Combinatorics, Probability and Computing (2014) 23, 1010–1027. © Cambridge University Press 2014
doi:10.1017/S0963548314000297

The Total External Branch Length of Beta-Coalescents†

IULIA DAHMER§ GÖTZ KERSTING and ANTON WAKOLBINGER

Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany
(e-mail: {dahmer,kersting,wakolbinger}@math.uni-frankfurt.de)

Received 22 December 2012; revised 2 August 2013; first published online 10 July 2014

For $1 < \alpha < 2$ we derive the asymptotic distribution of the total length of *external* branches of a Beta$(2-\alpha, \alpha)$-coalescent as the number n of leaves becomes large. It turns out that the fluctuations of the external branch length follow those of $\tau_n^{2-\alpha}$ over the entire parameter regime, where τ_n denotes the random number of coalescences that bring the n lineages down to one. This is in contrast to the fluctuation behaviour of the total branch length, which exhibits a transition at $\alpha_0 = (1+\sqrt{5})/2$ ([18]).

2010 *Mathematics subject classification*: Primary 60K35
Secondary 60F05, 60J10

1. Introduction and main results

The family of Beta coalescents belongs to the so-called Λ-coalescents introduced by Pitman [21] and Sagitov [22]: see the survey by N. Berestycki [1]. These are characterized by a probability measure $\Lambda(dp)$ on $[0,1]$. For $\Lambda = \delta_0$, one recovers the classical Kingman coalescent. For Λ having no mass in 0, they can be thought as modelling a random gene tree within a species in which single reproduction events affect a non-vanishing fraction of the population (see [8, 13, 26] for applications to certain maritime species). The basic ingredient for the coalescent dynamics is a Poisson process on $\mathbb{R} \times [0,1]$ with intensity measure $dt\, \nu(dp) = dt\, p^{-2}\Lambda(dp)$. With this Poisson process as a random input, the (Λ, n)-coalescent arises as follows. Initially there are n lineages (or 'particles'), and whenever a Poisson point (t, p) arrives, each of the lineages that exist at time t independently takes part in the corresponding coalescence event with probability p. By assumption, the pair coalescence rate equals 1, so the events that are relevant for coalescences within the sample

† Work partially supported by the DFG Priority Programme SPP 1590 'Probabilistic Structures in Evolution'.
§ Supported by the German Academic Exchange Service (DAAD).

arrive at finite rate. (More formally, this can be viewed as a partition-valued Markov process: see below.)

For the choice

$$\Lambda(dp) = \text{const}\left(\frac{p}{1-p}\right)^{1-\alpha} dp, \quad 0 < \alpha < 2,$$

the probability measure Λ is the Beta$(2-\alpha, \alpha)$-distribution. For $\alpha = 1$ one obtains the Bolthausen–Sznitman coalescent [7], and in the limit $\alpha \to 2$ one retrieves Kingman's coalescent. Beta coalescents play a prominent role because of their intimate connections to α-stable continuum branching processes (see [6]) and also figure as prominent alternatives to Kingman's coalescent from a statistical point of view (see [5, 24]).

Recently there has been considerable progress in investigating the asymptotic distribution (as the sample size $n \to \infty$) of interesting functionals of Beta coalescents, such as the total tree length L_n or the total length ℓ_n of all *external branches*, *i.e.*, those branches which end in a leaf. These quantities are also of interest in population genetics, since in the infinite sites model L_n figures as the (random) Poisson parameter of the total number of segregating sites in the sample, whereas ℓ_n is the (random) Poisson parameter of the total number of mutations carried by single individuals. Mathematicians interested in further reading in population genetics are referred to the monographs by Durrett [12] and Wakeley [25].

In this paper we will investigate the asymptotic distribution of the suitably normalized total external branch length ℓ_n in a (Beta$(2-\alpha, \alpha)$, n)-coalescent, with $1 < \alpha < 2$. In order to state our main result and put it into context with previous research, we give a few more formal definitions.

With a labelling of the initial particles by the numbers $1, \ldots, n$, the merging process is described by a sequence $\Pi_0, \ldots, \Pi_{\tau_n}$ of partitions of $\{1, \ldots, n\}$. Here Π_k consists of the classes of a (random) equivalence relation $\sim_k$ on $\{1, \ldots, n\}$, where $i \sim_k j$ states that $i, j \in \{1, \ldots, n\}$ have coalesced into one particle after k merging events. In particular, $\Pi_0 = \{\{1\}, \ldots, \{n\}\}$ and $\Pi_{\tau_n} = \{\{1, \ldots, n\}\}$, where τ_n denotes the (random) total number of merging events.

The process proceeds in continuous time. At times $0 = T_0 < T_1 < \cdots < T_{\tau_n}$ particles merge, and $n = X_0 > X_1 > \cdots > X_{\tau_n} = 1$ are the corresponding numbers of particles. Thus

$$X_k = \#\Pi_k.$$

Note that the times T_k and numbers X_k depend on n (which we suppress in the notation). For convenience we put $X_k = 1$ for $k > \tau_n$. The total branch length of the tree is given by

$$L_n = \sum_{k=0}^{\tau_n - 1} X_k (T_{k+1} - T_k),$$

whereas the *total length ℓ_n of all external branches* can be written as

$$\ell_n = \sum_{k=0}^{\tau_n - 1} Y_k (T_{k+1} - T_k), \tag{1.1}$$

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

with

$$Y_k = \#\{i \leqslant n : \{i\} \in \Pi_k\},$$

which is the number of external branches, still present up to time T_k. Note that $Y_{\tau_n} = 0$. For definiteness let $Y_k = 0$ for $k > \tau_n$.

The asymptotic distribution of L_n and ℓ_n has been studied in various publications. Drmota, Iksanov, Möhle and Rösler [11] and Iksanov and Möhle [15] studied the total length of the Bolthausen–Sznitman coalescent. Möhle's investigation [20] covers the case $0 < \alpha < 1$. In that case there is no substantial difference in the asymptotic behaviour of L_n and ℓ_n, more precisely,

$$\frac{L_n}{n} \xrightarrow{d} S \quad \text{and} \quad \frac{\ell_n}{n} \xrightarrow{d} S,$$

with

$$S = \int_0^\infty \exp(-\xi_t)\, dt,$$

where $(\xi_t)_{t \geqslant 0}$ denotes a certain driftless subordinator, depending on α.

In case $1 < \alpha < 2$ the situation is notably different. For this case Berestycki, Berestycki and Schweinsberg [4, 3] (see Theorems 1.9 and 9 therein) give the following results:

$$\frac{L_n}{n^{2-\alpha}} \to \frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha} \quad \text{and} \quad \frac{\ell_n}{L_n} \to 2-\alpha$$

in probability. See also Berestycki, Berestycki and Limic [2], who proved that these convergence results hold almost surely, and Dhersin and Yuan [10] who also investigated the convergence of moments of ℓ_n. We shall see that the difference to the case $0 < \alpha < 1$ becomes even more visible on the level of fluctuations.

Let ς denote a real-valued stable random variable with index $1 < \alpha < 2$, which is normalized by the properties

$$\mathbf{E}(\varsigma) = 0, \quad \mathbf{P}(\varsigma > x) = o(x^{-\alpha}), \quad \mathbf{P}(\varsigma < -x) \sim x^{-\alpha} \tag{1.2}$$

for $x \to \infty$. Thus ς is maximally skewed among the stable distributions of index α.

