Electronic Journal of Probability

Electron. J. Probab. **24** (2019), no. 134, 1–36.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP354

External branch lengths of Λ-coalescents without a dust component*

Christina S. Diehl[†] Götz Kersting[†]

Abstract

Λ-coalescents model genealogies of samples of individuals from a large population by means of a family tree. The tree's leaves represent the individuals, and the lengths of the adjacent edges indicate the individuals' time durations up to some common ancestor. These edges are called external branches. We consider typical external branches under the broad assumption that the coalescent has no dust component and maximal external branches under further regularity assumptions. As it transpires, 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, in addition to the asymptotic independence of several typical external lengths, holds 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. Firstly, we analyze a class of Λ-coalescents coming down from infinity and with regularly varying μ. 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. Secondly, 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.

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.
AMS MSC 2010: Primary 60J75, Secondary 60F05; 60J27; 92D25.
Submitted to EJP on November 20, 2018, final version accepted on August 15, 2019.
Supersedes arXiv:1811.07653.

*Work partially supported by the DFG Priority Programme SPP 1590 "Probabilistic Structures in Evolution".
[†]Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany.
E-mail: diehl@math.uni-frankfurt.de,kersting@math.uni-frankfurt.de

1 Introduction and main results

In population genetics, family trees stemming from a sample out of a big population are modeled by coalescents. The prominent Kingman coalescent [22] found widespread applications in biology. More recently, the Bolthausen-Sznitman coalescent, originating from statistical mechanics [3], has gained in importance in analyzing genealogies of populations undergoing selection [5, 8, 26, 32]. Unlike Kingman's coalescent, the Bolthausen-Sznitman coalescent allows multiple mergers. The larger class of Beta-coalescents has found increasing interest, e.g., in the study of marine species [34, 27]. All these instances are covered by the notion of Λ-coalescents as introduced by Pitman [28] and Sagitov [30] in 1999. Today, general properties of this extensive class have become more transparent [20, 12].

In this paper, we deal with the lengths of external branches of Λ-coalescents under the broad assumption that the coalescent has no dust component, which applies to all the 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 [24, 17, 7, 18, 11].

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

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

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

$$\mu(b) \;:=\; \sum_{k=2}^{b} (k-1) \binom{b}{k} \lambda_{b,k}, \qquad b \geq 2.$$

We call this quantity the rate of decrease as it is the rate at which the number of blocks is decreasing on average. Note that a merger of k blocks corresponds to a decline of $k-1$ blocks. The importance of μ also became apparent from other publications [31, 23, 12]. In particular, the assumption of absence of a dust component may be expressed in this term. Originally characterized by the condition

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

(see [28]), it can be equivalently specified by the requirement

$$\frac{\mu(n)}{n} \;\to\; \infty$$

as $n \to \infty$ (see Lemma 1 (iii) of [12]).

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 [21] 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 [36] and are known to be a characteristic for trees with multiple mergers [13]; e.g., external branch lengths have been used to discriminate between different coalescents in the context of HIV trees [37] (see also [35]). Of course, such considerations have rather theoretical value as long as singleton polymorphisms cannot be distinguished from sequencing errors.

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

$$T_i^n \ := \ \inf\{t \ge 0 : \ \{i\} \notin \Pi_n(t)\}.$$

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_i^n for $1 \le i \le n$. The result clarifies the magnitude of T^n in full generality.

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

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

as $n \to \infty$.

Among others, this theorem excludes the possibility of T^n converging to a positive constant in probability. In [19] 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 $1 < \alpha < 2$ were considered and the duration was given as $n^{1-\alpha}$. Our theorem shows that for this quantity the term $n/\mu(n)$ is a suitable measure for Λ-coalescents without a dust component.

Asymptotic independence of the external branch lengths holds as well in full generality for dustless coalescents. In light of the waiting times, which the different external branches have in common, this may be an unexpected result. However, this dependence vanishes in the limit. Then it becomes crucial whether two external branches end in the same merger. Such an event is asymptotically negligible only in the dustless case. This heuristic motivates the following result.

Theorem 1.2. *A Λ-coalescent has no dust component if and only if for fixed $k \in \mathbb{N}$ and for any sequence of numbers $t_1^n, \ldots, t_k^n \ge 0$, $n \ge 2$, we have*

$$\mathbf{P}\left(T_1^n \le t_1^n, \ldots, T_k^n \le t_k^n\right) \ = \ \mathbf{P}\left(T_1^n \le t_1^n\right) \cdots \mathbf{P}\left(T_k^n \le t_k^n\right) + o(1)$$

as $n \to \infty$.

In the dustless case, one has $T_i^n \to 0$ in probability for $1 \le i \le k$ and one reasonably restricts to the case $t_i^n \to 0$ as $n \to \infty$. The statement that the asymptotic independence fails for coalescents with a dust component goes back to Möhle (see equation (10) of [24]).

In order to achieve convergence in distribution of the scaled lengths, stronger assumptions are required on the rate of decrease, namely that μ is a regularly varying sequence. A characterization of this property is given in Proposition 3.2 below. Let δ_0 denote the Dirac measure at zero.

Theorem 1.3. *For a Λ-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 to δ_0 as $n \to \infty$ if and only if μ is regularly varying at infinity. Then its exponent α of regular variation fulfills $1 \le \alpha \le 2$ and we have*

(i) *for $1 < \alpha \le 2$,*

$$\mathbf{P}\left(\frac{\mu(n)}{n}\, T^n > t\right) \ \longrightarrow \ \frac{1}{(1+(\alpha-1)\,t)^{\frac{\alpha}{\alpha-1}}}\,, \qquad t \ge 0,$$

(ii) *for* $\alpha = 1$*,*

$$\mathbf{P}\left(\frac{\mu(n)}{n}\, T^n > t\right) \;\longrightarrow\; e^{-t}, \qquad t \geq 0,$$

as $n \to \infty$*.*

In particular, this theorem includes the special cases known from the literature. Blum and François [2], as well as Caliebe et al. [6], studied Kingman's coalescent. For the Bolthausen-Sznitman coalescent, Freund and Möhle [15] showed asymptotic exponentiality of the external branch length. This result was generalized by Yuan [38]. A class of coalescents containing the Beta$(2-\alpha,\alpha)$-coalescent with $1<\alpha<2$ was analyzed by Dhersin et al. [9].

Combining Theorem 1.2 and 1.3 yields the following corollary:

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

$$\frac{\mu(n)}{n}\,(T_1^n,\,\dots,\,T_k^n) \;\xrightarrow{d}\; (T_1,\,\dots,\,T_k)$$

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

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

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

Example 1.5. For $k \in \mathbb{N}$, let $T_1,\,\dots,\,T_k$ be the i.i.d. random variables from Corollary 1.4.

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

$$n\,(T_1^n,\dots,T_k^n) \;\xrightarrow{d}\; (T_1,\dots,T_k)$$

as $n \to \infty$. This statement covers (after scaling) the Kingman case. Note that $\Lambda|_{(0,1]}$ does not affect the limit.

(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,

$$n^{1-a}\,(T_1^n,\dots,T_k^n) \;\xrightarrow{d}\; (T_1,\dots,T_k)$$

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 [33]). Note that the constant b does not appear in the limit.

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

$$\log n\,(T_1^n,\dots,T_k^n) \;\xrightarrow{d}\; (T_1,\dots,T_k) \tag{1.2}$$

as $n \to \infty$. This contains the Bolthausen-Sznitman coalescent (see Corollary 1.7 of Dhersin and Möhle [10]). Again, the constant b does not show up in the limit.

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 Λ-coalescents having regularly varying rate of decrease μ, additionally to the absence of a dust component. It transpires that one has to distinguish between two cases.

First, we treat the case of μ 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

$$\mu(s_n) \;=\; \frac{\mu(n)}{n}\,. \tag{1.3}$$

Note that $\mu(n)/n$ is a strictly increasing and, in the dustless case, diverging sequence (see Lemma 3.1 (ii) and (iv) below), which directly transfers to the sequence $(s_n)_{n\geq 2}$. Also note in view of Lemma 3.1 (ii) below that

$$s_n \;=\; o(n) \tag{1.4}$$

as $n \to \infty$.

Example 1.6. (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 μ 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 Φ^n on $(0, \infty)$ via

$$\Phi^n(B) \;:=\; \#\left\{ i \leq n : \; \frac{\mu(n)}{ns_n}\, T_i^n \in B \right\}$$

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

Theorem 1.7. *Assume that the Λ-coalescent has a regularly varying rate of decrease μ with exponent $\alpha \in (1, 2]$. Then, as $n \to \infty$, the point process Φ^n converges in distribution to a Poisson point process Φ 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 Φ accumulate at the origin. On the other hand, we have $\int_1^\infty \phi(x)dx < \infty$, meaning 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 [17] using a different method.

In particular, letting $T^n_{\langle 1\rangle}$ be the maximal length of the external branches, we obtain for $x > 0$,

$$\mathbf{P}\left(\frac{\mu(n)}{ns_n} T^n_{\langle 1\rangle} \;\leq\; x \right) \;\to\; e^{-((\alpha-1)x)^{-\frac{\alpha}{\alpha-1}}}$$

as $n \to \infty$, i.e., the properly scaled $T^n_{\langle 1\rangle}$ is asymptotically Fréchet-distributed.

Corollary 1.4 shows that the external branch lengths behave for large n as i.i.d. random variables. This observation is emphasized by Theorem 1.7 because the maximal values of i.i.d. random variables with the densities stated in Corollary 1.4 have the exact limiting behavior as given in Theorem 1.7 (including the scaling constants s_n).

This heuristic fails for the Bolthausen-Sznitman coalescent, which we shall now address. 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 Ψ^n on the whole real line given by

$$\Psi^n(B) \;:=\; \#\left\{ i \leq n : \; \log\log{(n)}(T_i^n - t_n) \in B \right\}$$

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

Theorem 1.8. *For the Bolthausen-Sznitman coalescent, the point process Ψ^n converges in distribution as $n \to \infty$ to a Cox point process Ψ 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 9.1 below). Recall that $G := -\log E$ has a standard Gumbel distribution.

Again, letting $T^n_{\langle 1 \rangle}$ be the maximum of $T^n_1, \ldots, T^n_n$, we obtain

$$\mathbf{P}\Big(\log\log(n)(T^n_{\langle 1\rangle} - t_n) \ \le \ x\Big) \ \longrightarrow \ \int_0^\infty e^{-ye^{-x}} e^{-y}\, dy \ = \ \frac{1}{1+e^{-x}} \tag{1.5}$$

as $n \to \infty$. Notably, we arrive at a limit that is non-standard in the extreme value theory of i.i.d. random variables, namely, the so-called logistic distribution.

We point out that the limiting point process Ψ 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, note that (1.5) implies

$$\frac{T^n_{\langle 1\rangle}}{\log\log n} \ = \ 1 + o_p(1)$$

as $n \to \infty$, where $o_p(1)$ denotes a sequence of random variables converging to 0 in probability. In particular, $T^n_{\langle 1\rangle} \to \infty$ in probability. Hence, we pass with this theorem to the situation where very large mergers affect 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 μ 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 $\widetilde{\tau}_n$ (the moment of the most recent common ancestor) of the Bolthausen-Sznitman coalescent:

$$\widetilde{\tau}_n - \log\log n \ \stackrel{d}{\longrightarrow} \ G$$

as $n \to \infty$ (see Goldschmidt and Martin [16]). However, this shift remains unscaled. Apparently, these two Gumbel distributed variables under consideration build up within different parts of the coalescent tree.

Before closing this introduction, we provide some hints concerning the proofs. For the first three theorems, we make use of an asymptotic representation for the tail probabilities of the external branch lengths. Remarkably, this representation involves, solely, the rate of decrease μ, though in a somewhat implicit, twofold manner. The proofs of the three theorems consist in working out the consequences of these circumstances. The representation is given in Theorem 4.1 and relies largely on different approximation formulae derived in [12]. We recall the required statements in Section 2.

The proofs of the last two theorems incorporate Corollary 1.4 as one ingredient. The idea is to implement stopping times $\widetilde{\rho}_{c,n}$ with the property that, at that moment, a positive number of external branches is still extant that is of order 1 uniformly in n. 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 $t > \widetilde{\tau}_n$. In the case of regularly varying μ with exponent $1 < \alpha \leq 2$, we will show that

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

with arbitrary $c > 0$ is the right choice. Next, we split the external lengths T_i^n into the times $\check{T}_i^n$ up to the moment $\widetilde{\rho}_{c,n}$ and the residual times $\widehat{T}_i^n$. Formally, we have

$$\check{T}_i^n \;:=\; T_i^n \wedge \widetilde{\rho}_{c,n} \qquad \text{and} \qquad \widehat{T}_i^n \;:=\; T_i^n - \check{T}_i^n.$$

We shall see that $\check{T}_i^n$ is of negligible size compared to $\widehat{T}_i^n$ for large values of c. On the other hand, with increasing c, also the number of extant external branches tends to infinity uniformly in n. Corollary 1.4 tells us that the $\widehat{T}_i^n$ behave approximately like i.i.d. random variables. Therefore, one expects that the classical extreme value theory applies in our context. These are the ingredients of the proof of Theorem 1.7.

