

# The size of the last merger and time reversal in $\Lambda$ -coalescents

Götz Kersting<sup>a,1</sup>, Jason Schweinsberg<sup>b,2</sup> and Anton Wakolbinger<sup>a,1</sup>

<sup>a</sup>*Institut für Mathematik, Goethe Universität, Frankfurt am Main, Germany.*

*E-mail: kersting@math.uni-frankfurt.de; wakolbin@math.uni-frankfurt.de*

<sup>b</sup>*Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0112, USA.*

*E-mail: jschwein@math.ucsd.edu*

Received 3 January 2017; revised 16 May 2017; accepted 23 May 2017

**Abstract.** We consider the number of blocks involved in the last merger of a  $\Lambda$ -coalescent started with  $n$  blocks. We give conditions under which, as  $n \rightarrow \infty$ , the sequence of these random variables (a) is tight, (b) converges in distribution to a finite random variable or (c) converges to infinity in probability. Our conditions are optimal for  $\Lambda$ -coalescents that have a dust component. For general  $\Lambda$ , we relate the three cases to the existence, uniqueness and non-existence of invariant measures for the dynamics of the block-counting process, and in case (b) investigate the time-reversal of the block-counting process back from the time of the last merger.

**Résumé.** Nous considérons le nombre de blocs impliqués dans le dernier regroupement d'un  $\Lambda$ -coalescent issu de  $n$  blocs. Nous donnons des conditions sous lesquelles, quand  $n$  tend vers l'infini, la suite de variables aléatoires (a) est tendue (b) converge en loi vers une variable aléatoire finie ou (c) converge vers l'infini en probabilité. Nos conditions sont optimales pour les  $\Lambda$ -coalescents qui ont une composante de poussière. Pour un  $\Lambda$  général, nous associons ces trois cas à l'existence, l'unicité et la non-existence d'une mesure invariante pour la dynamique du processus de comptage des blocs. Dans le cas (b), nous étudions le retourné en temps du processus de comptage des blocs depuis de le temps de dernier regroupement.

*MSC:* Primary 60J27; secondary 60K05; 60G51

*Keywords:*  $\Lambda$ -coalescent; Block-counting process; Renewal theory; Subordinator

## 1. Introduction and main results

We consider coalescents with multiple mergers, also known as  $\Lambda$ -coalescents, which were introduced in 1999 by Pitman [12] and Sagitov [13]. If  $\Lambda$  is a finite measure on  $[0, 1]$ , then the  $\Lambda$ -coalescent started with  $n$  blocks is a continuous-time Markov chain  $(\Pi_n(t), t \geq 0)$  taking its values in the set of partitions of  $\{1, \dots, n\}$ . It has the property that whenever there are  $b$  blocks, each possible transition that involves merging  $k \geq 2$  of the blocks into a single block happens at rate

$$\lambda_{b,k} = \int_0^1 p^{k-2} (1-p)^{b-k} \Lambda(dp), \quad (1)$$

and these are the only possible transitions. One can also define the  $\Lambda$ -coalescent started with infinitely many blocks, which is a continuous-time Markov process  $(\Pi_\infty(t), t \geq 0)$  taking its values in the set of partitions of the positive

<sup>1</sup>Supported in part by DFG priority program 1590.

<sup>2</sup>Supported in part by NSF Grant DMS-1206195.

integers such that for all  $n$ , the restriction of  $(\Pi_\infty(t), t \geq 0)$  to the integers  $\{1, \dots, n\}$  has the same law as  $(\Pi_n(t), t \geq 0)$ .

Let  $N_n(t)$  be the number of blocks in the partition  $\Pi_n(t)$ . Denote by  $T_n = \inf\{t : N_n(t) = 1\}$  the time of the last merger. In this paper, we are interested in the distribution of

$$L_n := N_n(T_n-),$$

the number of blocks that coalesce during the last merger. The asymptotic behaviour of the distribution of  $L_n$  depends on how much mass the measure  $\Lambda$  has in the vicinity of point 1. Here it turns out to be decisive whether or not the finiteness condition

$$\int_0^1 |\log(1 - p)| \Lambda(dp) < \infty \tag{2}$$

is valid. We shall prove that (2) together with a logarithmic nonlattice property implies convergence on the sequence  $(L_n)$  in distribution. Without additional assumptions condition (2) entails tightness of  $(L_n)$ , but in general not convergence. In the presence of dust, (2) turns out to be necessary for tightness of  $(L_n)$ . When the  $\Lambda$ -coalescent comes down from infinity, which means that almost surely  $N_\infty(t) < \infty$  for all  $t > 0$ , we have  $T_\infty < \infty$  almost surely. See [14] for a necessary and sufficient condition for the  $\Lambda$ -coalescent to come down from infinity. In this case the distribution of  $L_n$  converges as  $n \rightarrow \infty$  to the distribution of  $N_\infty(T_\infty-)$ .

A second issue is the characterisation of the limit distribution of  $L_n$  in case of convergence by means of invariant measures  $\mu$ . Let

$$\rho_{ij} := \binom{i}{i-j+1} \lambda_{i,i-j+1}, \quad \rho_i := \sum_{j=1}^{i-1} \rho_{ij}, \quad 1 \leq j < i.$$

Then  $\rho_{ij}$  is the rate at which  $N_n$  jumps from state  $i$  to  $j$ , and  $\rho_i$  is the total rate of a jump from  $i$ . In particular note that  $\rho_{i1} = \lambda_{i,i}$ . We consider locally finite measures  $\mu = (\mu_i)_{i \geq 2}$  on  $\{2, 3, \dots\}$  which fulfill the equations

$$\sum_{j=i+1}^\infty \mu_j \rho_{ji} = \mu_i \rho_i, \quad i \geq 2, \quad \text{and} \quad \sum_{j=2}^\infty \mu_j \rho_{j1} = 1. \tag{3}$$

Note that for such measures  $\mu$  we have  $\mu_i > 0$  for all  $i \geq 2$ . The first property in (3) says that the measure  $\mu$  on  $\{2, 3, \dots\}$  is  $\rho$ -invariant: for each  $i \geq 2$  the flow of mass into the state  $i \geq 2$  equals the flow out of  $i$ . The second property says that the total flow out of the set  $\{2, 3, \dots\}$  equals one. We shall address questions of existence and uniqueness of solutions to (3) and shall in particular prove that in case of convergence of  $L_n$  the limiting distribution has weights  $\mu_i \rho_{i1}$ ,  $i \geq 2$ , with  $(\mu_i)_{i \geq 2}$  being the unique solution of (3). Moreover this representation of the limit will allow us to identify the time-reversal of the block-counting process.

Hénard [8] and Möhle [11] were able to calculate the limiting distribution for  $L_n$  when  $\Lambda$  is the beta distribution with parameters  $2 - \alpha$  and  $\alpha$  for  $0 < \alpha < 2$ . Note that this coalescent process comes down from infinity only when  $1 < \alpha < 2$ . Earlier, Goldschmidt and Martin [6] had calculated this distribution for the Bolthausen-Sznitman coalescent, which is the case  $\alpha = 1$ . Abraham and Delmas found this limit for  $\alpha = 1/2$  in [1], and for all  $\alpha \in (0, 1/2]$  in [2].

We are now going to present our main results. Throughout, we will assume that  $\Lambda$  is a nonzero, finite measure on  $[0, 1]$ . Theorem 1 concerns tightness.

**Theorem 1.** *Suppose that condition (2) is satisfied. Then the sequence  $(L_n)_{n \geq 1}$  is tight.*

Under an additional regularity condition, we are able to show that the distribution of the number of blocks involved in the last merger tends to a limit as  $n \rightarrow \infty$ . We call the measure  $\Lambda$  *log-nonlattice* if

$$\forall d > 0: \sum_{z=1}^\infty \Lambda(\{1 - e^{-zd}\}) < \Lambda((0, 1]).$$

**Theorem 2.** *Suppose (2) holds, and  $\Lambda$  is log-nonlattice. Then the sequence  $(L_n)_{n \geq 1}$  converges in distribution.*

In this theorem the log-nonlattice assumption cannot be completely avoided. Indeed we shall show below that when  $\Lambda$  has all its mass at one single point within  $(0, 1)$ , the sequence  $(L_n)_{n \geq 1}$ , though tight, does not converge in distribution. It is natural to conjecture that in the lattice case we always will experience such non-convergence.

The next theorem shows that condition (2) is necessary for tightness of the size of the last merger in the presence of dust.

**Theorem 3.** *Suppose*

$$\int_0^1 p^{-1} \Lambda(dp) < \infty \tag{4}$$

and hence in particular  $\Lambda(\{0\}) = 0$ . Also suppose

$$\int_0^1 |\log(1 - p)| \Lambda(dp) = \infty. \tag{5}$$

Then for all positive integers  $\ell$ , we have

$$\lim_{n \rightarrow \infty} P(L_n \leq \ell) = 0. \tag{6}$$

It was shown in [12] that (4) is the condition under which the  $\Lambda$ -coalescent has a dust component, which means that for all  $t > 0$ , the partition  $\Pi_\infty(t)$  contains singleton blocks almost surely. We can see from the statements of Theorems 1 and 3 that when  $\Lambda$  satisfies (4), the condition (5) is necessary and sufficient for (6) to hold. Therefore, the only case that remains open is the case when the  $\Lambda$ -coalescent fails to come down from infinity but there is no dust component. In that case, we expect that it is possible that (5) holds but (6) fails to hold.

The central tool for the proof of Theorem 3 is a uniform approximation of  $\log N_n(t)$  by the solution of an SDE driven by a subordinator, see Theorem 10 in Section 3 and its corollaries. These results can be seen as refinement and generalization of the subordinator approximation by Gnedin, Iksanov, and Marynych [5] in the presence of a dust component, see Remark 13 below.

Whenever the random variables  $L_n$  converge in distribution, it is natural to ask whether convergence in distribution holds for the block-counting processes  $N_n = (N_n(t))_{t \geq 0}$  as  $n \rightarrow \infty$  in any finite observation window around state 1. An appropriate description is by means of time-reversal. As a tool we use  $\rho$ -invariant measures satisfying equations (3). Existence and uniqueness of such measures are closely related to the asymptotic behaviour of the sequence of distributions of the last merger sizes  $L_n$ .

