now we come to the proof of (4.19). By Fubini's Theorem and (4.22),

$$\mathbb{E}[(\Lambda_u^N - \Lambda_u)^2] = 2 \int_0^u \int_0^v \mathbb{E}\big[\big(S_v^N - S_v - \mathbb{E}[S_v^N - S_v]\big)\big(S_w^N - S_w - \mathbb{E}[S_w^N - S_w]\big)\big] \, dw dv$$
$$= 2 \int_0^u \int_0^v \mathbb{COV}[S_v - S_v^N, S_w - S_w^N] dw dv$$
$$\leq 2 \int_0^\infty \int_0^v \mathbb{COV}[S_v - S_v^N, S_w - S_w^N] dw dv = \mathbb{E}[(\Lambda_\infty^N - \Lambda_\infty)^2] \xrightarrow{N \to \infty} 0$$

by Proposition 4.4, since $(\Lambda_\infty^N - \Lambda_\infty)^2$ is distributed as B_N^2 in that Proposition. $\square$

Lemma 4.10. *For the random variables Λ_u,*

$$\mathbb{V}[\Lambda_u] \overset{u \to 0}{\sim} \frac{2}{3}u. \tag{4.23}$$

Proof. By Lemma 4.6, we see that for $w \leq v$

$$\mathbb{COV}[S_v, S_w] \overset{w \leq v \to 0}{\sim} \sqrt{\frac{2}{3w} \frac{2}{3v} \frac{w^3}{v^3}} = \frac{2}{3}\frac{w}{v^2}, \tag{4.24}$$

so we may write

$$\mathbb{V}[\Lambda_u] = 2 \int_0^u \int_0^v \mathbb{COV}[S_v, S_w] dw dv \tag{4.25}$$

since the right hand side exists. Now, we compute directly

$$\mathbb{V}[\Lambda_u] = 2 \int_0^u \int_0^v \mathbb{COV}[S_w, S_v] dw dv \overset{u \to 0}{\sim} \frac{4}{3} \int_0^u \int_0^v \frac{w}{v^2} dw dv = \frac{2}{3}u.$$

$\square$

Remark 4.11. Note that Λ_u is an integral over approximately Gaussian random variables for small u. In addition, for $s \leq t$,

$$\mathbb{COV}\Big[\frac{\Lambda_{su}}{\sqrt{\frac{2}{3}su}}, \frac{\Lambda_{tu}}{\sqrt{\frac{2}{3}tu}}\Big] = \frac{3}{2u\sqrt{st}} \int_0^{su} \int_0^{tu} \mathbb{COV}[S_v, S_w]dwdv$$

$$\overset{u \to 0}{\sim} \frac{3}{2u\sqrt{st}}\Big(2 \int_0^{su} \int_0^v \frac{2}{3}\frac{w}{v^2}dwdv + \frac{2}{3} \int_0^{su} wdw \int_{su}^{tu} \frac{1}{v^2}dv\Big)$$

$$= \frac{3}{2}\sqrt{\frac{s}{t}} - \frac{1}{2}\sqrt{\frac{s^3}{t^3}}.$$

Hence, as an extension of (4.23), we see that the finite dimensional distributions of $\Big(\Lambda_{tu}/\sqrt{\frac{2}{3}tu}\Big)_{t \geq 0}$ converge as $u \to 0$ to those of a centered Gaussian process $(A_t)_{t \geq 0}$ with covariance $\mathbb{COV}[A_s, A_t] = \frac{3}{2}\sqrt{\frac{s}{t}} - \frac{1}{2}\sqrt{\frac{s^3}{t^3}}$ for $s \leq t$.

5 Proof of Theorem 1

We will prove Theorem 1 in four steps. For convergence of $\mathscr{L}^N$ we need to show (see e.g. [EK86, Lemma 3.4.3])

(a) The sequence of processes $\mathscr{L}^N$ is tight in $\mathbb{D}$,

(b) The finite-dimensional distributions of the sequence $\mathscr{L}^N$ converge.

