External branch lengths of Λ-coalescents without a dust component

Christina S. Diehl* and Görtz Kersting*

November 20, 2018

Abstract

Λ-coalescents model genealogies of samples of individuals from a large population, and the individuals' time durations up to some common ancestor are given by the external branch lengths of the family tree. We consider typical external branches under the minimal assumption that the coalescent has no dust component, and maximal external branches under further regularity assumptions. As it turns out, the crucial characteristic is the coalescent’s rate of decrease $\mu(b)$, $b \geq 2$. The magnitude of a typical external branch is asymptotically given by $n/\mu(n)$, where $n$ denotes the sample size. This result and also asymptotic independence of several typical external lengths hold in full generality, while convergence in distribution of the scaled external lengths requires that $\mu(n)$ is regularly varying at infinity.

For the maximal lengths, we distinguish two cases. First, we analyze a class of Λ-coalescents coming down from infinity and with regularly varying $\mu$. Here the scaled external lengths behave as the maximal values of $n$ i.i.d. random variables, and their limit is captured by a Poisson point process on the positive real line. Second, we turn to the Bolthausen-Sznitman coalescent, where the picture changes. Now the limiting behavior of the normalized external lengths is given by a Cox point process, which can be expressed by a randomly shifted Poisson point process.

AMS 2010 subject classification: 60J75 (primary), 60F05, 60J27, 92D25

Keywords: Λ-coalescent, dustless coalescent, Bolthausen-Sznitman coalescent, Beta-coalescent, Kingman’s coalescent, external branch lengths, Poisson point process, Cox point process, weak limit law

*Institut für Mathematik, Goethe Universität, Frankfurt am Main, Germany
diehl@math.uni-frankfurt.de, kersting@math.uni-frankfurt.de

Work partially supported by the DFG Priority Programme SPP 1590 “Probabilistic Structures in Evolution”
1 Introduction and main results

In population genetics, family trees stemming from a sample out of a big population are modeled by coalescents nowadays. The prominent Kingman coalescent [21] found widespread applications in biology. More recently, the Bolthausen-Sznitman coalescent, originating from statistical mechanics [3], gained in importance in order to analyze genealogies of populations undergoing selection [4, 7, 25, 31]. Unlike Kingman’s coalescent, the Bolthausen-Sznitman coalescent allows multiple mergers. The larger class of Beta-coalescents found interest, e.g., in the study of marine species [33, 26]. All these instances are covered by the notion of Λ-coalescents as introduced by Pitman [27] and Sagitov [29] in 1999. Today general properties of this broad class become more transparent [19, 11].

In this paper, we deal with the lengths of external branches of Λ-coalescents under the minimal assumption that the coalescent has no dust component, which applies to all cases mentioned above. We shall treat external branches of typical and, under additional regularity assumptions, of maximal length. For the total external length see the publications [23, 16, 6, 17, 10].

Λ-coalescents are Markov processes \((Π(t), t \geq 0)\) taking values in the set of partitions of \(\mathbb{N}\), where \(\Lambda\) denotes a non-vanishing finite measure on the unit interval \([0,1]\). Its restrictions \((Π_n(t), t \geq 0)\) to the sets \(\{1,\ldots,n\}\) are called \(n\)-coalescents. They are continuous-time Markov chains characterized by the following dynamics: Given the event that \(Π_n(t)\) is a partition consisting of \(b \geq 2\) blocks, \(k\) specified blocks merge at rate

\[ \lambda_{b,k} := \int_{[0,1]} p^k (1 - p)^{b-k} \frac{\Lambda(dp)}{p^2}, \quad 2 \leq k \leq b, \]

to a single one. In this paper, the crucial characteristic of Λ-coalescents is their rate of decrease \(\mu = (\mu(b))_{b \geq 2}\) defined as

\[ \mu(b) := \sum_{k=2}^{b} (k-1) \binom{b}{k} \lambda_{b,k}, \quad b \geq 2. \]

This notion captures that a merger of \(k\) blocks corresponds to a decline of \(k-1\) blocks. Its importance became also apparent from other publications [30, 22, 11]. In particular, the assumption of absence of a dust component may be expressed in this terms. Originally characterized by the condition

\[ \int_{[0,1]} \frac{\Lambda(dp)}{p} = \infty, \]

(see [27]), it can be equivalently specified by the requirement

\[ \frac{\mu(n)}{n} \to \infty \]

as \(n \to \infty\) (see Lemma 1 (iii) of [11]).
An $n$-coalescent can be thought of as a random rooted tree with $n$ labeled leaves representing the individuals of a sample. Its branches specify ancestral lineages of the individuals or their ancestors. The branch lengths give the time spans until the occurrence of new common ancestors. Branches ending in a leaf are called external branches. If mutations under the infinite sites model [20] are added in these considerations, the importance of external branches is revealed. This is due to the fact that mutations on external branches only affect a single individual of the sample. Longer external branches result thereby in an excess of singleton polymorphisms [35] and are known to be a characteristic for trees with multiple mergers [12]; e.g., external branch lengths have been used to discriminate between different coalescents in the context of HIV trees [36] (see also [34]).

Now we turn to the main results of this paper. For $1 \leq i \leq n$, the length of the external branch ending in leaf $i$ within an $n$-coalescent is defined as

$$
T^n_i := \inf \{ t \geq 0 : \{i\} \notin \Pi_n(t) \}.
$$

In the first theorem, we consider the length $T^n$ of a randomly chosen external branch. Based on the exchangeability, $T^n$ is equal in distribution to $T^n_i$ for $1 \leq i \leq n$. The result clarifies the magnitude of $T^n$ in full generality.

**Theorem 1.1.** For a $\Lambda$-coalescent without a dust component, we have for $t \geq 0$,

$$
e^{-2t} + o(1) \leq \mathbb{P} \left( \frac{\mu(n)}{n} T^n > t \right) \leq \frac{1}{1 + t} + o(1)
$$

as $n \to \infty$.

Among others, this theorem excludes the possibility that $T^n$ converges to a positive constant in probability. In [18] the order of $T^n$ was interpreted as the duration of a generation, namely the time at which a specific lineage out of the $n$ present ones takes part in a merging event. In that paper, only Beta($2 - \alpha, \alpha$)-coalescents with $1 < \alpha < 2$ have been considered, and the duration was given as $n^{1 - \alpha}$. Our theorem shows that for this heuristics $n/\mu(n)$ is a suitable choice in general.

Asymptotic independence of the external branch lengths holds as well in full generality for dustless coalescents.

**Theorem 1.2.** For a $\Lambda$-coalescent without a dust component, for fixed $k \in \mathbb{N}$ and for any sequence of numbers $t^n_1, \ldots, t^n_k \geq 0$, $n \geq 2$, we have

$$
\mathbb{P} \left( T^n_1 \leq t^n_1, \ldots, T^n_k \leq t^n_k \right) = \mathbb{P} \left( T^n_1 \leq t^n_1 \right) \cdots \mathbb{P} \left( T^n_k \leq t^n_k \right) + o(1)
$$

as $n \to \infty$. 
In the light of the waiting times, which the different external branches have in common, this might be an unexpected result. We point out that for each coalescent with a dust component, Möhle showed that this asymptotic independence does not hold (see equation (11) of [23]).

In order to achieve convergence in distribution of the scaled lengths, stronger assumptions are required on the rate of decrease, namely that $\mu$ is a regularly varying sequence. A characterization of this property is given in Proposition 2.2 below. Let $\delta_0$ denote the Dirac measure at zero.

**Theorem 1.3.** For a $\Lambda$-coalescent without a dust component, there is a sequence $(\gamma_n)_{n \in \mathbb{N}}$ such that $\gamma_n T^n$ converges in distribution to a probability measure unequal $\delta_0$ as $n \to \infty$ if and only if $\mu$ is regularly varying at infinity. Then its exponent $\alpha$ of regular variation fulfills $1 \leq \alpha \leq 2$, and we have

(i) for $1 < \alpha \leq 2$, 
\[ P \left( \frac{\mu(n)}{n} T^n > t \right) \to \frac{1}{(1 + (\alpha - 1) t)^{\alpha - 1}}, \quad t \geq 0, \]

(ii) for $\alpha = 1$, 
\[ P \left( \frac{\mu(n)}{n} T^n > t \right) \to e^{-t}, \quad t \geq 0, \]

as $n \to \infty$.

In particular, this theorem contains the special cases known from the literature. Blum and François [2] as well as Caliebe et al. [5] studied Kingman’s coalescent. For the Bolthausen-Sznitman coalescent, Freund and Möhle [14] showed asymptotic exponentiality of the external branch length. This result was somewhat generalized by Yuan [37]. A class of coalescents containing the Beta($2 - \alpha$, $\alpha$)-coalescent with $1 < \alpha < 2$ was analyzed by Dhersin et al. [8].

Combining Theorem 1.2 and 1.3 yields the following corollary:

**Corollary 1.4.** Suppose that the $\Lambda$-coalescent lacks a dust component and has regularly varying rate of decrease $\mu$ with exponent $\alpha \in [1, 2]$. Then for fixed $k \in \mathbb{N}$, we have

\[ \frac{\mu(n)}{n} (T_1^n, \ldots, T_k^n) \overset{d}{\to} (T_1, \ldots, T_k) \]

as $n \to \infty$, where $T_1, \ldots, T_k$ are i.i.d. random variables each having the density

\[ f(t) dt = \frac{\alpha}{(1 + (\alpha - 1) t)^{1+\frac{\alpha - 1}{\alpha}}}, \quad t \geq 0, \quad (1.1) \]

for $1 < \alpha \leq 2$ and a standard exponential distribution for $\alpha = 1$. 

4
For Beta-coalescents this corollary reads as follows:

**Example.** For $k \in \mathbb{N}$, let $T_1, \ldots, T_k$ be the i.i.d. random variables from Corollary 1.4.

(i) If $\Lambda (\{0\}) = 2$, then $\mu(n) \sim n^2$ and consequently

\[
n(T_1^n, \ldots, T_k^n) \xrightarrow{d} (T_1, \ldots, T_k)
\]
as $n \to \infty$. This statement covers (after scaling) the Kingman case.

(ii) If $\Lambda (dp) = c_a p^{a-1} (1 - p)^{b-1} dp$ for $0 < a < 1$, $b > 0$ and $c_a := (1 - a)(2 - a)/\Gamma(a)$, then $\mu(n) \sim n^{2-a}$ and therefore

\[
n^{1-a} (T_1^n, \ldots, T_k^n) \xrightarrow{d} (T_1, \ldots, T_k)
\]
as $n \to \infty$. After scaling, this includes the Beta$(2 - \alpha, \alpha)$-coalescent with $1 < \alpha < 2$ (see Theorem 1.1 of Siri-Jégousse and Yuan [32]).

(iii) If $\Lambda (dp) = (1 - p)^{b-1} dp$ with $b > 0$, then we have $\mu(n) \sim n \log n$ implying

\[
\log n (T_1^n, \ldots, T_k^n) \xrightarrow{d} (T_1, \ldots, T_k)
\]
as $n \to \infty$. This contains the Bolthausen-Sznitman coalescent (see Corollary 1.7 of Dhersin and Möhle [9]).

In the second part of this paper, we change perspective and examine the external branch lengths ordered by size downwards from their maximal value. In this context, an approach via a point process description is appropriate. Here we consider $\Lambda$-coalescents having regularly varying rate of decrease $\mu$, additionally to the absence of a dust component. It turns out that one has to distinguish two cases.

First, we treat the case of $\mu$ being regularly varying with exponent $\alpha \in (1, 2]$ (implying that the coalescent comes down from infinity). We introduce the sequence $(s_n)_{n \geq 2}$ given by

\[
\mu(s_n) = \frac{\mu(n)}{n}
\]

Note that $\mu(n)/n$ is a strictly increasing and, in the dustless case, diverging sequence (see Lemma 2.1 (ii) and (iv) below), which directly transfers to the sequence $(s_n)_{n \geq 2}$. Also note in view of Lemma 2.1 (ii) below that

\[
s_n = o(n)
\]
as $n \to \infty$. 
Example. (i) If $\mu(n) \sim n^\alpha$ with $\alpha \in (1, 2]$, then we have $s_n \sim n^{(\alpha - 1)/\alpha}$ as $n \to \infty$.

(ii) If $\mu$ is regularly varying with exponent $\alpha \in (1, 2]$, then the sequence $s_n$ is regularly varying with exponent $(\alpha - 1)/\alpha$.

We define point processes $\Phi^n$ on $(0, \infty)$ via

$$\Phi^n(B) := \# \{ i \leq n : \frac{\mu(n)}{ns_n} T^n_i \in B \}$$

for Borel sets $B \subset (0, \infty)$.

**Theorem 1.5.** Assume that the $\Lambda$-coalescent has a regularly varying rate of decrease $\mu$ with exponent $\alpha \in (1, 2]$. Then, as $n \to \infty$, the point process $\Phi^n$ converges in distribution to a Poisson point process $\Phi$ on $(0, \infty)$ with intensity measure

$$\phi(dx) = \frac{\alpha}{((\alpha - 1)x)^{1+\frac{\alpha}{\alpha-1}}} dx.$$ 

Note that $\int_0^1 \phi(x) dx = \infty$, which means that the points from the limit $\Phi$ accumulate at the origin. On the other hand, we have $\int_1^\infty \phi(x) dx < \infty$ saying that the points can be arranged in decreasing order. Thus, the theorem focuses on the maximal external lengths showing that the longest external branches differ from a typical one by the factor $s_n$ in order of magnitude (see Corollary 1.4). For Kingman’s coalescent, this result was obtained by Janson and Kersting [16] using a different method.

Corollary 1.4 shows that the external branch lengths behave for large $n$ as i.i.d. random variables. This observation is emphasized by the above theorem because the maximal values of i.i.d. random variables with the densities stated in Corollary 1.4 have just the limiting behavior as given in the above theorem (including the scaling constants $s_n$).

