The Annals of Applied Probability
2019, Vol. 29, No. 5, 2700–2743
https://doi.org/10.1214/19-AAP1462

TREE LENGTHS FOR GENERAL Λ-COALESCENTS AND THE ASYMPTOTIC SITE FREQUENCY SPECTRUM AROUND THE BOLTHAUSEN–SZNITMAN COALESCENT[1]

BY CHRISTINA S. DIEHL AND GÖTZ KERSTING

Goethe Universität

We study tree lengths in Λ-coalescents without a dust component from a sample of n individuals. For the total length of all branches and the total length of all external branches, we present laws of large numbers in full generality. The other results treat regularly varying coalescents with exponent 1, which cover the Bolthausen–Sznitman coalescent. The theorems contain laws of large numbers for the total length of all internal branches and of internal branches of order a (i.e., branches carrying a individuals out of the sample). These results immediately transform to sampling formulas in the infinite sites model. In particular, we obtain the asymptotic site frequency spectrum of the Bolthausen–Sznitman coalescent. The proofs rely on a new technique to obtain laws of large numbers for certain functionals of decreasing Markov chains.

1. Introduction and main results. Λ-coalescents are established models for family trees of a sample of individuals from some large population. Its most prominent representative, the Kingman coalescent [18], is widely used in population genetics. More recently, the Bolthausen–Sznitman coalescent [6] gained attention for models including selection. The class of Beta-coalescents with parameter $1 < \alpha < 2$ has been applied to marine populations [5]. In this paper, we investigate branch lengths and sampling formulas in the infinite sites model for the whole class of Λ-coalescents without a dust component, which covers all these special cases.

Λ-coalescents have been introduced by Pitman [20] and Sagitov [21] as Markov processes whose states are partitions of $\mathbb{N}$ and whose evolution may be imagined as a random tree. In this paper, we identify a Λ-coalescent with an induced sequence of n-coalescents, $n \in \mathbb{N}$, by restricting partitions to the subsets $\{1, \dots, n\}$ of $\mathbb{N}$. These n-coalescents are considered to be Markovian models for the family trees of a sample of n individuals. If such a tree contains $b \in \{2, \dots, n\}$ lineages at the moment $t \geq 0$ backwards in time (representing the ancestors living at that

Received April 2018; revised October 2018.
[1]Supported in part by the DFG Priority Programme SPP 1590 "Probabilistic Structures in Evolution".
MSC2010 subject classifications. Primary 60J75; secondary 60F05, 60J27, 92D25.
Key words and phrases. Λ-coalescent, Bolthausen–Sznitman coalescent, law of large numbers, tree length, infinite sites model, sampling formula, site frequency spectrum, decreasing Markov chain.

moment), then it is assumed that of those, $k \in \{2, \dots, b\}$ specified lines merge at rate

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

to one line. Here, Λ denotes any finite, nonvanishing measure on $[0, 1]$. The resulting tree consists of n leaves, τ_n merging events, and a root at time $\tilde{\tau}_n > 0$, representing the MCRA (most recent common ancestor). Its branches have lengths which specify lifetimes. There are external branches ending in leaves on the one hand, and internal branches ending in mergers on the other. For the detailed partition valued picture, we refer to the survey [4].

In the sequel, we work with the Markovian counting process $N_n = (N_n(t))_{t \geq 0}$, where $N_n(t)$ denotes the number of lineages at time $t \geq 0$. Thus, $N_n(0) = n$ and $N_n(\tilde{\tau}_n) = 1$. For convenience, we set $N_n(t) := 1$ for $t > \tilde{\tau}_n$. Then for $b = 2, \dots, n$ the numbers

$$\lambda(b) = \sum_{k=2}^{b} \binom{b}{k} \lambda_{b,k}$$

give the jump rates of the process N_n and

$$\mu(b) = \sum_{k=2}^{b} (k-1) \binom{b}{k} \lambda_{b,k}$$

its rate of decrease since a merger of k lineages results in a downward jump of the block counting process of size $k - 1$. These two sequences can be naturally extended to positive continuous functions $\lambda, \mu : [2, \infty) \to \mathbb{R}$ (see the formulas (4) and (5) below).

In this paper, we focus on the class of Λ-coalescents without a dust component. To put it briefly, in this case the rate, at which within the n-coalescents a single lineage merges with some others, diverges as the sample size n tends to infinity. They are characterized by the condition (Pitman [20])

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

In particular, they cover Λ-coalescents coming down from infinity. These are the coalescents with the property that the absorption times $\tilde{\tau}_n$ are bounded in probability uniformly in n. They are distinguished by the criterion (Schweinsberg [22])

$$\int_2^{\infty} \frac{dx}{\mu(x)} < \infty.$$

We shall analyze the lengths of the whole n-coalescents as well as of different parts. They play an important role in the infinite sites model introduced by Kimura [17]. In this model, each mutation affects a different site in the DNA. The mutations are distributed along the branches of the coalescent depending on their appearance in the past. Mathematically, they build a homogeneous Poisson point process with rate $\theta > 0$. Their total number $S(n)$ counts the segregating sites in the sample of size n underlying the coalescent and are thus closely tied to its total length ℓ_n. Mutations which are located on an external branch appear only once in the sample, these are the singleton polymorphisms (see Wakeley [25], page 103). Accordingly, their number is linked to the total length $\widehat{\ell}_n$ of all external branches. Likewise, the number of mutations visible repeatedly in the sample corresponds to the total internal length $\check{\ell}_n$.

NOTATION. For two sequences A_n and B_n of positive random variables, we write $A_n \overset{1}{\sim} B_n$ and $A_n \overset{P}{\sim} B_n$ if the sequence A_n/B_n converges to 1 in the L_1-sense or in probability, respectively. The notation $A_n = O_P(B_n)$ means that the sequence A_n/B_n is tight and $A_n = o_P(B_n)$ that A_n/B_n converges to 0 in probability.

Now we come to the main results of this paper. As a first notion let

$$\ell_n := \int_0^{\tilde{\tau}_n} N_n(t)\,dt, \quad n \geq 1,$$

be the total length of the coalescent tree.

THEOREM 1. *Assume that the Λ-coalescent has no dust component. Then as $n \to \infty$*

$$\ell_n \overset{1}{\sim} \int_2^n \frac{x}{\mu(x)}\,dx. \tag{1}$$

This is an intuitive approximation: if the counting process N_n takes the value x, then there are currently x lines and $1/\mu(x)$, the reciprocal of the rate of decrease, indicates how long on average N_n will stay close to x. Observe that the right-hand integral diverges for $n \to \infty$: because of Lemma 1(i) below, the function $\mu(x)/(x(x-1))$ is decreasing and, therefore, $\int_2^n \frac{x}{\mu(x)}\,dx \geq \frac{2}{\mu(2)} \log(n-1)$. This lower bound is attained by the Kingman coalescent.

Berestycki et al. [2] conjectured Theorem 1 and proved it for Λ-coalescents coming down from infinity in the framework of convergence in probability by using instead of $\mu(x)$ the asymptotically equivalent quantity

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

For the Kingman coalescent, the result was earlier obtained by Watterson [26] and for the Bolthausen–Sznitman coalescent by Drmota et al. [10].

In the context of the infinite sites model, Theorem 1 may be restated directly in terms of the number $S(n)$ of segregating sites as

$$S(n) \overset{1}{\sim} \theta \int_2^n \frac{x}{\mu(x)}\,dx,$$

due to the assumption that mutations appear according to a homogeneous Poisson point process with rate θ.

Next, we define the total external length

$$\widehat{\ell}_n := \int_0^{\tilde{\tau}_n} \widehat{N}_n(t)\,dt,$$

where $\widehat{N}_n(t)$ denotes the number of external branches extant at time t. In contrast to the previous theorem, the following result contains a statement only of convergence in probability (but see Theorem 4 below, which shows that, indeed, L_1-convergence holds under the stronger conditions of that theorem).

THEOREM 2. *Suppose that the Λ-coalescent has no dust component. Then as* $n \to \infty$

$$\widehat{\ell}_n \overset{P}{\sim} \frac{n^2}{\mu(n)}. \tag{2}$$

An intuitive explanation goes as follows: $\mu(n)/n$ is the rate of decrease per individual at time 0. Therefore, it is plausible that each external branch length is of order $n/\mu(n)$, and the external branch length in total results approximately in $n^2/\mu(n)$. The sequence $\mu(n)/n^2$, $n \geq 1$, has the limit $\Lambda(\{0\})/2$ (see formula (8) below). Thus, the total external lengths diverge in probability if and only if $\Lambda(\{0\}) = 0$. We point out that $x/\mu(x)$ is a decreasing function (see Lemma 1(i) below), therefore, the integral appearing in (1) exceeds the corresponding term in (2).

The proof rests on a close relation between the functions λ and μ, which seems unobserved until now. To describe it, let us introduce another function $\kappa : [2, \infty) \to \mathbb{R}$ given by

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

which could be named the rate of decrease per capita. Then we have the approximation

$$\lambda(x) \sim x^2 \kappa'(x)$$

as $x \to \infty$ (see Lemma 1(ii) below).

In the special case of Beta-coalescents, Theorem 2 follows from [7] and [9]. For Λ-coalescents with a dust component, the picture is rather different. Then ℓ_n/n as well as $\widehat{\ell}_n/n$ converge in distribution to one and the same nondegenerate limit law (see [19]). Fluctuation results on the total length or the total external length of Λ-coalescents with no dust component are known only in special cases [7, 10, 13, 14, 24].

Theorem 2 again allows a reformulation in terms of the infinite site model. Letting $M_1(n)$ be the number of singletons, we obtain for $\Lambda(\{0\}) = 0$,

$$M_1(n) \overset{P}{\sim} \theta \frac{n^2}{\mu(n)},$$

whereas for $\Lambda(\{0\}) > 0$ it follows that $M_1(n)$ is asymptotically Poisson with parameter $2\theta/\Lambda(\{0\})$.

REMARK. From our proofs, we will gain further insight into the structure of the coalescents. For $0 < c < 1$, let

$$\tilde{\rho}_n := \inf\{t \geq 0 : N_n(t) \leq cn\}$$

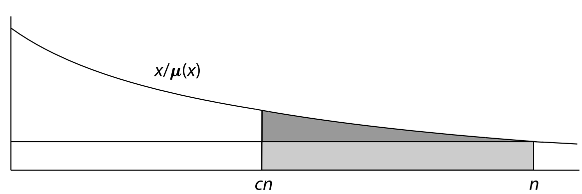
be the first moment when the number of lineages falls below cn, and let

$$\ell_n^* := \int_0^{\tilde{\rho}_n} N_n(t)\, dt, \qquad \widehat{\ell}_n^* := \int_0^{\tilde{\rho}_n} \widehat{N}_n(t)\, dt$$

be the respective lengths up to this moment. Then, from (23) and (32) below, we have

$$\ell_n^* \overset{1}{\sim} \int_{cn}^{n} \frac{x}{\mu(x)}\, dx \quad \text{and} \quad \widehat{\ell}_n^* \overset{P}{\sim} (1-c)\frac{n^2}{\mu(n)}.$$

The picture provides an illustration.



ℓ_n^* and $\widehat{\ell}_n^*$ are the areas of the total grey region and of the lighter part, respectively.

It is tempting to expect an analogous result for the total internal length

$$\check{\ell}_n := \ell_n - \widehat{\ell}_n.$$

This is certainly true in cases where the total external length $\widehat{\ell}_n$ does not exceed the total internal length $\check{\ell}_n$. Then we have

$$\check{\ell}_n \overset{P}{\sim} \int_2^n \frac{x}{\mu(x)}\, dx - \frac{n^2}{\mu(n)} \sim \int_2^n \left(\frac{x}{\mu(x)} - \frac{n}{\mu(n)}\right) dx.$$

To provide some examples, let us introduce the following class of coalescents.

DEFINITION. We call the Λ-coalescent regularly varying with exponent $0 \le \alpha \le 2$ if $\Lambda(\{0\}) = 0$ and if, as $y \to 0$,

$$\int_{(y,1]} \frac{\Lambda(dp)}{p^2} \sim y^{-\alpha} L\left(\frac{1}{y}\right)$$

with a function L that is slowly varying at infinity.

This generalizes the notion of Berestycki et al. [2] of a strong regularly varying coalescent.

EXAMPLES. (i) The Kingman case: if $\Lambda(\{0\}) > 0$, then $\mu(x)/x^2 \sim \Lambda(\{0\})/2$ and, therefore,

$$\ell_n \overset{1}{\sim} \frac{2\log n}{\Lambda(\{0\})} \quad \text{and} \quad \widehat{\ell}_n \overset{P}{\sim} \frac{2}{\Lambda(\{0\})}$$

as $n \to \infty$. Here, the internal total length completely dominates the external ones, and we have

$$\check{\ell}_n \overset{1}{\sim} \frac{2\log n}{\Lambda(\{0\})}.$$

(ii) Regularly varying coalescents with exponent $1 < \alpha < 2$ come down from infinity, as can be easily checked by the above criterion. They fulfill (see Lemma 2(ii) below)

$$\mu(x) \sim \frac{\Gamma(2-\alpha)}{\alpha - 1} x^{\alpha} L(x)$$

as $x \to \infty$. Hence (having in mind how to integrate regularly varying functions (see Theorem 1(b), Section VIII.9 in Feller [11]),

$$\ell_n \overset{1}{\sim} \frac{\alpha - 1}{(2-\alpha)\Gamma(2-\alpha)} \frac{n^{2-\alpha}}{L(n)} \quad \text{and} \quad \widehat{\ell}_n \overset{P}{\sim} \frac{\alpha - 1}{\Gamma(2-\alpha)} \frac{n^{2-\alpha}}{L(n)}$$

and consequently

$$\check{\ell}_n \overset{P}{\sim} \frac{(\alpha-1)^2}{(2-\alpha)\Gamma(2-\alpha)} \frac{n^{2-\alpha}}{L(n)} \overset{P}{\sim} \frac{\alpha - 1}{2-\alpha} \widehat{\ell}_n.$$

Here, the external and the internal length are of the same order. As an application, one may use the quantity $2 - \hat{\ell}_n/\ell_n$ as an estimator for the parameter α. In the special case of Beta$(2-\alpha, \alpha)$-coalescents, this was already discussed in more detail in Example 9 of [16].

In some other cases, $\check{\ell}_n$ is of smaller order than $\widehat{\ell}_n$. Then a large part of internal length will be located close to the coalescent tree's root, where extremely large mergers may take over (which is definitely the case for Λ-coalescents not coming down from infinity). Still the internal length may obey the law of large numbers suggested above. The following theorem presents a situation of special interest. Define for a slowly varying function L and for $x \geq 1$

$$L^*(x) := \int_1^x \frac{L(y)}{y}\,dy.$$

THEOREM 3. *Assume that the Λ-coalescent has no dust component and is regularly varying with exponent $\alpha = 1$. Then*

$$\check{\ell}_n \overset{1}{\sim} \int_2^n \left(\frac{x}{\mu(x)} - \frac{n}{\mu(n)}\right) dx$$

and

$$\int_2^n \left(\frac{x}{\mu(x)} - \frac{n}{\mu(n)}\right) dx \sim \frac{nL(n)}{L^*(n)^2}$$

as $n \to \infty$.