The main work is to show tightness of $(\mathscr{L}^N)_{N \in \mathbb{N}}$ in $\mathbb{D}$, carried out in Steps 1 and 2. For this, it is enough to show (see [EK86, Theorem 3.8.6 and Theorem 3.8.8])

$$(\mathscr{L}_0^N)_{N \in \mathbb{N}} \text{ is tight in } \mathbb{R} \tag{5.1}$$

and there exists $\beta > 0$ and $\theta > 1$ such that[4] for all $t \in \mathbb{R}$,

$$\limsup_{N \to \infty} \mathbb{E}[1 \wedge (\mathscr{L}_{t+h}^N - \mathscr{L}_t^N)^\beta \wedge (\mathscr{L}_t^N - \mathscr{L}_{t-h}^N)^\beta] \lesssim h^\theta. \tag{5.2}$$

While (5.1) is true by Proposition 2.1, the main work is to show (5.2).

The proof of Theorem 1 is structured as follows: Step 1 prepares the proof of (5.2) by giving the basis for an estimate on the smaller jump, $\mathscr{L}_t^N - \mathscr{L}_{t-h}^N$ or $\mathscr{L}_{t+h}^N - \mathscr{L}_t^N$. In Step 2 we use this estimate to show (5.2). In Step 3 we show convergence of finite-dimensional distributions. Step 4 then shows (2.2). Finally, Proposition 2.1 already shows that $\frac{1}{2}\mathscr{L}_t$ is Gumbel distributed for all $t \in \mathbb{R}$.

Step 1 (Evolution of inter-coalescence times)**.** Consider the graphical representation of a Moran model given in Figure 1. For any time $t \in \mathbb{R}$ the tree $\mathscr{T}_t^N$ can be identified with a random subset of $(-\infty; t] \times \{1, ..., N\}$ which we continue to denote by $\mathscr{T}_t^N$. For any fixed t we define the inter-coalescence times

$$X_i^N(t) := \text{ length of the time interval in which } \mathscr{T}_t^N \text{ has } i \text{ lines}$$

for $i = 2, ..., N$. For any two time points t and $t + h$, we put

$$F_{(t,t+h]}^N := \min\{i = 2, ..., N : X_i^N(t) \neq X_i^N(t + h)\}.$$

---

[4]For functions $a, b : \mathbb{R}_+ \to \mathbb{R}_+$ we write $a(h) \lesssim b(h)$ iff there is a $C > 0$, independent of any other parameter, such that $a(h) \leq C \cdot b(h)$ for all $h > 0$.

For the distribution of $F^{\mathrm{N}}_{(t,t+h]}$, we claim that

$$\mathbb{P}[F^{\mathrm{N}}_{(t,t+h]} \geq f] = e^{-\binom{f}{2}h}, \quad f = 2, 3, \ldots \tag{5.3}$$

To see this, it is important to note that $\{F^{\mathrm{N}}_{(t,t+h]} \geq f\}$ is the event that none of the f oldest families of the population of size N, which are present at time t, gets extinct by time $t + h$. To obtain the probability of this event consider the random time of the first loss of one of these f oldest families. The distribution of the frequencies of the f oldest families in the population is the quasi-stationary distribution of the Moran model with population size N and f different types. Moreover, the first time at which one of the f oldest families dies out is a boundary hitting time of a Markov process started in its quasi-stationary distribution and hence must be exponentially distributed. When the boundary is hit, there is a new set of f oldest families, whose frequencies are again in the quasi-stationary distribution for the f-type Moran model. This reasoning implies that the times for which one of the f oldest families dies out is a Poisson process.