This heuristics fails for the Bolthausen-Sznitman coalescent, which we address now. For $n \in \mathbb{N}$, define the quantity

$$t_n := \log \log n - \log \log \log n + \frac{\log \log \log n}{\log \log n},$$

where we put $t_n := 0$ if the right-hand side is negative or not well-defined. Here we consider the point processes $\Psi^n$ on the whole real line given by

$$\Psi^n(B) := \# \{ i \leq n : \log \log (n)(T^n_i - t_n) \in B \}$$

for Borel sets $B \subset \mathbb{R}$. As before, we focus on the maximal values of $\Psi^n$. 

6
Theorem 1.6. For the Bolthausen-Sznitman coalescent, the point process $\Psi^n$ converges in distribution as $n \to \infty$ to a Cox point process $\Psi$ on $\mathbb{R}$ directed by the random measure

$$\psi(dx) = E e^{-x} dx,$$

where $E$ denotes a standard exponential random variable.

Observe that this random density may be rewritten as

$$e^{-x+\log E} dx.$$

This means that the limiting point process can also be considered as a Poisson point process with intensity measure $e^{-x} dx$ shifted by the independent amount $\log E$. This alternative representation will be used in the theorem’s proof (see Theorem 8.1 below). Recall that $G := -\log E$ has a standard Gumbel distribution.

We point out that the limiting point process $\Psi$ no longer coincides with the limiting Poisson point process as obtained for the maximal values of $n$ independent exponential random variables. The same turns out to be true for the scaling sequences. In order to explain these findings, let us consider the external branch with maximal length $T_{(1)}^n$. Then Theorem 1.6 implies that

$$\frac{T_{(1)}^n}{\log \log n} = 1 + c_p(1)$$

as $n \to \infty$, where $c_p(1)$ denotes a sequence of random variables converging to 0 in probability. In particular, $T_{(1)}^n \to \infty$ in probability. Hence, we pass with this theorem to the situation where very large mergers effect the maximal external lengths. Then circumstances change and new techniques are required. For this reason, we have to confine ourselves to the Bolthausen-Sznitman coalescent in the case of regularly varying $\mu$ with exponent $\alpha = 1$.

It is interesting to note that an asymptotic shift by a Gumbel distributed variable also shows up in the absorption time $\tilde{\tau}_n$ (the moment of the most recent common ancestor) of the Bolthausen-Sznitman coalescent:

$$\tilde{\tau}_n - \log \log n \xrightarrow{d} G$$

as $n \to \infty$ (see Goldschmidt and Martin [15]). However, this shift remains unscaled. Apparently, these two Gumbel distributed variables under consideration built up within different parts of the coalescent tree.

Before closing this introduction, we give some hints concerning the proofs. For the first three theorems, we make use of an asymptotic representation of the tail probabilities of the external
branch lengths, which is given in Theorem 3.1 below. Remarkably, this representation involves solely the rate of decrease $\mu$. This fact largely relies on different approximation formulas derived in [11].

The proofs of the last two theorems incorporate Corollary 1.4 as one ingredient. The idea is to implement stopping times $\tilde{\rho}_{c,n}$ with the property that at that moment a positive number of external branches is still extant which remains bounded as $n \to \infty$. To these remaining branches, the results of Corollary 1.4 are applied taking the strong Markov property into account. More precisely, let

$$N_n = (N_n(t), t \geq 0)$$

be the block counting process of the $n$-coalescent, where

$$N_n(t) := \#\Pi_n(t)$$

states the number of lineages present at time $t \geq 0$. For definiteness, we put $N_n(t) = 1$ for $t > \tilde{\tau}_n$. In the case of $1 < \alpha \leq 2$, we set

$$\tilde{\rho}_{c,n} := \inf \{t \geq 0 : N_n(t) \leq cs_n\}$$

with some $c > 0$. Next, we split the external lengths $T_{i}^n$ into the times $\tilde{T}_{i}^n$ up to the moment $\tilde{\rho}_{c,n}$ and the residual times $\hat{T}_{i}^n$. Formally, we have

$$\tilde{T}_{i}^n := T_{i}^n \wedge \tilde{\rho}_{c,n} \quad \text{and} \quad \hat{T}_{i}^n := T_{i}^n - \tilde{T}_{i}^n.$$

![Figure 1: The stopping time $\tilde{\rho}_{c,n}$ subdividing the external branch ending in leaf $i$ into two parts of length $\tilde{T}_{i}^n$ and $\hat{T}_{i}^n$.](image)
The approach for the Bolthausen-Sznitman coalescent is essentially the same. Here the role of \( \tilde{\rho}_{c,n} \) is taken by \( t_{c,n} \wedge \tilde{\tau}_n \), where

\[
t_{c,n} := t_n - \frac{\log c}{\log \log n}
\]

for some \( c > 1 \). Thus, for the Bolthausen-Sznitman coalescent, the external lengths \( T_i^n \) are split into

\[
\tilde{T}_i^n := T_i^n \wedge t_{c,n} \quad \text{and} \quad \hat{T}_i^n := T_i^n - \tilde{T}_i^n.
\]

The paper is organized as follows: Section 2 summarizes several properties of the rate of decrease. The fundamental asymptotic expression of the external tail properties is developed in Section 3. Section 4 and 5 contain the proofs of Theorem 1.1 to 1.3. In Section 6, we prepare the proofs of the remaining theorems by establishing a formula for factorial moments of the number of external branches. Section 7 and 8 contain the proofs of Theorem 1.5 and 1.6.

2 Properties of the rate of decrease

We now have a closer look at the rate of decrease \( \mu \) introduced in the first section. Defining

\[
\mu(x) := \int_{[0,1]} (xp - 1 + (1 - p)x) \frac{\Lambda(dp)}{p^2},
\]

we extend \( \mu \) to all real values \( x \geq 1 \), where the integrand’s value at \( p = 0 \) is understood to be \( x(x - 1)/2 \).

The next lemma summarizes some required properties of \( \mu \).

**Lemma 2.1.** The rate of decrease and its derivatives have the following properties:

(i) \( \mu(x) \) has derivatives of any order with finite values, also at \( x = 1 \). Moreover, \( \mu \) and \( \mu' \) are both non-negative and strictly increasing, while \( \mu'' \) is a non-negative and decreasing function.

(ii) For \( 1 < x \leq y \),

\[
\frac{x(x - 1)}{y(y - 1)} \leq \frac{\mu(x)}{\mu(y)} \leq \frac{x}{y}.
\]

(iii) For \( x > 1 \),

\[
\mu'(1) \leq \frac{\mu(x)}{x - 1} \leq \mu'(x) \quad \text{and} \quad \mu''(x) \leq \frac{\mu'(x)}{x - 1}.
\]
(iv) In the dustless case, \( \frac{\mu(x)}{x} \to \infty \) as \( x \to \infty \).

**Proof.** (i) Let

\[ \mu_2(x) := \int_{[0,1]} (1-p)^x \log(1-p) \frac{\Lambda(dp)}{p^2}, \]

which is a \( C^\infty \)-function for \( x > 0 \). Set

\[ \mu_1(x) := \int_1^x \mu_2(y) dy + \int_{[0,1]} (p + (1-p) \log(1-p)) \frac{\Lambda(dp)}{p^2} \]

\[ = \int_{[0,1]} ((1-p)^x \log(1-p) + p) \frac{\Lambda(dp)}{p^2} \]

Note that the second integral in the first is finite and non-negative just as its integrand. Then we have

\[ \mu(x) = \int_1^x \mu_1(y) dy. \]

Thus, \( \mu_1(x) = \mu'(x) \) and \( \mu_2(x) = \mu''(x) \) for \( x \geq 1 \). From these formulas our claim follows.

(ii) The inequalities are equivalent to the fact that \( \mu(x)/x \) is increasing and \( \mu(x)/(x(x-1)) \) is decreasing as follows from formulas (7) and (8) of [11].

(iii) The monotonicity properties from (i) and \( \mu(1) = 0 \) yield for \( x \geq 1 \),

\[ \mu'(1)(x-1) \leq \mu(1) + \int_1^x \mu'(y) dy \leq \mu'(x)(x-1). \]

Similarly, we get \( \mu''(x)(x-1) \leq \mu'(x) \) because \( \mu'(1) \geq 0 \).

(iv) See Lemma 1 (iii) of [11].

In order to characterize regular variation of \( \mu \), we introduce the function

\[ H(u) := \frac{\Lambda\{0\}}{2} + \int_0^u h(z) dz, \quad 0 \leq u \leq 1, \]

where

\[ h(z) := \int_z^1 \int_{[y,1]} \frac{\Lambda(dp)}{p^2} dy, \quad 0 \leq z \leq 1. \]

Note that \( H \) is a finite function because we have

\[ H(1) = \frac{\Lambda\{0\}}{2} + \int_0^1 \int_0^y dz dy \frac{\Lambda(dp)}{p^2} = \frac{\Lambda([0,1])}{2} < \infty. \] (2.2)
Proposition 2.2. For a $\Lambda$-coalescent without a dust component, the following statements hold:

(i) $\mu(x)$ is regularly varying at infinity if and only if $H(u)$ is regularly varying at the origin. Then $\mu$ has an exponent $\alpha \in [1, 2]$, and we have

$$\mu(x) \sim \Gamma(3 - \alpha) x^2 H(x^{-1})$$

as $x \to \infty$.

(ii) $\mu(x)$ is regularly varying at infinity with some exponent $\alpha \in (1, 2)$ if and only if the function $\int_{(y, 1)} p^{-2} \Lambda(dp)$ is regularly varying at the origin with an exponent $\alpha \in (1, 2)$. Then we have

$$\mu(x) \sim \frac{\Gamma(2 - \alpha)}{\alpha - 1} \int_{x^{-1}}^{1} \frac{\Lambda(dp)}{p^2}$$

as $x \to \infty$.

The last statement brings the regular variation of $\mu$ together with the notion of regularly varying $\Lambda$-coalescents as introduced in [11].

For the proof of this proposition, we apply the following characterization of regular variation.

Lemma 2.3. Let $V(z), z > 0$, be a positive function with an ultimately monotone derivative $v(z)$, and let $\eta \neq 0$. Then $V$ is regularly varying at the origin with exponent $\eta$ if and only if $|v|$ is regularly varying at the origin with exponent $\eta - 1$, and

$$z v(z) \sim \eta V(z)$$

as $z \to 0^+$. 

Proof. For $\eta > 0$, we have $V(0+) = 0$ and therefore $V(z) = \int_{0}^{z} v(y)dy$. For $\eta < 0$, we use the equation $V(z) = \int_{0}^{1} (-v(y))dy + V(1)$ instead, here it holds $V(0+) = \infty$. Now our claim follows from well known results for regularly varying functions at infinity (see [28] as well as Theorem 1 (a) and (b) in Section VIII.9 [13]). The proofs translate one-to-one to regularly varying functions at the origin. \hfill \Box

Proof of Proposition 2.2. (i) From the definition (2.1), we obtain by double partial integration (see formula (8) of [11]) that

$$\frac{\mu(x)}{x(x - 1)} = \frac{\Lambda(\{0\})}{2} + \int_{0}^{1} (1 - z)^{x-2} h(z) dz.$$

(2.4)
If $\Lambda(\{0\}) > 0$, then our claim is obvious because the first term of the right-hand side of (2.4) dominates the integral as $x \to \infty$ implying $\mu(x)/x^2 \sim \Lambda(\{0\})/2 = H(0)$ and therefore $\alpha = 2$. Thus, let us assume that $\Lambda(\{0\}) = 0$. Let

$$\mathcal{L}(x) := \int_0^1 e^{-zx} h(z) \, dz$$

be the Laplace transform of $H$. In view of a Tauberian theorem (see Theorem 3 and Theorem 2 in Section XIII.5 of [13]), it is sufficient to prove that

$$\mathcal{L}(x) \sim \frac{\mu(x)}{x^2}$$

as $x \to \infty$. For $\frac{1}{2} < \delta < 1$, let us consider the decomposition

$$\frac{\mu(x)}{x(x-1)} = \int_0^{x-\delta} (1-z)^{x-2} h(z) \, dz + \int_{x-\delta}^1 (1-z)^{x-2} h(z) \, dz. \quad (2.6)$$

Because of $\delta < 1$ and (2.2), we have

$$\int_{x-\delta}^1 (1-z)^{x-2} h(z) \, dz \leq (1-x^{-\delta})^{x-2} \int_{x-\delta}^1 h(z) \, dz \leq e^{-x^{-\delta}(x-2)} H(1) = o(x^{-1}) \quad (2.7)$$

as $x \to \infty$. In particular, the second integral in the decomposition (2.6) can be neglected in the limit $x \to \infty$ since $\mu(x)/(x(x-1)) \geq \mu'(1)/x$ due to Lemma 2.1 (iii). As to the first integral in (2.6), observe for $\delta > \frac{1}{2}$ that

$$\frac{(1-z)^{x-2}}{e^{-zx}} = e^{-O(x^{1-2\delta})} \to 1$$

uniformly for $z \in [0, x^{-\delta}]$ as $x \to \infty$ and therefore

$$\int_0^{x-\delta} (1-z)^{x-2} h(z) \, dz \sim \int_0^{x-\delta} e^{-zx} h(z) \, dz. \quad (2.8)$$

Also note that

$$\int_{x-\delta}^1 e^{-zx} h(z) \, dz \leq e^{-x^{-\delta}} H(1) = o(x^{-1}) \quad (2.9)$$

as $x \to \infty$. Combining (2.6) to (2.9) entails

$$\int_0^1 (1-z)^{x-2} h(z) \, dz \sim \mathcal{L}(x).$$
Hence, along with formula (2.4), this proves the asymptotics in (2.5). Moreover, from Lemma 2.1 (ii) we get $1 \leq \alpha \leq 2$.