**Theorem 4.**

- (i) *If  $L_n \rightarrow \infty$  in probability as  $n \rightarrow \infty$ , then there is no solution to (3).*
- (ii) *If there is a probability measure  $\pi = (\pi_i)_{i \geq 2}$  on  $\{2, 3, \dots\}$  and a sequence of positive numbers  $\alpha_n, n \geq 1$ , not converging to 0, such that as  $n \rightarrow \infty$*

$$P(L_n = i) \sim \alpha_n \pi_i$$

*for all  $i \geq 2$ , then the measure  $\mu = (\mu_i)_{i \geq 2}$  given by  $\mu_i \rho_{i1} = \pi_i, i \geq 2$ , is the unique solution to (3). In particular, if the sequence  $(L_n)_{n \geq 1}$  converges in distribution to a finite random variable  $L_\infty$ , then*

$$P(L_\infty = i) = \mu_i \rho_{i1} = \mu_i \lambda_{i,i}, \quad i \geq 2.$$

- (iii) *In all other cases, there exist at least two different solutions of (3). In particular we have at least two solutions if the sequence  $(L_n)_{n \geq 1}$  is tight, but not convergent in distribution.*

In the case of a coalescent coming down from infinity, as already stated above, item (ii) applies. In the presence of dust the three cases all occur (see Theorem 2, Theorem 3, and Section 5). At first sight one may expect that the condition  $P(L_n = i) \sim \alpha_n \pi_i$  in item (ii) will occur only with  $\alpha_n \rightarrow 1$ , that is the random variables  $L_n$  converge in distribution. At the moment, however, we cannot exclude the possibility that the sequence  $(\alpha_n)$  is not convergent.

Theorem 4 will allow us to treat the time-reversal  $\hat{N}_n = (\hat{N}_n(t))_{t \geq 0}$  of the block-counting process  $N_n$ . This process is defined as the càdlàg process given by

$$\hat{N}_n(t) := \begin{cases} N_n((T_n - t)-) & \text{for } 0 \leq t < T_n, \\ n & \text{for } t \geq T_n. \end{cases}$$

In particular we have  $\hat{N}_n(0) = L_n$ .

**Theorem 5.** *If the sequence  $(L_n)_{n \geq 1}$  converges in distribution, then also the sequence of processes  $(\hat{N}_n)_{n \geq 1}$  converges in distribution in Skorohod space. The limit  $\hat{N}_\infty$  is a Markov jump process with values in  $\{2, 3, \dots\}$  and jump rates*

$$\hat{\rho}_{ij} := \frac{\mu_j \rho_{ji}}{\mu_i}, \quad i < j,$$

where the  $\mu_i$  are the weights of the  $\rho$ -invariant measure from Theorem 4(ii).

**Remark 6.** For the Kingman coalescent a direct computation shows that the solution of (3) is given by

$$\mu_i = \frac{2}{i(i-1)}, \quad i \geq 2.$$

For  $\Lambda = \text{Beta}(2 - \alpha, \alpha)$  with  $\alpha \in (0, 2)$ , Hénard [8] and Möhle [11] obtain

$$P(L_\infty = i) = \begin{cases} (-1)^{i-1} \alpha \binom{\alpha-1}{i-1} \int_{[0,1]} \frac{x^{i-1}}{1-(1-x)^{1-\alpha}} dx & \text{if } \alpha \neq 1, \\ -\frac{1}{i-1} \int_{[0,1]} \frac{x^{i-1}}{\log(1-x)} dx & \text{if } \alpha = 1. \end{cases}$$

Since in this case  $\lambda_{i,i} = \frac{B(i-\alpha,\alpha)}{B(2-\alpha,\alpha)}$ , we obtain from Theorem 4 an expression for the  $\rho$ -invariant measure  $\mu$  obeying (3).

The rest of this paper is organized as follows. We prove Theorem 1 in Section 2. In Section 3, we show how to approximate the number of blocks in the  $\Lambda$ -coalescent by means of a subordinator when (4) holds. We prove Theorem 2 in Section 4. In Section 5 we give an example in which  $(L_n)_{n \geq 1}$  is tight but does not converge in distribution because the log-nonlattice assumption in Theorem 2 fails. We then derive Theorem 3 in Section 6, and we prove Theorems 4 and 5 in Section 7.

## 2. Proof of Theorem 1

It will be useful throughout the paper to work with a Poisson process construction of the  $\Lambda$ -coalescent. The construction that we will give is a slight variation of the original such construction provided by Pitman in [12].

Assume  $\Lambda(\{0\}) = 0$ . Let  $\Psi$  be a Poisson point process on  $(0, \infty) \times (0, 1] \times [0, 1]^n$  with intensity

$$dt \times p^{-2} \Lambda(dp) \times du_1 \times \dots \times du_n.$$

Let  $\Pi_n(0) = \{\{1\}, \dots, \{n\}\}$  be the partition of the integers  $1, \dots, n$  into singletons. Suppose  $(t, p, u_1, \dots, u_n)$  is a point of  $\Psi$ , and  $\Pi_n(t-)$  consists of the blocks  $B_1, \dots, B_b$ , ranked in order by their smallest element. Then  $\Pi_n(t)$  is obtained from  $\Pi_n(t-)$  by merging together all of the blocks  $B_i$  for which  $u_i \leq p$  into a single block. These are the only times that mergers occur. This construction is well-defined because almost surely for any fixed  $t_0 < \infty$ , there are only finitely many points  $(t, p, u_1, \dots, u_n)$  of  $\Psi$  for which  $t \leq t_0$  and at least two of  $u_1, \dots, u_n$  are less than or equal to  $p$ . The resulting process  $\Pi_n = (\Pi_n(t), t \geq 0)$  is the  $\Lambda$ -coalescent. When  $(t, p, u_1, \dots, u_n)$  is a point of  $\Psi$ , we say that a  $p$ -merger occurs at time  $t$ .

We will need the following simple lemma pertaining to the rate at which the number of blocks decreases.

**Lemma 7.** Consider the  $\Lambda$ -coalescent  $\Pi_n$  started with  $n$  blocks and let  $0 < \gamma < 1$ . Let  $W_n = \inf\{t \geq 0 : N_n(t) \leq \gamma n\}$ . Then there exists a positive constant  $C$ , depending on  $\Lambda$  and  $\gamma$  but not on  $n$ , such that  $E[W_n] \leq C$  for all  $n \geq 2$ .

**Proof.** For  $2 \leq k \leq n$ , the probability that  $k$  is the smallest integer in one of the blocks of  $\Pi_n(t)$  is bounded above by the probability that the integers 1 and  $k$  do not merge before time  $t$ , which is  $e^{-\lambda_{2,2}t}$ . Therefore,

$$E[N_n(t)] \leq 1 + (n - 1)e^{-\lambda_{2,2}t}.$$

Thus, using Markov's Inequality,

$$P(W_n > t) = P(N_n(t) > \gamma n) \leq \frac{E[N_n(t)]}{\gamma n} \leq \frac{1}{\gamma n} + \frac{(n - 1)e^{-\lambda_{2,2}t}}{\gamma n}.$$

Because  $\lambda_{2,2} = \Lambda([0, 1]) > 0$  by assumption, there exists  $t_0 > 0$  such that  $P(W_n > t_0) \leq 1/2$  for sufficiently large  $n$ . By increasing the value of  $t_0$  if necessary, we can arrange for this inequality to hold for all  $n \geq 2$ . Then by repeatedly applying the Markov property, we get  $P(W_n > mt_0) \leq 2^{-m}$  for all positive integers  $m$ . It follows that  $E[W_n] \leq 2t_0$  for all  $n \geq 2$ , which gives the result.  $\square$

**Lemma 8.** Let  $B_{b,p}$  have a binomial distribution with parameters  $b$  and  $p$ . Then for all  $k, x > 0$

$$P(B_{b,p} \geq b - k) \leq 2p^{\lfloor b/2k \rfloor} \tag{7}$$

and

$$P(B_{b,p} \geq x) \leq p^x 2^b. \tag{8}$$

Moreover,

$$E\left[\frac{1}{B_{b,p} + 1}\right] = \frac{1 - (1 - p)^{b+1}}{(b + 1)p}. \tag{9}$$

**Proof.** To prove (7), let  $\xi_1, \dots, \xi_b$  be independent random variables with  $P(\xi_i = 1) = p$  and  $P(\xi_i = 0) = 1 - p$ . Observe that

$$P\left(\bigcup_{i=1}^j \{\xi_i = 0\} \mid \sum_{i=1}^b \xi_i \geq b - k\right) \leq j P\left(\xi_1 = 0 \mid \sum_{i=1}^b \xi_i \geq b - k\right) \leq \frac{jk}{b}.$$

In particular, if  $j \leq b/2k$ , then the right-hand side is less than 1/2 and, taking complements, we get

$$P\left(\xi_1 = \dots = \xi_j = 1 \mid \sum_{i=1}^b \xi_i \geq b - k\right) \geq \frac{1}{2}.$$

It follows by taking  $j = \lfloor b/2k \rfloor$  that

$$P\left(\sum_{i=1}^b \xi_i \geq b - k\right) \leq 2P(\xi_1 = \dots = \xi_j = 1) = 2p^{\lfloor b/2k \rfloor},$$

which gives (7).

To show (8) we obtain from an exponential Markov inequality that

$$P(B_{b,p} \geq x) \leq e^{-\lambda x} (1 + pe^\lambda)^b \tag{10}$$

with  $\lambda > 0$ . Putting  $\lambda = -\log p$  the inequality follows.

Finally, we have

$$E\left[\frac{1}{B_{b,p} + 1}\right] = \sum_{k=0}^b \frac{1}{k+1} \binom{b}{k} p^k (1-p)^{b-k} = \frac{1}{(b+1)p} \sum_{k=0}^b \binom{b+1}{k+1} p^{k+1} (1-p)^{b-k},$$

which equals the right-hand side of (9). □

Theorem 1 is an immediate consequence of Proposition 9 below when  $m = 1$ . (We state this proposition in a more general form, which we will use in the proof of Theorem 2.)