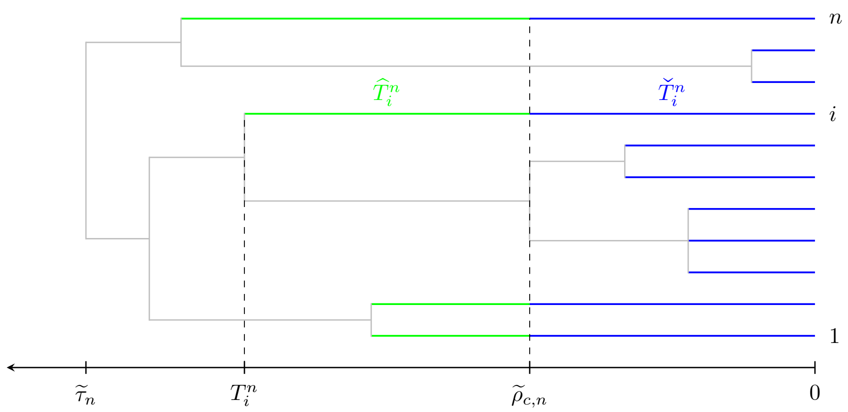

Figure 1: The stopping time $\widetilde{\rho}_{c,n}$ subdividing the external branch ending in leaf i into two parts of length $\check{T}_i^n$ and $\widehat{T}_i^n$, respectively.

The approach for the Bolthausen-Sznitman coalescent is essentially the same. However, new obstacles appear. In contrast to the previous case $\alpha > 1$, the lengths of the maximal branches now diverge in probability. As a consequence, in the case $\alpha = 1$, we have in general no longer control over the stopping times $\widetilde{\rho}_{c,n}$ as defined above. Fortunately, Möhle [25] provides for the Bolthausen-Sznitman coalescent a precise asymptotic description of the block counting process N_n by means of the Mittag-Leffler process, which applies also in the large time regime. Adapted to this result, the role of $\widetilde{\rho}_{c,n}$ is taken by $t_{c,n} \wedge \widetilde{\tau}_n$, 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

$$\check{T}_i^n \;:=\; T_i^n \wedge t_{c,n} \qquad \text{and} \qquad \widehat{T}_i^n \;:=\; T_i^n - \check{T}_i^n.$$

In contrast to the case $\alpha > 1$, the part $\breve{T}_i^n$ does not disappear for $c \to \infty$ but is asymptotically Gumbel-distributed and shows up in the above mentioned independent shift.

The paper is organized as follows: In Section 2 we recapitulate some laws of large numbers from [12]. Section 3 summarizes several properties of the rate of decrease. The fundamental asymptotic expression of the external tail properties is developed in Section 4. Sections 5 and 6 contain the proofs of Theorem 1.1 to 1.3. In Section 7 we prepare the proofs of the remaining theorems by establishing a formula for factorial moments of the number of external branches. Sections 8 and 9 include the proofs of Theorem 1.7 and 1.8.

2 Some laws of large numbers

In this section, we report on some laws of large numbers from the recent publication [12], which are a main tool in the subsequent proofs. Let $X = (X_j)_{j\in\mathbb{N}_0}$ denote the Markov chain embedded in the block-counting process N_n, i.e., X_j denotes the number of branches after j merging events. (For convenience, we suppress n in the notation of X.) Also, let

$$\rho_r := \min\{j \geq 0 : X_j \leq r\}$$

for numbers $r > 0$. We are dealing with laws of large numbers for functionals of the form

$$\sum_{j=0}^{\rho_{r_n}-1} f(X_j)$$

with some suitable positive function f and some sequence $(r_n)_{n\geq 1}$ of positive numbers. These laws of large numbers build on two approximation steps. First, letting

$$\Delta X_{j+1} := X_j - X_{j+1} \quad \text{and} \quad \nu(b) := \mathbb{E}[\Delta X_{j+1} \mid X_j = b]$$

for $j \geq 1$, we notice that for large n,

$$\sum_{j=0}^{\rho_r-1} f(X_j) \;\approx\; \sum_{j=0}^{\rho_r-1} f(X_j)\frac{\Delta X_{j+1}}{\nu(X_j)}.$$

The rationale of this approximation consists in the observation that the difference of both sums stems from the martingale difference sequence $f(X_j)(1 - \Delta X_{j+1}/\nu(X_j))$, $j \geq 0$, and, therefore, is of a comparatively negligible order. Second, we remark that

$$\sum_{j=0}^{\rho_r-1} \frac{f(X_j)}{\nu(X_j)}\Delta X_{j+1} \;\approx\; \int_r^n \frac{f(x)}{\nu(x)}\,dx,$$

with $\nu(x)$ extending the numbers $\nu(b)$ to real numbers $x \geq 2$. Here, we regard the left-hand sum as a Riemann approximation of the right-hand integral and take $X_{\rho_r} \approx r$ into account. Altogether,

$$\sum_{i=0}^{\rho_r-1} f(X_i) \;\approx\; \int_r^n \frac{f(x)}{\nu(x)}.$$

In order to estimate the errors and, in particular, the martingale's quadratic variation, different assumptions are required. For details we refer to [12] and deal here only with the two cases that we use later in our proofs.

The first case concerns the time

$$\widetilde{\rho}_r := \inf\{t \geq 0 :\; N_n(t) \;\leq\; r\},$$

when the block-counting process drops below r. Letting W_j be the period of stay of N_n at state X_j (again suppressing n in the notation), we have

$$\widetilde{\rho}_r \;=\; \sum_{j=0}^{\rho_r-1} W_j \;\approx\; \sum_{j=0}^{\rho_r-1} \mathbb{E}[W_j \mid N_n] \;=\; \sum_{j=0}^{\rho_r-1} \frac{1}{\lambda(X_j)},$$

where $\lambda(b) := \sum_{2\le k\le b} \lambda_{b,k}$ is the jump rate of the block counting process. Also, $\nu(b) = \mu(b)/\lambda(b)$. Therefore, putting $f(x) = \lambda(x)^{-1}$, we are led to the approximation formula

$$\rho_r \;\approx\; \int_r^n \frac{dx}{\mu(x)}\;.$$

More precisely, we have the following law of large numbers.

Proposition 2.1. *Assume that the* Λ*-coalescent is dustless. Let* $\gamma < 1$ *and let* $2 \le r_n \le \gamma n$, $n \ge 1$, *be numbers such that*

$$\int_{r_n}^n \frac{dx}{\mu(x)} \;\to\; 0$$

as $n \to \infty$*. Then*

$$\widetilde{\rho}_{r_n} \;=\; (1 + o_P(1)) \int_{r_n}^n \frac{dx}{\mu(x)}$$

as $n \to \infty$*.*

The role of the assumptions is easily understood: The condition $\int_{r_n}^n \frac{dx}{\mu(x)} \to 0$ implies that $\widetilde{\rho}_{r_n} \to 0$ in probability, i.e., we are in the small time regime. This is required to avoid very large jumps ΔX_{j+1} of order X_{j+1}, which would ruin the above Riemann approximation. The condition $r_n \le \gamma n$ guarantees that $\widetilde{\rho}_{r_n}$ is sufficiently large to allow for a law of large numbers.

Secondly, we turn to the case $f(x) = x^{-1}$. Here, we point out that, as $x \to \infty$,

$$\frac{1}{\nu(x)} \;\sim\; x\frac{d}{dx}\log\frac{\mu(x)}{x},$$

which follows from [12, Lemma 1 (ii)]. Hence,

$$\int_r^n \frac{dx}{x\nu(x)} \;\approx\; \log\left(\frac{\mu(n)}{n}\frac{r}{\mu(r)}\right)$$

and we have the following law of large numbers.

Proposition 2.2. *Under the assumptions of the previous proposition, we have*

$$\sum_{j=0}^{\rho_{r_n}-1} \frac{1}{X_j} \;=\; (1 + o_P(1))\log\left(\frac{\mu(n)}{n}\frac{r_n}{\mu(r_n)}\right) \quad \textit{and} \quad \sum_{j=0}^{\rho_{r_n}-1} \frac{1}{X_j} \;=\; \log\left(\frac{\mu(n)}{n}\frac{r_n}{\mu(r_n)}\right) + o_P(1)$$

as $n \to \infty$*.*

For the proofs of these propositions, see [12, Section 3].

3 Properties of the rate of decrease

We now have a closer look at the rate of decrease μ introduced in the first section. Defining

$$\mu(x) \;:=\; \int_{[0,1]} (xp - 1 + (1-p)^x)\,\frac{\Lambda(dp)}{p^2}, \tag{3.1}$$

we extent μ to all real values $x \ge 1$, where the integrand's value at $p = 0$ is understood to be $x(x-1)/2$.

The next lemma summarizes some required properties of μ.

Lemma 3.1. *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,* μ *and* μ' *are both non-negative and strictly increasing, while* μ'' *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* $x > 1$*,*

$$\mu'(1) \;\leq\; \frac{\mu(x)}{x-1} \;\leq\; \mu'(x) \qquad \textit{and} \qquad \mu''(x) \;\leq\; \frac{\mu'(x)}{x-1}.$$

(iv) *In the dustless case,*

$$\frac{\mu(x)}{x} \;\to\; \infty$$

as $x \to \infty$*.*

Proof. (i) Let

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

which is a $\mathcal{C}^\infty$-function for $x > 0$. Set

$$\begin{aligned} \mu_1(x) \;:=&\; \int_1^x \mu_2(y)dy + \int_{[0,1]} \left(p + (1-p)\log(1-p)\right) \frac{\Lambda(dp)}{p^2} \\ =&\; \int_{[0,1]} \left((1-p)^x \log(1-p) + p\right) \frac{\Lambda(dp)}{p^2}\,. \end{aligned}$$

Note that the second integral in the first line 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 formulae our claim follows.

(ii) The inequalities are equivalent to the fact that $\mu(x)/x$ is increasing and that $\mu(x)/(x(x-1))$ is decreasing, as follows from formulae (7) and (8) of [12].

(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 [12]. □

In order to characterize regular variation of μ, 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

$$H(1) \;=\; \frac{\Lambda(\{0\})}{2} + \int_0^1 \int_0^p \int_0^y dz\,dy\,\frac{\Lambda(dp)}{p^2} \;=\; \frac{\Lambda([0,1])}{2} \;<\; \infty\,. \tag{3.2}$$

Proposition 3.2. *For a Λ-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 μ has an exponent $\alpha \in [1,2]$ and we have*

$$\mu(x) \ \sim \ \Gamma(3-\alpha)\, x^2\, H\left(x^{-1}\right) \tag{3.3}$$

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 μ together with the notion of regularly varying Λ-coalescents as introduced in [12].

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

Lemma 3.3. *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 η 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 $V(z) = \int_z^1(-v(y))dy + V(1)$ instead; here it holds $V(0+) = \infty$. Now, our claim follows from well known results for regularly varying functions at infinity (see [29] and Theorem 1 (a) and (b) in Section VIII.9 [14]). The proofs translate one-to-one to regularly varying functions at the origin. □

Proof of Proposition 3.2. (i) From the definition (3.1), we obtain by double partial integration (see formula (8) of [12]) that

$$\frac{\mu(x)}{x(x-1)} \ = \ \frac{\Lambda(\{0\})}{2} + \int_0^1 (1-z)^{x-2} h(z)\, dz. \tag{3.4}$$

If $\Lambda(\{0\}) > 0$, then our claim is obvious because the first term of the right-hand side of (3.4) 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 [14]), it is sufficient to prove that

$$\mathcal{L}(x) \ \sim \ \frac{\mu(x)}{x^2} \tag{3.5}$$

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

$$\frac{\mu(x)}{x(x-1)} \ = \ \int_0^{x^{-\delta}} (1-z)^{x-2} h(z)\, dz + \int_{x^{-\delta}}^1 (1-z)^{x-2} h(z)\, dz. \tag{3.6}$$

Because of $\delta < 1$ and (3.2), we have

$$\int_{x^{-\delta}}^{1} (1-z)^{x-2} h(z)\,dz \;\le\; (1-x^{-\delta})^{x-2} \int_{x^{-\delta}}^{1} h(z)\,dz \;\le\; e^{-x^{-\delta}(x-2)}\, H(1) \;=\; o\left(x^{-1}\right) \quad (3.7)$$

as $x \to \infty$. In particular, the second integral in the decomposition (3.6) can be neglected in the limit $x \to \infty$ since $\mu(x)/(x(x-1)) \ge \mu'(1)/x$ in view of Lemma 3.1 (iii). As to the first integral in (3.6), observe for $\delta > \frac{1}{2}$ that

$$-\log \frac{(1-z)^{x-2}}{e^{-zx}} \;=\; \mathcal{O}(x^{1-2\delta}) \;\to\; 0$$

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

$$\int_0^{x^{-\delta}} (1-z)^{x-2} h(z)\,dz \;\sim\; \int_0^{x^{-\delta}} e^{-zx} h(z)\,dz. \quad (3.8)$$

Also, note that

$$\int_{x^{-\delta}}^{1} e^{-zx} h(z) dz \;\le\; e^{-x^{1-\delta}} H(1) \;=\; o\left(x^{-1}\right) \quad (3.9)$$

as $x \to \infty$. Combining (3.6) to (3.9) entails

$$\int_0^1 (1-z)^{x-2} h(z) dz \;\sim\; \mathcal{L}(x).$$

Hence, along with formula (3.4), this proves the asymptotics in (3.5). Moreover, from Lemma 3.1 (ii), we get $1 \le \alpha \le 2$.

