Dynamics of the evolving
Bolthausen-Sznitman coalecent[*]

Jason Schweinsberg[†]

Abstract

Consider a population of fixed size that evolves over time. At each time, the genealogical structure of the population can be described by a coalescent tree whose branches are traced back to the most recent common ancestor of the population. As time goes forward, the genealogy of the population evolves, leading to what is known as an evolving coalescent. We will study the evolving coalescent for populations whose genealogy can be described by the Bolthausen-Sznitman coalescent. We obtain the limiting behavior of the evolution of the time back to the most recent common ancestor and the total length of the branches in the tree. By similar methods, we also obtain a new result concerning the number of blocks in the Bolthausen-Sznitman coalescent.

1 Introduction

Consider a haploid population of fixed size n that evolves over time. The genealogy of the population at time t can be represented by a coalescent process $(\Pi(s), s \geq 0)$ taking its values in the set of partitions of $\{1, \ldots, n\}$, which is defined so that integers i and j are in the same block of $\Pi(s)$ if and only if the ith and jth individuals in the population at time t have the same ancestor at time $t - s$. The genealogical structure encoded by the coalescent process can also be represented as a tree $\mathcal{T}_n(t)$. The shape of this tree changes over time as the population evolves, leading to what was called in [23] an evolving coalescent. The associated tree-valued stochastic process, for infinite as well as finite populations, was constructed and studied by Greven, Pfaffelhuber, and Winter [19]. Depperschmidt, Greven, and Pfaffelhuber [11] incorporated mutation and selection into the model.

Rather than studying the full tree-valued process, one can follow the evolution of certain properties of the tree that are of interest. One such property is the time back

[†]University of California at San Diego, USA. E-mail: jschwein@math.ucsd.edu

to the most recent common ancestor (MRCA) of the population. This evolution can be described by a process $(A_n(t), t \geq 0)$, where

$$t - A_n(t) = \sup\{s : \text{ one individual at time } s \text{ is the ancestor of all individuals at time } t\}.$$

Note that $A_n(t)$ is the height of the tree $\mathcal{T}_n(t)$. The process $(A_n(t), t \geq 0)$ increases linearly at speed one between jumps, and jumps downward when a new, more recent MRCA is established. A new MRCA is established when the population consists of two families, which coalesce at the time of the MRCA, and then one of these two families dies out. One can also consider the process $(L_n(t), t \geq 0)$, where $L_n(t)$ denotes the sum of the lengths of all branches in the tree $\mathcal{T}_n(t)$. This process is of interest because, assuming that mutations occur at a constant rate θ along each branch of the coalescent tree, $L_n(t)$ should be roughly proportional to the number of distinct mutations observed in the population at time t.

A natural population model to consider is the Moran model [22]. In this model, the population size stays fixed at n. Changes in the population occur at times of a homogeneous Poisson process, and we will scale time so that these changes occur at rate $n(n-1)/2$. At the time of each such change, one of the n individuals is chosen at random to give birth to a new offspring, and independently one of the other $n-1$ individuals is chosen at random to be killed. For any fixed $t \in \mathbb{R}$, the genealogy of the population follows Kingman's coalescent [20], meaning that each pair of lineages merges at rate one and no other transitions are possible. An analogous construction for infinite populations can be carried out using the lookdown construction of Donnelly and Kurtz [14]. The associated evolving coalescent was studied by Pfaffelhuber and Wakolbinger [23]. They showed that the jumps of the process $(A(t), t \geq 0)$ that follows the time back to the MRCA occur at times of a homogeneous Poisson process, but that the process $(A(t), t \geq 0)$ is not Markov. They also calculated the distributions of some other quantities, such as the number of individuals in the population at time t who will have descendants in the population when the next MRCA is established and the number of individuals in the population who will become the MRCA of the population in the future. Delmas, Dhersin, and Siri-Jegousse [12] extended these results by considering also the distribution of the sizes of the two oldest families at time t. Simon and Derrida [32] did some further work related to the evolution of the time back to the MRCA for populations with genealogies governed by Kingman's coalescent, and considered correlations between the time back to the MRCA and a measure of genetic diversity. Pfaffelhuber, Wakolbinger, and Weisshaupt [24] studied the evolution of the total branch length. They showed that the sequence of processes $(L_n(t) - 2\log n, t \in \mathbb{R})$ converges weakly as $n \to \infty$ to a limit process which is a stationary process with infinite infinitesimal variance.

Evans and Ralph [17] studied the dynamics of the time back to the MRCA in a population in which a single "immortal particle" produces offspring at times of a Poisson process, and descendants of the offspring eventually die out. In this setting, the process $(A(t), t \geq 0)$ is a Markov process whose jump rates and stationary distribution can be calculated explicitly. An example of a process that fits into this framework is the α-stable continuous-state branching process conditioned on nonextinction with $1 < \alpha \leq 2$.

The goal of the present paper is to determine the dynamics of the time back to the MRCA and the total branch length for populations whose genealogy is given by the Bolthausen-Sznitman coalescent. The Bolthausen-Sznitman coalescent, which was introduced in [6], is an example of a coalescent with multiple mergers [25, 28], in which it is possible for many lineages to merge at once. More precisely, the Bolthausen-Sznitman coalecent started with n blocks is a continuous-time Markov chain taking its values in the set of partitions of $\{1, \ldots, n\}$ such that whenever the partition has b blocks,

each transition that involves the merger of k blocks into one is happening at rate

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}\,dx = \frac{(k-2)!(b-k)!}{(b-1)!}. \tag{1.1}$$

This means that the total rate of all transitions when there are b blocks is

$$\lambda_b = \sum_{k=2}^{b} \binom{b}{k}\lambda_{b,k}.$$

It is well-known (see, for example, [25]) that this process has the property of sampling consistency, meaning that if $m < n$, then the process restricted to the integers $\{1, \ldots, m\}$ is a Bolthausen-Sznitman coalescent started with m blocks.

The questions about the evolution of the time back to the MRCA and the total branch length could also be considered for other coalescents with multiple mergers besides the Bolthausen-Sznitman coalescent. However, some of the tools used here, especially the recursive tree construction introduced in section 2, are specific to the Bolthausen-Sznitman coalescent and do not have a counterpart for other coalescents with multiple mergers. Consequently, our results and methods are restricted to this case. Nevertheless, the Bolthausen-Sznitman coalescent has arisen in a variety of settings. It has been shown recently to describe the genealogy of certain populations undergoing selection [3, 8]. The Bolthausen-Sznitman coalescent also describes the genealogical structure of Neveu's continuous-state branching process [5], certain Galton-Watson processes with heavy-tailed offspring distributions [31], and Derrida's generalized random energy model [6, 7].

We describe the population model that we will study in section 1.1. In section 1.2, we state our main result concerning the dynamics of the time back to the MRCA. In section 1.3, we state our main result concerning the total branch length. By similar methods, we also obtain a new result concerning the number of blocks of the Bolthausen-Sznitman coalescent. We state this result in section 1.4. The rest of the paper is devoted to proofs.

1.1 A population model

We now define more precisely the population model that we will study in this paper. We assume that for all times $t \in \mathbb{R}$, there are exactly n individuals in the population, labeled by the integers $1, \ldots, n$. Changes in the population occur at times of a homogeneous Poisson process on $\mathbb{R}$ with rate $n-1$. At the time of such a change, one particle is chosen at random to give birth to a random number ξ of new offspring, with

$$P(\xi = k) = \frac{n}{n-1} \cdot \frac{1}{k(k+1)}, \quad k = 1, 2, \ldots, n-1. \tag{1.2}$$

Then ξ of the $n-1$ individuals who did not give birth are chosen at random to be killed, and the new individuals take over the labels of the individuals who were killed.

We now give a representation of the genealogy of this population. For each $s \in \mathbb{R}$ and $t \geq 0$, let $\Pi_n(s,t)$ be the partition of $1, \ldots, n$ such that i and j are in the same block of $\Pi_n(s,t)$ if and only if the individuals at time s labeled i and j are descended from the same ancestor at time $(s-t)-$. We consider the population at time $(s-t)-$, rather than exactly at time $s-t$, to ensure that for each $s \in \mathbb{R}$, the process $(\Pi_n(s,t), t \geq 0)$ is right continuous. As long as there is no change in the population at time s, the partition $\Pi_n(s,0)$ consists of n singletons. However, if k new individuals are born at time s, then $\Pi_n(s,0)$ will consist of one block of size $k+1$ and $n-k-1$ singleton blocks.

Proposition 1.1. *Fix $s \in \mathbb{R}$. Then $(\Pi_n(s,t), t \geq 0)$ is the Bolthausen-Sznitman coalescent started with n blocks.*

Proof. With probability one, there is no change in the population at time s, and $\Pi_n(s,0)$ consists of n singletons. When ancestral lines are followed backwards in time, an event in which $k-1$ individuals are born becomes an event in which k randomly chosen lineages merge, because the ancestral lines of the $k-1$ children merge with that of the parent. The rate of events in which $k-1$ individuals are born is

$$(n-1)P(\xi = k-1) = \frac{n}{k(k-1)} = \binom{n}{k}\lambda_{n,k}, \tag{1.3}$$

where $\lambda_{n,k}$ comes from (1.1). Therefore, $(\Pi_n(s,t), t \geq 0)$ follows the dynamics of the Bolthausen-Sznitman coalescent up to the time of the first merger. That the process continues to follow the dynamics of the Bolthausen-Sznitman coalescent after this time is a consequence of the exchangeability of the population model and the sampling consistency of the Bolthausen-Sznitman coalescent. □

For $s \in \mathbb{R}$ and $t \geq 0$, let $N_n(s,t)$ be the number of blocks of the partition $\Pi_n(s,t)$. Let

$$A_n(s) = \inf\{t : N_n(s,t) = 1\}, \tag{1.4}$$

which is the time back to the MRCA of the population and corresponds to the height of the tree $\mathcal{T}_n(s)$ that represents the genealogy of the population at time s. Let

$$L_n(s) = \int_0^\infty N_n(s,t)\mathbf{1}_{\{N_n(s,t)>1\}},$$

which is the sum of the lengths of all branches in the tree $\mathcal{T}_n(s)$.

1.2 Time back to the MRCA

We consider here how the time back to the MRCA of the population evolves over time. Proposition 3.4 of [18] states that if Y has the exponential distribution with mean 1, then for each fixed $t \in \mathbb{R}$, we have

$$A_n(t) - \log\log n \Rightarrow -\log Y, \tag{1.5}$$

where $\Rightarrow$ denotes convergence in distribution as $n \to \infty$. We are interested here in finding the limit of the stochastic processes $(A_n(t), t \geq 0)$ as $n \to \infty$.

We now construct the limit process $(A(t), t \in \mathbb{R})$ and its time-reversal $(R(t), t \in \mathbb{R})$ from a Poisson point process. The construction is similar to constructions in [16] and [17]. Let N be a Poisson process on $\mathbb{R}^2$ with intensity $\lambda \times \nu$, where λ is Lebesgue measure and $\nu(dy) = e^{-y}\,dy$. Let M consist of the points $\{(t,y) : (-t,y) \text{ is a point of } N\}$. Note that M is also a Poisson process on $\mathbb{R}^2$ with the same intensity as N. For each $(t,x) \in \mathbb{R}^2$, define the wedges

$$W(t,x) = \{(s,y) \in \mathbb{R}^2 : s \leq t \text{ and } y \geq x+t-s\},$$
$$W'(t,x) = \{(s,y) \in \mathbb{R}^2 : s > t \text{ and } y \geq x+s-t\}.$$

Then let

$$R(t) = \sup\{x : \text{ there is a point of } N \text{ in } W(t,x)\},$$
$$A(t) = \sup\{x : \text{ there is a point of } M \text{ in } W'(t,x)\}.$$

To see that these quantities are well-defined, observe that if $t \in \mathbb{R}$ and $x \in \mathbb{R}$, then the event that there is no point of N in $W(t,x)$ has probability

$$\exp\left(-\int_{-\infty}^t \int_{x+t-s}^\infty e^{-y}\,dy\,ds\right) = \exp(-e^{-x}). \tag{1.6}$$

Note also that both $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ are right continuous, and because of the relationship between N and M, we have $A(t) = R(-t)$ for all t such that there is no point in M whose first coordinate is t. If (t, z) is a point of M and $R((-t)-) < z$, then $R(-t) = z$ and $A(t) = R((-t)-)$.

The figure below shows how $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ are constructed from the Poisson point process. The process $(R(t), t \in \mathbb{R})$ decreases linearly at speed one between jumps but jumps up to the level of any point of N that appears above it. That is, if (t, y) is a point of N and $R(t-) < y$, then $R(t) = y$. The process $(A(t), t \in \mathbb{R})$ increases linearly at speed one between jumps. When the trajectory of the process encounters a point of M at time t, the process jumps downward to the highest level y such that there is a point of M on the half-line starting at (t, y) and extending diagonally upward and to the right, in the direction of the vector $(1, 1)$.
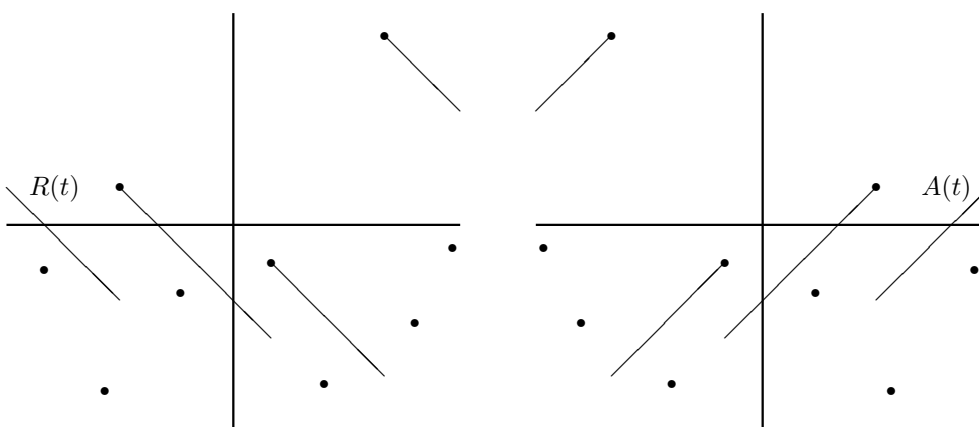

Figure 1: The processes $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$

The following theorem is our main result concerning the time back to the MRCA for the evolving Bolthausen-Sznitman coalescent. This result is proved in Section 2 by considering the processes in reversed time and obtaining convergence to $(R(t), t \geq 0)$. Note that unlike in the case of Kingman's coalescent, the limit process is Markov.

Theorem 1.2. *As $n \to \infty$, the sequence of processes $((A_n(t) - \log \log n), t \geq 0)$ converges weakly to $(A(t), t \geq 0)$.*

Remark 1.3. From the form of the Poisson process M, we see that $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ are stationary processes. Note that if Y has the exponential distribution with mean one, then $P(-\log Y \leq x) = \exp(-e^{-x})$. Therefore, it follows from (1.6) that the stationary distribution of $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ is the same as the distribution of $-\log Y$. This result also follows from Theorem 1.2 and (1.5).

Remark 1.4. The processes $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ are both examples of piecewise deterministic Markov processes, a class of processes whose theory was developed by Davis [9, 10]. These processes are characterized by their deterministic behavior between jump times, which is linear drift for the processes $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$, and their jump rates. The jump rates for $(R(t), t \in \mathbb{R})$ can easily be read off the Poisson process N. The process $(R(t), t \in \mathbb{R})$ jumps away from x at rate e^{-x}, and when it jumps away from x, the distribution of the location to which it jumps has density $e^{x-y}\mathbf{1}_{\{y \geq x\}}$. This means that the rate of jumps from x to y is given by $q(x, y) = e^{-y}\mathbf{1}_{\{y \geq x\}}$.

To obtain the jump rates for $(A(t), t \in \mathbb{R})$, note that if $A(t) = x$ then there is a point of M at $(t + s, x + s)$ for some $s \geq 0$ but no points above this diagonal line. Because the

density of the intensity measure of M at $(t + s, x + s)$ is $e^{-(x+s)}$, we see that conditional on $A(t) = x$, the distribution of the time before the next jump is exponential with mean 1. That is, for all $x \in \mathbb{R}$, the process $(A(t), t \in \mathbb{R})$ jumps away from x at rate one, which implies that the jump times of $(A(t), t \in \mathbb{R})$ form a homogeneous Poisson process of rate one on $\mathbb{R}$. If the process jumps away from x at time t, then the probability that it jumps below y is the probability that there is no point of M in the trapezoidal region $\{(s, z) : s > t \text{ and } y + s - t \le z \le x + s - t\}$, which is

$$\exp\left(- \int_t^\infty \int_{y+s-t}^{x+s-t} e^{-z} \, dz \, ds \right) = \exp(e^{-x} - e^{-y}).$$

Differentiating with respect to y, we see that the rate of jumps from x to y is given by $r(x, y) = \exp(e^{-x} - e^{-y} - y)\mathbf{1}_{\{y \le x\}}$.

As a check on these formulas, let $\pi(y) = e^{-y}e^{-e^{-y}}$ for $y \in \mathbb{R}$, which is the density of the stationary distribution, obtained by differentiating the right-hand side of (1.6). Then note that $\pi(x)q(x, y) = \pi(y)r(y, x)$ for all $x, y \in \mathbb{R}$, as expected given that $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ are related by time reversal.

1.3 Total branch length

Theorem 5.2 of [13] establishes that for each fixed $t \in \mathbb{R}$,

$$\frac{(\log n)^2}{n}\left(L_n(t) - \frac{n}{\log n} - \frac{n \log \log n}{(\log n)^2} \right) \Rightarrow X, \tag{1.7}$$

where, using γ to denote Euler's constant,

$$E[e^{iuX}] = \exp\left(- \frac{\pi}{2}|u| + iu \log|u| \right)$$

$$= \exp\left(iu(1 - \gamma) - \int_{-\infty}^0 \left(1 - e^{iux} + iux\mathbf{1}_{\{|x| \le 1\}}\right)x^{-2} \, dx \right). \tag{1.8}$$

We consider now the stochastic process $L_n = (L_n(t), t \ge 0)$. If there are no changes in the population between times t and $t+s$, then the tree $\mathcal{T}_n(t+s)$ is obtained by starting with the tree $\mathcal{T}_n(s)$ and then adding a segment of length s to each of the n branches. Consequently, the process L_n increases at speed n between jumps. However, if k individuals die at time t, then to obtain the tree $\mathcal{T}_n(t)$ from $\mathcal{T}_n(t-)$, we must remove the k external branches corresponding to the k individuals who died. This causes a downward jump in L_n. Our main result concerning the dynamics of the total branch length for the Bolthausen-Sznitman coalescent is that the processes $(L_n(t), t \ge 0)$, properly centered and scaled, converge to a stable process of Ornstein-Uhlenbeck type.

We now review some facts about processes of Ornstein-Uhlenbeck type, which can be found in Chapter 17 of [30]. Suppose $(Z(t), t \ge 0)$ is a Lévy process such that

$$E[e^{iuZ(t)}] = \exp\left(iaut - \frac{bu^2 t}{2} - t \int_{-\infty}^\infty \left(1 - e^{iux} + iux\mathbf{1}_{\{|x| \le 1\}}\right) \nu(dx) \right).$$

Given $c > 0$ and a random variable $X(0)$, there is a unique process $(X(t), t \ge 0)$ having paths that are almost surely right continuous with left limits such that

$$X(t) = X(0) + Z(t) - c \int_0^t X(s) \, ds \tag{1.9}$$

almost surely. Following [30], we call $(X(t), t \ge 0)$ the process of Ornstein-Uhlenbeck type generated by (a, b, ν, c). As long as

$$\int_{|x| > 2} \log|x| \, \nu(dx) < \infty,$$

the process $(X(t), t \geq 0)$ has a unique stationary distribution μ, with characteristic function

$$\int_{-\infty}^{\infty} e^{iux} \mu(dx) = \exp\left(i\alpha u - \frac{\beta u^2}{2} - \int_{-\infty}^{\infty} \left(1 - e^{iux} + iux\mathbf{1}_{\{|x|\leq 1\}}\right) \rho(dx)\right), \qquad (1.10)$$

where $\beta = b/2c$, $\alpha = (a + \nu((1,\infty)) - \nu((-\infty,-1)))/c$, and for all Borel sets B,

$$\rho(B) = \frac{1}{c} \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathbf{1}_B(e^{-s}y) \, ds \, \nu(dy)$$

(see Theorem 17.5 of [30]). In this case, if $X(0)$ has distribution μ, then the process $(X(t), t \geq 0)$ is stationary.

Theorem 1.5. *Let ν be the measure on $\mathbb{R}$ whose density with respect to Lebesgue measure is given by $x^{-2}\mathbf{1}_{(-\infty,0)}(x)$. Let $(L(t), t \geq 0)$ be a stationary process of Ornstein-Uhlenbeck type generated by $(2 - \gamma, 0, \nu, 1)$. As $n \to \infty$, the sequence of processes*

$$\left(\frac{(\log n)^2}{n}\left(L_n\left(\frac{t}{\log n}\right) - \frac{n}{\log n} - \frac{n \log \log n}{(\log n)^2}\right), t \geq 0\right)$$

converges weakly to $(L(t), t \geq 0)$.

Remark 1.6. The stationary distribution of $(L(t), t \geq 0)$ has characteristic function given by (1.10) with $\alpha = 1 - \gamma$, $\beta = 0$, and $\rho = \nu$, which matches the right-hand side of (1.8). To see this, observe that $\nu((1,\infty)) = 0$ and $\nu((-\infty,1)) = 1$, so $\alpha = (2-\gamma)-1 = 1-\gamma$. Also, for all $z > 0$,

$$\rho((-\infty,-z]) = \int_{-\infty}^{0} \int_{0}^{\infty} \mathbf{1}_{(-\infty,-z]}(e^{-s}y)y^{-2} \, ds \, dy$$

$$= \int_{0}^{\infty} \int_{-\infty}^{-ze^s} y^{-2} \, dy \, ds = \frac{1}{z} = \nu((-\infty,-z]). \qquad (1.11)$$

Thus, the convergence implied by Theorem 1.5 for each fixed t is consistent with the result (1.7).

1.4 Number of blocks

The techniques used to establish Theorem 1.5 can also be used to prove a new result about how the number of blocks of the Bolthausen-Sznitman coalescent changes over time. Because the number of blocks for other coalescents with multiple mergers has been studied in some depth (see, for example, [1, 2]), we believe that this result may be of independent interest.

Theorem 1.7. *Let $(\Pi_n(t), t \geq 0)$ be a Bolthausen-Sznitman coalescent started with n blocks. Let $N_n(t)$ be the number of blocks of $\Pi_n(t)$, and let*

$$X_n(t) = \frac{\log n}{n}\left(N_n\left(\frac{t}{\log n}\right) - ne^{-t} - \frac{nte^{-t}\log\log n}{\log n}\right).$$

Let $(S(t), t \geq 0)$ be a stable Lévy process satisfying

$$E[e^{iuS(t)}] = \exp\left(-\frac{\pi t}{2}|u| + itu\log|u|\right).$$

As $n \to \infty$, the sequence of processes $(X_n(t), t \geq 0)$ converges weakly to

$$\left(e^{-t}S(t) + \frac{e^{-t}t^2}{2}, t \geq 0\right).$$

2 Proof of Theorem 1.2

2.1 Construction from random recursive trees

Our proof of Theorem 1.2 uses a connection between the Bolthausen-Sznitman coalescent and random recursive trees which was discovered by Goldschmidt and Martin [18]. Suppose $\ell_1, \ldots, \ell_n$ are labels, with an ordering so that $\ell_1 < \cdots < \ell_n$. A random recursive tree with vertices labeled $\ell_1, \ldots, \ell_n$ can be constructed inductively as follows. The vertex labeled ℓ_1 is the root. For $k \geq 2$, once the vertices labeled $\ell_1, \ldots, \ell_{k-1}$ have been placed in the tree, the vertex labeled ℓ_k is attached to a vertex chosen uniformly at random from those labeled $\ell_1, \ldots, \ell_{k-1}$.