Further, let

$$c_1 = \alpha(\alpha-1)\Gamma(\alpha), \quad c_2 = \frac{\alpha(2-\alpha)(\alpha-1)^{1/\alpha+1}\Gamma(\alpha)}{\Gamma(2-\alpha)^{1/\alpha}}.$$

Theorem 1.1. *For the Beta-coalescent with* $1 < \alpha < 2$*,*

$$\frac{\ell_n - c_1 n^{2-\alpha}}{n^{1/\alpha+1-\alpha}} \xrightarrow{d} c_2\varsigma. \tag{1.3}$$

The corresponding statement for the Kingman coalescent can be found in Janson and Kersting [17]: up to a logarithmic correction in scale the analogue of (1.3) also holds for the case $\alpha = 2$, but with a normal limiting distribution.

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

It is of interest to contrast Theorem 1.1 with the corresponding statement for the total length L_n, obtained in Kersting [18]. For convenience we reproduce it here.

Theorem 1.2. *Let $c_i' = c_i/(2-\alpha)$, $i = 1,2$. Then, for $1 < \alpha < 2$, we have the following.*

(i) *If $1 < \alpha < \frac{1}{2}(1+\sqrt{5})$ (thus $1+\alpha-\alpha^2 > 0$), then*

$$\frac{L_n - c_1' n^{2-\alpha}}{n^{1/\alpha+1-\alpha}} \xrightarrow{d} \frac{c_2' \varsigma}{(1+\alpha-\alpha^2)^{1/\alpha}}.$$

(ii) *If $\alpha = \frac{1}{2}(1+\sqrt{5})$, then*

$$\frac{L_n - c_1' n^{2-\alpha}}{(\log n)^{1/\alpha}} \xrightarrow{d} c_2' \varsigma.$$

(iii) *If $\frac{1}{2}(1+\sqrt{5}) < \alpha < 2$, then*

$$L_n - c_1' n^{2-\alpha} \xrightarrow{d} \eta,$$

where η is a non-degenerate random variable.

The contrast between the two theorems is striking. The different regimes in Theorem 1.2 and the transition at the golden ratio $\alpha_0 = \frac{1}{2}(1+\sqrt{5})$ are no longer present in Theorem 1.1. Instead we observe that $|\ell_n - c_1 n^{2-\alpha}|$ converges to 0 or ∞, depending on whether $\alpha > \alpha_0$ or $\alpha < \alpha_0$. This can be understood as follows. There are two sources of randomness at work in L_n. On the one hand fluctuations arise from the randomness within the 'topology' of the tree, more precisely from the random variables $X_0 > X_1 > \cdots > X_{\tau_n}$. These fluctuations are of order $n^{1/\alpha+1-\alpha}$. On the other hand, the waiting times $T_{k+1} - T_k$ also contribute to the random fluctuations. Most substantial are those fluctuations arising close to the root of the tree, that is, the times $T_{k+1} - T_k$ with k close to τ_n. These fluctuations are of order 1. Now either of these fluctuations may dominate the other, depending on the sign of $1/\alpha + 1 - \alpha$. This is reflected in Theorem 1.2.

On the other hand the above-mentioned fluctuations of order one do not show up asymptotically in ℓ_n. This is simply due to the fact that branches close to the root of the tree are typically *internal* branches.

The fact that for $\alpha > \alpha_0 = 1.61$ the external length ℓ_n is decoupled from the total length L_n is illustrated by the plots in Figure 1.

As a by-product, we prove the following result on the decrease of the numbers X_k of branches and Y_k of external branches.

Theorem 1.3. *For $1 < \alpha < 2$ as $n \to \infty$,*

$$\max_{1 \leqslant j \leqslant \tau_n} \left| \frac{X_{\tau_n - j}}{n} - \frac{j}{\tau_n} \right| = o_P(1) \quad \textit{and} \quad \max_{1 \leqslant j \leqslant \tau_n} \left| \frac{Y_{\tau_n - j}}{n} - \left(\frac{j}{\tau_n}\right)^{\alpha} \right| = o_P(1).$$

This is illustrated in Figure 2.

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

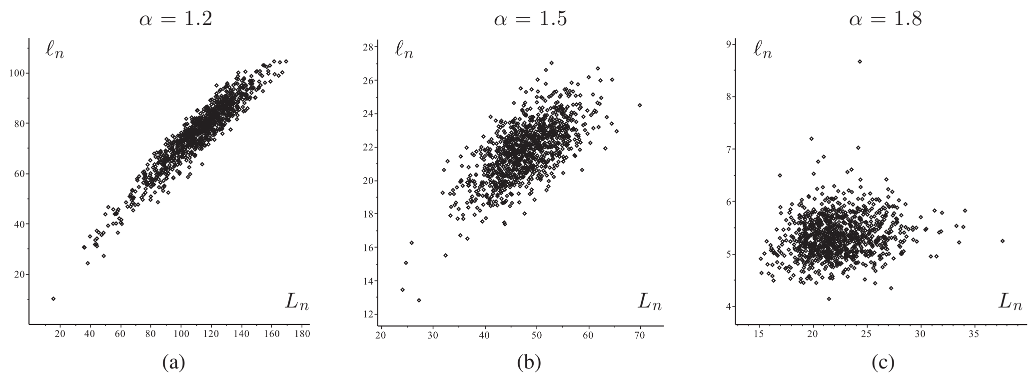

Figure 1. External versus total length. Each plot is based on 1000 coalescent realizations with $n = 1000$.

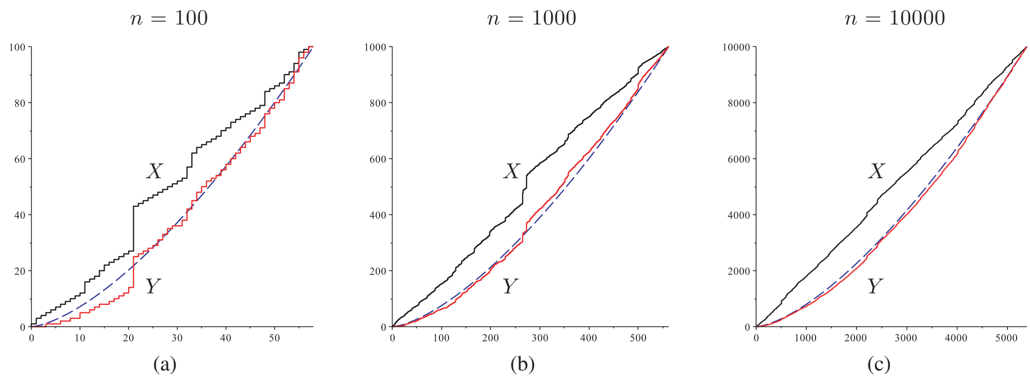

Figure 2. (Colour online) Simulation of $X = (X_{\tau_n - j})$ and $Y = (Y_{\tau_n - j})$ for $\alpha = 1.5$ and (a) $n = 100$, (b) $n = 1000$ and (c) $n = 10000$. The dashed line shows the curve $n(j/\tau_n)^\alpha$.

2. Heuristics and outline

The proof of Theorem 1.1 is given in the next section. It reveals an unexpected link between ℓ_n and the total number of mergers τ_n. In this section we explain the heuristics; this may also serve as a guide throughout the next section.