(ii) If $1 < \alpha < 2$, then $\Lambda(\{0\}) = 0$. Lemma 2.3 provides that for $\alpha < 2$ the function $H(u)$ is regularly varying with exponent $2 - \alpha$ iff $h(u)$ is regularly varying with exponent $1 - \alpha$, and then

$$(2 - \alpha)H(u) \sim uh(u)$$

as $u \to 0^+$. Applying Lemma 2.3 once more for $\alpha > 1$, $h(u)$ is regularly varying with exponent $1 - \alpha$ iff $\int_{(u,1]} \frac{\Lambda(dp)}{p^2}$ is regularly varying with exponent $-\alpha$, and then

$$(\alpha - 1)h(u) \sim u \int_{(u,1]} \frac{\Lambda(dp)}{p^2}$$

as $u \to 0^+$. Bringing both asymptotics together with statement (i) finishes the proof.

3 The length of a random external branch

Recall that $T_n$ denotes the length of an external branch picked at random. The following result on its distribution function does not only play a decisive role in the proofs of Theorem 1.1 and 1.3 but is of interest on its own. It shows that the distribution of $T_n$ is primarily determined by the rate function $\mu$.

Theorem 3.1. For a $\Lambda$-coalescent without a dust component and a sequence $(r_n)_{n \in \mathbb{N}}$ satisfying $1 < r_n \leq n$ for all $n \in \mathbb{N}$, we have

$$\mathbb{P}(T_n > \int_{r_n}^n \frac{dx}{\mu(x)}) = \frac{\mu(r_n)}{\mu(n)} + o(1)$$

as $n \to \infty$. Moreover,

$$\left(\frac{r_n}{n}\right)^2 + o(1) \leq \mathbb{P}(T_n > \int_{r_n}^n \frac{dx}{\mu(x)}) \leq \frac{r_n}{n} + o(1)$$

as $n \to \infty$.

For the proof, we introduce some notations. Recall from the introduction that the continuous-time Markov process $N_n$ denotes the block counting process. For the embedded discrete-time Markov chain, we use the notation $X = (X_j)_{j \in \mathbb{N}_0}$, where $X_j$ denotes the number of branches after $j$ merging events. In particular, we have $X_0 = n$, and we set $X_j = 1$ for $j \geq \tau_n$, where $\tau_n$ is defined as the total number of merging events. (For convenience, we suppress $n$ in the notation.
of $X$.) The waiting time of the process $N_n$ in state $X_j$ is referred to as $W_j$ for $0 \leq j \leq \tau_n - 1$. The number of merging events until the external branch ending in leaf $i \in \{1, \ldots, n\}$ coalesces is given by
\[
\zeta^n_i := \min \{ j \geq 1 : \{i\} \notin \Pi_n(W_0 + \cdots + W_{j-1}) \}.
\]
Similarly, $\zeta^n$ denotes the corresponding number of a random external branch with length $T^n$.

**Proof of Theorem 3.1.** For later purpose, we show the stronger statement
\[
\mathbb{P} \left( T^n > \int_{r_n}^{x} \frac{dx}{\mu(x)} \bigg| N_n \right) = \frac{\mu(r_n)}{\mu(n)} + o_P(1) \tag{3.3}
\]
as $n \to \infty$. It implies (3.1) by taking expectations and using dominated convergence. The statement (3.2) is a direct consequence in view of Lemma 2.1 (ii).

In order to prove (3.3), note that, by the standard subsubsequence argument and the metrizability of the convergence in probability, we can assume that $r_n/n$ converges to some value $q$ with $0 \leq q \leq 1$. We distinguish three different cases of asymptotic behavior of the sequence $r_n/n$:

(a) We begin with the case $r_n \sim qn$ as $n \to \infty$, where $0 < q < 1$. Then there exist $q_1, q_2 \in (0, 1)$ such that $q_1 n \leq r_n \leq q_2 n$ for all $n \in \mathbb{N}$ but finitely many.

From Lemma 4 of [11], we know that
\[
\mathbb{P} \left( \zeta^n \geq k \mid N_n \right) = \frac{X_k - 1}{n - 1} \prod_{j=0}^{k-1} \left( 1 - \frac{1}{X_j} \right) \text{ a.s.}
\]
for $k \geq 1$. Note that $\sum_{j=0}^{k-1} X_j^{-2} \leq \sum_{m=X_{k-1}}^{\infty} m^{-2} \leq 2 (X_{k-1})^{-1}$ to obtain via a Taylor expansion that
\[
\mathbb{P} \left( \zeta^n \geq k \mid N_n \right) = \frac{X_k - 1}{n - 1} \exp \left( - \sum_{j=0}^{k-1} \frac{1}{X_j} + o \left( X_{k-1}^{-1} \right) \right) \tag{3.4}
\]
as $n \to \infty$.

Now we have a closer look at the stopping times
\[
\rho_{r_n} := \min \{ j \geq 0 : X_j \leq r_n \}
\]
and their analogues
\[
\tilde{\rho}_{r_n} := \inf \{ t \geq 0 : N_n(t) \leq r_n \}.
\]
From Lemma 2.1 (i) and (iii), we know that the function $\mu(x)$ is increasing in $x$ and that $x/\mu(x)$
converges in the dustless case to 0 as \( x \to \infty \). In view of \( r_n \geq q_1 n \), we therefore have

\[
\int_{r_n}^{n} \frac{dx}{\mu(x)} \leq \frac{n - r_n}{\mu(r_n)} \leq \left( \frac{1}{q_1} - 1 \right) \frac{r_n}{\mu(r_n)} = o(1).
\]

Thus, Lemma 3 of [11] implies the convergence of \( \tilde{\rho}_{r_n} \) in probability to 0 and

\[
X_{\rho_{r_n}} = r_n + O_P(\Delta X_{\rho_{r_n}}) = r_n + o_P(X_{\rho_{r_n}}).
\]

Hence, we may apply Proposition 3 of [11] yielding

\[
\sum_{j=0}^{\rho_{r_n}-1} \frac{1}{X_j} = \log \left( \frac{\mu(n)}{n} \frac{X_{\rho_{r_n}}}{\mu(X_{\rho_{r_n}})} \right) + o_P(1).
\]

Using Lemma 2.1 (ii) once more, equation (3.4) implies

\[
P(\zeta_n \geq \rho_{r_n} | N_n) = \frac{X_{\rho_{r_n}} - 1}{n} \frac{\mu(X_{\rho_{r_n}})}{\mu(n)} \frac{n}{1 + o_P(1)} = \frac{\mu(r_n)}{\mu(n)} + o_P(1).
\] (3.5)

In order to transfer this equality to the continuous-time setting, we first show that for each \( \varepsilon \in (0, 1) \) there is a \( \delta > 0 \) such that

\[
(1 + \delta) \int_{(1+\varepsilon)r_n}^{n} \frac{dx}{\mu(x)} < \int_{r_n}^{n} \frac{dx}{\mu(x)} < (1 - \delta) \int_{(1-\varepsilon)r_n}^{n} \frac{dx}{\mu(x)}
\] (3.6)

for large \( n \in \mathbb{N} \). For the proof of the left-hand inequality, note that due to Lemma 2.1 (ii) we have

\[
\frac{1}{n - (1+\varepsilon)r_n} \int_{(1+\varepsilon)r_n}^{n} \frac{dx}{\mu(x)} \leq \frac{1}{n - r_n} \int_{r_n}^{n} \frac{dx}{\mu(x)}
\]

implying with \( q_1 n \leq r_n \) that

\[
\frac{1}{1 - \varepsilon \frac{q_1}{1-q_1}} \int_{(1+\varepsilon)r_n}^{n} \frac{dx}{\mu(x)} \leq \frac{1}{1 - \varepsilon \frac{r_n}{n-r_n}} \int_{r_n}^{n} \frac{dx}{\mu(x)} \leq \int_{r_n}^{n} \frac{dx}{\mu(x)}.
\]

These inequalities show how to choose \( \delta > 0 \). The right-hand inequality in (3.6) follows along the same lines.

Finally, the choice \( f \equiv 1 \) in Proposition 2 of [11] provides for sufficiently small \( \varepsilon > 0 \),

\[
\tilde{\rho}_{r_n(1+\varepsilon)} = \int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} (1 + o_P(1))
\] (3.7)
as \( n \to \infty \). Combining (3.5) to (3.7) yields

\[
\mathbb{P}\left( T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right) 
\leq \mathbb{P}\left( T^n \geq (1 + \delta) \int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right)
\leq \mathbb{P}\left( T^n \geq \tilde{\rho}_r \int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right) + \mathbb{P}\left( (1 + \delta) \int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} < \tilde{\rho}_r \int_{r_n(1+\varepsilon)}^{n} \mu(x) \bigg| N_n \right)
= \mathbb{P}\left( \zeta^n \geq \rho_{r_n(1+\varepsilon)} \big| N_n \right) + o_P(1)
= \frac{\mu(r_n(1+\varepsilon))}{\mu(n)} + o_P(1)
\leq \frac{\mu(r_n)}{\mu(n)} (1 + \varepsilon)^2 + o_P(1),
\]

where we used Lemma 2.1 (ii) for the last inequality. With this estimate holding for all \( \varepsilon > 0 \), we end up with

\[
\mathbb{P}\left( T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right) \leq \frac{\mu(r_n)}{\mu(n)} + o_P(1)
\]
as \( n \to \infty \). The reverse inequality can be shown in the same way so that we obtain equation (3.3).

(b) Now we turn to the two remaining cases \( r_n \sim n \) and \( r_n = o(n) \). In view of Lemma 2.1 (ii), the asymptotics \( r_n \sim n \) implies \( \mu(r_n) \sim \mu(n) \), i.e., the right-hand side of (3.3) converges to 1. Furthermore, the sequence \( (r'_n)_{n \in \mathbb{N}} := (qr_n)_{n \in \mathbb{N}}, 0 < q < 1 \), fulfills the requirements of part (a). With respect to Lemma 2.1 (ii), part (a) therefore entails for all \( q \in (0, 1) \),

\[
\mathbb{P}\left( T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right) \geq \mathbb{P}\left( T^n > \int_{r'_n}^{n} \frac{dx}{\mu(x)} \bigg| N_n \right) \geq \frac{\mu(qn)}{\mu(n)} + o_P(1) \geq q^2 + o_P(1)
\]
as \( n \to \infty \). Hence, the left-hand side of (3.3) also converges in probability to 1. Similarly, the convergence of both sides of (3.3) to 0 can be shown for \( r_n = o(n) \). \( \square \)
4 Proofs of Theorem 1.1 and 1.2

Proof of Theorem 1.1. Let \( r_n \) be as required in Theorem 3.1. Applying Lemma 2.1 (ii), we obtain
\[
\int_{r_n}^{n} \frac{dx}{x} \leq \frac{\mu(n)}{n} \int_{r_n}^{n} \frac{dx}{\mu(x)} \leq \int_{r_n}^{n} \frac{n-1}{x(x-1)} dx.
\]
Observing
\[
\int_{r_n}^{n} \frac{dx}{x} = \log \frac{n}{r_n}
\]
and
\[
\int_{r_n}^{n} \frac{n-1}{x(x-1)} dx = (n-1) \log \frac{r_n-nr_n}{n-nr_n},
\]
Theorem 3.1 entails
\[
P \left( \frac{\mu(n)}{n} T^n > \log \frac{n}{r_n} \right) \geq \left( \frac{r_n}{n} \right)^2 + o(1) \tag{4.1}
\]
and
\[
P \left( \frac{\mu(n)}{n} T^n > (n-1) \log \frac{r_n-nr_n}{n-nr_n} \right) \leq \frac{r_n}{n} + o(1) \tag{4.2}
\]
as \( n \to \infty \), respectively.

Now let \( t \geq 0 \). Using equation (4.1) for
\[
r_n = ne^{-t},
\]
while choosing
\[
r_n = \frac{ne^{t/(n-1)}}{1 + n(e^{t/(n-1)} - 1)}
\]
in (4.2), we arrive at
\[
e^{-2t} + o(1) \leq P \left( \frac{\mu(n)}{n} T^n > t \right) \leq \frac{e^{t/(n-1)}}{1 + n(e^{t/(n-1)} - 1)} + o(1) = \frac{1}{1 + t} (1 + o(1)),
\]
as required. \( \Box \)

Proof of Theorem 1.2. Similar to the proof of Theorem 3.1, we are first considering the discrete version \( \zeta^n_i \) of \( T^n_i \) for \( 1 \leq i \leq k \) to prove
\[
P \left( \zeta^n_0 \geq I^n_0, \ldots, \zeta^n_k \geq I^n_k \mid N_n \right) = P \left( \zeta^n_0 \geq I^n_0 \mid N_n \right) \cdots P \left( \zeta^n_k \geq I^n_k \mid N_n \right) + o(1) \quad a.s. \tag{4.3}
\]
as \( n \to \infty \), where \( 0 =: I^n_0 \leq I^n_1 \leq \cdots \leq I^n_k \) are random variables measurable with respect to the
σ-fields σ( \( N_n \)). Denote by \( \zeta_A \) the number of mergers until some external branch out of the set \( A \subseteq \{1, \ldots, n\} \) coalesces, and let \( a := \#A \). Given \( \Delta X_j \), the \( j \)-th merging amounts to choosing \( \Delta X_j + 1 \) branches uniformly at random out of the \( X_j \) present ones implying

\[
P( \zeta_A \geq m \mid N_n) = \frac{(X_m - 1) \cdots (X_m - a)}{(n - 1) \cdots (n - a)} \prod_{j=0}^{m-1} \left( 1 - \frac{a}{X_j} \right) \quad \text{a.s.} \quad (4.4)
\]

for \( m \geq 1 \) (for details see (28) of [11]). Let \( \bar{\zeta}_{i, \ldots, k} := \zeta_{i, \ldots, k} - \zeta_{i-1, \ldots, k} \) for \( 2 \leq i \leq k \). Moreover, let \( \bar{N}_{X_i}(t) := N_n(t + W_0 + \cdots + W_{j-1}) \), in particular, \( \bar{N}_{X_i}(t) := N_n(t) \).