**Proposition 9.** *Suppose that (2) holds. Then for all  $\varepsilon > 0$ , there exists a positive integer  $K_\varepsilon$  such that  $P(m < N_n(t) \leq K_\varepsilon m$  for some  $t \geq 0$ )  $> 1 - \varepsilon$  for all integers  $m$  and  $n$  such that  $1 \leq m < n$ .*

**Proof.** For  $K \geq 2$ , let  $A_{m,n}$  be the complement of the event that  $m < N_n(t) \leq Km$  for some  $t \geq 0$ . If  $A_{m,n}$  occurs, then for some nonnegative integer  $\ell$ , a single merger takes the coalescent from between  $2^\ell Km + 1$  and  $2^{\ell+1} Km$  blocks down to  $m$  blocks or fewer.

Suppose there are  $b$  blocks in the  $\Lambda$ -coalescent at some time, where  $b \geq 2^\ell Km + 1$ , and then a  $p$ -merger occurs. For the  $p$ -merger to take the coalescent down to  $m$  blocks or fewer, the number of blocks that participate in the merger must be at least  $b - m + 1$ . By (7), if  $m \geq 2$ , then the probability that this occurs is bounded above by

$$2p^{\lfloor b/2(m-1) \rfloor} \leq 2p^{\lfloor (2^\ell Km + 1)/(2(m-1)) \rfloor} \leq 2p^{\lfloor 2^\ell (K/2) \rfloor} \leq 2p^{2^\ell (K/2) - 1}.$$

If  $m = 1$ , this probability is bounded above by  $p^b \leq 2p^{2^\ell (K/2) - 1}$ . Because, from the Poisson process construction of the  $\Lambda$ -coalescent, we know that  $p$ -mergers take place at rate  $p^{-2} \Lambda(dp)$ , it follows that the rate of events that take the coalescent down to  $m$  blocks or fewer is bounded above by

$$2 \int_0^1 p^{2^\ell (K/2) - 3} \Lambda(dp).$$

By Lemma 7, the expected amount of time for which the number of blocks is between  $2^\ell Km + 1$  and  $2^{\ell+1} Km$  is bounded above by  $C$  for all  $\ell$ . Therefore,

$$\begin{aligned} P(A_{m,n}) &\leq \sum_{\ell=0}^{\infty} 2C \int_0^1 p^{2^\ell (K/2) - 3} \Lambda(dp) \\ &= 2C \int_0^1 \sum_{\ell=0}^{\infty} p^{2^\ell (K/2) - 3} \Lambda(dp) \\ &\leq 2C \int_0^1 \sum_{\ell=0}^{\infty} p^{2^\ell ((K/2) - 3)} \Lambda(dp). \end{aligned}$$

For any  $a > 0$  and any  $x \in (0, 1)$ , we have

$$\sum_{\ell=0}^{\infty} x^{2^\ell a} = x^a + \sum_{\ell=1}^{\infty} \sum_{j=2^{\ell-1}+1}^{2^\ell} \frac{x^{2^\ell a}}{2^{\ell-1}} \leq x^a + \sum_{\ell=1}^{\infty} \sum_{j=2^{\ell-1}+1}^{2^\ell} \frac{2x^{ja}}{j} = 2 \sum_{j=1}^{\infty} \frac{x^{ja}}{j} = 2|\log(1 - x^a)|.$$

Therefore, if  $1 \leq m < n$ , then for  $K > 6$

$$P(A_{m,n}) \leq 4C \int_0^1 |\log(1 - p^{(K/2) - 3})| \Lambda(dp).$$

It follows from (2) and the Dominated Convergence Theorem that this expression tends to zero as  $K \rightarrow \infty$ , which gives the result. □

### 3. An approximation in the case of dust

Condition (4) allows us to approximate the number of blocks in the  $\Lambda$ -coalescent by a subordinator. For this, we will use the construction of the  $\Lambda$ -coalescent from the Poisson point process  $\Psi$  introduced at the beginning of Section 2. Let  $\phi : (0, \infty) \times (0, 1] \times [0, 1]^n \rightarrow (0, \infty) \times (0, \infty]$  be the function defined by

$$\phi(t, p, u_1, \dots, u_n) = (t, -\log(1 - p)).$$

Now  $\phi(\Psi)$  is a Poisson point process, and we can define a pure jump subordinator  $(S(t), t \geq 0)$  having the property that  $S(0) = 0$  and, if  $(t, x)$  is a point of  $\phi(\Psi)$ , then  $S(t) = S(t-) + x$ . This subordinator first appeared in the work of Pitman [12] and was used to approximate the block-counting process by Gnedin et al. [5] and Möhle [10]. The next theorem provides a refinement.

Define

$$f(y) := \int_0^1 \frac{1 - (1 - p)^{e^y}}{e^y} \frac{\Lambda(dp)}{p^2}, \quad y \in \mathbb{R}. \tag{11}$$

From (4), we see that  $f(y)$  is finite for all  $y \in \mathbb{R}$ . Also  $f$  is decreasing with  $\lim_{y \rightarrow \infty} f(y) = 0$ , because for fixed  $p$  the integrand has this behaviour. Let  $Y_n = (Y_n(t))_{t \geq 0}$  be the solution of the SDE

$$\log n - S(t) = Y_n(t) - \int_0^t f(Y_n(s)) ds, \quad t \geq 0. \tag{12}$$

Our goal is to show that for coalescents with dust the log of the block-counting process follows closely the process  $Y_n$ , up to the time when  $N_n$  has nearly reached the state 1. The drift  $f(Y_n(t)) dt$  appears because a merging of  $b$  out of  $N_n(t)$  lines results in a decrease by  $b - 1$  and not by  $b$  lines, see equation (23) below. For this purpose, we define for any  $k > 1$

$$\tau_{k,n} := \inf\{t \geq 0 : N_n(t) < k\}. \tag{13}$$

**Theorem 10.** *Under assumption (4), for all  $\varepsilon > 0$  there is an integer  $k \geq 2$  such that for all  $n$ ,*

$$P\left(\sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\log N_n(t) - Y_n(t)| \leq \varepsilon\right) > 1 - \varepsilon. \tag{14}$$

Note that (14) controls the distance between  $Y_n$  and  $\log N_n$  up to the first time point when  $N_n$  jumps below  $k$ . This time point is excluded only if the jump leads directly to 1, i.e. on the event  $\{\tau_{k,n} = T_n\}$ .

Before proving this theorem let us derive some consequences.

**Corollary 11.** *Under assumption (4), for all  $\varepsilon > 0$  there is an integer  $\ell$  such that*

$$P\left(\sup_{0 \leq t < T_n} |\log N_n(t) - Y_n(t)| \leq \ell\right) > 1 - \varepsilon. \tag{15}$$

**Proof.** For  $\tau_{k,n} < t < T_n$  and  $|\log N_n(\tau_{k,n}) - Y_n(\tau_{k,n})| \leq \varepsilon$  we have, since  $f(x) \geq 0$ ,

$$Y_n(t) \geq S(\tau_{k,n}) - S(t) + Y_n(\tau_{k,n}) \geq S(\tau_{k,n}) - S(T_n) - \varepsilon.$$

Hence, since  $f$  is decreasing,

$$\begin{aligned} |Y_n(t) - Y_n(\tau_{k,n})| &\leq S(T_n) - S(\tau_{k,n}) + \int_{\tau_{k,n}}^{T_n} f(Y_n(s)) ds \\ &\leq S(T_n) - S(\tau_{k,n}) + f(S(\tau_{k,n}) - S(T_n) - \varepsilon)(T_n - \tau_{k,n}) \end{aligned}$$

and therefore

$$\begin{aligned} |\log N_n(t) - Y_n(t)| &\leq |\log N_n(t) - \log N_n(\tau_{k,n})| + |\log N_n(\tau_{k,n}) - Y_n(\tau_{k,n})| + |Y_n(\tau_{k,n}) - Y_n(t)| \\ &\leq \log k + \varepsilon + S(T_n) - S(\tau_{k,n}) + f(S(\tau_{k,n}) - S(T_n) - \varepsilon)(T_n - \tau_{k,n}). \end{aligned}$$

By the strong Markov property,  $T_n - \tau_{k,n}$  is stochastically bounded from above by  $T_k$  and similarly  $S(T_n) - S(\tau_{k,n})$  by  $S(T_k)$ . Therefore  $\sup_{\tau_{k,n} < t < T_n} |\log N_n(t) - Y_n(t)|$  is stochastically bounded on the event  $|\log N_n(\tau_{k,n}) - Y_n(\tau_{k,n})| \leq \varepsilon$ . The claim now follows from Theorem 10.  $\square$

Since  $f(x) \rightarrow 0$  for  $x \rightarrow \infty$ , the processes  $Y_n$  and  $\log n - S$  are in view of (12) close to each other, and one may wonder whether also  $\log n - S$  is suitable to approximate the log of the block-counting process. This works under a stronger condition.