(ii) If $1 < \alpha < 2$, then $\Lambda(\{0\}) = 0$. Lemma 3.3 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

$$(2-\alpha) H(u) \;\sim\; u h(u)$$

as $u \to 0^+$. Applying Lemma 3.3 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. □

4 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 1.2 but is also of interest on its own. It shows that the distribution of T^n is primarily determined by the rate function μ.

Theorem 4.1. *For a Λ-coalescent without a dust component and a sequence $(r_n)_{n\in\mathbb{N}}$ satisfying $1 < r_n \le n$ for all $n \in \mathbb{N}$, we have*

$$\mathbf{P}\left(T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)}\right) \;=\; \frac{\mu(r_n)}{\mu(n)} + o(1) \quad (4.1)$$

as $n \to \infty$. Furthermore,

$$\left(\frac{r_n}{n}\right)^2 + o(1) \;\le\; \mathbf{P}\left(T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)}\right) \;\le\; \frac{r_n}{n} + o(1) \quad (4.2)$$

as $n \to \infty$.

Observe that, according to Proposition 2.1, the integral $\int_{r_n}^n \frac{dx}{\mu(x)}$ is the asymptotic time needed to go from n to r_n lineages.

For the proof, we recall our notations. $N_n = (N_n(t))_{t\geq 0}$ denotes the block counting process, with the embedded Markov chain $X = (X_j)_{j\in\mathbb{N}_0}$. In particular, we have $N_n(0) = X_0 = n$ and we set $X_j = 1$ for $j \geq \tau_n$, where τ_n is defined as the total number of merging events. The waiting time of the process N_n in state X_j is again 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 $i \in \{1, \dots, n\}$ coalesces is given by

$$\zeta_i^n \ := \ \max\{j \geq 0 : \ \{i\} \in \Pi_n(W_0 + \cdots + W_{j-1})\}.$$

Similarly, ζ^n denotes the corresponding number of a random external branch with length T^n.

Proof of Theorem 4.1. For later purposes, we show the stronger statement

$$\mathbf{P}\left(T^n > \int_{r_n}^n \frac{dx}{\mu(x)} \,\middle|\, N_n\right) \ = \ \frac{\mu(r_n)}{\mu(n)} + o_P(1) \tag{4.3}$$

as $n \to \infty$. It implies (4.1) by taking expectations and using dominated convergence. Also note that, in view of Lemma 3.1 (ii), the statement (4.2) is a direct consequence of (4.1).

In order to prove (4.3), 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.

Let us first consider the discrete embedded setting and afterwards insert the time component. Since there are $\Delta X_0 + 1$ branches involved in the first merger, we have

$$\mathbf{P}(\zeta^n \geq 1 \mid N_n) \ = \ 1 - \frac{\Delta X_0 + 1}{X_0} \ = \ \frac{X_1 - 1}{X_0} \quad a.s.$$

Iterating this formula, it follows

$$\mathbf{P}(\zeta^n \geq k \mid N_n) \ = \ \prod_{j=0}^{k-1} \frac{X_{j+1} - 1}{X_j} \ = \ \frac{X_k - 1}{n - 1} \prod_{j=0}^{k-1} \left(1 - \frac{1}{X_j}\right) \quad a.s.$$

for $k \geq 1$. For a combinatorial treatment of this formula, see [12, Lemma 4]. 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

$$\mathbf{P}(\zeta^n \geq k \mid N_n) \ = \ \frac{X_k - 1}{n - 1} \exp\left(-\sum_{j=0}^{k-1} \frac{1}{X_j} + \mathcal{O}\left(X_{k-1}^{-1}\right)\right) \qquad a.s. \tag{4.4}$$

as $n \to \infty$.

We like to evaluate this quantity at the stopping times

$$\rho_{r_n} \ := \ \min\{j \geq 0 : \ X_j \leq r_n\}.$$

From Lemma 3.1 (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 $r_n \geq q_1 n$, therefore, we have

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

Hence, we may apply Proposition 2.2 and obtain

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

Also, Lemma 3 of [12] implies

$$X_{\rho_{r_n}} \;=\; r_n + \mathcal{O}_P\left(\Delta X_{\rho_{r_n}}\right) \;=\; r_n + o_P(X_{\rho_{r_n}}).$$

Inserting these two estimates into equation (4.4) and using Lemma 3.1 (ii), it follows

$$\mathbf{P}\left(\zeta^n \geq \rho_{r_n} \mid N_n\right) \;=\; \frac{X_{\rho_{r_n}}-1}{n-1}\frac{\mu\left(X_{\rho_{r_n}}\right)}{X_{\rho_{r_n}}}\frac{n}{\mu\left(n\right)}(1+o_P(1)) \;=\; \frac{\mu\left(r_n\right)}{\mu\left(n\right)} + o_P(1). \tag{4.5}$$

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

$$(1+\delta)\int_{(1+\varepsilon)r_n}^{n} \frac{dx}{\mu(x)} \;<\; \int_{r_n}^{n} \frac{dx}{\mu(x)} \;<\; (1-\delta)\int_{(1-\varepsilon)r_n}^{n} \frac{dx}{\mu(x)} \tag{4.6}$$

for large $n \in \mathbb{N}$. For the proof of the left-hand inequality, note by Lemma 3.1 (ii) that

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

implying, with $q_1 n \leq r_n$,

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

These inequalities show how to choose $\delta > 0$. The right-hand inequality in (4.6) follows along the same lines.

Now, recalling the notion

$$\widetilde{\rho}_{r_n} \;:=\; \inf\{t \geq 0 :\; N_n(t) \leq r_n\},$$

Proposition 2.1 gives for sufficiently small $\varepsilon > 0$ the formula

$$\widetilde{\rho}_{r_n(1+\varepsilon)} \;=\; \int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu\left(x\right)}\left(1+o_P(1)\right) \tag{4.7}$$

as $n \to \infty$. Combining (4.5) to (4.7) yields

$$\begin{aligned}
&\mathbf{P}\Big(T^n > \int_{r_n}^{n} \frac{dx}{\mu(x)} \,\Big|\, N_n\Big) \\
&\quad\leq\; \mathbf{P}\Big(T^n \geq (1+\delta)\int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} \,\Big|\, N_n\Big) \\
&\quad\leq\; \mathbf{P}\left(T^n \geq \widetilde{\rho}_{r_n(1+\varepsilon)} \big|\, N_n\right) + \mathbf{P}\Big(\,(1+\delta)\int_{r_n(1+\varepsilon)}^{n} \frac{dx}{\mu(x)} < \widetilde{\rho}_{r_n(1+\varepsilon)} \,\Big|\, N_n\Big) \\
&\quad=\; \mathbf{P}\left(\zeta^n \geq \rho_{r_n(1+\varepsilon)} \big|\, N_n\right) + o_P(1) \\
&\quad=\; \frac{\mu(r_n(1+\varepsilon))}{\mu(n)} + o_P(1) \\
&\quad\leq\; \frac{\mu(r_n)}{\mu(n)}\left(1+\varepsilon\right)^2 + o_P(1),
\end{aligned}$$

where we used Lemma 3.1 (ii) for the last inequality. With this estimate holding for all $\varepsilon > 0$, we end up with

$$\mathbf{P}\left(T^n > \int_{r_n}^n \frac{dx}{\mu(x)} \middle| N_n\right) \;\leq\; \frac{\mu(r_n)}{\mu(n)} + o_P(1)$$

as $n \to \infty$. The reverse inequality can be shown in the same way so that we obtain equation (4.3).

(b) Now, we turn to the two remaining cases $r_n \sim n$ and $r_n = o(n)$. In view of Lemma 3.1 (ii), the asymptotics $r_n \sim n$ implies $\mu(r_n) \sim \mu(n)$, i.e., the right-hand side of (4.3) 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 3.1 (ii), part (a), therefore, entails for all $q \in (0,1)$,

$$\mathbf{P}\left(T^n > \int_{r_n}^n \frac{dx}{\mu(x)} \middle| N_n\right) \;\geq\; \mathbf{P}\left(T^n > \int_{r'_n}^n \frac{dx}{\mu(x)} \middle| N_n\right) \;\geq\; \frac{\mu(qn)}{\mu(n)} + o_P(1) \;\geq\; q^2 + o_P(1)$$

as $n \to \infty$. Hence, the left-hand side of (4.3) also converges to 1 in probability. Similarly, the convergence of both sides of (4.3) to 0 can be shown for $r_n = o(n)$. □

5 Proofs of Theorem 1.1 and 1.2

Proof of Theorem 1.1. Let r_n be as required in Theorem 4.1. Applying Lemma 3.1 (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 4.1 entails

$$\mathbf{P}\left(\frac{\mu(n)}{n}T^n > \log\frac{n}{r_n}\right) \;\geq\; \left(\frac{r_n}{n}\right)^2 + o(1) \tag{5.1}$$

and

$$\mathbf{P}\left(\frac{\mu(n)}{n}T^n > (n-1)\log\frac{r_n - nr_n}{n - nr_n}\right) \;\leq\; \frac{r_n}{n} + o(1) \tag{5.2}$$

as $n \to \infty$, respectively.

Now, let $t \geq 0$. Using equation (5.1) for

$$r_n \;=\; ne^{-t},$$

while choosing

$$r_n \;=\; \frac{ne^{t/(n-1)}}{1 + n(e^{t/(n-1)} - 1)}$$

in (5.2), we arrive at

$$e^{-2t} + o(1) \;\leq\; \mathbf{P}\left(\frac{\mu(n)}{n}T^n > t\right) \;\leq\; \frac{e^{t/(n-1)}}{1 + n(e^{t/(n-1)} - 1)} + o(1) \;=\; \frac{1}{1+t}\left(1 + o(1)\right),$$

as required. □

Proof of Theorem 1.2. First, we treat the dustless case. Similar to the proof of Theorem 4.1, we initially consider the discrete version ζ_i^n of T_i^n for $1 \le i \le k$ to prove

$$\mathbf{P}\left(\zeta_1^n \ge I_1^n, \ldots, \zeta_k^n \ge I_k^n \mid N_n\right) = \mathbf{P}\left(\zeta_1^n \ge I_1^n \mid N_n\right) \cdots \mathbf{P}\left(\zeta_k^n \ge I_k^n \mid N_n\right) + o_P(1) \quad (5.3)$$

as $n \to \infty$, where $0 =: I_0^n \le I_1^n \le \cdots \le I_k^n$ are random variables measurable with respect to the σ-fields $\sigma(N_n)$. Denote by ζ_A the number of mergers until some external branch out of the set $A \subseteq \{1, \ldots, n\}$ coalesces and let $a := \# A$. Given ΔX_j, the j-th merging amounts to choosing $\Delta X_j + 1$ branches uniformly at random out of the X_j present ones, implying

$$\mathbf{P}\left(\zeta_A \ge m \mid N_n\right) = \frac{(X_m - 1)\cdots(X_m - a)}{(n-1)\cdots(n-a)} \prod_{j=0}^{m-1}\left(1 - \frac{a}{X_j}\right) \qquad a.s. \quad (5.4)$$

for $m \ge 1$ (for details see (28) of [12]). Let $\bar{\zeta}_{\{1,\ldots,k\}} := \zeta_{\{1,\ldots,k\}}$ and $\bar{\zeta}_{\{i,\ldots,k\}} := \zeta_{\{i,\ldots,k\}} - \zeta_{\{i-1,\ldots,k\}}$ for $2 \le i \le k$. Furthermore, let $\bar{N}_{X_j}(t) := N_n(t + W_0 + \cdots + W_{j-1})$, in particular, $\bar{N}_{X_0}(t) := N_n(t)$. The Markov property and (5.4) provide

$$\begin{aligned}
&\mathbf{P}\Big(\zeta_1^n \ge I_1^n, \ldots, \zeta_k^n \ge I_k^n \Big| N_n\Big) \\
&\quad = \prod_{i=1}^{k} \mathbf{P}\Big(\bar{\zeta}_{\{i,\ldots,k\}} \ge I_i^n - I_{i-1}^n \Big| \bar{N}_{X_{I_{i-1}^n}}\Big) \\
&\quad = \prod_{i=1}^{k}\left[\frac{(X_{I_i^n} - 1)\cdots(X_{I_i^n} - k + i - 1)}{(X_{I_{i-1}^n} - 1)\cdots(X_{I_{i-1}^n} - k + i - 1)} \prod_{j=I_{i-1}^n}^{I_i^n - 1}\left(1 - \frac{k-i+1}{X_j}\right)\right] \\
&\quad = \prod_{i=1}^{k}\left[\frac{(X_{I_i^n} - k + i - 1)}{(n - k + i - 1)} \prod_{j=I_{i-1}^n}^{I_i^n - 1}\left(1 - \frac{k-i+1}{X_j}\right)\right] \qquad a.s.
\end{aligned}$$