Consider a random recursive tree in which the labels are disjoint subsets of $\mathbb{N}$, where $\ell_i < \ell_j$ if the smallest element of ℓ_i is less than the smallest element of ℓ_j. Suppose e is an edge in the tree connecting vertices x and y, where x is closer to the root than y. We can cut the tree at the edge e by deleting the edge e from the tree as well as the entire subtree below e. That is, we remove all vertices z such that the shortest path from the root to z goes through x. All integers that are in labels of vertices that are removed from the tree are then added to the label of x. When a random recursive tree is cut at a randomly chosen edge, the remaining tree is a random recursive tree on the new set of labels (see Proposition 2.1 of [18]).

To establish the connection with the Bolthausen-Sznitman coalescent, start with a random recursive tree on n vertices, labeled with the one-element sets $\{1\}, \ldots, \{n\}$. Then to each edge, assign an independent exponential random variable with mean 1, whose value gives the time at which the edge is cut. For all $t \geq 0$, let $\Pi_n(t)$ denote the partition of $\{1, \ldots, n\}$ such that i and j are in the same block of $\Pi_n(t)$ if and only if the integers i and j are in the same vertex label at time t. Then $(\Pi_n(t), t \geq 0)$ is the Bolthausen-Sznitman coalescent started with n blocks (see Proposition 2.2 of [18]). Because the last transition always involves deleting an edge adjacent to the root, the time back to the MRCA for this Bolthausen-Sznitman coalescent is the maximum of the exponential random variables assigned to the edges adjacent to the root. In [18], Goldschmidt and Martin used this fact to prove (1.5).

We now use recursive trees to construct an evolving Bolthausen-Sznitman coalescent in reversed time. Note that the dynamics of an evolving Bolthausen-Sznitman coalescent in reversed time are the same as the dynamics of an ordinary Bolthausen-Sznitman coalescent, except that whenever k lineages are lost due to a merger, these lineages are replaced by k new lineages. We account for this in the construction by adding new vertices to the tree to replace those lost when an edge is cut.

Vertex labels, instead of being subsets of $\mathbb{N}$, will be ordered pairs (t, i), where $t \geq 0$ and $i \in \{1, \ldots, n\}$. We can order such labels by saying that $(s, i) < (t, j)$ if either $s < t$ or if $s = t$ and $i < j$, so the notion of a recursive tree with labels of this form is well-defined.

To carry out our construction, begin with a random recursive tree with n vertices labeled $(0, 1), \ldots, (0, n)$. Assign an independent exponential random variable with mean 1 to each edge. Denote the resulting tree by $T_n(0)$. The tree-valued process $(T_n(t), t \geq 0)$ then evolves in time as follows. The edge labels decrease linearly at speed one. Suppose that, at time τ, one of the edge labels hits zero. Suppose the edge whose label hits zero connects (s_0, i_0) to a subtree consisting of vertices labeled $(s_1, i_1), \ldots, (s_k, i_k)$. We then cut the tree at this edge, thereby removing the edge along with the k vertices below it. We next pick an integer j uniformly at random from $\{0, 1, \ldots, k\}$, and we make (s_j, i_j) the new label of the vertex that was previously labeled (s_0, i_0). We then add k new vertices to the tree, labeled $(\tau, i_0), \ldots, (\tau, i_{j-1}), (\tau, i_{j+1}), \ldots, (\tau, i_k)$, one at a time in increasing order of their labels, by attaching them to randomly chosen vertices of the existing tree. Note that the tree changes each time a vertex is added, so once m

of the k vertices have been replaced, there are $n - k + m$ different vertices to which the next vertex could be attached. Finally, the k new edges are assigned independent exponential random variables, and the resulting tree is $T_n(\tau)$. We call $i_0, i_1, \ldots, i_k$ the vertices involved in the cut at time τ, and we call i_j the vertex preserved by the cut at time τ. The process then continues to evolve according to the same rules.

To see that the process $(T_n(t), t \geq 0)$ encodes an evolving Bolthausen-Sznitman coalescent, we will show how to couple the process with the population model defined in section 1.1. The vertices involved in a cut at time τ will correspond to lineages in the population that merge at time $-\tau$. Let $0 < \tau_1 < \tau_2 < \ldots$ be the times at which edges are cut from the tree. Set $\tau_0 = 0$, and let $\eta_m = \tau_m - \tau_{m-1}$ for $m \geq 1$. For $m = 1, 2, \ldots$, let S_m be the set consisting of the vertices involved in the cut at time τ_m, and let $|S_m|$ be the cardinality of S_m. Let B_m denote the vertex preserved by the cut at time τ_m. We can now construct a population model, defined over the time period $(-\infty, 0]$, as follows. For all times $t \in (-\infty, 0]$, there are n individuals in the population labeled $1, \ldots, n$. Changes in the population occur at the times $-\tau_m$ for $m = 1, 2, \ldots$. At time $-\tau_m$, the individuals with labels in $S_m \setminus B_m$ are killed, and the individual B_m gives birth to $|S_m| - 1$ new offspring who take over the labels of the individuals who were killed.

In the population model defined in section 1.1, changes in the population happen at times of a Poisson process with rate $\lambda_n = n - 1$. Also, from (1.2) we see that the probability that, at the time of such a change, individual j gives birth while individuals $i_1, \ldots, i_k$ are killed is

$$\frac{1}{n} \binom{n-1}{k}^{-1} P(\xi = k) = \frac{(k-1)!(n-k-1)!}{(n-1)!(n-1)(k+1)}$$

because, given that $\xi = k$, there are n choices for the individual who could give birth and $\binom{n-1}{k}$ choices for the individuals who could be killed. Therefore, the following result shows that the population model defined from $(T_n(t), t \geq 0)$ has the same law as the population process defined in section 1.1.

Proposition 2.1. *For all positive integers m, the distribution of η_m is exponential with rate $n - 1$. For all positive integers m and all distinct integers $j, i_1, \ldots, i_k$ in $\{1, \ldots, n\}$, we have*

$$P(B_m = j, S_m = \{j, i_1, \ldots, i_k\}) = \frac{(k-1)!(n-k-1)!}{(n-1)!(n-1)(k+1)}, \tag{2.1}$$

and the pair (B_m, S_m) is independent of η_m. Also, the triples $(\eta_1, B_1, S_1), (\eta_2, B_2, S_2), \ldots$ are independent.

Proof. For $m \geq 0$, let $\mathcal{F}_m$ be the σ-field generated by $(\eta_1, B_1, S_1), \ldots, (\eta_m, B_m, S_m)$. We claim that for all integers $m \geq 0$, the conditional distribution of $T_n(\tau_m)$ given $\mathcal{F}_m$ is the distribution of a random recursive tree with its set of labels, with independent mean one exponential random variables assigned to the edges. We prove this claim by induction. Because $T_n(0)$ is a random recursive tree, the result is immediate when $m = 0$. Suppose the result holds for some $m \geq 0$. Recall that when a random recursive tree is cut at a randomly chosen edge, causing k vertices to be removed, the remaining tree is a random recursive tree on the new set of $n - k$ labels. Therefore, using the induction hypothesis, after $|S_{m+1}| - 1$ vertices are removed from the tree $T_n(\tau_m)$, the distribution of the remaining tree is that of a random recursive tree on the remaining $n - |S_{m+1}| + 1$ labels. Because $T_n(\tau_{m+1})$ is obtained from this tree by adding $|S_{m+1}| - 1$ new vertices to randomly chosen edges, it follows that, ignoring the edge labels, the conditional distribution of $T_n(\tau_{m+1})$ given $\mathcal{F}_{m+1}$ is that of a random recursive tree on n vertices with the given set of vertex labels. Furthermore, by the memoryless property of the exponential distribution, the labels assigned to the edges of $T_n(\tau_{m+1})$ are independent

mean one exponential random variables, conditional on $\mathcal{F}_{m+1}$. The claim follows by induction.

Suppose $m \geq 0$. Because, conditional on $\mathcal{F}_m$, the labels assigned to the edges of $T_n(\tau_m)$ are independent mean one exponential random variables, the conditional distribution of η_{m+1} given $\mathcal{F}_m$ is the distribution of the minimum of $n-1$ mean one exponential random variables, which is the exponential distribution with rate $n-1$. Also, because, conditional on $\mathcal{F}_m$, the tree $T_n(\tau_m)$ is a random recursive tree with its given set of labels, it follows from the result of Goldschmidt and Martin that if $j, i_1, \ldots, i_k$ are distinct integers in $\{1, \ldots, n\}$, then

$$P(S_{m+1} = \{j, i_1, \ldots, i_k\}|\mathcal{F}_m) = \frac{\lambda_{n,k+1}}{\lambda_n} = \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}.$$

Because B_{m+1} is chosen uniformly at random from the set S_{m+1}, it follows that

$$P(B_{m+1} = j, S_{m+1} = \{j, i_1, \ldots, i_k\}|\mathcal{F}_m) = \frac{(k-1)!(n-k-1)!}{(n-1)!(n-1)(k+1)}.$$

Furthermore, (B_{m+1}, S_{m+1}) and η_{m+1} are conditionally independent given $\mathcal{F}_m$. These observations imply the proposition. □

Let $M_n(t)$ denote the maximum of the exponential random variables assigned to the edges that are adjacent to the root in the tree $T_n(t)$. Then $M_n(t)$ is the time back to the MRCA of the population at time $-t$, as long as there is no change in the population exactly at time $-t$. More precisely, defining A_n as in (1.4) for the population model defined from $(T_n(t), t \geq 0)$, we have

$$M_n(t) = A_n((-t)-)$$

for all $t \geq 0$. Therefore, if we let $R_n(t) = M_n(t) - \log\log n$, then $(R_n(t), t \geq 0)$ would have the same law as $(A_n(-t) - \log\log n, t \geq 0)$ if the process $(A_n(-t), t \geq 0)$ were modified at the jump times to make it right-continuous rather than left-continuous. Consequently, if we can show that the processes $(R_n(t), t \geq 0)$ converge weakly to $(R(t), t \geq 0)$, then in view of the stationarity of the population model and the relationship between $(R(t), t \in \mathbb{R})$ and $(A(t), t \in \mathbb{R})$ via time reversal, it will follow that the processes $(A_n(t) - \log\log n, t \geq 0)$ must converge weakly to $(A(t), t \geq 0)$, which is the content of Theorem 1.2.

2.2 A heuristic argument

To understand heuristically why Theorem 1.2 is true, note that if $R_n(t) \leq z$, then the process R_n jumps above z only when a new vertex is attached to the root, and the random variable assigned to the new edge is greater than $\log\log n + z$. Because the number of blocks in the Bolthausen-Sznitman coalescent decreases by $k-1$ whenever k blocks merge into one, the rate of events that cause $k-1$ blocks to be lost is $\binom{n}{k}\lambda_{n,k}$. Therefore, the total rate of decrease in the number of blocks, and thus the rate at which new vertices are being added to the tree, is

$$\eta_n = \sum_{k=2}^{n} (k-1)\binom{n}{k}\lambda_{n,k} = \sum_{k=2}^{n} \frac{n}{k} \approx n\log n.$$

As long as not too many vertices are cut away from the tree at once, the probability that a new vertex attaches to the root is approximately $1/n$. The probability that the

exponential random variable assigned to the new edge is greater than $\log \log n + z$ is $e^{-z}/\log n$. Hence, the rate at which the process R_n jumps above z is approximately

$$(n \log n)\left(\frac{1}{n}\right)\left(\frac{e^{-z}}{\log n}\right) = e^{-z},$$

in agreement with the dynamics of the process $(R(t), t \geq 0)$. The rest of the proof consists of making these ideas rigorous.

2.3 Lemmas pertaining to random recursive trees

We prove here two lemmas related to random recursive trees that will be used later in the proof of Theorem 1.2.

Lemma 2.2. *For all $\varepsilon > 0$ and $z \in \mathbb{R}$, the probability that there exists $u \in [0, \varepsilon]$ such that $R_n(u) \neq R_n(u-)$ and $R_n(u) > z$ is at most $\varepsilon(1 + 2e^{-z})$.*

Proof. The total rate of decrease in the number of blocks for the Bolthausen-Sznitman coalescent is given by

$$\eta_n = \sum_{k=2}^{n} \frac{n}{k} \leq n \log n. \tag{2.2}$$

The rate of events in which at least half the blocks are lost is

$$\sum_{k=\lceil n/2 \rceil + 1}^{n} \binom{n}{k} \lambda_{n,k} = \sum_{k=\lceil n/2 \rceil + 1}^{n} \frac{n}{k(k-1)} = n\left(\frac{1}{\lceil n/2 \rceil} - \frac{1}{n}\right) \leq 1.$$

By the construction, for each block that is lost due to a merger in the Bolthausen-Sznitman coalescent, a new vertex is added to the tree. As long as the merger causes at most half the blocks to disappear, the probability that each new vertex attaches to the root is bounded by $2/n$. Furthermore, when a new vertex attaches to the root, the probability that it causes the process $(R_n(t), t \geq 0)$ to jump above z is the same as the probability that an exponential random variable with mean 1 is greater than $\log \log n + z$, which is $e^{-(\log \log n + z)} = e^{-z}/\log n$. Consequently, in view of (2.2), the rate of events that cause the process $(R_n(t), t \geq 0)$ to jump above z is at most

$$1 + \eta_n \cdot \frac{2}{n} \cdot \frac{e^{-z}}{\log n} \leq 1 + 2e^{-z}.$$

Thus, the probability that such an event happens before time ε is bounded above by $1 - e^{-\varepsilon(1+2e^{-z})} \leq \varepsilon(1 + 2e^{-z})$, which is the desired conclusion. $\square$

Lemma 2.3. *Consider a random recursive tree with vertices labeled $1, \ldots, n$. Let d_k be the depth of the vertex labeled k, which is the number of edges on the path from the root to k, and let $D_n = d_1 + \cdots + d_n$. Then*

$$E[D_n] = n\left(\sum_{k=1}^{n} \frac{1}{k} - 1\right). \tag{2.3}$$

Also, there exists a positive constant C such that

$$\mathrm{Var}(D_n) \leq Cn^2. \tag{2.4}$$

Proof. We first prove (2.3) by induction. Because the vertex labeled 1 is the root vertex, which has depth zero, clearly $E[D_1] = 0$, verifying (2.3). Suppose (2.3) holds for $n = m - 1$, where $m \geq 2$. Let $\mathcal{F}_{m-1} = \sigma(d_1, \ldots, d_{m-1})$. Recall that the random recursive

tree can be constructed so that the vertex labeled m is attached to one of the previous $m-1$ vertices at random. Consequently, the level of the vertex labeled m is the level of a randomly chosen previous vertex incremented by one, so

$$E[d_m | \mathcal{F}_{m-1}] = 1 + \frac{d_1 + \cdots + d_{m-1}}{m-1} = 1 + \frac{D_{m-1}}{m-1} \tag{2.5}$$

and thus

$$E[d_m] = 1 + \frac{E[D_{m-1}]}{m-1}.$$

Therefore, using the induction hypothesis,

$$E[D_m] = E[D_{m-1}] + E[d_m] = 1 + \frac{m}{m-1} E[D_{m-1}] = 1 + m\left(\sum_{k=1}^{m-1} \frac{1}{k} - 1\right) = m\left(\sum_{k=1}^{m} \frac{1}{k} - 1\right),$$

which implies that (2.3) holds for all $n \in \mathbb{N}$.

We next show by induction that

$$\mathrm{Var}(D_n) \leq n^2 \sum_{k=1}^{n} \frac{E[d_k^2]}{k^2}. \tag{2.6}$$

The result is clear when $n = 1$ because $\mathrm{Var}(D_1) = 0$. Suppose the claim holds for $n = m - 1$, where $m \geq 2$. By (2.5), $E[D_m | \mathcal{F}_{m-1}] = 1 + mD_{m-1}/(m-1)$, so using the induction hypothesis,

$$
\begin{aligned}
\mathrm{Var}(D_m) &= E[\mathrm{Var}(D_m | \mathcal{F}_{m-1})] + \mathrm{Var}(E[D_m | \mathcal{F}_{m-1}]) \\
&= E[\mathrm{Var}(d_m | \mathcal{F}_{m-1})] + \mathrm{Var}\left(\frac{m}{m-1} D_{m-1}\right) \\
&\leq E[E[d_m^2 | \mathcal{F}_{m-1}]] + m^2 \sum_{k=1}^{m-1} \frac{E[d_k^2]}{k^2} \\
&= m^2 \sum_{k=1}^{m} \frac{E[d_k^2]}{k^2}.
\end{aligned}
$$

Now (2.6) follows by induction.

It remains to bound $E[d_k^2]$. Suppose $k \geq 2$. If $d_k = j \geq 2$, then there is a sequence of numbers $1 = i_0 < i_1 < \cdots < i_j = k$ such that during the construction of the random recursive tree, vertex i_ℓ attaches to vertex $i_{\ell-1}$ for $\ell = 1, \ldots, j$. Because the vertex i_ℓ has a choice of $i_\ell - 1$ vertices to which it can attach, the probability of this event is $1/[(i_1 - 1) \ldots (i_{j-1} - 1)(k-1)]$. Thus,

$$
\begin{aligned}
P(d_k = j) &= \frac{1}{k-1} \sum_{1 < i_1 < \cdots < i_{j-1} < k} \frac{1}{(i_1 - 1) \ldots (i_{j-1} - 1)} \\
&\leq \frac{1}{(k-1)(j-1)!} \left(\sum_{i=2}^{k-1} \frac{1}{i-1}\right)^{j-1} \\
&\leq \frac{(1 + \log k)^{j-1}}{(k-1)(j-1)!}.
\end{aligned}
$$

Combining this bound with the trivial bound that $P(d_k = j) \leq 1$ for $j = 1, 2$ gives

$$
\begin{aligned}
E[d_k^2] &= 5 + \frac{1}{k-1} \sum_{j=3}^{k-1} \frac{j^2 (1 + \log k)^{j-1}}{(j-1)!} \\
&= 5 + \frac{(1 + \log k)^2}{k-1} \sum_{j=3}^{k-1} \frac{j^2}{(j-1)(j-2)} \cdot \frac{(1 + \log k)^{j-3}}{(j-3)!} \\
&\leq 5 + \frac{9(1 + \log k)^2}{2(k-1)} e^{1 + \log k} \\
&\leq C(\log k)^2
\end{aligned}
\tag{2.7}
$$

for some positive constant C. Combining this bound with (2.6) gives (2.4). $\qquad\square$

2.4 Generator of the limit process

From the Poisson process construction described in the introduction, it is clear that $(R(t), t \geq 0)$ is a Markov process. Furthermore, it is easy to describe its transition semigroup. Suppose $R(0) = x$. Then for $y > x - t$, we have $R(t) \leq y$ when there are no points in the Poisson process N above the line segment from $(0, y+t)$ to (t, y). It follows that

$$
P(R(t) \leq y) = \exp\left(- \int_0^t \int_{y+t-s}^\infty e^{-z} \, dz \, ds \right) = \exp\left(- e^{-y}(1 - e^{-t}) \right),
$$

and

$$
P(R(t) = x - t) = \exp\left(- e^{-(x-t)}(1 - e^{-t}) \right).
$$

Let $E = [-\infty, \infty)$, and for $x, y \in E$, let $d(x, y) = |e^x - e^y|$. Then (E, d) is a complete separable metric space. Let $C_0(E)$ be the set of continuous real-valued functions on $[-\infty, \infty)$ that vanish at infinity with the norm $\|f\| = \sup_{x \in E} |f(x)|$. Note that if $f \in C_0(E)$, then $\lim_{x \to -\infty} f(x)$ exists and equals $f(-\infty)$, and $\lim_{x \to \infty} f(x) = 0$. If $f \in C_0(E)$ and $x \in E$, define

$$
P_t f(x) = \int_{x-t}^\infty f(y) e^{-e^{-y}(1-e^{-t})} e^{-y}(1 - e^{-t}) \, dy + f(x-t) e^{-e^{-(x-t)}(1-e^{-t})},
\tag{2.8}
$$

so that $E[f(R(t))] = P_t f(x)$ when $R(0) = x$. Note that the definition of $P_t f(x)$ makes sense when $x = -\infty$, in which case the second term is zero. It is easily checked that $P_t f \in C_0(E)$ and that $P_t f \to f$ as $t \to 0$. Consequently, $(P_t)_{t \geq 0}$ is a Feller semigroup on $C_0(E)$, and $(R(t), t \geq 0)$ is a Feller process with semigroup $(P_t)_{t \geq 0}$.

The following result characterizes the infinitesimal generator of $(R(t), t \geq 0)$ and describes a core for the generator. We see from the form of the generator in (2.9) that for $y > x$, the process jumps from x to y at rate e^{-y}. See also chapter 26 of [10] for a full characterization of the domain of the extended generator.

Lemma 2.4. *Let A be the infinitesimal generator associated with $(P_t)_{t \geq 0}$. Let $\mathcal{C}$ be the collection of functions f that are constant on $[-\infty, z]$ for some $z > -\infty$ and have the property that f, f', and f'' are in $C_0(E)$ when we define f' and f'' in the usual way on $(-\infty, \infty)$ and set $f'(-\infty) = f''(-\infty) = 0$. Then for all $f \in \mathcal{C}$ and all $x \in E$,*

$$
Af(x) = -f'(x) + \int_x^\infty e^{-y}(f(y) - f(x)) \, dy.
\tag{2.9}
$$

Furthermore, $\mathcal{C}$ is a core for A.

Proof. Fix $f \in \mathcal{C}$. Choose a real number z such that f is constant on $[-\infty, z]$. By (2.8), for all $x \in E$,

$$\frac{P_t f(x) - f(x)}{t} = \frac{(f(x-t) - f(x))e^{-e^{-(x-t)}(1-e^{-t})}}{t}$$
$$+ \frac{1}{t} \int_{x-t}^{\infty} (f(y) - f(x))e^{-e^{-y}(1-e^{-t})}e^{-y}(1-e^{-t}) \, dy. \qquad (2.10)$$

We need to show that the right-hand side of (2.10) converges to the right-hand side of (2.9) uniformly in x as $t \to 0$. Because f'' is bounded,

$$\frac{f(x-t) - f(x)}{t} \to -f'(x)$$

uniformly in x as $t \to 0$. Also,

$$\left| \frac{f(x-t) - f(x)}{t} \left(e^{-e^{-(x-t)}(1-e^{-t})} - 1 \right) \right| \leq \left| \frac{f(x-t) - f(x)}{t} \right| e^{-(x-t)}t \leq \|f'\| t e^t e^{-z},$$

which tends to zero uniformly in x as $t \to 0$. Because

$$\left| 1 - \frac{e^{-e^{-y}(1-e^{-t})}(1-e^{-t})}{t} \right| \leq \left| 1 - e^{-e^{-y}(1-e^{-t})} \right| + e^{-e^{-y}(1-e^{-t})} \left| 1 - \frac{1-e^{-t}}{t} \right| \leq e^{-y} t + \frac{t}{2},$$

we have

$$\left| \int_x^{\infty} (f(y) - f(x))e^{-y} \left(1 - \frac{e^{-e^{-y}(1-e^{-t})}(1-e^{-t})}{t} \right) dy \right| \leq \int_z^{\infty} 2\|f\| e^{-y} \left(e^{-y} t + \frac{t}{2} \right) dy,$$

which tends to zero uniformly in x as $t \to 0$. Likewise,

$$\frac{1}{t} \int_{x-t}^{x} (f(y) - f(x))e^{-e^{-y}(1-e^{-t})}e^{-y}(1-e^{-t}) \, dy$$

tends to zero uniformly in x as $t \to 0$. Therefore,

$$\frac{1}{t} \int_{x-t}^{\infty} (f(y) - f(x))e^{-e^{-y}(1-e^{-t})}e^{-y}(1-e^{-t}) \, dy \to \int_x^{\infty} e^{-y}(f(y) - f(x)) \, dy$$

uniformly in x as $t \to 0$. Equation (2.9) follows.