For $m \geqslant 2$, the rate at which a coalescence happens within m lineages is denoted by λ_m. We thus have

$$L_n \stackrel{d}{=} \sum_{k=0}^{\tau_n - 1} X_k \, W_k / \lambda_{X_k} \tag{2.1}$$

and

$$\ell_n \stackrel{d}{=} \sum_{k=0}^{\tau_n - 1} Y_k \, W_k / \lambda_{X_k}, \tag{2.2}$$

where $W_0, W_1, \dots$ are i.i.d. standard exponential distributed random variables. From Lemma 2.2 in [9] we have, for $m \to \infty$,

$$\lambda_m = \frac{1}{\alpha\Gamma(\alpha)} m^\alpha + O(m^{\alpha-1}). \tag{2.3}$$

Combining (2.1) and (2.3) suggests, writing informally,

$$L_n \approx \alpha\Gamma(\alpha) \sum_{k=0}^{\tau_n - 1} X_k^{1-\alpha}.$$

On the other hand, a coupling of (X_k) with a renewal process shows that

$$X_k \approx \gamma \cdot (\tau_n - k), \tag{2.4}$$

where $\gamma = 1/(\alpha - 1)$ (see Lemma 3.3 for the exact statement). This might suggest that

$$L_n \approx \alpha\Gamma(\alpha) \frac{\gamma^{1-\alpha}}{2-\alpha} \tau_n^{2-\alpha}. \tag{2.5}$$

Before discussing the (in-)validity of these asymptotics, let us turn to the external length. Since, given $X = (X_k)$, the random variables Y_k result from a hypergeometric sampling (see equation (3.12) below), we obtain (see (3.14))

$$\mathbf{E}[Y_k \mid X] = X_k \prod_{i=1}^{k} (1 - 1/X_i).$$

Replacing Y_k in (2.2) by this conditional expectation and again using (2.3), we are led to

$$\ell_n \approx \alpha\Gamma(\alpha) \sum_{k=0}^{\tau_n - 1} X_k^{1-\alpha} \prod_{i=1}^{k} (1 - 1/X_i). \tag{2.6}$$

Plugging in (2.4), using the asymptotics

$$\prod_{i=1}^{k} \left(1 - \frac{1/\gamma}{\tau_n - i}\right) \sim (\tau_n/(\tau_n - k))^{-1/\gamma}$$

and the cancellation of $(\tau_n - k)^{1/\gamma}$ against $(\tau_n - k)^{1-\alpha}$, we arrive at

$$\ell_n \approx \alpha\Gamma(\alpha)\gamma^{1-\alpha}\tau_n^{2-\alpha} \tag{2.7}$$

(see Lemma 3.2 for a precise statement). The asymptotics of τ_n are (see Lemma 3.1)

$$\tau_n \approx \frac{n}{\gamma} + n^{1/\alpha} \text{const } \varsigma.$$

Using a Taylor expansion, one hence obtains (see (3.1) for the exact statement)

$$\tau_n^{2-\alpha} \approx \left(\frac{n}{\gamma}\right)^{2-\alpha} + n^{1-\alpha+1/\alpha} \text{const}' \ \varsigma.$$

This shows that (as $n \to \infty$) the fluctuations of $\tau_n^{2-\alpha}$ decrease with n for $\alpha > \alpha_0$ and increase for $\alpha < \alpha_0$. Since (as discussed above) the fluctuations of L_n are always at least of order one, the approximation (2.5) cannot capture the fluctuations. In contrast, the

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

approximation (2.7) turns out to be sufficiently fine on the level of fluctuations as well. The reason for this is the asymptotic flatness of the function

$$(x_k) \mapsto \sum_{k=0}^{\tau-1} x_k^{1-\alpha} \prod_{i=1}^{k} (1 - 1/x_i)$$

in the point $(x_k) = (\gamma(\tau - k))$, as revealed in the proof of Lemma 3.2.

3. Proofs

Theorem 1.1 is a consequence of the following two lemmas.

The first lemma on the number of mergers has been independently obtained by Delmas, Dhersin and Siri-Jégousse [9], Gnedin and Yakubovich [14] and Iksanov and Möhle [16]. (Note that they use a different normalization of the limit law.) We let

$$\gamma = \frac{1}{\alpha - 1}.$$

Lemma 3.1. *Let* $1 < \alpha < 2$. *Then, for* $n \to \infty$,

$$\frac{\tau_n - n/\gamma}{n^{1/\alpha}} \xrightarrow{d} \frac{\varsigma}{\gamma^{1/\alpha+1}\Gamma(2-\alpha)^{1/\alpha}},$$

with ς *as described in Section 1.*

In short, the proof of this lemma goes along the following lines. The negative increments $X_{k-1} - X_k$ can be asymptotically replaced by i.i.d. positive random variables V_k, $k \geqslant 1$, with expectation γ and tail behaviour

$$\mathbf{P}(V_k \geqslant v) \sim \frac{1}{\Gamma(2-\alpha)} v^{-\alpha}.$$

From the theory of stable laws for $k \to \infty$,

$$\frac{X_0 - X_k - \gamma k}{k^{1/\alpha}} \sim \frac{V_1 + \cdots + V_k - \gamma k}{k^{1/\alpha}} \xrightarrow{d} \frac{-\varsigma}{\Gamma(2-\alpha)^{1/\alpha}}.$$

The choice $k = \tau_n$ gives

$$\frac{n - 1 - \gamma\tau_n}{\tau_n^{1/\alpha}} \xrightarrow{d} \frac{-\varsigma}{\Gamma(2-\alpha)^{1/\alpha}}$$

and thus $\tau_n \sim n/\gamma$. This implies the lemma. For details we refer to [9, 14, 16].

The asymptotic expansion in the next lemma discloses the link between ℓ_n and τ_n.

Lemma 3.2. *Let* $1 < \alpha < 2$ *and* $\varepsilon > 0$. *Then, for* $n \to \infty$,

$$\ell_n = \alpha\Gamma(\alpha)(\alpha - 1)^{\alpha-1}\tau_n^{2-\alpha} + O_P\left(n^{2/\alpha-\alpha+\varepsilon} + n^{3/2-\alpha+\varepsilon}\right).$$

Before proving this key lemma we show that Theorem 1.1 is a direct consequence of the two preceding lemmas.

Proof of Theorem 1.1. From Lemma 3.1 and a Taylor expansion, we obtain

$$\tau_n^{2-\alpha} = \left(\frac{n}{\gamma}\right)^{2-\alpha} + (2-\alpha)\left(\frac{n}{\gamma}\right)^{1-\alpha}\left(\tau_n - \frac{n}{\gamma}\right) + O_P(n^{-\alpha+2/\alpha}). \tag{3.1}$$

Using $1 < \alpha < 2$ and Lemma 3.2,

$$\ell_n = \alpha(\alpha-1)\Gamma(\alpha)n^{2-\alpha} + \alpha(2-\alpha)\Gamma(\alpha)n^{1-\alpha}\left(\tau_n - \frac{n}{\gamma}\right) + O_P(n^{1/\alpha+1-\alpha-\varepsilon}),$$

for some $\varepsilon > 0$. This implies the theorem. □

Now we prepare the proof of Lemma 3.2 by deriving two other lemmas.

Lemma 3.3. *For any* $\varepsilon > 0$,

$$\max_{1\leqslant j\leqslant \tau_n} j^{-1/\alpha-\varepsilon}|X_{\tau_n-j} - \gamma j| = O_P(1),$$

as $n \to \infty$.