For $1 \le i \le k$, note that

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

and

$$\frac{X_{I_i^n} - k + i - 1}{n - k + i - 1} = \frac{X_{I_i^n} - 1}{n-1} + \mathcal{O}\left(n^{-1}\right)$$

to obtain

$$\begin{aligned}
&\mathbf{P}\big(\zeta_1^n \ge I_1^n, \ldots, \zeta_k^n \ge I_k^n \mid N_n\big) \\
&\quad = \prod_{i=1}^{k}\left[\left(\frac{X_{I_i^n} - 1}{n-1} + \mathcal{O}\left(n^{-1}\right)\right)\left(\prod_{j=I_{i-1}^n}^{I_i^n - 1}\left(1 - \frac{1}{X_j}\right)^{k-i+1} + \mathcal{O}\left(\left(X_{I_i^n} - 1\right)^{-1}\right)\right)\right] \\
&\quad = \prod_{i=1}^{k}\left[\frac{X_{I_i^n} - 1}{n-1} \prod_{j=I_{i-1}^n}^{I_i^n - 1}\left(1 - \frac{1}{X_j}\right)^{k-i+1}\right] + o_P(1) \\
&\quad = \prod_{i=1}^{k}\left[\frac{X_{I_i^n} - 1}{n-1} \prod_{j=0}^{I_i^n - 1}\left(1 - \frac{1}{X_j}\right)\right] + o_P(1)
\end{aligned}$$

as $n \to \infty$, where the rightmost $\mathcal{O}(\cdot)$-term in the first line stems from the fact that $X_{I_i^n} < X_j$ for all $j < I_i^n$. Furthermore, from (5.4) with $A = \{i\}$, we know that

$$\mathbf{P}\left(\zeta_i^n \geq I_i^n \mid N_n\right) \;=\; \frac{X_{I_i^n} - 1}{n-1} \prod_{j=0}^{I_i^n - 1} \left(1 - \frac{1}{X_j}\right) \qquad a.s.$$

so that we arrive at equation (5.3).

Now, based on exchangeability, it is no loss to assume that $0 \leq t_1^n \leq \cdots \leq t_k^n$. So inserting

$$I_i^n \;:=\; \min\left\{k \geq 1 : \sum_{j=0}^{k-1} W_j > t_i^n\right\} \wedge \tau_n$$

in (5.3) yields

$$\begin{aligned}\mathbf{P}\left(T_1^n > t_1^n, \ldots, T_k^n > t_k^n \mid N_n\right) \;&=\; \mathbf{P}\left(\zeta_1^n \geq I_1^n, \ldots, \zeta_k^n \geq I_k^n \mid N_n\right) \\ &=\; \prod_{i=1}^k \mathbf{P}\left(\zeta_i^n \geq I_i^n \mid N_n\right) + o_P(1) \\ &=\; \prod_{i=1}^k \mathbf{P}\left(T_i^n > t_i^n \mid N_n\right) + o_P(1)\end{aligned}$$

as $n \to \infty$. For $1 \leq i \leq k$, let $1 < r_i^n \leq n$ be defined implicitly via

$$t_i^n \;=\; \int_{r_i^n}^n \frac{dx}{\mu(x)}\,.$$

From Lemma 3.1 (iii), we know that $\int_1^n \frac{dx}{\mu(x)} = \infty$; therefore, r_i^n is well-defined. In the dustless case, consequently, we may apply formula (4.3) to obtain

$$\begin{aligned}\mathbf{P}\left(T_1^n > t_1^n, \ldots, T_k^n > t_k^n \mid N_n\right) \;&=\; \prod_{i=1}^k \mathbf{P}\left(T_i^n > t_i^n \mid N_n\right) + o_P(1) \\ &=\; \prod_{i=1}^k \frac{\mu(r_i^n)}{\mu(n)} + o_P(1)\end{aligned}$$

as $n \to \infty$. Taking expectations in this equation yields, via dominated convergence, the theorem's claim for Λ-coalescents without a dust component.

For Λ-coalescents with dust, we use for $t > 0$ the formula

$$\lim_{n\to\infty} \mathbf{P}\left(T_1^n > t, \ldots, T_k^n > t\right) \;=\; \mathbf{E}\left[S_t^k\right],$$

with non-degenerative positive random variables S_t (see (10) in [24]). For $k \geq 2$, Jensen's inequality implies

$$\lim_{n\to\infty} \mathbf{P}\left(T_1^n > t, \ldots, T_k^n > t\right) \;>\; \mathbf{E}\left[S_t\right]^k \;=\; \lim_{n\to\infty} \mathbf{P}\left(T_1^n > t, \ldots, T_k^n > t\right).$$

This finishes the proof. □

6 Proof of Theorem 1.3

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

$$\mu(x) \;=\; x^\alpha L(x), \tag{6.1}$$

where L is a slowly varying function. Let $r_n := qn$ with $0 < q \le 1$. The statement of Theorem 4.1 then boils down to

$$\mathbf{P}\left(\frac{\mu(n)}{n}T^n > \frac{1}{n}\int_{qn}^{n}\frac{\mu(n)}{\mu(x)}dx\right) = q^\alpha + o(1) \tag{6.2}$$

as $n \to \infty$. From (6.1) we obtain

$$n^{-1}\int_{qn}^{n}\frac{\mu(n)}{\mu(x)}dx \sim \begin{cases} -\log q & \text{for} \quad \alpha = 1 \\ \frac{1}{\alpha-1}\left(q^{-(\alpha-1)}-1\right) & \text{for} \quad 1 < \alpha \le 2 \end{cases}$$

as $n \to \infty$. Thus, choosing, for given $t \ge 0$,

$$q = \begin{cases} e^{-t} & \text{for} \quad \alpha = 1 \\ (1+(\alpha-1)\,t)^{-\frac{1}{\alpha-1}} & \text{for} \quad 1 < \alpha \le 2 \end{cases}$$

in equation (6.2) yields the claim.

(b) Now, 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 to δ_0 with cumulative distribution function $F = 1 - \bar{F}$, i.e.,

$$\mathbf{P}\left(\gamma_n T^n > t\right) \overset{n\to\infty}{\longrightarrow} \bar{F}(t) \tag{6.3}$$

for $t \ge 0$, $t \notin D$, where D denotes the set of discontinuities of $\bar{F}$. In view of Theorem 1.1, note that $0 < \bar{F}(t) < 1$ for all $t > 0$. In order to prove that μ is regularly varying, we bring together the assumption (6.3) with the statement of Theorem 4.1, which requires several steps.

For this purpose, similar to the proof of Theorem 1.2, we define the numbers $r_n(t)$ for $t \ge 0$ implicitly via

$$t = \gamma_n \int_{r_n(t)}^{n}\frac{dx}{\mu(x)}\,. \tag{6.4}$$

Let us first solve this implicit equation. Applying formula (4.3) and (6.3), we obtain

$$\frac{\mu(r_n(t))}{\mu(n)} = \bar{F}(t) + o(1) \tag{6.5}$$

for all $t \ge 0$, $t \notin D$, as $n \to \infty$. Differentiating both sides of (6.4) with respect to t and using Lemma 3.1 (i) yields

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

In conjunction with (6.5), it follows that

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

and, by dominated convergence,

$$r_n(t) = n - \frac{\mu(n)}{\gamma_n}\left(\int_0^t \bar{F}(s)ds + o\left(1\right)\right) \tag{6.6}$$

as $n \to \infty$.

Next, we show that $\gamma_n \sim \ c\mu'(n)$ for some $c > 0$. By Theorem 1.1, it follows that there exist $0 < c_1 \le c_2 < \infty$ with

$$c_1 \, \frac{\mu(n)}{n} \ \le \ \gamma_n \ \le \ c_2 \, \frac{\mu(n)}{n}, \qquad n \ge 2. \tag{6.7}$$

Furthermore, from equation (6.6) and a Taylor expansion, we get

$$\mu(r_n(t)) \ = \ \mu(n) + \mu'(n)\,(r_n(t)-n) + \frac{1}{2}\mu''\,(\xi_n)\,(r_n(t)-n)^2\,,$$

where $r_n(t) \le \xi_n \le n$. Dividing this equation by $\mu(n)$, using (6.5) and (6.6), and rearranging terms, we obtain

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

as $n \to \infty$. From Lemma 3.1 (iii) and (i), we get $\mu''(\xi_n) \le \mu'(\xi_n)/(\xi_n{-}1) \le \mu'(n)/(r_n(t){-}1)$. Moreover, equation (6.6) with (6.7) yields $r_n(t) - 1 \ge n/2 + o(n)$ for t sufficiently small. Taking (6.7) once more into account, we obtain that, for given $\varepsilon > 0$ and t sufficiently small,

$$\begin{aligned}\left|1 - \bar F(t) + o(1) - \frac{\mu'(n)}{\gamma_n}\int_0^t \bar F(s)ds\,(1+o(1))\right| \ &\le \ \frac{\mu'(n)}{c_1\gamma_n}\left(\int_0^t \bar F(s)ds\right)^2 (1+o(1)) \\ &\le \ \varepsilon\,\frac{\mu'(n)}{\gamma_n}\left(\int_0^t \bar F(s)ds\right)(1+o(1))\end{aligned}$$

or, equivalently for $t > 0$,

$$\left|\frac{\gamma_n}{\mu'(n)} \ - \ \frac{\int_0^t \bar F(s)ds}{1-\bar F(t)}\,(1+o(1))\right| \ \le \ \varepsilon\,\frac{\int_0^t \bar F(s)ds}{1-\bar F(t)}\,(1+o(1))\,.$$

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

We now remove γ_n from our equations by setting $\gamma_n = \mu'(n)$, without loss of generality. With this choice, (6.7) changes into

$$c_1\,\frac{\mu(n)}{n} \ \le \ \mu'(n) \ \le \ c_2\,\frac{\mu(n)}{n}, \qquad n \ge 2.$$

Also, inserting (6.6) and (6.7) in (6.5) 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 $n \to \infty$. Let us suitably remodel these formulae. By the monotonicity properties of μ and μ' in view of Lemma 3.1 (i), we may proceed to

$$c_3\frac{\mu(x)}{x} \ \le \ \mu'(x) \ \le \ c_4\frac{\mu(x)}{x}, \quad x \ge 2, \tag{6.8}$$

for suitable $0 < c_3 \le c_4 < \infty$, and

$$\begin{aligned}\mu(x)\bar F(t) \ &= \ \mu\left(x - \frac{\mu(x)}{\mu'(x)}\int_0^t \bar F(s)ds + o(x)\right)(1+o(1)) \\ &= \ \mu\left(x - \frac{\mu(x)}{\mu'(x)}\int_0^t \bar F(s)ds + o(x)\right)\end{aligned} \tag{6.9}$$

as $x \to \infty$, where we pushed the $(1 + o(1))$-term into μ by means of Lemma 3.1 (ii). This equation suggests to pass to the inverse of μ. From Lemma 3.1 (i) we know that $\mu(x)$ has an inverse $\nu(y)$. For this function, formula (6.8) translates into

$$\frac{\nu(y)}{c_4 y} \;\leq\; \nu'(y) \;\leq\; \frac{\nu(y)}{c_3 y}. \tag{6.10}$$

Also, applying ν to equation (6.9), both inside and outside, we get

$$\nu\left(y\bar{F}(t)\right) \;=\; \nu(y) - y\,\nu'(y)\int_0^t \bar{F}(s)ds + o(\nu(y)).$$

This equation allows us, in a next step, to further analyse $\bar{F}$. With $0 \leq u < v$, $u, v \notin D$, it follows that

$$\nu\left(\bar{F}(u)y\right) - \nu\left(\bar{F}(v)y\right) \;=\; y\,\nu'(y)\int_u^v \bar{F}(s)ds\,(1 + o(1)) \tag{6.11}$$

as $y \to \infty$. This equation immediately implies that $\bar{F}(v) < \bar{F}(u)$ for all $u < v$. It also shows that $\bar{F}$ has no jump discontinuities, i.e., $D = \emptyset$. Indeed, by the mean value theorem and because $\nu'(y) = 1/\mu'(\nu(y))$ is decreasing in view of Lemma 3.1 (i), we have for $0 \leq u < v$ that

$$\nu\left(\bar{F}(u)y\right) - \nu\left(\bar{F}(v)y\right) \;\geq\; \nu'(y\bar{F}(u))y\left(\bar{F}(u) - \bar{F}(v)\right) \;\geq\; \nu'(y)y\left(\bar{F}(u) - \bar{F}(v)\right).$$

Thus, also assuming $u, v \notin D$, (6.11) yields

$$\bar{F}(u) - \bar{F}(v) \;\leq\; \int_u^v \bar{F}(s)ds \;\leq\; v - u,$$

which implies $D = \emptyset$.