The Markov property and (4.4) provide

\[
P(\zeta^n_1 \geq I^n_1, \ldots, \zeta^n_k \geq I^n_k \mid N_n)
= \prod_{i=1}^k P(\bar{\zeta}_{i, \ldots, k} \geq I^n_i - I^n_{i-1} \mid \bar{N}_{X_i}^{I^n_i - 1})
= \prod_{i=1}^k \left[ \frac{(X_{I^n_1} - 1) \cdots (X_{I^n_1} - k + i - 1)}{(X_{I^n_1} - 1) \cdots (X_{I^n_1} - k + i - 1)} \prod_{j=I^n_{i-1}}^{I^n_i - 1} \left( 1 - \frac{k - i + 1}{X_j} \right) \right]
= \prod_{i=1}^k \left[ \frac{(X_{I^n_i} - k + i - 1)}{(n - k + i - 1)} \prod_{j=I^n_{i-1}}^{I^n_i - 1} \left( 1 - \frac{k - i + 1}{X_j} \right) \right] \quad \text{a.s.}
\]

For \( 1 \leq i \leq k \), note that

\[
\left( 1 - \frac{k - i + 1}{X_j} \right) = \left( 1 - \frac{1}{X_j} \right)^{k-1} + \mathcal{O} \left( X_j^{-1} \right)
\]

and

\[
\frac{X_{I^n_i} - k + i - 1}{n - k + i - 1} = \frac{X_{I^n_i} - 1}{n - 1} + \mathcal{O} (n^{-1})
\]

to obtain

\[
P(\zeta^n_1 \geq I^n_1, \ldots, \zeta^n_k \geq I^n_k \mid N_n)
= \prod_{i=1}^k \left[ \frac{X_{I^n_i} - 1}{n - 1} + \mathcal{O} (n^{-1}) \right] \left( \prod_{j=I^n_{i-1}}^{I^n_i - 1} \left( 1 - \frac{1}{X_j} \right)^{k-1} + \mathcal{O} ((X_{I^n_i} - 1)^{-1}) \right)
= \prod_{i=1}^k \left[ \frac{X_{I^n_i} - 1}{n - 1} \prod_{j=I^n_{i-1}}^{I^n_i - 1} \left( 1 - \frac{1}{X_j} \right)^{k-1} \right] + o(1)
\]
\[
= \prod_{i=1}^{k} \left[ \frac{X_{I^n_i} - 1}{n-1} \prod_{j=0}^{I^n_i-1} \left( 1 - \frac{1}{X_j} \right) \right] + o(1) \quad \text{a.s.}
\]
as \( n \to \infty \), where the rightmost \( O(\cdot) \) term in the first line stems from the fact that \( X_{I^n_i} < X_j \) for all \( j < I^n_i \). Further, from (4.4) with \( A = \{ i \} \), we know that
\[
P \left( \zeta^n_i \geq I^n_i \mid N_n \right) = \frac{X_{I^n_i} - 1}{n-1} \prod_{j=0}^{I^n_i-1} \left( 1 - \frac{1}{X_j} \right) \quad \text{a.s.}
\]
so that we receive equation (4.3).

Now based on exchangeability, it is no loss to assume that \( 0 \leq t^n_1 \leq \cdots \leq t^n_k \). So inserting
\[
I^n_i := \min \left\{ k \geq 1 : \sum_{j=0}^{k-1} W_j > t^n_i \right\} \wedge \tau_n
\]
in (4.3) yields
\[
P \left( T^n_1 > t^n_1, \ldots, T^n_k > t^n_k \mid N_n \right) = \prod_{i=1}^{k} P \left( \zeta^n_i \geq I^n_i \mid N_n \right) + o(1)
\]
\[
= \prod_{i=1}^{k} P \left( T^n_i > t^n_i \mid N_n \right) + o(1) \quad \text{a.s.}
\]
as \( n \to \infty \). For \( 1 \leq i \leq k \), let \( 1 < r^n_i \leq n \) be defined implicitly via
\[
t^n_i = \int_{r^n_i}^{n} \frac{dx}{\mu(x)}.
\]

From Lemma 2.1 (iii) we know that \( \int_{1}^{n} \frac{dx}{\mu(x)} = \infty \); therefore, \( r^n_i \) is well-defined. Thus, we may apply formula (3.3) to obtain
\[
P \left( T^n_1 > t^n_1, \ldots, T^n_k > t^n_k \mid N_n \right) = \prod_{i=1}^{k} P \left( T^n_i > t^n_i \mid N_n \right) + o_P(1)
\]
\[
= \prod_{i=1}^{k} \frac{\mu(r^n_i)}{\mu(n)} + o_P(1)
\]
as \( n \to \infty \). Taking expectations in this equation yields via dominated convergence the theorem’s claim. \( \square \)
5 Proof of Theorem 1.3

(a) First suppose that $\gamma_n T^n$ converges for some positive sequence $(\gamma_n)_{n \in \mathbb{N}}$ in distribution as $n \to \infty$ to a probability measure unequal $\delta_0$ with cumulative distribution function $F = 1 - \bar{F}$, i.e.,

$$
P(\gamma_n T^n > t) \xrightarrow{n \to \infty} \bar{F}(t)
$$

(5.1)

for $t \geq 0$, $t \notin D$, where $D$ denotes the set of discontinuities of $\bar{F}$. We first show that $\gamma_n \sim c \mu'(n)$ for some $c > 0$. From Theorem 1.1 it follows that there exist $0 < c_1 \leq c_2 < \infty$ with

$$
c_1 \frac{\mu(n)}{n} \leq \gamma_n \leq c_2 \frac{\mu(n)}{n}, \quad n \geq 2,
$$

(5.2)

and that $0 < \bar{F}(t) < 1$ for all $t > 0$. Similarly as in the proof of Theorem 1.2, define $r_n(t)$ for $t \geq 0$ implicitly via

$$
t = \gamma_n \int_{r_n(t)}^{n} \frac{dx}{\mu(x)}.
$$

(5.3)

Applying formula (3.3) and (5.1), we obtain

$$
\frac{\mu(r_n(t))}{\mu(n)} = \bar{F}(t) + o(1)
$$

(5.4)

for all $t \geq 0$, $t \notin D$, as $n \to \infty$. Differentiating both sides of (5.3) with respect to $t$ and using Lemma 2.1 (i) yields

$$
\left| \frac{\gamma_n r_n'(t)}{\mu(n)} \right| = \frac{\mu(r_n(t))}{\mu(n)} \leq 1.
$$

In conjunction with (5.4), it follows that

$$
\frac{\gamma_n r_n'(t)}{\mu(n)} = -\bar{F}(t) + o(1)
$$

and, by dominated convergence and (5.2),

$$
r_n(t) = n - \frac{\mu(n)}{\gamma_n} \left( \int_0^t \bar{F}(s) ds + o(1) \right) = n - \frac{\mu(n)}{\gamma_n} \int_0^t \bar{F}(s) ds + o(n)
$$

(5.5)

as $n \to \infty$. Furthermore, from a Taylor expansion, we get

$$
\mu(r_n(t)) = \mu(n) + \mu'(n) (r_n(t) - n) + \frac{1}{2} \mu''(\zeta_n) (r_n(t) - n)^2,
$$
where \( r_n(t) \leq \xi_n \leq n \). Dividing this equation by \( \mu(n) \), using (5.4) and (5.5) as well as rearranging terms, we obtain

\[
\left| 1 - \bar{F}(t) + o(1) - \frac{\mu'(n)}{\gamma_n} \int_0^t \bar{F}(s) ds (1 + o(1)) \right| = \frac{\mu''(n) \mu(n)}{2 \gamma_n^2} \left( \int_0^t \bar{F}(s) ds \right)^2 (1 + o(1))
\]

as \( n \to \infty \). From Lemma 2.1 (iii) and (i), we get \( \mu''(\xi_n) \leq \mu'(\xi_n)/\xi_n - 1 \leq \mu'(n)/(r_n(t) - 1) \). Moreover, equation (5.5) with (5.2) yield \( r_n(t) - 1 \geq n/2 + o(n) \) for \( t \) sufficiently small. Taking (5.2) once more into account, we obtain that for given \( \varepsilon > 0 \) and \( t \) sufficiently small,

\[
\left| 1 - \bar{F}(t) + o(1) - \frac{\mu'(n)}{\gamma_n} \int_0^t \bar{F}(s) ds (1 + o(1)) \right| \leq \varepsilon \frac{\mu'(n)}{\gamma_n} \left( \int_0^t \bar{F}(s) ds \right) (1 + o(1))
\]

or equivalently, for \( t > 0 \),

\[
\left| \gamma_n \frac{\mu'(n)}{\mu'(x)} - \frac{\int_0^t \bar{F}(s) ds}{1 - \bar{F}(t)} (1 + o(1)) \right| \leq \varepsilon \frac{\int_0^t \bar{F}(s) ds}{1 - \bar{F}(t)} (1 + o(1)) .
\]

The right-hand quotient is finite and positive for all \( t > 0 \), which implies our claim \( \gamma_n \sim c \mu'(n) \) for some \( c > 0 \).

Without loss of generality, we now set \( \gamma_n = \mu'(n) \). With this choice, inserting (5.5) in (5.4) yields

\[
\mu(n) \bar{F}(t) (1 + o(1)) = \mu(r_n(t)) = \mu \left( n - \frac{\mu(n)}{\mu'(n)} \int_0^t \bar{F}(s) ds + o(n) \right)
\]

as \( n \to \infty \). In view of the monotonicity properties of \( \mu \) and \( \mu' \) due to Lemma 2.1 (i), we may proceed to

\[
\mu(x) \bar{F}(t) = \mu \left( x - \frac{\mu(x)}{\mu'(x)} \int_0^t \bar{F}(s) ds + o(x) \right) (1 + o(1))
\]

\[
= \mu \left( x - \frac{\mu(x)}{\mu'(x)} \int_0^t \bar{F}(s) ds + o(x) \right)
\]

as \( x \to \infty \). Similarly, (5.2) yields

\[
d_1 \frac{\mu(x)}{x} \leq \mu'(x) \leq d_2 \frac{\mu(x)}{x} , \quad x \geq 2,
\]
for suitable $0 < d_1 \leq d_2 < \infty$. From Lemma 2.1 (i) we know that $\mu(x)$ has an inverse $\nu(y)$. Let us now show that $\nu$ is regularly varying. For the inverse, the last equation translates into

$$\frac{\nu(y)}{d_2 y} \leq \nu'(y) \leq \frac{\nu(y)}{d_1 y}. \quad (5.8)$$

Applying $\nu$ to equation (5.7), both inside and outside, we get

$$\nu(yF(t)) = \nu(y) - y\nu'(y) \int_0^t F(s)ds + o(\nu(y)).$$

With $0 \leq u < v$, $u, v \notin D$, it follows that

$$\nu(F(u)y) - \nu(F(v)y) \geq \nu'(yF(u))y(F(u) - F(v)) \geq \nu'(y)(F(u) - F(v)). \quad (5.9)$$

Thus, also assuming $u, v \notin D$, (5.9) yields

$$F(u) - F(v) \leq \int_u^v F(s)ds \leq v - u,$$

which implies $D = \emptyset$. By a Taylor expansion, we get

$$\nu(F(v)y) - \nu(F(u)y) = -\nu'(F(u)y)y(F(u) - F(v)) + \frac{1}{2} \nu''(y)\xi(y)^2(F(u) - F(v))^2,$$

where $F(v)y \leq \xi(y) \leq F(u)y$. Dividing this equation by $y\nu'(y)$, using formula (5.9) and rearranging terms, it follows

$$\left| \int_u^v F(s)ds(1 + o(1)) - \frac{\nu'(F(u)y)}{\nu'(y)}(F(u) - F(v)) \right| = \frac{1}{2} \frac{\nu''(y)}{\nu'(y)}(F(u) - F(v))^2.$$

Note that from Lemma 2.1 (iii) we have

$$\left| \nu''(y) \right| = \nu'(y)^2 \frac{\mu''(\nu(y))}{\mu'(\nu(y))} \leq \frac{\nu'(y)^2}{\nu(y) - 1}.$$
Hence, together with (5.8) we obtain

\[
\left| \int_u^v \tilde{F}(s)ds(1+c(1)) - \frac{\nu'(\tilde{F}(u)y)}{\nu'(y)}(\tilde{F}(u) - \tilde{F}(v)) \right| \\
\leq \frac{1}{2} \frac{\nu'(\xi_y)^2 y}{(\nu(\xi_y) - 1)\nu'(y)}(\tilde{F}(u) - \tilde{F}(v))^2 \\
\leq \frac{1}{2d_1^2} \frac{\nu(\xi_y)^2 y}{(\nu(\xi_y) - 1)\nu'(y)\xi_y}(\tilde{F}(u) - \tilde{F}(v))^2 \\
\leq \frac{\nu(\tilde{F}(u)y)y}{d_1^2 \nu'(y)(\tilde{F}(v)y)^2}(\tilde{F}(u) - \tilde{F}(v))^2 \\
\leq \frac{d_2 \nu'(\tilde{F}(u)y)y^2 \tilde{F}(u)}{d_1^2 \nu'(y)(\tilde{F}(v)y)^2}(\tilde{F}(u) - \tilde{F}(v))^2 \\
\leq \varepsilon \frac{\nu'(\tilde{F}(u)y)}{\nu'(y)}(\tilde{F}(u) - \tilde{F}(v))
\]

for each \( \varepsilon > 0 \) and for \( v \) close enough to \( u \), or equivalently

\[
\left| \frac{\nu'(y)}{\nu'(\tilde{F}(u)y)} - \frac{\tilde{F}(u) - \tilde{F}(v)}{\int_u^v \tilde{F}(s)ds}(1+c(1)) \right| \leq \varepsilon \frac{\tilde{F}(u) - \tilde{F}(v)}{\int_u^v \tilde{F}(s)ds}(1+c(1)).
\]

Again, since the right-hand quotient is finite and positive for all \( u < v \), this estimate implies that \( \nu'(y)/\nu'(\tilde{F}(u)y) \) has a positive finite limit as \( y \to \infty \). Because \( \tilde{F}(u) \) takes all values between 0 and 1, \( \nu'(y) \) is regularly varying. From the Lemma in Section VIII.9 of [13], we then get the regular variation of \( \nu \) with exponent \( \eta \geq 0 \). It fulfills \( \frac{1}{2} \leq \eta \leq 1 \) as Lemma 2.1 (ii) yields

\[
a\sqrt{y} \leq \nu(y) \leq by
\]

for some \( a, b > 0 \). Hence, \( \mu \) as the inverse function of \( \nu \) is regularly varying with exponent \( \alpha \in (1, 2) \) (see Theorem 1.5.12 of [1]).