Below we prove (Lemma 2(ii)) that the function L^* is slowly varying at infinity, too, and that $L(x) = o(L^*(x))$ as $x \to \infty$. In comparison with Theorem 2, we see that $\check{\ell}_n = o_P(\widehat{\ell}_n)$ for regularly varying coalescents with exponent 1. In particular, for the Bolthausen–Sznitman coalescent Theorems 2 and 3 yield

$$\widehat{\ell}_n \overset{P}{\sim} \frac{n}{\log n} \quad \text{and} \quad \check{\ell}_n \overset{1}{\sim} \frac{n}{\log^2 n}.$$

These approximations were already obtained in Dhersin and Möhle [9] and Kersting, Pardo and Siri-Jégousse [15], respectively.

Our last object of investigation concerns lengths of higher order. A branch of order $a \geq 2$ is by definition an internal branch carrying a subtree with a leaves out of the original sample. In this context we consider external branches as branches of order 1. Denote the number of branches of order $a \geq 1$ present at time $t \geq 0$ as $\widehat{N}_{n,a}(t)$, notably $\widehat{N}_{n,1}(t) = \widehat{N}_n(t)$. Then the total length of all these branches is given by

$$\widehat{\ell}_{n,a} = \int_0^{\tilde{\tau}_n} \widehat{N}_{n,a}(t)\,dt, \quad a \geq 1.$$

THEOREM 4. *Suppose that the Λ-coalescent has no dust component and is regularly varying with exponent $\alpha = 1$. Then*

$$\widehat{\ell}_{n,1} \overset{1}{\sim} \frac{n}{L^*(n)},$$

whereas for $a \geq 2$

$$\widehat{\ell}_{n,a} \overset{1}{\sim} \frac{1}{(a-1)a} \frac{nL(n)}{L^*(n)^2}$$

as $n \to \infty$.

These formulas have interesting applications to the site frequency spectrum. It consists of the counts $M_a(n)$, $a \geq 1$, specifying the numbers of mutations located on branches of order a, which can be distinguished in the various DNAs of the sample. The theorem yields

$$M_1(n) \stackrel{1}{\sim} \theta \frac{n}{L^*(n)} \quad \text{and} \quad M_a(n) \stackrel{1}{\sim} \frac{\theta}{(a-1)a} \frac{nL(n)}{L^*(n)^2}, \quad a \geq 2.$$

A corresponding result in the so-called infinite allele model was obtained by Basdevant and Goldschmidt [1]. They deal with the allele frequency spectrum and consider the special case of the Bolthausen–Sznitman coalescent. Analogue results for Beta-coalescents coming down from infinity were presented by Berestycki, Berestycki and Schweinsberg [3] and, more generally, for strongly regular varying coalescents with exponent $1 < \alpha < 2$ by Berestycki, Berestycki and Limic [2]. We conjecture that these results can be further extended to regular varying coalescents. Computational procedures for the general site frequency spectrum were established by Spence et al. [23].

Theorem 4 illustrates that mutations mainly show up on external branches, whereas the tree structure of the coalescent becomes visible only at branches of higher order. This reflects that, for regularly varying coalescents with exponent $\alpha = 1$, mergers occur preferentially at a later time and close to the MRCA. Our proof will show that, with probability asymptotically equal to 1, any mutation seen in exactly $a \geq 2$ individuals stems from an internal branch of order a arose from one single merger. This is similar to the findings of Basdevant and Goldschmidt.

Closing this Introduction, we briefly discuss our methods of proof, which differ from other approaches in the literature. They rest upon L_2-considerations and elementary martingale estimates, and they may well find further applications as indicated in two examples below. These methods deal with the time-discrete Markov chain $n = X_0 > X_1 > \cdots > X_{\tau_n} = 1$, embedded in the Markov process N_n (or more generally with decreasing Markov chains). Let

$$\Delta_i := X_{i-1} - X_i, \quad i \geq 1,$$

denote its downwards jump size resulting from the ith merger. We shall present different laws of large numbers for expressions of the form $\sum_{i=0}^{\rho_n - 1} f(X_i)$, with some function $f : [2, \infty) \to \mathbb{R}$ and with stopping times ρ_n of the form

$$\rho_n := \min\{i \geq 0 : X_i \leq r_n\},$$

where r_n, $n \geq 1$, is some sequence of positive numbers.

Our approach specifies the following intuition:

$$\sum_{i=0}^{\rho_n - 1} f(X_i) \approx \sum_{i=0}^{\rho_n - 1} f(X_i) \frac{\Delta_{i+1}}{\nu(X_i)} \approx \int_{r_n}^{n} f(x) \frac{dx}{\nu(x)} \tag{3}$$

with

$$\nu(x) := \frac{\mu(x)}{\lambda(x)}, \quad x \geq 2.$$

Thus,

$$\nu(b) = \sum_{k=2}^{b}(k-1)\binom{b}{k}\frac{\lambda_{b,k}}{\lambda(b)} = \sum_{k=2}^{b}(k-1)\mathbf{P}(\Delta_{i+1} = k-1 \mid X_i = b)$$

or

$$\nu(b) = \mathbf{E}[\Delta_{i+1} \mid X_i = b], \quad b \geq 2.$$

The rationale behind this intuition is that the differences of both sums in (3) are small, because they stem from a martingale, whereas the second sum may be considered as a Riemann approximation of the right-hand integral. This latter approximation requires that the jump sizes Δ_{i+1} are small compared to the values X_i of the Markov chain and is only ensured if the time

$$\tilde{\rho}_n := \inf\{t \geq 0 : N_n(t) = X_{\rho_n}\} = \inf\{t \geq 0 : N_n(t) \leq r_n\}$$

of entrance into the interval $[1, r_n]$ converges to 0 in probability (observe that these random times generalize the notion used in the above remark). Thus, we strive toward small time approximations.

The quadratic variation of the previously mentioned martingale will be estimated along the following lines: since $\Delta_{i+1} = o_P(X_i)$ in the range of the small time approximation, we have

$$\begin{aligned}\sum_{i=0}^{\rho_n-1} f(X_i)^2\frac{\Delta_{i+1}^2}{\nu(X_i)^2} &= o_P\left(\sum_{i=0}^{\rho_n-1} f(X_i)^2\frac{X_i\Delta_{i+1}}{\nu(X_i)^2}\right)\\ &= o_P\left(\max_{r_n\leq x\leq n}\frac{xf(x)}{\nu(x)}\sum_{i=0}^{\rho_n-1} f(X_i)\frac{\Delta_{i+1}}{\nu(X_i)}\right).\end{aligned}$$

Under suitable conditions not only the right-hand sum but also the maximum is of order $\int_{r_n}^{n} f(x)\frac{dx}{\nu(x)}$ resulting in

$$\sum_{i=0}^{\rho_n-1} f(X_i)^2\frac{\Delta_{i+1}^2}{\nu(X_i)^2} = o_P\left(\left(\int_{r_n}^{n} f(x)\frac{dx}{\nu(x)}\right)^2\right).$$

According to this pattern, we may control the martingale's quadratic variation and its fluctuations.

The paper is organized as follows: Section 2 deals in detail with the rate functions. Section 3 contains our general laws of large numbers. Finally, our theorems are proved in Sections 4 to 7.

2. Properties of the rate functions. We extend the above sequences $\lambda(b)$ and $\mu(b)$, $b \geq 2$, of rates to positive continuous functions $\lambda, \mu : [2, \infty) \to \mathbb{R}$ via the definitions

$$\lambda(x) := \int_{[0,1]} \bigl(1 - (1-p)^x - xp(1-p)^{x-1}\bigr) \frac{\Lambda(dp)}{p^2}, \tag{4}$$

$$\mu(x) := \int_{[0,1]} \bigl(xp - 1 + (1-p)^x\bigr) \frac{\Lambda(dp)}{p^2}, \tag{5}$$

with $x \geq 2$. Note that $\mu(x) \geq \lambda(x)$ for all $x \geq 2$, since the integrands fulfill the corresponding inequality. Recall our notion

$$\kappa(x) := \frac{\mu(x)}{x}, \quad x \geq 2.$$

LEMMA 1. (i) *The functions* $\lambda(x)$ *and* $\kappa(x)$, $x \geq 2$, *are increasing in* x, *and the functions* $\lambda(x)/(x(x-1))$ *and* $\kappa(x)/(x-1)$, $x \geq 2$, *are decreasing*.

(ii) *For any* $0 < \delta < 1$ *as* $x \to \infty$

$$\lambda(x) = x^2 \kappa'(x)\bigl(1 + O(x^{-\delta})\bigr).$$

(iii) *The* Λ*-coalescent has no dust component if and only if* $\kappa(x) \to \infty$ *as* $x \to \infty$ *and then* $\kappa(x) = o(\lambda(x))$.

PROOF. (i) The function $(1-p)^x + xp(1-p)^{x-1} = (1-p)^{x-1}(1+(x-1)p)$, $x \geq 2$, respectively, its logarithm, has a negative derivative. From (4), we thus obtain monotonicity of $\lambda(x)$. Moreover, a partial integration yields

$$\frac{\lambda(x)}{x(x-1)} = \frac{\Lambda(\{0\})}{2} + \int_0^1 y(1-y)^{x-2} \int_{(y,1]} \frac{\Lambda(dp)}{p^2}\, dy \tag{6}$$

implying that $\lambda(x)/(x(x-1))$ is decreasing for $x \geq 2$.

Similarly, from (5) and a partial integration we have

$$\kappa(x) = \frac{\mu(x)}{x} = \frac{x-1}{2}\Lambda(\{0\}) + \int_0^1 \bigl(1 - (1-y)^{x-1}\bigr) \int_{(y,1]} \frac{\Lambda(dp)}{p^2}\, dy, \tag{7}$$

which is increasing in x, and by another partial integration

$$\frac{\mu(x)}{x(x-1)} = \frac{\Lambda(\{0\})}{2} + \int_0^1 (1-z)^{x-2} \int_z^1 \int_{(y,1]} \frac{\Lambda(dp)}{p^2}\, dy\, dz, \tag{8}$$

a decreasing function in x.

(ii) From (7), we have

$$\kappa'(x) = \frac{\Lambda(\{0\})}{2} + \int_0^1 (1-y)^{x-1} \log\frac{1}{1-y} \int_{(y,1]} \frac{\Lambda(dp)}{p^2}\, dy. \tag{9}$$

From concavity $(1-y)\log\frac{1}{1-y} \le y$ for $y \in [0,1]$, also $(1-y)\log\frac{1}{1-y} \sim y + O(y^2)$ for $y \to 0$. Thus, combining (6) and (9), for $0<\delta<1$ and a suitable $c \ge 1$

$$\begin{aligned}
0 &\le \frac{\lambda(x)}{x(x-1)} - \kappa'(x) \\
&\le c\int_0^{x^{-\delta}} y^2(1-y)^{x-2}\int_{(y,1]}\frac{\Lambda(dp)}{p^2}\,dy \\
&\quad + \int_{x^{-\delta}}^1 y(1-y)^{x-2}\int_{(y,1]}\frac{\Lambda(dp)}{p^2}\,dy \\
&\le cx^{-\delta}\int_0^{x^{-\delta}} y(1-y)^{x-2}\int_y^1\frac{\Lambda(dp)}{p^2}\,dy \\
&\quad + \left(1-x^{-\delta}\right)^{x-2}\int_{x^{-\delta}}^1 y\int_{(y,1]}\frac{\Lambda(dp)}{p^2}\,dy \\
&\le cx^{-\delta}\int_0^1 y(1-y)^{x-2}\int_y^1\frac{\Lambda(dp)}{p^2}\,dy \\
&\quad + e^{-x^{-\delta}(x-2)}\int_0^1 y\int_{(y,1]}\frac{\Lambda(dp)}{p^2}\,dy \\
&\le cx^{-\delta}\frac{\lambda(x)}{x(x-1)} + e^{2-x^{1-\delta}}\Lambda\big([0,1]\big).
\end{aligned}$$

Since λ is an increasing function, this implies

$$\kappa'(x) = \frac{\lambda(x)}{x^2} + O\left(\frac{\lambda(x)}{x^{2+\delta}}\right),$$

which is equivalent to our claim.

(iii) From (5) and the definition of κ, it follows by monotone convergence that

$$\kappa(x) \to \int_{[0,1]}\frac{\Lambda(dp)}{p}.$$

This formula implies our first claim. As to the second one, we have for $2 \le x_0 \le x$,

$$\kappa(x) = \kappa(x_0) + \int_{x_0}^x \kappa'(y)\,dy.$$

From part (ii) and since λ is increasing, if x_0 is sufficiently large,

$$\kappa(x) \le \kappa(x_0) + 2\int_{x_0}^x \frac{\lambda(y)}{y^2}\,dy \le \kappa(x_0) + 2\lambda(x)\int_{x_0}^x\frac{dy}{y^2} \le \kappa(x_0) + 2\frac{\lambda(x)}{x_0}.$$

Thus, from $\kappa(x) \to \infty$ it follows that $\lambda(x) \to \infty$ as $x \to \infty$, and we obtain for any $x_0 \geq 2$

$$\limsup_{x\to\infty} \frac{\kappa(x)}{\lambda(x)} \leq \frac{2}{x_0}.$$

This implies our second claim. □

Preliminary consequences of the previous results are contained in the next lemma.

LEMMA 2. *Let the Λ-coalescent be regularly varying with exponent α.*

(i) *If $0 \leq \alpha < 2$, then*

$$\lambda(x) \sim \Gamma(2-\alpha)x^{\alpha}L(x)$$

as $x \to \infty$, and for $k \geq 2$

$$\binom{b}{k}\frac{\lambda_{b,k}}{\lambda(b)} \to \frac{\alpha}{\Gamma(2-\alpha)}\frac{\Gamma(k-\alpha)}{k!}$$

as $b \to \infty$.

(ii) *The Λ-coalescent has no dust component if and only if $\int_1^\infty x^{\alpha-2}L(x)\,dx = \infty$. Then $\alpha \geq 1$ and we have*

$$\kappa(x) \sim \begin{cases} \dfrac{\Gamma(2-\alpha)}{\alpha-1}x^{\alpha-1}L(x) & \text{for } 1 < \alpha < 2, \\ L^*(x) & \text{for } \alpha = 1, \end{cases}$$

as $x \to \infty$, with a function L^* given by

$$L^*(x) := \int_1^x \frac{L(y)}{y}\,dy, \qquad x \geq 1.$$

L^* is slowly varying at infinity and satisfies $L(x) = o(L^*(x))$ as $x \to \infty$.

For $1 < \alpha < 2$, the convergence result on $\lambda_{b,k}$ has already been obtained by Delmas et al. [8]. In this case, they also have asymptotic estimates on $\lambda(x)$ and $\mu(x)$ under the assumption that the slowly varying function L is constant.

PROOF OF LEMMA 2. (i) Let $1 \leq k \leq b$ be natural numbers. Starting from the identity,

$$1-(1-p)^b - bp(1-p)^{b-1} - \cdots - \binom{b}{k}p^k(1-p)^{b-k}$$
$$= \binom{b}{k}(b-k)\int_0^p y^k(1-y)^{b-k-1}\,dy,$$

we obtain

$$\lambda(b) - \sum_{j=2}^{k} \binom{b}{j} \lambda_{b,j} = \binom{b}{k}(b-k) \int_{[0,1]} \int_0^p y^k (1-y)^{b-k-1}\, dy \frac{\Lambda(dp)}{p^2}$$
$$= \binom{b}{k}(b-k) \int_0^1 y^k (1-y)^{b-k-1} \int_{(y,1]} \frac{\Lambda(dp)}{p^2}\, dy.$$

Taking account of the definition of regular varying coalescents, it is no loss of generality to specify the function L in such a way that $\int_{(y,1]} p^{-2}\Lambda(dp) = y^{-\alpha} L(1/y)$ for $0 < y \le 1$. It follows that

$$\lambda(b) - \sum_{j=2}^{k} \binom{b}{j} \lambda_{b,j}$$
$$= \binom{b}{k}(b-k) \int_0^1 y^{k-\alpha}(1-y)^{b-k-1} L\left(\frac{1}{y}\right) dy$$
$$= \binom{b}{k}(b-k) b^{\alpha-k-1} L(b) \int_0^b z^{k-\alpha}\left(1 - \frac{z}{b}\right)^{b-k-1} \frac{L(b/z)}{L(b)}\, dz,$$

where for $k = 1$ we set the value of the left-hand sum equal to 0.