Now, we are ready to show that ν and, therefore, μ is regularly varying. By a Taylor expansion, we get

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

where $\bar{F}(v)y \leq \xi_y \leq \bar{F}(u)y$. Dividing this equation by $y\nu'(y)$, using formula (6.11) and rearranging terms, it follows for $y \to \infty$ that

$$\left|\int_u^v \bar{F}(s)ds(1 + o(1)) \;-\; \frac{\nu'(\bar{F}(u)y)}{\nu'(y)}\left(\bar{F}(u) - \bar{F}(v)\right)\right| \;=\; \frac{1}{2}\frac{\nu''(\xi_y)y}{\nu'(y)}(\bar{F}(u) - \bar{F}(v))^2. \tag{6.12}$$

Next, let us bound the right-hand term. Note that from Lemma 3.1 (iii) we have, for y sufficiently large,

$$|\nu''(y)| \;=\; \nu'(y)^2\frac{\mu''(\nu(y))}{\mu'(\nu(y))} \;\leq\; \frac{\nu'(y)^2}{\nu(y) - 1} \;\leq\; \frac{2\nu'(y)^2}{\nu(y)}.$$

Hence, using (6.10) twice and $\bar{F}(v)y \leq \xi_y \leq \bar{F}(u)y$, it follows for y sufficiently large that

$$\frac{1}{2}\nu''(\xi_y) \;\leq\; \frac{\nu'(\xi_y)^2}{\nu(\xi_y)} \;\leq\; \frac{1}{c_3^2}\frac{\nu(\xi_y)}{\xi_y^2} \;\leq\; \frac{\nu(\bar{F}(u)y)}{\bar{F}(v)^2y^2} \;\leq\; \frac{c_4}{c_3^2}\frac{\nu'(\bar{F}(u)y)\bar{F}(u)}{\bar{F}(v)^2y}.$$

Now, for given $u > 0$ and $\varepsilon > 0$, because of the continuity and strict monotonicity of $\bar{F}$, we get

$$\frac{1}{2}\nu''(\xi_y) \;\leq\; \varepsilon\frac{\nu'(\bar{F}(u)y)}{y(\bar{F}(u) - \bar{F}(v))}$$

if only the (positive) difference $v - u$ is sufficiently small. Inserting into (6.12), we get

$$\left| \int_u^v \bar{F}(s) ds (1 + o(1)) \;-\; \frac{\nu'(\bar{F}(u)y)}{\nu'(y)} \left(\bar{F}(u) - \bar{F}(v) \right) \right| \;\le\; \varepsilon \, \frac{\nu'(\bar{F}(u)y)}{\nu'(y)} (\bar{F}(u) - \bar{F}(v))$$

or, equivalently for $y \to \infty$,

$$\Big| \frac{\nu'(y)}{\nu'(\bar{F}(u)y)} \;-\; \frac{\bar{F}(u) - \bar{F}(v)}{\int_u^v \bar{F}(s) ds} (1 + o(1)) \Big| \;\le\; \varepsilon \, \frac{\bar{F}(u) - \bar{F}(v)}{\int_u^v \bar{F}(s) ds} (1 + o(1)).$$

Again, since the right-hand quotient is finite and positive for all $u < v$, this estimate implies that $\nu'(y)/\nu'(\bar{F}(u)y)$ has a positive finite limit as $y \to \infty$. Because $\bar{F}(u)$ takes all values between 0 and 1, $\nu'(y)$ is regularly varying. From the Lemma in Section VIII.9 of [14], we then obtain the regular variation of ν with some exponent $\eta \ge 0$. It fulfills $\frac{1}{2} \le \eta \le 1$ as Lemma 3.1 (ii) yields

$$a\sqrt{y} \;\le\; \nu(y) \;\le\; by$$

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

7 Moment calculations for external branches of Λ-coalescents

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

$$Y_j \;:=\; \#\left\{ 1 \le i \le n : \; \{i\} \in \Pi_n \left(W_0 + \cdots + W_{j-1} \right) \right\}.$$

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 Λ-coalescents (also covering coalescents with a 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 τ_n is the total number of merging events.

Lemma 7.1. *Consider a general Λ-coalescent and let ρ be a $\sigma(N_n)$-measurable random variable with $0 \le \rho \le \tau_n$ a.s.*

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

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

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

$$\mathbf{Var}\left(Y_\rho \,|\, N_n \right) \;\le\; \mathbf{E}\left[Y_\rho \,|\, N_n \right] \qquad a.s.$$

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 [7]:
Given N_n and $Y_0, \ldots, Y_{\rho-1}$, the $\Delta X_\rho + 1$ lineages coalescing at the ρ-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 $\Delta Y_\rho := Y_{\rho-1} - Y_\rho$ has a hypergeometric distribution with parameters $X_{\rho-1}$, $Y_{\rho-1}$ and $\Delta X_\rho + 1$. From the formula of the i-th factorial moment of a hypergeometric distributed random variable, we obtain

$$\mathbf{E}\left[(\Delta Y_\rho)_i \mid N_n, Y_0, \ldots, Y_{k-1} \right] \;=\; (\Delta X_\rho + 1)_i \, \frac{(Y_{\rho-1})_i}{(X_{\rho-1})_i} \qquad a.s. \tag{7.1}$$

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

$$(a-b)_r \;=\; (a)_r \sum_{i=0}^{r} \binom{r}{i} (-1)^i \,\frac{(b)_i}{(a)_i} \tag{7.2}$$

for $a, b \in \mathbb{R}$ and $r \in \mathbb{N}$. It follows from the Chu–Vandermonde identity (formula 1.5.7 in [4])

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

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

$$\begin{aligned}
(a-b)_r \;&=\; (-1)^r (b+r-1-a)_r \\
&=\; (-1)^r \sum_{i=0}^{r} \binom{r}{i} (b)_i \,(r-1-a)_{r-i} \\
&=\; (-1)^r \sum_{i=0}^{r} \binom{r}{i} (b)_i \,(-1)^{r-i} \frac{(a)_r}{(a)_i} \,.
\end{aligned}$$

Returning to the number of external branches, we obtain from the identity (7.2) 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 (7.1), we arrive at

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

Furthermore, combining the binomial identity (7.2) with the definition of ΔX_ρ, 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,

$$\mathbf{E}\left[(Y_\rho)_r \mid N_n, Y_0, \ldots, Y_{\rho-1}\right] \;=\; (Y_{\rho-1})_r \,\frac{(X_\rho - 1)_r}{(X_{\rho-1})_r} \qquad a.s.$$

and, finally,

$$\frac{\mathbf{E}\left[(Y_\rho)_r \mid N_n\right]}{(X_\rho)_r} \;=\; \frac{\mathbf{E}\left[(Y_{\rho-1})_r \mid N_n\right]}{(X_{\rho-1})_r} \frac{(X_\rho - 1)_r}{(X_\rho)_r} \;=\; \frac{\mathbf{E}\left[(Y_{\rho-1})_r \mid N_n\right]}{(X_{\rho-1})_r} \left(1 - \frac{r}{X_\rho}\right) \quad a.s.$$

The proof now finishes by iteration and taking $\mathbf{E}\left[Y_0 \,|N_n\,\right] = Y_0 = X_0$ into account.

(ii) The inequality for the conditional variance follows from the representation in (i) with $r = 1$ and $r = 2$:

$$\begin{aligned}
\mathbf{Var}\left(Y_\rho \,|N_n\,\right) \;&=\; X_\rho \left(X_\rho - 1\right) \prod_{j=1}^{\rho} \left(1 - \frac{2}{X_j}\right) - X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)^2 + X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right) \\
&\le\; X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{2}{X_j}\right) - X_\rho^2 \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right)^2 + X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right) \\
&\le\; X_\rho \prod_{j=1}^{\rho} \left(1 - \frac{1}{X_j}\right) \;=\; \mathbf{E}\left[Y_\rho \,|N_n\,\right] \qquad a.s.
\end{aligned}$$

This finishes the proof. □

8 Proof of Theorem 1.7

In order to study Λ-coalescents having a regularly varying rate of decrease μ with exponent $\alpha \in (1, 2]$, we define

$$\kappa(x) \ := \ \frac{\mu(x)}{x}, \qquad x \geq 1,$$

for convenience. For $k \in \mathbb{N}$ and for real-valued random variables $Z_1, \dots, Z_k$, denote the reversed order statistics by

$$Z_{\langle 1 \rangle} \ \geq \ \cdots \ \geq \ Z_{\langle k \rangle}.$$

We now prove the following theorem that is equivalent to Theorem 1.7. Recall the definition of s_n in (1.3).

Theorem 8.1. *Suppose that the* Λ*-coalescent has a regularly varying rate* μ *with exponent* $1 < \alpha \leq 2$ *and fix* $\ell \in \mathbb{N}$*. Then, as* $n \to \infty$*, the following convergence holds:*

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

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

For the rest of this section, keep the stopping times

$$\widetilde{\rho}_{c,n} \ := \ \inf \{ t \geq 0 : \ N_n(t) \leq c s_n \} \tag{8.1}$$

in mind and define their discrete equivalents

$$\rho_{c,n} \ := \ \min \{ j \geq 0 : \ X_j \leq c s_n \} \tag{8.2}$$

for $c > 0$. Later, we shall apply Proposition 2.2 to the latter stopping times, in view of (1.4) and

$$\int_{c s_n}^{n} \frac{dx}{\mu(x)} \ = \ \mathcal{O}\left(\int_{c s_n}^{n} x^{-\alpha + \varepsilon} dx \right) \ = \ \mathcal{O}\left(s_n^{1-\alpha+\varepsilon} \right) \ = \ o(1) \tag{8.3}$$

for $0 < \varepsilon < \alpha - 1$ (because of μ being regularly varying with exponent α).

The next proposition deals with properties of the stopping times from (8.1) and (8.2). It justifies the choice of s_n, it shows that $X_{\rho_{c,n}}$ diverges at the same rate as s_n and that $Y_{\rho_{c,n}}$ is uniformly bounded in n. In particular, it reveals that for large c there are with high probability external branches still present up to the times $\widetilde{\rho}_{c,n}$.

Proposition 8.2. *Assume that the* Λ*-coalescent has a regularly varying rate* μ *with exponent* $\alpha \in (1, 2]$*. Then we have:*

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

$$\lim_{n \to \infty} \mathbf{P} \left(\kappa(s_n)\, \widetilde{\rho}_{c,n} \geq \varepsilon \right) \ = \ 0.$$

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

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

(iii) *For each* $\varepsilon > 0$*,*

$$\limsup_{n \to \infty} \mathbf{P} \left(\left| c^{-\alpha}\, Y_{\rho_{c,n}} - 1 \right| \geq \varepsilon \right) \ \overset{c \to \infty}{\longrightarrow} \ 0.$$

Proof. (i) Because μ is regularly varying with exponent $\alpha > 1$, we have

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

as $n \to \infty$. Now, Proposition 2.2 implies that

$$\kappa(s_n)\,\widetilde{\rho}_{c,n} \;\le\; \frac{1}{\alpha-1}c^{1-\alpha}(1+o_P(1)),$$

which entails the claim.

(ii) Because of (8.3), we may use Lemma 3 (ii) of [12]. Hence, in conjunction with the definition of $\rho_{c,n}$, we obtain

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

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

(iii) We first prove that

$$\mathbf{E}\left[Y_{\rho_{c,n}}|N_n\right] \;=\; c^{\alpha} + o_P\left(1\right) \tag{8.4}$$

as $n \to \infty$. Lemma 7.1 (i), together with a Taylor expansion as in (4.4), provides

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

as $n \to \infty$. Furthermore, (1.4) and (8.3) allow us to apply Proposition 2.2, yielding

$$\sum_{j=0}^{\rho_{c,n}-1}\frac{1}{X_j} \;=\; \log\left(\frac{\kappa(n)}{\kappa(X_{\rho_{c,n}})}\right) + o_P(1) \tag{8.5}$$

as $n \to \infty$. Combining statement (ii) with Lemma 3.1 (ii), therefore, we arrive at

$$\mathbf{E}\left[Y_{\rho_{c,n}} \mid N_n\right] \;=\; n\,\frac{\mu\big(X_{\rho_{c,n}}\big)}{\mu(n)}\left(1+o_P\left(1\right)\right) \;=\; n\,\frac{\mu(cs_n)}{\mu(n)}\left(1+o_P\left(1\right)\right)$$

so that the regular variation of μ and the definition of s_n imply (8.4). Thus, in the upper bound

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

with $\varepsilon > 0$, the first right-hand probability converges to 0. For the second one, Chebyshev's inequality and Lemma 7.1 (ii) imply that

$$\begin{aligned}
\mathbf{P}\big(\big|Y_{\rho_{c,n}} - \mathbf{E}\big[Y_{\rho_{c,n}}|N_n\big]\big| \ge \varepsilon\, c^{\alpha}\big) \;&=\; \mathbf{E}\left[\mathbf{P}\left(\left|Y_{\rho_{c,n}} - \mathbf{E}\left[Y_{\rho_{c,n}}|N_n\right]\right| \ge \varepsilon\, c^{\alpha} \mid N_n\right)\right] \\
&\le\; \mathbf{E}\left[\frac{\mathbf{Var}\left(Y_{\rho_{c,n}}|N_n\right)}{\varepsilon^2 c^{2\alpha}} \wedge 1\right] \\
&\le\; \mathbf{E}\left[\frac{\mathbf{E}\left[Y_{\rho_{c,n}} \mid N_n\right]}{\varepsilon^2\, c^{2\alpha}} \wedge 1\right].
\end{aligned}$$

From (8.4) and dominated convergence, we conclude

$$\mathbf{P}\big(\big|Y_{\rho_{c,n}} - \mathbf{E}\big[Y_{\rho_{c,n}}|N_n\big]\big| \ge \varepsilon\, c^{\alpha}\big) \;\le\; \varepsilon^{-2}c^{-\alpha} + o(1)$$

as $n \to \infty$, which provides the claim. □

For the following lemma, let us recall the subdivided external branch lengths

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

for $1 \le i \le n$ and let

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

Lemma 8.3. *Suppose that the* Λ*-coalescent has a regularly varying rate* μ *with exponent* $\alpha \in (1, 2]$*. Then, for* $\ell, y \in \mathbb{N}$*, there exist random variables* $U_{1,y} \ge \ldots \ge U_{\ell,y}$ *such that the following convergence results hold:*

(i) *For any bounded continuous function* $g : \mathbb{R}^\ell \to \mathbb{R}$ *and for fixed* $y \ge \ell$*, as* $n \to \infty$*,*

$$\mathbf{E}\left[g\left(\kappa(cs_n)\widehat{T}_{\langle 1\rangle}^n, \ldots, \kappa(cs_n)\widehat{T}_{\langle \ell\rangle}^n\right) \bigm| Y_{\rho_{c,n}} = y, X_{\rho_{c,n}}\right] \longrightarrow \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right]$$

in probability.