(b) Now suppose that \( \mu(x) \) is regularly varying with exponent \( \alpha \in [1, 2] \), i.e., we have

\[
\mu(x) = x^\alpha L(x), \quad (5.10)
\]

where \( L \) is a slowly varying function. Let \( r_n := qn \) with \( 0 < q \leq 1 \). The statement of Theorem 3.1 then boils down to

\[
\mathbb{P} \left( \frac{\mu(n)}{n} T^n > \frac{1}{n} \int_{qn}^{\mu(n)} \frac{\mu(x)}{\mu(x)}dx \right) = q^\alpha + o(1) \quad (5.11)
\]
as \( n \to \infty \). From (5.10) we obtain

\[
\frac{1}{n} \int_n^\infty \frac{\mu(n)}{\mu(x)} \, dx \sim \begin{cases}
-\log q & \text{for } \alpha = 1 \\
\frac{1}{\alpha - 1} (q^{-(\alpha - 1)} - 1) & \text{for } 1 < \alpha \leq 2
\end{cases}
\]
as \( n \to \infty \). Thus, choosing, for given \( t \geq 0 \),

\[
q = \begin{cases}
e^{-t} & \text{for } \alpha = 1 \\
(1 + (\alpha - 1)t)^{-\frac{1}{\alpha - 1}} & \text{for } 1 < \alpha \leq 2
\end{cases}
\]
in equation (5.11) yields the claim. \( \square \)

6 Moment calculations for external branches of \( \Lambda \)-coalescents

In this section, we consider the number of external branches \( Y_j \) after \( j \) merging events:

\[
Y_j := \# \{ 1 \leq i \leq n : \{ i \} \in \Pi_n (W_0 + \cdots + W_{j-1}) \}.
\]

In particular, we set \( Y_0 = n \) and \( Y_j = 0 \) for \( j > \tau_n \). (Again we suppress \( n \) in the notation for convenience.) We provide a representation of the conditional moments of the number of external branches for general \( \Lambda \)-coalescents (also covering coalescents with dust component). For this purpose, we use the notation \((x)_r := x(x-1)\cdots(x-r+1)\) for falling factorials with \( x \in \mathbb{R} \) and \( r \in \mathbb{N} \). Recall that \( \tau_n \) is the total number of merging events.

**Lemma 6.1.** Consider a general \( \Lambda \)-coalescent, and let \( \rho \) be a \( \sigma(N_n) \)-measurable random variable with \( 0 \leq \rho \leq \tau_n \) a.s.

(i) For a natural number \( r \), the \( r \)-th factorial moment, given \( N_n \), can be expressed as

\[
\mathbb{E} \left[ (Y_\rho)_r \mid N_n \right] = (X_\rho)_r \prod_{j=1}^\rho \left( 1 - \frac{r}{X_j} \right) = (X_\rho - 1)_r \frac{n}{n - r} \prod_{j=0}^{\rho-1} \left( 1 - \frac{r}{X_j} \right) \quad \text{a.s.}
\]

(ii) For the conditional variance, the following inequality holds:

\[
\text{Var} (Y_\rho \mid N_n) \leq \mathbb{E} [Y_\rho \mid N_n] \quad \text{a.s.}
\]
Proof. (i) First, we recall a link between the external branches and the hypergeometric distribution based on the Markov property and exchangeability properties of the Λ-coalescent, as already described for Beta-coalescents in [6]:

Given \( N_n \) and \( Y_0, \ldots, Y_{\rho-1} \), the \( \Delta X_{\rho} + 1 \) lineages coalescing at the \( \rho \)-th merging event are chosen uniformly at random among the \( X_{\rho-1} \) present ones. For the external branches, this means that, given \( N_n \) and \( Y_0, \ldots, Y_{\rho-1} \), the decrement \( \Delta Y_{\rho} := Y_{\rho} - Y_{\rho-1} \) has a hypergeometric distribution with parameters \( X_{\rho-1}, Y_{\rho-1} \) and \( \Delta X_{\rho} + 1 \). In view of the formula of the \( i \)-th factorial moment of a hypergeometric distributed random variable, we obtain

\[
E \left[ (\Delta Y_{\rho})_i \mid N_n, Y_0, \ldots, Y_{\rho-1} \right] = (\Delta X_{\rho} + 1)_i \frac{(Y_{\rho-1})_i}{(X_{\rho-1})_i} \quad \text{a.s.} \tag{6.1}
\]

Next, we look closer at the falling factorials. We have the following binomial identity

\[
(a - b)_r = (a)_r \sum_{i=0}^{r} \binom{r}{i} (-1)^i \frac{(b)_i}{(a)_i} \tag{6.2}
\]

for \( a, b \in \mathbb{R} \) and \( r \in \mathbb{N} \). It follows from the Chu-Vandermonde identity

\[
(x + y)_r = \sum_{i=0}^{r} \binom{r}{i} (x)_i (y)_{r-i}
\]

with \( x, y \in \mathbb{R} \) and the calculation

\[
(a - b)_r = (-1)^r (b + r - 1 - a)_r \\
= (-1)^r \sum_{i=0}^{r} \binom{r}{i} (b)_i (r - 1 - a)_{r-i} \\
= (-1)^r \sum_{i=0}^{r} \binom{r}{i} (b)_i (-1)^{r-i} \frac{(a)_r}{(a)_i}.
\]

Returning to the number of external branches, we get from the identity (6.2) that

\[
(Y_{\rho})_r = (Y_{\rho-1})_r \sum_{i=0}^{r} \binom{r}{i} (-1)^i \frac{(\Delta Y_{\rho})_i}{(Y_{\rho-1})_i}.
\]

With equation (6.1), we arrive at

\[
E \left[ (Y_{\rho})_r \mid N_n, Y_0, \ldots, Y_{\rho-1} \right] = (Y_{\rho-1})_r \sum_{i=0}^{r} \binom{r}{i} (-1)^i \frac{(\Delta X_{\rho} + 1)_i}{(X_{\rho-1})_i} \quad \text{a.s.}
\]

25
Furthermore, combining the binomial identity (6.2) with the definition of \( \Delta X_\rho \), we have
\[
(X_\rho - 1)_r = (X_{\rho-1})_r \sum_{i=0}^{r} \binom{r}{i} (-1)^i \frac{(\Delta X_\rho + 1)_i}{(X_{\rho-1})_i}.
\]
Thus,
\[
\mathbb{E}[(Y_\rho)_r | N_n, Y_0, \ldots, Y_{\rho-1}] = (Y_{\rho-1})_r \frac{(X_\rho - 1)_r}{(X_{\rho-1})_r} \quad \text{a.s.}
\]
and finally
\[
\frac{\mathbb{E}[(Y_\rho)_r | N_n]}{(X_\rho)_r} = \frac{\mathbb{E}[(Y_{\rho-1})_r | N_n]}{(X_{\rho-1})_r} \frac{(X_\rho - 1)_r}{(X_{\rho-1})_r} \frac{1 - r}{X_\rho} \quad \text{a.s.}
\]
The proof now finishes by iteration and taking \( \mathbb{E}[Y_0 | N_n] = Y_0 = X_0 \) into account.

(ii) The inequality for the conditional variance follows from the representation in (i) with \( r = 1 \) and \( r = 2 \):
\[
\text{Var}(Y_\rho | N_n) = X_\rho (X_\rho - 1) \prod_{j=1}^{\rho} \left(1 - \frac{2}{X_j}\right) - X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)^2 + X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)
\]
\[
\leq X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{2}{X_j}\right) - X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)^2 + X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)
\]
\[
\leq X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right) \quad \text{a.s.}
\]
This finishes the proof. \( \square \)

7 Proof of Theorem 1.5

In order to study \( \Lambda \)-coalescents having a regularly varying rate of decrease \( \mu \) with exponent \( \alpha \in (1, 2] \), define
\[
\kappa(x) := \frac{\mu(x)}{x}, \quad x \geq 1,
\]
for convenience. For \( k \in \mathbb{N} \) and for real-valued random variables \( Z_1, \ldots, Z_k \), denote the reversed order statistics by
\[
Z_{(1)} \geq \cdots \geq Z_{(k)}.
\]
We now prove the following theorem which is equivalent to Theorem 1.5:
Theorem 7.1. Suppose that the $\Lambda$-coalescent has regularly varying rate $\mu$ with exponent $1 < \alpha \leq 2$, and fix $\ell \in \mathbb{N}$. Then, as $n \to \infty$, the following convergence holds:

$$
\kappa(s_n) \left(T^n_{(1)}, \ldots, T^n_{(\ell)}\right) \xrightarrow{d} (U_1, \ldots, U_\ell),
$$

where $U_1 > \cdots > U_\ell$ are the points in decreasing order of a Poisson point process $\Phi$ on $(0, \infty)$ with intensity measure $\phi(dx) = \alpha((\alpha - 1)x)^{-1-\alpha/\alpha} dx$.

For the rest of this section, keep the stopping times

$$
\tilde{\rho}_{c,n} := \inf \{t \geq 0 : N_n(t) \leq cs_n\}
$$

in mind, and define their discrete equivalents

$$
\rho_{c,n} := \min \{j \geq 0 : X_j \leq cs_n\}
$$

for $c > 0$. Note that

$$
\mathbb{E}[\tilde{\rho}_{c,n}] \to 0
$$

as $n \to \infty$. This follows from Lemma 3 (i) of [11] and the estimate

$$
\int_{cs_n}^n \frac{dx}{\mu(x)} = \mathcal{O} \left( \int_{cs_n}^n x^{-\alpha+\epsilon} dx \right) = \mathcal{O} \left(s_n^{1-\alpha+\epsilon}\right) = o(1)
$$

for $0 < \epsilon < \alpha - 1$, because $\mu$ is regularly varying with exponent $\alpha$.

The next lemma deals with properties of the stopping times from (7.1) and (7.2). In particular, it reveals that external branches are still present up to the times $\tilde{\rho}_{c,n}$.

Lemma 7.2. Assume the $\Lambda$-coalescent has regularly varying rate $\mu$ with exponent $\alpha \in (1, 2]$. Then we have:

(i) For each $\epsilon > 0$, there exists $c_\epsilon > 0$ such that for all $c \geq c_\epsilon$,

$$
\lim_{n \to \infty} \mathbb{P} \left( \kappa(s_n) \tilde{\rho}_{c,n} \geq \epsilon \right) = 0.
$$

(ii) For each $c > 0$, as $n \to \infty$,

$$
X_{\rho_{c,n}} = cs_n + o_p(s_n).
$$

(iii) For each $\epsilon > 0$,

$$
\limsup_{n \to \infty} \mathbb{P} \left( \left| e^{-\alpha} Y_{\rho_{c,n}} - 1 \right| \geq \epsilon \right) \xrightarrow{c \to \infty} 0.
$$
Proof. (i) Because $\mu$ is regularly varying with exponent $\alpha > 1$, we have

$$
\int_{cs_n}^{\infty} \frac{dx}{\mu(x)} \sim \frac{1}{\alpha - 1} \frac{cs_n}{\mu(cs_n)} \sim \frac{1}{\alpha - 1} e^{1-\alpha} \frac{1}{\kappa(s_n)}
$$

as $n \to \infty$. In view of (1.3) and (7.3), we may apply Proposition 2 of [11] with $f \equiv 1$ saying that $\tilde{\rho}_{c,n} = \int_{cs_n}^{n} \frac{dx}{\mu(x)} (1 + \sigma_P(1))$ as $n \to \infty$. This implies that

$$
\kappa(s_n) \tilde{\rho}_{c,n} \leq \frac{1}{\alpha - 1} e^{1-\alpha}(1 + \sigma_P(1)),
$$

which entails the claim.

(ii) Because of (7.3), we may use Lemma 3 (ii) of [11]. In conjunction with the definition of $\rho_{c,n}$, we therefore obtain

$$
\frac{X_{\rho_{c,n}}}{X_{\rho_{c,n}-1}} = 1 - \frac{\Delta X_{\rho_{c,n}}}{X_{\rho_{c,n}-1}} = 1 + \sigma_P(1)
$$

as $n \to \infty$. This implies the statement because of $X_{\rho_{c,n}} \leq cs_n < X_{\rho_{c,n}-1}$.