To obtain the rate of this Poisson process, consider the limit $h \to 0$. In this limit, $F^{\mathrm{N}}_{(t,t+h]} < f$ if and only if one of the f oldest families consists of only one individual which is killed upon the next resampling event. In other words, such a resampling event breaks off the external branch belonging to one of the f oldest families which consists of only a single individual. Given the occurrence of a resampling event, this event has the same probability as the event $\{F^{\mathrm{N}} < f\}$, with F^{N} from Section 4.2. For $h \to 0$ we can assume that there is at most one resampling event in $(t, t + h]$. Therefore,

$$\mathbb{P}[F^{\mathrm{N}}_{(t,t+h]} < f] \overset{h \to 0}{\sim} \binom{\mathrm{N}}{2} h \cdot \frac{\binom{f}{2}}{\binom{\mathrm{N}}{2}} = \binom{f}{2} h$$

and (5.3) follows. Moreover, since the events of loss of one of the f oldest families is a Poisson process, $F^{\mathrm{N}}_{(t-h,t]}$ and $F^{\mathrm{N}}_{(t,t+h]}$ are independent, and thus

$$\mathbb{P}[F^{\mathrm{N}}_{(t-h,t]} \vee F^{\mathrm{N}}_{(t,t+h]} < f] = (1 - e^{-\binom{f}{2}h})^2 \lesssim f^4 h^2 \wedge 1. \tag{5.4}$$

Step 2 (Proof of (5.2)). We will now show (5.2) for $\theta = \frac{10}{9}$ and $\beta = 10$. We write

$$\begin{aligned}
&\mathbb{E}[1 \wedge (\mathscr{L}^{\mathrm{N}}_{t+h} - \mathscr{L}^{\mathrm{N}}_t)^{10} \wedge (\mathscr{L}^{\mathrm{N}}_t - \mathscr{L}^{\mathrm{N}}_{t-h})^{10}] \\
&\quad \leq \mathbb{E}[1 \wedge (\mathscr{L}^{\mathrm{N}}_{t+h} - \mathscr{L}^{\mathrm{N}}_t)^{10}; F^{\mathrm{N}}_{(t,t+h]} \geq F^{\mathrm{N}}_{(t-h,t]}] + \mathbb{E}[1 \wedge (\mathscr{L}^{\mathrm{N}}_t - \mathscr{L}^{\mathrm{N}}_{t-h})^{10}; F^{\mathrm{N}}_{(t-h,t]} \geq F^{\mathrm{N}}_{(t,t+h]}].
\end{aligned} \tag{5.5}$$

We will show how to bound the first term on the right hand, since the bound for the second term is obtained in the same manner. We observe

$$\begin{aligned}
&\mathbb{E}[1 \wedge (\mathscr{L}^{\mathrm{N}}_{t+h} - \mathscr{L}^{\mathrm{N}}_t)^{10}; F^{\mathrm{N}}_{(t,t+h]} \geq F^{\mathrm{N}}_{(t-h,t]}] \\
&\quad \leq \mathbb{P}[F^{\mathrm{N}}_{(t-h,t]} \leq F^{\mathrm{N}}_{(t,t+h]} \leq h^{-2/9}] + \mathbb{E}[(\mathscr{L}^{\mathrm{N}}_{t+h} - \mathscr{L}^{\mathrm{N}}_t)^{10}; F^{\mathrm{N}}_{(t,t+h]} \geq h^{-2/9}] \\
&\quad \leq \mathbb{P}[F^{\mathrm{N}}_{(t-h,t]} \vee F^{\mathrm{N}}_{(t,t+h]} \leq h^{-2/9}] \\
&\qquad\qquad + \mathbb{E}\Big[\Big(\sum_{i=2}^{\mathrm{N}} i\Big(X^{\mathrm{N}}_i(t+h) - \tfrac{1}{\binom{i}{2}}\Big) - \sum_{i=2}^{\mathrm{N}} i\Big(X^{\mathrm{N}}_i(t) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^{10}; F^{\mathrm{N}}_{(t,t+h]} \geq h^{-2/9}\Big] \\
&\quad \lesssim h^{10/9} + \mathbb{E}\Big[\Big(\sum_{i=\lfloor h^{-2/9}\rfloor}^{\mathrm{N}} i\Big(X^{\mathrm{N}}_i(t+h) - \tfrac{1}{\binom{i}{2}}\Big) - \sum_{i=\lfloor h^{-2/9}\rfloor}^{\mathrm{N}} i\Big(X^{\mathrm{N}}_i(t) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^{10}\Big],
\end{aligned} \tag{5.6}$$