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

$$y^{-\beta}\left(U_{1,y}, \ldots, U_{\ell,y}\right) \xrightarrow{d} \left(U_1, \ldots, U_\ell\right),$$

where $U_1 > \cdots > U_\ell$ *are the points of the Poisson point process of Theorem 8.1.*

Proof. (i) Let

$$\bar{g}_y(x, z) := \mathbf{E}\left[g\left(\kappa(z)\widehat{T}_{\langle 1\rangle}^n, \ldots, \kappa(z)\widehat{T}_{\langle \ell\rangle}^n\right) \bigm| Y_{\rho_{c,n}} = y, X_{\rho_{c,n}} = x\right]$$

for $x > y, z \ge 2$. Observe that by the strong Markov property, given the events $X_{\rho_{c,n}} = x$ and $Y_{\rho_{c,n}} = y$, the y remaining external branches evolve as y ordinary external branches out of a sample of x many individuals. From these y external branches, we consider the ℓ largest ones. Hence, since κ is regularly varying, Corollary 1.4 yields that

$$\bar{g}_y(x, z) \longrightarrow \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right]$$

as $x \to \infty$ and $z/x \to 1$. From established formulae for order statistics of i.i.d random variables, $(U_{1,y}, \ldots, U_{\ell,y})$ has the density

$$\ell!\binom{y}{\ell} F\left(u_\ell\right)^{y-\ell} \prod_{i=1}^{\ell} f\left(u_i\right) du_1 \cdots du_\ell, \tag{8.6}$$

with $u_1 \ge \cdots \ge u_\ell \ge 0$, where f is the density from formula (1.1) and F its cumulative distribution function.

Now, it follows from Skorohod's representation theorem that one can construct random variables X_n' on a common probability space with the properties that X_n' and $X_{\rho_{c,n}}$ have the same distribution for each $n \ge 1$ and that, in view of Proposition 8.2 (ii), the random variables X_n'/cs_n converge to 1 a.s. It follows

$$\bar{g}_y(X_n', cs_n) \longrightarrow \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right] \qquad a.s.$$

and, therefore,

$$\bar{g}_y(X_{\rho_{c,n}}, cs_n) \longrightarrow \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right]$$

in probability, which is our claim.

(ii) Note that

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

and

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

Consequently,

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

being the density of $y^{-\beta}\left(U_{1,y},\dots,U_{\ell,y}\right)$, has the limit

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

as $y\to\infty$. Indeed, this is the joint density of the rightmost points $U_1>\cdots>U_\ell$ of the Poisson point process given in Theorem 8.1. □

Proof of Theorem 8.1. The proof consists of two parts. First, we consider $(\widehat{T}^n_{\langle 1\rangle},\dots,\widehat{T}^n_{\langle \ell\rangle})$ in the limits $n\to\infty$ and then $c\to\infty$, which gives already the limit of our theorem. Consequently, in the second step it remains to show that $(\breve{T}^n_{\langle 1\rangle},\dots,\breve{T}^n_{\langle \ell\rangle})$ can asymptotically be neglected.

In the first step, we normalize $\widehat{T}^n_{\langle j\rangle}$ not by $\kappa(s_n)$ but by the factor $Y^{-\beta}_{\rho_{c,n}}\kappa(cs_n)$, which is equivalent in the limit $c\to\infty$ because of Proposition 8.2 (iii). Thus, we set

$$V_{c,n}\;:=\;\kappa(cs_n)\left(\widehat{T}^n_{\langle 1\rangle},\dots,\widehat{T}^n_{\langle \ell\rangle}\right).$$

Let $g:\mathbb{R}^\ell\to\mathbb{R}$ be a continuous function and assume that $\max|g|\le 1$. For $c>0$, we obtain via the law of total expectation and Lemma 8.3 (i) that

$$\begin{aligned}
&\Big|\mathbf{E}\Big[g\Big(Y^{-\beta}_{\rho_{c,n}}V_{c,n}\Big)\,\Big|\,X_{\rho_{c,n}}\Big]\;-\;\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|\\
&\quad\le\sum_{c/2\le y\le 2c}\Big|\mathbf{E}\left[g\left(y^{-\beta}V_{c,n}\right)\,\middle|\,Y_{\rho_{c,n}}=y,\,X_{\rho_{c,n}}\right]-\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|\cdot\mathbf{P}\left(Y_{\rho_{c,n}}=y\,|\,X_{\rho_{c,n}}\right)\\
&\qquad\quad+\;2\mathbf{P}\left(\left|Y_{\rho_{c,n}}-c^\alpha\right|\ge c^\alpha/2\,|\,X_{\rho_{c,n}}\right)\\
&\quad\le\;\max_{c/2\le y\le 2c}\Big|\mathbf{E}\left[g\left(y^{-\beta}V_{c,n}\right)\,\middle|\,Y_{\rho_{c,n}}=y,\,X_{\rho_{c,n}}\right]\;-\;\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|\\
&\qquad\quad+\;2\mathbf{P}\left(\left|Y_{\rho_{c,n}}-c^\alpha\right|\ge c^\alpha/2\,|\,X_{\rho_{c,n}}\right)\\
&\quad\le\;\max_{c/2\le y\le 2c}\Big|\mathbf{E}\left[g\left(y^{-\beta}U_{1,y},\dots,y^{-\beta}U_{\ell,y}\right)\right]\;-\;\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|+o_P(1)\\
&\qquad\quad+\;2\mathbf{P}\left(\left|Y_{\rho_{c,n}}-c^\alpha\right|\ge c^\alpha/2\,|\,X_{\rho_{c,n}}\right)
\end{aligned}$$

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

$$\begin{aligned}
&\Big|\mathbf{E}\Big[g\Big(Y^{-\beta}_{\rho_{c,n}}V_{c,n}\Big)\Big]\;-\;\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|\\
&\qquad\le\;\max_{c/2\le y\le 2c}\Big|\mathbf{E}\left[g\left(y^{-\beta}U_{1,y},\dots,y^{-\beta}U_{\ell,y}\right)\right]\;-\;\mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\Big|\;+\;o(1)\\
&\qquad\quad+\;2\mathbf{P}\left(\left|Y_{\rho_{c,n}}-c^\alpha\right|\ge c^\alpha/2\right)
\end{aligned}$$

as $n \to \infty$. Then Lemma 8.3 (ii) and Proposition 8.2 (iii) entail

$$\limsup_{n\to\infty} \Big| \mathbf{E}\Big[g\Big(Y_{\rho_{c,n}}^{-\beta}\, V_{c,n} \Big) \Big] - \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right] \Big| \overset{c\to\infty}{\longrightarrow} 0. \tag{8.7}$$

This finishes the first part of our proof. For the second one, we additionally assume that g is a Lipschitz continuous function with Lipschitz constant 1 (in each coordinate) and prove that

$$\mathbf{E}\left[g\left(\kappa(s_n)\, T^n_{\langle 1\rangle}, \dots, \kappa(s_n)\, T^n_{\langle \ell\rangle} \right) \right] \overset{n\to\infty}{\longrightarrow} \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right], \tag{8.8}$$

which implies the theorem's statement. For $\varepsilon > 0$, we have

$$\begin{aligned}
&\big| \mathbf{E}\big[g\big(\kappa(s_n)\, T^n_{\langle 1\rangle}, \dots, \kappa(s_n)\, T^n_{\langle \ell\rangle} \big) \big] \; - \; \mathbf{E}\big[g\big(U_1, \dots, U_\ell \big) \big] \big| \\
&\quad \le \; \Big| \mathbf{E}\left[g\left(\kappa(s_n)\, \widehat{T}^n_{\langle 1\rangle}, \dots, \kappa(s_n)\, \widehat{T}^n_{\langle \ell\rangle} \right) \right] \; - \; \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right] \Big| + \sum_{i=1}^{\ell} \mathbf{E}\left[\kappa(s_n) \check{T}^n_{\langle i\rangle} \wedge 2 \right] \\
&\quad \le \; \Big| \mathbf{E}\left[g\left(Y_{\rho_{c,n}}^{-\beta}\, V_{c,n} \right) \right] \; - \; \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right] \Big| \\
&\qquad + \; \sum_{i=1}^{\ell} \mathbf{E}\left[\Big| \Big(Y_{\rho_{c,n}}^{-\beta} \kappa(cs_n) - \kappa(s_n) \Big)\, \widehat{T}^n_{\langle i\rangle} \Big| \wedge 2 \right] + \ell\, \mathbf{E}\left[\kappa(s_n) \check{T}^n_{\langle 1\rangle} \wedge 2 \right] \\
&\quad \le \; \Big| \mathbf{E}\left[g\left(Y_{\rho_{c,n}}^{-\beta}\, V_{c,n} \right) \right] \; - \; \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right] \Big| \\
&\qquad + \; \ell\, \mathbf{E}\left[\left(\varepsilon \kappa(cs_n) Y_{\rho_{c,n}}^{-\beta} \widehat{T}_{\langle 1\rangle} \right) \wedge 2 \right] + \; 2\ell\, \mathbf{P}\left(\Big| Y_{\rho_{c,n}}^{-\beta} \kappa(cs_n) - \kappa(s_n) \Big| \ge \varepsilon \kappa(cs_n) Y_{\rho_{c,n}}^{-\beta} \right) \\
&\qquad + \; \ell\varepsilon + 2\ell\, \mathbf{P}\left(\kappa(s_n)\, \check{T}_{\langle 1\rangle} \ge \varepsilon \right)
\end{aligned}$$

and, consequently,

$$\begin{aligned}
&\limsup_{n\to\infty} \big| \mathbf{E}\big[g\big(\kappa(s_n)\, T^n_{\langle 1\rangle}, \dots, \kappa(s_n)\, T^n_{\langle \ell\rangle} \big) \big] - \mathbf{E}\big[g\big(U_1, \dots, U_\ell \big) \big] \big| \\
&\quad \le \; \limsup_{n\to\infty} \Big| \mathbf{E}\left[g\left(Y_{\rho_{c,n}}^{-\beta}\, V_{c,n} \right) \right] - \mathbf{E}\left[g\left(U_1, \dots, U_\ell \right) \right] \Big| \\
&\qquad + \; \ell\, \limsup_{n\to\infty} \Big| \mathbf{E}\left[\left(\varepsilon \kappa(cs_n) Y_{\rho_{c,n}}^{-\beta} \widehat{T}_{\langle 1\rangle} \right) \wedge 2 \right] - \mathbf{E}\left[(\varepsilon U_1) \wedge 2 \right] \Big| + \ell\, \mathbf{E}\left[(\varepsilon U_1) \wedge 2 \right] \\
&\qquad + \; 2\ell \limsup_{n\to\infty} \mathbf{P}\left(\left| 1 - \frac{\kappa(s_n)}{\kappa(cs_n)} Y_{\rho_{c,n}}^{\beta} \right| \ge \varepsilon \right) \\
&\qquad + \; \ell\varepsilon + 2\ell\, \limsup_{n\to\infty} \mathbf{P}\left(\kappa(s_n)\, \widetilde{\rho}_{c,n} \ge \varepsilon \right).
\end{aligned}$$