It remains to show that $\mathcal{C}$ is a core for A. It is easy to see that $\mathcal{C}$ is dense in $C_0(E)$. Suppose $f \in \mathcal{C}$, and choose z so that f is constant on $[-\infty, z]$. For all $x \leq z + t$,

$$P_t f(x) = \int_z^{\infty} f(y)e^{-e^{-y}(1-e^{-t})}e^{-y}(1-e^{-t}) \, dy + f(z)e^{-e^{-z}(1-e^{-t})}.$$

Thus, $P_t f$ is constant on $[-\infty, z+t]$. By differentiating the right-hand side of (2.8), we see that the first and second derivatives of $P_t f$ are continuous and vanish at infinity. Thus, $P_t f \in \mathcal{C}$. It follows from Proposition 3.3 in Chapter 1 of [15] that $\mathcal{C}$ is a core for A. $\qquad \square$

2.5 Convergence of finite-dimensional distributions

We show here that the finite-dimensional distributions of the processes $(R_n(t), t \geq 0)$ defined from random recursive trees in section 2.1 converge as $n \to \infty$ to the finite-dimensional distributions of $(R(t), t \geq 0)$.

Let $\ell_1(t) < \cdots < \ell_n(t)$ be the labels of the tree $T_n(t)$ defined in section 2.1. For $i = 2, \ldots, n$, let $H_i(t) = j$ if $\ell_j(t)$ is the parent of $\ell_i(t)$ in the tree $T_n(t)$. Let $\mathcal{G}_n(t)$

be the σ-field generated by the random variables $H_2(t), \ldots, H_n(t)$ and the exponential random variables assigned to the edges adjacent to the root. Informally, this means that $\mathcal{G}_n(t)$ includes all the information about the tree at time t except for the values of the exponential random variables assigned to the edges that are not adjacent to the root. Let $\mathcal{F}_n(t)$ be the smallest σ-field containing $\cup_{s \in [0,t]} \mathcal{G}_n(s)$. Note that $R_n(t)$ is $\mathcal{G}_n(t)$-measurable for all $t \geq 0$, and thus the process $(R_n(t), t \geq 0)$ is adapted to the filtration $\mathcal{F}_n = (\mathcal{F}_n(t), t \geq 0)$. Because $(R_n(t), t \geq 0)$ is right continuous, it follows that $(R_n(t), t \geq 0)$ is $\mathcal{F}_n$-progressive (see p. 50 of [15]).

Fix a function $f \in \mathcal{C}$. Let $\varepsilon_n = n^{-4}$ for all $n \in \mathbb{N}$. Let

$$\xi_n(t) = \frac{1}{\varepsilon_n} \int_0^{\varepsilon_n} E[f(R_n(t+s))|\mathcal{F}_n(t)]\, ds$$

and

$$\varphi_n(t) = \frac{1}{\varepsilon_n} E[f(R_n(t+\varepsilon_n)) - f(R_n(t))|\mathcal{F}_n(t)].$$

By (1.5), $R_n(0)$ converges in distribution to $R(0)$ as $n \to \infty$. By Theorem 8.2 and parts (a) and (b) of Remark 8.3 in chapter 4 of [15], to show that the finite-dimensional distributions of $(R_n(t), t \geq 0)$ converge to those of $(R(t), t \geq 0)$, it suffices to show that the following hold for all $t \geq 0$:

$$\sup_n \sup_{0 \leq s \leq t} E[|\xi_n(s)|] < \infty, \tag{2.11}$$

$$\sup_n \sup_{0 \leq s \leq t} E[|\varphi_n(s)|] < \infty, \tag{2.12}$$

$$\lim_{n \to \infty} E[|\xi_n(t) - f(R_n(t))|] = 0, \tag{2.13}$$

$$\lim_{n \to \infty} E[|\varphi_n(t) - (Af)(R_n(t))|] = 0. \tag{2.14}$$

Note that (2.11) is obvious because $|\xi_n(s)| \leq \|f\|$ for all $s \geq 0$. To show (2.12), choose $z > -\infty$ such that f is constant on $[-\infty, z]$. For $s \geq 0$, let J_s be the event that there exists a time $u \in [s, s+\varepsilon_n]$ such that $R_n(u) \neq R_n(u-)$ and $R_n(u) > z$. Because the process $(R_n(t), t \geq 0)$ decreases at speed one between jumps, we have $R_n(s+\varepsilon_n) - R_n(s) = \varepsilon_n$ if the process does not jump between times s and $s + \varepsilon_n$, which implies that

$$|f(R_n(s+\varepsilon_n)) - f(R_n(s))| \leq \varepsilon_n \|f'\|. \tag{2.15}$$

If the process makes one or more jumps between times s and $s + \varepsilon_n$ but never jumps above z, then $R_n(u) \leq z + \varepsilon_n$ for all $u \in [s, s+\varepsilon_n]$, and once again (2.15) holds because f is constant on $[-\infty, z]$. It follows that (2.15) holds on the event J_s^c. Because the process $(R_n(t), t \geq 0)$ is stationary, we have $P(J_s) = P(J_0)$ for all $s \geq 0$. Therefore, by Lemma 2.2,

$$E[|\varphi_n(s)|] \leq \frac{1}{\varepsilon_n} E[|f(R_n(s+\varepsilon_n)) - f(R_n(s))|\mathbf{1}_{J_s^c}] + \frac{1}{\varepsilon_n} E[|f(R_n(s+\varepsilon_n)) - f(R_n(s))|\mathbf{1}_{J_s}]$$

$$\leq \|f'\| + \frac{2}{\varepsilon_n} \|f\| P(J_s)$$

$$\leq \|f'\| + (2 + 4e^{-z})\|f\|,$$

which proves (2.12).

Next, observe that

$$|\xi_n(t) - f(R_n(t))| = \frac{1}{\varepsilon_n} \left| \int_0^{\varepsilon_n} E[f(R_n(t+s)) - f(R_n(t))|\mathcal{F}_n(t)]\, ds \right|$$

$$\leq \varepsilon_n \|f'\| + 2\|f\| P(J_t|\mathcal{F}_n(t)). \tag{2.16}$$

Thus, by Lemma 2.2, taking expectations of both sides gives

$$E[|\xi_n(t) - f(R_n(t))|] \leq \varepsilon_n \|f'\| + 2\|f\|P(J_t) \leq \varepsilon_n \|f'\| + 2\varepsilon_n \|f\|(1 + 2e^{-z}) \to 0$$

as $n \to \infty$, which gives (2.13). It remains only to show (2.14).

When the tree is cut, we call the event a small cut if fewer than $n/(\log n)^{1/2}$ vertices are removed as a result of the cut, and a large cut otherwise. We define the following five events. Recall that z has been chosen so that f is constant on $[-\infty, z]$.

- Let A_1 be the event that between times t and $t + \varepsilon_n$, there is a small cut at some edge not adjacent to the root, one of the new edges attaches to the root and is assigned a label greater than $\log \log n + z$, and this is the only edge that attaches to the root between times t and $t + \varepsilon_n$ and gets a label greater than $\log \log n + z$.
- Let A_2 be the event that between times t and $t + \varepsilon_n$, there is a large cut during which one of the new edges attaches to the root and is assigned a label greater than $\log \log n + z$.
- Let A_3 be the event that between times t and $t + \varepsilon_n$, there is an event in which the tree is cut at some edge adjacent to the root, and one of the new edges attaches to the root and is assigned a label greater than $\log \log n + z$.
- Let A_4 be the event that between times t and $t + \varepsilon_n$, two or more new edges attach to the root and are assigned labels greater than $\log \log n + z$.
- Let $A_5 = A_1^c \cap A_2^c \cap A_3^c \cap A_4^c$.

The next lemma shows that A_2, A_3, and A_4 are unlikely to occur, which means that jumps of the process R_n between times t and $t + \varepsilon_n$ will occur primarily on the event A_1.

Lemma 2.5. *We have*
$$\lim_{n \to \infty} \frac{P(A_2 \cup A_3 \cup A_4)}{\varepsilon_n} = 0.$$

Proof. To bound $P(A_2)$, note that an event during which $k-1$ vertices are removed from the tree corresponds to a transition in the Bolthausen-Sznitman coalescent in which k blocks merge into one. Such events happen at rate $\binom{n}{k}\lambda_{n,k} = n/[k(k-1)]$ by (1.3). When such an event occurs, the expected number of vertices that reattach to the root is $1/(n-k+1) + \cdots + 1/(n-1)$. When a vertex reattaches to the root, the probability that its label exceeds $\log \log n + z$ is $e^{-(\log \log n + z)} = e^{-z}/\log n$. Thus, for sufficiently large n,

$$P(A_2) \leq \varepsilon_n \sum_{k=\lceil n/(\log n)^{1/2}+1\rceil}^{n} \frac{n}{k(k-1)} \left(\sum_{j=n-k+1}^{n-1} \frac{1}{j} \right) \frac{e^{-z}}{\log n}$$

$$\leq \frac{\varepsilon_n e^{-z}}{\log n} \left(\sum_{k=\lceil n/(\log n)^{1/2}+1\rceil}^{\lfloor n/2\rfloor+1} \frac{n}{k(k-1)} \cdot \frac{2(k-1)}{n} + \sum_{k=\lfloor n/2\rfloor+2}^{n} \frac{n}{k(k-1)} \left(\sum_{j=n-k+1}^{n-1} \frac{1}{j} \right) \right)$$

$$\leq \frac{\varepsilon_n e^{-z}}{\log n} \left(2\left(\log n - \log \left(\frac{n}{(\log n)^{1/2}} \right) \right) + \frac{4}{n} \sum_{j=1}^{n} \sum_{k=n-j+1}^{n} \frac{1}{j} \right)$$

$$= \frac{\varepsilon_n e^{-z}(\log \log n + 4)}{\log n}. \tag{2.17}$$

To bound $P(A_3)$, note that an event in which the tree is cut at some edge adjacent to the root corresponds to a merger in the Bolthausen-Sznitman coalescent that involves the block containing the integer 1. By the sampling consistency of the Bolthausen-Sznitman coalescent, any pair of blocks merges at rate $\lambda_2 = 1$. Consequently, the expected number of blocks that are removed from the tree between times t and $t + \varepsilon_n$

when the tree is cut at an edge adjacent to the root is $\varepsilon_n(n-1)$. For sufficiently large n, provided that fewer than $n/(\log n)^{1/2}$ vertices are removed as a result of the cut, the probability that a given vertex reattaches to the root is at most $2/n$. The probability that the label on the new edge exceeds $\log\log n + z$ is $e^{-z}/\log n$ as before. Thus, for sufficiently large n

$$P(A_3 \cap A_2^c) \leq \varepsilon_n(n-1) \cdot \frac{2}{n} \cdot \frac{e^{-z}}{\log n} \leq \frac{2e^{-z}\varepsilon_n}{\log n}. \tag{2.18}$$

To bound $P(A_4)$, note that there are two ways that A_4 can occur. Either the tree can be cut two or more times between times t and $t+\varepsilon_n$, or two or more edges can reattach to the root after a single cut. Because cuts of the tree happen at times of a Poisson process of rate $\lambda_n = n-1$, the probability that two or more cuts happen between times t and $t+\varepsilon_n$ is at most $\varepsilon_n^2(n-1)^2$. If there is an event in which $k-1$ vertices are removed from the tree following a cut, there are $\binom{k-1}{2}$ pairs of vertices removed, and A_4 will occur if both vertices in one or more of these pairs reattach to the root. On A_2^c, for sufficiently large n, the chance that two given vertices reattach to the root is at most $4/n^2$, and each new edge independently has probability $e^{-z}/(\log n)$ of having a label greater than $\log\log n + z$. Thus,

$$P(A_4 \cap A_2^c) \leq \varepsilon_n^2(n-1)^2 + \varepsilon_n \sum_{k=3}^{\lceil n/(\log n)^{1/2}\rceil} \frac{n}{k(k-1)} \binom{k-1}{2} \frac{4e^{-2z}}{n^2(\log n)^2}$$

$$\leq \varepsilon_n^2(n-1)^2 + \frac{2e^{-2z}\varepsilon_n}{(\log n)^{5/2}}. \tag{2.19}$$

The result follows from (2.17), (2.18), and (2.19). $\qquad\square$

Lemma 2.6. *We have*

$$P(A_1) \leq 2e^{-z}\varepsilon_n$$

for sufficiently large n. Furthermore,

$$\lim_{n\to\infty} E\left[\left|\frac{e^z}{\varepsilon_n}P(A_1|\mathcal{F}_n(t)) - 1\right|\right] = 0.$$

Proof. Recall that in our construction using trees, we labeled the vertices by ordered pairs (t,i) rather than using the vertex labels $1,\ldots,n$. However, for the purposes of this proof, we will arbitrarily number the vertices at time t by the integers $1,\ldots,n$, with the root being vertex 1. For $k \geq 2$, let d_k be the depth of vertex k, which is the number of edges on the path from the root to k. Let v_k be the number of edges along the path from the root to k that are not adjacent to the root but that have at least $n/(\log n)^{1/2}$ vertices below them. That is, v_k is the number of edges e along this path not adjacent to the root such that if we cut the tree at the edge e, it would be classified as a large cut. Define $D_n(t) = d_2 + \cdots + d_n$ and $V_n(t) = v_2 + \cdots + v_n$. For each $k = 2,\ldots,n$, we will separately bound the probability that A_1 occurs and that k is the vertex that reattaches to the root with an edge label of at least $\log\log n + z$.

Note that there are $d_k - 1 - v_k$ edges not adjacent to the root such that, if the tree were cut at that edge, the vertex labeled k would be removed from the tree and this would be a small cut. The probability that one of these edges is cut before time $t+\varepsilon_n$ is $1 - e^{-\varepsilon_n(d_k-1-v_k)}$, which is between $\varepsilon_n(d_k-1-v_k) - \varepsilon_n^2(d_k-1-v_k)^2/2$ and $\varepsilon_n(d_k-1-v_k)$. The probability that the vertex labeled k reattaches to the root is between $1/n$ and $1/(n - n/\sqrt{\log n})$, and the probability that the new edge label is at least $\log\log n + z$ is

$e^{-z}/\log n$. Therefore,

$$
\begin{aligned}
P(A_1|\mathcal{F}_n(t)) &\leq \sum_{k=2}^{n} \varepsilon_n(d_k - 1 - v_k) \cdot \frac{1}{n - n/\sqrt{\log n}} \cdot \frac{e^{-z}}{\log n} \\
&= \frac{\varepsilon_n(D_n(t) - (n-1) - V_n(t))e^{-z}}{n \log n}\left(\frac{\sqrt{\log n}}{\sqrt{\log n} - 1}\right) \\
&\leq \frac{\varepsilon_n D_n(t)e^{-z}}{n \log n}\left(\frac{\sqrt{\log n}}{\sqrt{\log n} - 1}\right).
\end{aligned}
\tag{2.20}
$$

Because $E[D_n(t)] \leq n \log n$ by Lemma 2.3, it follows that $P(A_1) \leq 2e^{-z}\varepsilon_n$ for all n large enough that $\sqrt{\log n}/(\sqrt{\log n} - 1) \leq 2$.

Suppose the vertex k is cut from the tree. The probability that vertex k and some other vertex both reattach to the root with new edge labels greater than $\log\log n + z$ is at most

$$
(n-2)\left(\frac{1}{n - n/\sqrt{\log n}}\right)^2\left(\frac{e^{-z}}{\log n}\right)^2 \leq \frac{2e^{-2z}}{n(\log n)^2}
$$

for sufficiently large n. Also, the probability that there are two cuts to the tree before time $t + \varepsilon_n$ is at most $(n-1)^2\varepsilon_n^2$. Combining these observations, we get

$$
\begin{aligned}
P(A_1&|\mathcal{F}_n(t)) \\
&\geq \sum_{k=2}^{n}\left(\left(\varepsilon_n(d_k - 1 - v_k) - \frac{\varepsilon_n^2(d_k - 1 - v_k)^2}{2}\right)\left(\frac{e^{-z}}{n \log n} - \frac{2e^{-2z}}{n(\log n)^2}\right) - \varepsilon_n^2(n-1)^2\right) \\
&\geq \frac{\varepsilon_n(D_n(t) - (n-1) - V_n(t))e^{-z}}{n \log n} - \frac{\varepsilon_n^2 D_n(t)^2 e^{-z}}{2n \log n} - \frac{2\varepsilon_n D_n(t)e^{-2z}}{n(\log n)^2} - \varepsilon_n^2 n^3.
\end{aligned}
\tag{2.21}
$$

Combining (2.20) and (2.21) gives

$$
\begin{aligned}
\left|\frac{e^z}{\varepsilon_n}P(A_1|\mathcal{F}_n(t)) - 1\right| &\leq \left|\frac{D_n(t)}{n \log n} - 1\right| + \frac{(n-1) + V_n(t)}{n \log n} \\
&\quad + \frac{D_n(t)}{n \log n}\left(\frac{1}{\sqrt{\log n} - 1}\right) + \frac{\varepsilon_n D_n(t)^2}{2n \log n} + \frac{2D_n(t)e^{-z}}{n(\log n)^2} + \varepsilon_n e^z n^3.
\end{aligned}
\tag{2.22}
$$

We need to show that the six terms on the right-hand side of (2.22) tend to zero in expectation as $n \to \infty$. By Lemma 2.3 and the Cauchy-Schwarz Inequality,

$$
\begin{aligned}
E\left[\left|\frac{D_n(t)}{n \log n} - 1\right|\right] &\leq \frac{1}{n \log n}E\left[|D_n(t) - E[D_n(t)]|\right] + \left|\frac{E[D_n(t)]}{n \log n} - 1\right| \\
&\leq \frac{\text{Var}(D_n(t))^{1/2}}{n \log n} + \left|\frac{E[D_n(t)]}{n \log n} - 1\right| \to 0
\end{aligned}
\tag{2.23}
$$

as $n \to \infty$. Because $E[V_n(t)]$ is at most n times the rate of transitions in the Bolthausen-Sznitman coalescent that cause at least $n/(\log n)^{1/2}$ blocks to be lost, we have

$$
E[V_n(t)] \leq n \sum_{k=\lceil n/(\log n)^{1/2}+1\rceil}^{n}\binom{n}{k}\lambda_{n,k} = n\sum_{k=\lceil n/(\log n)^{1/2}+1\rceil}^{n}\frac{n}{k(k-1)} \leq n\sqrt{\log n},
$$

from which it follows that the expected value of the second term on the right-hand side of (2.22) tends to zero as $n \to \infty$. Using Lemma 2.3 and the fact that $\varepsilon_n = n^{-4}$, it is easily checked that the expectations of the last four terms on the right-hand side of (2.22) tend to zero as $n \to \infty$. $\qquad\square$

Proposition 2.7. *The finite-dimensional distributions of the processes $(R_n(t), t \geq 0)$ converge as $n \to \infty$ to the finite-dimensional distributions of $(R(t), t \geq 0)$.*

Proof. Recall that it remains only to show (2.14), which is equivalent to showing that

$$\lim_{n \to \infty} E\left[\left|\frac{1}{\varepsilon_n} E[f(R_n(t + \varepsilon_n)) - f(R_n(t))|\mathcal{F}_n(t)]\right.\right.$$
$$\left.\left. + f'(R_n(t)) - \int_{R_n(t)}^{\infty} e^{-y}(f(y) - f(R_n(t)))\, dy\right|\right] = 0, \qquad (2.24)$$

where f is an arbitrary function in $\mathcal{C}$ which is constant on $[-\infty, z]$. We evaluate the expression $f(R_n(t + \varepsilon_n)) - f(R_n(t))$ on the three disjoint events A_1, $A_2 \cup A_3 \cup A_4$, and A_5.

By Lemma 2.5,

$$E\left[\left|\frac{1}{\varepsilon_n} E[(f(R_n(t+\varepsilon_n)) - f(R_n(t)))\mathbf{1}_{A_2 \cup A_3 \cup A_4}|\mathcal{F}_n(t)]\right|\right] \leq \frac{2\|f\|}{\varepsilon_n} P(A_2 \cup A_3 \cup A_4) \to 0 \quad (2.25)$$

as $n \to \infty$. Note that $f(R_n(t + \varepsilon_n)) = f(R_n(t) - \varepsilon_n)$ on A_5 because either the process $(R_n(t), t \geq 0)$ does not jump between times t and $t + \varepsilon_n$, in which case $R_n(t + \varepsilon_n) = R_n(t) - \varepsilon_n$, or else $R_n(t + \varepsilon_n) \leq z$, in which case $f(R_n(t + \varepsilon_n)) = f(R_n(t) - \varepsilon_n) = f(z)$. Therefore, using Lemmas 2.5 and 2.6,

$$E\left[\left|\frac{1}{\varepsilon_n} E[(f(R_n(t + \varepsilon_n)) - f(R_n(t)))\mathbf{1}_{A_5}|\mathcal{F}_n(t)] + f'(R_n(t))\right|\right]$$

$$= E\left[\left|E\left[\left(\frac{f(R_n(t) - \varepsilon_n) - f(R_n(t))}{\varepsilon_n} + f'(R_n(t))\right)\mathbf{1}_{A_5} + f'(R_n(t))\mathbf{1}_{A_5^c}\middle|\mathcal{F}_n(t)\right]\right|\right]$$

$$\leq E\left[\left|\frac{f(R_n(t) - \varepsilon_n) - f(R_n(t))}{\varepsilon_n} + f'(R_n(t))\right|\right] + \|f'\|P(A_5^c)$$

$$= E\left[\left|\int_{R_n(t) - \varepsilon_n}^{R_n(t)} \frac{f'(s) - f'(R_n(t))}{\varepsilon_n}\, ds\right|\right] + \|f'\|P(A_1 \cup A_2 \cup A_3 \cup A_4)$$

$$\leq \varepsilon_n\|f''\| + \|f'\|(P(A_1) + P(A_2) + P(A_3) + P(A_4)) \to 0 \qquad (2.26)$$

as $n \to \infty$.

On A_1, there is a unique time $\tau \in [t, t + \varepsilon_n]$ such that at time τ, a new edge attaches to the root and is assigned a label greater than $\log \log n + z$. Denote by K the value of this edge label minus $\log \log n$, and let $J = \max\{K, R_n(t)\}$. Conditional on A_1 and $\mathcal{F}_n(t)$, the distribution of K has a density given by $k(y) = e^{z-y}\mathbf{1}_{\{y>z\}}$. Therefore,

$$E[(f(R_n(t + \varepsilon_n)) - f(R_n(t))\mathbf{1}_{A_1}|\mathcal{F}_n(t)]$$

$$= E[(f(R_n(t + \varepsilon_n)) - f(J))\mathbf{1}_{A_1}|\mathcal{F}_n(t)] + E[(f(J) - f(R_n(t))\mathbf{1}_{A_1}|\mathcal{F}_n(t)]$$

$$= E[(f(R_n(t + \varepsilon_n)) - f(J))\mathbf{1}_{A_1}|\mathcal{F}_n(t)] + \left(\int_{R_n(t)}^{\infty} e^{z-y}(f(y) - f(R_n(t))\, dy\right)P(A_1|\mathcal{F}_n(t)).$$

Note that $J - \varepsilon_n \leq R_n(t + \varepsilon_n) \leq J$ on A_1, so

$$|E[f(R_n(t + \varepsilon_n)) - f(J))\mathbf{1}_{A_1}|\mathcal{F}_n(t)]| \leq \varepsilon_n\|f'\|P(A_1|\mathcal{F}_n(t)).$$

It follows that

$$E\left[\left|\frac{1}{\varepsilon_n} E[f(R_n(t + \varepsilon_n)) - f(R_n(t))\mathbf{1}_{A_1}|\mathcal{F}_n(t)] - \int_{R_n(t)}^{\infty} e^{-y}(f(y) - f(R_n(t))\, dy\right|\right]$$

$$\leq \|f'\|P(A_1) + E\left[\left|\left(\int_{R_n(t)}^{\infty} e^{-y}(f(y) - f(R_n(t))\, dy\right)\left(\frac{e^z P(A_1|\mathcal{F}_n(t))}{\varepsilon_n} - 1\right)\right|\right]$$

$$\leq \|f'\|P(A_1) + 2\|f\|e^{-z}E\left[\left|\frac{e^z}{\varepsilon_n} P(A_1|\mathcal{F}_n(t)) - 1\right|\right]. \qquad (2.27)$$

It now follows from Lemma 2.6 that the right-hand side of (2.27) tends to zero as $n \to \infty$. The result now follows by combining this observation with (2.25) and (2.26). $\qquad\square$

2.6 Tightness

Here we show that the sequence of processes $(R_n)_{n=1}^\infty$ is relatively compact. By Theorem 7.8 in Chapter 3 of [15], this result in combination with Proposition 2.7 implies that the processes $(R_n(t), t \geq 0)$ converge weakly to $(R(t), t \geq 0)$.