Proof. In the proof of Lemma 3.3 we make heavy use of two coupling devices from [18] which reduce our problem to a problem concerning sums of independent and identically distributed random variables in the domain of attraction of an α-stable law. We start by briefly recalling the ideas and results from [18] that are relevant for our purpose.

We consider a random variable V with values in $\mathbb{N}$ and distribution given by

$$\mathbf{P}(V = k) = \frac{\alpha}{\Gamma(2-\alpha)}\frac{\Gamma(k+1-\alpha)}{\Gamma(k+2)}, \quad k \geqslant 1. \tag{3.2}$$

It follows that $\mathbf{E}(V) = 1/(\alpha-1) = \gamma$, and by Stirling's approximation

$$\mathbf{P}(V \geqslant k) \sim \frac{1}{\Gamma(2-\alpha)}k^{-\alpha}. \tag{3.3}$$

Kersting [18] shows that the random variable V stochastically dominates the jump size $U = X_1 - X_0$, that is, for all $k \geqslant 1$,

$$\mathbf{P}(U \geqslant k \mid X_0 = n) \leqslant \mathbf{P}(V \geqslant k). \tag{3.4}$$

The first coupling that we will use is between two random counting measures on $\{2, 3, \ldots\}$. These are defined as follows. The *coalescent's point process downwards from n* (or CPP(n)) is the point process on $\{2, 3, \ldots\}$ defined by

$$\mu_n = \sum_{i=0}^{\tau_n - 1} \delta_{X_i}.$$

A *coalescent's point process downwards from* ∞ (or CPP) is a point process μ on $\{2, 3, \ldots\}$ with the properties $\mu(\{2, 3, \ldots\}) = \infty$, $\mu(\{n\}) \in \{0, 1\}$ for any $n \geqslant 2$ a.s., and for any $n \geqslant 2$, given the event $\mu(\{n\}) = 1$ and $\mu_{[n+1,\infty)}$, the point process $\mu_{[2,n]}$ is a CPP(n) a.s. Here we denoted by μ_I for an interval I the point process given by $\mu_I(B) = \mu(B \cap I)$, $B \subset \{2, 3, \ldots\}$. Note that CPP($n$) can be defined for all coalescents and that CPP can be defined for all coalescents that come down from infinity.

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

Observe that the points of μ_n and those of μ are exactly the values of the block counting processes of the n-Beta$(2-\alpha,\alpha)$-coalescent and those of the Beta$(2-\alpha,\alpha)$-coalescent starting with infinitely many leaves, respectively, evaluated at their jump times (it is known that the Beta$(2-\alpha,\alpha)$-coalescent comes down from infinity: see [23]).

Let $\cdots > S_2 > S_1 \geqslant 2$ denote the points of the process μ. Thus

$$\mu = \sum_{i\geqslant 1} \delta_{S_i}.$$

It is shown in [18] that one can couple μ and μ_n on a common probability space in such a way that $X_{\tau_n - j} = S_j$ for all $j \leqslant \tau_n - \eta_n$, where the random variables η_n are stochastically bounded uniformly in n: see Lemma 6 therein.

In the second coupling from [18] the process μ is coupled to a stationary renewal process v with points $\cdots > R_2 > R_1 \geqslant 2$ such that for $j \geqslant 1$ the increments $R_{j+1} - R_j$ are independent and identically distributed with

$$R_{j+1} - R_j \stackrel{d}{=} V,$$

with V defined in (3.2). In order to ensure stationarity, R_1 is considered such that

$$\mathbf{P}(R_1 = r) = \frac{\mathbf{P}(V \geqslant r-1)}{\mathbf{E}(V)}, \quad r = 2,3,\ldots.$$

Therefore we can write $R_j = R_1 + V_2 + \cdots + V_j$, where $V_2, V_3 \ldots$ are independent copies of the random variable V. Using a Borel–Cantelli argument, Lemma 9 in [18] shows that, for any $\beta \in \mathbb{R}$ and any $\varepsilon > 0$,

$$\sum_{k=2}^{n} k^{-\beta}(V_k - \gamma) = \vartheta_n + o(n^{1/\alpha - \beta + \varepsilon}) \text{ a.s.,} \tag{3.5}$$

where ϑ_n is a.s. convergent.

Now let

$$\rho(n) = \max\{j \geqslant 0 : R_j \leqslant n\}$$

(with the convention $R_0 = 1$) and observe that since $V_2, V_3, \ldots$ are i.i.d. with expectation γ, we have

$$\rho(n) \sim \gamma^{-1} n \text{ a.s.} \tag{3.6}$$

Using (3.5) for $\beta = 0$, we obtain

$$\begin{aligned}\max_{\rho(2^{r-1}) \leqslant j \leqslant \rho(2^r)} j^{-1/\alpha - \varepsilon}|R_j - \gamma j| &\leqslant \rho(2^{r-1})^{-1/\alpha-\varepsilon} \max_{\rho(2^{r-1}) \leqslant j \leqslant \rho(2^r)} \left| \sum_{k=2}^{j} (V_k - \gamma) + R_1 - \gamma \right| \\ &\leqslant C \cdot \rho(2^{r-1})^{-1/\alpha - \varepsilon} \cdot \rho(2^r)^{1/\alpha + \varepsilon/2},\end{aligned}$$

where C is a finite random variable. Now, using (3.6) we obtain

$$\max_{\rho(2^{r-1}) \leqslant j \leqslant \rho(2^r)} j^{-1/\alpha - \varepsilon}|R_j - \gamma j| = O(2^{-\varepsilon r/2}) \text{ a.s.,} \tag{3.7}$$

where here and below, for a sequence of random variables Z_r and a sequence of numbers $g(r)$, the statement

$$Z_r = O(g(r)) \text{ a.s.}$$

means the existence of a finite random variable C such that $Z_r \leqslant Cg(r)$ a.s. for all r.

We are now ready to prove Lemma 3.3. First note that due to monotonicity of (X_k), for any $\eta \in \mathbb{N}$,

$$\max_{1\leqslant j\leqslant \tau_n} j^{-1/\alpha-\varepsilon}|X_{\tau_n-j}-\gamma j| \leqslant \max_{1\leqslant j\leqslant \tau_n-\eta} j^{-1/\alpha-\varepsilon}|X_{\tau_n-j}-\gamma j| + X_0 - X_\eta + \gamma\eta.$$

Using (3.4) and (3.3), it is easy to see that $X_0 - X_\eta$ is stochastically bounded uniformly in n. Therefore it is sufficient to prove the claim for the quantity

$$\max_{1\leqslant j\leqslant \tau_n-\eta_n} j^{-1/\alpha-\varepsilon}|X_{\tau_n-j}-\gamma j|,$$

where η_n is any sequence of $\mathbb{N}$-valued random variables that is stochastically bounded.

Now using the coupling between the point processes μ_n and μ, it follows that it is sufficient to estimate

$$\max_{1\leqslant j\leqslant \sigma(n)} j^{-1/\alpha-\varepsilon}|S_j-\gamma j|,$$

with

$$\sigma(n) = \max\{j \geqslant 0 : S_j \leqslant n\},$$

where we put $S_0 = 1$.

We shall consider

$$M_r = \max_{\sigma(2^{r-1})<j\leqslant\sigma(2^r)} j^{-1/\alpha-\varepsilon}|S_j-\gamma j|$$

for $r \geqslant 2$.