(iii) We first prove that

$$
\mathbb{E}[Y_{\rho_{c,n}} | N_n] = c^\alpha + \sigma_P(1) \quad (7.4)
$$

as $n \to \infty$. Lemma 6.1 (i), together with a Taylor expansion as in (3.4), provides

$$
\mathbb{E}[Y_{\rho_{c,n}} | N_n] = (X_{\rho_{c,n}} - 1) \exp \left( - \sum_{j=0}^{\rho_{c,n}-1} \frac{1}{X_j} + \mathcal{O} \left( \frac{X_{\rho_{c,n}-1}^{-1}}{X_{\rho_{c,n}}} \right) \right)
$$

as $n \to \infty$. Furthermore, (7.3) allows us to apply Proposition 3 of [11] yielding

$$
\sum_{j=0}^{\rho_{c,n}-1} \frac{1}{X_j} = \log \left( \frac{\kappa(n)}{\kappa(X_{\rho_{c,n}})} \right) + \sigma_P(1) \quad (7.5)
$$

as $n \to \infty$. Combining statement (ii) with Lemma 2.1 (ii), we therefore arrive at

$$
\mathbb{E}[Y_{\rho_{c,n}} | N_n] = n \frac{\mu(X_{\rho_{c,n}})}{\mu(n)} (1 + \sigma_P(1)) = n \frac{\mu(cs_n)}{\mu(n)} (1 + \sigma_P(1))
$$

so that the regular variation of $\mu$ and the definition of $s_n$ imply (7.4). Thus, in the upper bound

$$
P \left( |Y_{\rho_{c,n}} - c^\alpha| \geq \varepsilon c^\alpha \right) \leq P \left( |\mathbb{E}[Y_{\rho_{c,n}} | N_n] - c^\alpha| \geq \frac{\varepsilon}{2} c^\alpha \right)
$$

$$
+ P \left( |Y_{\rho_{c,n}} - \mathbb{E}[Y_{\rho_{c,n}} | N_n]| \geq \frac{\varepsilon}{2} c^\alpha \right)
$$

28
with $\varepsilon > 0$, the first right-hand probability converges to 0. For the second one, Chebyshev’s inequality and Lemma 6.1 (ii) imply that

$$
P(\left| Y_{\rho c,n} - E[Y_{\rho c,n} | N_n] \right| \geq \varepsilon c^\alpha) = E\left[ P(\left| Y_{\rho c,n} - E[Y_{\rho c,n} | N_n] \right| \geq \varepsilon c^\alpha | N_n) \right]
$$

$$\leq E \left[ \frac{\text{Var}(Y_{\rho c,n} | N_n)}{\varepsilon^2 c^{2\alpha}} \wedge 1 \right]
$$

$$\leq E \left[ \frac{E[Y_{\rho c,n} | N_n]}{\varepsilon^2 c^{2\alpha}} \wedge 1 \right].$$

From (7.4) and dominated convergence, we conclude

$$P(\left| Y_{\rho c,n} - E[Y_{\rho c,n} | N_n] \right| \geq \varepsilon c^\alpha) \leq \varepsilon^{-2} c^{-\alpha} + o(1)$$

as $n \to \infty$, which gives the claim. \(\square\)

For the following lemma, recall the subdivided external branch lengths

$$\tilde{T}_i^n := T_i^n \wedge \tilde{\rho}_{c,n} \quad \text{and} \quad \hat{T}_i^n := T_i^n - \tilde{T}_i^n$$

for $1 \leq i \leq n$, and let

$$\beta := \frac{\alpha - 1}{\alpha}.$$

**Lemma 7.3.** Suppose the $\Lambda$-coalescent has regularly varying rate $\mu$ with exponent $\alpha \in (1, 2]$. Then for $\ell, y \in \mathbb{N}$, there exist random variables $U_{1,y} \geq \ldots \geq U_{\ell,y}$ such that the following convergence results hold:

(i) For fixed $y \geq \ell$, as $n \to \infty$,

$$\mathcal{L}\left( \kappa(cs_n) \left( \hat{T}^n_{(1)}, \ldots, \hat{T}^n_{(\ell)} \right) \left| Y_{\rho c,n} = y, X_{\rho c,n} \right. \right) \to \mathcal{L}(U_{1,y}, \ldots, U_{\ell,y})$$

in probability, where convergence takes place in the space of probability measures on $\mathbb{R}^\ell$.

(ii) For fixed $\ell \in \mathbb{N}$, as $y \to \infty$,

$$y^{-\beta} (U_{1,y}, \ldots, U_{\ell,y}) \overset{d}{\to} (U_1, \ldots, U_\ell),$$

where $U_1 > \ldots > U_\ell$ are the points of the Poisson point process of Theorem 7.1.
Proof. (i) Observe that due to the strong Markov property, given $X_{\rho,c,n}$ and the event $Y_{\rho,c,n} = y$, the $y$ remaining external branches evolve as $y$ ordinary external branches out of a sample of $X_{\rho,c,n}$ many individuals. From these $y$ external branches, we now consider the $\ell$ largest ones. In view of Lemma 7.2 (i), we have $X_{\rho,c,n} \rightarrow \infty$ in probability as $n \rightarrow \infty$. Hence, Corollary 1.4, together with established formulas for order statistics of i.i.d random variables, yields that

$$\ell! \binom{y}{\ell} F(\alpha(y \beta u_\ell))^{y-\ell} \prod_{i=1}^{\ell} f(\alpha(y \beta u_i)) du_1 \cdots du_\ell, \quad (7.6)$$

with $u_1 \geq \cdots \geq u_\ell \geq 0$, is the limit of the conditional distributions of $\kappa(X_{\rho,c,n})(\hat{T}_1,\ldots,\hat{T}_\ell)$ as $n \rightarrow \infty$, where $f$ is the density from formula (1.1) and $F$ its cumulative distribution function. Our claim now follows due to Lemma 7.2 (ii) and Lemma 2.1 (ii).

(ii) Note that

$$y^{\beta+1} f(y^\beta u) = y^{\beta+1} \alpha \left(1 + (\alpha - 1)uy^\beta\right)^{-1-1/\beta} \overset{y \rightarrow \infty}{\longrightarrow} \alpha ((\alpha - 1)u)^{-1-1/\beta}$$

and

$$F(y^\beta u)^{y-\ell} = \left[1 - \left(1 + (\alpha - 1)y^\beta u\right)^{-1/\beta}\right]^{y-\ell} \overset{y \rightarrow \infty}{\longrightarrow} \exp\left(-((\alpha - 1)u)^{-1/\beta}\right).$$

Consequently,

$$\ell! \binom{y}{\ell} F(y^\beta u_\ell)^{y-\ell} \prod_{i=1}^{\ell} \left[f(y^\beta u_i) y^\beta du_i\right],$$

being the density of $y^{-\beta}(U_1,\ldots,U_\ell)$, has the limit

$$\exp\left(-((\alpha - 1)u_\ell)^{-1/\beta}\right) \prod_{i=1}^{\ell} \alpha ((\alpha - 1)u_i)^{-1-1/\beta} du_1 \cdots du_\ell$$

as $y \rightarrow \infty$. Indeed, this is the joint density of the rightmost points $U_1 > \cdots > U_\ell$ of the Poisson point process given in Theorem 7.1. \hfill \Box

**Proof of Theorem 7.1.** For convenience, we use in this proof the notation

$$V_{c,n} := \kappa(c_{n\rho}) \left(\hat{T}_1,\ldots,\hat{T}_\ell\right).$$

Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuous function, and assume that $\max |g| \leq 1$. For $c > 0$, we obtain
via the law of total expectation and Lemma 7.3 (i) that
\[
\left| \mathbb{E} \left[ g \left( Y_{\rho,c,n}^{-\beta} V_{c,n} \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| \\
\leq \sum_{c/2 \leq y \leq 2c} \left| \mathbb{E} \left[ g \left( y^{-\beta} V_{c,n} \right) \left| Y_{\rho,c,n} = y, X_{\rho,c,n} \right. \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \cdot \mathbb{P} \left( \left| Y_{\rho,c,n} - c^\alpha \right| \geq c^\alpha/2 \right| X_{\rho,c,n} \right) \\
+ 2 \mathbb{P} \left( \left| Y_{\rho,c,n} - c^\alpha \right| \geq c^\alpha/2 \right| X_{\rho,c,n} \right) \\
\leq \max_{c/2 \leq y \leq 2c} \left| \mathbb{E} \left[ g \left( y^{-\beta} V_{c,n} \right) \left| Y_{\rho,c,n} = y, X_{\rho,c,n} \right. \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| \\
+ 2 \mathbb{P} \left( \left| Y_{\rho,c,n} - c^\alpha \right| \geq c^\alpha/2 \right| X_{\rho,c,n} \right) \\
\leq \max_{c/2 \leq y \leq 2c} \mathbb{E} \left[ g \left( y^{-\beta} U_1, \ldots, y^{-\beta} U_\ell, y \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] + o_P(1) \\
+ 2 \mathbb{P} \left( \left| Y_{\rho,c,n} - c^\alpha \right| \geq c^\alpha/2 \right) \\
a.s.
\]

as \( n \to \infty \). Without loss of generality, we may assume that the \( o_P(\cdot) \) term is bounded by 1. Hence, taking expectations, applying Jensen’s inequality to the left-hand side and using dominated convergence, we obtain

\[
\left| \mathbb{E} \left[ g \left( Y_{\rho,c,n}^{-\beta} V_{c,n} \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| \\
\leq \max_{c/2 \leq y \leq 2c} \mathbb{E} \left[ g \left( y^{-\beta} U_1, \ldots, y^{-\beta} U_\ell, y \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] + o(1) \\
+ 2 \mathbb{P} \left( \left| Y_{\rho,c,n} - c^\alpha \right| \geq c^\alpha/2 \right)
\]
as \( n \to \infty \). Then Lemma 7.3 (ii) and Lemma 7.2 (iii) entail

\[
\limsup_{n \to \infty} \left| \mathbb{E} \left[ g \left( Y_{\rho,c,n}^{-\beta} V_{c,n} \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| \xrightarrow{c.s.} 0.
\]

(7.7)

After this preparatory work, we now additionally assume that \( g \) is a Lipschitz continuous function with Lipschitz constant 1 (in each coordinate) and prove that

\[
\mathbb{E} \left[ g \left( \kappa(s_n) T_{(1)}^n, \ldots, \kappa(s_n) T_{(\ell)}^n \right) \right] \xrightarrow{n \to \infty} \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right],
\]

(7.8)

which implies the theorem’s statement. For \( \varepsilon > 0 \), we have

\[
\left| \mathbb{E} \left[ g \left( \kappa(s_n) T_{(1)}^n, \ldots, \kappa(s_n) T_{(\ell)}^n \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| \\
\leq \left| \mathbb{E} \left[ g \left( \kappa(s_n) \hat{T}_{(1)}^n, \ldots, \kappa(s_n) \hat{T}_{(\ell)}^n \right) \right] - \mathbb{E} \left[ g \left( U_1, \ldots, U_\ell \right) \right] \right| + \sum_{i=1}^\ell \mathbb{E} \left[ \kappa(s_n) \hat{T}_{(i)}^n \vee 2 \right]
\]

31
\[
\begin{align*}
&\leq |E[g\left(Y^{-\beta}_{\rho_{c,n}} V_{c,n}\right)] - E[g(U_1,\ldots,U_\ell)]| \\
&\quad + \sum_{i=1}^{\ell} E\left[\left|\left(Y^{-\beta}_{\rho_{c,n}} \kappa(cs_n) - \kappa(s_n)\right)\tilde{T}_n^{(i)}\right|\wedge 2\right] \\
&\quad + \ell E\left[\kappa(s_n)\tilde{T}_n^{(1)}\wedge 2\right] \\
&\leq |E[g\left(Y^{-\beta}_{\rho_{c,n}} V_{c,n}\right)] - E[g(U_1,\ldots,U_\ell)]| \\
&\quad + \ell E\left[\left(\varepsilon \kappa(cs_n)Y^{-\beta}_{\rho_{c,n}}\tilde{T}_n^{(1)}\right)\wedge 2\right] + 2\ell P\left(\left|Y^{-\beta}_{\rho_{c,n}} \kappa(cs_n) - \kappa(s_n)\right| \geq \varepsilon \kappa(cs_n)Y^{-\beta}_{\rho_{c,n}}\right) \\
&\quad + \ell \varepsilon + 2\ell P\left(\kappa(s_n)\tilde{T}_n^{(1)} \geq \varepsilon\right)
\end{align*}
\]

and consequently

\[
\begin{align*}
&\limsup_{n \to \infty} |E[g(\kappa(s_n) T_n^{(1)},\ldots,\kappa(s_n) T_n^{(\ell)})] - E[g(U_1,\ldots,U_\ell)]| \\
&\quad \leq \limsup_{n \to \infty} |E[g\left(Y^{-\beta}_{\rho_{c,n}} V_{c,n}\right)] - E[g(U_1,\ldots,U_\ell)]| \\
&\quad + \ell \limsup_{n \to \infty} E\left[\left(\varepsilon \kappa(cs_n)Y^{-\beta}_{\rho_{c,n}}\tilde{T}_n^{(1)}\right)\wedge 2\right] - E\left[\left(\varepsilon U_1\right)\wedge 2\right] + \ell E\left[\left(\varepsilon U_1\right)\wedge 2\right] \\
&\quad + 2\ell \limsup_{n \to \infty} P\left(1 - \frac{\kappa(s_n)}{\kappa(cs_n)} Y^{-\beta}_{\rho_{c,n}} \geq \varepsilon\right) \\
&\quad + \ell \varepsilon + 2\ell \limsup_{n \to \infty} P\left(\kappa(s_n)\tilde{\rho}_{c,n} \geq \varepsilon\right).
\end{align*}
\]

We now use (7.7) for the first two right-hand terms and Lemma 7.2 (iii) for the first probability taking \(\kappa(cs_n)/\kappa(s_n) \sim c^{\alpha-1} = c^{\alpha \beta}\) into account. To the other probability, we apply Lemma 7.2 (i). Hence, passing to the limit as \(c \to \infty\) yields

\[
\limsup_{n \to \infty} |E[g(\kappa(s_n) T_n^{(1)},\ldots,\kappa(s_n) T_n^{(\ell)})] - E[g(U_1,\ldots,U_\ell)]| \leq \ell E\left[\left(\varepsilon U_1\right)\wedge 2\right] + \ell \varepsilon.
\]

Finally, taking the limit \(\varepsilon \to 0\) and using dominated convergence gives the claim. \(\square\)

8 Proof of Theorem 1.6

Recall the notation of the reversed order statistics \(Z_{(1)} \geq Z_{(2)} \geq \cdots\) of real-valued random variables as introduced in the previous section and the definition

\[
t_n := \log \log n - \log \log \log n + \log \log \log n/\log \log n.
\]

In this section, we prove the following equivalent version of Theorem 1.6:
Theorem 8.1. For the Bolthausen-Sznitman coalescent, the following convergence holds: For \( \ell \in \mathbb{N} \),

\[
\log \log n \left( T_{(1)}^n - t_n \ldots T_{(\ell)}^n - t_n \right) \xrightarrow{d} (U_1 - G, \ldots, U_\ell - G)
\]

as \( n \to \infty \), where \( U_1 > \cdots > U_\ell \) are the \( \ell \) maximal points in decreasing order of a Poisson point process on \( \mathbb{R} \) with intensity measure \( e^{-x} \, dx \) and \( G \) is an independent standard Gumbel distributed random variable.