For $\delta > 0$ and $t > 0$, let

$$w_n(\delta, t) = \inf_{\{t_i\}} \max_i \sup_{r, s \in [t_{i-1}, t_i)} |R_n(r) - R_n(s)|,$$

where the infimum is taken over all partitions of the form $0 = t_0 < t_1 < \cdots < t_m = t$ such that $t_i - t_{i-1} > \delta$ for $i = 1, \ldots, m$. By Corollary 7.4 in Chapter 3 of [15], the sequence $(R_n)_{n=1}^\infty$ is relatively compact provided that for all $\varepsilon > 0$ and $t > 0$, there exists $\delta > 0$ such that

$$\limsup_{n \to \infty} P(w_n(\delta, t) \geq \varepsilon) \leq \varepsilon. \tag{2.28}$$

Therefore, we can conclude the proof of Theorem 1.2 by showing (2.28).

Fix $t > 0$. Choose $\varepsilon \in (0, 1)$, and choose $y < 0$ such that $P(R_n(0) \leq y + t) < \varepsilon/6$. Note that if $R_n(0) > y + t$, then $R_n(s) > y$ for all $s \in [0, t]$. Let

$$\delta = \frac{e^{2y} \varepsilon}{36 \max\{t, 1\}}.$$

Let $0 < \tau_{1,n} < \tau_{2,n} < \ldots$ denote the jump times of $(R_n(s), s \geq 0)$. As long as $\tau_{j,n} - \tau_{j-1,n} > \delta$ for all j such that $\tau_{j,n} \leq t$ and there are no jump times in $[0, \delta]$ or $[t - \delta, t]$, it is easy to choose the times $t_0, \ldots, t_n$ such that $\delta < t_i - t_{i-1} < 2\delta$ for $i = 1, \ldots, n$ and for all j such that $\tau_{j,n} \leq t$, we have $\tau_{j,n} = t_i$ for some i. That is, there is one of the t_i at every jump time of the process. In this case, whenever $r, s \in [t_{i-1}, t_i)$, we have $|R_n(r) - R_n(s)| = |r - s| \leq 2\delta < \varepsilon$.

By Lemmas 2.5 and 2.6 with δ in place of ε_n and y in place of z, we have

$$\limsup_{n \to \infty} P\big(R_n(s) \neq R_n(s-) \text{ for some } s \in [0, \delta] \cup [t - \delta, t]\big)$$

$$\leq \limsup_{n \to \infty} P(R_n(0) \leq y + t) + 2e^{-y}\delta + 2e^{-y}\delta < \frac{\varepsilon}{3}. \tag{2.29}$$

Note that if $\tau_{j,n} - \tau_{j-1,n} \leq \delta$ for some j such that $\tau_{j,n} \leq t$, then there exists a nonnegative integer $k \leq t/\delta - 1$ such that $k\delta \leq \tau_{j-1,n} < \tau_{j,n} \leq \min\{t, (k+2)\delta\}$. For two jumps of the process $(R_n(s), s \geq 0)$ to fall within the interval $[k\delta, \min\{t, (k+2)\delta\}]$, one of the following four events must occur:

- We have $R_n(s) \leq y$ for some $s \in [0, t]$.
- Between times $k\delta$ and $(k+2)\delta$, there is a large cut, and one of the new edges attaches to the root and is assigned a label greater than $\log \log n + y$.
- Between times $k\delta$ and $(k+2)\delta$, more than $3\delta n \log n$ vertices are removed from the tree during small cuts.
- Of the first $\lfloor 3\delta n \log n \rfloor$ vertices, after time $k\delta$, that are removed from the tree during small cuts, two or more reattach to the root with new edge labels greater than $\log \log n + y$.

We have already bounded the probability of the first event by $\varepsilon/6$. The other three events depend on k. The probability of the second event tends to zero as $n \to \infty$ by

(2.17) with 2δ in place of ε_n and y in place of z. To bound the probability of the third event, note that mergers in the Bolthausen-Sznitman coalescent in which $k-1$ blocks are lost occur at rate $\binom{n}{k}\lambda_{n,k} = n/(k(k-1))$. Therefore, if N denotes the number of vertices removed during small cuts between times $k\delta$ and $(k+2)\delta$, we have

$$E[N] = 2\delta \sum_{k=2}^{\lceil n/(\log n)^{1/2}\rceil} (k-1) \cdot \frac{n}{k(k-1)} = 2\delta \sum_{k=2}^{\lceil n/(\log n)^{1/2}\rceil} \frac{n}{k} \leq 2\delta n \log n$$

and

$$\mathrm{Var}(N) = 2\delta \sum_{k=2}^{\lceil n/(\log n)^{1/2}\rceil} (k-1)^2 \cdot \frac{n}{k(k-1)} \leq 2\delta n^2.$$

Therefore, by Chebyshev's Inequality,

$$P(N > 3\delta n \log n) \leq P(|N - E[N]| > \delta n \log n) \leq \frac{2\delta n^2}{\delta^2 n^2 (\log n)^2} \to 0$$

as $n \to \infty$. Finally, concerning the fourth event, note that when a vertex is reattached after being removed during a small cut, the probability that it reattaches to the root is at most $2/n$, and the probability that it is assigned a label greater than $\log\log n + y$ is $e^{-y}/(\log n)$. Therefore, since there are at most $(3\delta n \log n)^2/2$ pairs of vertices to consider, the probability of the fourth event for a particular k is at most

$$\frac{(3\delta n \log n)^2}{2} \cdot \frac{4e^{-2y}}{(n\log n)^2} \leq 18\delta^2 e^{-2y}.$$

Since there are at most t/δ possible values of k to consider, the probability that the fourth event occurs for some k is at most $18t\delta e^{-2y} \leq \varepsilon/2$. Therefore,

$$\limsup_{n\to\infty} P(\tau_{j,n} - \tau_{j-1,n} \leq \delta \text{ for some } j \text{ such that } \tau_{j,n} \leq t) \leq \frac{\varepsilon}{6} + \frac{\varepsilon}{2} \leq \frac{2\varepsilon}{3}.$$

Combining this result with (2.29) gives (2.28) and completes the proof of Theorem 1.2.

3 Proof of Theorem 1.7

We obtain Theorems 1.5 and 1.7 using a very different approach. Rather than using recursive trees, we couple the population model with a family of stable processes by constructing both from a Poisson process. We describe this construction in section 3.1. We then prove Theorem 1.7 in section 3.4, and we prove Theorem 1.5 in section 4.

Throughout the rest of the paper, $T > 0$ will be an arbitrary positive constant, and

$$T_n = 2\log\log n.$$

Also, $\to_p$ will denote convergence in probability as $n \to \infty$.

3.1 A Poisson process construction

Fix a positive integer n. Let Ψ be a Poisson point process on $\mathbb{R} \times (0, \infty)$ whose intensity measure is given by $dt \times y^{-2} \, dy$. Then define Θ to be the image of Ψ under the map $(t, y) \mapsto (-t/\log n, y/\log n)$, restricted to $\mathbb{R} \times (0, 1]$. That is, if (t, y) is a point of Ψ with $y \leq \log n$, then $(-t/\log n, y/\log n)$ is a point of Ψ. We claim that Θ is a Poisson point process on $\mathbb{R} \times (0, 1]$ with intensity measure $dt \times y^{-2} \, dy$. To see this, observe that if $a < b$

and $0 < x < 1$, then the expected number of points of Θ in the region $[a, b] \times [x, 1]$ is the expected number of points of Ψ in the region $[-b \log n, -a \log n] \times [x \log n, \log n]$, which is

$$
\int_{-b \log n}^{-a \log n} \int_{x \log n}^{\log n} y^{-2} \, dy \, dt = (b - a) \log n \left(\frac{1}{x \log n} - \frac{1}{\log n} \right)
$$
$$
= (b - a) \left(\frac{1}{x} - 1 \right) = \int_a^b \int_x^1 y^{-2} \, dy \, dt.
$$

We now construct a population model consisting of n individuals labelled $1, \ldots, n$. We independently attach to each point (t_i, y_i) of Θ random variables $U_{i,1}, \ldots, U_{i,n}$, which are independent and have the uniform distribution on $(0, 1)$. If zero or one of the random variables $U_{i,1}, \ldots, U_{i,n}$ is less than y_i, then there is no change in the population at time t_i. However, if $k \geq 2$ of the random variables $U_{i,1}, \ldots, U_{i,n}$ are less than y_i and the k smallest of these random variables are $U_{i,j_1} < \cdots < U_{i,j_k}$, then at time t_i, the individuals labeled $j_2, \ldots, j_k$ are killed, and the individual labeled j_1 gives birth to $k - 1$ new offspring, which assume the labels $j_2, \ldots, j_k$.

To see that this is equivalent to the population model described in the introduction, note that if (t_i, y_i) is a point of Θ, the probability that exactly k new offspring are born at time t_i is $\binom{n}{k+1} y_i^{k+1} (1 - y_i)^{n-k-1}$. Thus, the rate of events in which exactly k new offspring are born is

$$
\binom{n}{k+1} \int_0^1 y^{k+1} (1 - y)^{n-k-1} \cdot y^{-2} \, dy = \frac{n}{k(k+1)},
$$

which matches (1.2) because changes in the population occur at rate $n - 1$.

For $s \in \mathbb{R}$ and $t \geq 0$, let $N_n(s, t)$ denote the number of individuals in the population at time $(s - t)-$ who have a descendant alive in the population at time s. That is, $N_n(s, t)$ is the number of ancestral lines remaining after time t if we trace back the ancestral lines of the individuals in the population at time s. Note that because we consider the population at time $(s - t)-$ rather than at time $s - t$ when defining $N_n(s, t)$, the process $(N_n(s, t), t \geq 0)$ is right continuous. Also, note that $N(s, 0) = n$ as long as there is no change in the population at time s, but if k individuals are killed at time s and replaced by new offspring, then $N(s, 0) = n - k$. Let $N_n(t) = N_n(0, t)$. Because the genealogy of this population is given by the Bolthausen-Sznitman coalescent started with n blocks, the process $(N_n(t), t \geq 0)$ has the same law as the process defined in the statement of Theorem 1.7. Let

$$
L_n(s) = \int_0^\infty N_n(s, t) \mathbf{1}_{\{N_n(s,t) > 1\}} \, dt. \tag{3.1}
$$

Then $L_n(s)$ is the total branch length for the coalescent tree representing the genealogy of the population at time s, so $(L_n(s), s \geq 0)$ has the same law as the process considered in Theorem 1.5.

Next, we use the Poisson process Ψ to construct, for each $s \in \mathbb{R}$, a stable process $(S(s, t), t \geq 0)$ with characteristic exponent

$$
E[e^{iuS(s,t)}] = \exp\left(-\frac{\pi t}{2} |u| + itu \log |u| \right)
$$
$$
= \exp\left(itu(1 - \gamma) - t \int_{-\infty}^0 \left(1 - e^{iux} + iux \mathbf{1}_{\{|x| \leq 1\}} \right) x^{-2} \, dx \right), \tag{3.2}
$$

where γ denotes Euler's constant. Because Ψ has only countably many points, we can enumerate the points of Ψ as $(s_j, x_j)_{j=1}^\infty$. If $s_j > -s$, then the process $(S(s, t), t \geq 0)$ will have a jump of size $-x_j$ at time $s_j + s$. Also, there is a linear drift. To make this

construction precise, we use a standard approximation procedure that is described, for example, in Section I.1 of [4]. Define

$$\varepsilon_n = e^{-\sqrt{\log n}},$$

which implies that for all $a > 0$,

$$\lim_{n \to \infty} (\log n)^a \varepsilon_n = 0$$

and

$$\lim_{n \to \infty} n^a \varepsilon_n = \infty.$$

Let $R_n(s,t)$ be the set of all j such that $-s < s_j \leq -s + t$ and $x_j > \varepsilon_n$. For all $s \in \mathbb{R}$ and $t \geq 0$, let

$$S_n(s,t) = t(1 - \gamma - \log \varepsilon_n) - \sum_{j \in R_n(s,t)} x_j. \tag{3.3}$$

Note that $\int_{-\infty}^0 x \mathbf{1}_{\{\varepsilon_n < |x| \leq 1\}} \, x^{-2} \, dx = \log \varepsilon_n$. Therefore, for any fixed $T > 0$ and any integers $m, n > N$, the proof of Theorem 1 in Section I.1 of [4] (see the bottom of p. 14) gives

$$E\left[\sup_{0 \leq t \leq T} (S_n(s,t) - S_m(s,t))^2 \right] \leq 4T \int_0^\infty x^2 \mathbf{1}_{\{|x| \leq \varepsilon_N\}} \, x^{-2} \, dx \leq 4T\varepsilon_N. \tag{3.4}$$

It follows that for each $s \in \mathbb{R}$, there is a limit process $(S(s,t), t \geq 0)$ satisfying (3.2) such that for each fixed $T > 0$,

$$E\left[\sup_{0 \leq t \leq T} (S_n(s,t) - S(s,t))^2 \right] \leq 4T\varepsilon_n \to 0 \tag{3.5}$$

as $n \to \infty$. Let $S_n(t) = S_n(0,t)$ and $S(t) = S(0,t)$ for all $t \geq 0$.

3.2 Bounds on stable processes

The three lemmas below collect some bounds on these stable processes that will be needed later.

Lemma 3.1. *Fix $\varepsilon > 0$. Then there exists $K > 0$ such that*

$$P\left(\sup_{0 \leq s \leq T} \sup_{t \geq 0} e^{-t} |S(s,t)| > K \right) < \varepsilon. \tag{3.6}$$

Additionally, we have

$$\lim_{n \to \infty} P\left(\sup_{0 \leq s \leq T} \sup_{0 \leq t \leq T_n + s} |S(s,t)| > 5(\log \log n)^2 \right) = 0. \tag{3.7}$$

Also,

$$\lim_{n \to \infty} P(|S(T_n)| \leq T_n^2) = 1 \tag{3.8}$$

and

$$\sup_{0 \leq s \leq T} \int_{T_n + s}^\infty e^{-t} S(s,t) \, dt \to_p 0. \tag{3.9}$$

Proof. The process $(S(t), t \geq 0)$ is a stable process of index 1. Therefore, by Proposition 48.10 of [30], we have

$$\limsup_{t \to \infty} t^{-\alpha} |S(t)| = 0$$

almost surely for all $\alpha > 1$. Therefore, there exists a random time U with $P(U < \infty) = 1$ such that $|S(t)| \leq t^2$ for all $t \geq U$. This implies (3.8).

Let $L = \sup_{0 \leq t \leq U} |S(t)|$ and $M = \sup_{0 \leq t \leq T} |S(T, t)|$. Note that $L < \infty$ and $M < \infty$ almost surely. Suppose $0 \leq s \leq T$. By the construction, if $0 \leq t \leq s$ then

$$S(s, t) = S(T, T - s + t) - S(T, T - s), \tag{3.10}$$

while if $t \geq s$ then

$$S(s, t) = S(s, s) + S(0, t - s) = S(T, T) - S(T, T - s) + S(0, t - s). \tag{3.11}$$

Therefore,

$$|S(s, t)| \leq 2M + L + t^2. \tag{3.12}$$

Since $e^{-t} t^2 \leq 1$ for all $t \geq 0$, it follows that

$$\sup_{0 \leq s \leq T} \sup_{t \geq 0} e^{-t} |S(s, t)| \leq 2M + L + 1.$$

Now choose K large enough that $P(2M + L + 1 > K) < \varepsilon$ to obtain (3.6). Furthermore, (3.12) implies that

$$\sup_{0 \leq s \leq T} \sup_{0 \leq t \leq T_n + s} |S(s, t)| \leq 2M + L + (2 \log \log n + T)^2,$$

and (3.7) follows because $P(2M + L + (2 \log \log n + T)^2 > 5(\log \log n)^2) \to 0$ as $n \to \infty$. On the event that $2M + L \leq T_n^2$, which has probability tending to one as $n \to \infty$, equation (3.12) gives that for $0 \leq s \leq T$,

$$\int_{T_n + s}^{\infty} e^{-t} |S(s, t)| \, dt \leq 2 \int_{T_n}^{\infty} e^{-t} t^2 \, dt \to 0$$

as $n \to \infty$, which gives (3.9). $\qquad\square$

Lemma 3.2. *We have*

$$\sup_{0 \leq s \leq T} \sup_{0 \leq t \leq T_n + s} |S_n(s, t) - S(s, t)| \to_p 0.$$

Proof. Suppose $0 \leq s \leq T$. By (3.10), which holds also with S_n in place of S, if $0 \leq t \leq s$ then

$$|S_n(s, t) - S(s, t)| = |S_n(T, T - s + t) - S_n(T, T - s) - S(T, T - s + t) + S(T, T - s)|$$
$$\leq 2 \sup_{0 \leq u \leq T} |S_n(T, u) - S(T, u)|.$$

By (3.11), if $s \leq t \leq T_n + s$, then

$$|S_n(s, t) - S(s, t)|$$
$$= |S_n(T, T) - S_n(T, T - s) + S_n(0, t - s) - S(T, T) + S(T, T - s) - S(0, t - s)|$$
$$\leq 2 \sup_{0 \leq u \leq T} |S_n(T, u) - S(T, u)| + \sup_{0 \leq u \leq T_n} |S_n(0, u) - S(0, u)|.$$

Using that $(a + b)^2 \leq 2a^2 + 2b^2$ and applying (3.5), we get

$$E\left[\left(\sup_{0 \leq s \leq T} \sup_{0 \leq t \leq T_n + s} |S_n(s, t) - S(s, t)|\right)^2\right] \leq 32T\varepsilon_n + 8T_n \varepsilon_n \to 0$$

as $n \to \infty$, which implies the result. $\qquad\square$

Lemma 3.3. *Suppose $s \in \mathbb{R}$ and $(\delta_n)_{n=1}^{\infty}$ is a sequence of numbers with $\delta_n \sqrt{\log n} \to 0$ as $n \to \infty$. Let $\theta > 0$, and let $A(\theta)$ be the event that there are no points of Ψ in $[-s, -s + \delta_n] \times [\theta, \infty)$. Then for sufficiently large n,*

$$P\left(\sup_{0 \le t \le \delta_n} |S_n(s,t)| \ge \theta^{1/3} \Big| A(\theta) \right) \le 4\delta_n \theta^{1/3}.$$

Proof. By (3.3), we have

$$\sup_{0 \le t \le \delta_n} |S_n(s,t)| \le \delta_n(1 - \gamma - \log \varepsilon_n) + \sum_{j \in R_n(s,t)} x_j.$$

Because $|\log \varepsilon_n| = \sqrt{\log n}$ and $\delta_n \sqrt{\log n} \to 0$, we have $\delta_n(1 - \gamma - \log \varepsilon_n) \le \theta^{1/3}/4$ for sufficiently large n. Furthermore,

$$E\left[\sum_{j \in R_n(s,t)} x_j \Big| A(\theta) \right] = \delta_n \int_{\varepsilon_n}^{\theta} x \cdot x^{-2} \, dx = \delta_n(\log \theta - \log \varepsilon_n) \to 0$$

as $n \to \infty$ and

$$\mathrm{Var}\left(\sum_{j \in R_n(s,t)} x_j \Big| A(\theta) \right) = \delta_n \int_{\varepsilon_n}^{\theta} x^2 \cdot x^{-2} \, dx \le \delta_n \theta.$$

Therefore, by Chebyshev's Inequality, for sufficiently large n,

$$P\left(\sup_{0 \le t \le \delta_n} |S_n(s,t)| \ge \theta^{1/3} \Big| A(\theta) \right) \le P\left(\sum_{j \in R_n(s,t)} x_j \ge \frac{3\theta^{1/3}}{4} \Big| A(\theta) \right)$$

$$\le P\left(\Big| \sum_{j \in R_n(s,t)} x_j - E\Big[\sum_{j \in R_n(s,t)} x_j \Big| A(\theta) \Big] \Big| \ge \frac{\theta^{1/3}}{2} \Big| A(\theta) \right)$$

$$\le \delta_n \theta \left(\frac{4}{\theta^{2/3}} \right), \tag{3.13}$$

which gives the result. $\qquad\square$

3.3 Rate of decrease in the number of blocks

We record here some results about the rate at which the number of blocks decreases in the Bolthausen-Sznitman coalescent. Recall that the process $(N_n(t), t \ge 0)$ tracks the evolution of the number of blocks over time. Because the number of blocks decreases by $k - 1$ whenever k blocks merge into one, the rate at which the number of blocks is decreasing when there are b blocks is

$$\eta(b) = \sum_{k=2}^{b} (k-1)\binom{b}{k} \lambda_{b,k}. \tag{3.14}$$

Considering the process from the perspective of the construction in section 3.1, suppose there is a point $(-t, y)$ in the Poisson process Θ. If $N_n(t-) = b$, then $N_n(t-) - N_n(t)$ is one less than the number out of b independent uniformly distributed random variables that are less than or equal to y, unless all of the random variables are greater than y in which case $N_n(t-) - N_n(t) = 0$. Therefore, the expected decrease in the process N_n at time t is $by - 1 + (1-y)^b$. It follows that

$$\eta(b) = \int_0^1 \big(by - 1 + (1-y)^b \big) y^{-2} \, dy.$$

We will be interested also in the process obtained by removing from Θ the points whose second coordinate exceeds $\varepsilon_n/\log n$. In this case, the genealogy of the population is given not by the Bolthausen-Sznitman coalescent but by a different coalescent process in which the largest merger events are suppressed. In particular, when there are b blocks, the rate at which k blocks merge into one is given by

$$\lambda^*_{b,k} = \int_0^{\varepsilon_n/\log n} x^{k-2}(1-x)^{b-k}\,dx \tag{3.15}$$

and the rate at which the number of blocks is decreasing is given by

$$\eta^*(b) = \sum_{k=2}^b (k-1)\binom{b}{k}\lambda^*_{b,k} = \int_0^{\varepsilon_n/\log n}\big(by-1+(1-y)^b\big)y^{-2}\,dy.$$

Also, let

$$v^*(b) = \sum_{k=2}^b (k-1)^2\binom{b}{k}\lambda^*_{b,k}. \tag{3.16}$$

Lemma 3.4. *For all positive integers b and n with $2 \le b \le n$, we have*

$$\big|\eta^*(b) - b(\log b - \log\log n + \log\varepsilon_n + \gamma - 1)\big| \le \frac{\log n}{\varepsilon_n} + 1, \tag{3.17}$$

where γ denotes Euler's constant. Also,

$$v^*(b) \le \frac{b^2\varepsilon_n}{\log n}. \tag{3.18}$$

Proof. Using (3.14) and (1.3), we get

$$\eta(b) = b\sum_{k=2}^b \frac{1}{k}. \tag{3.19}$$

Now

$$\lim_{b\to\infty}\left(\sum_{k=2}^b \frac{1}{k} - \log b\right) = \gamma - 1$$

and

$$0 \le \big(\log b - \log(b-1)\big) - \frac{1}{b} = \int_{b-1}^b \left(\frac{1}{x} - \frac{1}{b}\right)dx \le \frac{1}{b-1} - \frac{1}{b}.$$

Therefore,

$$\left|\left(\sum_{k=2}^b \frac{1}{k} - \log b\right) - (\gamma-1)\right| \le \sum_{k=b+1}^\infty \left(\frac{1}{k-1} - \frac{1}{k}\right) = \frac{1}{b}. \tag{3.20}$$

By (3.19) and (3.20),

$$\big|\eta(b) - b(\log b + \gamma - 1)\big| \le b\left|\sum_{k=2}^b \frac{1}{k} - (\log b + \gamma - 1)\right| \le 1 \tag{3.21}$$

for all b. It follows that

$$\begin{aligned}
\eta^*(b) &= \int_0^{\varepsilon_n/\log n}\big(by-1+(1-y)^b\big)y^{-2}\,dy.\\
&= \int_0^1 \big(by-1+(1-y)^b\big)y^{-2}\,dy - \int_{\varepsilon_n/\log n}^1 \big(by-1+(1-y)^b\big)y^{-2}\,dy\\
&= \eta(b) + b\log\left(\frac{\varepsilon_n}{\log n}\right) + \int_{\varepsilon_n/\log n}^1 \big(1-(1-y)^b\big)y^{-2}\,dy.
\end{aligned} \tag{3.22}$$

Because the last integral is bounded by $\int_{\varepsilon_n/\log n}^{\infty} y^{-2}\,dy = (\log n)/\varepsilon_n$, the result (3.17) follows by combining (3.21) and (3.22).