where the last inequality follows from (5.4) and the fact that $X_i(t + h) = X_i(t)$ for $i < \lfloor h^{-2/9}\rfloor$ by definition of $F^{\mathrm{N}}_{(t,t+h]}$. It remains to bound the second term in the last line. For this, we define for $n = 1, 2, \ldots$

$$a_n(h) := \sum_{i=\lfloor h^{-2/9}\rfloor}^{\mathrm{N}} \mathbb{E}\Big[\Big(i\Big(X^{\mathrm{N}}_i(t) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^n\Big]$$

and observe that

$$a_1(h) = 0, \qquad a_n(h) \lesssim h^{2(n-1)/9}, \tag{5.7}$$

since the nth central moment of an exponentially distributed random variable with parameter λ is proportional to λ^{-n}. In addition, we use that $(x-y)^n \leq (2x)^n + (2y)^n$ for even n and all $x, y \in \mathbb{R}$, and independence of $X_2^N(t), ..., X_N^N(t)$ as well as of $X_2^N(t+h), ..., X_N^N(t+h)$ to obtain

$$\mathbb{E}\Big[\Big(\sum_{i=\lfloor h^{-2/9}\rfloor}^{N} i\Big(X_i^N(t+h) - \tfrac{1}{\binom{i}{2}}\Big) - \sum_{i=\lfloor h^{-2/9}\rfloor}^{N} i\Big(X_i^N(t) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^{10}\Big]$$

$$\lesssim \mathbb{E}\Big[\Big(\sum_{i=\lfloor h^{-2/9}\rfloor}^{N} i\Big(X_i^N(t+h) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^{10}\Big] + \mathbb{E}\Big[\Big(\sum_{i=\lfloor h^{-2/9}\rfloor}^{N} i\Big(X_i^N(t) - \tfrac{1}{\binom{i}{2}}\Big)\Big)^{10}\Big] \tag{5.8}$$

$$\lesssim \sum_{k=1}^{10} \sum_{\substack{n_1,...,n_k \\ n_1+...+n_k=10}}^{10} a_{n_1}(h) \cdots a_{n_k}(h) \lesssim (a_2(h))^5 \lesssim h^{10/9}.$$

by (5.7). Plugging (5.8) and (5.6) into (5.5) shows that (5.2) holds with $\beta = 10$ and $\theta = \frac{10}{9}$.

Step 3 (Convergence of finite-dimensional distributions). Fix $t_1 < ... < t_n$. We will show that $(\mathscr{L}_{t_1}^N, ..., \mathscr{L}_{t_n}^N)$ converges weakly for $N \to \infty$. The strategy is to define a probability space on which all $\mathscr{L}_{t_i}^N$, $i = 1, ..., n$, $N = 2, 3, ...$, are defined.

Consider a coalescent, started with infinitely many lines which are numbered by $1_n, 2_n, ...$ for some time $t_n - t_{n-1}$. Denote the number of the ancestors at time $t_n - t_{n-1}$ by S_n . Number the lines going back from these by $1_{n-1}, ..., (S_n)_{n-1}$ and augment them by lines numbered $(S_n + 1)_{n-1}, (S_n + 2)_{n-1}, ...$. Let these infinitely many lines coalesce for some time $t_{n-1} - t_{n-2}$, number the S_{n-1} ancestors at time $t_{n-1} - t_{n-2}$ by $1_{n-2}, ..., (S_{n-1})_{n-2}$ and augment their lines by lines numbered $(S_{n-1} + 1)_{n-2}, (S_{n-1} + 2)_{n-2}, ...$. In this way we get iteratively n genealogies for an infinite population back from times $t_1, ..., t_n$. Considering the compensated tree lengths of lines numbered $1_i, ..., N_i$ gives the compensated tree length of a population of size N at time t_i, $i = 1, ..., n$. Moreover, as shown in Lemma 3.1, these tree lengths converge in L^2 as $N \to \infty$ for each $i = 1, ..., n$. Since L^2-convergence implies convergence in probability, which, in turn, implies weak convergence, we are done.