**Corollary 12.** *Under the assumption*

$$\int_0^1 |\log p| \frac{\Lambda(dp)}{p} < \infty, \tag{16}$$

for all  $\varepsilon > 0$  there is an integer  $k \geq 2$  such that for all  $n$ ,

$$P\left(\sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\log N_n(t) - \log n + S(t)| \leq \varepsilon\right) > 1 - \varepsilon. \tag{17}$$

**Proof.** For  $z \geq 1$  we have  $1 - (1 - p)^z \leq pz \wedge 1$ . Therefore with  $z = e^y$

$$\begin{aligned} \int_0^\infty f(y) dy &= \int_0^1 \int_0^\infty \frac{1 - (1 - p)^{e^y}}{e^y} dy \frac{\Lambda(dp)}{p^2} \\ &\leq \int_0^1 \left( \int_0^{|\log p|} p dy + \int_{|\log p|}^\infty e^{-y} dy \right) \frac{\Lambda(dp)}{p^2} \\ &= \int_0^1 (|\log p| + 1) \frac{\Lambda(dp)}{p} < \infty. \end{aligned}$$

For any integer  $i$  we have on the event  $\sup_{t < \tau_{2^i, n}} |\log N_n(t) - Y_n(t)| \leq \varepsilon$  because of the monotonicity of  $f$ ,

$$\begin{aligned} \int_0^{\tau_{2^i, n}} f(Y_n(s)) ds &\leq \sum_{j \geq i} \int_{\tau_{2^{j+1}, n}}^{\tau_{2^j, n}} f(\log N_n(s) - \varepsilon) ds \\ &\leq \sum_{j \geq i} f(j \log 2 - \varepsilon) (\tau_{2^j, n} - \tau_{2^{j+1}, n}). \end{aligned}$$

From Lemma 7 and the strong Markov property there is a  $C > 0$  such that

$$E \left[ \int_0^{\tau_{2^i, n}} f(Y_n(s)) ds \right] \leq C \sum_{j \geq i} f(j \log 2 - \varepsilon) \leq \frac{C}{\log 2} \int_{(i-1) \log 2 - \varepsilon}^\infty f(y) dy.$$

Choosing  $i$  large enough this bound may be made arbitrarily small. In view of (12) and Theorem 10 our claim follows.  $\square$

**Remark 13.** Gnedin, Iskanov, and Marynych [5] also studied the absorption time  $T_n$  by coupling with a subordinator. The hypothesis of Lemma 4.2 in [5] is that

$$\int_0^1 \left( \int_0^x v(y) dy \right) x^{-1} dx < \infty,$$



where  $\nu(y) = \int_y^1 x^{-2} \Lambda(dx)$ . This condition is equivalent to (16). To see this, note that

$$\begin{aligned} \int_0^1 (|\log x| + 1)x^{-1} \Lambda(dx) &= \int_0^1 (-x \log x + x)x^{-2} \Lambda(dx) = \int_0^1 \left( \int_0^x (-\log y) dy \right) x^{-2} \Lambda(dx) \\ &= \int_0^1 (-\log y) \left( \int_y^1 x^{-2} \Lambda(dx) \right) dy = \int_0^1 \left( \int_y^1 z^{-1} dz \right) \nu(y) dy \\ &= \int_0^1 \left( \int_0^z \nu(y) dy \right) z^{-1} dz. \end{aligned}$$

We now come to the proof of Theorem 10. It requires two preparatory lemmas.

**Lemma 14.** *Suppose  $X$  has a binomial distribution with parameters  $b$  and  $p$ . Then*

$$\log\left(\frac{X+1}{b+1}\right) - \log p = \frac{1}{p} \left( \frac{X+1}{b+1} - p - \frac{1-p}{b+1} \right) + R, \tag{18}$$

where

$$E[|R|] \leq \frac{1-p}{(b+1)p}.$$

**Proof.** By the Mean Value Theorem, if  $x > 0$  and  $y > 0$ , then there exists a positive number  $z$  between  $x$  and  $y$  such that  $\log x - \log y = z^{-1}(x - y)$ . Therefore, there exists a random variable  $Z$  between  $(X+1)/(b+1)$  and  $p$  such that

$$\log\left(\frac{X+1}{b+1}\right) - \log p = \frac{1}{Z} \left( \frac{X+1}{b+1} - p \right) = \frac{1}{p} \left( \frac{X+1}{b+1} - p \right) - R',$$

where

$$R' = \left( \frac{1}{p} - \frac{1}{Z} \right) \left( \frac{X+1}{b+1} - p \right).$$

Clearly  $R' \geq 0$ . It remains to bound  $E[R']$ . Because  $Z$  must be between  $(X+1)/(b+1)$  and  $p$ , we see that  $|1/Z - 1/p|$  can be bounded from above by substituting  $(X+1)/(b+1)$  in place of  $Z$ . We get

$$R' \leq \left( \frac{1}{p} - \frac{b+1}{X+1} \right) \left( \frac{X+1}{b+1} - p \right) = \frac{X+1}{(b+1)p} + \frac{(b+1)p}{X+1} - 2.$$

Now by (9),

$$E\left[ \frac{1}{X+1} \right] \leq \frac{1}{(b+1)p}.$$

Therefore,

$$E[R'] \leq \frac{bp+1}{(b+1)p} - 1 = \frac{1-p}{(b+1)p}.$$

Letting  $R = \frac{1-p}{(b+1)p} - R'$  proves the lemma. □

**Lemma 15.** *Suppose  $\Lambda((0, 1]) > 0$ , and define  $\tau_{k,n}$  as in (13). Then there exists a positive constant  $C_1$ , depending on  $\Lambda$  but not on  $n$ , such that for all  $2 \leq k \leq n$ ,*

$$E\left[ \int_0^{\tau_{k,n}} \frac{1}{N_n(s)} ds \right] \leq \frac{C_1}{k}. \tag{19}$$

**Proof.** Because  $\Lambda((0, 1]) > 0$ , there exist positive numbers  $r$  and  $d$  such that  $\Lambda([r, 1]) = d$ . This means that  $p$ -mergers with  $p \geq r$  occur at rate  $d$ . Let  $a \in (0, r \wedge 1/2)$  and  $c \in (0, d)$ . By the Law of Large Numbers, there exists a positive integer  $m$  such that for  $b \geq m$ , whenever the coalescent has  $b$  blocks, the rate of mergers that will bring the coalescent down to fewer than  $(1 - a)b$  blocks is at least  $c$ . Let  $e_b$  be the expected time, when the coalescent starts with  $b$  blocks, before the number of blocks drops below  $(1 - a)b$ . Let

$$C = \max \left\{ \frac{1}{c}, e_2, \dots, e_m \right\}.$$

Then, for all  $b \geq 2$ , if the coalescent starts with  $b$  blocks, the expected time before the number of blocks drops below  $(1 - a)b$  is at most  $C$ . For positive integers  $j$ , let

$$B_j = \{b \in \mathbb{N} : (1 - a)^{-(j-1)}k \leq b < (1 - a)^{-j}k\}.$$

Then the expected Lebesgue measure of  $\{t : N_n(t) \in B_j\}$  is at most  $C$ . Therefore,

$$E \left[ \int_0^{\tau_{k,n}} \frac{1}{N_n(s)} ds \right] \leq \sum_{j=1}^{\infty} \frac{C(1 - a)^{j-1}}{k} = \frac{C}{ak},$$

which implies (19) with  $C_1 = C/a$ . □

**Proof of Theorem 10.** Again we construct the  $\Lambda$ -coalescent from the Poisson point process  $\Psi$ , as described at the beginning of Section 2. Enumerate the points of  $\Psi$  as  $((t_i, p_i, u_{1,i}, \dots, u_{n,i}))_{i=1}^{\infty}$ . For each  $i \in \mathbb{N}$ , let

$$X_i = \sum_{j=1}^{N_n(t_i-)} \mathbb{1}_{\{u_{j,i} > p_i\}},$$

which is the number of extant lines that are not included in the merger at time  $t_i$ . Conditional on  $p_i$  and  $N_n(t_i-)$ , the distribution of  $X_i$  is binomial with parameters  $N_n(t_i-)$  and  $1 - p_i$ . Also, for all  $i \in \mathbb{N}$ , we have  $N_n(t_i) = X_i + \mathbb{1}_{\{X_i < N_n(t_i-)\}}$ . Dividing both sides by  $N_n(t_i-)$  and taking logs, we get

$$\log N_n(t_i) - \log N_n(t_i-) = \log \left( \frac{X_i + \mathbb{1}_{\{X_i < N_n(t_i-)\}}}{N_n(t_i-)} \right).$$

Also,

$$S(t_i) - S(t_i-) = -\log(1 - p_i).$$

It follows that for  $t > 0$ ,

$$\log N_n(t) - (\log n - S(t)) = \sum_{i=1}^{\infty} \left( \log \left( \frac{X_i + \mathbb{1}_{\{X_i < N_n(t_i-)\}}}{N_n(t_i-)} \right) - \log(1 - p_i) \right) \mathbb{1}_{\{t_i \leq t\}}.$$

Noting

$$\log \left( \frac{X_i + \mathbb{1}_{\{X_i < N_n(t_i-)\}}}{N_n(t_i-)} \right) = \log \left( \frac{X_i + 1}{N_n(t_i-) + 1} \right) + \mathbb{1}_{\{X_i < N_n(t_i-)\}} \log \frac{N_n(t_i-) + 1}{N_n(t_i-)}$$

and letting

$$U_n(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\{X_i < N_n(t_i-)\}} \log \frac{N_n(t_i-) + 1}{N_n(t_i-)} \mathbb{1}_{\{t_i \leq t\}},$$

we can write

$$\begin{aligned} & \log N_n(t) - (\log n - S(t)) \\ &= \sum_{i=1}^{\infty} \left( \frac{1}{1-p_i} \left( \frac{X_i + 1}{N_n(t_i-) + 1} - (1-p_i) - \frac{p_i}{N_n(t_i-) + 1} \right) + R_i \right) \mathbb{1}_{\{t_i \leq t\}} + U_n(t), \end{aligned}$$

where  $R_i$  is defined as in (18), with  $N_n(t_i-)$  in place of  $n$ ,  $X_i$  in place of  $X$ , and  $1-p_i$  in place of  $p$ .