We now use (8.7) for the first two right-hand terms and Proposition 8.2 (iii) for the first probability also taking $\kappa(cs_n)/\kappa(s_n) \sim c^{\alpha-1} = c^{\alpha\beta}$ into account. To the other probability, we apply Proposition 8.2 (i). Hence, passing to the limit as $c \to \infty$ yields

$$\limsup_{n\to\infty} \big| \mathbf{E}\big[g\big(\kappa(s_n)\, T^n_{\langle 1\rangle}, \dots, \kappa(s_n)\, T^n_{\langle \ell\rangle} \big) \big] - \mathbf{E}\big[g\big(U_1, \dots, U_\ell \big) \big] \big| \; \le \; \ell\, \mathbf{E}\left[(\varepsilon U_1) \wedge 2 \right] + \ell\varepsilon.$$

Finally, taking the limit $\varepsilon \to 0$ and using dominated convergence provides the claim. □

9 Proof of Theorem 1.8

Recall the notation of the reversed order statistics $Z_{\langle 1\rangle} \ge Z_{\langle 2\rangle} \ge \cdots$ 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 1.8:

Theorem 9.1. *For the Bolthausen-Sznitman coalescent, the following convergence holds: For $\ell \in \mathbb{N}$,*

$$\log\log n \left(T^n_{\langle 1\rangle} - t_n \;\ldots, T^n_{\langle \ell\rangle} - t_n\right) \;\xrightarrow{d}\; (U_1 - G, \ldots, U_\ell - G)$$

as $n \to \infty$, where $U_1 > \cdots > U_\ell$ are the ℓ 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 9.2. *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 [25] provides

$$\mathbf{E}\left[N_n(t)^{(r)}\right] \;=\; \frac{\Gamma(r+1)}{\Gamma(1+re^{-t})}\frac{\Gamma(n+re^{-t})}{\Gamma(n)}.$$

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

$$\frac{\Gamma(n+re^{-t})}{\Gamma(n)} \;=\; n^{re^{-t}}\,(1+o(1))$$

and, consequently,

$$\mathbf{E}\left[N_n(t)^{(r)}\right] \;=\; \frac{\Gamma(r+1)}{\Gamma(1+re^{-t})}\; n^{re^{-t}}\,(1+o(1))$$

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

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

as $n \to \infty$.

Now, observe

$$\begin{aligned} e^{-t_{c,n}}\log n \;&=\; \exp\left(-\frac{\log\log\log n}{\log\log n} + \frac{\log c}{\log\log n}\right)\log\log n \\ &=\; \log\log n - \log\log\log n + \log c + o(1) \\ &=\; t_{c,n} + \log c + o(1). \end{aligned}$$

Equivalently,

$$n^{e^{-t_{c,n}}} \;=\; ce^{t_{c,n}}\,(1+o(1))$$

and, therefore,

$$e^{-rt_{c,n}}\mathbf{E}\left[N_n(t_{c,n})^{(r)}\right] \;\to\; c^r r! \tag{9.1}$$

as $n \to \infty$.

Furthermore, because of

$$N_n(t)^r \ \leq \ N_n(t)^{(r)} \ \leq \ N_n(t)^r + 2^r r^r N_n(t)^{r-1} \ \leq \ N_n(t)^r + 2^r r^r N_n(t)^{(r-1)},$$

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, (9.1) transfers to

$$e^{-rt_{c,n}} \mathbf{E}\left[N_n(t_{c,n})^r\right] \ \longrightarrow \ c^r r!$$

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 $t_{c,n}$. Let

$$M_n(t) \ := \ \#\left\{i \geq 1 : \ \{i\} \in \Pi_n(t)\right\}, \qquad t \geq 0,$$

which is the number of external branches at time t. Also recall

$$\widehat{T}^n_{\langle i \rangle} \ := \ (T^n_{\langle i \rangle} - t_{c,n})^+.$$

Lemma 9.3. *For $\ell, y \in \mathbb{N}$, there exist random variables $U_{1,y} \geq \cdots \geq U_{\ell,y}$ such that the following convergence results hold:*

(i) *For any bounded continuous function $g : \mathbb{R}^\ell \to \mathbb{R}$ and for fixed natural numbers $\ell \leq y$, as $n \to \infty$,*

$$\mathbf{E}\left[g\left(\log\log(n)\left(\widehat{T}^n_{\langle 1 \rangle}, \ldots, \widehat{T}^n_{\langle \ell \rangle}\right)\right) \,\middle|\, N_n(t_{c,n}), M_n(t_{c,n}) = y\right] \ \longrightarrow \ \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right]$$

in probability.

(ii) *For fixed ℓ, as $y \to \infty$,*

$$\left(U_{1,y} - \log y, \ldots, U_{\ell,y} - \log y\right) \ \xrightarrow{d} \ \left(U_1, \ldots, U_\ell\right),$$

where $U_1 > \cdots > U_\ell$ are the points of the Poisson point process of Theorem 9.1.

Proof. (i) We proceed in the same vein as in the proof of Lemma 8.3 (i). The strong Markov property, Corollary 1.4 (see also formula (1.2) in the first example) and Lemma 9.2 yield that

$$\mathbf{E}\left[g\left(z\left(\widehat{T}^n_{\langle 1 \rangle}, \ldots, \widehat{T}^n_{\langle \ell \rangle}\right)\right) \,\middle|\, N_n(t_{c,n}) = x, M_n(t_{c,n}) = y\right] \ \longrightarrow \ \mathbf{E}\left[g\left(U_{1,y}, \ldots, U_{\ell,y}\right)\right]$$

as $x \to \infty$ and $z/\log x \to 1$, where $(U_{1,y}, \ldots, U_{\ell,y})$ has the density

$$\ell! \binom{y}{\ell} \left(1 - e^{-u_\ell}\right)^{y-\ell} \prod_{i=1}^{\ell} e^{-u_i} du_1 \cdots du_\ell \tag{9.2}$$

for $u_1 \geq \cdots \geq u_\ell$. Moreover, from Lemma 9.2, we obtain

$$\log\left(N_n(t_{c,n})\right) \ = \ t_{c,n} + \mathcal{O}_P(1) \ = \ \log\log n + o_P\left(\log\log n\right)$$

as $n \to \infty$. Thus, replacing x and z above by $N_n(t_{c,n})$ and $\log\log n$, respectively, and invoking Skorohod's representation theorem once more, our claim follows.

(ii) Shifting the distribution from (9.2) by $\log y$, we arrive at the densities

$$\ell! \binom{y}{\ell} \left(1 - \frac{e^{-u_\ell}}{y}\right)^{y-\ell} y^{-\ell} \prod_{i=1}^{\ell} e^{-u_i} du_1 \cdots du_\ell$$

and their limit

$$e^{-e^{-u_\ell}} \prod_{i=1}^{\ell} e^{-u_i} du_i$$

as $y \to \infty$, which is the joint density of $U_1, \dots, U_\ell$. 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 Proposition 2.2 is no longer helpful and we may not simply apply (8.5). As a substitute, we shall use the following lemma.

Lemma 9.4. *As $n \to \infty$,*

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

Proof. Let $\mathcal{F}_k := \sigma\left(X, W_0, \dots, W_{k-1}\right)$ and

$$Z_k \ := \ \sum_{j=0}^{k\wedge\tau_n-1} \left(W_j - \frac{1}{X_j - 1} \right), \qquad k \geq 0.$$

In particular, we have $Z_0 = 0$. Given $\mathcal{F}_j$ 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 [28]). Thus, $(Z_k)_{k\in\mathbb{N}}$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$ with (predictable) quadratic variation

$$\langle Z \rangle_k \ := \ \sum_{j=0}^{k\wedge\tau_n-1} \mathbf{E}\left[(Z_{j+1} - Z_j)^2 \middle| \mathcal{F}_j \right] \ = \ \sum_{j=0}^{k\wedge\tau_n-1} \frac{1}{(X_j-1)^2} \qquad a.s.$$

Applying Doob's optional sampling theorem to the martingale $Z_k^2 - \langle Z \rangle_k$ yields

$$\mathbf{E}\left[Z_{\rho_{c,n}}^2 \right] \ = \ \mathbf{E}\left[\langle Z \rangle_{\rho_{c,n}} \right] \ = \ \mathbf{E}\left[\sum_{j=0}^{\rho_{c,n}-1} \frac{1}{(X_j-1)^2} \right] \ \leq \ \mathbf{E}\left[\sum_{k=X_{\rho_{c,n}-1}}^{\infty} \frac{1}{(k-1)^2} \right] \tag{9.3}$$

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

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

By Lemma 9.2 and dominated convergence, the right-hand term converges to 0 as $n \to \infty$, implying

$$\sum_{j=0}^{\rho_{c,n}-1} \left(W_j - \frac{1}{X_j} \right) \ = \ Z_{\rho_{c,n}} + \mathcal{O}_P\left(\frac{4}{X_{\rho_{c,n}-1}} \right) \ = \ o_P(1)$$

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 $N_n(t_{c,n})$. Because of the property that exponential times lack memory, the residual time is exponential with parameter $N_n(t_{c,n})$. Thus, by Lemma 9.2, the residual time converges to 0 in probability, which provides our claim. □

Lemma 9.5. *For the number of external branches at time $t_{c,n}$, we have the following results:*

(i) *For $c > 1$,*

$$\mathbf{E}\left[M_n(t_{c,n}) \mid N_n\right] \xrightarrow{d} c\,E$$

as $n \to \infty$, where E denotes a standard exponential random variable.

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

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

as well as

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

Proof. (i) Using the representation from Lemma 7.1 (i) and a Taylor expansion as in (4.4), we get

$$\mathbf{E}\left[Y_{\rho_{c,n}-1} \mid N_n\right] = X_{\rho_{c,n}-1} \exp\left(-\sum_{j=1}^{\rho_{c,n}-1} \frac{1}{X_j} + \mathcal{O}_P\left(X^{-1}_{\rho_{c,n}-1}\right)\right)$$

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

$$\mathbf{E}\left[M_n(t_{c,n}) \mid N_n\right] = N_n(t_{c,n}) \exp\left(-\sum_{j=1}^{\rho_{c,n}-1} \frac{1}{X_j} + \mathcal{O}_P\left(N_n\left(t_{c,n}\right)^{-1}\right)\right). \tag{9.4}$$

From Lemma 9.4 and Lemma 9.2, it follows

$$\mathbf{E}\left[M_n(t_{c,n}) \mid N_n\right] = N_n(t_{c,n}) \exp\left(-t_{c,n} + o_P(1)\right).$$

Hence, Lemma 9.2 implies our claim.

(ii) Chebyshev's inequality and Lemma 7.1 (ii) provide

$$\begin{aligned}
\mathbf{P}&\left(\left|M_n(t_{c,n}) - \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]\right| > c^{1/2+\varepsilon}\right)\\
&= \mathbf{E}\left[\mathbf{P}\left(\left|M_n(t_{c,n}) - \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]\right| > c^{1/2+\varepsilon}\middle|N_n\right)\right]\\
&\le \mathbf{E}\left[\frac{\mathbf{Var}\left(M_n(t_{c,n}) \mid N_n\right)}{c^{1+2\varepsilon}} \wedge 1\right]\\
&\le \mathbf{E}\left[\frac{\mathbf{E}\left(M_n(t_{c,n}) \mid N_n\right)}{c^{1+2\varepsilon}} \wedge 1\right].
\end{aligned}$$

From statement (i) it follows that

$$\limsup_{n\to\infty} \mathbf{P}\left(\left|M_n(t_{c,n}) - \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]\right| > c^{1/2+\varepsilon}\right) \le \mathbf{E}\left[\frac{cE}{c^{1+2\varepsilon}} \wedge 1\right] \le c^{-2\varepsilon},$$

which entails the first claim.

Similarly, Markov's inequality yields

$$\limsup_{n\to\infty} \mathbf{P}\left(M_n(t_{c,n}) > c^{1+\varepsilon}\right) \;\le\; \limsup_{n\to\infty} \mathbf{E}\left[\frac{\mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]}{c^{1+\varepsilon}} \wedge 1\right] \;\le\; c^{-\varepsilon},$$

giving the second claim.

Furthermore, we have

$$\begin{aligned}\mathbf{P}\big(M_n(t_{c,n}) < c^{1-\varepsilon}\big) \;\le\; & \mathbf{P}\left(\mathbf{E}\left[M_n(t_{c,n})\,|N_n\right] < 2c^{1-\varepsilon}\right) \\ & + \mathbf{P}\left(\left|M_n(t_{c,n}) - \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]\right| > c^{1-\varepsilon}\right)\end{aligned}$$

and, consequently, in view of part (i),

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}\big(M_n(t_{c,n}) < c^{1-\varepsilon}\big) \;\le\; & \mathbf{P}\left(E < 2c^{-\varepsilon}\right) \\ & + \limsup_{n\to\infty} \mathbf{P}\left(\left|M_n(t_{c,n}) - \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]\right| > c^{1-\varepsilon}\right).\end{aligned}$$

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, by means of the first claim of part (ii). $\square$

With these preparations, we now turn to the proof of Theorem 9.1.