Step 4 (Decomposition of $\mathscr{L}_t^N - \mathscr{L}_0^N$ and proof of (2.2)). Recall the graphical representation of a Moran model from Figure 1. Using the random set $\mathscr{T}_t^N \subseteq (-\infty, t] \times \{1, ..., N\}$ introduced at the beginning of Step 1, we have the representation $(\mathscr{L}_t^N)_{t\in\mathbb{R}} \overset{d}{=} (\lambda^N(\mathscr{T}_t^N) - 2\log N)_{t\in\mathbb{R}}$ where λ^N is Lebesgue measure on $\mathbb{R} \times \{1, ..., N\}$. We set

$$\begin{aligned}
A_{0,t}^N &:= \lambda^N(\mathscr{T}_t^N \setminus \mathscr{T}_0^N) - \mathbb{E}[\lambda^N(\mathscr{T}_t^N \setminus \mathscr{T}_0^N)], \\
B_{0,t}^N &:= \lambda^N(\mathscr{T}_0^N \setminus \mathscr{T}_t^N) - \mathbb{E}[\lambda^N(\mathscr{T}_0^N \setminus \mathscr{T}_t^N)],
\end{aligned} \tag{5.9}$$

compare Figure 3. Note that $\mathbb{E}[\lambda^N(\mathscr{T}_t^N \setminus \mathscr{T}_0^N)] = \mathbb{E}[\lambda^N(\mathscr{T}_0^N \setminus \mathscr{T}_t^N)]$ due to stationarity, and thus

$$\mathscr{L}_t^N - \mathscr{L}_0^N \overset{d}{=} A_{0,t}^N - B_{0,t}^N. \tag{5.10}$$

For the infinitesimal variance, we find by the convergence of finite dimensional distributions and (5.10) that

$$\mathbb{E}[(\mathscr{L}_t - \mathscr{L}_0)^2] = \lim_{N\to\infty}(\mathscr{L}_t^N - \mathscr{L}_0^N)^2] = \lim_{N\to\infty} \mathbb{E}[(A_{0,t}^N - B_{0,t}^N)^2]. \tag{5.11}$$

From (5.9) and (4.18) we conclude that $A_{0,t}^N \overset{d}{=} \Lambda_t^N$, again see Figure 3. We have, using the L^2-convergence from Lemma 4.9 and Lemma 4.10

$$\lim_{N\to\infty} \mathbb{V}[A_{0,t}^N] = \lim_{N\to\infty} \mathbb{V}[\Lambda_t^N] = \mathbb{V}[\Lambda_t] \overset{t\to 0}{\sim} \tfrac{2}{3}t. \tag{5.12}$$

For the variance of $B_{0,t}^{\mathrm{N}}$, note that $B_{0,t}^{\mathrm{N}} \stackrel{d}{=} B_{\mathrm{N}} - B_{S_t^{\mathrm{N}}}$, where B_{N} and S_t^{N} are as in Sections 4.3 and 4.4, with $\mathcal{T} := \mathcal{T}_0$ and S_t^{N} independent. We thus have for fixed t and $\mathrm{N} \to \infty$ because of Proposition 4.4

$$\mathbb{V}[B_{0,t}^{\mathrm{N}}] = \mathbb{V}[B_{\mathrm{N}} - B_{S_t^{\mathrm{N}}}] \sim \mathbb{V}[B_{S_t^{\mathrm{N}}}] \sim \mathbb{V}[B_{S_t}]. \tag{5.13}$$