We now use the coupling of μ to the stationary point process ν. Let us define (following [18]) for $r \geqslant 2$ the following random variables:

$$N_r := \sigma(2^r) - \sigma(2^{r-1}), \quad N_r' := \rho(2^r) - \rho(2^{r-1})$$

and

$$D_r := \max_{0\leqslant i\leqslant N_r\wedge N_r'} |S_{\sigma(2^r)-i} - R_{\rho(2^r)-i}|.$$

Now equation (24) and Lemma 8 in [18] say that

$$|N_r - N_r'| \leqslant D_r$$

and

$$\mathbf{P}(D_r > t) \leqslant ct^{1-\alpha}, \quad t > 0,$$

for some $c > 0$. Hence it follows by the Borel–Cantelli lemma that

$$D_r = O(r^{\gamma+\varepsilon}) \text{ a.s.}, \tag{3.8}$$

and therefore, from $|\sigma(2^r) - \rho(2^r)| \leqslant D_r + \cdots + D_1 + 1$,

$$|\sigma(2^r) - \rho(2^r)| = O(r^{\gamma+1+\varepsilon}) \text{ a.s.} \tag{3.9}$$

for any $\varepsilon > 0$. Equation (3.6) implies

$$\sigma(2^r) \sim \gamma^{-1} 2^r \text{ a.s.} \tag{3.10}$$

for $r \to \infty$. By the triangle inequality (and reparametrizing by $j = \sigma(2^r) - i$)),

$$\begin{aligned}\sigma(2^{r-1})^{1/\alpha+\varepsilon} M_r &\leqslant \max_{\sigma(2^{r-1})<j\leqslant\sigma(2^r)} |S_j - \gamma j| \\ &\leqslant \max_{0\leqslant i<N_r} |S_{\sigma(2^r)-i} - R_{(\rho(2^r)-i)\vee\rho(2^{r-1})}| \\ &\quad + \max_{\rho(2^{r-1})\leqslant j\leqslant\rho(2^r)} |R_j - \gamma j| \\ &\quad + \gamma \max_{0\leqslant i<N_r} |(\rho(2^r) - i) \vee \rho(2^{r-1}) - (\sigma(2^r) - i)|.\end{aligned}$$

If $N_r' < N_r$, the first maximum on the right-hand side is not attained for $i > N_r'$, because then

$$|S_{\sigma(2^r)-i} - R_{(\rho(2^r)-i)\vee\rho(2^{r-1})}| = S_{\sigma(2^r)-i} - R_{\rho(2^{r-1})}$$

is decreasing for $i \geqslant N_r'$. Also, the third maximum is attained either for $i = 0$ or $i = N_r$. This implies

$$\begin{aligned}\sigma(2^{r-1})^{1/\alpha+\varepsilon} M_r \leqslant &\max_{0\leqslant i\leqslant(N_r-1)\wedge N_r'} |S_{\sigma(2^r)-i} - R_{\rho(2^r)-i}| \\ &+ \gamma|\sigma(2^r) - \rho(2^r)| + \gamma|\sigma(2^{r-1}) - \rho(2^{r-1})| \\ &+ \max_{\rho(2^{r-1})\leqslant j\leqslant\rho(2^r)} |R_j - \gamma j|.\end{aligned}$$

From (3.8), (3.9) and (3.10) for ε sufficiently small,

$$\begin{aligned}M_r &= \sigma(2^{r-1})^{-1/\alpha-\varepsilon} \max_{\rho(2^{r-1})\leqslant j\leqslant\rho(2^r)} |R_j - \gamma j| + O(2^{-r/\alpha} r^{1+\gamma+\varepsilon}) \\ &\leqslant 2 \max_{\rho(2^{r-1})\leqslant j\leqslant\rho(2^r)} j^{-1/\alpha-\varepsilon} |R_j - \gamma j| + O(2^{-\varepsilon r}).\end{aligned}$$

Using (3.7), we infer

$$M_r = O(2^{-\varepsilon r/2}) \text{ a.s.}$$

Now for $n \in \mathbb{N}$, choosing $s \in \mathbb{N}$ such that $2^{s-1} < n \leqslant 2^s$, since $\sigma(n) \leqslant n$

$$\max_{1\leqslant j\leqslant\sigma(n)} j^{-1/\alpha-\varepsilon} |S_j - \gamma j| \leqslant \sum_{r=1}^{s} M_r = O\left(\sum_{r=1}^{s} 2^{-\varepsilon r/2}\right) = O(1) \text{ a.s.}$$

This gives the claim. □

Now observe the following. Not only is $X = (X_0, \ldots, X_{\tau_n})$ a Markov chain, but given X, also the sequence $Y = (Y_0, \ldots, Y_{\tau_n})$. More precisely, let U_k be the number of particles collapsing to one particle in the kth merging event, and thus

$$X_k = X_{k-1} - U_k + 1. \tag{3.11}$$

For $k > \tau_n$ we have $U_k = 1$, since by definition $1 = X_{\tau_n} = X_{\tau_{n+1}} = \cdots$. The numbers of external branches, $n = Y_0 > \cdots > Y_{\tau_n} = 0$, follow the recursion

$$Y_k = Y_{k-1} - H_k \tag{3.12}$$

with

$$(H_k \mid X, Y_0, \ldots, Y_{k-1}) \stackrel{d}{=} \mathrm{Hyp}(X_{k-1}, Y_{k-1}, U_k),$$

where $\mathrm{Hyp}(N, M, v)$ denotes the hypergeometric distribution with parameters N, M, v. (These formulas also hold for $k > \tau_n$.) Thus

$$\mathbf{E}[Y_k \mid X, Y_0, \ldots, Y_{k-1}] = Y_{k-1} - U_k \frac{Y_{k-1}}{X_{k-1}} = \frac{X_k - 1}{X_{k-1}} Y_{k-1}$$

and

$$\frac{\mathbf{E}[Y_k \mid X]}{X_k} = \frac{X_k - 1}{X_k X_{k-1}} \mathbf{E}[Y_{k-1} \mid X] = \frac{\mathbf{E}[Y_{k-1} \mid X]}{X_{k-1}} \left(1 - \frac{1}{X_k}\right). \tag{3.13}$$

Iterating this formula and observing that $\mathbf{E}[Y_0 \mid X] = Y_0 = X_0$, we obtain

$$\mathbf{E}[Y_k \mid X] = X_k \Pi_0^k, \tag{3.14}$$

where

$$\Pi_j^k := \prod_{i=j+1}^{k} \left(1 - \frac{1}{X_i}\right), \quad 0 \leqslant j \leqslant k,$$

and in particular, $\Pi_k^k = 1$.

Next we study the auxiliary quantity Π_j^k.

Lemma 3.4. *Let* $\varepsilon > 0$. *Then, for* $0 \leqslant j \leqslant k < \tau_n$ *as* $n \to \infty$,

$$\Pi_j^k = \left(\frac{\tau_n - k}{\tau_n - j}\right)^{1/\gamma} \left(1 + \sum_{i=j+1}^{k} \frac{X_i - \gamma(\tau_n - i)}{\gamma^2(\tau_n - i)^2} + O_P\left((\tau_n - k)^{2/\alpha - 2 + \varepsilon}\right)\right) \tag{3.15}$$

and

$$\Pi_j^k = \left(\frac{\tau_n - k}{\tau_n - j}\right)^{1/\gamma} \left(1 + O_P\left((\tau_n - k)^{1/\alpha - 1 + \varepsilon}\right)\right), \tag{3.16}$$

where the $O_P(\cdot)$ *terms hold uniformly for all* $0 \leqslant j \leqslant k < \tau_n$.