We now break this sum into pieces. Let  $\varepsilon > 0$ , and let  $J = \{i \in \mathbb{N} : p_i \leq 1 - \varepsilon/(4N_n(t_i-))\}$ . For  $t \geq 0$ , let

$$M_n(t) = \sum_{i=1}^{\infty} \frac{1}{1-p_i} \left( \frac{X_i + 1}{N_n(t_i-) + 1} - (1-p_i) - \frac{p_i}{N_n(t_i-) + 1} \right) \mathbb{1}_{\{t_i \leq t \wedge T_n\}} \mathbb{1}_{\{i \in J\}}$$

and

$$V_n(t) = \sum_{i=1}^{\infty} R_i \mathbb{1}_{\{t_i \leq t \wedge T_n\}} \mathbb{1}_{\{i \in J\}}.$$

The probability that  $N_n(t_i) = 1$ , conditional on  $N_n(t_i-)$  and on the event  $\{i \notin J\}$ , is at least  $1 - \varepsilon/4$ . Therefore,

$$P(\log N_n(t) - U_n(t) - (\log n - S(t)) = M_n(t) + V_n(t) \text{ for all } t < T_n) \geq 1 - \varepsilon/4,$$

which means that for  $k > 1$

$$\begin{aligned} & P\left( \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\log N_n(t) - U_n(t) - (\log n - S(t))| > \frac{\varepsilon}{4} \right) \\ & \leq \frac{\varepsilon}{4} + P\left( \sup_{t \leq \tau_{k,n}} |M_n(t)| > \frac{\varepsilon}{8} \right) + P\left( \sup_{t \leq \tau_{k,n}} |V_n(t)| > \frac{\varepsilon}{8} \right). \end{aligned} \tag{20}$$

Conditional on  $p_i$  and  $N_n(t_i-)$ , the random variable

$$\frac{1}{1-p_i} \left( \frac{X_i + (1-p_i)}{N_n(t_i-) + 1} - (1-p_i) \right)$$

has mean zero and variance

$$\frac{N_n(t_i-) p_i}{(N_n(t_i-) + 1)^2 (1-p_i)}.$$

In particular, the process  $(M_n(t), t \geq 0)$  is a martingale. Recalling the definition of  $\tau_{k,n}$  from (13) and putting  $l_p := \lceil \varepsilon/(4(1-p)) \rceil$ , we get for the bracket process  $\langle M_n \rangle$

$$\begin{aligned} \langle M_n \rangle(\tau_{k,n}) & \leq \int_0^{\tau_{k,n}} \int_0^{1-\varepsilon/(4N_n(s))} \frac{p}{(N_n(s) + 1)(1-p)} \frac{\Lambda(dp)}{p^2} ds \\ & \leq \int_0^1 \frac{1}{1-p} \left( \int_0^{\tau_{k,n}} \frac{1}{N_n(s)} \mathbb{1}_{\{N_n(s) \geq \varepsilon/(4(1-p))\}} ds \right) \frac{\Lambda(dp)}{p} \\ & \leq \int_0^1 \frac{1}{1-p} \left( \int_0^{\tau_{k,n} \wedge \tau_{l_p,n}} \frac{1}{N_n(s)} ds \right) \frac{\Lambda(dp)}{p}. \end{aligned}$$

Combining this result with (19) and using  $\tau_{k,n} \wedge \tau_{l_p,n} = \tau_{k \vee l_p, n}$  we obtain

$$E[\langle M_n \rangle(\tau_{k,n})] \leq \int_0^1 \frac{1}{1-p} \cdot C_1 \left( \frac{1}{k} \wedge \frac{4(1-p)}{\varepsilon} \right) \cdot \frac{\Lambda(dp)}{p},$$

which is finite by (4) and goes to 0 for  $k \rightarrow \infty$ . Therefore, by the  $L^2$  Maximum Inequality for martingales and Markov’s inequality, we get that for  $k$  sufficiently large

$$E \left[ \sup_{t \leq \tau_{k,n}} |M_n(t)|^2 \right] \leq \frac{\varepsilon^3}{4 \cdot 64} \text{ and } P \left( \sup_{t \leq \tau_{k,n}} |M_n(t)| > \frac{\varepsilon}{8} \right) \leq \frac{\varepsilon}{4}. \tag{21}$$

We now consider the process  $(V_n(t), t \geq 0)$ . By Lemma 14,

$$\begin{aligned} E \left[ \sup_{t \leq \tau_{k,n}} |V_n(t)| \right] &\leq E \left[ \sum_{i=1}^{\infty} |R_i| \mathbb{1}_{\{t_i \leq \tau_{k,n}\}} \mathbb{1}_{\{i \in J\}} \right] \\ &\leq E \left[ \int_0^{\tau_{k,n}} \int_0^{1-\varepsilon/(4N_n(s))} \frac{p}{(N_n(s) + 1)(1-p)} \frac{\Lambda(dp)}{p^2} ds \right]. \end{aligned}$$

Thus as above, if  $k$  is sufficiently large,

$$E \left[ \sup_{t \leq \tau_{k,n}} |V_n(t)| \right] \leq \frac{\varepsilon^2}{32} \text{ and } P \left( \sup_{t \leq \tau_{k,n}} |V_n(t)| > \frac{\varepsilon}{8} \right) \leq \frac{\varepsilon}{4}.$$

Together with (20) and (21) we arrive at

$$P \left( \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\log N_n(t) - U_n(t) - (\log n - S(t))| > \frac{\varepsilon}{4} \right) \leq \frac{3\varepsilon}{4}. \tag{22}$$

Now we approximate  $U_n(t)$  by  $\int_0^t f(\log N_n(s)) ds$ , uniformly for  $t \leq \tau_{k,n}$ . Note that by (4), there are only finitely many  $t_i$  such that  $t_i \leq T_n$  and  $X_i < N_n(t_i -)$ . Denote these points by  $s_1 < \dots < s_m$ , and also set  $s_0 = 0$  and  $s_{m+1} = \infty$ . Note that  $s_m = T_n$ . When the coalescent has  $b$  blocks, the points  $s_i$  appear at rate

$$\rho(b) = \int_0^1 \sum_{k=1}^b \binom{b}{k} p^k (1-p)^{b-k} \frac{\Lambda(dp)}{p^2} = \int_0^1 (1 - (1-p)^b) \frac{\Lambda(dp)}{p^2}. \tag{23}$$

Therefore, the random variables  $G_i = (s_{i+1} - s_i) \rho(N_n(s_i))$  for  $0 \leq i \leq m - 1$  are independent standard exponential random variables, also independent of the process  $N_n(s_j), j \geq 1$ . Recalling (11), we have  $\rho(b) = bf(\log b)$ . Now for  $t \leq T_n$

$$\begin{aligned} \int_0^t f(\log N_n(s)) ds &= \sum_{i=0}^{m-1} f(\log N_n(s_i)) ((s_{i+1} - s_i) \mathbb{1}_{\{s_{i+1} \leq t\}} + (t - s_i) \mathbb{1}_{\{s_i < t < s_{i+1}\}}) \\ &= \sum_{i=0}^{m-1} \frac{G_i}{N_n(s_i)} \left( \mathbb{1}_{\{s_{i+1} \leq t\}} + \frac{t - s_i}{s_{i+1} - s_i} \mathbb{1}_{\{s_i < t < s_{i+1}\}} \right). \end{aligned}$$

Consequently, since  $U_n(t) = \sum_{i=0}^{m-1} \log((N_n(s_i) + 1)/N_n(s_i)) \mathbb{1}_{\{s_{i+1} \leq t\}}$ ,

$$\begin{aligned} \int_0^t f(\log N_n(s)) ds - U_n(t) &= \sum_{i=0}^{m-1} \frac{G_i - 1}{N_n(s_i)} \mathbb{1}_{\{s_{i+1} \leq t\}} + \sum_{i=0}^{m-1} \frac{G_i}{N_n(s_i)} \frac{t - s_i}{s_{i+1} - s_i} \mathbb{1}_{\{s_i < t < s_{i+1}\}} \\ &\quad + \sum_{i=0}^{m-1} \left( \frac{1}{N_n(s_i)} - \log \frac{N_n(s_i) + 1}{N_n(s_i)} \right) \mathbb{1}_{\{s_{i+1} \leq t\}}. \end{aligned}$$

Using that the second sum has just one non-vanishing summand, and that  $x - \log(1 + x) \leq x^2$  for  $x \geq 0$ , we have for  $t \leq \tau_{k,n}$

$$\begin{aligned} & \left| \int_0^t f(\log N_n(s)) ds - U_n(t) \right| \\ & \leq \left| \sum_{i=0}^{m-1} \frac{G_i - 1}{N_n(s_i)} \mathbb{1}_{\{s_{i+1} \leq t\}} \right| + \max_{0 \leq i \leq m-1} \frac{G_i}{N_n(s_i)} \mathbb{1}_{\{s_i < \tau_{k,n}\}} + \sum_{i=0}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{s_i < \tau_{k,n}\}}. \end{aligned} \tag{24}$$

We show that for  $k$  sufficiently large the supremum over  $t \leq \tau_{k,n}$  of the right-hand side gets arbitrarily small in probability, uniformly in  $n$ . To this end we deal with the three summands on the r.h.s. of (24) in reverse order.

First we have

$$\sum_{i=0}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{s_i < \tau_{k,n}\}} \leq \sum_{j=k}^n \frac{1}{j^2} + \sum_{i=1}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{N_n(s_i) = N_n(s_{i-1})\}} \mathbb{1}_{\{s_i < \tau_{k,n}\}}$$

and so by Lemma 15

$$\begin{aligned} E \left[ \sum_{i=0}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{s_i < \tau_{k,n}\}} \right] & \leq \frac{2}{k} + E \left[ \int_0^1 \int_0^{\tau_{k,n}} \frac{N_n(s) p(1-p)^{N_n(s)-1}}{N_n(s)^2} ds \frac{\Lambda(dp)}{p^2} \right] \\ & \leq \frac{2}{k} + \int_0^1 E \left[ \int_0^{\tau_{k,n}} \frac{1}{N_n(s)} ds \right] \frac{\Lambda(dp)}{p} \\ & \leq \frac{1}{k} \left( 2 + C_1 \int_0^1 \frac{\Lambda(dp)}{p} \right). \end{aligned} \tag{25}$$

Second, since  $E[G_i^2] = 2$ , we have for  $u > 0$

$$\begin{aligned} P \left( \max_{0 \leq i \leq m-1} \frac{G_i}{N_n(s_i)} \mathbb{1}_{\{s_i < \tau_{k,n}\}} > u \right) & \leq E \left[ \sum_{i=0}^{m-1} P \left( \frac{G_i}{N_n(s_i)} \mathbb{1}_{\{s_i < \tau_{k,n}\}} > u \mid N_n(s_i), i \geq 1 \right) \right] \\ & \leq \frac{1}{u^2} E \left[ \sum_{i=0}^{m-1} \frac{2}{N_n(s_i)^2} \mathbb{1}_{\{s_i < \tau_{k,n}\}} \right] \\ & \leq \frac{2}{u^2 k} \left( 2 + C_1 \int_0^1 \frac{\Lambda(dp)}{p} \right), \end{aligned}$$

where we used (25) in the last inequality.