Since $t \cdot S_t \to 2$ almost surely as $t \to 0$ (see Lemma 4.6), we conclude from (5.13) and Proposition 4.4, (4.8), that

$$\lim_{\mathrm{N} \to \infty} \mathbb{V}[B_{0,t}^{\mathrm{N}}] = \mathbb{V}[B_{S_t}] \stackrel{t \to 0}{\sim} 4t |\log t|. \tag{5.14}$$

Finally, combining (5.11), (5.12) and (5.14) we arrive at

$$\mathbb{E}[(\mathscr{L}_t - \mathscr{L}_0)^2] \stackrel{t \to 0}{\sim} 4t \,|\log t|.$$

This completes the proof of Theorem 1.

6   Proofs of strong convergence results

In this section we prove Propositions 3.1 and 3.2.

6.1   Proof of Proposition 3.1

From Proposition 2.1 we know that Λ_1^{N} converges weakly as $\mathrm{N} \to \infty$ to a random variable Λ such that $\frac{1}{2}\Lambda$ is Gumbel distributed. Since Λ_1^{N} is a sum of independent random variables, Kolmogorov's three series criterion shows that the convergence holds almost surely as well. Moreover, since seconds moments converge in (3.4), the convergence also holds in L^2.

Next, we will show that

$$\Lambda_1^{\mathrm{N}} - \Lambda_2^{\mathrm{N}} \xrightarrow{\mathrm{N} \to \infty} 0$$

in L^2. Together with the L^2-convergence of Λ_1^{N} this gives (3.5). We compute directly, using that $\Lambda_1^{\mathrm{N}} \stackrel{d}{=} \Lambda_2^{\mathrm{N}}$ and Lemma 4.1,

$$
\begin{aligned}
\mathbb{E}\big[(\Lambda_1^{\mathrm{N}} - \Lambda_2^{\mathrm{N}})^2\big] &= \mathbb{E}\Big[\Big(\sum_{i=2}^{\infty} K_i^{\mathrm{N}} X_i - \sum_{i=2}^{\mathrm{N}} i X_i\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i=2}^{\infty} K_i^{\mathrm{N}} X_i\Big)^2\Big] + \mathbb{E}\Big[\Big(\sum_{i=2}^{\mathrm{N}} i X_i\Big)^2\Big] - 2\mathbb{E}\Big[\sum_{i=2}^{\infty}\sum_{j=2}^{\mathrm{N}} K_i^{\mathrm{N}} j X_i X_j\Big] \\
&= 2\Big(\mathbb{E}\Big[\Big(\sum_{i=2}^{\mathrm{N}} i X_i\Big)^2\Big] - \sum_{i=2}^{\infty}\sum_{j=2}^{\mathrm{N}} \frac{ij\mathrm{N}}{\mathrm{N}+i-1} \mathbb{E}[X_i X_j]\Big) \\
&= 2\Big(\mathbb{V}\Big[\sum_{i=2}^{\mathrm{N}} i X_i\Big] + \Big(\mathbb{E}\Big[\sum_{i=2}^{\mathrm{N}} i X_i\Big]\Big)^2 - \sum_{i=\mathrm{N}+1}^{\infty}\sum_{j=2}^{\mathrm{N}} \frac{ij\mathrm{N}}{\mathrm{N}+i-1}\frac{4}{i(i-1)j(j-1)}(1+\delta_{ij})\Big) \\
&= 2\Big(4\sum_{i=1}^{\mathrm{N}-1}\frac{1}{i^2} + \Big(2\sum_{i=1}^{\mathrm{N}-1}\frac{1}{i}\Big)^2 - 4\sum_{j=1}^{\mathrm{N}-1}\frac{1}{j}\sum_{i=1}^{\infty}\frac{\mathrm{N}}{i(\mathrm{N}+i)} - 4\sum_{i=1}^{\mathrm{N}-1}\frac{\mathrm{N}}{i^2(\mathrm{N}+i)}\Big) \\
&= 8\sum_{i=1}^{\mathrm{N}-1}\frac{1}{i^2}\Big(1 - \frac{\mathrm{N}}{\mathrm{N}+i}\Big) + 8\Big(\sum_{i=1}^{\mathrm{N}-1}\frac{1}{i}\Big)^2 - 8\sum_{j=1}^{\mathrm{N}-1}\frac{1}{j}\sum_{i=1}^{\infty}\Big(\frac{1}{i} - \frac{1}{\mathrm{N}+i}\Big) \\
&= 8\sum_{i=1}^{\mathrm{N}-1}\frac{1}{i(\mathrm{N}+i)} = \frac{8}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}-1}\Big(\frac{1}{i} - \frac{1}{\mathrm{N}+i}\Big) \stackrel{\mathrm{N} \to \infty}{\sim} 8\frac{\log \mathrm{N}}{\mathrm{N}} \xrightarrow{\mathrm{N} \to \infty} 0.
\end{aligned}
$$