Since the function L is slowly varying, the right-hand integrand converges pointwise to the limit $z^{k-\alpha} e^{-z}$. Also, by the fundamental representation theorem of slowly varying functions (see Feller [11] Section VIII.9, corollary), we have $L(b) \sim c \exp(\int_0^b \eta(z) z^{-1}\, dz)$ with a constant $c > 0$ and a function $\eta(z) = o(z)$ as $z \to \infty$. This implies that for any $\varepsilon > 0$ we have $L(b/z) \le z^{\varepsilon} L(b)$ for $z \ge 1$ and $L(b/z) \le z^{-\varepsilon} L(b)$ for $z \le 1$ if only b is sufficiently large. Hence, for large b we may dominate the above integrand by the function $z^{k-\alpha} \max(z^{\varepsilon}, z^{-\varepsilon}) e^{-z/2}$. By the assumption $\alpha < 2$, it is integrable for $k \ge 1$ if we choose ε small enough. Thus, by dominated convergence,

$$\lambda(b) - \sum_{j=2}^{k} \binom{b}{j} \lambda_{b,j} \sim \frac{b^{\alpha} L(b)}{k!} \int_0^{\infty} z^{k-\alpha} e^{-z}\, dz = \frac{b^{\alpha} L(b) \Gamma(k-\alpha+1)}{k!}$$

as $b \to \infty$, or in other terms

$$\lambda(b) \sim \Gamma(2-\alpha) b^{\alpha} L(b) \quad \text{and} \quad \binom{b}{k} \lambda_{b,k} \sim \frac{\alpha \Gamma(k-\alpha)}{k!} b^{\alpha} L(b)$$

for $k \ge 2$. This implies our claim.

(ii) From part (i) and Lemma 1(ii), we obtain

$$\kappa'(x) \sim \Gamma(2-\alpha) x^{\alpha-2} L(x)$$

as $x \to \infty$. In view of Lemma 1(iii), we are in the dustless case if and only if $\kappa(x) \to \infty$, that is if and only if the integral $\int_1^\infty x^{\alpha-2} L(x)\,dx$ is divergent. This implies $\alpha \geq 1$ and the claimed asymptotic formulas for κ.

Finally, since L is slowly varying at infinity, we have for $c > 1$,

$$L^*(cx) - L^*(x) = \int_x^{cx} \frac{L(y)}{y}\,dy \sim L(x) \int_x^{cx} \frac{1}{y}\,dy = L(x) \log c \tag{10}$$

as $x \to \infty$. Because L^* is increasing, this implies for any $d > 1$ and large x,

$$0 \leq \frac{L^*(cx)}{L^*(x)} - 1 \leq \frac{L^*(cx) - L^*(x)}{L^*(x) - L^*(x/d)} \sim \frac{\log c}{\log d}$$

as $x \to \infty$. Since d may be chosen arbitrarily large, it follows that L^* is slowly varying. Consequently, choosing $c = e$ in (10),

$$\frac{L(x)}{L^*(x)} \sim \frac{L^*(ex) - L^*(x)}{L^*(x)} = o(1).$$

This completes our proof. □

EXAMPLES. We consider regularly varying Λ-coalescents with exponent $\alpha = 1$.

(i) Let $L(x) = (\log x)^\delta$ with exponent $\delta > -1$. Then

$$L^*(x) = \int_0^{\log x} y^\delta\,dy = \frac{1}{\delta+1} (\log x)^{\delta+1} = \frac{1}{\delta+1} L(x) \log x.$$

For $-1 < \delta \leq 0$, these coalescents neither have a dust component nor come down from infinity, and for $\delta > 0$ they come down from infinity. The case $\delta = 0$ covers the Bolthausen–Sznitman coalescent.

(ii) Let $L(x) = e^{(\log x)^\delta}$ with $0 < \delta < 1$. Then

$$L^*(x) = \int_0^{\log x} e^{y^\delta}\,dy \sim \frac{1}{\delta} e^{(\log x)^\delta} (\log x)^{1-\delta} = \frac{1}{\delta} L(x) (\log x)^{1-\delta}.$$

These coalescents come down from infinity.

3. Some general laws of large numbers.

The laws of large numbers in this section apply not only to branch lengths of Λ-coalescents. As explained in the Introduction, they concern the discrete-time Markov chains $X = (X_i)_{i \in \mathbb{N}_0}$ embedded in the lineage counting processes. For notational ease, we do not account here for the dependence of the chains X on n. Thus, $n = X_0 > X_1 > \cdots > X_{\tau_n - 1} > X_{\tau_n} = 1$ denote the states which the process N_n is successively visiting, and $\tau_n := \min\{i \geq 0 : X_i = 1\}$ is the respective number of merging events. For convenience, we set $X_i = 1$ for all natural numbers $i > \tau_n$. Also, for $i \geq 1$ let W_i be the waiting times of the process N_n in the states X_i and

$$\Delta_i := X_{i-1} - X_i.$$

Recall that

$$\mathbf{E}[\Delta_{i+1} \mid X_i = b] = \nu(b) := \frac{\mu(b)}{\lambda(b)}.$$

For sequences $r_n, s_n \geq 2$, $n \geq 1$, of positive numbers, we set

$$\rho_n := \min\{i \geq 0 : X_i \leq r_n\}, \qquad \sigma_n := \min\{i \geq 0 : X_i \leq s_n\},$$
$$\tilde{\rho}_n := \inf\{t \geq 0 : N_n(t) \leq r_n\}.$$

PROPOSITION 1. *Assume that the Λ-coalescent has no dust component. Let $f : [2, \infty) \to \mathbb{R}$ be a nonnegative function with the property that $x^\beta f(x)$ is increasing and $x^{-\beta} f(x)$ is decreasing in x for some $\beta > 0$. Let $2 \leq r_n \leq s_n \leq n$, $n \geq 1$, be two sequences of numbers fulfilling*

$$r_n \leq \gamma s_n \tag{11}$$

for all $n \geq 1$ and some $\gamma < 1$. Also, assume that

$$\tilde{\rho}_n \xrightarrow{P} 0 \tag{12}$$

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

$$\sum_{i=\sigma_n}^{\rho_n - 1} f(X_i) \stackrel{1}{\sim} \int_{r_n}^{s_n} f(x) \frac{dx}{\nu(x)}.$$

Also, as $n \to \infty$

$$\mathbf{E}\left[\sum_{i=\rho_n}^{\tau_n - 1} f(X_i)\right] = O\left(\int_2^{r_n} f(x) \frac{dx}{\nu(x)}\right).$$

Due to the assumption (12), this result addresses the coalescent's evolution in only a short initial time interval. This takes double effect: first, as seen from the next lemma, the chain is kept away from the occurrence of huge jumps Δ_i being of the same order as the chain's values. Second, the chain X is prevented from taking values that are too small where larger fluctuations may become obstructive. Note that, for any Λ-coalescent, the passage times $\inf\{t \geq 0 : N_n(t) \leq r\}$ below some number $r > 1$ are bounded away from 0 uniformly in $n \in \mathbb{N}$. Therefore, the assumption (12) enforces that

$$r_n \to \infty \tag{13}$$

as $n \to \infty$. Actually, both requirements are equivalent if the coalescent comes down from infinity, otherwise the assumption (12) is the more incisive one.

EXAMPLE (Total number of mergers). If we choose $f \equiv 1$ and $s_n = n$, then $\sum_{i=0}^{\rho_n-1} f(X_i)$ is equal to the number of mergers up to time $\tilde{\rho}_n$. For Λ-coalescents coming down from infinity, we may consider any divergent sequence $r_n \le \gamma n$. In particular, since $\int_2^\infty \frac{dx}{\nu(x)} = \infty$, we may choose the r_n in such a way that $r_n = o(\int_{r_n}^n \frac{dx}{\nu(x)})$. This implies that the number of mergers after the moment $\tilde{\rho}_n$ are of negligible order and that, for the total number τ_n of mergers, we have

$$\tau_n \stackrel{1}{\sim} \int_2^n \frac{dx}{\nu(x)}.$$

Aside from Λ-coalescents coming down from infinity, the scope of this formula is unclear. For the Bolthausen–Sznitman coalescent it is valid (see [12]).

The proof of Proposition 1 is prepared by the next lemma.

LEMMA 3. (i) *If the Λ-coalescent has no dust component and if*

$$\int_{r_n}^n \mu(x)^{-1}\,dx \to 0,$$

then as $n \to \infty$,

$$\mathbf{E}[\tilde{\rho}_n] \to 0.$$

(ii) *If $\tilde{\rho}_n \xrightarrow{P} 0$, then for any $\eta > 0$ as $n \to \infty$*

$$\mathbf{P}(\Delta_{i+1} > \eta X_i \textit{ for some } i < \rho_n) \to 0.$$

PROOF. (i) Given X, the waiting times W_i are exponential with expectation $1/\lambda(X_i)$. Therefore,

$$\mathbf{E}[\tilde{\rho}_n] = \mathbf{E}\left[\sum_{i=0}^{\rho_n-1} W_i\right] = \sum_{i=0}^{n-1} \mathbf{E}\left[\frac{1}{\lambda(X_i)}; X_i > r_n\right].$$

Also, $\mathbf{E}[\Delta_{i+1} \mid X_i] = \nu(X_i)$ a.s. Thus, according to the Markov property,

$$\mathbf{E}[\tilde{\rho}_n] = \sum_{i=0}^{n-1} \mathbf{E}\left[\frac{\Delta_{i+1}}{\lambda(X_i)\nu(X_i)}; X_i > r_n\right] = \mathbf{E}\left[\sum_{i=0}^{\rho_n-1} \frac{\Delta_{i+1}}{\mu(X_i)}\right].$$

From Lemma 1(i), we know that $\mu(x) = x\kappa(x)$ is increasing. Also, $\Delta_{i+1} \le X_i$, and hence

$$\mathbf{E}[\tilde{\rho}_n] \le \mathbf{E}\left[\sum_{i=0}^{\rho_n-2} \int_{X_{i+1}}^{X_i} \frac{dx}{\mu(x)} + \frac{X_{\rho_n-1}}{\mu(X_{\rho_n-1})}\right] = \mathbf{E}\left[\int_{X_{\rho_n-1}}^n \frac{dx}{\mu(x)} + \frac{1}{\kappa(X_{\rho_n-1})}\right],$$

and since κ is an increasing function, we end up with the estimate

$$\mathbf{E}[\tilde{\rho}_n] \le \int_{r_n}^n \frac{dx}{\mu(x)} + \frac{1}{\kappa(r_n)}.$$

Our assumptions imply $r_n \to \infty$. Therefore, since there is no dust component, we have $\kappa(r_n) \to \infty$ by Lemma 1(iii). This entails our claim.

(ii) For $b \geq 2/\eta$, we have

$$\begin{aligned}
&\mathbf{P}(\Delta_{i+1} > \eta X_i \mid X_i = b) \\
&= \sum_{k>\eta b} \frac{1}{\lambda(b)} \int_{[0,1]} \binom{b}{k} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2} \\
&\leq \frac{1}{\lambda(b)} \int_{[0,1]} \sum_{k>\eta b} \frac{2}{\eta^2} \binom{b-2}{k-2} p^{k-2} (1-p)^{(b-2)-(k-2)} \Lambda(dp) \\
&\leq \frac{c}{\lambda(b)},
\end{aligned}$$

with $c := 2\Lambda([0,1])/\eta^2$.

Let $\delta_0 := 0$ and $\delta_i := W_0 + \cdots + W_{i-1}$, $i \geq 1$, which is the moment of the ith jump. Then

$$\begin{aligned}
&\mathbf{P}(\Delta_{i+1} > \eta X_i \text{ for some } i < \rho_n, \tilde{\rho}_n \leq 1) \\
&\quad \leq \mathbf{P}(\Delta_{i+1} > \eta X_i, X_i > r_n, \delta_i \leq 1 \text{ for some } i < n) \\
&\quad \leq \sum_{i=0}^{n-1} \mathbf{E}\big[\mathbf{P}(\Delta_{i+1} > \eta X_i \mid X_i); X_i > r_n, \delta_i \leq 1\big] \\
&\quad \leq \sum_{i=0}^{n-1} \mathbf{E}\left[\frac{c}{\lambda(X_i)}; X_i > r_n, \delta_i \leq 1\right].
\end{aligned}$$

Also, because $\lambda(x)$ is increasing, for $X_i \geq 2$

$$\begin{aligned}
\mathbf{E}[W_i I_{\{W_i \leq 1\}} \mid X_i] &= \int_0^1 t\lambda(X_i) e^{-\lambda(X_i)t}\, dt \\
&= \frac{1}{\lambda(X_i)} \int_0^{\lambda(X_i)} u e^{-u}\, du \geq \frac{d}{\lambda(X_i)}
\end{aligned}$$

with $d := \int_0^{\lambda(2)} u e^{-u}\, du > 0$. This allows for the estimate

$$\begin{aligned}
&\mathbf{P}(\Delta_{i+1} > \eta X_i \text{ for some } i < \rho_n, \tilde{\rho}_n \leq 1) \\
&\quad \leq \frac{c}{d} \sum_{i=0}^{n-1} \mathbf{E}[W_i; X_i > r_n, W_i \leq 1, W_0 + \cdots + W_{i-1} \leq 1] \\
&\quad \leq \frac{c}{d} \mathbf{E}\left[\sum_{i=0}^{\rho_n - 1} W_i I_{\{W_0 + \cdots + W_i \leq 2\}}\right]
\end{aligned}$$

$$\leq \frac{c}{d}\mathbf{E}\left[2 \wedge \sum_{i=0}^{\rho_n-1} W_i\right]$$
$$= \frac{c}{d}\mathbf{E}[2 \wedge \tilde{\rho}_n],$$

and consequently

$$\mathbf{P}(\Delta_{i+1} > \eta X_i \text{ for some } i < \rho_n) \leq \frac{c}{d}\mathbf{E}[\tilde{\rho}_n \wedge 2] + \mathbf{P}(\tilde{\rho}_n > 1).$$

Thus, by assumption and dominated convergence, our claim follows. □

PROOF OF PROPOSITION 1. (i) We start with some preliminary estimates. Since $x^{-\beta} f(x)$ is a decreasing function, we have that for $0 < \delta < 1$, $0 \leq y \leq \delta z$ and $y \leq z - 2$,

$$yf(z) \leq z^\beta \int_{z-y}^{z} x^{-\beta} f(x)\,dx \leq z^\beta (z-y)^{-\beta} \int_{z-y}^{z} f(x)\,dx$$
$$\leq (1-\delta)^{-\beta} \int_{z-y}^{z} f(x)\,dx.$$

Similarly, using the fact that $x^\beta f(x)$ is increasing, we obtain a lower bound. Altogether, for $0 < \delta < 1$, $0 \leq y \leq \delta z$ and $y \leq z - 2$,

$$(1-\delta)^\beta \int_{z-y}^{z} f(x)\,dx \leq yf(z) \leq (1-\delta)^{-\beta} \int_{z-y}^{z} f(x)\,dx. \tag{14}$$

Moreover, for any $\varepsilon > 0$ there is an $\eta > 0$ such that

$$(1-\varepsilon) \int_{(1-\eta)r_n}^{s_n} f(x)\,dx \leq \int_{r_n}^{s_n} f(x)\,dx \leq (1+\varepsilon) \int_{r_n}^{(1-\eta)s_n} f(x)\,dx \tag{15}$$

for all n. We prove the left-hand inequality. Let a be the affine function mapping the interval $[(1-\eta)r_n, r_n]$ onto $[(1-\eta)r_n, r_n/\gamma]$. Substituting $y = a(x)$, this implies

$$dy = \frac{\eta + \gamma^{-1} - 1}{\eta}\,dx.$$

Moreover, since $\gamma < 1$, $a(x) \geq x$ for $x \geq (1-\eta)r_n$. Therefore, by monotonicity of x^β and $x^\beta f(x)$, we have with γ as in (11)

$$\int_{(1-\eta)r_n}^{r_n} f(x)\,dx \leq \big((1-\eta)r_n\big)^{-\beta} \int_{(1-\eta)r_n}^{r_n} f(x)x^\beta\,dx$$
$$\leq \big((1-\eta)r_n\big)^{-\beta} \int_{(1-\eta)r_n}^{r_n} f\big(a(x)\big)a(x)^\beta\,dx$$

$$= \big((1-\eta)r_n\big)^{-\beta} \frac{\eta}{\eta+\gamma^{-1}-1} \int_{(1-\eta)r_n}^{r_n/\gamma} f(y) y^{\beta}\, dy$$

$$\le \big(\gamma(1-\eta)\big)^{-\beta} \frac{\eta}{\eta+\gamma^{-1}-1} \int_{(1-\eta)r_n}^{r_n/\gamma} f(y)\, dy.$$

From condition (11), it follows that

$$\int_{(1-\eta)r_n}^{r_n} f(x)\, dx \le \big(\gamma(1-\eta)\big)^{-\beta} \frac{\eta}{\eta+\gamma^{-1}-1} \int_{(1-\eta)r_n}^{s_n} f(x)\, dx$$

$$\le \varepsilon \int_{(1-\eta)r_n}^{s_n} f(x)\, dx$$

for sufficiently small $\eta > 0$. This implies the left-hand inequality of (15). The other one follows similarly, now using the monotonicity of $x^{-\beta} f(x)$.