Third let

$$M'_n(t) = \sum_{i=0}^{m-1} \frac{G_i - 1}{N_n(s_i)} \mathbb{1}_{\{s_{i+1} \leq t\}}.$$

Then  $(M'_n(t), t \geq 0)$  is a martingale with

$$E \left[ (M'_n)'(\tau_{k,n}) \right] = E \left[ \sum_{i=0}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{s_{i+1} \leq \tau_{k,n}\}} \right] \leq E \left[ \sum_{i=0}^{m-1} \frac{1}{N_n(s_i)^2} \mathbb{1}_{\{s_i < \tau_{k,n}\}} \right],$$

and again by means of the  $L_2$  Maximum inequality and (25)

$$E \left[ \sup_{t \leq \tau_{k,n}} \left| \sum_{i=0}^{m-1} \frac{G_i - 1}{N_n(s_i)} \mathbb{1}_{\{s_{i+1} \leq t\}} \right|^2 \right] \leq \frac{4}{k} \left( 2 + C_1 \int_0^1 \frac{\Lambda(dp)}{p} \right).$$

Using these three estimates we obtain from (24) that for any  $\varepsilon > 0$

$$P \left( \sup_{t \leq \tau_{k,n}} \left| \int_0^t f(\log N_n(s)) ds - U_n(t) \right| > \frac{\varepsilon}{4} \right) \leq \frac{\varepsilon}{4},$$

if  $k$  is sufficiently large. Combining this bound with (22) we arrive at the formula

$$P \left( \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} \left| \log N_n(t) - \int_0^t f(\log N_n(s)) ds - (\log n - S(t)) \right| > \frac{\varepsilon}{2} \right) \leq \varepsilon. \tag{26}$$

To finish the proof we define for  $t \geq 0$

$$\begin{aligned} \Delta_n(t) &:= \log N_n(t) - \int_0^t f(\log N_n(s)) ds - (\log n - S(t)) \\ &= \log N_n(t) - \int_0^t f(\log N_n(s)) ds - \left( Y_n(t) - \int_0^t f(Y_n(s)) ds \right). \end{aligned}$$

For fixed  $t$  and  $n$  we consider the event  $A_{\geq} := \{t < T_n, t \leq \tau_{k,n}, \log N_n(t) \geq Y_n(t)\}$  and define the random time

$$\sigma_t := \sup\{s \leq t : \log N_n(s) \leq Y_n(s)\}.$$

Then on the event  $A_{\geq}$  we have  $\log N_n(\sigma_t-) - Y_n(\sigma_t-) \leq 0$  and  $f(\log N_n(s)) - f(Y_n(s)) \leq 0$  for  $s > \sigma_t$ , since  $f$  is decreasing. Thus, on  $A_{\geq}$ ,

$$\begin{aligned} 0 &\leq \log N_n(t) - Y_n(t) \\ &= \log N_n(\sigma_t-) - Y_n(\sigma_t-) + \int_{\sigma_t}^t (f(\log N_n(s)) - f(Y_n(s))) ds + \Delta_n(t) - \Delta_n(\sigma_t-) \\ &\leq \Delta_n(t) - \Delta_n(\sigma_t-) \\ &\leq 2 \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\Delta_n(t)|. \end{aligned}$$

Similarly on  $A_{\leq} := \{t < T_n, t \leq \tau_{k,n}, \log N_n(t) \leq Y_n(t)\}$ ,

$$0 \leq Y_n(t) - \log N_n(t) \leq 2 \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\Delta_n(t)|.$$

Recalling (26), this implies that for sufficiently large  $k$ ,

$$P \left( \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\log N_n(t) - Y_n(t)| > \varepsilon \right) \leq P \left( \sup_{t \in [0, \tau_{k,n}] \cap [0, T_n]} |\Delta_n(t)| > \frac{\varepsilon}{2} \right) \leq \varepsilon,$$

which was the claim. □

#### 4. Proof of Theorem 2

In this section we prove Theorem 2. First we provide a lemma which gives a uniform lower bound for the probability that the block-counting process does not jump over certain intervals.

**Lemma 16.** *Assume (2) and that  $\Lambda$  is log-nonlattice. Fix  $0 < \delta < 1$  and  $K > 1$ . Suppose  $m < n \leq Km$ . Then there exist constants  $C > 0$  and  $\alpha \in (0, 1]$ , depending on  $\delta$  and  $K$  but not on  $m$  or  $n$ , such that  $P((1 - \delta)\alpha m \leq N_n(t) \leq \alpha m \text{ for some } t \geq 0) \geq C$ .*

**Proof.** We distinguish two cases. First assume that for all  $\eta > 0$  we have  $\Lambda((0, \eta]) > 0$ . Let  $\eta = 4^{-2K/\delta}$  and define  $N'_n, N''_n$  to be the block-counting processes belonging to the two coalescents arising by restricting  $\Lambda$  to the intervals either  $[0, \eta]$  or  $(\eta, 1]$ , and using the same Poisson process  $\Psi$ . The processes  $N'_n, N''_n$  are independent, therefore for any  $u > 0$

$$\begin{aligned} &P((1 - \delta)m \leq N_n(t) \leq m \text{ for some } t \geq 0) \\ &\geq P\left(N''_n(u) = n, N'_n(u) \leq (1 - \delta)m, \sup_{t \leq u} (N'_n(t-) - N'_n(t)) \leq \delta m\right) \\ &\geq P\left(N''_n(u) = n, N'_n(u) \leq (1 - \delta)n/K, \sup_{t \leq u} (N'_n(t-) - N'_n(t)) \leq \delta n/K\right) \\ &\geq P(N''_n(u) = n)P(N'_n(u) \leq (1 - \delta)n/K) - P\left(\sup_{t \leq u} (N'_n(t-) - N'_n(t)) > \delta n/K\right). \end{aligned}$$

By assumption the process  $N'_n$  is non-degenerate. Thus in view of Lemma 7 the expectation of  $W'_n := \min\{t \geq 0 : N'_n(t) \leq (1 - \delta)n/K\}$  is bounded by a constant  $\kappa$ , depending on  $\delta$  and  $K$  but not on  $n$ . Choosing  $u = 2\kappa$  we obtain from Markov's Inequality

$$P(N'_n(2\kappa) > (1 - \delta)n/K) = P(W'_n \geq 2\kappa) \leq \frac{1}{2\kappa} E[W'_n] \leq \frac{1}{2}.$$

Moreover

$$P(N''_n(2\kappa) = n) \geq e^{-2\kappa \int_{\eta}^1 p^{-2\Lambda(dp)} > 0}.$$

Finally, for the rate at which  $N'_n$  performs at time  $t$  a jump of size larger than  $\delta n/K$ , we obtain from (8) and from the choice of  $\eta$  for  $n \geq 4K/\delta$  the bound

$$\begin{aligned} \int_0^{\eta} P(B_{N'_n(t-), p} > \delta n/K) \frac{\Lambda(dp)}{p^2} &\leq \int_0^{\eta} p^{\delta n/K} 2^{N'_n(t-)} \frac{\Lambda(dp)}{p^2} \\ &\leq \eta^{\delta n/(2K)} 2^n \Lambda([0, 1]) = 2^{-n} \Lambda([0, 1]). \end{aligned}$$

Therefore

$$P\left(\sup_{t \leq 2\kappa} (N'_n(t-) - N'_n(t)) > \delta n/K\right) \leq 2\kappa 2^{-n} \Lambda([0, 1]).$$

Putting our estimates together we arrive at

$$P((1 - \delta)m \leq N_n(t) \leq m \text{ for some } t \geq 0) \geq \frac{1}{4} e^{-2\kappa \int_{\eta}^1 p^{-2\Lambda(dp)} > 0}$$

for  $n$  sufficiently large and any  $m$  with  $m < n \leq Km$ . A further lowering of this bound makes the estimate valid for all  $n$ . Letting  $\alpha = 1$  our claim follows.

For the second part of the proof let  $\Lambda([0, \eta]) = 0$  for some  $\eta > 0$ . Then (16) is satisfied such that we may resort to Corollary 12. Note that our log-nonlattice assumption means that the random walk  $(S(i), i \in \mathbb{N}_0)$  is non-lattice in the usual sense. Condition (2) implies  $E[S(1)] < \infty$ . Therefore the classical renewal theorem implies that with  $\alpha$  sufficiently small there is a constant  $0 < C \leq 1/2$  depending on  $\delta$  such that for all  $s \geq 0$

$$P\left(\exists i \in \mathbb{N}_0 : s - \log \alpha - \frac{1}{3} \log(1 - \delta) \leq S(i) \leq s - \log \alpha - \frac{2}{3} \log(1 - \delta)\right) \geq 2C,$$

and consequently for  $m < n$  (letting  $s = \log n - \log m$ )

$$P\left(\exists t \geq 0 : \frac{2}{3} \log(1 - \delta) + \log \alpha m \leq \log n - S(t) \leq \frac{1}{3} \log(1 - \delta) + \log \alpha m\right) \geq 2C. \tag{27}$$

Next, choose  $k$  according to Corollary 12 so that (17) holds with  $\varepsilon = \frac{1}{4}C \wedge \frac{1}{3}|\log(1 - \delta)|$ . Let  $k$  be so large that by Theorem 1, we have  $P(\tau_{k,n} = T_n) = P(N_n(T_n-) \geq k) \leq \frac{1}{4}C$  for all  $n$ . Then

$$P\left(\sup_{t \leq \tau_{k,n}} |\log N_n(t) - \log n + S(t)| \geq \frac{1}{3}|\log(1 - \delta)|\right) \leq \frac{1}{2}C. \tag{28}$$

In particular with  $t = \tau_{k,n}$ , since  $k \geq N_n(\tau_{k,n})$ ,

$$P\left(\log n - S(\tau_{k,n}) \geq \log k - \frac{1}{3} \log(1 - \delta)\right) \leq \frac{1}{2}C$$

and hence for  $n$  sufficiently large, because  $m \geq n/K$ , and because of a.s. monotonicity of  $S$ ,

$$P\left(\forall t > \tau_{k,n} : \log n - S(t) < \log \alpha m + \frac{2}{3} \log(1 - \delta)\right) \geq 1 - \frac{1}{2}C.$$