6.2 Proof of Proposition 3.2

We recall [DK96, Lemma A2.1]:

Proposition 6.1. *Let $(X^n)_{n=1,2,...}$ be a sequence of processes with sample paths in $\mathbb{D}$, defined on the same probability space. Suppose that $(X^n)_{n=1,2,...}$ is relatively compact in $\mathbb{D}$ (in the sense of convergence in distribution) and that for a dense set $H \subseteq \mathbb{R}$, $(X_t^n)_{n=1,2,...}$ converges in probability in $\mathbb{R}$ for each $t \in H$. Then, there is a process X such that $d_{\mathrm{Sk}}(X_n, X) \xrightarrow{n\to\infty} 0$ in probability.*

We use this Proposition for $(\mathscr{L}^{\mathrm{ld,N}})_{N=2,3,...}$. First, $\mathscr{L}^{\mathrm{ld,N}} \stackrel{d}{=} \mathscr{L}^{\mathrm{N}}$ with $\mathscr{L}^{\mathrm{N}}$ as in Theorem 1. Hence, as Theorem 1 shows, $(\mathscr{L}^{\mathrm{ld,N}})_{N=2,3,...}$ converges weakly. In particular, the sequence is relatively compact in $\mathbb{D}$.

For all $t \in \mathbb{R}$, we have that $\mathscr{T}^{\mathrm{ld,N}} \stackrel{d}{=} \mathscr{T}^{\mathrm{N}}$ with $\mathscr{T}^{\mathrm{N}}$ from (3.3). Consequently, $\mathscr{L}^{\mathrm{ld,N}} \stackrel{d}{=} \Lambda_2^{\mathrm{N}}$ and there exists a random variable $\mathscr{L}_t^{\mathrm{ld}}$ such that $\mathscr{L}_t^{\mathrm{ld,N}} - \mathscr{L}_t^{\mathrm{ld}} \xrightarrow{\mathrm{N}\to\infty} 0$ in L^2. Since the L^2-convergence implies convergence in probability we have proved Proposition 3.2.

Acknowledgements

References

[Ald99]   D. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.

[BBS08]   J. Berestycki, N. Berestycki, and J. Schweinsberg. Small-time behavior of beta coalescents. *Annales de l'Institut Henri Poincare (B)*, 44(2):214–238, 2008.

[BF05]   M. G. Blum and O. Francois. Minimal clade size and external branch length under the neutral coalescent. *Adv. Appl. Probab.*, 37:647–662, 2005.

[CNKR07] A. Caliebe, R. Neininger, M. Krawczak, and U. Rösler. On the length distribution of external branches in coalescence trees: genetic diversity within species. *Theo. Pop. Biol.*, 72(2):245–252, 2007.