We note that $(x-1)/x^2$ is a decreasing function for $x \ge 2$. Therefore, in view of Lemma 1(i) the functions $\lambda(x)$ and $\mu(x)$ are increasing and $\lambda(x)/x^3$ and $\mu(x)/x^3$ are decreasing for $x \ge 2$. Thus, the function $f(x)/\nu(x) = f(x)\lambda(x)/\mu(x)$ fulfills the same assumptions as $f(x)$, with β replaced by $\beta + 3$. Accordingly, we shall use the preceding estimates with $f(x)$ replaced by $f(x)/\nu(x)$ (and β replaced by $\beta + 3$).

(ii) Next, we develop Riemann approximations of certain random sums. For $0 < \eta < 1$ and $b \ge 2$, let

$$\mu_\eta(b) := \sum_{2 \le k \le \eta b} (k-1) \binom{b}{k} \int_{[0,1]} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2}.$$

In view of the well-known formula for the second factorial moment of binomials,

$$\mu(b) - \mu_\eta(b) = \sum_{\eta b < k \le b} (k-1) \binom{b}{k} \int_{[0,1]} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2}$$

$$\le \frac{1}{\eta b} \int_{[0,1]} \sum_{k=0}^{b} k(k-1) \binom{b}{k} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2}$$

$$= \frac{1}{\eta b} \int_{[0,1]} b(b-1) p^2 \frac{\Lambda(dp)}{p^2} = \frac{(b-1)\Lambda([0,1])}{\eta}.$$

From Lemma 1(iii), we have $\mu(b)/b \to \infty$ in the dustless case. Hence, for any $\eta > 0$ we obtain

$$\frac{\mu_\eta(b)}{\mu(b)} \to 1 \tag{16}$$

as $b \to \infty$.

Also, let

$$\nu_\eta(b) := \frac{\mu_\eta(b)}{\lambda(b)} = \mathbf{E}[\Delta_{i+1} I_{\{\Delta_{i+1} \le \eta b\}} \mid X_i = b].$$

Note that $\nu_\eta(b) > 0$ for $b \ge 2/\eta$. Therefore, for $0 < \eta < 1$ and natural numbers n satisfying $r_n \ge 2/\eta$, we may define the random variables

$$R_n = R_{n,\eta} := \sum_{i=\sigma_n}^{\rho_n - 1} f(X_i) \frac{\Delta_{i+1} I_{\{\Delta_{i+1} \le \eta X_i\}}}{\nu_\eta(X_i)}.$$

Given η, these random variables are, in view of (13), well defined up to finitely many n. We shall use them below as an intermediate approximation to $\sum_{i=\sigma_n}^{\rho_n - 1} f(X_i)$. We estimate R_n from above and below. From (14) with $z = X_i$, $y = \Delta_{i+1}$ and $\delta = \eta$ and with $f(x)/\nu(x)$ replacing $f(x)$, we have on the event that $\Delta_{i+1} \le \eta X_i$,

$$\frac{f(X_i)}{\nu(X_i)} \Delta_{i+1} \le (1-\eta)^{-\beta-3} \int_{X_{i+1}}^{X_i} f(x) \frac{dx}{\nu(x)},$$

consequently

$$R_n \le (1-\eta)^{-\beta-3} \sum_{i=\sigma_n}^{\rho_n - 1} \frac{\nu(X_i)}{\nu_\eta(X_i)} I_{\{\Delta_{i+1} \le \eta X_i\}} \int_{X_{i+1}}^{X_i} f(x) \frac{dx}{\nu(x)}$$

and, by definition of ρ_n,

$$R_n \le (1-\eta)^{-\beta-3} \sup_{b \ge r_n} \frac{\mu(b)}{\mu_\eta(b)} \int_{r_n(1-\eta)}^{s_n} f(x) \frac{dx}{\nu(x)}. \tag{17}$$

Therefore, in view of (15) and (16) and since $r_n \to \infty$, there is for given $\varepsilon > 0$ an $\eta > 0$ fulfilling

$$R_n \le (1+\varepsilon) \int_{r_n}^{s_n} f(x) \frac{dx}{\nu(x)} \tag{18}$$

for all n sufficiently large. Similarly, for given $\varepsilon > 0$ there is an $\eta > 0$ satisfying for large n the inequality

$$R_n \ge (1-\varepsilon) \int_{r_n}^{s_n} f(x) \frac{dx}{\nu(x)} \quad \text{on the event } \{\Delta_{i+1} \le \eta X_i \text{ for all } i < \rho_n\}. \tag{19}$$

(iii) Now observe that the random variables $M_0 := 0$ and

$$M_k := \sum_{i=0}^{k \wedge \tau_n - 1} \left(f(X_i) \frac{\Delta_{i+1} I_{\{\Delta_{i+1} \le \eta X_i\}}}{\nu_\eta(X_i)} - f(X_i) \right), \quad k \ge 1,$$

build a martingale $M = (M_k)_{k\geq 0}$. The optional sampling theorem yields

$$\mathbf{E}\Bigg[\Bigg(R_n - \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\Bigg)^2\Bigg] = \mathbf{E}\big[(M_{\rho_n} - M_{\sigma_n})^2\big]$$

$$\leq \mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)^2\Delta_{i+1}^2}{\nu_\eta(X_i)^2} I_{\{\Delta_{i+1}\leq \eta X_i\}}\Bigg]$$

$$\leq \eta\mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)^2 X_i\Delta_{i+1}}{\nu_\eta(X_i)^2} I_{\{\Delta_{i+1}\leq \eta X_i\}}\Bigg].$$

Letting x_n be the point where the function $xf(x)/\nu_\eta(x)$ takes its maximal value within the interval $[r_n, s_n]$, it follows

$$\mathbf{E}\Bigg[\Bigg(R_n - \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\Bigg)^2\Bigg] \leq \eta\frac{x_n f(x_n)}{\nu_\eta(x_n)}\mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)\Delta_{i+1}}{\nu_\eta(X_i)} I_{\{\Delta_{i+1}\leq \eta X_i\}}\Bigg]$$

$$= \eta\frac{x_n f(x_n)}{\nu_\eta(x_n)}\mathbf{E}[R_n].$$

By the assumption (11), it follows that there are numbers ξ_n, $n \geq 1$, such that

$$r_n \leq \gamma\xi_n \leq x_n \leq \xi_n \leq s_n.$$

Using the monotonicity of $x^{\beta+3}f(x)/\nu(x)$ and (14) with $z = \xi_n$, $y = (1-\gamma)\xi_n$, $\delta = 1-\gamma$, it follows that

$$(1-\gamma)\frac{x_n f(x_n)}{\nu(x_n)} \leq \left(\frac{\xi_n}{x_n}\right)^{\beta+2}(1-\gamma)\frac{\xi_n f(\xi_n)}{\nu(\xi_n)} \leq \gamma^{-2\beta-5}\int_{\gamma\xi_n}^{\xi_n} f(x)\frac{dx}{\nu(x)}$$

$$\leq \gamma^{-2\beta-5}\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}.$$

Hence,

$$\mathbf{E}\Bigg[\Bigg(R_n - \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\Bigg)^2\Bigg] \leq \eta\frac{\gamma^{-2\beta-5}}{1-\gamma}\sup_{x\geq 2/\eta}\frac{\nu(x)}{\nu_\eta(x)}\mathbf{E}[R_n]\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}.$$

Finally, the formulas (16) and (18) yield that for any $\varepsilon > 0$ there is an $\eta > 0$ fulfilling

$$\mathbf{E}\Bigg[\Bigg(R_n - \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\Bigg)^2\Bigg] \leq \varepsilon\bigg(\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\bigg)^2. \tag{20}$$

(iv) Putting pieces together, we obtain for given $\varepsilon > 0$ and $\eta > 0$

$$\begin{aligned}
&\mathbf{P}\biggl(\biggl|\sum_{i=\sigma_n}^{\rho_n-1} f(X_i) - \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr| \ge \varepsilon \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr)\\
&\quad\le \mathbf{P}\biggl(\biggl|R_n - \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\biggr| \ge \frac{\varepsilon}{2}\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr)\\
&\qquad + \mathbf{P}\biggl(\biggl|R_n - \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr| \ge \frac{\varepsilon}{2}\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}, \Delta_{i+1} \le \eta X_i \text{ for all } i < \rho_n\biggr)\\
&\qquad + \mathbf{P}(\Delta_{i+1} > \eta X_i \text{ for some } i < \rho_n).
\end{aligned}$$

From (20) for suitably chosen $\eta > 0$ the first right-hand term becomes smaller than ε and from (18) and (19) the second one vanishes for large n. Consulting also Lemma 3(ii), we arrive at

$$\mathbf{P}\biggl(\biggl|\sum_{i=\sigma_n}^{\rho_n-1} f(X_i) - \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr| \ge \varepsilon \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}\biggr) \le \varepsilon$$

for n large enough. This means that

$$\sum_{i=\sigma_n}^{\rho_n-1} f(X_i) \overset{P}{\sim} \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}.$$

To show L_1-convergence, it is by a convergence criterion due to F. Riesz sufficient to have

$$\mathbf{E}\biggl[\sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\biggr] \sim \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}$$

as $n \to \infty$. From convergence in probability and Fatou's lemma, we have

$$\liminf_{n\to\infty} \mathbf{E}\biggl[\sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\biggr] \Big/ \int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)} \ge 1.$$

On the other hand, (18) yields for given $\varepsilon > 0$, suitable $\eta > 0$ and large n,

$$\mathbf{E}\biggl[\sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\biggr] = \mathbf{E}[R_n] \le (1+\varepsilon)\int_{r_n}^{s_n} f(x)\frac{dx}{\nu(x)}.$$

This gives our first claim.

(v) For the second claim, we again use the random variables R_n, now for some $\eta > 0$ (say $\eta = 1/2$) with $r_n = r := 2/\eta$. Recall that for this choice the R_n are well defined for all n. From inequality (17), we obtain the estimate

$$R_n \le (1-\eta)^{-\beta-3} \sup_{b \ge 2/\eta} \frac{\mu(b)}{\mu_\eta(b)} \int_2^{s_n} f(x)\frac{dx}{\nu(x)}.$$

It follows

$$\mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\tau_n-1} f(X_i)\Bigg] \le \sum_{b<2/\eta} f(b) + \mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\rho_n-1} f(X_i)\Bigg] = \sum_{b<2/\eta} f(b) + \mathbf{E}[R_n].$$

Putting both estimates together and then replacing σ_n by ρ_n (which is just a change in notation), we arrive at our second claim. □

The next proposition presents a version of Proposition 1 in continuous time. As before, let $\tilde{\rho}_n$ and $\tilde{\sigma}_n$ denote the times when the process N_n falls below r_n and s_n, respectively, while $\tilde{\tau}_n$ is the absorption time of N_n.

PROPOSITION 2. *Under the assumptions of Proposition 1, as $n \to \infty$,*

$$\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\big(N_n(t)\big)\,dt \overset{1}{\sim} \int_{r_n}^{s_n} f(x)\frac{dx}{\mu(x)}.$$

Also, as $n \to \infty$,

$$\mathbf{E}\bigg[\int_{\tilde{\rho}_n}^{\tilde{\tau}_n} f\big(N_n(t)\big)\,dt\bigg] = O\bigg(\int_2^{r_n} f(x)\frac{dx}{\mu(x)}\bigg).$$

EXAMPLE. For $f(x) \equiv 1$, we obtain under the assumptions of Proposition 2

$$\tilde{\rho}_n - \tilde{\sigma}_n \overset{1}{\sim} \int_{r_n}^{s_n} \frac{dx}{\mu(x)},$$

in particular,

$$\tilde{\rho}_n \overset{1}{\sim} \int_{r_n}^{n} \frac{dx}{\mu(x)}$$

as $n \to \infty$.