To prove (3.18), note that if $(-t, y)$ is a point of Θ and $N_n(t-) = b$, then because $(k-1)^2 \leq 2\binom{k}{2}$ for $k \geq 2$, the expected square of the decrease in the process N_n at time t is at most twice the expected number of pairs out of b independent uniformly distributed random variables having the property that both random variables are less than or equal to y, which is $2\binom{b}{2}y^2$. Therefore,

$$v^*(b) \leq \int_0^{\varepsilon_n/\log n} 2\binom{b}{2} y^2 \cdot y^{-2}\,dy \leq \frac{b^2 \varepsilon_n}{\log n},$$

as claimed. $\qquad\square$

3.4 The coupling

In this section, we prove Theorem 1.7. We use the construction and notation of section 3.1. As in the statement of Theorem 1.7, let

$$X_n(t) = \frac{\log n}{n}\left(N_n\left(\frac{t}{\log n}\right) - ne^{-t} - \frac{nte^{-t}\log\log n}{\log n}\right) \tag{3.23}$$

for all $t \geq 0$. Also, let

$$Y_n(t) = e^{-t} S_n(t) + \frac{e^{-t} t^2}{2}$$

and

$$Y(t) = e^{-t} S(t) + \frac{e^{-t} t^2}{2}.$$

We need to show that the processes $(X_n(t), t \geq 0)$ converge weakly to $(Y(t), t \geq 0)$. Because

$$E\left[\sup_{0 \leq t \leq T_n} (Y_n(t) - Y(t))^2\right] \leq 4T_n \varepsilon_n \to 0$$

as $n \to \infty$ by (3.5), it suffices to show that

$$\sup_{0 \leq t \leq T_n} |X_n(t) - Y_n(t)| \to_p 0. \tag{3.24}$$

In fact, to prove Theorem 1.7, it would suffice to show (3.24) with the arbitrary fixed constant T in place of T_n, but it will be helpful for the proof of Theorem 1.5 to control the difference between X_n and Y_n up to time T_n.

Let $0 < \tau_1 < \cdots < \tau_{J_n} < T_n$ be the jump times of the process $(S_n(t), t \geq 0)$ before time T_n. Let $\tau_0 = 0$ and $\tau_{J_n+1} = T_n$. Note that the τ_i depend on n even though we do not record this dependence in the notation. This means that there are points (τ_i, y_i) in Ψ with $y_i \geq \varepsilon_n$ for $i = 1, \ldots, J_n$. Also, the process Θ, which is used to construct the population process, contains the points $(-\tau_i/\log n, y_i/\log n)$ but Θ contains no points in the regions $(-\tau_{i+1}/\log n, -\tau_i/\log n) \times [\varepsilon_n/\log n, 1]$. Therefore, conditional on $\tau_1, \ldots, \tau_{J_n}$, between times $\tau_i/\log n$ and $\tau_{i+1}/\log n$, the process N_n follows the dynamics of the number of blocks in a coalescent process with transition rates given by (3.15).

For $i = 0, 1, \ldots, J_n$ and $t \in [0, \tau_{i+1} - \tau_i)$, let

$$M_{i,n}(t) = N_n\left(\frac{\tau_i + t}{\log n}\right) - N_n\left(\frac{\tau_i}{\log n}\right) + \int_{\tau_i/\log n}^{(\tau_i+t)/\log n} \eta^*(N_n(s))\,ds. \tag{3.25}$$

By standard results about compensators for Markov jump processes (see, for example, Theorem 9.15 in [21]), the process $(M_{i,n}(t), 0 \leq t < \tau_{i+1} - \tau_i)$ is a martingale. Note

also that $\tau_{i+1} - \tau_i$ is exponentially distributed with mean ε_n and is independent of the evolution of the process $M_{i,n}$ before time $\tau_{i+1} - \tau_i$. Next, for $\tau_j \leq t < \tau_{j+1}$, let

$$M_n(t) = \sum_{i=0}^{j-1} M_{i,n}((\tau_{i+1} - \tau_i)-) + M_{j,n}(t - \tau_j). \tag{3.26}$$

Then the process $(M_n(t), 0 \leq t < T_n)$ is a martingale.

Lemma 3.5. *We have*

$$P\left(\frac{\log n}{n} \sup_{0 \leq t \leq T_n} |M_n(t)| > \frac{1}{\log n}\right) \leq 4(\log n)^2 T_n \varepsilon_n.$$

Proof. By standard results about compensated Markov jump processes (see, for example, Corollary 9.17 of [21]),

$$\text{Var}(M_n(T_n)) = E\left[\int_0^{T_n / \log n} v^*(N_n(t)) \, dt\right],$$

where v^* was defined in (3.16). Therefore, by (3.18),

$$\text{Var}(M_n(T_n)) \leq \frac{n^2 T_n \varepsilon_n}{(\log n)^2}.$$

By the L^2 Maximum Inequality for martingales,

$$E\left[\sup_{0 \leq t \leq T_n} M_n(t)^2\right] \leq \frac{4n^2 T_n \varepsilon_n}{(\log n)^2}.$$

Thus, by Markov's Inequality,

$$P\left(\frac{\log n}{n} \sup_{0 \leq t \leq T_n} |M_n(t)| > \frac{1}{\log n}\right) = P\left(\sup_{0 \leq t \leq T_n} M_n(t)^2 > \frac{n^2}{(\log n)^4}\right) \leq 4(\log n)^2 T_n \varepsilon_n,$$

as claimed. $\qquad\square$

For $0 \leq t \leq T_n$, let

$$Z_n(t) = X_n(t) - \frac{\log n}{n} M_n(t) = \frac{\log n}{n}\left(N_n\left(\frac{t}{\log n}\right) - M_n(t) - ne^{-t} - \frac{nte^{-t} \log\log n}{\log n}\right). \tag{3.27}$$

Lemma 3.5 implies that

$$\sup_{0 \leq t \leq T_n} \frac{\log n}{n} |M_n(t)| \to_p 0$$

as $n \to \infty$. Therefore, to show (3.24), it suffices to compare the processes Z_n and Y_n.

For $s \geq 0$, let $(\tau_1^s, y_1^s), (\tau_2^s, y_2^s), \ldots, (\tau_{J_n^s}^s, y_n^s)$ denote the points of Ψ in $[-s, T_n] \times [\varepsilon_n, \infty)$, ranked so that $\tau_1^s < \cdots < \tau_{J_n^s}^s$. Let $\tau_0^s = -s$ and $\tau_{J_n^s + 1}^s = T_n$. Note that $\tau_i^0 = \tau_i$ for $i = 0, 1, \ldots, J_n + 1$. Also let $y_i = y_i^0$ for $i = 1, \ldots, J_n$. In the next Lemma, we show that several events hold with high probability. This result will allow us to assume that these events holds throughout much of the rest of the paper.

Lemma 3.6. *Let $A_{1,n}$ be the event that $J_n^T \leq n^{1/4}$. Let $A_{2,n}$ be the event that*

$$\sum_{i=1}^{J_n^T} y_i^T \leq (\log n)^{3/4}.$$

Let $A_{3,n}$ be the event that $\tau_{i+1}^T - \tau_i^T \leq \varepsilon_n \log n$ for $i = 0, 1, \ldots, J_n^T$. Let $A_n = A_{1,n} \cap A_{2,n} \cap A_{3,n}$. Then $\lim_{n \to \infty} P(A_n) = 1$.

Proof. Note that
$$E[J_n^T] = (T_n + T) \int_{\varepsilon_n}^{\infty} y^{-2} \, dy = \frac{(T_n + T)}{\varepsilon_n}.$$

It follows from Markov's Inequality that
$$\limsup_{n \to \infty} P(A_{1,n}^c) \leq \limsup_{n \to \infty} \frac{(T_n + T)}{n^{1/4} \varepsilon_n} = 0,$$

and thus $\lim_{n \to \infty} P(A_{1,n}) = 1$.

We have
$$P\left(\max_{1 \leq i \leq J_n^T} y_i > \log n \right) \leq (T_n + T) \int_{\log n}^{\infty} y^{-2} \, dy = \frac{T_n + T}{\log n} \to 0$$

as $n \to \infty$. Also,
$$E\left[\sum_{i=1}^{J_n^T} y_i^T \,\bigg|\, \max_{1 \leq i \leq J_n^T} y_i^T \leq \log n \right] = (T_n + T) \int_{\varepsilon_n}^{\log n} y \cdot y^{-2} \, dy = (T_n + T)(\log \log n - \log \varepsilon_n).$$

Because $-\log \varepsilon_n = (\log n)^{1/2}$ and thus $(T_n + T)(\log \log n - \log \varepsilon_n)/(\log n)^{3/4} \to 0$ as $n \to \infty$, it now follows from Markov's Inequality that $\lim_{n \to \infty} P(A_{2,n}) = 1$.

To estimate $P(A_{3,n})$, let $a_k = -T + (k \varepsilon_n \log n)/2$ for $k = 0, 1, \ldots$, and let $K = \min\{k : a_k > T_n\}$. Note that if $\tau_{i+1}^T - \tau_i^T > \varepsilon_n \log n$, then some interval of the form $[a_{k-1}, a_k]$ with $1 \leq k \leq K$ must not contain any of the points τ_j^T. The probability that $[a_{k-1}, a_k]$ does not contain any of the τ_j^T is
$$\exp\left(-(a_k - a_{k-1}) \int_{\varepsilon_n}^{\infty} y^{-2} \, dy \right) = \exp\left(-\frac{\varepsilon_n \log n}{2} \cdot \frac{1}{\varepsilon_n} \right) = n^{-1/2}.$$

It follows that for sufficiently large n,
$$P(A_{3,n}^c) \leq K n^{-1/2} \leq \left(\frac{2(T_n + T)}{\varepsilon_n \log n} + 1 \right) n^{-1/2} \to 0$$

as $n \to \infty$. Therefore, $\lim_{n \to \infty} P(A_{3,n}) = 1$, which completes the proof. $\square$

The next lemma shows that the processes Y_n and Z_n typically jump by approximately the same amount at the jump times τ_i.

Lemma 3.7. *We have*
$$P\left(A_n \cap \left\{ \sum_{i=1}^{J_n} \left| (Z_n(\tau_i) - Z_n(\tau_i-)) - (Y_n(\tau_i) - Y_n(\tau_i-)) \right| \mathbf{1}_{\{|X_n(\tau_i-)| \leq \log \log n\}} \right.\right.$$
$$\left.\left. > \frac{1}{(\log n)^{1/8}} \right\} \right) \leq \frac{T_n (\log n)^{3/2}}{n^{1/2} \varepsilon_n}$$

for sufficiently large n.

Proof. Recall that (τ_i, y_i) is a point of Ψ for $i = 1, \ldots, J_n$. Thus, $S_n(\tau_i) - S_n(\tau_i-) = -y_i$, so
$$Y_n(\tau_i) - Y_n(\tau_i-) = -e^{-\tau_i} y_i. \tag{3.28}$$

Let $\tau_i = T_n$ and $y_i = 0$ on $\{i > J_n\}$, and let $\mathcal{G}_i = \sigma(\tau_i, N_n((\tau_i/\log n)-), y_i)$. Then on $\{i \leq J_n\}$,
$$N_n\left(\frac{\tau_i}{\log n} \right) - N_n\left(\frac{\tau_i}{\log n} - \right) = \min\{0, 1 - B_i\}, \tag{3.29}$$

where, conditional on $\mathcal{G}_i$, the distribution of B_i is the binomial distribution with parameters $N_n((\tau_i/\log n)-)$ and $y_i/\log n$. Here B_i represents the number of lineages out of $N_n((\tau_i/\log n)-)$ that merge at time $\tau_i/\log n$. We have $\mathrm{Var}(B_i|\mathcal{G}_i) \le n y_i/\log n$ on $\{i \le J_n\}$, so by the Cauchy-Schwarz Inequality,

$$E\left[\left|B_i - N_n\left(\frac{\tau_i}{\log n}-\right)\frac{y_i}{\log n}\right|\Big|\mathcal{G}_i\right] \le E\left[\left(B_i - N_n\left(\frac{\tau_i}{\log n}-\right)\frac{y_i}{\log n}\right)^2\Big|\mathcal{G}_i\right]^{1/2} \le \left(\frac{ny_i}{\log n}\right)^{1/2}$$

on $\{i \le J_n\}$. Multiplying both sides by $\mathbf{1}_{\{i \le J_n\}}\mathbf{1}_{\{y_i \le \log n\}}$, which is $\mathcal{G}_i$-measurable, and taking expectations gives

$$E\left[\left|B_i - N_n\left(\frac{\tau_i}{\log n}-\right)\frac{y_i}{\log n}\right|\mathbf{1}_{\{i \le J_n\}}\mathbf{1}_{\{y_i \le \log n\}}\right] \le n^{1/2}P(i \le J_n).$$

Summing over i and observing that $A_{2,n} \cap \{i \le J_n\} \subset \{y_i \le \log n\}$ for all i, we get

$$E\left[\mathbf{1}_{A_{2,n}}\sum_{i=1}^{J_n}\left|B_i - N_n\left(\frac{\tau_i}{\log n}-\right)\frac{y_i}{\log n}\right|\right] \le n^{1/2}E[J_n] = n^{1/2}T_n\int_{\varepsilon_n}^{\infty}y^{-2}\,dy = \frac{n^{1/2}T_n}{\varepsilon_n}.$$

Thus,

$$P\left(A_{2,n} \cap \left\{\frac{\log n}{n}\sum_{i=1}^{J_n}\left|B_i - N_n\left(\frac{\tau_i}{\log n}-\right)\frac{y_i}{\log n}\right| > \frac{1}{(\log n)^{1/2}}\right\}\right) \le \frac{T_n(\log n)^{3/2}}{n^{1/2}\varepsilon_n}. \quad (3.30)$$

Also,

$$\left|\frac{y_i}{n}N_n\left(\frac{\tau_i}{\log n}-\right) - e^{-\tau_i}y_i\right| = \left|\frac{y_i}{n}\left(\frac{n}{\log n}X_n(\tau_i-) + ne^{-\tau_i} + \frac{n\tau_i e^{-\tau_i}\log\log n}{\log n}\right) - e^{-\tau_i}y_i\right|$$

$$= y_i\left|\frac{X_n(\tau_i-)}{\log n} + \frac{\tau_i e^{-\tau_i}\log\log n}{\log n}\right|,$$

which on the event $\{|X_n(\tau_i-)| \le \log\log n\}$ is bounded by $2y_i(\log\log n)/\log n$. Thus, on $A_{2,n}$,

$$\sum_{i=1}^{J_n}\left|\frac{y_i}{n}N_n\left(\frac{\tau_i}{\log n}-\right) - e^{-\tau_i}y_i\right|\mathbf{1}_{\{|X_n(\tau_i-)|\le\log\log n\}} \le \frac{2\log\log n}{\log n}\sum_{i=1}^{J_n}y_i \le \frac{2\log\log n}{(\log n)^{1/4}}. \quad (3.31)$$

By (3.29),

$$\left|Z_n(\tau_i) - Z_n(\tau_i-) + \frac{\log n}{n}B_i\right| = \left|\frac{\log n}{n}\left(N_n\left(\frac{\tau_i}{\log n}\right) - N_n\left(\frac{\tau_i}{\log n}-\right) + B_i\right)\right| \le \frac{\log n}{n},$$

and so on $A_{1,n}$,

$$\sum_{i=1}^{J_n}\left|Z_n(\tau_i) - Z_n(\tau_i-) + \frac{\log n}{n}B_i\right| \le \frac{\log n}{n^{3/4}}. \quad (3.32)$$

Combining (3.30), (3.31), and (3.32), we get

$$P\left(A_n \cap \left\{\sum_{i=1}^{J_n}\left|(Z_n(\tau_i) - Z_n(\tau_i-)) + e^{-\tau_i}y_i\right|\mathbf{1}_{\{|X_n(\tau_i-)|\le\log\log n\}}\right.\right.$$

$$\left.\left. > \frac{1}{(\log n)^{1/2}} + \frac{2\log\log n}{(\log n)^{1/4}} + \frac{\log n}{n^{3/4}}\right\}\right) \le \frac{T_n(\log n)^{3/2}}{n^{1/2}\varepsilon_n}. \quad (3.33)$$

The lemma follows by combining (3.33) with (3.28) and using that

$$\frac{1}{(\log n)^{1/2}} + \frac{2\log\log n}{(\log n)^{1/4}} + \frac{\log n}{n^{3/4}} \le \frac{1}{(\log n)^{1/8}}$$

for sufficiently large n. $\qquad\qquad\square$

Lemmas 3.8 and 3.10 below pertain to the behavior of the processes Y_n and Z_n in between the jump times τ_i.

Lemma 3.8. *If $0 \le h < \tau_{i+1} - \tau_i \le \varepsilon_n \log n$ and n is sufficiently large, then on $\{i \le J_n\}$, we have*

$$\left| Y_n(\tau_i + h) - Y_n(\tau_i) - he^{-\tau_i}(1 - \gamma - \log \varepsilon_n + \tau_i) + \int_{\tau_i}^{\tau_i + h} Y_n(s)\, ds \right| \mathbf{1}_{\{S_n(\tau_i) \le \frac{1}{4} \log n\}} \le h^2 \log n.$$

Proof. By the construction of the process S_n, we have

$$S_n(\tau_i + h) - S_n(\tau_i) = h(1 - \gamma - \log \varepsilon_n).$$

Using $O(h^2)$ to denote an expression whose absolute value is at most h^2, Taylor's Theorem gives $e^{-(\tau_i + h)} = e^{-\tau_i} - he^{-\tau_i} + O(h^2)$ and

$$\frac{1}{2} e^{-(\tau_i + h)}(\tau_i + h)^2 = \frac{1}{2} e^{-\tau_i} \tau_i^2 + h\left(\tau_i e^{-\tau_i} - \frac{1}{2} \tau_i^2 e^{-\tau_i} \right) + O(h^2).$$

Therefore,

$$\begin{aligned}
Y_n(\tau_i + h) - Y_n(\tau_i) &= e^{-(\tau_i + h)} S_n(\tau_i + h) - e^{-\tau_i} S_n(\tau_i) + \frac{e^{-(\tau_i + h)}(\tau_i + h)^2}{2} - \frac{e^{-\tau_i} \tau_i^2}{2} \\
&= (e^{-\tau_i} - he^{-\tau_i} + O(h^2))(S_n(\tau_i) + h(1 - \gamma - \log \varepsilon_n)) - e^{-\tau_i} S_n(\tau_i) \\
&\qquad + h\left(\tau_i e^{-\tau_i} - \frac{1}{2} \tau_i^2 e^{-\tau_i} \right) + O(h^2) \\
&= he^{-\tau_i}(1 - \gamma - \log \varepsilon_n + \tau_i) - h\left(e^{-\tau_i} S_n(\tau_i) + \frac{e^{-\tau_i} \tau_i^2}{2} \right) + O(h^2) \\
&\qquad - h^2 e^{-\tau_i}(1 - \gamma - \log \varepsilon_n) + O(h^2)(S_n(\tau_i) + h(1 - \gamma - \log \varepsilon_n)).
\end{aligned}$$

On the event that $S_n(\tau_i) \le \frac{1}{4} \log n$, for sufficiently large n, the absolute value of the sum of the first two terms is bounded above by $\frac{1}{2} h \log n$, while the absolute value of the sum of the last three terms is bounded above by $\frac{1}{2} h^2 \log n$. Therefore, for sufficiently large n, on the event that $S_n(\tau_i) \le \frac{1}{4} \log n$, we have

$$\left| Y_n(\tau_i + h) - Y_n(\tau_i) - he^{-\tau_i}(1 - \gamma - \log \varepsilon_n + \tau_i) + hY_n(\tau_i) \right| \le \frac{1}{2} h^2 \log n. \tag{3.34}$$

Also, for sufficiently large n, if $0 \le s \le h$, then

$$|Y_n(\tau_i + s) - Y_n(\tau_i)| \le \frac{1}{2} s \log n + \frac{1}{2} s^2 \log n \le s \log n$$

on the event that $S_n(\tau_i) \le \frac{1}{4} \log n$ because $h \le \varepsilon_n \log n \le 1$. Therefore,

$$\left| \int_{\tau_i}^{\tau_i + h} Y_n(s)\, ds - hY_n(\tau_i) \right| \le \int_0^h |Y_n(\tau_i + s) - Y_n(\tau_i)|\, ds \le \int_0^h s \log n\, ds = \frac{1}{2} h^2 \log n. \tag{3.35}$$

By combining (3.34) and (3.35), we arrive at the statement of the lemma. $\qquad \square$

Lemma 3.9. *Suppose $0 \le h < \tau_{i+1} - \tau_i \le \varepsilon_n \log n$ and $\tau_i \le s \le \tau_i + h$. If n is sufficiently large and $|X_n(s)| \le \log \log n$, then*

$$N_n\left(\frac{s}{\log n} \right) \left| \log N_n\left(\frac{s}{\log n} \right) - \log(ne^{-\tau_i}) \right| \le \frac{7n(\log \log n)^2}{\log n}. \tag{3.36}$$

Proof. In view of (3.23), if $\tau_i \le s \le \tau_i + h$, then, using that $h \le \varepsilon_n \log n \le (\log \log n)/\log n$ for sufficiently large n and using the assumption that $|X_n(s)| \le \log \log n$, we get

$$
\begin{aligned}
\left| N_n\left(\frac{s}{\log n}\right) - ne^{-\tau_i} \right| &= \left| \frac{nX_n(s)}{\log n} + ne^{-s} + \frac{nse^{-s}\log\log n}{\log n} - ne^{-\tau_i} \right| \\
&\le \left| ne^{-s} - ne^{-\tau_i} \right| + \frac{nse^{-s}\log\log n}{\log n} + \left| \frac{nX_n(s)}{\log n} \right| \\
&\le nh + \frac{n\log\log n}{\log n} + \frac{n\log\log n}{\log n} \\
&\le \frac{3n\log\log n}{\log n}
\end{aligned}
\tag{3.37}
$$

for sufficiently large n.