Recall for \( c > 1 \) the notion

\[
t_{c,n} := t_n - \frac{\log c}{\log \log n}.
\]

Lemma 8.2. Let \( E \) be a standard exponential random variable. Then, as \( n \to \infty \), we have for \( c > 1 \),

\[
e^{-t_{c,n}} N_n(t_{c,n}) \xrightarrow{d} cE.
\]

Proof. We first consider \( N_n(t)^{(r)} := N_n(t)(N_n(t)+1)\cdots(N_n(t)+r-1) \) for \( r \in \mathbb{N} \). For these ascending factorials, Lemma 3.1 of [24] provides

\[
\mathbb{E} \left[ N_n(t)^{(r)} \right] = \frac{\Gamma (r + 1) \, \Gamma (n + r - t)}{\Gamma (1 + r - t) \, \Gamma (n)}.
\]

The Sterling approximation with remainder term yields uniformly in \( t \geq 0 \),

\[
\frac{\Gamma (n + r - t)}{\Gamma (n)} = n^{r-t} (1 + \Theta (1)),
\]

and consequently

\[
\mathbb{E} \left[ N_n(t)^{(r)} \right] = \frac{\Gamma (r + 1)}{\Gamma (1 + r - t)} \, n^{r-t} (1 + \Theta (1))
\]

uniformly in \( t \geq 0 \) as \( n \to \infty \). Inserting \( t_{c,n} \) in this equation entails

\[
n^{-r-t_{c,n}} \mathbb{E} \left[ N_n(t_{c,n})^{(r)} \right] \to r!
\]

as \( n \to \infty \).

Now observe

\[
e^{-t_{c,n}} \log n = \exp \left( -\frac{\log \log n}{\log \log n} + \frac{\log c}{\log \log n} \right) \log \log n
\]

\[
= \log \log n - \log \log \log n + \log c + \Theta (1)
\]

\[
= t_{c,n} + \log c + \Theta (1).
\]
Equivalently, 
\[ n^{e^{-tc,n}} = ce^{tc,n}(1 + o(1)), \]
and therefore
\[ e^{-rtc,n}E[N_{n}(tc,n)^{(r)}] \to c^{r}r! \quad (8.1) \]
as \( n \to \infty \).

Furthermore, because of
\[ N_{n}(t)^{(r)} \leq N_{n}(t)^{(r)} \leq N_{n}(t)^{(r)} + 2^{r}r^{r}N_{n}(t)^{(r-1)}, \]
we have
\[ N_{n}(t)^{(r)} - 2^{r}r^{r}N_{n}(t)^{(r-1)} \leq N_{n}(t)^{(r)} \leq N_{n}(t)^{(r)}. \]
Thus, (8.1) transfers to
\[ e^{-rtc,n}E[N_{n}(tc,n)^{(r)}] \to c^{r}r! \]
as \( n \to \infty \), and our claim follows by method of moments.

The following lemma provides the asymptotic behavior of the joint probability distribution of
the lengths of the longest external branches starting at time \( tc,n \). Let
\[ M_{n}(t) := \# \{ i \geq 1 : \{ i \} \in \Pi_{n}(t) \}, \quad t \geq 0, \]
which is the number of external branches at time \( t \). Also recall
\[ \hat{T}_{(i)}^{m} := (T_{(i)}^{m} - tc,n)^{+}. \]

**Lemma 8.3.** For \( \ell, m \in \mathbb{N} \), there exist random variables \( U_{1,m} \geq \cdots \geq U_{\ell,m} \) such that the following convergence results hold:

(i) For fixed natural numbers \( \ell \leq m \), as \( n \to \infty \),
\[ \mathcal{L} \left( \log \log (n) \left( \hat{T}_{(1)}^{m}, \ldots, \hat{T}_{(\ell)}^{m} \right) \mid M_{n}(tc,n) = m, N_{n}(tc,n) \right) \to \mathcal{L} (U_{1,m}, \ldots, U_{\ell,m}) \]
in probability, where convergence takes place in the space of probability measures on \( \mathbb{R}^{\ell} \).

(ii) For fixed \( \ell \), as \( m \to \infty \),
\[ (U_{1,m} - \log m, \ldots, U_{\ell,m} - \log m) \overset{d}{\to} (U_{1}, \ldots, U_{\ell}), \]
where \( U_{1} > \cdots > U_{\ell} \) are the points of the Poisson point process of Theorem 8.1.
Proof. (i) We proceed in the same vein as in the proof of Lemma 7.3 (i). The strong Markov property, Corollary 1.4 (see also formula (1.2) in the first example) and Lemma 8.2 yield that the random vector \( \log N_n(t_{c,n}) \left( \hat{T}_{n,1}, \ldots, \hat{T}_{n,\ell} \right) \), given \( N_n(t_{c,n}) \) and the event \( M_{t_{c,n}} = m \), has a limiting distribution as \( n \to \infty \) with density

\[
\ell! \left( \frac{m}{\ell} \right) \left( 1 - e^{-u} \right)^{m-\ell} \prod_{i=1}^{\ell} e^{-u_i} du_1 \cdots du_\ell
\]

for \( u_1 \geq \cdots \geq u_\ell \). Moreover, from Lemma 8.2, we obtain

\[
\log (N_n(t_{c,n})) = t_{c,n} + O_P(1) = \log \log n + o_P(\log \log n)
\]
as \( n \to \infty \). This implies our claim.

(ii) Shifting the distribution from (8.2) by \( \log m \), we arrive at the densities

\[
\ell! \left( \frac{m}{\ell} \right) \left( 1 - e^{-u} \right)^{m-\ell} \prod_{i=1}^{\ell} e^{-u_i} du_1 \cdots du_\ell
\]

and its limit

\[
e^{-e^{-u}} \prod_{i=1}^{\ell} e^{-u_i} du_i
\]
as \( m \to \infty \), which is the joint density of \( U_1, \ldots, U_\ell \). This finishes the proof.

Next, we introduce the notion

\[
\rho_{c,n} := \min \left\{ k \geq 1 : \sum_{j=0}^{k-1} W_j > t_{c,n} \right\} \wedge \tau_n.
\]

It is important to note that in the case of the Bolthausen-Sznitman coalescent (7.3) is no longer valid, and consequently we may not simply apply (7.5). As a substitute, we shall use the following lemma.

**Lemma 8.4.** As \( n \to \infty \),

\[
\sum_{j=0}^{\rho_{c,n}-1} \frac{1}{X_j} = t_{c,n} + o_P(1).
\]

**Proof.** Let \( \mathcal{F}_k := \sigma(X, W_0, \ldots, W_{k-1}) \) and

\[
Z_k := \sum_{j=0}^{k/\tau_n-1} \left( W_j - \frac{1}{X_j - 1} \right), \quad k \geq 0.
\]
In particular, we have $Z_0 = 0$. Given $F_j$ and $X_j = b$ with $b \geq 2$, the waiting time $W_j$ in the Bolthausen-Sznitman coalescent is exponential with rate parameter $b - 1$ (see (47) in [27]). Thus, $(Z_k)_{k \in \mathbb{N}}$ is a martingale with respect to the filtration $(F_k)_{k \in \mathbb{N}}$ with (predictable) quadratic variation

$$
(Z)_k^2 := \sum_{j=0}^{k \wedge \tau_n-1} E \left[ (Z_j - Z_j)^2 | F_j \right] = \sum_{j=0}^{k \wedge \tau_n-1} \frac{1}{(X_j - 1)^2} \quad \text{a.s.}
$$

Applying Doob’s optional sampling theorem to the martingale $Z^2_k - (Z)_k$ yields

$$
E \left[ Z^2_{\rho_{c,n}} \right] = E \left[ (Z)_{\rho_{c,n}} \right] = E \left[ \sum_{j=0}^{\rho_{c,n}-1} \frac{1}{(X_j - 1)^2} \right] \leq E \left[ \sum_{k=X_{\rho_{c,n}-1}}^{\infty} \frac{1}{(k - 1)^2} \right] \quad \text{(8.3)}
$$

and therefore, because of $X_{\rho_{c,n}-1} = N_n(t_{c,n})$ a.s.,

$$
E \left[ Z^2_{\rho_{c,n}} \right] \leq E \left[ \frac{4}{N_n(t_{c,n})} \right].
$$

By Lemma 8.2 and dominated convergence, the right-hand term converges to 0 as $n \to \infty$ implying

$$
\sum_{j=0}^{\rho_{c,n}-1} \left( W_j - \frac{1}{X_j} \right) = Z_{\rho_{c,n}} + O_P \left( \frac{4}{X_{\rho_{c,n}-1}} \right) = o_P(1)
$$
as $n \to \infty$. Finally, the quantity $\sum_{j=0}^{\rho_{c,n}-1} W_j - t_{c,n}$ is the residual time the process $N_n$ spends in the state $N_n(t_{c,n})$. Due to the property that exponential times lack memory, the residual time is exponential with parameter $N_n(t_{c,n})$. Thus, in view of Lemma 8.2, the residual time converges to 0 in probability. This finishes the proof. \qed

**Lemma 8.5.** For the number of external branches at time $t_{c,n}$, we have the following results:

(i) For $c > 1$,

$$
E \left[ M_n(t_{c,n}) \mid N_n \right] \xrightarrow{d} c E
$$
as $n \to \infty$, where $E$ denotes a standard exponential random variable.

(ii) For $\varepsilon > 0$, as $c \to \infty$,

$$
\limsup_{n \to \infty} P \left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) \mid N_n]| > c^{1/2+\varepsilon} \right) \to 0
$$
as well as

$$
\limsup_{n \to \infty} P \left( M_n(t_{c,n}) > c^{1+\varepsilon} \right) \to 0 \quad \text{and} \quad \limsup_{n \to \infty} P \left( M_n(t_{c,n}) < c^{1-\varepsilon} \right) \to 0.
$$

36
Proof. (i) Using the representation from Lemma 6.1 (i) and a Taylor expansion as in (3.4), we get
\[
E[Y_{\rho_{c,n}^{-1}, N_n}] = X_{\rho_{c,n}^{-1}} \exp \left( - \sum_{j=1}^{\rho_{c,n}^{-1}} \frac{1}{X_j} + O_P \left(X_{\rho_{c,n}^{-1}}^{-1}\right) \right)
\]
as \(n \to \infty\). Recall that the definition of \(\rho_{c,n}\) entails \(N_n(t_{c,n}) = X_{\rho_{c,n}^{-1}}\) and \(M_n(t_{c,n}) = Y_{\rho_{c,n}^{-1}}\) a.s. Thus, we obtain
\[
E[M_n(t_{c,n}) | N_n] = N_n(t_{c,n}) \exp \left( - \sum_{j=1}^{\rho_{c,n}^{-1}} \frac{1}{X_j} + O_P \left(N_n(t_{c,n})^{-1}\right) \right).
\]
(8.4)
From Lemma 8.4 and Lemma 8.2, it follows
\[
E[M_n(t_{c,n}) | N_n] = N_n(t_{c,n}) \exp (-t_{c,n} + O_P (1)).
\]

Hence, Lemma 8.2 implies our claim.

(ii) Chebyshev’s inequality and Lemma 6.1 (ii) provide
\[
P \left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) | N_n]| > c^{1/2+\varepsilon} \right)
\]
\[
= E \left[ P \left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) | N_n]| > c^{1/2+\varepsilon} \right) | N_n \right]
\]
\[
\leq E \left[ \frac{\Var(M_n(t_{c,n}) | N_n)}{c^{1+2\varepsilon}} \wedge 1 \right]
\]
\[
\leq E \left[ \frac{E(M_n(t_{c,n}) | N_n)}{c^{1+2\varepsilon}} \wedge 1 \right].
\]
From statement (i) it follows that
\[
\limsup_{n \to \infty} P \left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) | N_n]| > c^{1/2+\varepsilon} \right) \leq E \left[ \frac{cE}{c^{1+2\varepsilon}} \wedge 1 \right] \leq c^{-2\varepsilon},
\]
which yields the first claim.

Similarly, Markov’s inequality yields
\[
\limsup_{n \to \infty} P \left( M_n(t_{c,n}) > c^{1+\varepsilon} \right) \leq \limsup_{n \to \infty} E \left[ \frac{E(M_n(t_{c,n}) | N_n)}{c^{1+\varepsilon}} \wedge 1 \right] \leq c^{-\varepsilon}
\]
giving the second claim.
Further, we have

\[
P(M_n(t_{c,n}) < c^{1-\varepsilon}) \leq P\left( E[M_n(t_{c,n}) | N_n] < 2c^{1-\varepsilon} \right)
+ P\left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) | N_n]| > c^{1-\varepsilon} \right),
\]

and consequently, in view of part (i),

\[
\limsup_{n \to \infty} P(M_n(t_{c,n}) < c^{1-\varepsilon}) \leq P(E < 2c^{-\varepsilon}) + \limsup_{n \to \infty} P\left( |M_n(t_{c,n}) - E[M_n(t_{c,n}) | N_n]| > c^{1-\varepsilon} \right).
\]

The first right-hand term converges to 0 as \( c \to \infty \). Also, as we may assume \( \varepsilon < 1/2 \), the second term goes to 0 in view of the first claim of part (ii).