PROOF OF PROPOSITION 2. We have

$$\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\big(N_n(t)\big)\,dt = \sum_{i=\sigma_n}^{\rho_n-1} f(X_i)W_i$$

and

$$\mathbf{E}\Bigg[\sum_{i=\sigma_n}^{\rho_n-1} f(X_i)W_i \,\Bigg|\, X\Bigg] = \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)}.$$

Thus, for any $\eta > 0$, by the Markov property,

$$\mathbf{E}\biggl[\biggl(\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\bigl(N_n(t)\bigr)\,dt - \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)}\biggr)^2\biggr] = \mathbf{E}\biggl[\sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)^2}{\lambda(X_i)^2}\biggr] = \mathbf{E}[R_n],$$

where we now set

$$R_n := \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)^2}{\lambda(X_i)^2} \frac{\Delta_{i+1} I_{\{\Delta_{i+1} \le \eta X_i\}}}{\nu_\eta(X_i)}.$$

Using (18) with $\varepsilon = 1$, it follows

$$\begin{aligned}\mathbf{E}\biggl[\biggl(\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\bigl(N_n(t)\bigr)\,dt - \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)}\biggr)^2\biggr] &\le 2\int_{r_n}^{s_n} \frac{f(x)^2}{\lambda(x)^2 \nu(x)}\,dx \\ &= 2\int_{r_n}^{s_n} \frac{f(x)^2}{\lambda(x)\mu(x)}\,dx.\end{aligned}$$

Furthermore, since we are in the dustless case, due to Lemma 1(iii) we have

$$\mathbf{E}\biggl[\biggl(\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\bigl(N_n(t)\bigr)\,dt - \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)}\biggr)^2\biggr] \le o\biggl(\int_{r_n}^{s_n} \frac{x f(x)^2}{\mu(x)^2}\,dx\biggr).$$

Letting x_n be the point where $xf(x)/\mu(x)$ takes its maximum in the interval $[r_n, s_n]$, we obtain in the same manner as in the preceding proof

$$\begin{aligned}\mathbf{E}\biggl[\biggl(\int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} f\bigl(N_n(t)\bigr)\,dt - \sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)}\biggr)^2\biggr] &= o\biggl(\frac{x_n f(x_n)}{\mu(x_n)} \int_{r_n}^{s_n} \frac{f(x)}{\mu(x)}\,dx\biggr) \\ &= o\biggl(\biggl(\int_{r_n}^{s_n} \frac{f(x)}{\mu(x)}\,dx\biggr)^2\biggr).\end{aligned}$$

On the other hand, Proposition 1 implies

$$\sum_{i=\sigma_n}^{\rho_n-1} \frac{f(X_i)}{\lambda(X_i)} \overset{1}{\sim} \int_{r_n}^{s_n} \frac{f(x)}{\lambda(x)} \frac{dx}{\nu(x)} = \int_{r_n}^{s_n} \frac{f(x)}{\mu(x)}\,dx.$$

These last two formulas imply our first statement. Moreover,

$$\mathbf{E}\biggl[\int_{\tilde{\rho}_n}^{\tilde{\tau}_n} f\bigl(N_n(t)\bigr)\,dt\biggr] = \mathbf{E}\biggl[\sum_{i=\rho_n}^{\tau_n-1} \frac{f(X_i)}{\lambda(X_i)}\biggr]$$

and, therefore, our second claim follows from the second statement of Proposition 1. □

We now turn to the special case that $f(x) = 1/x$, $x \geq 2$, where Proposition 1 can be considerably extended. Here, we are content with the case $s_n = n$. Observe that the two following statements do not imply each other.

PROPOSITION 3. *Assume that the* Λ*-coalescent has no dust component. Let* $2 \leq r_n \leq n$, $n \geq 1$, *be a sequence of numbers fulfilling*

$$r_n \leq \gamma n$$

for all n *sufficiently large and some* $\gamma < 1$. *Also, assume that*

$$\tilde{\rho}_n \xrightarrow{P} 0$$

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

$$\sum_{i=0}^{\rho_n - 1} \frac{1}{X_i} \overset{1}{\sim} \log \frac{\kappa(n)}{\kappa(r_n)}$$

and

$$\max_{1 \leq j \leq \rho_n} \left| \sum_{i=0}^{j-1} \frac{1}{X_i} - \log \frac{\kappa(n)}{\kappa(X_j)} \right| = o_P(1).$$

PROOF. (i) The first statement is a special case of Proposition 1: due to Lemma 1(ii) we have $1/(x\nu(x)) = \lambda(x)/(x\mu(x)) \sim \kappa'(x)/\kappa(x)$ as $x \to \infty$. Therefore,

$$\int_{r_n}^{n} \frac{dx}{x\nu(x)} \sim \int_{r_n}^{n} \frac{\kappa'(x)\,dx}{\kappa(x)} = \log \frac{\kappa(n)}{\kappa(r_n)}.$$

(ii) For the proof of the second statement, we proceed along similar lines as in the proof of Proposition 1. Using the second factorial moment of a binomial distribution, we have as a first estimate

$$\begin{aligned}
\mathbf{E}\left[\frac{\Delta_{i+1}^2}{X_i^2} \,\Big|\, X_i = b\right] &= \frac{1}{b^2\lambda(b)} \sum_{k=2}^{b} (k-1)^2 \binom{b}{k} \int_{[0,1]} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2} \\
&\leq \frac{1}{b^2\lambda(b)} \int_{[0,1]} \sum_{k=0}^{b} k(k-1) \binom{b}{k} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2} \\
&= \frac{1}{b^2\lambda(b)} \int_{[0,1]} p^2 b(b-1) \frac{\Lambda(dp)}{p^2} \\
&\leq \frac{\Lambda([0,1])}{\lambda(b)}.
\end{aligned}$$

This bound yields

$$\mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{\Delta_{i+1}^2}{X_i^2\nu(X_i)}\right]\le\sum_{i=0}^{n-1}\mathbf{E}\left[\frac{1}{\mu(X_i)};X_i>r_n\right]\Lambda([0,1])$$
$$=\sum_{i=0}^{n-1}\mathbf{E}\left[\frac{\Delta_{i+1}}{\mu(X_i)\nu_{1/2}(X_i)}I_{\{\Delta_{i+1}\le X_i/2\}};X_i>r_n\right]\Lambda([0,1])$$
$$\le c\mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{\Delta_{i+1}}{\mu(X_i)\nu(X_i)}I_{\{\Delta_{i+1}\le X_i/2\}}\right]$$

for large n, with $c:=\Lambda([0,1])\sup_{b\ge 4}\nu(b)/\nu_{1/2}(b)<\infty$. In view of Lemma 1(i), the function

$$x\mapsto(x-1)\mu(x)\nu(x)/x=\kappa(x)^2/\big(\lambda(x)/x(x-1)\big)$$

is increasing, which entails $\Delta_{i+1}/(\mu(X_i)\nu(X_i))\le\int_{X_{i+1}}^{X_i}x((x-1)\mu(x)\nu(x))^{-1}\,dx$. By means of Lemma 1(ii), $\mu(x)\nu(x)\sim\kappa(x)^2/\kappa'(x)$. Therefore,

$$\mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{\Delta_{i+1}^2}{X_i^2\nu(X_i)}\right]\le c\int_{r_n/2}^{n}\frac{x\,dx}{(x-1)\mu(x)\nu(x)}\sim c\int_{r_n/2}^{n}\frac{\kappa'(x)}{\kappa^2(x)}\,dx\le\frac{c}{\kappa(r_n/2)},$$

and consequently, since $\kappa(x)\to\infty$ in the dustless case,

$$\mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{\Delta_{i+1}^2}{X_i^2\nu(X_i)}\right]=o(1).$$

(iii) We now consider the martingale $M=(M_k)_{k\ge 0}$ given by $M_0=0$ a.s. and

$$M_k:=\sum_{i=0}^{k\wedge\tau_n-1}\left(\frac{\Delta_{i+1}}{(X_i-1)\nu(X_i)}-\frac{1}{X_i-1}\right),\quad k\ge 1.$$

By means of the optional sampling theorem, since ρ_n is a stopping time, we have, because $\nu(b)\ge 1$ for all $b\ge 2$,

$$\mathbf{E}[M_{\rho_n}^2]\le\mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{\Delta_{i+1}^2}{X_i^2\nu(X_i)^2}\right]=o(1).$$

Thus, by means of Doob's maximal inequality,

$$\max_{1\le j\le\rho_n}\left|\sum_{i=0}^{j-1}\frac{\Delta_{i+1}}{(X_i-1)\nu(X_i)}-\sum_{i=0}^{j-1}\frac{1}{X_i-1}\right|=o_P(1).$$

Also, for $j\le\rho_n$,

$$0\le\sum_{i=0}^{j-1}\frac{1}{X_i-1}-\sum_{i=0}^{j-1}\frac{1}{X_i}\le\sum_{m=X_{\rho_n-1}}^{\infty}\left(\frac{1}{m-1}-\frac{1}{m}\right)=\frac{1}{X_{\rho_n-1}-1}\le\frac{1}{r_n-1},$$

and consequently,

$$\max_{1\le j\le \rho_n}\left|\sum_{i=0}^{j-1}\frac{\Delta_{i+1}}{(X_i-1)\nu(X_i)}-\sum_{i=0}^{j-1}\frac{1}{X_i}\right|=o_P(1). \tag{21}$$

(iv) Note that, in view of Lemma 1(i), the function $(x-1)\nu(x)=\kappa(x)/(\lambda(x)/x(x-1))$ is increasing and $\nu(x)/(x-1)^2=(x/(x-1))\cdot(\kappa(x)/(x-1))/\lambda(x)$ is decreasing. For $X_i\ge 2$, this yields

$$\begin{aligned}
0&\le \frac{1}{(X_{i+1}-1)\nu(X_{i+1})}-\frac{1}{(X_i-1)\nu(X_i)}\\
&=\frac{(X_{i+1}-1)^2}{\nu(X_{i+1})}\frac{1}{(X_{i+1}-1)^3}-\frac{(X_i-1)^2}{\nu(X_i)}\frac{1}{(X_i-1)^3}\\
&\le \frac{(X_i-1)^2}{\nu(X_i)}\left(\frac{1}{(X_{i+1}-1)^3}-\frac{1}{(X_i-1)^3}\right)\\
&=\frac{((X_i-1)^2+(X_i-1)(X_{i+1}-1)+(X_{i+1}-1)^2)\Delta_{i+1}}{\nu(X_i)(X_i-1)(X_{i+1}-1)^3}\\
&\le \frac{3(X_i-1)}{\nu(X_i)(X_{i+1}-1)^3}\Delta_{i+1}\\
&\le \frac{24X_i}{\nu(X_i)X_{i+1}^3}\Delta_{i+1}.
\end{aligned}$$

It follows

$$\begin{aligned}
&\mathbf{E}\Biggl[\max_{1\le j\le\rho_n}\left|\sum_{i=0}^{j-1}\frac{\Delta_{i+1}}{(X_i-1)\nu(X_i)}-\sum_{i=0}^{j-1}\frac{\Delta_{i+1}}{(X_{i+1}-1)\nu(X_{i+1})}\right|;\\
&\qquad \Delta_{i+1}\le X_i/2 \text{ for all } i<\rho_n\Biggr]\\
&\quad\le \mathbf{E}\Biggl[\sum_{i=0}^{\rho_n-1}\left|\frac{1}{(X_i-1)\nu(X_i)}-\frac{1}{(X_{i+1}-1)\nu(X_{i+1})}\right|\Delta_{i+1};\\
&\qquad \Delta_{i+1}\le X_i/2 \text{ for all } i<\rho_n\Biggr]\\
&\quad\le \mathbf{E}\left[\sum_{i=0}^{\rho_n-1}\frac{8\cdot 24}{X_i^2\nu(X_i)}\Delta_{i+1}^2\right]=o(1).
\end{aligned}$$

Also, since $(x-1)\nu(x)$ is increasing,

$$\frac{1}{(X_i-1)\nu(X_i)}\Delta_{i+1}\le\int_{X_{i+1}}^{X_i}\frac{dx}{(x-1)\nu(x)}\le\frac{1}{(X_{i+1}-1)\nu(X_{i+1})}\Delta_{i+1}$$

and we obtain

$$\max_{1\le j\le \rho_n}\left|\sum_{i=0}^{j-1}\frac{1}{(X_i-1)\nu(X_i)}\Delta_{i+1}-\int_{X_j}^{n}\frac{dx}{(x-1)\nu(x)}\right| = o_P(1) \tag{22}$$

on the event that $\Delta_{i+1}\le X_i/2$ for all $i<\rho_n$. Lemma 3(ii) says that the complementary event has an asymptotically vanishing probability.

(v) Finally, by Lemma 1(ii) with $r_n\le y\le n$ for $\delta<1$,

$$\begin{aligned}\int_y^n\frac{dx}{(x-1)\nu(x)} &= \int_y^n\frac{\lambda(x)/x^2}{\mu(x)/x}\,dx+\int_y^n\frac{dx}{x(x-1)\nu(x)}\\ &= \int_y^n\frac{\kappa'(x)}{\kappa(x)}\,dx+O\biggl(\int_{r_n}^n\frac{\kappa'(x)\,dx}{\kappa(x)x^\delta}\biggr)+O(r_n^{-1}),\end{aligned}$$

and recalling $r_n\to\infty$,

$$\int_{r_n}^n\frac{\kappa'(x)\,dx}{\kappa(x)x^\delta}\sim\int_{r_n}^n\frac{dx}{x\nu(x)x^\delta}\le\int_{r_n}^n\frac{dx}{x^{1+\delta}}=o(1).$$

Altogether, we obtain

$$\max_{r_n\le y\le n}\left|\int_y^n\frac{dx}{(x-1)\nu(x)}-\log\frac{\kappa(n)}{\kappa(y)}\right|=o(1).$$

Combining this formula with (21) and (22) and recalling the definition of ρ_n, we arrive at

$$\max_{1\le j\le\rho_n}\left|\sum_{i=0}^{j-1}\frac{1}{X_i}-\log\frac{\kappa(n)}{\kappa(X_j)}\right|=o_P(1).$$

This is our claim. □

4. Proof of Theorem 1.

Under the assumptions of Proposition 2, we have

$$\ell_n^* := \int_0^{\tilde\rho_n} N_n(t)\,dt \overset{1}{\sim} \int_{r_n}^n\frac{x}{\mu(x)}\,dx. \tag{23}$$

In particular, this formula holds for $r_n:=cn$ with $0<c<1$ as anticipated in the Introduction's remark. Here, the assumptions of Proposition 2 are satisfied because of Lemma 3(i) and

$$\int_{cn}^n\frac{dx}{\mu(x)}\le(1-c)\frac{n}{\mu(cn)}=o(1),$$

which in turn is valid in view of Lemma 1(iii).

In order to fill the gap up to ℓ_n, we construct a distinguished sequence of real numbers. We construct the numbers $2\le r_n\le n$, $n\ge 1$, satisfying

$$\int_{r_n}^n\frac{dx}{\mu(x)}\to 0\quad\text{and}\quad\int_2^{r_n}\frac{x}{\mu(x)}\,dx=o\biggl(\int_{r_n}^n\frac{x}{\mu(x)}\,dx\biggr) \tag{24}$$

as $n \to \infty$. From Lemma 3(i), we get $\tilde{\rho}_n = o_P(1)$. Also, since, by Lemma 1(i), $x/\mu(x)$ is decreasing,

$$\int_2^{r_n} \frac{x}{\mu(x)}\,dx \ge \frac{r_n(r_n-2)}{\mu(r_n)} \quad \text{and} \quad \int_{r_n}^{n} \frac{x}{\mu(x)}\,dx \le \frac{r_n(n-r_n)}{\mu(r_n)}.$$

Therefore, the second statement in (24) entails $r_n - 2 = o(n - r_n)$ as $n \to \infty$, and consequently $r_n = o(n)$. Hence, the sequence r_n, $n \ge 1$, fulfills all requirements of Proposition 2.

For the construction of the numbers r_n, note that, from Lemma 1(i), we have $x/\mu(x) \ge 2/(\mu(2)(x-1))$ for $x \ge 2$, and consequently,

$$\int_2^{\infty} \frac{x}{\mu(x)}\,dx = \infty. \tag{25}$$

We distinguish two cases. If $\int_2^\infty \frac{dx}{\mu(x)} < \infty$, then the required sequence is easily obtained, because the first condition of (24) is fulfilled for any divergent sequence $r_n \le n$ and the second one by reason of (25), if only r_n is diverging slowly enough.