We consider two cases. First, suppose $ne^{-\tau_i} \ge (3n\log\log n)/\log n$. Because (3.37) then implies that $ne^{-\tau_i} \ge \frac{1}{2}N_n(s/\log n)$, we get, using (3.37) and the fact that $\frac{d}{dx}\log x = 1/x$,

$$
\left| \log N_n\left(\frac{s}{\log n}\right) - \log(ne^{-\tau_i}) \right| \le \left(\frac{3n\log\log n}{\log n} \right) \left(\frac{1}{2}N_n\left(\frac{s}{\log n}\right) \right)^{-1}.
$$

Therefore,

$$
N_n\left(\frac{s}{\log n}\right)\left| \log N_n\left(\frac{s}{\log n}\right) - \log(ne^{-\tau_i}) \right| \le \frac{6n\log\log n}{\log n}.
$$

Next, suppose that $ne^{-\tau_i} \le (3n\log\log n)/\log n$. To determine the value of $N_n(s/\log n)$ that maximizes the left-hand side of (3.36), we consider the function $f(x) = x(\log x - \log a)$, where $a > 0$ is a constant. Note that $f'(x) = 1 + \log(x/a)$, which is positive when $x > a/e$ and negative when $x < a/e$. Thus, $f(x)$ is negative when $x < a$ and reaches its minimum when $x = a/e$, and $f(x)$ is positive and increasing for $x > a$. Therefore, (3.36) will hold in general provided that it holds if $ne^{-\tau_i}/e$ or $ne^{-\tau_i} + (3n\log\log n)/\log n$ is plugged in for $N_n(s/(\log n))$. Note that by (3.37), we need not consider larger values. In the former case, the expression that we get is

$$
\frac{ne^{-\tau_i}}{e}\left| \log\left(\frac{ne^{-\tau_i}}{e} \right) - \log(ne^{-\tau_i}) \right| = \frac{ne^{-\tau_i}}{e} \le \frac{3n\log\log n}{e\log n}.
$$

In the latter case, since $ne^{-\tau_i} \ge ne^{-T_n} = n/(\log n)^2$, the expression that we get is bounded above by

$$
\begin{aligned}
\frac{6n\log\log n}{\log n} &\left(\log\left(\frac{6n\log\log n}{\log n} \right) - \log\left(\frac{n}{(\log n)^2} \right) \right) \\
&\le \frac{6n\log\log n}{\log n}\left(\log 6 + \log\log\log n + \log\log n \right).
\end{aligned}
$$

Thus, (3.36) holds for sufficiently large n. $\qquad\square$

Lemma 3.10. *If $0 \le h < \tau_{i+1} - \tau_i \le \varepsilon_n \log n$ and n is sufficiently large, then on $\{i \le J_n\}$, we have*

$$
\begin{aligned}
\bigg| Z_n(\tau_i + h) &- Z_n(\tau_i) - he^{-\tau_i}(1 - \gamma - \log\varepsilon_n + \tau_i) \\
&+ \int_{\tau_i}^{\tau_i+h} X_n(t)\,dt \bigg|\mathbf{1}_{\{|X_n(s)| \le \log\log n \,\forall\, s \in [\tau_i, \tau_i+h]\}} \le h^2(\log n)^{1/2} + \frac{3h\log\log n}{(\log n)^{1/2}}.
\end{aligned}
$$

Proof. Throughout the proof, we will work on the event that $i \leq J_n$ and $|X_n(s)| \leq \log \log n$ for all $s \in [\tau_i, \tau_i + h]$. By the definition (3.27) of Z_n,

$$
Z_n(\tau_i + h) - Z_n(\tau_i) = \frac{\log n}{n} \left(N_n \left(\frac{\tau_i + h}{\log n} \right) - N_n \left(\frac{\tau_i}{\log n} \right) - M_n(\tau_i + h) + M_n(\tau_i) \right.
$$

$$
\left. - ne^{-(\tau_i + h)} + ne^{-\tau_i} - \frac{n(\tau_i + h)e^{-(\tau_i+h)} \log \log n}{\log n} + \frac{n\tau_i e^{-\tau_i} \log \log n}{\log n} \right).
$$

Using the definition of M_n in (3.25) and (3.26), we get

$$
Z_n(\tau_i + h) - Z_n(\tau_i) = -\frac{\log n}{n} \int_{\tau_i/\log n}^{(\tau_i+h)/\log n} \eta^*(N_n(t))\, dt + (\log n)(e^{-\tau_i} - e^{-(\tau_i+h)})
$$

$$
+ (\log \log n)(\tau_i e^{-\tau_i} - (\tau_i + h)e^{-(\tau_i+h)}). \tag{3.38}
$$

To estimate the integral, we need to estimate $\eta^*(N_n(t))$. Equation (3.17), which holds for sufficiently large n even when $b = 1$ if we define $\eta^*(1) = 0$, gives

$$
\left| \eta^*(N_n(t)) - N_n(t)(\log N_n(t) - \log \log n + \log \varepsilon_n + \gamma - 1) \right| \leq \frac{\log n}{\varepsilon_n} + 1. \tag{3.39}
$$

It follows from (3.39) and Lemma 3.9, after making the substitution $t = s/\log n$, that if $\tau_i/\log n \leq t \leq (\tau_i + h)/\log n$, then for sufficiently large n,

$$
\left| \eta^*(N_n(t)) - N_n(t)(\log n - \log \log n + \log \varepsilon_n + \gamma - 1 - \tau_i) \right| \leq \frac{8n(\log \log n)^2}{\log n}. \tag{3.40}
$$

Using (3.23), we have

$$
\int_{\tau_i/\log n}^{(\tau_i+h)/\log n} N_n(t)\, dt = \frac{1}{\log n} \int_{\tau_i}^{\tau_i+h} N_n \left(\frac{s}{\log n} \right) ds
$$

$$
= \frac{1}{\log n} \int_{\tau_i}^{\tau_i+h} ne^{-s} + \frac{nse^{-s} \log \log n}{\log n} + \frac{nX_n(s)}{\log n}\, ds
$$

$$
= \frac{n(e^{-\tau_i} - e^{-(\tau_i+h)})}{\log n} + \frac{n \log \log n}{(\log n)^2} \big((\tau_i + 1)e^{-\tau_i}
$$

$$
- (\tau_i + h + 1)e^{-(\tau_i+h)} \big) + \frac{n}{(\log n)^2} \int_{\tau_i}^{\tau_i+h} X_n(s)\, ds.
$$

Combining this result with (3.40) and letting ξ denote an expression whose absolute value is less than $8h(\log \log n)^2/(\log n)$, we get

$$
\frac{\log n}{n} \int_{\tau_i/\log n}^{(\tau_i+h)/\log n} \eta^*(N_n(t))\, dt
$$

$$
= \frac{(\log n)(\log n - \log \log n + \log \varepsilon_n + \gamma - 1 - \tau_i)}{n} \int_{\tau_i/\log n}^{(\tau_i+h)/\log n} N_n(t)\, dt + \xi
$$

$$
= (\log n - \log \log n + \log \varepsilon_n + \gamma - 1 - \tau_i)(e^{-\tau_i} - e^{-(\tau_i+h)})
$$

$$
+ (\log \log n)\big((\tau_i + 1)e^{-\tau_i} - (\tau_i + h + 1)e^{-(\tau_i+h)} \big) + \int_{\tau_i}^{\tau_i+h} X_n(s)\, ds
$$

$$
+ \left(\frac{(-\log \log n + \log \varepsilon_n + \gamma - 1 - \tau_i)(\log \log n)}{\log n} \right)
$$

$$
\times \big((\tau_i + 1)e^{-\tau_i} - (\tau_i + h + 1)e^{-(\tau_i+h)} \big)
$$

$$
+ \left(\frac{-\log \log n + \log \varepsilon_n + \gamma - 1 - \tau_i}{\log n} \right) \int_{\tau_i}^{\tau_i+h} X_n(s)\, ds + \xi. \tag{3.41}
$$

Recall that $\log \varepsilon_n = -(\log n)^{1/2}$. One can show using calculus that if $g(x) = (x+1)e^{-x}$, then $|g'(x)| \le 1$ for all $x > 0$. Consequently, we have

$$|(\tau_i + 1)e^{-\tau_i} - (\tau_i + h + 1)e^{-(\tau_i + h)}| \le h.$$

Also, because we are assuming $|X_n(s)| \le \log \log n$ for all $s \in [\tau_i, \tau_i + h]$, we have $|\int_{\tau_i}^{\tau_i + h} X_n(s)\, ds| \le h \log \log n$. It follows from these observations that the sum of the terms on the last two lines of (3.41) is bounded by $3h(\log \log n)(\log n)^{-1/2}$ for sufficiently large n. Therefore,

$$
\begin{aligned}
\frac{\log n}{n} \int_{\tau_i / \log n}^{(\tau_i + h)/\log n} & \eta^*(N_n(t))\, dt \\
&= (\log n)(e^{-\tau_i} - e^{-(\tau_i + h)}) + (\log \log n)(\tau_i e^{-\tau_i} - (\tau_i + h)e^{-(\tau_i + h)}) \\
&\quad + (\log \varepsilon_n + \gamma - 1 - \tau_i)(e^{-\tau_i} - e^{-(\tau_i + h)}) + \int_{\tau_i}^{\tau_i + h} X_n(s)\, ds + \xi',
\end{aligned}
$$

where $|\xi'| \le 3h(\log \log n)(\log n)^{-1/2}$. Combining this result with (3.38), we get

$$Z_n(\tau_i + h) - Z_n(\tau_i) = (1 - \gamma - \log \varepsilon_n + \tau_i)(e^{-\tau_i} - e^{-(\tau_i + h)}) - \int_{\tau_i}^{\tau_i + h} X_n(s)\, ds - \xi'.$$

Because $-\log \varepsilon_n = (\log n)^{1/2}$ and $|(e^{-\tau_i} - e^{-(\tau_i + h)}) - he^{-\tau_i}| \le h^2/2$, the result follows. □

Lemma 3.11. *Define the events*

$$
\begin{aligned}
B_{1,n} &= \left\{ \sup_{0 \le s \le T} \sup_{t \ge 0} e^{-t} |S(s,t)| \le \frac{1}{2} \log \log n \right\}. \\
B_{2,n} &= \left\{ \sup_{0 \le s \le T} \sup_{0 \le t \le T_n + s} |S(s,t)| \le 5(\log \log n)^2 \right\}. \\
B_{3,n} &= \left\{ \sup_{0 \le s \le T} \sup_{0 \le t \le T_n + s} |S_n(s,t) - S(s,t)| \le 1 \right\}.
\end{aligned}
$$

Let $B_n = B_{1,n} \cap B_{2,n} \cap B_{3,n}$. Then $\lim_{n \to \infty} P(B_n) = 1$.

Proof. Note that $B_{1,n}$ is the complement of the event on the right-hand side of (3.6) with $\frac{1}{2} \log \log n$ in place of K, and $B_{2,n}$ is the complement of the event in (3.7). Therefore, by Lemma 3.1, we have $\lim_{n \to \infty} P(B_{1,n}) = \lim_{n \to \infty} P(B_{2,n}) = 1$. We also have $\lim_{n \to \infty} P(B_{3,n}) = 1$ by Lemma 3.2. □

Recall that the event A_n was defined in Lemma 3.6. The next lemma shows that when A_n and B_n occur, it is unlikely that the processes X_n and Y_n are ever far apart before time T_n. In view of Lemmas 3.6 and 3.11, this result implies (3.24), and therefore Theorem 1.7.

Lemma 3.12. *We have*

$$P\left(A_n \cap B_n \cap \left\{ \sup_{0 \le t \le T_n} |X_n(t) - Y_n(t)| > \frac{4}{(\log n)^{1/16}} \right\} \right) \le \frac{T_n (\log n)^{3/2}}{n^{1/2} \varepsilon_n} + 4(\log n)^2 T_n \varepsilon_n$$

for sufficiently large n.

Proof. We claim that on the event

$$A_n \cap B_n \cap \left\{ \frac{\log n}{n} \sup_{0 \le t \le T_n} |M_n(t)| \le \frac{1}{\log n} \right\}$$

$$\cap \left\{ \sum_{i=1}^{J_n} \left| (Z_n(\tau_i) - Z_n(\tau_i-)) - (Y_n(\tau_i) - Y_n(\tau_i-)) \right| \mathbf{1}_{\{|X_n(\tau_i-)| \le \log \log n\}} \le \frac{1}{(\log n)^{1/8}} \right\},$$

(3.42)

we have $|Z_n(t) - Y_n(t)| \le 3/(\log n)^{1/16}$ for all $t \le T_n$, and therefore $|X_n(t) - Y_n(t)| \le 4/(\log n)^{1/16}$ for all $t \le T_n$ by (3.27). Thus, by Lemmas 3.5 and 3.7, this claim implies the result.

To prove the claim, let $\zeta = \inf\{t \le T_n : |Z_n(t) - Y_n(t)| > 3/(\log n)^{1/16}\}$. We assume that we are working on the event in (3.42), and we must show that ζ is the infimum of the empty set, and thus that $\zeta = \infty$. On $B_{1,n} \cap B_{3,n}$, we have

$$|Y_n(t)| \le \frac{1}{2} \log \log n + \frac{e^{-t}t^2}{2} + |Y_n(t) - Y(t)| \le \frac{1}{2} \log \log n + 2$$

for all $t \le T_n$. Therefore, if $t < \zeta$ and n is sufficiently large, then

$$|X_n(t)| \le |Y_n(t)| + |Z_n(t) - Y_n(t)| + \frac{\log n}{n}|M_n(t)|$$

$$\le \frac{1}{2} \log \log n + 2 + \frac{3}{(\log n)^{1/16}} + \frac{1}{\log n}$$

$$\le \log \log n.$$

(3.43)

Note also that $S_n(t) \le \frac{1}{4} \log n$ for all $t \le T_n$ on $B_{2,n} \cap B_{3,n}$ for sufficiently large n, and $\tau_{i+1} - \tau_i \le \varepsilon_n \log n$ for $i = 0, 1, \ldots, J_n$ on $A_{3,n}$. Therefore, Lemmas 3.8 and 3.10 imply that if $\tau_i + h < \zeta$ and $h < \tau_{i+1} - \tau_i$, then

$$\left| (Z_n(\tau_i + h) - Z_n(\tau_i)) - (Y_n(\tau_i + h) - Y_n(\tau_i)) + \int_{\tau_i}^{\tau_i + h} (X_n(t) - Y_n(t)) \, dt \right|$$

$$\le \frac{3h \log \log n}{(\log n)^{1/2}} + h^2 (\log n)^{1/2} + h^2 \log n.$$

Since

$$\int_{\tau_i}^{\tau_i + h} |X_n(t) - Z_n(t)| \, dy = \int_{\tau_i}^{\tau_i + h} \frac{\log n}{n} |M_n(t)| \le \frac{h}{\log n},$$

it follows that

$$\left| (Z_n(\tau_i + h) - Z_n(\tau_i)) - (Y_n(\tau_i + h) - Y_n(\tau_i)) + \int_{\tau_i}^{\tau_i + h} (Z_n(t) - Y_n(t)) \, dt \right| \le \frac{4h \log \log n}{(\log n)^{1/2}} \quad (3.44)$$

for sufficiently large n.

Define the sets

$$I_1 = \left\{ i \in \{0, 1, \ldots, J_n + 1\} : |Z_n(\tau_i) - Y_n(\tau_i)| \le \frac{1}{(\log n)^{1/16}} \right\},$$

$$I_2 = \left\{ i \in \{0, 1, \ldots, J_n + 1\} : Z_n(\tau_i) - Y_n(\tau_i) > \frac{1}{(\log n)^{1/16}} \right\},$$

$$I_3 = \left\{ i \in \{0, 1, \ldots, J_n + 1\} : Z_n(\tau_i) - Y_n(\tau_i) < -\frac{1}{(\log n)^{1/16}} \right\}.$$

Clearly, we have $I_1 \cup I_2 \cup I_3 = \{0, 1, \ldots, J_n + 1\}$. By right continuity and the fact that $Z_n(0) = Y_n(0) = 0$, we have $\zeta > 0 = \tau_0$. Suppose that $\zeta \in (\tau_i, \tau_{i+1})$ for some $i \in I_1$. Then, by (3.44) and the fact that $\tau_{i+1} - \tau_i \le \varepsilon_n \log n$,

$$\begin{aligned}
|Z_n(\zeta) - Y_n(\zeta)| &\le |Z_n(\tau_i) - Y_n(\tau_i)| + \int_{\tau_i}^{\zeta} |Z_n(t) - Y_n(t)| \, dt + \frac{4(\zeta - \tau_i) \log \log n}{(\log n)^{1/2}} \\
&\le \frac{1}{(\log n)^{1/16}} + \varepsilon_n \log n \cdot \frac{3}{(\log n)^{1/16}} + 4\varepsilon_n (\log n)^{1/2} \log \log n, \quad (3.45)
\end{aligned}$$

a contradiction for sufficiently large n because $|Z_n(\zeta) - Y_n(\zeta)| \ge 3/(\log n)^{1/16}$ by right continuity. Thus, if $i \in I_1$, then $\zeta \ge \tau_{i+1}$. Therefore, if $i \in I_1$, then reasoning as in (3.45) and using the fact that we are working on the event (3.42), we get

$$\begin{aligned}
|Z_n(\tau_{i+1}) - Y_n(\tau_{i+1})| &\le |Z_n(\tau_i) - Y_n(\tau_i)| + \int_{\tau_i}^{\tau_{i+1}} |Z_n(t) - Y_n(t)| \, dt + \frac{4(\tau_{i+1} - \tau_i) \log \log n}{(\log n)^{1/2}} \\
&\quad + |(Z_n(\tau_{i+1}) - Z_n(\tau_{i+1}-)) - (Y_n(\tau_{i+1}) - Y_n(\tau_{i+1}-))| \\
&\le \frac{1}{(\log n)^{1/16}} + \varepsilon_n \log n \cdot \frac{3}{(\log n)^{1/16}} \\
&\quad + 4\varepsilon_n (\log n)^{1/2} \log \log n + \frac{1}{(\log n)^{1/8}} \\
&\le \frac{2}{(\log n)^{1/16}} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (3.46)
\end{aligned}$$

for sufficiently large n, which implies that $\zeta \ne \tau_{i+1}$.

Next, suppose that $i \in I_2$. Let $\rho = \inf\{t > \tau_i : Z_n(t) \le Y_n(t)\}$. Suppose $\rho \le \min\{\zeta, \tau_{i+1}\}$. Then $0 \le Z_n(s) - Y_n(s) \le 3/(\log n)^{1/16}$ for all $s < \rho$. Therefore by (3.44),

$$\begin{aligned}
Z_n(\rho) - Y_n(\rho) &\ge Z_n(\tau_i) - Y_n(\tau_i) - \int_{\tau_i}^{\rho} (Z_n(t) - Y_n(t)) \, dt - \frac{4(\rho - \tau_i) \log \log n}{(\log n)^{1/2}} - \frac{1}{(\log n)^{1/8}} \\
&\ge \frac{1}{(\log n)^{1/16}} - \varepsilon_n \log n \cdot \frac{3}{(\log n)^{1/16}} - 4\varepsilon_n (\log n)^{1/2} \log \log n - \frac{1}{(\log n)^{1/8}},
\end{aligned}$$
$$(3.47)$$

which is positive for sufficiently large n, a contradiction. Therefore, for sufficiently large n, if $i \in I_2$, then Z_n must stay greater than Y_n from time τ_i until after time $\min\{\zeta, \tau_{i+1}\}$. In particular, if $i \in I_2$ and $\zeta \ge \tau_{i+1}$, then $i + 1 \in I_1 \cup I_2$. By the same argument with the roles of Y_n and Z_n reversed, if $i \in I_3$ and $\zeta \ge \tau_{i+1}$, then $i + 1 \in I_1 \cup I_3$. It follows from these observations and (3.46) that the only way we could have $\zeta \in (\tau_i, \tau_{i+1}]$ with $i \in I_2$ is if there exists $j < i$ such that $j \in I_1$, $j + 1, j + 2, \ldots, i \in I_2$, and

$$Z_n(\zeta) - Y_n(\zeta) > Z_n(\tau_{j+1}) - Y_n(\tau_{j+1}) + \frac{1}{(\log n)^{1/16}}.$$

However, if this is true, then $Z_n(t) > Y_n(t)$ for all $t \in [\tau_{j+1}, \zeta]$. Therefore, using equations (3.43) and (3.44) and the fact that we are working on the event in (3.42), we have

$$\begin{aligned}
Z_n(\zeta) - Y_n(\zeta) &\le Z_n(\tau_{j+1}) - Y_n(\tau_{j+1}) + \sum_{k=j+2}^{i+1} \left| (Z_n(\tau_k) - Z_n(\tau_k-)) - (Y_n(\tau_k) - Y_n(\tau_k-)) \right| \\
&\quad + \frac{4(\zeta - \tau_{j+1})(\log \log n)}{(\log n)^{1/2}} \\
&\le Z_n(\tau_{j+1}) - Y_n(\tau_{j+1}) + \frac{1}{(\log n)^{1/8}} + \frac{4T_n(\log \log n)}{(\log n)^{1/2}}.
\end{aligned}$$

For sufficiently large n, the sum of the last two terms on the right-hand side is less than $1/(\log n)^{1/16}$, a contradiction. Hence, we can not have $\zeta \in (\tau_i, \tau_{i+1}]$ for $i \in I_2$, and the same argument with the roles of Y_n and Z_n reversed establishes that we can not have $\zeta \in (\tau_i, \tau_{i+1}]$ with $i \in I_3$. We conclude that $\zeta = \infty$, which completes the proof. $\qquad\square$

3.5 Extension to arbitrary starting times

The result (3.24) pertains to the evolution of the number of lineages when we trace back the ancestral lines of the individuals in the population at time zero. Of course, by stationarity, the analogous result holds if we instead trace back the ancestral lines of the individuals in the population at some other time $s \geq 0$. However, to help with the proof of Theorem 1.5 in the next section, we will need a stronger version of this result that will make it possible to show that the approximation works well simultaneously at many times. The key to this result is that the events A_n and B_n were defined so that the bounds that hold on these events are valid simultaneously for all $s \in [0, T]$.

Recall that $N_n(s,t)$ denotes the number of individuals at time $(s-t)-$ with a descendant alive in the population at time s. Let $\tilde{N}_n(s,t) = N_n(s,t)\mathbf{1}_{\{N(s,t)>1\}}$, which will be convenient because

$$L_n(s) = \int_0^\infty \tilde{N}_n(s,t)\,dt. \qquad (3.48)$$

Let

$$X_n(s,t) = \frac{\log n}{n}\left(\tilde{N}_n\left(\frac{s}{\log n}, \frac{t}{\log n} \right) - ne^{-t} - \frac{nte^{-t}\log\log n}{\log n} \right).$$

Also, let

$$Y_n(s,t) = e^{-t}S_n(s,t) + \frac{e^{-t}t^2}{2}.$$

Note that $X_n(0,t) = X_n(t)$ for all $t \leq (\log n)\inf\{s : N_n(s) = 1\}$ and $Y_n(0,t) = Y_n(t)$ for all $t \geq 0$.

We have the following extension of Lemma 3.12. The result follows from the same argument that gives Lemma 3.12. We have replaced $4/(\log n)^{1/16}$ by $5/(\log n)^{1/16}$ to account for the error in replacing $N_n(s,t)$ by $\tilde{N}_n(s,t)$. Indeed, it would be possible to use $4/(\log n)^{1/16} + (\log n)/n$ because $|N_n(s,t) - \tilde{N}_n(s,t)| \leq 1$ for all s and t. Also, because an interval of time $T_n + s$ rather than T_n must be considered when adapting the proofs of Lemmas 3.5 and 3.7, the bounds involving T_n have been replaced by bounds involving $T_n + T$.

Lemma 3.13. *For sufficiently large n, we have*

$$P\left(A_n \cap B_n \cap \left\{ \sup_{0 \leq t \leq T_n + s} |X_n(s,t) - Y_n(s,t)| > \frac{5}{(\log n)^{1/16}} \right\} \right)$$

$$\leq \frac{(T_n + T)(\log n)^{3/2}}{n^{1/2}\varepsilon_n} + 4(\log n)^2(T_n + T)\varepsilon_n$$

for each fixed $s \in [0, T]$.

4 Proof of Theorem 1.5

Recall that for each $s \in \mathbb{R}$, a process $(S(s,t), t \geq 0)$ was defined in section 3.1. Let $J(s) = y$ if $(-s, y)$ is a point of Ψ, and let $J(s) = 0$ otherwise. For each $s \in \mathbb{R}$ and $t \geq 0$, let

$$Y(s,t) = e^{-t}S(s,t) + \frac{e^{-t}t^2}{2},$$

and then let

$$L(s) = -J(s) + \int_0^\infty Y(s,t)\,dt.$$

Note that $P(J(s) = 0) = 1$ for each $s \geq 0$, but the $-J(s)$ term keeps the process $(L(s), s \geq 0)$ right continuous at the jump times and thus ensures that $(L(s), s \geq 0)$ has paths that are almost surely right continuous with left limits. To see this, note that if $(-s,y)$ is a point of Ψ, then $S(s+,t) = S(s,t) - y$ and thus $Y(s+,t) = Y(s,t) - e^{-t}y$ for all $t > 0$. Therefore,

$$\int_0^\infty Y(s+,t)\,dt = \int_0^\infty Y(s,t)\,dt - y.$$

The next result shows that $(L(s), s \geq 0)$ has the same law as the process defined in the statement of Theorem 1.5.