Proof. Using a Taylor expansion of $\log(1 + h)$, we obtain that on the event $\{k < \tau_n\}$

$$\Pi_j^k = \exp\left(-\sum_{i=j+1}^{k} \frac{1}{X_i} + O\left(\sum_{i=j+1}^{k} \frac{1}{X_i^2}\right)\right)$$

with the error term uniformly bounded in n and $0 \leqslant j \leqslant k < \tau_n$. In view of a Taylor expansion together with Lemma 3.3 and the fact that

$$\sum_{m=r}^{s-1} \frac{1}{m} = \log \frac{s}{r} + O\left(\frac{1}{r}\right),$$

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

we obtain

$$\begin{aligned}\sum_{i=j+1}^{k}\frac{1}{X_i}&=\sum_{i=j+1}^{k}\left(\frac{1}{\gamma(\tau_n-i)}-\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)^2}+O_P\left(\frac{(X_i-\gamma(\tau_n-i))^2}{(\tau_n-i)^3}\right)\right)\\&=\frac{1}{\gamma}\log\frac{\tau_n-j}{\tau_n-k}-\sum_{i=j+1}^{k}\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)^2}+O_P\left((\tau_n-k)^{2/\alpha-2+\varepsilon}\right),\end{aligned}$$

where the $O_P(\cdot)$ holds uniformly for all $0\leqslant j\leqslant k<\tau_n$. Also, from Lemma 3.3,

$$\sum_{i=j+1}^{k}\frac{1}{X_i^2}=O_P\left(\sum_{i=j+1}^{k}\frac{1}{(\tau_n-i)^2}\right)=O_P\left((\tau_n-k)^{-1}\right),$$

and therefore, for $0\leqslant j\leqslant k<\tau_n$, since $1<\alpha<2$,

$$\Pi_j^k=\left(\frac{\tau_n-k}{\tau_n-j}\right)^{1/\gamma}\exp\left(\sum_{i=j+1}^{k}\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)^2}+O_P\left((\tau_n-k)^{2/\alpha-2+\varepsilon}\right)\right).$$

This shows (3.15). Finally, from Lemma 3.3, since $\alpha>1$,

$$\sum_{i=j+1}^{k}\frac{X_i-\gamma(\tau_n-i)}{(\tau_n-i)^2}=O_P\left(\sum_{i=j+1}^{k}(\tau_n-i)^{1/\alpha-2+\varepsilon}\right)=O_P\left((\tau_n-k)^{1/\alpha-1+\varepsilon}\right),$$

and hence (3.16) also follows. □

Proof of Lemma 3.2. We shall approximate ℓ_n step by step. First let

$$\mathfrak{l}_n=\sum_{k=0}^{\tau_n-1}\frac{\mathbf{E}[Y_k\mid X]}{X_k^\alpha}=\sum_{k=0}^{\tau_n-1}X_k^{1-\alpha}\Pi_0^k.$$

To obtain an approximation to this expression we use the following expansion, implied by Lemma 3.3. Uniformly for $k<\tau_n$,

$$\begin{aligned}X_k^{1-\alpha}&=(\gamma(\tau_n-k))^{1-\alpha}+(1-\alpha)(\gamma(\tau_n-k))^{-\alpha}(X_k-\gamma(\tau_n-k))\\&\quad+O_P\left((\tau_n-k)^{-\alpha-1}(X_k-\gamma(\tau_n-k))^2\right)\end{aligned}$$

and thus

$$X_k^{1-\alpha}=(\gamma(\tau_n-k))^{1-\alpha}\left(1-\frac{X_k-\gamma(\tau_n-k)}{\gamma^2(\tau_n-k)}+O_P\left((\tau_n-k)^{2/\alpha-2+\varepsilon}\right)\right).$$

Together with Lemma 3.4, we obtain (taking the first summand $n^{1-\alpha}$ apart)

$$\begin{aligned}\mathfrak{l}_n=n^{1-\alpha}+(\gamma\tau_n)^{1-\alpha}\sum_{k=1}^{\tau_n-1}\Bigg(1&+\sum_{i=1}^{k}\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)^2}\\&-\frac{X_k-\gamma(\tau_n-k)}{\gamma^2(\tau_n-k)}+O_P\left((\tau_n-k)^{2/\alpha-2+\varepsilon}\right)\Bigg).\end{aligned}$$

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

Now

$$\sum_{k=1}^{\tau_n-1}\sum_{i=1}^{k}\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)^2}=\sum_{i=1}^{\tau_n-1}\frac{X_i-\gamma(\tau_n-i)}{\gamma^2(\tau_n-i)},$$

such that (surprisingly) matters simplify to

$$\mathfrak{l}_n=n^{1-\alpha}+(\gamma\tau_n)^{1-\alpha}\big(\tau_n+O_P(\tau_n^{2/\alpha-1+\varepsilon})\big)$$

or, using Lemma 3.1,

$$\mathfrak{l}_n=\gamma^{1-\alpha}\tau_n^{2-\alpha}+O_P(n^{2/\alpha-\alpha+\varepsilon}). \tag{3.17}$$

Next we consider

$$\mathfrak{l}_n'=\sum_{k=0}^{\tau_n-1}\frac{Y_k}{X_k^{\alpha}}.$$

Putting $G_k=Y_k-\mathbf{E}[Y_k\mid X]$, we obtain, since $G_0=0$,

$$\mathfrak{l}_n'-\mathfrak{l}_n=\sum_{k=1}^{\tau_n-1}\frac{G_k}{X_k^{\alpha}}.$$

Now, from (3.12) and (3.11),

$$Y_k=Y_{k-1}\frac{X_k-1}{X_{k-1}}-\tilde{H}_k,\quad \text{with } \tilde{H}_k=H_k-U_k\frac{Y_{k-1}}{X_{k-1}},$$

and thus, from (3.13),

$$G_k=G_{k-1}\frac{X_k-1}{X_{k-1}}-\tilde{H}_k$$

or

$$\frac{G_k}{X_k\Pi_0^k}=\frac{G_{k-1}}{X_{k-1}\Pi_0^{k-1}}-\frac{\tilde{H}_k}{X_k\Pi_0^k}.$$

Iterating this relation, we obtain because of $G_0=0$

$$G_k=-X_k\Pi_0^k\sum_{i=1}^{k}\frac{\tilde{H}_i}{X_i\Pi_0^i}, \tag{3.18}$$

and thus

$$\mathfrak{l}_n-\mathfrak{l}_n'=\sum_{k=1}^{\tau_n-1}\frac{\Pi_0^k}{X_k^{\alpha-1}}\sum_{i=1}^{k}\frac{\tilde{H}_i}{X_i\Pi_0^i}=\sum_{i=1}^{\tau_n-1}\frac{\tilde{H}_i}{X_i}\sum_{k=i}^{\tau_n-1}\frac{\Pi_i^k}{X_k^{\alpha-1}}.$$