Thus, let us assume that $\int_2^\infty \frac{dx}{\mu(x)} = \infty$, and let $r_{n,m} \ge 2$ for given $m \in \mathbb{N}$ be the solution of the equation

$$\int_{r_{n,m}}^{n} \frac{dx}{\mu(x)} = \frac{1}{m},$$

which exists for $n \ge 3$ and $m \ge 1/\int_2^3 \frac{dx}{\mu(x)}$. Since $\int_2^\infty \frac{dx}{\mu(x)} = \infty$, we have $r_{n,m} \to \infty$ as $n \to \infty$. It follows

$$\int_{r_{n,m}}^{n} \frac{x}{\mu(x)}\,dx \ge r_{n,m}/m \quad \text{and} \quad \int_2^{r_{n,m}} \frac{x}{\mu(x)}\,dx = o(r_{n,m})$$

as $n \to \infty$ because of $x = o(\mu(x))$ from Lemma 1(iii). Therefore, there are natural numbers $n_1 < n_2 < \cdots$ such that

$$\int_2^{r_{n,m}} \frac{x}{\mu(x)}\,dx \le \frac{1}{m} \int_{r_{n,m}}^{n} \frac{x}{\mu(x)}\,dx$$

for all $n \ge n_m$. Now, letting $r_n := r_{n,m}$ for $n = n_m, \dots, n_{m+1} - 1$, we obtain

$$\int_{r_n}^{n} \frac{dx}{\mu(x)} \le \frac{1}{m} \quad \text{and} \quad \int_2^{r_n} \frac{x}{\mu(x)}\,dx \le \frac{1}{m} \int_{r_n}^{n} \frac{x}{\mu(x)}\,dx$$

for all $n \ge n_m$. This implies (24).

Applying now Proposition 2 with $f(x) = x$, we obtain from (24)

$$\int_0^{\tilde{\rho}_n} N_n(t)\,dt \overset{1}{\sim} \int_{r_n}^{n} \frac{x}{\mu(x)}\,dx \sim \int_2^{n} \frac{x}{\mu(x)}\,dx$$

and

$$\mathbf{E}\left[\int_{\tilde{\rho}_n}^{\tilde{\tau}_n} N_n(t)\,dt\right] = O\left(\int_2^{r_n} \frac{x}{\mu(x)}\,dx\right) = o\left(\int_2^{n} \frac{x}{\mu(x)}\,dx\right).$$

This implies our claim.

5. Proof of Theorem 2. Again, let r_n, $n \geq 1$, be any sequence fulfilling the assumptions of Proposition 1. We investigate the lengths

$$\widehat{\ell}_n^* := \int_0^{\tilde{\rho}_n} \widehat{N}_n(t)\, dt = \sum_{i=0}^{\rho_n - 1} W_i Y_i, \tag{26}$$

which are the total lengths of the external branches up to the ρ_nth merger. Here, Y_i, $i \geq 0$, denotes the number of external branches extant after the first i merging events and W_i as above the waiting time at the state X_i. In the proof, we approximate $\widehat{\ell}_n^*$ by its conditional expectation given the block-counting process N_n, which in turn can be handled by means of Propositions 2 and 3. We shall employ the representation

$$Y_i = \sum_{k=1}^{n} I_{\{\zeta_k \geq i\}}, \tag{27}$$

where ζ_k denotes the number of coalescent events before the kth external branch (out of n) merges with some other branches within the coalescent.

LEMMA 4. *For $i, j \geq 0$, $k, l = 1, \ldots, n$, we have*

$$\mathbf{P}(\zeta_k \geq i \mid N_n) = \frac{X_i - 1}{n - 1} \prod_{m=0}^{i-1} \left(1 - \frac{1}{X_m}\right) \quad a.s.$$

and for $k \neq l$,

$$\mathbf{P}(\zeta_k \geq i, \zeta_l \geq j \mid N_n) \leq \mathbf{P}(\zeta_k \geq i \mid N_n)\mathbf{P}(\zeta_l \geq j \mid N_n) \quad a.s.$$

PROOF. Let A be a subset of $\{1, \ldots, n\}$ with $a \geq 1$ elements, and let ζ_A be the number of mergers before one of the branches ending in A gets involved in a merging event. Given Δ_1, the first merger consists of a uniformly random choice of $\Delta_1 + 1$ members out of $X_0 = n$ elements. Therefore, we have

$$\begin{aligned}\mathbf{P}(\zeta_A \geq 1 \mid N_n) &= \frac{\binom{X_0 - a}{\Delta_1 + 1}}{\binom{X_0}{\Delta_1 + 1}} = \frac{(X_0 - a) \cdots (X_1 - a)}{X_0 \cdots X_1} \\ &= \frac{(X_1 - 1) \cdots (X_1 - a)}{X_0 \cdots (X_0 - a + 1)} \quad \text{a.s.}\end{aligned}$$

or

$$\mathbf{P}(\zeta_A \geq 1 \mid N_n) = \frac{(X_1 - 1) \cdots (X_1 - a)}{(X_0 - 1) \cdots (X_0 - a)} \left(1 - \frac{a}{X_0}\right) \quad \text{a.s.}$$

Because of the Markov property, we may iterate this formula yielding

$$\mathbf{P}(\zeta_A \geq i \mid N_n) = \frac{(X_i - 1) \cdots (X_i - a)}{(X_0 - 1) \cdots (X_0 - a)} \prod_{m=0}^{i-1} \left(1 - \frac{a}{X_m}\right) \quad \text{a.s.} \tag{28}$$

In particular, our first claim follows with $A = \{k\}$ and $a = 1$.

Similarly, for $k \neq l$ and $i \leq j$ with $\zeta'_{\{l\}} := \zeta_{\{l\}} - \zeta_{\{k,l\}}$ and $N'_{X_i}(t) := N_n(t + W_0 + \cdots + W_{i-1})$, $t \geq 0$, by means of the Markov property,

$$\begin{aligned}
\mathbf{P}(\zeta_k \geq i, \zeta_l \geq j \mid N_n) &= \mathbf{P}(\zeta_{\{k,l\}} \geq i \mid N_n)\mathbf{P}\big(\zeta'_{\{l\}} \geq j - i \mid N'_{X_i}\big) \\
&= \frac{(X_i - 1)(X_i - 2)}{(X_0 - 1)(X_0 - 2)} \prod_{m=0}^{i-1} \left(1 - \frac{2}{X_m}\right) \\
&\quad \times \frac{X_j - 1}{X_i - 1} \prod_{m=i}^{j-1} \left(1 - \frac{1}{X_m}\right) \quad \text{a.s.}
\end{aligned}$$

Since $X_i \leq X_0$ and $(1 - 2/X_m) \leq (1 - 1/X_m)^2$, this implies

$$\begin{aligned}
\mathbf{P}(\zeta_k \geq i, \zeta_l \geq j \mid N_n) &\leq \frac{(X_i - 1)^2}{(X_0 - 1)^2} \prod_{m=0}^{i-1} \left(1 - \frac{1}{X_m}\right)^2 \times \frac{X_j - 1}{X_i - 1} \prod_{m=i}^{j-1} \left(1 - \frac{1}{X_m}\right) \\
&= \mathbf{P}(\zeta_k \geq i \mid N_n)\mathbf{P}(\zeta_l \geq j \mid N_n) \quad \text{a.s.,}
\end{aligned}$$

which is our second claim. $\square$

PROOF OF THEOREM 2. (i) First, we consider $\mathbf{E}[\widehat{\ell}_n^* \mid N_n]$. Due to (27) and Lemma 4, we have

$$\begin{aligned}
\mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] &= \sum_{i=0}^{\rho_n - 1} \sum_{k=1}^{n} W_i \mathbf{P}(\zeta_k \geq i \mid N_n) \\
&= \frac{n}{n-1} \sum_{i=0}^{\rho_n - 1} W_i (X_i - 1) \prod_{m=0}^{i-1} \left(1 - \frac{1}{X_m}\right) \quad \text{a.s.}
\end{aligned} \tag{29}$$

Since $\sum_{m=0}^{i-1} X_m^{-2} \leq \sum_{a=X_{i-1}}^{\infty} a^{-2} \leq (X_{i-1} - 1)^{-1} \leq (r_n - 1)^{-1}$ for $i \leq \rho_n$ and in view of Proposition 3,

$$\prod_{m=0}^{i-1} \left(1 - \frac{1}{X_m}\right) = \exp\left(-\sum_{m=0}^{i-1} \frac{1}{X_m} + O\big(r_n^{-1}\big)\right) = \frac{\kappa(X_i)}{\kappa(n)} \exp\big(o_P(1)\big),$$

where the $o_P(1)$ may be taken uniformly in $i < \rho_n$ in the sense of Proposition 3. Thus, we obtain

$$\mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] \overset{P}{\sim} \frac{1}{\kappa(n)} \sum_{i=0}^{\rho_n - 1} W_i (X_i - 1)\kappa(X_i) = \frac{1}{\kappa(n)} \int_0^{\tilde{\rho}_n} f\big(N_n(t)\big)\, dt$$

with $f(x) := (x - 1)\kappa(x)$. This function satisfies the assumption of Proposition 2. Because of $f(x) \sim \mu(x)$ and $r_n \to \infty$, we obtain

$$\mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] \overset{P}{\sim} \frac{1}{\kappa(n)} \int_{r_n}^{n} (x - 1)\kappa(x) \frac{dx}{\mu(x)} \sim \frac{n - r_n}{\kappa(n)} = \frac{n(n - r_n)}{\mu(n)}. \tag{30}$$

(ii) Next, we have

$$\begin{aligned}
&\mathbf{E}\big[\big(\widehat{\ell}_n^* - \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big]\big)^2 \mid N_n\big] \\
&\quad = \mathbf{E}\Bigg[\Bigg(\sum_{i=0}^{\rho_n-1}\sum_{k=1}^{n}\big(W_i I_{\{\zeta_k \ge i\}} - W_i \mathbf{P}(\zeta_k \ge i \mid N_n)\big)\Bigg)^2 \,\Bigg|\, N_n\Bigg] \\
&\quad = \sum_{i,j=0}^{\rho_n-1}\sum_{k,l=1}^{n} W_i W_j \big(\mathbf{P}(\zeta_k \ge i, \zeta_l \ge j \mid N_n) - \mathbf{P}(\zeta_k \ge i \mid N_n)\mathbf{P}(\zeta_l \ge j \mid N_n)\big) \quad \text{a.s.}
\end{aligned}$$

Applying Lemma 4, it follows

$$\begin{aligned}
\mathbf{E}\big[\big(\widehat{\ell}_n^* - \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big]\big)^2 \mid N_n\big] &\le \sum_{i,j=0}^{\rho_n-1}\sum_{k=1}^{n} W_i W_j \mathbf{P}(\zeta_k \ge i \vee j \mid N_n) \\
&\le \sum_{i,j=0}^{\rho_n-1} W_i W_j \sum_{k=1}^{n} \mathbf{P}(\zeta_k \ge i \mid N_n) \\
&= \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] \sum_{j=0}^{\rho_n-1} W_j \\
&= \tilde{\rho}_n \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] \quad \text{a.s.}
\end{aligned}$$

Since, by assumption, $\tilde{\rho}_n = o_P(1)$ and $r_n \le \gamma n$, (30) yields

$$\mathbf{E}\big[\big(\widehat{\ell}_n^* - \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big]\big)^2 \mid N_n\big] = o_P\bigg(\frac{n(n-r_n)}{\mu(n)}\bigg) = o_P\bigg(\frac{n^2}{\mu(n)}\bigg).$$

Because of Lemma 1(i), $n(n-1)/\mu(n)$ is increasing, which implies

$$\widehat{\ell}_n^* - \mathbf{E}\big[\widehat{\ell}_n^* \mid N_n\big] = o_P\bigg(\frac{n}{\sqrt{\mu(n)}}\bigg) = o_P\bigg(\frac{n^2}{\mu(n)}\bigg), \tag{31}$$

and in view of (30),

$$\widehat{\ell}_n^* \overset{P}{\sim} \frac{n(n-r_n)}{\mu(n)}. \tag{32}$$

In particular, as discussed in the proof of Theorem 1 and addressed in the Introduction, this approximation is valid for the sequence $r_n = cn$ with $0 < c < 1$.

(iii) Finally, let us switch to the numbers r_n, $n \ge 1$, constructed in the proof of Theorem 1 and fulfilling (24) as well as $r_n = o(n)$. As above,

$$\begin{aligned}
\mathbf{E}\big[\widehat{\ell}_n - \widehat{\ell}_n^* \mid N_n\big] &= \frac{n}{n-1}\sum_{i=\rho_n}^{\tau_n-1} W_i(X_i - 1)\prod_{m=0}^{i-1}\bigg(1 - \frac{1}{X_m}\bigg) \\
&\le 2\exp\Bigg(-\sum_{m=0}^{\rho_n-1}\frac{1}{X_m}\Bigg)\sum_{i=\rho_n}^{\tau_n-1} W_i X_i \quad \text{a.s.}
\end{aligned}$$

From the Markov property and Theorem 1, applied to the coalescent with initial value X_{ρ_n}, we obtain

$$\sum_{i=\rho_n}^{\tau_n-1} W_i X_i \overset{P}{\sim} \int_2^{X_{\rho_n}} \frac{x}{\mu(x)}\,dx \le \int_2^{r_n} \frac{x}{\mu(x)}\,dx.$$

By (24) and by monotonicity of $x/\mu(x)$, it follows

$$\sum_{i=\rho_n}^{\tau_n-1} W_i X_i = o_P\left(\int_{r_n}^{n} \frac{x}{\mu(x)}\,dx\right) = o_P\left(\frac{r_n}{\mu(r_n)} n\right).$$

Moreover, from Proposition 3,

$$\exp\left(-\sum_{m=0}^{\rho_n-1} \frac{1}{X_m}\right) \overset{P}{\sim} \frac{\mu(r_n)}{r_n}\frac{n}{\mu(n)},$$

and we arrive at

$$\mathbf{E}[\widehat{\ell}_n - \widehat{\ell}_n^* \mid N_n] = o_P\left(\frac{n^2}{\mu(n)}\right).$$

Hence,

$$\widehat{\ell}_n - \widehat{\ell}_n^* = o_P\left(\frac{n^2}{\mu(n)}\right).$$

This estimate, in combination with (32) and $r_n = o(n)$, proves our theorem. □

6. Proof of Theorem 3. (i) As to the second claim,

$$\int_2^n \left(\frac{x}{\mu(x)} - \frac{n}{\mu(n)}\right) dx = \int_2^n \int_x^n \frac{\kappa'(y)}{\kappa(y)^2}\,dy\,dx = \int_2^n (y-2)\frac{\kappa'(y)}{\kappa(y)^2}\,dy,$$

and by Lemma 1(ii) and Lemma 2,

$$\int_2^n \left(\frac{x}{\mu(x)} - \frac{n}{\mu(n)}\right) dx \sim \int_2^n y\frac{\lambda(y)}{\mu(y)^2}\,dy \sim \int_2^n \frac{L(y)}{L^*(y)^2}\,dy \sim \frac{nL(n)}{L^*(n)^2}.$$

(ii) Turning to the first claim, we now strive for a lower bound for $\check{\ell}_n$. We resort to the definitions (23) and (26), and set

$$\check{\ell}_n^* := \ell_n^* - \widehat{\ell}_n^*.$$

Again, we first investigate its conditional expectation given N_n. Note that $\mathbf{E}[\ell_n^* \mid N_n] = \ell_n^*$ a.s. and, therefore, in view of (31),

$$\check{\ell}_n^* - \mathbf{E}[\check{\ell}_n^* \mid N_n] = \mathbf{E}[\widehat{\ell}_n^* \mid N_n] - \widehat{\ell}_n^* = o_P\left(\frac{n}{\sqrt{\mu(n)}}\right).$$

Lemma 1(i) implies $\mu(n) \geq n\mu(2)/2$ for $n \geq 2$, hence

$$\breve{\ell}_n^* - \mathbf{E}[\breve{\ell}_n^* \mid N_n] = o_P(n^{1/2}). \tag{33}$$