Proposition 4.1. *Let ν be the measure on $\mathbb{R}$ whose density with respect to Lebesgue measure is given by $x^{-2}\mathbf{1}_{(-\infty,0)}(x)$. The process $(L(s), s \geq 0)$ is a process of Ornstein-Uhlenbeck type generated by $(2 - \gamma, 0, \nu, 1)$.*

Proof. Fix $t \geq 0$. Note that if $0 \leq s \leq t$ and $u \geq 0$, then $S(s,u) = S(t,t-s+u) - S(t,t-s)$. We use the notation $\int_0^\infty f(x)\,dS_t(x)$ to denote the stochastic integral of $f(x)$ with respect to the stable process $(S(t,x), x \geq 0)$. If $0 \leq s \leq t$, then

$$\begin{aligned}
L(s) &= 1 - J(s) + \int_0^\infty e^{-u}S(s,u)\,du \\
&= 1 - J(s) + \int_0^\infty e^{-u}\left(\int_{t-s}^{t-s+u} dS_t(x)\right)du \\
&= 1 - J(s) + \int_{t-s}^\infty \left(\int_{x-t+s}^\infty e^{-u}\,du\right)dS_t(x) \\
&= 1 - J(s) + \int_{t-s}^\infty e^{-(x-t+s)}\,dS_t(x),
\end{aligned} \tag{4.1}$$

where in the next-to-last step we used Fubini's Theorem for general stochastic integrals (see, for example, Theorem 45 in Part IV of [27] and the remark on p. 161 in [27] that the measure μ in that theorem can be taken to be σ-finite rather than finite). Note that $J(0) = 0$ almost surely, and almost surely $\int_0^t J(s)\,ds = 0$ for all t. Therefore, using Fubini's Theorem again and (4.1), we get

$$\begin{aligned}
L(t) - L(0) + \int_0^t L(s)\,ds &= t - J(t) + \int_0^\infty e^{-x}\,dS_t(x) - \int_t^\infty e^{-(x-t)}\,dS_t(x) \\
&\quad + \int_0^t \int_{t-s}^\infty e^{-(x-t+s)}\,dS_t(x)\,ds \\
&= t - J(t) + \int_0^\infty e^{-x}\,dS_t(x) - \int_t^\infty e^{-(x-t)}\,dS_t(x) \\
&\quad + \int_0^\infty \int_{\max\{0,t-x\}}^t e^{-(x-t+s)}\,ds\,dS_t(x) \\
&= t - J(t) + \int_0^t \left(e^{-x} + \int_{t-x}^t e^{-(x-t+s)}\,ds\right)dS_t(x) \\
&\quad + \int_t^\infty \left(e^{-x} - e^{-(x-t)} + \int_0^t e^{-(x-t+s)}\,ds\right)dS_t(x)
\end{aligned} \tag{4.2}$$

for all $t \geq 0$ almost surely. Because

$$e^{-x} + \int_{t-x}^{t} e^{-(x-t+s)} \, ds = 1$$

and

$$e^{-x} - e^{-(x-t)} + \int_{0}^{t} e^{-(x-t+s)} \, ds = 0,$$

it follows that

$$L(t) - L(0) + \int_{0}^{t} L(s) \, ds = t - J(t) + \int_{0}^{t} dS_t(x) = t - J(t) + S(t,t)$$

for all $t \geq 0$ almost surely. Therefore, if we define $Z(t) = S(t,t) + t - J(t)$ for all $t \geq 0$, then

$$L(t) = L(0) + Z(t) - \int_{0}^{t} L(s) \, ds, \tag{4.3}$$

which is exactly (1.9) with $c = 1$ and L in place of X.

By the symmetry of the construction in section 3.1, the processes $(S(0,t), t \geq 0)$ and $(S(t,t) - J(t), t \geq 0)$ have the same law. Indeed, if one reflects the points of the Poisson process about the vertical axis, so that a point at (t,x) is moved to $(-t,x)$, and then follows the procedure used to construct $(S(0,t), t \geq 0)$, one obtains the process $(S(t,t) - J(t), t \geq 0)$ provided that $J(0) = 0$. Therefore, the process $(S(t,t) - J(t), t \geq 0)$ is a stable process whose characteristic exponent is given by the expression in (3.2). Since $(Z(t), t \geq 0)$ differs from $(S(t,t) - J(t), t \geq 0)$ only by the addition of a linear drift of rate 1, it follows that $(Z(t), t \geq 0)$ is a stable process whose characteristic exponent is obtained by replacing $1 - \gamma$ with $2 - \gamma$ on the right-hand side of (3.2). This observation, combined with (4.3), implies the result. $\square$

Remark 4.2. The process $(L(s), s \geq 0)$ is clearly stationary by construction. Thus, it follows from Propositon 4.1 and the theory of processes of Ornstein-Uhlenbeck type reviewed in the introduction that the distribution of $L(0)$, and therefore the distribution of $L(s)$ for any fixed $s \geq 0$, has a characteristic function given by the right-hand side of (1.8). To observe this result more directly, use the Integration by Parts Formula (see, for example, Corollary 8.7 of [21]) to write that almost surely

$$L(0) = \int_{0}^{\infty} Y(t) \, dt = 1 + \int_{0}^{\infty} e^{-t} S(t) \, dt = 1 + \int_{0}^{\infty} e^{-t} dS(t). \tag{4.4}$$

The distribution of the stable integral on the right-hand side of (4.4) can be evaluated using the theory developed in Chapter 3 of [29]. In this case, we apply Proposition 3.4.1 of [29] with m being $\pi/2$ times Lebesgue measure, $E = [0, \infty)$, $\sigma = \pi/2$, $\mu = 0$, $\beta = -1$, and $f(t) = e^{-t}$. We get $\sigma_f = \pi/2$, $\beta_f = -1$, and $\mu_f = -\int_{0}^{\infty} t e^{-t} \, dt = -1$ to recover the result.

By (3.48), we have

$$\frac{(\log n)^2}{n} \left(L_n \left(\frac{s}{\log n} \right) - \frac{n}{\log n} - \frac{n \log \log n}{(\log n)^2} \right)$$

$$= \frac{(\log n)^2}{n} \int_{0}^{\infty} \frac{1}{\log n} \tilde{N}_n \left(\frac{s}{\log n}, \frac{t}{\log n} \right) - \frac{n e^{-t}}{\log n} - \frac{n t e^{-t} \log \log n}{(\log n)^2} \, dt = \int_{0}^{\infty} X_n(s,t) \, dt.$$

Consequently, the following proposition will imply Theorem 1.5.

Proposition 4.3. *We have*

$$\sup_{0 \le s \le T} \left| \int_0^\infty X_n(s,t)\, dt - L(s) \right| \to_p 0.$$

To prove Proposition 4.3, we need to compare $\int_0^\infty X_n(s,t)\, dt$ and $\int_0^\infty Y(s,t)\, dt$. Lemmas 4.4 and 4.6 below will show that it suffices to consider the integrals of X_n and Y up to time $T_n + s$, and Lemma 4.7 will allow us to replace Y by Y_n.

Lemma 4.4. *We have*

$$\sup_{0 \le s \le T} \int_{T_n+s}^\infty Y(s,t)\, dt \to_p 0.$$

Proof. We have

$$\sup_{0 \le s \le T} \int_{T_n+s}^\infty e^{-t} S(s,t)\, dt \to_p 0$$

by (3.9). Also, it is clear that

$$\lim_{n \to \infty} \int_{T_n+s}^\infty \frac{e^{-t} t^2}{2}\, dt = 0.$$

The result is immediate. □

Lemma 4.5. *For all $\varepsilon > 0$, there exists a positive constant C such that*

$$\liminf_{n \to \infty} P\left(N_n\left(\frac{T_n}{\log n} \right) \le \frac{Cn}{(\log n)^2} \right) > 1 - \varepsilon.$$

Proof. Let $(\Pi_n(t), t \ge 0)$ be a Bolthausen-Sznitman coalescent started with n blocks. There is a well-known method for constructing a random partition having the same distribution as $\Pi_n(t)$. Let $\alpha = e^{-t}$. Then let $J_1 \ge J_2 \ge \dots$ denote the points of a Poisson point process on $(0, \infty)$ whose intensity measure is given by $x^{-1-\alpha}\, dx$. Let $S = \sum_{i=1}^\infty J_i$, and divide the interval $[0,1]$ into disjoint subintervals of lengths $J_1/S, J_2/S, \dots$. Let $U_1, U_2, \dots$ be i.i.d. random variables having the uniform distribution on $[0,1]$. Then $\Pi_n(t)$ has the same distribution as the partition of $\{1, \dots, n\}$ such that i and j are in the same block if and only if U_i and U_j land in the same subinterval of $[0,1]$; see, for example, section 4 of [5] or section 2.4 of [25]. Therefore, $N_n(t)$ has the same distribution as the number of blocks of $\Pi_n(t)$. Thus, by (3.13) of [26],

$$E[N_n(t)] = \frac{\Gamma(n+\alpha)}{\alpha \Gamma(\alpha) \Gamma(n)}.$$

Therefore, by Stirling's Formula, there exist positive constants C_1 and C_2 such that for all $n \ge 2$ and $t \ge 0$,

$$E[N_n(t)] \le \frac{C_1 (n-1+\alpha)^{n+\alpha-1/2} e^{-(n-1+\alpha)}}{\alpha \Gamma(\alpha)(n-1)^{n-1/2} e^{-(n-1)}} \le \frac{C_2}{\alpha \Gamma(\alpha)} n^\alpha. \tag{4.5}$$

Now, take $t = T_n/(\log n)$, so that $\alpha = e^{-T_n/(\log n)}$. Because

$$\alpha \le 1 - \frac{T_n}{\log n} + \frac{T_n^2}{2(\log n)^2},$$

we have

$$n^\alpha \le n e^{-T_n + T_n^2/2 \log n} \le \frac{C_3 n}{(\log n)^2} \tag{4.6}$$

for some positive constant C_3. Combining (4.5) and (4.6), we get that

$$E\left[N_n\left(\frac{T_n}{\log n} \right) \right] \le \frac{C_4 n}{(\log n)^2}$$

for some positive constant C_4. The result now follows from Markov's Inequality. □

Lemma 4.6. *We have*

$$\sup_{0 \le s \le T} \int_{T_n+s}^{\infty} X_n(s,t)\, dt \to_p 0.$$

Proof. It is easy to check that

$$\lim_{n \to \infty} \frac{\log n}{n} \int_{T_n}^{\infty} \left(ne^{-t} + \frac{nte^{-t}\log\log n}{\log n} \right) dt = 0.$$

Therefore, it suffices to show that

$$\sup_{0 \le s \le T} \frac{\log n}{n} \int_{T_n+s}^{\infty} \tilde{N}_n \left(\frac{s}{\log n}, \frac{t}{\log n} \right) dt \to_p 0.$$

Note that

$$N_n \left(\frac{s}{\log n}, \frac{t}{\log n} \right) \le N_n \left(0, \frac{t-s}{\log n} \right)$$

because the number of individuals in the population at time $((s-t)/\log n)-$ who have descendants in the population at time 0 will be at least as large as the number of individuals in the population at time $((s-t)/\log n)-$ who have descendants in the population at the later time $s/(\log n)$. It follows that

$$\sup_{0 \le s \le T} \frac{\log n}{n} \int_{T_n+s}^{\infty} \tilde{N}_n \left(\frac{s}{\log n}, \frac{t}{\log n} \right) dt \le \sup_{0 \le s \le T} \frac{\log n}{n} \int_{T_n+s}^{\infty} \tilde{N}_n \left(0, \frac{t-s}{\log n} \right) dt$$

$$= \frac{\log n}{n} \int_{T_n}^{\infty} \tilde{N}_n \left(0, \frac{t}{\log n} \right) dt$$

$$= \frac{(\log n)^2}{n} \int_{T_n/\log n}^{\infty} \tilde{N}_n(0,t)\, dt. \qquad (4.7)$$

Conditional on $N_n(0, T_n/\log n) = m \ge 2$, the distribution of $\int_{T_n/\log n}^{\infty} \tilde{N}_n(0,t)\, dt$ is the same as the distribution of $L_m(0)$. Let $\varepsilon > 0$. By Lemma 4.5, there is a positive constant C such that $P(N_n(0, T_n/\log n) \le Cn/(\log n)^2) > 1-\varepsilon$ for sufficiently large n. Conditional on the event that $N_n(0, T_n/\log n) \le Cn/(\log n)^2$, the distribution of $\int_{T_n/\log n}^{\infty} \tilde{N}_n(0,t)\, dt$ is stochastically dominated by the distribution of $L_m(0)$ for $m = \lfloor Cn/(\log n)^2 \rfloor$. Thus, by (1.7), there is a positive constant K such that

$$P \left(\int_{T_n/\log n}^{\infty} N_n(0,t)\, dt \le \frac{Kn}{(\log n)^3} \right) > 1 - 2\varepsilon$$

for sufficiently large n. Hence, the right-hand side of (4.7) converges in probability to zero, which gives the result. $\qquad \square$

Lemma 4.7. *We have*

$$\sup_{0 \le s \le T} \int_0^{T_n+s} |Y_n(s,t) - Y(s,t)|\, dt \to_p 0.$$

Proof. Note that

$$\int_0^{T_n+s} |Y_n(s,t) - Y(s,t)|\, dt = \int_0^{T_n+s} e^{-t} |S_n(s,t) - S(s,t)|\, dt$$

$$\le \sup_{0 \le t \le T_n+s} |S_n(s,t) - S(s,t)|. \qquad (4.8)$$

Therefore,

$$\sup_{0 \le s \le T} \int_0^{T_n+s} |Y_n(s,t) - Y(s,t)|\, dt \le \sup_{0 \le s \le T} \sup_{0 \le t \le T_n+s} |S_n(s,t) - S(s,t)| \to_p 0$$

by Lemma 3.2. $\qquad \square$

Choose fixed times $0 = s_0 < s_1 < \cdots < s_m = T$ such that $1/(\log n)^2 \leq s_{i+1} - s_i \leq 2/(\log n)^2$ for $i = 0, 1, \ldots, m-1$. This is clearly possible for sufficiently large n, and $m \leq T(\log n)^2$. Let $\varepsilon > 0$, and let $(-u_1, y_1), \ldots, (-u_k, y_k)$ denote the points of Ψ in the region $[-T, 0] \times [\varepsilon^3, \infty)$. Note that with probability one, there are only finitely many points of Ψ in this region. For $i = 0, 1, \ldots, m-1$, let G_i be the event that none of the u_j falls in $[s_i, s_{i+1}]$, and let H_i be the event that exactly one of the u_j falls in $[s_i, s_{i+1}]$ and that this point u_j is in (s_i, s_{i+1}). Note that almost surely none of the u_j land on one of the points $s_0, \ldots, s_m$, and with probability tending to one as $n \to \infty$, no two of the points u_j fall in the same interval $[s_i, s_{i+1}]$. Consequently, we have

$$\lim_{n \to \infty} P\left(\bigcap_{i=0}^{m-1} (G_i \cup H_i) \right) = 1. \tag{4.9}$$

On H_i, let $j(i)$ be the value of j such that $u_j \in (s_i, s_{i+1})$.

Lemma 4.8. *We have*

$$\limsup_{n \to \infty} P\left(\sup_{0 \leq i \leq m-1} \sup_{s \in [s_i, s_{i+1}]} \left| \int_0^{T_n + s} Y_n(s, t)\, dt - \int_0^{T_n + s_{i+1}} Y_n(s_{i+1}, t)\, dt \right| \mathbf{1}_{G_i} \geq 4\varepsilon \right) \leq 4T\varepsilon.$$

Proof. If $s_i \leq s \leq s_{i+1}$ and $t \geq 0$, then

$$S_n(s, t) = S_n(s_{i+1}, t + s_{i+1} - s) - S_n(s_{i+1}, s_{i+1} - s).$$

Therefore,

$$
\begin{aligned}
\int_0^{T_n+s} Y_n(s,t)\, dt &= \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, t + s_{i+1} - s)\, dt \\
&\quad - \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1} - s)\, dt + \int_0^{T_n+s} \frac{e^{-t} t^2}{2}\, dt \\
&= e^{s_{i+1}-s} \int_0^{T_n+s_{i+1}} e^{-t} S_n(s_{i+1}, t)\, dt - e^{s_{i+1}-s} \int_0^{s_{i+1}-s} e^{-t} S_n(s_{i+1}, t)\, dt \\
&\quad - \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1} - s)\, dt + \int_0^{T_n+s} \frac{e^{-t} t^2}{2}\, dt \\
&= \int_0^{T_n+s_{i+1}} Y_n(s_{i+1}, t)\, dt + (e^{s_{i+1}-s} - 1) \int_0^{T_n+s_{i+1}} e^{-t} S_n(s_{i+1}, t)\, dt \\
&\quad - e^{s_{i+1}-s} \int_0^{s_{i+1}-s} e^{-t} S_n(s_{i+1}, t)\, dt \\
&\quad - \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1} - s)\, dt - \int_{T_n+s}^{T_n+s_{i+1}} \frac{e^{-t} t^2}{2}\, dt. \tag{4.10}
\end{aligned}
$$

We must bound the last four terms on the right-hand side of (4.10). Observe that

$e^{s_{i+1}-s_i} - 1 \leq 3/(\log n)^2$ for sufficiently large n, so

$$\left| (e^{s_{i+1}-s} - 1) \int_0^{T_n+s_{i+1}} e^{-t} S_n(s_{i+1}, t) \, dt \right|$$

$$\leq \frac{3}{(\log n)^2} \left(1 + \left| \int_0^{T_n+s_{i+1}} Y_n(s_{i+1}, t) \, dt \right| \right)$$

$$\leq \frac{3}{(\log n)^2} \left(1 + \int_0^{T_n+s_{i+1}} |Y_n(s_{i+1}, t) - Y(s_{i+1}, t)| \, dt \right.$$

$$\left. + \left| \int_{T_n+s_{i+1}}^{\infty} Y(s_{i+1}, t) \, dt \right| + \left| \int_0^{\infty} Y(s_{i+1}, t) \, dt \right| \right)$$

$$\leq \frac{3}{(\log n)^2} \left(1 + \sup_{0 \leq s \leq T} \int_0^{T_n+s} |Y_n(s, t) - Y(s, t)| \, dt \right.$$

$$\left. + \sup_{0 \leq s \leq T} \left| \int_{T_n+s}^{\infty} Y(s, t) \, dt \right| + \sup_{0 \leq s \leq T} |L(s)| + \sup_{0 \leq s \leq T} |J(s)| \right). \quad (4.11)$$

Also, $e^{s_{i+1}-s} \leq 2$ for sufficiently large n, which means

$$\left| e^{s_{i+1}-s} \int_0^{s_{i+1}-s} e^{-t} S_n(s_{i+1}, t) \, dt \right| \leq 2(s_{i+1} - s_i) \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)|$$

$$\leq \frac{4}{(\log n)^2} \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)|. \quad (4.12)$$

Likewise,

$$\left| \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1} - s) \, dt \right| \leq \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)|. \quad (4.13)$$

Finally,

$$\left| \int_{T_n+s}^{T_n+s_{i+1}} \frac{e^{-t} t^2}{2} \, dt \right| \leq \frac{1}{(\log n)^2} \quad (4.14)$$

for sufficiently large n. Combining (4.11), (4.12), (4.13), and (4.14) with (4.10), we get that for sufficiently large n,

$$\sup_{0 \leq i \leq m-1} \sup_{s \in [s_i, s_{i+1}]} \left| \int_0^{T_n+s} Y_n(s, t) \, dt - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1}, t) \, dt \right|$$

$$\leq \frac{4}{(\log n)^2} + 2 \sup_{0 \leq i \leq m-1} \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)|$$

$$+ \frac{3}{(\log n)^2} \left(\sup_{0 \leq s \leq T} \int_0^{T_n+s} |Y_n(s, t) - Y(s, t)| \, dt \right.$$

$$\left. + \sup_{0 \leq s \leq T} \left| \int_{T_n+s}^{\infty} Y(s, t) \, dt \right| + \sup_{0 \leq s \leq T} |L(s)| + \sup_{0 \leq s \leq T} |J(s)| \right). \quad (4.15)$$

Note that $\sup_{0 \leq s \leq T} |J(s)|$ is almost surely finite by properties of the Poisson process Ψ, and $\sup_{0 \leq s \leq T} |L(s)|$ is almost surely finite by Proposition 4.1. Combining these observations with Lemmas 4.4 and 4.7, we see that the probability that the third term on the right-hand side of (4.15) is less than ε tends to zero as $n \to \infty$. Clearly $4/(\log n)^2 < \varepsilon$ for sufficiently large n. To control the second term, apply Lemma 3.3 with $\delta_n = s_{i+1} - s_i$ and $\theta = \varepsilon^3$ to get

$$P\left(\sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)| \mathbf{1}_{G_i} > \varepsilon \right) \leq P\left(\sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1}, t)| > \varepsilon \Big| G_i \right) \leq 4(s_{i+1} - s_i)\varepsilon,$$

and thus

$$P\left(2 \sup_{0 \leq i \leq m-1} \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1},t)|\mathbf{1}_{G_i} > 2\varepsilon\right) \leq 4\varepsilon \sum_{i=0}^{m-1} (s_{i+1}-s_i) = 4T\varepsilon.$$

Combining these bounds with (4.15) gives the result. □

Lemma 4.9. *We have*

$$\lim_{n \to \infty} P\left(\sup_{0 \leq i \leq m-1} \sup_{s \in [s_i, s_{i+1}]} \left| \int_0^{T_n+s} Y_n(s,t)\, dt - J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} \right.\right.$$
$$\left.\left. - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\, dt \right| \mathbf{1}_{H_i} \geq 5\varepsilon\right) = 0.$$

Proof. Note that (4.10) still holds in this setting. Consequently, if $s_i \leq s \leq s_{i+1}$, then on the event H_i, we have

$$\left| \int_0^{T_n+s} Y_n(s,t)\, dt - J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\, dt \right|$$

$$\leq \left| (e^{s_{i+1}-s} - 1) \int_0^{T_n+s_{i+1}} e^{-t} S_n(s_{i+1},t)\, dt \right| + \left| e^{s_{i+1}-s} \int_0^{s_{i+1}-s} e^{-t} S_n(s_{i+1},t)\, dt \right|$$

$$+ \left| \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1}-s)\, dt + J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} \right| + \left| \int_{T_n+s}^{T_n+s_{i+1}} \frac{e^{-t} t^2}{2}\, dt \right|.$$
(4.16)

The bounds (4.11), (4.12), and (4.14) also hold. In place of (4.13), observe that if $s_i \leq s \leq s_{i+1}$, then on the event H_i,

$$\left| \int_0^{T_n+s} e^{-t} S_n(s_{i+1}, s_{i+1}-s)\, dt + J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} \right|$$

$$= \left| S_n(s_{i+1}, s_{i+1}-s)\left(1 - \int_{T_n+s}^{\infty} e^{-t}\, dt\right) + J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} \right|$$

$$\leq \left| S_n(s_{i+1}, s_{i+1}-s) + J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} \right| + \left| S_n(s_{i+1}, s_{i+1}-s)e^{-(T_n+s)} \right|$$

$$\leq \sup_{0 \leq t \leq s_{i+1}-s_i} \left| S_n(s_{i+1},t) + J(u_{j(i)})\mathbf{1}_{\{t \geq s_{i+1}-u_{j(i)}\}} \right| + \frac{1}{(\log n)^2} \sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1},t)|.$$
(4.17)

Also, on the event H_i,

$$\sup_{0 \leq t \leq s_{i+1}-s_i} |S_n(s_{i+1},t)| \leq |J(u_{j(i)})| + \sup_{0 \leq t \leq s_{i+1}-s_i} \left| S_n(s_{i+1},t) - J(u_{j(i)})\mathbf{1}_{\{t \geq s_{i+1}-u_{j(i)}\}} \right|.$$
(4.18)