Intersecting this event with the event in (27) we obtain

$$P\left(\exists t \leq \tau_{k,n} : \frac{2}{3} \log(1 - \delta) + \log \alpha m \leq \log n - S(t) \leq \frac{1}{3} \log(1 - \delta) + \log \alpha m\right) \geq \frac{3}{2}C.$$

Hence from (28) it follows for  $n$  sufficiently large and  $m < n \leq Km$

$$P(\exists t \leq \tau_{k,n} : \log(1 - \delta) + \log \alpha m \leq \log N_n(t) \leq \log \alpha m) \geq C.$$

Again by suitably lowering the constant  $C$  this estimate holds for all  $n$ , which then translates into our claim. □

**Proof of Theorem 2.** We prove this result by coupling. Let  $\varepsilon > 0$ . It suffices to show that there exists a positive integer  $n_0$  such that if  $n_0 < n_1 < n_2$ , then we can construct  $\Lambda$ -coalescents  $(\Pi_{n_1}(t), t \geq 0)$  and  $(\Pi_{n_2}(t), t \geq 0)$  started with  $n_1$  and  $n_2$  blocks respectively such that

$$P(N_{n_1}(T_{n_1}-) = N_{n_2}(T_{n_2}-)) > 1 - \varepsilon. \tag{29}$$

By Theorem 1, we can choose a positive integer  $\ell$  such that  $P(N_n(T_n-) \leq \ell) > 1 - \varepsilon/4$  for all  $n$ . Let  $C$  be the constant from Lemma 16 with  $\delta = \varepsilon/(4\ell)$  and with the constant  $K = K_{1/2}$  from Proposition 9. Choose a positive integer  $J$  large enough that

$$\left(1 - \frac{C^2}{4}\right)^J < \frac{\varepsilon}{2}.$$

Then for  $1 \leq j \leq J$ , let  $m_j = \lfloor n_0^{j/J} \rfloor$ . For  $1 \leq j \leq J$  and  $i \in \{1, 2\}$ , let  $A_{i,j}$  be the event that  $m_j < N_{n_i}(t) \leq Km_j$  for some  $t \geq 0$ , and let  $D_{i,j}$  be the event that  $(1 - \delta)\alpha m_j \leq N_{n_i}(t) \leq \alpha m_j$  for some  $t \geq 0$ , with the constant  $\alpha$  as in Lemma 16. It follows from Proposition 9 and Lemma 16 that for  $1 \leq j \leq J$  and  $i \in \{1, 2\}$ , we have

$$P(D_{i,j}) \geq P(D_{i,j} \cap A_{i,j}) = P(A_{i,j})P(D_{i,j}|A_{i,j}) \geq \frac{1}{2}C. \tag{30}$$

We will need to establish that a similar inequality holds when we condition on the events  $D_{i,k}$  for  $k > j$ . To this end, let  $U_{i,j} = 0$  for  $i \in \{1, 2\}$ , and for  $1 \leq j \leq J - 1$  and  $i \in \{1, 2\}$ , define the stopping time  $U_{i,j} = \inf\{t \geq 0 : N_{n_i}(t) \leq \alpha m_{j+1}\}$ . For  $1 \leq j \leq J$  and  $i \in \{1, 2\}$ , let  $G_{i,j} = \{N_{n_i}(U_{i,j}) > m_j\}$ . Let  $(\mathcal{F}_i(t), t \geq 0)$  be the natural filtration



associated with the process  $(\Pi_{n_i}(t), t \geq 0)$ . With  $N_{n_i}(U_{i,j})$  figuring as the new starting point, the reasoning leading to (30) implies that for  $1 \leq j \leq J$  and  $i \in \{1, 2\}$ , we have, on the event  $G_{i,j}$ ,

$$P(D_{i,j} | \mathcal{F}_i(U_{i,j})) \geq \frac{1}{2}C \quad \text{a.s.} \tag{31}$$

Because  $m_{j+1}/m_j \rightarrow \infty$  as  $n_0 \rightarrow \infty$ , it follows from Proposition 9 that

$$\lim_{n_0 \rightarrow \infty} P(G_{i,j}) = 1. \tag{32}$$

Since  $D_{i,k} \in \mathcal{F}_i(U_{i,j})$  for  $1 \leq j < k \leq J$  and  $i \in \{1, 2\}$ , the results (31) and (32) imply that if the processes  $(\Pi_{n_1}(t), t \geq 0)$  and  $(\Pi_{n_2}(t), t \geq 0)$  are independent, then

$$\limsup_{n_0 \rightarrow \infty} P\left(\bigcup_{j=1}^J (D_{1,j} \cap D_{2,j})\right) \geq 1 - \left(1 - \frac{C^2}{4}\right)^J > 1 - \frac{\varepsilon}{2}. \tag{33}$$

We now couple the processes  $(\Pi_{n_1}(t), t \geq 0)$  and  $(\Pi_{n_2}(t), t \geq 0)$ . We allow the two processes to evolve independently until the times  $U_{1,J-1}$  and  $U_{2,J-1}$  respectively. If  $D_{1,J} \cap D_{2,J}$  occurs, then we stop. Otherwise, we allow the processes to continue to evolve independently until the times  $U_{1,J-2}$  and  $U_{2,J-2}$  respectively. Then we stop if  $D_{1,J-1} \cap D_{2,J-1}$  occurs, and otherwise continue as before. According to (33), with probability at least  $1 - \varepsilon/2$ , we will eventually come to a value of  $j$  such that  $D_{1,j} \cap D_{2,j}$  occurs. In that case, the independent constructions will be stopped at the times  $U_{1,j-1}$  and  $U_{2,j-1}$  respectively, at which times both processes will have between  $(1 - \delta)\alpha m_j$  and  $\alpha m_j$  blocks.

We now suppose the independent constructions are stopped at the times  $U_{1,j-1}$  and  $U_{2,j-1}$ . Set  $n'_1 = N_{n_1}(U_{1,j-1})$  and  $n'_2 = N_{n_2}(U_{2,j-1})$ . Without loss of generality, assume  $n'_1 < n'_2$ . Let  $B_{1,1}, \dots, B_{1,n'_1}$  and  $B_{2,1}, \dots, B_{2,n'_2}$  denote the blocks of the partitions  $\Pi_{n_1}(U_{1,j-1})$  and  $\Pi_{n_2}(U_{2,j-1})$  respectively. We now construct  $(\Pi_{n_1}(U_{1,j-1} + t), t \geq 0)$  and  $(\Pi_{n_2}(U_{2,j-1} + t), t \geq 0)$  from the same Poisson point process  $\Psi$ , as described at the beginning of Section 2. This means both processes will have  $p$ -mergers at the same times, and the number of blocks in  $\Pi_{n_2}(U_{2,j-1} + t)$  that contain integers from one or more of the blocks  $B_{2,1}, \dots, B_{2,n'_1}$  will equal  $N_{n_1}(U_{1,j-1} + t)$ . Recall that  $T_{n_2}$  is the time of the last merger in  $(\Pi_{n_2}(t), t \geq 0)$ . Unless one or more blocks of  $\Pi_{n_2}(T_{n_2}-)$  contains only integers from the blocks  $B_{2,n'_1+1}, \dots, B_{2,n'_2}$ , we will have  $N_{n_1}(T_{n_1}-) = N_{n_2}(T_{n_2}-)$ . By the exchangeability of the coalescent dynamics, conditional on  $n'_1$  and  $n'_2$ , the probability that a particular block of  $\Pi_{n_2}(T_{n_2}-)$  contains only integers from the blocks  $B_{2,n'_1+1}, \dots, B_{2,n'_2}$  is at most  $(n'_2 - n'_1)/n'_2$ , which is at most  $\delta$  because we are assuming that  $D_{1,j} \cap D_{2,j}$  occurs. Therefore, recalling that  $\ell$  was chosen so that  $P(N_{n_2}(T_{n_2}-) > \ell) < \varepsilon/4$ , we have

$$P(N_{n_1}(T_{n_1}-) \neq N_{n_2}(T_{n_2}-)) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \ell\delta = \varepsilon,$$

which implies (29). □

### 5. Non-convergence for Eldon–Wakeley coalescents

To provide an example where the distribution of the size of the last merger does *not* converge as  $n \rightarrow \infty$ , we now focus on the class of coalescents proposed in [4] and thus assume that the measure  $\Lambda$  is concentrated in one point  $p \neq 0, 1$ . Because of Theorem 1, for such coalescents the size of the last merger is tight. We claim that still  $L_n$  does not converge in distribution as  $n \rightarrow \infty$ . There are obvious relations to non-convergence and periodicity phenomena in the so-called leader election, see e.g. Grübel and Hagemann [7] and references therein.

For notational convenience we restrict ourselves to the case  $\Lambda = p^2\delta_p$  and  $p = e^{-1}$ . Then the points of the Poisson point process  $\Psi$  are of the form  $(\sigma_i, p, u_1, \dots, u_n)$ ,  $i = 1, 2, \dots$ , where the numbers  $0 < \sigma_1 < \sigma_2 < \dots$  form a standard Poisson point process on  $\mathbb{R}_+$ . Define  $\tau_{k,n}$  as in (13).

We shall argue by contradiction, so let us assume that  $L_n$  does converge in distribution. Then, as shown in Theorem 5, the sequence of time-reversed Markov chains converges as  $n \rightarrow \infty$  in distribution to a limiting Markov chain.

This implies

$$\forall \varepsilon > 0 \exists k > 0: N_n(\tau_{k,n}) \xrightarrow{d} N_{\infty,k} \quad \text{with } P(N_{\infty,k} \geq 2) \geq 1 - \varepsilon. \tag{34}$$

Together with  $N_n$  we consider a process  $\bar{N}_n \geq N_n$  defined inductively as follows:  $\bar{N}_n(0) = N_n(0)$  and at times  $\sigma_i$  the random number  $\bar{N}_n(\sigma_i)$  is thinned according to  $p$  and afterwards is increased by one. Thinking of  $N_n$  and  $\bar{N}_n$  as numbers of lines, the difference between both processes only arises, when by a thinning no line of  $N_n$  is affected. Then  $N_n$  does not change its value but  $\bar{N}_n$  increases by 1. Given  $N_n(t) = m$  this takes place with probability  $q^m$  with  $q = 1 - p$ . This may occur several times, and, as long as  $N_n$  stays at level  $m$ , the expected increase of  $\bar{N}_n$  is bounded from above by  $q^m/(1 - q^m) \leq q^m/p$ . Therefore, given  $\varepsilon > 0$  there is a  $k$  such that

$$E[\bar{N}_n(\tau_{k,n}) - N_n(\tau_{k,n})] \leq \sum_{m \geq k} \frac{q^m}{p} = \frac{q^k}{p^2} \leq \varepsilon \quad \text{and} \quad P(\bar{N}_n(\tau_{k,n}) = N_n(\tau_{k,n})) \geq 1 - \varepsilon.$$

Combined with (34) we obtain that also for  $\bar{N}_n$  the size of the first jump to 1 converges in distribution with  $n \rightarrow \infty$ .