Now note that given $X,Y_0,\dots,Y_{i-1}$ the random variable $\tilde{H}_i$ has a hypergeometric distribution centred at expectation. In fact $(\tilde{H}_i)$ forms a sequence of martingale differences with respect to the filtration generated by $(X,Y_0,Y_1,\dots,Y_i)$, $i\geqslant 0$, and therefore $\tilde{H}_1,\tilde{H}_2,\dots$ are uncorrelated given X. Also, from the formula for the variance of a hypergeometric distribution,

$$\mathbf{E}[\tilde{H}_i^2\mid X,Y_0,\dots,Y_{i-1}]\leqslant U_i\frac{Y_{i-1}}{X_{i-1}}. \tag{3.19}$$

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

This entails

$$\mathbf{E}[(\ell_n' - \ell_n)^2 \mid X] \leqslant \sum_{i=1}^{\tau_n - 1} \frac{1}{X_i^2} \frac{U_i \mathbf{E}[Y_{i-1} \mid X]}{X_{i-1}} \left(\sum_{k=i}^{\tau_n-1} \frac{\Pi_i^k}{X_k^{\alpha-1}} \right)^2$$
$$= \sum_{i=1}^{\tau_n-1} U_i \frac{\Pi_0^{i-1}}{X_i^2} \left(\sum_{k=i}^{\tau_n-1} \frac{\Pi_i^k}{X_k^{\alpha-1}} \right)^2.$$

Using a Taylor expansion similar to the one in the proof of Lemma 3.4 and applying Lemma 3.3, we see that X_i^{-1} is of order $(\tau_n - i)^{-1}$ uniformly in $0 \leqslant i < \tau_n$. This together with Lemma 3.4 implies

$$\mathbf{E}[(\ell_n' - \ell_n)^2 \mid X]$$
$$= O_P\left(\sum_{i=1}^{\tau_n-1} U_i \left(\frac{\tau_n - i + 1}{\tau_n} \right)^{1/\gamma} \frac{1}{(\tau_n - i)^2} \left(\sum_{k=i}^{\tau_n-1} \left(\frac{\tau_n - k}{\tau_n - i} \right)^{1/\gamma} \frac{1}{(\tau_n - k)^{\alpha-1}} \right)^2 \right).$$

Since $1/\gamma = \alpha - 1$, this boils down to

$$\mathbf{E}[(\ell_n' - \ell_n)^2 \mid X] = O_P\left(\tau_n^{1-\alpha} \sum_{i=1}^{\tau_n-1} U_i (\tau_n - i)^{1-\alpha} \right).$$

Again from Lemma 3.3, X_{i-1} is of order $\tau_n - i$ uniformly in $0 \leqslant i < \tau_n$. Also, $U_i = X_{i-1} - X_i + 1 \leqslant 2(X_{i-1} - X_i)$, and therefore

$$\mathbf{E}[(\ell_n' - \ell_n)^2 \mid X] = O_P\left(\tau_n^{1-\alpha} \sum_{i=1}^{\tau_n-1} X_{i-1}^{1-\alpha} (X_{i-1} - X_i) \right).$$

Moreover,

$$\sum_{i=1}^{\tau_n-1} X_{i-1}^{1-\alpha}(X_{i-1} - X_i) \leqslant \int_{X_{\tau_n}}^{X_0} x^{1-\alpha}\, dx \leqslant \frac{1}{2-\alpha} X_0^{2-\alpha}.$$

Since $X_0 = n$, we arrive at $\mathbf{E}[(\ell_n' - \ell_n)^2 \mid X] = O_P(n^{3-2\alpha})$, which by Cauchy–Schwarz gives

$$\mathbf{E}[|\ell_n' - \ell_n| \mid X] = O_P(n^{3/2-\alpha}).$$

Markov's inequality and dominated convergence now imply that

$$\begin{aligned} &\lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}(|\ell_n' - \ell_n| > c\, n^{3/2-\alpha}) \\ &\quad = \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{E}\big[\mathbf{P}\big(|\ell_n' - \ell_n| > c\, n^{3/2-\alpha} \mid X\big)\big] \qquad (3.20) \\ &\quad \leqslant \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{E}\left[1 \wedge \left(\frac{1}{c} n^{-3/2+\alpha} \mathbf{E}[|\ell_n' - \ell_n| \mid X]\right)\right] = 0, \end{aligned}$$

and thus $\ell_n' - \ell_n = O_P(n^{3/2-\alpha})$. Combining this with (3.17), we obtain

$$\ell_n' = \gamma^{1-\alpha} \tau_n^{2-\alpha} + O_P(n^{2/\alpha - \alpha + \varepsilon} + n^{3/2-\alpha}) \qquad (3.21)$$

for all $\varepsilon > 0$.

Next we prepare to incorporate the waiting times between mergers into our expression. Let $W_1, W_2, \ldots$ be independent exponential random variables with mean 1, also

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

independent from (X, Y). Let

$$\ell''_n = \sum_{k=0}^{\tau_n - 1} \frac{Y_k}{X_k^\alpha} W_k.$$

Then

$$\mathbf{E}[(\ell''_n - \ell'_n)^2 \mid X, Y] = \sum_{k=0}^{\tau_n-1} \frac{Y_k^2}{X_k^{2\alpha}} \leqslant \sum_{k=0}^{\tau_n-1} \frac{Y_k}{X_k^{2\alpha-1}}$$

and therefore

$$\mathbf{E}[(\ell''_n - \ell'_n)^2 \mid X] \leqslant \sum_{k=0}^{\tau_n-1} \frac{\mathbf{E}[Y_k \mid X]}{X_k^{2\alpha-1}} = \sum_{k=0}^{\tau_n-1} X_k^{2-2\alpha} \Pi_0^k.$$

From Lemmas 3.3, 3.4 and $\alpha < 2$ it follows that

$$\mathbf{E}[(\ell''_n - \ell'_n)^2 \mid X] = O_P\left(\tau_n^{1-\alpha} \sum_{k=0}^{\tau_n-1} (\tau_n - k)^{1-\alpha}\right) = O_P(n^{3-2\alpha}).$$

As in (3.20), we may conclude $\ell''_n - \ell'_n = O_P(n^{3/2-\alpha})$ and consequently, from (3.21),

$$\ell''_n = \gamma^{1-\alpha} \tau_n^{2-\alpha} + O_P(n^{2/\alpha - \alpha + \varepsilon} + n^{3/2-\alpha}). \tag{3.22}$$

In a last step we add the coalescence rate λ_M into our formulas. From Lemmas 3.3 and 3.4,

$$\mathbf{E}\left[\frac{Y_k}{X_k^{\alpha+1}} W_k \mid X\right] = \frac{\Pi_0^k}{X_k^\alpha} = O_P(\tau_n^{1-\alpha}(\tau_n - k)^{-1}), \tag{3.23}$$

where the $O_P(\cdot)$ holds uniformly for all $0 \leqslant k \leqslant \tau_n$. Therefore, taking the sum in (3.23) and arguing as in (3.20),

$$\sum_{k=0}^{\tau_n-1} \frac{Y_k}{X_k^{\alpha+1}} W_k = O_P\left(\tau_n^{1-\alpha} \sum_{k=0}^{\tau_n-1} (\tau_n - k)^{-1}\right) = O_P(\tau_n^{1-\alpha} \log \tau_n).$$

Combined with (2.3) and (3.22), this gives

$$\sum_{k=0}^{\tau_n-1} \frac{Y_k}{\lambda_{X_k}} W_k = \alpha\Gamma(\alpha)\gamma^{1-\alpha}\tau_n^{2-\alpha} + O_P(n^{2/\alpha-\alpha+\varepsilon} + n^{3/2-\alpha+\varepsilon}).$$