Thus, on the event H_i, by (4.16), (4.17), (4.18), (4.11), (4.12), and (4.14), we have

$$\sup_{0 \leq i \leq m-1} \sup_{s \in [s_i, s_{i+1}]} \left| \int_0^{T_n+s} Y_n(s,t)\, dt - J(u_{j(i)})\mathbf{1}_{\{s \leq u_{j(i)}\}} - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\, dt \right|$$

$$\leq \frac{4}{(\log n)^2} + \frac{5}{(\log n)^2} \sup_{0 \leq s \leq T} |J(s)|$$

$$+ \left(1 + \frac{5}{(\log n)^2}\right) \sup_{0 \leq i \leq m-1} \sup_{0 \leq t \leq s_{i+1}-s_i} \left| S_n(s_{i+1},t) + J(u_{j(i)})\mathbf{1}_{\{t \geq s_{i+1}-u_{j(i)}\}} \right|$$

$$+ \frac{3}{(\log n)^2}\left(\sup_{0 \leq s \leq T} \int_0^{T_n+s} |Y_n(s,t) - Y(s,t)|\, dt \right.$$

$$\left. + \sup_{0 \leq s \leq T} \left| \int_{T_n+s}^{\infty} Y(s,t)\, dt \right| + \sup_{0 \leq s \leq T} |L(s)| + \sup_{0 \leq s \leq T} |J(s)| \right). \qquad (4.19)$$

By Lemmas 4.4 and 4.7, and the fact that $\sup_{0\le s\le T}|L(s)|$ and $\sup_{0\le s\le T}|J(s)|$ are finite almost surely, the probability that the first, second, or fourth term on the right-hand side of (4.19) is greater than ε tends to zero as $n\to\infty$. To bound the third term, we apply Lemma 3.3. Observe that the conditional distribution of

$$\sup_{0\le t\le s_{i+1}-s_i}\left|S_n(s_{i+1},t)+J(u_{j(i)})\mathbf{1}_{\{t\ge s_{i+1}-u_{j(i)}\}}\right|$$

given H_i is the same as the conditional distribution of $\sup_{0\le t\le\delta_n}|S_n(s,t)|$ given $A(\theta)$ in Lemma 3.3 if we take $\delta_n=s_{i+1}-s_i$ and $\theta=\varepsilon^3$. Because

$$P(H_i)\le(s_{i+1}-s_i)\int_{\varepsilon^3}^\infty x^{-2}\,dx\le\frac{2}{\varepsilon^3(\log n)^2}, \tag{4.20}$$

it follows from Lemma 3.3 that

$$P\left(2\sup_{0\le i\le m-1}\sup_{0\le t\le s_{i+1}-s_i}\left|S_n(s_{i+1},t)+J(u_{j(i)})\mathbf{1}_{\{t\ge s_{i+1}-u_{j(i)}\}}\right|\mathbf{1}_{H_i}>2\varepsilon\right)$$

$$\le\sum_{i=0}^{m-1}P(H_i)\cdot4(s_{i+1}-s_i)\varepsilon\le\frac{8T}{\varepsilon^2(\log n)^2},$$

which tends to zero as $n\to\infty$. The result follows. $\qquad\square$

Lemma 4.10. *If $s\ge0$ and $h\ge0$ with $0\le s+h\le T$, then*

$$\int_0^{T_n+s+h}X_n(s+h,t)\,dt\le\int_0^{T_n+s}X_n(s,t)\,dt+h\log n \tag{4.21}$$

and for sufficiently large n,

$$\int_0^{T_n+s+h}X_n(s+h,t)\,dt$$

$$\ge\int_0^{T_n+s}X_n(s,t)\,dt-\frac{2h}{\log n}-\frac{(\log n)(T_n+s)}{n}\left(n-N_n\left(\frac{s+h}{\log n},\frac{h}{\log n}\right)+1\right). \tag{4.22}$$

Proof. We have

$$\int_0^{T_n+s+h}X_n(s+h,t)\,dt=\frac{\log n}{n}\int_0^{T_n+s+h}\tilde{N}_n\left(\frac{s+h}{\log n},\frac{t}{\log n}\right)-ne^{-t}-\frac{nte^{-t}\log\log n}{\log n}\,dt$$

$$=\frac{\log n}{n}\int_0^{T_n+s}\tilde{N}_n\left(\frac{s}{\log n},\frac{t}{\log n}\right)-ne^{-t}-\frac{nte^{-t}\log\log n}{\log n}\,dt$$

$$-\frac{\log n}{n}\int_{T_n+s}^{T_n+s+h}\left(ne^{-t}+\frac{nte^{-t}\log\log n}{\log n}\right)\,dt$$

$$+\frac{\log n}{n}\int_0^{T_n+s+h}\tilde{N}_n\left(\frac{s+h}{\log n},\frac{t}{\log n}\right)\,dt$$

$$-\frac{\log n}{n}\int_0^{T_n+s}\tilde{N}_n\left(\frac{s}{\log n},\frac{t}{\log n}\right)\,dt$$

$$=\int_0^{T_n+s}X_n(s,t)\,dt-\frac{\log n}{n}\int_{T_n+s}^{T_n+s+h}\left(ne^{-t}+\frac{nte^{-t}\log\log n}{\log n}\right)\,dt$$

$$+\frac{\log n}{n}\int_0^h\tilde{N}_n\left(\frac{s+h}{\log n},\frac{t}{\log n}\right)\,dt$$

$$+\frac{\log n}{n}\int_0^{T_n+s}\tilde{N}_n\left(\frac{s+h}{\log n},\frac{t+h}{\log n}\right)-\tilde{N}_n\left(\frac{s}{\log n},\frac{t}{\log n}\right)\,dt. \tag{4.23}$$

Because $N_n(s,t) \leq n$ for all $s, t \geq 0$, the integral in the third term on the right-hand side of (4.23) is bounded by hn, so the term is bounded by $h \log n$. Also, for all $s, t, h \geq 0$, we have $N(s+h, t+h) \leq N(s,t)$ because the number of individuals in the population at time $((s+h) - (t+h))- = (s-t)-$ with descendants alive in the population at time $s+h$ is less than or equal to the number of individuals in the population at time $(s-t)-$ with descendants alive in the population at time s. Therefore, the fourth term in the right-hand side of (4.23) is nonpositive. Equation (4.21) follows.

Next, we claim that for all $s, t, h \geq 0$, we have

$$N_n(s,t) - N_n(s+h, t+h) \leq n - N_n(s+h, h).$$

To see this, note that $N_n(s,t) - N_n(s+h, t+h)$ is the number of individuals in the population at time $(s-t)-$ with descendants alive in the population at time s but not at time $s+h$. This is at most the number of individuals in the population at time s that do not have descendants alive in the population at time $s+h$, which is at most $n - N_n(s+h, h)$. Therefore, $\tilde{N}_n(s,t) - \tilde{N}_n(s+h, t+h) \leq n - N_n(s+h, h) + 1$. The result (4.22) follows from this observation and (4.23) because

$$\frac{\log n}{n} \int_{T_n+s}^{T_n+s+h} \left(ne^{-t} + \frac{nte^{-t} \log \log n}{\log n} \right) dt \leq (h \log n) \sup_{t \geq T_n} \left(e^{-t} + \frac{te^{-t} \log \log n}{\log n} \right) \leq \frac{2h}{\log n}$$

for sufficiently large n. $\qquad \square$

Proof of Proposition 4.3. By Lemmas 4.4, 4.6, and 4.7, it suffices to show that

$$\sup_{0 \leq s \leq T} \left| \int_0^{T_n+s} X_n(s,t)\, dt + J(s) - \int_0^{T_n+s} Y_n(s,t)\, dt \right| \to_p 0. \tag{4.24}$$

By Lemmas 3.6, 3.11, and 3.13, we have

$$\limsup_{n \to \infty} P\left(\int_0^{T_n+s_i} |X_n(s_i, t) - Y_n(s_i, t)|\, dt \geq \frac{5(T_n + s_i)}{(\log n)^{1/16}} \text{ for some } i = 0, 1, \ldots, m \right)$$

$$\leq \limsup_{n \to \infty} \left(P(A_n^c) + P(B_n^c) \right.$$

$$\left. + (1 + T(\log n)^2) \left(\frac{(T_n + T)(\log n)^{3/2}}{n^{1/2}\varepsilon_n} + 4(\log n)^2 (T_n + T)\varepsilon_n \right) \right) = 0.$$

Therefore, for all $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P\left(\sup_{0 \leq i \leq m} \left| \int_0^{T_n+s_i} X_n(s_i, t)\, dt - \int_0^{T_n+s_i} Y_n(s_i, t)\, dt \right| \geq \varepsilon \right) = 0. \tag{4.25}$$

We first consider the case in which G_i occurs. If $s \in [s_i, s_{i+1}]$, then by (4.21),

$$\int_0^{T_n+s} X_n(s,t) \leq \int_0^{T_n+s_i} X_n(s_i, t)\, dt + (s - s_i) \log n, \tag{4.26}$$

and therefore

$$\int_0^{T_n+s} X_n(s,t) \leq \int_0^{T_n+s} Y_n(s,t)\, dt + \left| \int_0^{T_n+s_i} X_n(s_i, t)\, dt - \int_0^{T_n+s_i} Y_n(s_i, t)\, dt \right|$$

$$+ \left| \int_0^{T_n+s_i} Y_n(s_i, t)\, dt - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1}, t)\, dt \right|$$

$$+ \left| \int_0^{T_n+s_{i+1}} Y_n(s_{i+1}, t)\, dt - \int_0^{T_n+s} Y_n(s,t)\, dt \right| + (s - s_i) \log n. \tag{4.27}$$

Likewise,

$$\int_0^{T_n+s} X_n(s,t) \geq \int_0^{T_n+s_{i+1}} X_n(s_{i+1},t)\,dt - (s_{i+1}-s)\log n, \tag{4.28}$$

and so

$$\int_0^{T_n+s} X_n(s,t) \geq \int_0^{T_n+s} Y_n(s,t)\,dt - \left| \int_0^{T_n+s_{i+1}} X_n(s_{i+1},t)\,dt - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|$$
$$- \left| \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt - \int_0^{T_n+s} Y_n(s,t)\,dt \right| - (s_{i+1}-s)\log n. \tag{4.29}$$

Since $s_{i+1}-s_i \leq 2/(\log n)^2$ for $i=0,1,\ldots,m-1$, it follows that

$$\sup_{0\leq i\leq m-1}\sup_{s\in[s_i,s_{i+1}]} \left| \int_0^{T+s} X_n(s,t)\,dt - \int_0^{T+s} Y_n(s,t)\,dt \right|$$
$$\leq \frac{2}{\log n} + \sup_{0\leq i\leq m} \left| \int_0^{T_n+s_i} X_n(s_i,t)\,dt - \int_0^{T_n+s_i} Y_n(s_i,t)\,dt \right|$$
$$+ 2 \sup_{0\leq i\leq m-1}\sup_{s\in[s_i,s_{i+1}]} \left| \int_0^{T_n+s} Y_n(s,t)\,dt - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|.$$

Clearly $2/\log n < \varepsilon$ for sufficiently large n. Therefore, by (4.25) and Lemma 4.8,

$$\limsup_{n\to\infty} P\left(\sup_{0\leq i\leq m-1}\sup_{s\in[s_i,s_{i+1}]} \left| \int_0^{T+s} X_n(s,t)\,dt - \int_0^{T+s} Y_n(s,t)\,dt \right| \mathbf{1}_{G_i} > 10\varepsilon \right) \leq 4T\varepsilon. \tag{4.30}$$

We next consider the case in which H_i occurs. Recall that $u_{j(i)} \in (s_i, s_{i+1})$ on H_i. Let

$$R_i = \lim_{s\uparrow u_{j(i)}} N_n\left(\frac{s}{\log n}, \frac{s-s_i}{\log n} \right).$$

That is, R_i is the number of individuals in the population at time $(s_i/\log n)-$ with descendants in the population at least until time $(u_{j(i)}/\log n)-$. Note that

$$N_n(s/\log n, (s-s_i)/\log n) \geq R_i$$

for all $s\in[s_i, u_{j(i)})$. Therefore, by (4.22), for $s\in[s_i, u_{j(i)})$ we have

$$\int_0^{T_n+s} X_n(s,t)\,dt \geq \int_0^{T_n+s_i} X_n(s_i,t)\,dt - \frac{2(s-s_i)}{\log n} - \frac{(\log n)(T_n+T)(n-R_i+1)}{n}.$$

Combining this result with (4.26) and the fact that $s_{i+1}-s_i \leq 2/(\log n)^2$ gives

$$\left| \int_0^{T_n+s} X_n(s,t)\,dt - \int_0^{T_n+s_i} X_n(s_i,t)\,dt \right| \leq \frac{2}{\log n} + \frac{4}{(\log n)^3} + \frac{(\log n)(T_n+T)(n-R_i+1)}{n}.$$

It follows that if $s\in[s_i, u_{j(i)})$, then

$$\left| \int_0^{T_n+s} X_n(s,t)\,dt - \int_0^{T_n+s} Y_n(s,t)\,dt \right|$$
$$\leq \frac{2}{\log n} + \frac{4}{(\log n)^3} + \frac{(\log n)(T_n+T)(n-R_i+1)}{n}$$
$$+ \left| \int_0^{T_n+s_i} X_n(s_i,t)\,dt - \int_0^{T_n+s_i} Y_n(s_i,t)\,dt \right|$$
$$+ \left| \int_0^{T_n+s_i} Y_n(s_i,t)\,dt - J(u_{j(i)}) - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|$$
$$+ \left| \int_0^{T_n+s} Y_n(s,t)\,dt - J(u_{j(i)}) - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|. \tag{4.31}$$

Likewise, let

$$S_i = N_n\left(\frac{s_{i+1}}{\log n}, \frac{s_{i+1} - u_{j(i)}}{\log n} -\right),$$

and note that $N_n(s_{i+1}/\log n, (s_{i+1} - s)/\log n) \geq S_i$ for all $s \in (u_{j(i)}, s_{i+1}]$. Therefore, if $s \in (u_{j(i)}, s_{i+1}]$, then (4.22) gives

$$\int_0^{T_n+s} X_n(s,t)\,dt \leq \int_0^{T_n+s_{i+1}} X_n(s_{i+1},t)\,dt + \frac{2(s_{i+1}-s)}{\log n} + \frac{(\log n)(T_n+T)(n-S_i+1)}{n}. \tag{4.32}$$

Because $X_n(u_{j(i)}, t) = \lim_{s\downarrow u_{j(i)}} X_n(s,t)$ for Lebesgue almost all $t > 0$, equation (4.32) holds for all $s \in [u_{j(i)}, s_{i+1}]$. Combining this result with (4.28) gives

$$\left| \int_0^{T_n+s} X_n(s,t)\,dt - \int_0^{T_n+s_{i+1}} X_n(s_{i+1},t)\,dt \right|$$
$$\leq \frac{2}{\log n} + \frac{4}{(\log n)^3} + \frac{(\log n)(T_n+T)(n-S_i+1)}{n}.$$

Therefore, if $s \in [u_{j(i)}, s_{i+1}]$, then

$$\left| \int_0^{T_n+s} X_n(s,t)\,dt + J(s) - \int_0^{T_n+s} Y_n(s,t)\,dt \right|$$
$$\leq \frac{2}{\log n} + \frac{4}{(\log n)^3} + \frac{(\log n)(T_n+T)(n-S_i+1)}{n}$$
$$+ \left| \int_0^{T_n+s_{i+1}} X_n(s_{i+1},t)\,dt - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|$$
$$+ \left| \int_0^{T_n+s} Y_n(s,t) - J(s) - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right|. \tag{4.33}$$

By (4.31) and (4.33), and the fact that $J(s) = 0$ if $s \in [s_i, u_{j(i)}) \cup (u_{j(i)}, s_{i+1}]$,

$$\sup_{0\leq i\leq m-1} \sup_{s\in[s_i,s_{i+1}]} \left| \int_0^{T+s} X_n(s,t)\,dt + J(s) - \int_0^{T_n+s} Y_n(s,t)\,dt \right| \mathbf{1}_{H_i}$$
$$\leq \frac{2}{\log n} + \frac{4}{(\log n)^3} + \sup_{0\leq i\leq m-1} \frac{(\log n)(T_n+T)(2n-R_i-S_i+2)\mathbf{1}_{H_i}}{n}$$
$$+ \sup_{0\leq i\leq m} \left| \int_0^{T_n+s_i} X_n(s_i,t)\,dt - \int_0^{T_n+s_i} Y_n(s_i,t)\,dt \right|$$
$$+ 2 \sup_{0\leq i\leq m-1} \sup_{s\in[s_i,s_{i+1}]} \left| \int_0^{T_n+s} Y_n(s,t) - J(u_{j(i)})\mathbf{1}_{\{s\leq u_{j(i)}\}} \right.$$
$$\left. - \int_0^{T_n+s_{i+1}} Y_n(s_{i+1},t)\,dt \right| \mathbf{1}_{H_i}. \tag{4.34}$$

We claim that the conditional distributions of $n - R_i$ and $n - S_i$ given H_i are each stochastically dominated by the distribution of $n - N_n(0, 2/(\log n)^3)$, which is the number of blocks lost by time $2/(\log n)^3$ in a Bolthausen-Sznitman coalescent started with n blocks. To see this, note that $n - S_i$ is the number of blocks that are lost by time $(s_{i+1} - u_{j(i)})/\log n \leq 2/(\log n)^3$ in a Bolthausen-Sznitman coalescent. Likewise $n - R_i$ is the number of blocks lost by time $(u_{j(i)} - s_i)/(\log n) \leq 2/(\log n)^3$ in a Bolthausen-Sznitman coalescent started with n lineages, if we disallow the instantaneous merger caused by the birth event at time $u_{j(i)}$. Because H_i requires that exactly one of the u_k falls in (s_i, s_{i+1}), the effect of conditioning on H_i is the same as suppressing all mergers

in which each lineage participates with probability greater than ε^3. This can only reduce the number of blocks lost. Thus, we have

$$E[n - R_i | H_i] \leq \frac{2\eta_n}{(\log n)^3} \leq \frac{2n}{(\log n)^2}$$

by (2.2). By the same argument, $E[n - S_i | H_i] \leq 2n/(\log n)^2$. Thus, using (4.20) and the fact that $m \leq T(\log n)^2$,

$$E\left[\sup_{0 \leq i \leq m-1} \frac{(\log n)(T_n + T)(2n - R_i - S_i + 2)\mathbf{1}_{H_i}}{n} \right]$$

$$\leq \frac{(\log n)(T_n + T)}{n} \sum_{i=0}^{m-1} P(H_i) \left(\frac{4n}{(\log n)^2} + 2 \right) \leq \frac{2(\log n)(T_n + T)m}{\varepsilon^3 n(\log n)^2} \left(\frac{4n}{(\log n)^2} + 2 \right) \to 0$$

as $n \to \infty$. Hence, by Markov's Inequality,

$$\frac{2}{\log n} + \frac{4}{(\log n)^3} + \sup_{0 \leq i \leq m-1} \frac{(\log n)(T_n + T)(2n - R_i - S_i + 2)\mathbf{1}_{H_i}}{n} \to_p 0.$$

Combining this result with (4.34), (4.25), and Lemma 4.9, we get

$$\limsup_{n \to \infty} P\left(\sup_{0 \leq i \leq m-1} \sup_{s \in [s_i, s_{i+1}]} \left| \int_0^{T+s} X_n(s,t)\, dt + J(s) - \int_0^{T+s} Y_n(s,t)\, dt \right| \mathbf{1}_{H_i} > 12\varepsilon \right) = 0.$$

(4.35)

The result (4.24) follows from (4.30), (4.35), and the fact that with probability tending to 1 as $n \to \infty$, we have $G_i \cup H_i$ for $i = 0, 1, \ldots, m-1$ by (4.9). □

References

[1] Berestycki, J., Berestycki, N., and Limic, V.: The Λ-coalescent speed of coming down from infinity. *Ann. Probab.* **38**, (2010), 207–233. MR-2599198

[2] Berestycki, J., Berestycki, N., and Schweinsberg, J.: Small-time behavior of beta coalescents. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, (2008), 214–238. MR-2446321

[3] Berestycki, J., Berestycki, N., and Schweinsberg, J.: The genealogy of branching Brownian motion with absorption, arXiv:1001.2337

[4] Bertoin, J.: *Lévy Processes.* Cambridge University Press, Cambridge (1996). MR-1406564

[5] Bertoin, J. and Le Gall, J.-F.: The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields* **117**, (2000), 249–266. MR-1771663

[6] Bolthausen, E. and Sznitman, A.-S.: On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197**, (2008), 247–276. MR-1652734

[7] Bovier, A. and Kurkova, I.: Much ado about Derrida's GREM. *Spin Glasses*, Lecture Notes in Math., 1900, Springer, Berlin (2007), 81–115. MR-2309599

[8] Brunet, E., Derrida, B., Mueller, A. H., and Munier, S.: Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76**, (2007), 041104. MR-2365627

[9] Davis, M. H. A.: Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46**, (1984), 353–388. MR-0790622

[10] Davis, M. H. A.: *Markov Models and Optimization.* Chapman and Hall, London (1993). MR-1283589

[11] Depperschmidt, A., Greven, A., and Pfaffelhuber, P.: Tree-valued Fleming-Viot dyamics with mutation and selection, arXiv:1101.0759

[12] Delmas, J.-F., Dhersin, J.-S., and Siri-Jegousse, A.: On the two oldest families for the Wright-Fisher process. *Electron. J. Probab.* **15**, (2010), 776–800. MR-2653183

[13] Drmota, M., Iksanov A., Möhle, M., and Rösler, U.: Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stochastic Process. Appl.* **117**, (2007), 1404–1421. MR-2353033

[14] Donnelly, P. and Kurtz, T. G.: Particle representations for measure-valued population models. *Ann. Probab.* **27**, (1999), 166–205. MR-1681126

[15] Ethier, S. N. and Kurtz, T. G.: *Markov Processes: Characterization and Convergence*. Wiley, New York (1986). MR-0838085

[16] Evans, S. N., Pitman, J., and Winter, A.: Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Relat. Fields* **134**, (2006), 81–126. MR-2221786

[17] Evans, S. N. and Ralph, P. L.: Dynamics of the time to the most recent common ancestor in a large branching population. *Ann. Appl. Probab.* **20**, (2010), 1–25. MR-2582640

[18] Goldschmidt, C. and Martin, J. B.: Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10**, (2005), 718–745. MR-2164028

[19] Greven, A., Pfaffelhuber, P., and Winter, A.: Tree-valued resampling dynamics: Martingale problems and applications, arXiv:0806.2224

[20] Kingman, J. F. C.: The coalescent. *Stochastic Process. Appl.* **13**, (1982), 235–248. MR-0671034

[21] Klebaner, F. C.: *Introduction to Stochastic Calculus with Applications*. 2nd ed. Imperial College Press, London (2005). MR-2160228

[22] Moran, P. A. P.: Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54**, (1958), 60–71. MR-0127989

[23] Pfaffelhuber, P. and Wakolbinger, A.: The process of most recent common ancestors in an evolving coalescent. *Stochastic Process. Appl.* **116**, (2006), 1836–1859. MR-2307061

[24] Pfaffelhuber, P., Wakolbinger, A., and Weisshaupt, H.: The tree length of an evolving coalescent. *Probab. Theory Related Fields* **151**, (2009), 529–557. MR-2851692

[25] Pitman, J.: Coalescents with multiple collisions. *Ann. Probab.* **27**, (1999), 1870–1902. MR-1742892

[26] Pitman, J.: *Combinatorial Stochastic Processes*. Lecture Notes in Math., 1875. Springer, Berlin (2006). MR-2245368

[27] Protter, P.: *Stochastic Integration and Differential Equations: A New Approach*. Springer, Berlin (1990). MR-1037262

[28] Sagitov, S.: The general coalescent with asynchronous mergers of ancestral lines. *J.* Appl. Probab. **36**, (1999), 1116-1125. MR-1742154

[29] Samorodnitsky, G. and Taqqu, M.S.: *S*table non-Gaussian Random Processes: Stochastic models with infinite variance. Chapman and Hall, New York (1994). MR-1280932

[30] Sato, K.: *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge (1999). MR-1739520

[31] Schweinsberg, J.: Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.* **106**, (2003), 107–139. MR-1983046

[32] Simon, D. and Derrida, B.: Evolution of the most recent common ancestor of a population with no selection. *J. Stat. Mech. Theory Exp.*, (2006), P05002.