We like to estimate $\mathbf{E}[\breve{\ell}_n^* \mid N_n]$ from below. For this purpose, we specify our choice of the numbers r_n. We fix $h \in \mathbb{N}$, and we define stopping times $0 = \rho_{n,h} \leq \rho_{n,h-1} \leq \cdots \leq \rho_{n,1}$ and the corresponding times $\tilde{\rho}_{n,g}$ as

$$\begin{aligned} \rho_{n,g} &:= \min\left\{i \geq 0 : X_i \leq \frac{g}{h}n\right\}, \\ \tilde{\rho}_{n,g} &:= \inf\left\{t \geq 0 : N_n(t) \leq \frac{g}{h}n\right\}, \quad g = 1, \dots, h. \end{aligned} \tag{34}$$

As we already argued in the proof of Theorem 1, we may apply Propositions 2 and 3 to these stopping times. We now proceed in the same manner as in (29). Using $X_i \leq n$, respectively $X_i \geq n(X_i - 1)/(n-1)$, we obtain

$$\begin{aligned} &\mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i(X_i - Y_i) \,\middle|\, N_n\right] \\ &\quad = \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i\left(X_i - \frac{n}{n-1}(X_i - 1)\prod_{m=0}^{i-1}\left(1 - \frac{1}{X_m}\right)\right) \\ &\quad \geq \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i X_i\left(1 - \prod_{m=0}^{i-1}\left(1 - \frac{1}{X_m}\right)\right) \\ &\quad \geq \left(1 - \prod_{m=0}^{\rho_{n,g}-1}\left(1 - \frac{1}{X_m}\right)\right)\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i X_i \\ &\quad \geq \left(1 - \exp\left(-\sum_{m=0}^{\rho_{n,g}-1}\frac{1}{X_m}\right)\right)\int_{\tilde{\rho}_{n,g}}^{\tilde{\rho}_{n,g-1}} N_n(t)\,dt \quad \text{a.s.} \end{aligned}$$

Proposition 2, together with Lemma 2(ii), implies

$$\int_{\tilde{\rho}_{n,g}}^{\tilde{\rho}_{n,g-1}} N_n(t)\,dt \overset{1}{\sim} \int_{(g-1)n/h}^{gn/h} \frac{x}{\mu(x)}\,dx \sim \frac{n}{hL^*(n)},$$

and Proposition 3 and Lemma 2(ii) yield

$$\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m} \overset{1}{\sim} \log\frac{\kappa(n)}{\kappa(gn/h)} \sim \log\frac{L^*(n)}{L^*(gn/h)}.$$

Because L^* is slowly varying and because of (10), we have

$$\log\frac{L^*(n)}{L^*(gn/h)} \sim \frac{L^*(n) - L^*(gn/h)}{L^*(gn/h)} \sim \frac{L(n)}{L^*(n)}\log\frac{h}{g}.$$

Hence, because of Lemma 2(ii),

$$\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m} \overset{1}{\sim} \frac{L(n)}{L^*(n)} \log \frac{h}{g} = o_P(1) \tag{35}$$

and

$$1 - \exp\Biggl(-\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m}\Biggr) \overset{P}{\sim} \frac{L(n)}{L^*(n)} \log \frac{h}{g}.$$

Altogether,

$$\mathbf{E}\Biggl[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i(X_i - Y_i) \Bigm| N_n\Biggr] \geq \bigl(1 + o_P(1)\bigr) \frac{nL(n)}{L^*(n)^2} \frac{1}{h} \log \frac{h}{g}$$

and, consequently,

$$\begin{aligned} \mathbf{E}\bigl[\breve{\ell}_n^* \mid N_n\bigr] &= \mathbf{E}\Biggl[\sum_{i=0}^{\rho_{n,1}-1} W_i(X_i - Y_i) \Bigm| N_n\Biggr] \\ &\geq \bigl(1 + o_P(1)\bigr) \frac{nL(n)}{L^*(n)^2} \sum_{g=2}^{h} \frac{1}{h} \log \frac{h}{g} \\ &\geq \bigl(1 + o_P(1)\bigr) \frac{nL(n)}{L^*(n)^2} \int_{2/h}^{1} \log \frac{1}{z}\, dz \\ &= \bigl(1 + o_P(1)\bigr) \frac{nL(n)}{L^*(n)^2} \biggl(1 - \frac{2}{h} - \frac{2}{h} \log \frac{2}{h}\biggr). \end{aligned}$$

In view of (33), this estimate transfers to $\breve{\ell}_n^*$. We have $\breve{\ell}_n \geq \breve{\ell}_n^*$ and, therefore, letting $h \to \infty$, we obtain

$$\breve{\ell}_n \geq \bigl(1 + o_P(1)\bigr) \frac{nL(n)}{L^*(n)^2} \tag{36}$$

as $n \to \infty$.

(iii) Coming to an upper bound, we have, in view of $\prod_{m=0}^{i-1}(1 - 1/X_m) \geq 1 - \sum_{m=0}^{i-1} 1/X_m$,

$$\begin{aligned} \mathbf{E}[\breve{\ell}_n] &= \mathbf{E}\bigl[\mathbf{E}[\breve{\ell}_n \mid N_n]\bigr] \\ &= \mathbf{E}\Biggl[\sum_{i=0}^{\tau_n-1} W_i\Biggl(X_i - \frac{n}{n-1}(X_i - 1) \prod_{m=0}^{i-1}\biggl(1 - \frac{1}{X_m}\biggr)\Biggr)\Biggr] \\ &\leq \mathbf{E}\Biggl[\sum_{i=0}^{\tau_n-1} W_i\Biggl(1 + \frac{n}{n-1} X_i \sum_{m=0}^{i-1} \frac{1}{X_m}\Biggr)\Biggr]. \end{aligned} \tag{37}$$

From Proposition 2, and since $x = o(\mu(x))$ (see Lemma 1(iii)),

$$\mathbf{E}\left[\sum_{i=0}^{\tau_n-1} W_i\right] = \mathbf{E}[\tilde{\tau}_n] = O\left(\int_2^n \frac{dx}{\mu(x)}\right) = o\left(\int_2^n \frac{dx}{x}\right) = o(\log n). \tag{38}$$

Furthermore, by means of the Markov property,

$$\begin{aligned}
\mathbf{E}\left[\sum_{i=0}^{\tau_n-1} W_i X_i \sum_{m=0}^{i-1} \frac{1}{X_m}\right] &= \sum_{i=0}^{n-1} \mathbf{E}\left[\frac{X_i}{\lambda(X_i)} \sum_{m=0}^{i-1} \frac{1}{X_m}; X_i > 2\right] \\
&= \sum_{i=0}^{n-1} \mathbf{E}\left[\frac{X_i \Delta_{i+1}}{\mu(X_i)} \sum_{m=0}^{i-1} \frac{1}{X_m}; X_i > 2\right] \\
&= \mathbf{E}\left[\sum_{m=0}^{\tau_n-1} \frac{1}{X_m} \sum_{i=m+1}^{\tau_n-1} \frac{X_i \Delta_{i+1}}{\mu(X_i)}\right].
\end{aligned}$$

Since $x/\mu(x)$ is decreasing, we have $\sum_{i=m+1}^{\tau_n-1} X_i \Delta_{i+1}/\mu(X_i) \le \int_1^{X_{m+1}} x/\mu(x)\,dx$, where we let $x/\mu(x) := 2/\mu(2)$ for $1 \le x \le 2$. Thus,

$$\mathbf{E}\left[\sum_{i=0}^{\tau_n-1} W_i X_i \sum_{m=0}^{i-1} \frac{1}{X_m}\right] \le \mathbf{E}\left[\sum_{m=0}^{\tau_n-1} \frac{1}{X_m} \int_1^{X_{m+1}} \frac{x}{\mu(x)}\,dx\right].$$

Invoking the Markov property once again, we obtain

$$\mathbf{E}\left[\sum_{i=0}^{\tau_n-1} W_i X_i \sum_{m=0}^{i-1} \frac{1}{X_m}\right] \le \mathbf{E}\left[\sum_{m=0}^{\tau_n-1} \frac{\Delta_{m+1}}{(X_m-1)\nu(X_m)} \int_1^{X_{m+1}} \frac{x}{\mu(x)}\,dx\right],$$

and taking now into account that the functions $(x-1)\nu(x)$ and $\int_1^x z/\mu(z)\,dz$ are increasing, we end up with

$$\mathbf{E}\left[\sum_{i=0}^{\tau_n-1} W_i X_i \sum_{m=0}^{i-1} \frac{1}{X_m}\right] \le \int_1^n \frac{1}{(x-1)\nu(x)} \int_1^x \frac{z}{\mu(z)}\,dz\,dx, \tag{39}$$

where we let $(x-1)\nu(x) := \nu(2)$ for $1 \le x \le 2$. Using Lemma 2, it follows

$$\begin{aligned}
\int_1^n \frac{1}{(x-1)\nu(x)} \int_1^x \frac{z}{\mu(z)}\,dz\,dx &\sim \int_2^n \frac{L(x)}{xL^*(x)} \int_2^x \frac{1}{L^*(z)}\,dz\,dx \\
&\sim \int_2^n \frac{L(x)}{xL^*(x)} \frac{x}{L^*(x)}\,dx \\
&\sim \frac{nL(n)}{L^*(n)^2}.
\end{aligned}$$

Together with (37), (38) and (39), we obtain all in all

$$\mathbf{E}[\check{\ell}_n] \le \big(1 + o(1)\big)\frac{nL(n)}{L^*(n)^2}.$$

Combining this result with (36), we get

$$\breve{\ell}_n \overset{P}{\sim} \frac{nL(n)}{L^*(n)^2},$$

and invoking once again the convergence criterion of F. Riesz, also L_1-convergence follows. This completes the proof of Theorem 3.

7. Proof of Theorem 4. (i) Concerning the first claim, we have, from Theorem 2 and Lemma 2(ii),

$$\widehat{\ell}_{n,1} = \widehat{\ell}_n \overset{P}{\sim} \frac{n^2}{\mu(n)} \sim \frac{n}{L^*(n)},$$

where L^* is slowly varying. In order to also obtain L_1-convergence, we use Theorem 1 and Lemma 2(ii), saying that

$$\ell_n \overset{1}{\sim} \int_2^n \frac{x}{\mu(x)}\,dx \sim \int_2^n \frac{dx}{L^*(x)} \sim \frac{n}{L^*(n)}.$$

In particular, the random variables $L^*(n)\ell_n/n$, $n \geq 1$, make a uniformly integrable sequence. Since $\widehat{\ell}_{n,1} \leq \ell_n$, this holds true also for the random variables $L^*(n)\widehat{\ell}_{n,1}/n$, $n \geq 1$. Therefore, the convergence in probability above converts to L_1-convergence.

(ii) Now we turn to the case $a \geq 2$. We preliminary study lengths of the form

$$\widehat{\ell}^*_{n,a} = \int_{\tilde{\sigma}_n}^{\tilde{\rho}_n} \widehat{N}_{n,a}(t)\,dt = \sum_{i=\sigma_n}^{\rho_n - 1} W_i Y_{i,a},$$

where $Y_{i,a}$ denotes the number of internal branches of order a present in the coalescent after i merging events. We are going to bound these numbers from below. Let A denote a subset of $\{1,\dots,n\}$ with a elements. For $1 \leq k \leq i$, let $E_{i,k,A}$ be the event that the external branches ending in A are not involved in the first $k-1$ mergers, next coalesce with the kth merger to one lineage without any other branch participating, and then remain untouched by merging events until the ith merger. These are disjoint events, which all contribute to $Y_{i,a}$. Thus,

$$Y_{i,a} \geq Y'_{i,a} := \sum_{k=1}^{i} \sum_{A} I_{E_{i,k,A}},$$

where the second sum is taken over all $A \subset \{1,\dots,n\}$ with a elements.

LEMMA 5. *Let both $A, A' \subset \{1,\dots,n\}$ have $a \geq 1$ elements. Then, for $1 \leq k \leq i$,*

$$\mathbf{P}(E_{i,k,A} \mid N_n) = \frac{a!}{(X_0-1)\cdots(X_0-a)} \frac{X_i}{X_{k-1}} \times \prod_{m=0}^{k-2}\left(1-\frac{a}{X_m}\right) \prod_{m=k}^{i}\left(1-\frac{1}{X_m}\right) I_{\{\Delta_k = a-1\}} \quad a.s.$$

and for $A \neq A'$ *or* $k \neq l$ *and for* $1 \leq j \leq \ell$

$$\mathbf{P}(E_{i,k,A} \cap E_{j,l,A'} \mid N_n) \leq \bigl(1 + O\bigl(X_{k-1}^{-1}\bigr) + O\bigl(X_{l-1}^{-1}\bigr)\bigr)\mathbf{P}(E_{i,k,A} \mid N_n)\mathbf{P}(E_{j,l,A'} \mid N_n) \quad \textit{a.s.}$$

PROOF. We proceed similarly as in the proof of Lemma 4. From (28) and the Markov property,

$$\mathbf{P}(E_{i,k,A} \mid N_n) = \frac{(X_{k-1}-1)\cdots(X_{k-1}-a)}{(X_0-1)\cdots(X_0-a)} \prod_{m=0}^{k-2}\left(1-\frac{a}{X_m}\right) \times \binom{X_{k-1}}{a}^{-1} I_{\{\Delta_k=a-1\}} \times \frac{X_i-1}{X_k-1}\prod_{m=k}^{i-1}\left(1-\frac{1}{X_m}\right) \quad \text{a.s.}$$

Note that $X_k - 1 = X_{k-1} - a$ on the event $\{\Delta_k = a-1\}$. Thus all factors containing X_{k-1} cancel up to the term X_{k-1} in the denominator. Replacing also $X_i - 1$ by $X_i(1 - 1/X_i)$, the first statement follows.