[Daw93]   D.A. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XXI–1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260, Berlin, 1993. Springer.

[DDSJ09a] J.-F. Delmas, J.-S. Dhersin, and A. Siri-Jegousse. Asymptotic results on the length of coalescent trees. *Ann. Appl. Prob.*, 18(3):997–1025, 2009.

[DDSJ09b] J.-F. Delmas, J.-S. Dhersin, and A. Siri-Jegousse. On the two oldest families for the Wright-Fisher process. *Preprint*, 2009.

[DIMR07] M. Drmota, A. Iksanov, M. Möhle, and U. Rösler. Asymptotic results concerning the total branch length of the Bolthausen−Sznitman coalescent. *Stoch. Proc. Appl.*, 117(10):1404–1421, 2007.

[DK96]   P. Donnelly and T.G. Kurtz. A countable representation of the Fleming Viot measure-valued diffusion. *Ann. Probab.*, 24(2):698–742, 1996.

[DK99]   P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27(1):166–205, 1999.

[Dur08]   R. Durrett. *Probability Models for DNA Sequence Evolution.* Springer, second edition, 2008.

[EK86]   S.N. Ethier and T. Kurtz. *Markov Processes. Characterization and Convergence.* John Wiley, New York, 1986.

[EPW06]   A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. An approximate sampling formula under genetic hitchhiking. *Ann. Appl. Probab.*, 15:685–729, 2006.

[Eth01]   A. Etheridge. *An introduction to superprocesses.* American Mathematical Society, 2001.

[Eva00]   S. Evans. Kingman's coalescent as a random metric space. In *Stochastic Models: Proceedings of the International Conference on Stochastic Models in Honour of Professor Donald A. Dawson, Ottawa, Canada, June 10-13, 1998 (L.G Gorostiza and B.G. Ivanoff eds.)*, Canad. Math. Soc., 2000.

[FL93]   Y.-X. Fu and W.-H. Li. Statistical tests of neutrality of mutations. *Genetics*, 133:693–709, 1993.

[GPW09]   A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics. Martingale problems and applications. *Submitted*, 2009.

[Kin82a] J. F. C. Kingman. The coalescent. *Stoch. Proc. Appl.*, 13(3):235–248, 1982.

[Kin82b] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, 19A:27–43, 1982.

[Möh06]   M. Möhle. On the number of segregating sites for populations with large family sizes. *Adv. Appl. Probab.*, 38:750–767, 2006.

[PW06]   P. Pfaffelhuber and A. Wakolbinger. The process of most recent common ancestors in an evolving coalescent. *Stoch. Proc. Appl.*, 116:1836–1859, 2006.

[RBY04]   E. M. Rauch and Y. Bar-Yam. Theory predicts the uneven distribution of genetic diversity within species. *Nature*, 431:449–452, 2004.

[STW84]   I. W. Saunders, S. Tavaré, and G. A. Watterson. On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Probab.*, 16:471–491, 1984.

[Taj90]   F. Tajima. Relationship between DNA polymorphism and fixation time. *Genetics*, 125:447–454, 1990.

[Tav84]   S. Tavaré. Line-of-descent and genealogical processes and their applications in population genetics models. *Theor. Pop. Biol.*, 26:119–164, 1984.

[Tav04]   S. Tavaré. *Ancestral Inference in Population Genetics, in: Lectures on Probability and Statistics 1188, in: Lecture Notes in Mathematics, vol. 1837.* Springer, 2004.

[Wak08]   J. Wakeley. *Coalescent Theory: An Introduction.* Roberts & Company, 2008.

[Wat75]   G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theo. Pop. Biol.*, 7:256–276, 1975.

[Wat82]   G.A. Watterson. Mutant substitutions at linked nucleotide sites. *Adv. Appl. Prob.*, 14:166–205, 1982.

[WH99]   C. Wiuf and J. Hein. Recombination as a point process along sequences. *Theo. Pop. Biol.*, 55:248–259, 1999.