With these preparations, we now turn to Theorem 8.1.

**Proof of Theorem 8.1.** Define the functions

\[
g(x_1, \ldots, x_{\ell}) := \exp(i(\theta_1 x_1 + \cdots + \theta_{\ell} x_{\ell})) \quad \text{and} \quad h(x) := \exp(i(\theta_1 + \cdots + \theta_\ell)x),
\]

where \( \theta_i \in \mathbb{R} \) for \( 1 \leq i \leq n \), and let

\[
V_{c,n} := \left( \log \log (n) \left( T_{\langle 1 \rangle} \right) - \log (T_{\langle 1 \rangle} - t_{c,n}), \ldots, \log \log (n) \left( T_{\langle \ell \rangle} \right) - \log (T_{\langle \ell \rangle} - t_{c,n}) \right).
\]

Recalling

\[
t_n = t_{c,n} + \frac{\log c}{\log \log n},
\]

we see that, on the event \( \{M_n(t_{c,n}) \geq \ell\} \), it holds \( T_{\langle j \rangle}^n = \hat{T}_{\langle j \rangle}^n + t_{c,n} \) and therefore

\[
\log \log (n)(T_{\langle j \rangle}^n - t_n) = \left( \log \log (n) \hat{T}_{\langle j \rangle}^n - \log (T_{\langle j \rangle} - t_{c,n}) \right) + \log \frac{M_n(t_{c,n})}{c}
\]

for \( 1 \leq j \leq \ell \). In conjunction with the independence of \((U_1, \ldots, U_{\ell})\) and the Gumbel random variable \( G \), it follows that

\[
|E \left[ g \left( \log \log (n) \left( T_{\langle 1 \rangle} - t_n \right), \ldots, \log \log (n) \left( T_{\langle \ell \rangle} - t_n \right) \right) \right] - E[g(U_1 - G, \ldots, U_\ell - G)]| \leq \left| E \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - E[g(U_1, \ldots, U_\ell)] E[h(-G)] \right| \quad \text{(8.5)}
+ 2P(M_n(t_{c,n}) < \ell).
\]
Lemma 8.3 (i)

We bound the three differences on the right-hand side. First, for \( c > 1 \), we have by means of Lemma 8.3 (i)

\[
\left| \mathbb{E} \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right| \\
\leq \left| \mathbb{E} \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right| \\
+ \left| \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right| \\
+ \left| \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right|.
\] (8.6)

We bound the three differences on the right-hand side. First, for \( c > 1 \), we have by means of Lemma 8.3 (i)

\[
\left| \mathbb{E} \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right| \\
\leq \sum_{\sqrt{c} \leq m \leq c^2} \left| \left( \mathbb{E} \left[ g(V_{c,n}) \mid M_n(t_{c,n}) = m, N_n(t_{c,n}) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \right) h \left( \log \frac{m}{c} \right) \right| \\
\cdot \mathbb{P} \left( M_n(t_{c,n}) = m \mid N_n(t_{c,n}) \right) \\
+ 2 \mathbb{P} \left( M_n(t_{c,n}) < \sqrt{c} \mid N_n(t_{c,n}) \right) + 2 \mathbb{P} \left( M_n(t_{c,n}) > c^2 \mid N_n(t_{c,n}) \right)
\]

\[
\leq \max_{\sqrt{c} \leq m \leq c^2} \left| \mathbb{E} \left[ g(U_{1,m} - \log m, \ldots, U_{\ell,m} - \log m) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \right| + \mathbb{O}(1)
\]

\[ + \sum_{\sqrt{c} \leq m \leq c^2} \mathbb{P} \left( M_n(t_{c,n}) < \sqrt{c} \mid N_n(t_{c,n}) \right) + 2 \mathbb{P} \left( M_n(t_{c,n}) > c^2 \mid N_n(t_{c,n}) \right) \]

as \( n \to \infty \). Without loss of generality, we may assume that the right-hand \( \mathbb{O}(\cdot) \) term is bounded by 1. Hence, taking expectations, we obtain via dominated convergence

\[
\left| \mathbb{E} \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] \right| \\
\leq \max_{\sqrt{c} \leq m \leq c^2} \left| \mathbb{E} \left[ g(U_{1,m} - \log m, \ldots, U_{\ell,m} - \log m) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_{\ell}) \right] \right| + \mathbb{O}(1) \quad (8.7)
\]

\[ + \sum_{\sqrt{c} \leq m \leq c^2} \mathbb{P} \left( M_n(t_{c,n}) < \sqrt{c} \right) + 2 \mathbb{P} \left( M_n(t_{c,n}) > c^2 \right). \]
Second, observe that the function \( h(\log x) \) is Lipschitz on the interval \([c^{-1/4}, \infty)\) with Lipschitz constant \( |\theta_1 + \cdots + \theta_\ell| c^{1/4} \). Thus,

\[
\left| \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ h \left( \log \mathbb{E} \left[ \frac{M_n(t_{c,n})}{c} | N_n \right] \right) \right] \right|
\]

\[
\leq \left| \mathbb{E} \left[ h \left( \log \frac{M_n(t_{c,n})}{c} \right) - h \left( \log \mathbb{E} \left[ \frac{M_n(t_{c,n})}{c} | N_n \right] \right) \right] ; M_{t_{c,n}} \wedge \mathbb{E} \left[ M_n(t_{c,n}) | N_n \right] \geq c^{3/4} \right|
\]

\[
+ 2 \mathbb{P} \left( M_{t_{c,n}} < c^{3/4} \right) + 2 \mathbb{P} \left( \mathbb{E} \left[ M_n(t_{c,n}) | N_n \right] < c^{3/4} \right)
\]

\[
\leq 2 \mathbb{P} \left( |M_n(t_{c,n}) - \mathbb{E} \left[ M_n(t_{c,n}) | N_n \right] | > c^{2/3} \right) + |\theta_1 + \cdots + \theta_\ell| c^{1/4 - 1/3}
\tag{8.8}
\]

\[
+ 2 \mathbb{P} \left( M_{t_{c,n}} < c^{3/4} \right) + 2 \mathbb{P} \left( \mathbb{E} \left[ M_n(t_{c,n}) | N_n \right] < c^{3/4} \right).
\]

Last, Lemma 8.5 (i) provides the convergence of the third difference of (8.6) to 0 as \( n \to \infty \) so that combining equation (8.5) to (8.8), using Lemma 8.5 and grouping terms yield

\[
\lim sup_{n \to \infty} \left| \mathbb{E} \left[ g(V_{c,n}) h \left( \log \frac{M_n(t_{c,n})}{c} \right) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_\ell) \right] \mathbb{E} \left[ h (-G) \right] \right|
\]

\[
\leq \max_{\sqrt{c} \leq m \leq c^2} \left| \mathbb{E} \left[ g(U_{1,m} - \log m, \ldots, U_{\ell,m} - \log m) \right] - \mathbb{E} \left[ g(U_1, \ldots, U_\ell) \right] \right|
\]

\[
+ 2 \lim sup_{n \to \infty} \mathbb{P} \left( M_{t_{c,n}} < \ell \right) + 2 \lim sup_{n \to \infty} \mathbb{P} \left( M_n(t_{c,n}) < \sqrt{c} \right) + 2 \lim sup_{n \to \infty} \mathbb{P} \left( M_n(t_{c,n}) < c^{3/4} \right)
\]

\[
+ 2 \lim sup_{n \to \infty} \mathbb{P} \left( M_n(t_{c,n}) > c^2 \right) + 2 \lim sup_{n \to \infty} \mathbb{P} \left( |M_n(t_{c,n}) - \mathbb{E} \left[ M_n(t_{c,n}) | N_n \right] | > c^{2/3} \right)
\]

\[
+ 2 \left( 1 - e^{-c^{-1/4}} \right) + |\theta_1 + \cdots + \theta_\ell| c^{-1/12}.
\]

Hence, taking the limit \( c \to \infty \), the right-hand terms converge to 0 in view of Lemma 8.3 (ii) and Lemma 8.5. \( \square \)
References

[1] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). Regular variation. Cambridge University Press, Cambridge.

[2] Blum, M. G. B. and François, O. (2005). Minimal clade size and external branch length under the neutral coalescent. Adv. in Appl. Probab. 37, 647–662.

[3] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle’s probability cascades and an abstract cavity method. Comm. Math. Phys. 197, 247–276.

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

[5] Caliebe, A., Neininge, R., Krawczak, M. and Röslter, U. (2007). On the length distribution of external branches in coalescence trees: genetic diversity within species. Theor. Popul. Biol. 72, 245–252.

[6] Dahmer, I., Kersting, G. and Wakolbinger, A. (2014). The total external branch length of Beta-coalescents. Combin. Probab. Comput. 23, 1010–1027.

[7] Desai, M. M., Walczak, A. M. and Fisher, D. S. (2013). Genetic Diversity and the Structure of Genealogies in Rapidly Adapting Populations. Genetics 193, 565–585.

[8] Dhersin, J.-S., Freund, F., Siri-Jégousse, A. and Yuan, L. (2013). On the length of an external branch in the Beta-coalescent. Stochastic Process. Appl. 123, 1691–1715.

[9] Dhersin, J.-S. and Möhle, M. (2013). On the external branches of coalescents with multiple collisions. Electron. J. Probab. 18, 1–11.

[10] Dhersin, J.-S. and Yuan, L. (2015). On the total length of external branches for Beta-coalescents. Adv. in Appl. Probab. 47, 693–714.

[11] Diehl, C. S. and Kersting, G. (2018). Tree lengths for general Λ-coalescents and the asymptotic site frequency spectrum around the Bolthausen-Sznitman coalescent. Submitted. Preprint available at arXiv: 1804.00961.

[12] Eldon, B., Birkner, M., Blath, J. and Freund, F. (2015). Can the Site-Frequency Spectrum Distinguish Exponential Population Growth from Multiple-Merger Coalescents? Genetics 199, 841–856.

[13] Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Vol. 2. John Wiley & Sons, New York.
[14] Freund, F. and Möhle, M. (2009). On the time back to the most recent common ancestor and the external branch length of the Bolthausen-Sznitman coalescent. *Markov Process. Related Fields* **15**, 7387–416.

[15] Goldschmidt, C. and Martin, J. (2005). Random Recursive Trees and the Bolthausen-Sznitman Coalescent. *Electron. J. Probab.* **10**, 718–745.

[16] Janson, S. and Kersting, G. (2011). On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16**, 2203–2218.

[17] Kersting, G., Pardo, J., and Siri-Jougusse, A. (2014). Total internal and external lengths of the Bolthausen-Sznitman coalescent. *J. Appl. Probab.*, **51**, 73–86.

[18] Kersting, G., Schweinsberg, J. and Wakolbinger, A. (2014). The evolving beta coalescent. *Electron. J. Probab.* **19**, 1–27.

[19] Kersting, G., Schweinsberg, J. and Wakolbinger, A. (2018). The size of the last merger and time reversal in Λ-coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **54**, 1527–1555.

[20] Kimura, M. (1969). The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics* **61**, 893–903.

[21] Kingman, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13**, 235–248.

[22] Limic, V. and Sturm, A. (2006). The spatial Λ-coalescent. *Electron. J. Probab.* **11**, 363–393.

[23] Möhle, M. (2010). Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent. *Stochastic Process. Appl.* **120**, 2159–2173.

[24] Möhle, M. (2015). The Mittag-Leffler process and a scaling limit for the block counting process of the Bolthausen-Sznitman coalescent. *ALEA Lat. Am. J. Probab. Math. Stat.* **12**, 35–53.

[25] Neher, R. A. and Hallatschek, O. (2013). Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci. USA* **10**, 437–442.

[26] Niwa, H.-S., Nashida, K. and Yanagimoto, T. (2016). Reproductive skew in Japanese sardine inferred from DNA sequences. *ICES J. Mar. Sci.* **73**, 2181–2189.

[27] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870–1902.
[28] Seneta, E. (1973). A Tauberian theorem of E. Landau and W. Feller. *Ann. Probab.* 1, 1057–1058.

[29] Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* 36, 1116–1125.

[30] Schweinsberg, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Probab.* 5, 1–11.

[31] Schweinsberg, J. (2017). Rigorous results for a population model with selection II: genealogy of the population. *Electron. J. Probab.* 38, 1–54.

[32] Siri-Jégousse, A. and Yuan, L. (2016). Asymptotics of the minimal clade size and related functionals of certain Beta-coalescents. *Acta Appl. Math.* 142, 127–148.

[33] Steinrücken, M., Birkner, M. and Blath, J. (2013). Analysis of DNA sequence variation within marine species using Beta-coalescents. *Theor. Popul. Biol.* 87, 15–24.

[34] Villandr, L., Labbe, A., Brenner, B., Roger, M. and Stephens, D. A. (2018). DM-PhyClus: a Bayesian phylogenetic algorithm for infectious disease transmission cluster inference. *BMC Bioinformatics* 19: 324.

[35] Wakeley, J., Nielsen, R., Liu-Cordero, S. N. and Ardlie, K. (2001). The Discovery of Single-Nucleotide Polymorphisms—and Inferences about Human Demographic History. *Am. J. Hum. Genet.* 69, 1332–1347.

[36] Wallstrom, T., Bhattacharya, T., Wilkins, J. and Fischer, W. (2016). Generalized Coalescents may be Necessary for Modeling Intrahost HIV Evolution, presented at 23rd International HIV Dynamics and Evolution, 2016-04-27 (Woods Hole, Massachusetts, United States). [https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-16-22791](https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-16-22791).

[37] Yuan, L. (2014). On the measure division construction of Λ-coalescents. *Markov Process. Related Fields.* 20, 229–264.