Now the left-hand quantity is equal to ℓ_n in distribution. This proves Lemma 3.2. □

Proof of Theorem 1.3. The first statement is a direct consequence of Lemma 3.3 with $\varepsilon = 1 - 1/\alpha$ and Lemma 3.1. For the proof of the second one we first note that it is sufficient to show that for $0 < c < 1$ and $k_n \sim c\tau_n$

$$\frac{Y_{k_n}}{n} \to (1-c)^\alpha \tag{3.24}$$

in probability as $n \to \infty$. Indeed, from (3.24) it follows for all natural numbers m that

$$M_m(n) := \max_{i=0,1,\ldots,m} \left| \frac{Y_{\lfloor (i/m)\tau_n \rfloor}}{n} - \left(1 - \frac{i}{m}\right)^\alpha \right| = o_P(1) \tag{3.25}$$

https://doi.org/10.1017/S0963548314000297 Published online by Cambridge University Press

as $n \to \infty$. Now for $c \in (0,1)$ let $i \in \{1,\dots,m\}$ obey $(i-1)/m \leqslant c < i/m$. Then, due to the monotonicity of $k \mapsto Y_k$, the random variable $Y_{\lfloor c\tau_n \rfloor}$ is sandwiched between $Y_{\lfloor (i/m)\tau_n \rfloor}$ and $Y_{\lfloor ((i-1)/m)\tau_n \rfloor}$, and we obtain from (3.25)

$$\sup_{c\in(0,1)}\left|\frac{Y_{\lfloor c\tau_n\rfloor}}{n}-(1-c)^\alpha\right| \leqslant 2M_m(n)+2\max_{i=0,1,\dots,m}\left|\left(1-\frac{i-1}{m}\right)^\alpha-\left(1-\frac{i}{m}\right)^\alpha\right|$$
$$\leqslant o_P(1)+2\frac{\alpha}{m}.$$

We now proceed to show (3.24). From (3.14), (3.16) and the first statement of the theorem,

$$\frac{\mathbf{E}[Y_{k_n}\mid X]}{n}=\frac{X_{k_n}}{n}\cdot\Pi_0^{k_n}\to(1-c)\cdot(1-c)^{\alpha-1}$$

in probability as $n\to\infty$. Thus it remains to prove that $G_{k_n}=Y_{k_n}-\mathbf{E}[Y_{k_n}\mid X]$ is $o_P(n)$. Now from (3.18), (3.19) and the explanation given there,

$$\mathbf{E}[G_k^2\mid X]\leqslant(X_k\Pi_0^k)^2\sum_{i=1}^k\frac{1}{(X_i\Pi_0^i)^2}U_i\Pi_0^{i-1}.$$

As in the proof of Lemma 3.2 (see the arguments following the inequality (3.19)), we mutually replace X_i and τ_n-i as well as Π_0^i and $(\tau_n-i)^{1/\gamma}/\tau_n^{1/\gamma}$ to obtain uniformly in k

$$\mathbf{E}[G_k^2\mid X]=O_P\left(\frac{(\tau_n-k)^{2\alpha}}{\tau_n^{\alpha-1}}\sum_{i=1}^k(\tau_n-i)^{-\alpha-1}U_i\right)$$
$$=O_P\left(\frac{(\tau_n-k)^{2\alpha}}{\tau_n^{\alpha-1}}\sum_{i=1}^k X_i^{-\alpha-1}(X_{i-1}-X_i)\right),$$

and uniformly in k

$$\mathbf{E}[G_k^2\mid X]=O_P\left(\frac{(\tau_n-k)^{2\alpha}}{\tau_n^{\alpha-1}}X_k^{-\alpha}\right)=O_P(\tau_n)=O_P(n).$$

By the same reasoning as in (3.20), this implies that G_{k_n} is $O_P(\sqrt{n})$, which gives our claim. □

Acknowledgement

We are grateful to the referees for valuable hints which led to an improvement of the paper.

References

[1] Berestycki, N. (2009) Recent progress in coalescent theory. *Enasios Mathemáticos* **16** 1–193.

[2] Berestycki, J., Berestycki, N. and Limic, V. (2012) Asymptotic sampling formulae for Lambda-coalescents. To appear in *Ann. Inst. H. Poincaré* arXiv:1201.6512

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

[4] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2008) Small time properties of Beta-coalescents. *Ann. Inst. H. Poincaré* **44** 214–238.

[5] Birkner, M. and Blath, J. (2008) Computing likelihoods for coalescents with multiple collisions in the infinitely-many-sites model. *J. Math. Biology* **57** 435–465.
[6] Birkner, M., Blath, J., Capaldo, M., Etheridge, A., Möhle, M., Schweinsberg, J. and Wakolbinger, A. (2005) Alpha-stable branching and Beta-coalescents. *Electron. J. Probab.* **10** 303–325.
[7] Bolthausen, E. and Sznitman, A.-S. (1998) On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197** 247–276.
[8] Boom, E. G., Boulding J. D. G. and Beckenbach A. T. (1994) Mitochondrial DNA variation in introduced populations of Pacific oyster, *Crassostrea gigas*, in British Columbia. *Canad. J. Fish. Aquat. Sci.* **51** 1608–1614.
[9] Delmas, J.-F., Dhersin, J.-S. and Siri-Jégousse, A. (2008) Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.* **18** 997–1025.
[10] Dhersin, J.-S. and Yuan, L. (2012) Asympotic behavior of the total length of external branches for Beta-coalescents. arXiv:1202.5859
[11] Drmota, M., Iksanov, A., Möhle, M. and Rösler, U. (2007) Asymptotic results about the total branch length of the Bolthausen–Sznitman coalescent. *Stoch. Proc. Appl.* **117** 1404–1421.
[12] Durrett, R. (2008) *Probability Models for DNA Sequence Evolution*, second edition, Springer.
[13] Eldon, B. and Wakeley, J. (2006) Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172** 2621–2633.
[14] Gnedin, A. and Yakubovich, Y. (2007) On the number of collisions in Λ-coalescents. *Electron. J. Probab.* **12** 1547–1567.
[15] Iksanov, A. and Möhle, M. (2007) A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.* **12** 28–35.
[16] Iksanov, A. and Möhle, M. (2008) On the number of jumps of random walks with a barrier. *Adv. Appl. Probab.* **40** 206–228.
[17] Janson, S. and Kersting, G. (2011) On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16** 2203–2218.
[18] Kersting, G. (2012) The asymptotic distribution of the length of Beta-coalescent trees. *Ann. Appl. Probab.* **22** 2086-2107.
[19] Kingman, J. F. C. (1982) The coalescent. *Stoch. Proc. Appl.* **13** 235–248.
[20] Möhle, M. (2010) Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent. *Stoch. Process. Appl.* **120** 2159–2173.
[21] Pitman, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* **27** 1870–1902.
[22] Sagitov, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** 1116–1125.
[23] Schweinsberg, J. (2000) A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Comm. Probab.* **5** 1–11.
[24] Steinrücken, M., Birkner, M. and Blath, J. (2013) Analysis of DNA sequence variation within marine species using Beta-coalescents. *Theoret. Popul. Biol.* **83** 20–29.
[25] Wakeley, J. (2008) *Coalescent Theory: An Introduction*, Roberts.
[26] Watterson, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theoret. Popul. Biol.* **7** 256–276.