For the second claim note that, in the case $A \neq A'$ and $A \cap A' \neq \varnothing$ or in the case $A = A'$ and $k \neq l$, the events $E_{i,k,A}$ and $E_{j,l,A'}$ are disjoint; thus making our claim obvious. Therefore, we may assume that $A \cap A' = \varnothing$. Let us consider the case $k < l < i \leq j$. Then, from (28) and the Markov property,

$$\begin{aligned}
&\mathbf{P}(E_{i,k,A} \cap E_{j,l,A'} \mid N_n) \\
&\quad = \frac{(X_{k-1}-1)\cdots(X_{k-1}-2a)}{(X_0-1)\cdots(X_0-2a)} \prod_{m=0}^{k-2}\left(1-\frac{2a}{X_m}\right) \\
&\quad\quad \times \binom{X_{k-1}}{a}^{-1} I_{\{\Delta_k=a-1\}} \times \frac{(X_{l-1}-1)\cdots(X_{l-1}-a-1)}{(X_k-1)\cdots(X_k-a-1)} \prod_{m=k}^{l-2}\left(1-\frac{a+1}{X_m}\right) \\
&\quad\quad \times \binom{X_{l-1}}{a}^{-1} I_{\{\Delta_l=a-1\}} \times \frac{(X_i-1)(X_i-2)}{(X_l-1)(X_l-2)} \prod_{m=l}^{i-1}\left(1-\frac{2}{X_m}\right) \\
&\quad\quad \times \frac{X_j-1}{X_i-1}\prod_{m=i}^{j-1}\left(1-\frac{1}{X_m}\right) \quad \text{a.s.}
\end{aligned}$$

Here, the product $(X_k - 1)\cdots(X_k - a - 1) = (X_{k-1} - a)\cdots(X_{k-1} - 2a)$ cancels out on the event $\{\Delta_k = a - 1\}$, and again only the factor X_{k-1} remains. Similarly, $X_l - 2$ and $X_{l-1} - a - 1$ cancel. By increasing some other terms, we get

$$\begin{aligned}
\mathbf{P}(E_{i,k,A} \cap E_{j,l,A'} \mid N_n) \leq{}& \frac{a!a!}{(X_0-1)\cdots(X_0-2a)} \prod_{m=0}^{k-2}\left(1-\frac{a}{X_m}\right)^2 \\
& \times I_{\{\Delta_k=a-1\}}\frac{1}{X_{k-1}} \prod_{m=k}^{l-2}\left(1-\frac{a}{X_m}\right)\left(1-\frac{1}{X_m}\right)
\end{aligned}$$

$$\times I_{\{\Delta_l=a-1\}} \frac{1}{X_{l-1}} \prod_{m=l}^{i-1} \left(1 - \frac{1}{X_m}\right)\left(1 - \frac{1}{X_m}\right)$$
$$\times (X_i - 1)(X_j - 1) \prod_{m=i}^{j-1} \left(1 - \frac{1}{X_m}\right) \quad \text{a.s.}$$

Replacing also $X_i - 1$ and $X_j - 1$ as above and comparing it to our first formula, we obtain

$$\begin{aligned}
&\mathbf{P}(E_{i,k,A} \cap E_{j,l,A'} \mid N_n) \\
&\quad \le \frac{(X_0 - 1)\cdots(X_0 - a)}{(X_0 - a - 1)\cdots(X_0 - 2a)} \left(1 - \frac{a}{X_{k-1}}\right)^{-1} \\
&\quad\quad \times \mathbf{P}(E_{i,k,A} \mid N_n)\mathbf{P}(E_{j,l,A'} \mid N_n) \quad \text{a.s.}
\end{aligned}$$

This implies our second claim. Other cases like $k < i < l < j$ are treated similarly. $\square$

(iii) Coming back to the proof of the theorem's second claim, we first consider, similar to above, conditional expectations given N_n. Recall from (34) the definition of $\rho_{n,g}$. From Lemma 5 for $g \ge 2$, we have

$$\begin{aligned}
&\mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\middle|\, N_n\right] \\
&\quad = \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i \sum_{k=1}^{i} \sum_{A} \mathbf{P}(E_{i,k,A} \mid N_n) \\
&\quad = \frac{X_0}{X_0 - a} \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i \sum_{k=1}^{i} \frac{X_i}{X_{k-1}} \prod_{m=0}^{k-2} \left(1 - \frac{a}{X_m}\right) \prod_{m=k}^{i} \left(1 - \frac{1}{X_m}\right) I_{\{\Delta_k=a-1\}} \quad \text{a.s.}
\end{aligned}$$

The products may be estimated from above by 1 and, by means of Proposition 3 and the bound $(1 - z) \ge \exp(-cz)$ for $z \le 1/2$ and a suitable $c > 0$, from below uniformly in i, k by

$$\prod_{m=0}^{\rho_{n,1}-1} \left(1 - \frac{a}{X_m}\right) \ge \exp\left(-ca \sum_{m=0}^{\rho_{n,1}-1} \frac{1}{X_m}\right) \overset{P}{\sim} \left(\frac{\kappa(n/h)}{\kappa(n)}\right)^{ca} \sim \left(\frac{L^*(n/h)}{L^*(n)}\right)^{ca} \sim 1.$$

Consequently, we may replace the products by 1 and obtain

$$\mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\middle|\, N_n\right] \overset{P}{\sim} \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i X_i \sum_{k=1}^{i} \frac{1}{X_{k-1}} I_{\{\Delta_k=a-1\}} \tag{40}$$

as $n \to \infty$.

From Lemma 2(i) and (35),

$$\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m} \mathbf{P}(\Delta_{m+1} = a-1) \overset{P}{\sim} \frac{1}{(a-1)a} \sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m} \overset{P}{\sim} \frac{1}{(a-1)a} \frac{L(n)}{L^*(n)} \log \frac{h}{g}.$$

By the Markov property,

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m}\left(I_{\{\Delta_{m+1}=a-1\}} - \mathbf{P}(\Delta_{m+1} = a-1)\right)\right)^2\right] \\
&\quad = \mathbf{E}\left[\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m^2}\left(I_{\{\Delta_{m+1}=a-1\}} - \mathbf{P}(\Delta_{m+1} = a-1)\right)^2\right] \\
&\quad \le \mathbf{E}\left[\frac{1}{X_{\rho_{n,g}-1} - 1}\right] \le \frac{1}{(gn/h) - 1}.
\end{aligned}$$

Therefore,

$$\sum_{m=0}^{\rho_{n,g}-1} \frac{1}{X_m} I_{\{\Delta_{m+1}=a-1\}} \overset{P}{\sim} \frac{1}{a(a-1)} \frac{L(n)}{L^*(n)} \log \frac{h}{g}.$$

Also, once more by Proposition 2 and Lemma 2(ii),

$$\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i X_i = \int_{\tilde{\rho}_{n,g}}^{\tilde{\rho}_{n,g-1}} N_n(t)\,dt \overset{1}{\sim} \int_{(g-1)n/h}^{gn/h} \frac{x}{\mu(x)}\,dx \sim \frac{n}{hL^*(n)}.$$

Together with (40), these formulas yield the following lower and upper bounds for $g \ge 2$:

$$\mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\middle|\, N_n\right] \ge \left(1 + o_P(1)\right) \frac{1}{a(a-1)} \frac{nL(n)}{L^*(n)^2} \frac{1}{h} \log \frac{h}{g} \tag{41}$$

and

$$\mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\middle|\, N_n\right] \le \left(1 + o_P(1)\right) \frac{1}{a(a-1)} \frac{nL(n)}{L^*(n)^2} \frac{1}{h} \log \frac{h}{g-1}. \tag{42}$$

(iv) Now we estimate the difference between $\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a}$ and its conditional expectation given N_n. We have

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} - \mathbf{E}\left[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\middle|\, N_n\right]\right)^2 \,\middle|\, N_n\right] \\
&\quad = \mathbf{E}\left[\Bigg(\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i \sum_{k=1}^{i} \sum_{A} \left(I_{E_{i,k,A}} - \mathbf{P}(E_{i,k,A} \mid N_n)\right)^2 \,\middle|\, N_n\right]
\end{aligned}$$

$$= \sum_{i,j=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i W_j \sum_{k=1}^{i} \sum_{l=1}^{j} \sum_{A,A'} \big(\mathbf{P}(E_{i,k,A} \cap E_{j,l,A'} \mid N_n) - \mathbf{P}(E_{i,k,A} \mid N_n) \mathbf{P}(E_{j,l,A'} \mid N_n) \big) \quad \text{a.s.}$$

For $A = A'$, $k = l$ and $i \le j$, we use the estimate

$$\mathbf{P}(E_{i,k,A} \cap E_{j,k,A} \mid N_n) \le \mathbf{P}(E_{i,k,A} \mid N_n) \quad \text{a.s.}$$

Taking also account of Lemma 5, we obtain with some $c > 0$,

$$\begin{aligned}
&\mathbf{E}\Biggl[\Biggl(\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} - \mathbf{E}\Biggl[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\Bigg|\, N_n\Biggr]\Biggr)^2 \,\Bigg|\, N_n\Biggr] \\
&\quad \le 2 \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} \sum_{j=i}^{\rho_{n,g-1}-1} \sum_{k=1}^{i} \sum_{A} W_i W_j \mathbf{P}(E_{i,k,A} \mid N_n) \\
&\qquad + c \sum_{i,j=\rho_{n,g}}^{\rho_{n,g-1}-1} \sum_{k=1}^{i} \sum_{l=1}^{j} \sum_{A,A'} W_i W_j \left(\frac{1}{X_k} + \frac{1}{X_l} \right) \mathbf{P}(E_{i,k,A} \mid N_n) \mathbf{P}(E_{j,l,A'} \mid N_n) \\
&\quad \le 2 \sum_{j=0}^{\rho_{n,g-1}-1} W_j \sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} \sum_{k=1}^{i} \sum_{A} W_i \mathbf{P}(E_{i,k,A} \mid N_n) \\
&\qquad + \frac{2c}{X_{\rho_{n,g-1}-1}} \left(\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} \sum_{k=1}^{i} \sum_{A} W_i \mathbf{P}(E_{i,k,A} \mid N_n) \right)^2 \\
&\quad = 2\tilde{\rho}_{n,g-1} \mathbf{E}\Biggl[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\Bigg|\, N_n\Biggr] + \frac{2ch}{n} \Biggl(\mathbf{E}\Biggl[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\Bigg|\, N_n\Biggr]\Biggr)^2 \quad \text{a.s.}
\end{aligned}$$

Using (42), $L_n = o(L^*(n))$ and the fact that $L^*(n)$ is increasing, this implies

$$\mathbf{E}\Biggl[\Biggl(\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} - \mathbf{E}\Biggl[\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \,\Bigg|\, N_n\Biggr]\Biggr)^2 \,\Bigg|\, N_n\Biggr] = o(n)$$

for $g \ge 2$, and thus, from (41),

$$\sum_{i=\rho_{n,g}}^{\rho_{n,g-1}-1} W_i Y'_{i,a} \ge \big(1 + o_P(1)\big) \frac{1}{a(a-1)} \frac{nL(n)}{L^*(n)^2} \frac{1}{h} \log \frac{h}{g}.$$

(v) The last formula implies for $a \geq 2$ that

$$\begin{aligned}
\widehat{\ell}_{n,a} &= \sum_{i=0}^{\tau_n-1} W_i Y_{i,a} \geq \sum_{i=\rho_{n,h-1}}^{\rho_{n,1}-1} W_i Y'_{i,a} \\
&\geq \big(1+o_P(1)\big)\frac{1}{a(a-1)}\frac{nL(n)}{L^*(n)^2}\sum_{g=2}^{h-1}\frac{1}{h}\log\frac{h}{g} \\
&\geq \big(1+o_P(1)\big)\frac{1}{a(a-1)}\frac{nL(n)}{L^*(n)^2}\int_{2/h}^{1}\log\frac{1}{z}\,dz.
\end{aligned}$$

Letting $h \to \infty$, we obtain the lower estimate

$$\widehat{\ell}_{n,a} \geq \big(1+o_P(1)\big)\frac{1}{a(a-1)}\frac{nL(n)}{L^*(n)^2}.$$

For an upper estimate, note that $\check{\ell}_n = \sum_{a\geq 2}\widehat{\ell}_{n,a}$. This formula and Theorem 3 imply for any natural number r,

$$\begin{aligned}
\widehat{\ell}_{n,a} &\leq \check{\ell}_n - \sum_{2\leq b\leq r, b\neq a}\widehat{\ell}_{b,n} \\
&\leq \big(1+o_P(1)\big)\frac{nL(n)}{L^*(n)^2}\bigg(1 - \sum_{2\leq b\leq r, b\neq a}\frac{1}{(b-1)b}\bigg) \\
&\overset{P}{\sim} \frac{nL(n)}{L^*(n)^2}\bigg(\frac{1}{r} + \frac{1}{(a-1)a}\bigg).
\end{aligned}$$

Letting $r \to \infty$ yields the upper estimates and we thus obtain altogether

$$\widehat{\ell}_{n,a} \overset{P}{\sim} \frac{1}{a(a-1)}\frac{nL(n)}{L^*(n)^2}.$$

In order to achieve also L_1-convergence, we deduce from Theorem 3 that the random variables $\check{\ell}_n L^*(n)^2/(nL(n))$ form a uniformly integrable sequence. Since $\widehat{\ell}_{n,a} \leq \check{\ell}_n$, the same holds true for the random variables $\widehat{\ell}_{n,a}L^*(n)^2/(nL(n))$. Thus, also L_1-convergence follows. This completes the proof of Theorem 4.

REFERENCES

[1] BASDEVANT, A.-L. and GOLDSCHMIDT, C. (2008). Asymptotics of the allele frequency spectrum associated with the Bolthausen–Sznitman coalescent. *Electron. J. Probab.* **13** 486–512. MR2386740

[2] BERESTYCKI, J., BERESTYCKI, N. and LIMIC, V. (2014). Asymptotic sampling formulae for Λ-coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 715–731. MR3224287

[3] BERESTYCKI, J., BERESTYCKI, N. and SCHWEINSBERG, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** 1835–1887. MR2349577

[4] BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory*. *Ensaios Matemáticos* [*Mathematical Surveys*] **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323

[5] BIRKNER, M., BLATH, J. and ELDON, B. (2013). Statistical properties of the site-frequency spectrum associated with Λ-coalescents. *Genetics* **195** 1037–1053.

[6] BOLTHAUSEN, E. and SZNITMAN, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197** 247–276. MR1652734

[7] DAHMER, I., KERSTING, G. and WAKOLBINGER, A. (2014). The total external branch length of beta-coalescents. *Combin. Probab. Comput.* **23** 1010–1027. MR3265836

[8] DELMAS, J.-F., DHERSIN, J.-S. and SIRI-JEGOUSSE, A. (2008). Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.* **18** 997–1025. MR2418236

[9] DHERSIN, J.-S. and MÖHLE, M. (2013). On the external branches of coalescents with multiple collisions. *Electron. J. Probab.* **18** no. 40, 11. MR3040550

[10] DRMOTA, M., IKSANOV, A., MÖHLE, M. and RÖSLER, U. (2007). Asymptotic results concerning the total branch length of the Bolthausen–Sznitman coalescent. *Stochastic Process. Appl.* **117** 1404–1421. MR2353033

[11] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*. *Vol. II*. Wiley, New York, NY. MR0270403

[12] GOLDSCHMIDT, C. and MARTIN, J. B. (2005). Random recursive trees and the Bolthausen–Sznitman coalescent. *Electron. J. Probab.* **10** 718–745. MR2164028

[13] JANSON, S. and KERSTING, G. (2011). On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16** 2203–2218. MR2861672

[14] KERSTING, G. (2012). The asymptotic distribution of the length of beta-coalescent trees. *Ann. Appl. Probab.* **22** 2086–2107. MR3025690

[15] KERSTING, G., PARDO, J. C. and SIRI-JÉGOUSSE, A. (2014). Total internal and external lengths of the Bolthausen–Sznitman coalescent. *J. Appl. Probab.* **51A** 73–86. MR3317351

[16] KERSTING, G., SCHWEINSBERG, J. and WAKOLBINGER, A. (2014). The evolving beta coalescent. *Electron. J. Probab.* **19** no. 64, 27. MR3238784

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

[18] KINGMAN, J. F. C. (1982). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43. MR0633178

[19] 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. MR2684740

[20] PITMAN, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27** 1870–1902. MR1742892

[21] SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** 1116–1125. MR1742154

[22] SCHWEINSBERG, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Probab.* **5** 1–11. MR1736720

[23] SPENCE, J. P., KAMM, J. A. and SONG, Y. S. (2016). The site frequency spectrum for general coalescents. *Genetics* **202** 1549–1561.

[24] TAVARÉ, S. (2004). Ancestral inference in population genetics. In *Lectures on Probability Theory and Statistics*. *Lecture Notes in Math.* **1837** 1–188. Springer, Berlin. MR2071630

[25] WAKELEY, J. (2009). *Coalescent Theory*: *An Introduction*. Roberts and Company Publishers, Greenwood Village, CO.

[26] WATTERSON, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7** 256–276. MR0366430

INSTITUT FÜR MATHEMATIK
GOETHE UNIVERSITÄT
60054 FRANKFURT AM MAIN
GERMANY
E-MAIL: diehl@math.uni-frankfurt.de
kersting@math.uni-frankfurt.de