Proof of Theorem 9.1. The strategy of this proof resembles that of Theorem 8.1. However, additional care is required to separate the impact of the parts $\check{T}_i^n$ and $\widehat{T}_i^n$. For this purpose, we consider the functions

$$g(x_1,\dots,x_\ell) \;:=\; \exp\left(i\left(\theta_1 x_1 + \dots + \theta_\ell x_\ell\right)\right) \qquad \text{and} \qquad h(x) \;:=\; \exp\left(i\left(\theta_1 + \dots + \theta_\ell\right)x\right),$$

where $\theta_i \in \mathbb{R}$ for $1 \le i \le n$. It is sufficient to prove

$$\mathbf{E}\left[g\left(\log\log(n)\left(T_{\langle 1\rangle}^n - t_n\right),\dots,\log\log(n)\left(T_{\langle \ell\rangle}^n - t_n\right)\right)\right] \;\longrightarrow\; \mathbf{E}\left[g\left(U_1 - G,\dots,U_\ell - G\right)\right]$$

as $n \to \infty$. We bound the difference of the terms on both sides. Recalling

$$t_n \;=\; t_{c,n} + \frac{\log c}{\log\log n},$$

we see that, on the event $\{M_n(t_{c,n}) \ge \ell\}$, it holds $T_{\langle i\rangle}^n \;=\; \widehat{T}_{\langle i\rangle}^n + t_{c,n}$ and, therefore,

$$\log\log(n)\big(T_{\langle j\rangle}^n - t_n\big) \;=\; \left(\log\log(n)\,\widehat{T}_{\langle j\rangle}^n - \log M_n(t_{c,n})\right) + \log\frac{M_n(t_{c,n})}{c} \tag{9.5}$$

for $1 \le j \le \ell$. In conjunction with the independence of $(U_1,\dots,U_\ell)$ and the Gumbel random variable G, it follows that

$$\begin{aligned}&\left|\mathbf{E}\left[g\left(\log\log(n)\left(T_{\langle 1\rangle}^n - t_n\right),\dots,\log\log(n)\left(T_{\langle \ell\rangle}^n - t_n\right)\right)\right] - \mathbf{E}\left[g\left(U_1 - G,\dots,U_\ell - G\right)\right]\right| \\ &\quad\le \left|\mathbf{E}\left[g\left(V_{c,n}\right)h\left(\log\frac{M_n(t_{c,n})}{c}\right)\right] - \mathbf{E}\left[g\left(U_1,\dots,U_\ell\right)\right]\mathbf{E}\left[h\left(-G\right)\right]\right| \\ &\qquad + 2\,\mathbf{P}\left(M_n(t_{c,n}) < \ell\right),\end{aligned} \tag{9.6}$$

where, in view of (9.5), we now set

$$V_{c,n} \ := \ \Big(\log\log(n)\,\widehat{T}^n_{\langle 1\rangle} - \log M_n(t_{c,n}), \dots, \log\log(n)\,\widehat{T}^n_{\langle \ell\rangle} - \log M_n(t_{c,n})\Big)\,.$$

Let us estimate the first term on the right-hand side of (9.6). We have

$$\begin{aligned}
&\Big|\mathbf{E}\Big[g\,(V_{c,n})\ h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\Big] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\mathbf{E}\,[h\,(-G)]\,\Big| \\
&\leq \ \Big|\mathbf{E}\Big[g\,(V_{c,n})\,h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\Big] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\mathbf{E}\Big[h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\Big]\Big| \\
&\qquad + \ \Big|\mathbf{E}\Big[h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\Big] \ - \ \mathbf{E}\Big[h\Big(\log\frac{\mathbf{E}\,[M_n(t_{c,n})\,|N_n]}{c}\Big)\Big]\Big| \\
&\qquad + \ \Big|\mathbf{E}\Big[h\Big(\log\frac{\mathbf{E}\,[M_n(t_{c,n})\,|N_n]}{c}\Big)\Big] \ - \ \mathbf{E}\,[h\,(-G)]\Big| \\
&=: \Delta'_{c,n} + \Delta''_{c,n} + \Delta'''_{c,n} \qquad \text{(say).}
\end{aligned}$$

We bound $\Delta'_{c,n}, \Delta''_{c,n}$ and $\Delta'''_{c,n}$ separately. For $\Delta'_{c,n}$, we first consider conditional expectations. For $c > 1$, we have

$$\begin{aligned}
&\Big|\mathbf{E}\Big[g\,(V_{c,n})\,h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\ \Big|\,N_n(t_{c,n})\Big] \\
&\qquad\qquad - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\mathbf{E}\Big[h\Big(\log\frac{M_n(t_{c,n})}{c}\Big)\ \Big|\,N_n(t_{c,n})\Big]\Big| \\
&\leq \sum_{\sqrt{c}\leq y\leq c^2}\Big|\Big(\mathbf{E}\,[\,g\,(V_{c,n})\,|\ N_n(t_{c,n}),\, M_n(t_{c,n})=y] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\Big)h\Big(\log\frac{y}{c}\Big)\Big| \\
&\qquad\quad \cdot \mathbf{P}\,\big(M_n(t_{c,n})=y\,\big|\,N_n(t_{c,n})\big) \\
&\qquad + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})<\sqrt{c}\,\big|\,N_n(t_{c,n})\big) + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})>c^2\,\big|\,N_n(t_{c,n})\big) \\
&\leq \max_{\sqrt{c}\leq y\leq c^2}\big|\mathbf{E}\,[\,g\,(V_{c,n})\,|\ N_n(t_{c,n}),\, M_n(t_{c,n})=y] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\big| \\
&\qquad + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})<\sqrt{c}\,|\,N_n(t_{c,n})\big) \ + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})>c^2\,|\,N_n(t_{c,n})\big) \\
&\leq \max_{\sqrt{c}\leq y\leq c^2}\big|\mathbf{E}\,[g\,(U_{1,y}-\log y,\dots,U_{\ell,y}-\log y)] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\big| + o_P(1) \\
&\qquad + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})<\sqrt{c}\,|\,N_n(t_{c,n})\big) \ + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})>c^2\,|\,N_n(t_{c,n})\big)
\end{aligned}$$

as $n\to\infty$, where we used Lemma 9.3 (i) in the last step. Without loss of generality, we may assume that the right-hand $o_P(\cdot)$-term is bounded by 1. Hence, taking expectations, we obtain via dominated convergence that

$$\begin{aligned}
\Delta'_{c,n} \ \leq \ &\max_{\sqrt{c}\leq y\leq c^2}\big|\mathbf{E}\,[g\,(U_{1,y}-\log y,\dots,U_{\ell,y}-\log y)] \ - \ \mathbf{E}\,[g\,(U_1,\dots,U_\ell)]\,\big| \ + \ o(1) \\
&+ \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})<\sqrt{c}\big) \ + \ 2\,\mathbf{P}\,\big(M_n(t_{c,n})>c^2\big)\,.
\end{aligned}$$

Secondly, 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,

$$\begin{aligned}
&\Delta''_{c,n} \\
&\quad \le \left| \mathbf{E}\left[h\left(\log \frac{M_n(t_{c,n})}{c}\right) - h\left(\log \frac{\mathbf{E}\left[M_n(t_{c,n})\,|N_n\right]}{c}\right) ;\, M_{t_{c,n}} \wedge \mathbf{E}\left[M_n(t_{c,n})\,|N_n\right] \ge c^{3/4} \right] \right| \\
&\qquad\quad + \; 2\,\mathbf{P}\left(M_n(t_{c,n}) < c^{3/4}\right) \; + \; 2\,\mathbf{P}\left(\mathbf{E}\left[M_n(t_{c,n})\,|N_n\right] < c^{3/4}\right) \\
&\quad \le \; 2\,\mathbf{P}\left(|M_n(t_{c,n}) \; - \; \mathbf{E}\left[M_n(t_{c,n})|\,N_n\right]| > c^{2/3}\right) \; + \; |\theta_1 + \cdots + \theta_\ell|\, c^{1/4-1/3} \qquad (9.7) \\
&\qquad\quad + \; 2\,\mathbf{P}\left(M_n(t_{c,n}) < c^{3/4}\right) \; + \; 2\,\mathbf{P}\left(\mathbf{E}\left[M_n(t_{c,n})\,|N_n\right] < c^{3/4}\right).
\end{aligned}$$

Lastly, Lemma 9.5 (i) provides the convergence of $\Delta'''_{c,n}$ to 0 as $n \to \infty$. Consequently, combining equation (9.6) to (9.7), using Lemma 9.5, and grouping terms yield

$$\begin{aligned}
&\limsup_{n\to\infty} \left| \mathbf{E}\left[g\left(V_{c,n}\right) h\left(\log \frac{M_n(t_{c,n})}{c}\right) \right] \; - \; \mathbf{E}\left[g\left(U_1, \ldots, U_\ell\right)\right] \mathbf{E}\left[h\left(-G\right)\right] \right| \\
&\le \max_{\sqrt{c} \le y \le c^2} \left| \mathbf{E}\left[g\left(U_{1,y} - \log y, \ldots, U_{\ell,y} - \log y\right)\right] \; - \; \mathbf{E}\left[g\left(U_1, \ldots, U_\ell\right)\right] \right| \\
&\quad + \; 2\, \limsup_{n\to\infty} \mathbf{P}\left(M_n(t_{c,n}) < \ell\right) \; + \; 2\, \limsup_{n\to\infty} \mathbf{P}\left(M_n(t_{c,n}) < \sqrt{c}\right) \\
&\quad + \; 2\, \limsup_{n\to\infty} \mathbf{P}\left(M_n(t_{c,n}) < c^{3/4}\right) \; + \; 2\, \limsup_{n\to\infty} \mathbf{P}\left(M_n(t_{c,n}) > c^2\right) \\
&\quad + \; 2\, \limsup_{n\to\infty} \mathbf{P}\left(|M_n(t_{c,n}) \; - \; \mathbf{E}\left[M_n(t_{c,n})|\,N_n\right]| > c^{2/3}\right) \\
&\quad + \; 2\left(1 - e^{-c^{-1/4}}\right) \; + \; |\theta_1 + \cdots + \theta_\ell|\, c^{-1/12}.
\end{aligned}$$

Finally, taking the limit $c \to \infty$, the right-hand terms converge to 0 by Lemma 9.3 (ii) and Lemma 9.5. This finishes the proof. □

References

[1] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press, Cambridge. MR-0898871

[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. MR-2156553

[3] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197**, 247–276. MR-1652734

[4] Boros, G. and Moll, V. (2004). *Irresistible Integrals: Symbolics, Analysis and Experiments in the Evaluation of Integrals*. Cambridge University Press, Cambridge. MR-2070237

[5] 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. MR-2365627

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

[7] Dahmer, I., Kersting, G. and Wakolbinger, A. (2014). The total external branch length of Beta-coalescents. *Combin. Probab. Comput.* **23**, 1010–1027. MR-3265836

[8] 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.

[9] 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. MR-3027896

[10] Dhersin, J.-S. and Möhle, M. (2013). On the external branches of coalescents with multiple collisions. *Electron. J. Probab.* **18**, 1–11. MR-3040550

[11] Dhersin, J.-S. and Yuan, L. (2015). On the total length of external branches for Beta-coalescents. *Adv. in Appl. Probab.* **47**, 693–714. MR-3406604

[12] Diehl, C. S. and Kersting, G. (2019). Tree lengths for general Λ-coalescents and the asymptotic site frequency spectrum around the Bolthausen-Sznitman coalescent. *Ann. Appl. Probab.* **29**, 2700–2743. MR-4019873

[13] 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.

[14] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. John Wiley & Sons, New York. MR-0270403

[15] 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. MR-2554368

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

[17] Janson, S. and Kersting, G. (2011). On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16**, 2203–2218. MR-2861672

[18] Kersting, G., Pardo, J., and Siri-Jégousse, A. (2014). Total internal and external lengths of the Bolthausen-Sznitman coalescent. *J. Appl. Probab.*, **51**, 73–86. MR-3317351

[19] Kersting, G., Schweinsberg, J. and Wakolbinger, A. (2014). The evolving beta coalescent. *Electron. J. Probab.* **19**, 1–27. MR-3238784

[20] 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. MR-3825890

[21] Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

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

[23] Limic, V. and Sturm, A. (2006). The spatial Λ-coalescent. *Electron. J. Probab.* **11**, 363–393. MR-2223040

[24] 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. MR-2684740

[25] 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. MR-3333734

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

[27] 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.

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

[29] Seneta, E. (1973). A Tauberian theorem of E. Landau and W. Feller. *Ann. Probab.* **1**, 1057–1058. MR-0358133

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

[31] Schweinsberg, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Probab.* **5**, 1–11. MR-1736720

[32] Schweinsberg, J. (2017). Rigorous results for a population model with selection II: genealogy of the population. *Electron. J. Probab.* **38**, 1–54. MR-3646064

[33] 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. MR-3466919

[34] 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.

[35] Villandré, 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.

[36] 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.

[37] 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.

[38] Yuan, L. (2014). On the measure division construction of Λ-coalescents. *Markov Process. Related Fields.* **20**, 229–264. MR-3241525

Acknowledgments. We are grateful to the anonymous referees for their insightful comments, which allowed us to improve the paper's presentation.