Now consider a representation  $\bar{N}_n = U_n + V_n$  with random variables  $U_n(0)$  and  $V_n(0)$  to be specified below, where at the times  $\sigma_i$  both  $U_n$  and  $V_n$  are thinned independently according to  $p$  and then  $V_n$  is enlarged by 1. Note that for independent  $U_n(0)$  and  $V_n(0)$  the Markov chains  $U_n$  and  $V_n$  are independent as well. Also  $U_n$  converges a.s. to zero, whereas  $V_n$  is an aperiodic, irreducible chain, which is positive recurrent in view of  $E[V_n(\sigma_{m+1}) - V_n(\sigma_m) | V_n(\sigma_m)] = 1 - pV_n(\sigma_m)$  a.s. Let  $\pi$  be its stationary distribution.

Let us study the case  $\bar{N}^\lambda = U^\lambda + V$  with independent Markov chains  $U^\lambda$  and  $V$ , both with the dynamics described above, where now  $U^\lambda(0)$  is Poisson( $e^\lambda$ )-distributed with  $\lambda \in \mathbb{R}$  and  $V(0)$  has the distribution  $\pi$ . Since  $p = e^{-1}$ , the random variable  $U^\lambda(\sigma_m)$  is Poisson( $e^{\lambda-m}$ )-distributed. Let  $\rho = \inf\{t : \bar{N}^\lambda(t) = 1\}$  and  $\rho' = \inf\{t : U^\lambda(t) = 0\}$ . Note that  $\rho' \leq \rho$ .

We now focus on the event  $\{\bar{N}^\lambda(\rho-) = 2\}$ . It can occur in two different ways, either  $\rho' = \rho$  or  $\rho' < \rho$ . The first instance takes place if and only if for some  $m \geq 0$  we have  $U^\lambda(\sigma_m) = 1$ ,  $U^\lambda(\sigma_{m+1}) = 0$ , and  $V(\sigma_m) = V(\sigma_{m+1}) = 1$ . By independence this event has probability

$$\pi(1)e^{-1} \sum_{m=0}^{\infty} e^{-e^{\lambda-m}} e^{\lambda-m} e^{-1}.$$

In case of the event  $\{\rho' < \rho\}$  we have  $V(\rho') \geq 2$  and  $V(\rho-) = 2$ . This will occur if and only if, defining  $h$  so that  $\rho' = \sigma_h$ , we have for some  $\ell > h$  that  $V(\sigma_i) \geq 2$  for  $i = h, h + 1, \dots, \ell - 2$ ,  $V(\sigma_{\ell-1}) = 2$ , and  $V(\sigma_\ell) = 1$ . By applying the strong Markov property at time  $\sigma_h$  and using the independence of the two chains, we see that, letting  $\sigma_0 = 0$ , the probability that this occurs is

$$\alpha := P(V(\sigma_0), \dots, V(\sigma_{\ell-2}) \geq 2, V(\sigma_{\ell-1}) = 2, V(\sigma_\ell) = 1 \text{ for some } \ell \geq 1).$$

Replacing  $\lambda$  by  $\lambda + n$  and letting  $n \rightarrow \infty$  we obtain

$$\lim_{n \rightarrow \infty} P(\bar{N}^{\lambda+n}(\rho-) = 2) = \alpha + \pi(1)e^{-2}f(\lambda) \quad \text{with } f(\lambda) := \sum_{m=-\infty}^{\infty} e^{-e^{\lambda-m}} e^{\lambda-m}.$$

The function  $f$  is smooth with period 1. By our assumption that  $L_n$  converges in distribution as  $n \rightarrow \infty$ , the function  $f$  does not depend on  $\lambda$ . To get a contradiction we compute its Fourier coefficients. They are given by

$$\hat{f}(k) = \int_{-\infty}^{\infty} e^{-e^\lambda} e^\lambda e^{-2\pi i k \lambda} d\lambda = E[e^{-2\pi i k G}],$$

where the distribution of  $G$  is standard Gumbel. The characteristic function of the standard Gumbel distribution is equal to  $\varphi(t) = \Gamma(1 - it)$ ,  $t \in \mathbb{R}$ . Also the gamma function is known to possess no zeros in the complex plane, thus none of the Fourier coefficients of  $f$  vanishes. Therefore  $f$  is non-constant, and we arrive at the promised contradiction.

### 6. Proof of Theorem 3

Our proof of Theorem 3 relies on an overshoot estimate for subordinators. The Renewal Theorem for subordinators (see, for example, Corollary 5.3 in [9]) implies that if  $(S(t), t \geq 0)$  is a subordinator and  $E[S(1)] = \infty$ , then for all  $y > 0$ ,

$$\lim_{x \rightarrow \infty} P(S(t) \in [x, x + y] \text{ for some } t) = 0.$$

To prove Theorem 3, we will need to establish a version of this result which holds for processes that can be obtained by adding a small state-dependent negative drift to a subordinator.

**Proposition 17.** *Let  $(S_t, t \geq 0)$  be a subordinator with  $E[S_1] = \infty$ . Let  $g : \mathbb{R} \rightarrow \mathbb{R}^+$  be a nonincreasing function such that*

$$\lim_{x \rightarrow \infty} g(x) = 0. \tag{35}$$

For all  $z > 0$ , define the process  $(Y_t^z)_{t \geq 0}$  to be the solution to the SDE

$$Y_t^z = z - \left( S_t - \int_0^t g(Y_s^z) ds \right). \tag{36}$$

For all  $y \in \mathbb{R}$ , let  $\tau_y^z = \inf\{t \geq 0 : Y_t^z \leq y\}$ . Then for all real numbers  $K > 0$ , we have

$$\lim_{z \rightarrow \infty} P(Y_{\tau_K^z}^z \in [-K, K]) = 0. \tag{37}$$

Equation (37) says that for any bounded interval the probability that  $Y^z$  jumps over the interval  $[-K, K]$  tends to one as the starting point  $z \rightarrow \infty$ .

**Proof.** We will prove this result by following some of the ideas from [3] in the proof of Blackwell’s Renewal Theorem in the infinite mean case. Let  $\beta_K^z = P(Y_{\tau_K^z}^z \in [-K, K])$ , and let

$$\beta_K = \limsup_{z \rightarrow \infty} \beta_K^z. \tag{38}$$

Seeking a contradiction, suppose  $\beta_K > 0$  for some  $K$ . Because  $\beta_K$  is a nondecreasing function of  $K$ , it suffices to obtain a contradiction when  $K$  is chosen to be a sufficiently large positive integer. We will choose  $K$  to be large enough to satisfy the following four conditions:

1. We require  $g(K) < K$ , which is true for sufficiently large  $K$  by (35).
2. We require

$$P(S_t \in (2(\ell - 1)K, 2\ell K] \text{ for some } t \geq 0) > 0 \tag{39}$$

for all positive integers  $\ell$ . Note that (39) may fail for small values of  $K$ , in particular when  $S_1$  has a lattice distribution, but will hold for sufficiently large  $K$ .

3. We require

$$P\left(\sup_{t \geq 0} (g(K)t - S_t) > 1\right) < \frac{1}{2}. \tag{40}$$

Note that this holds for sufficiently large  $K$  in view of (35) and the fact that  $t^{-1}S_t \rightarrow \infty$  as  $t \rightarrow \infty$  by the Law of Large Numbers for subordinators.

4. Let

$$\alpha_K = E[\inf\{t \geq 0 : S_t - g(K)t \geq 2\}], \tag{41}$$

which tends to a finite limit as  $K \rightarrow \infty$  by (35). We require

$$\frac{2\alpha_K(8K + 1)g(K)}{K} \leq \frac{\beta_K}{3}. \tag{42}$$

If  $\beta_K > 0$  for some  $K$ , then this condition holds for sufficiently large  $K$  by (35) and the fact that  $\beta_K$  is a nondecreasing function of  $K$ .

Because (37) does not depend on the behavior of the process after time  $\tau_K^z$ , we may consider instead the processes  $(Z_t^z)_{t \geq 0}$ , defined as the solution to the SDE

$$Z_t^z = z - \left( S_t - \int_0^{t \wedge \tau_K^z} g(Z_s^z) ds \right). \tag{43}$$

The processes  $Z^z$  and  $Y^z$  are the same until time  $\tau_K^z$ , which implies that

$$\beta_K^z = P(Y_{\tau_K^z}^z \in [-K, K]) = P(Z_{\tau_K^z}^z \in [-K, K]).$$

However, after time  $\tau_K^z$  the process  $Z^z$  is no longer affected by the drift term involving  $g$ . Because  $g$  is nonincreasing, we have  $Z_t^z \leq z - S_t + g(K)t$  for all  $t \geq 0$ . Therefore, (40) implies that

$$P\left(\sup_{t \geq 0} Z_t^z > z + 1\right) < \frac{1}{2}. \tag{44}$$

Let  $U^z$  denote the potential measure associated with the process  $Z^z$ , meaning that

$$U^z(A) = \int_0^\infty P(Z_t^z \in A) dt$$

for all Borel subsets  $A$  of  $\mathbb{R}$ . Suppose  $z > K$ , and  $n > K$  is a positive integer. If the process  $Z^z$  enters the interval  $(n - 1, n]$ , then it drops below  $n - 2$  after a time whose expectation is at most  $\alpha_K$ , and then by (44) and the strong Markov property, the probability that the process  $Z^z$  never returns to  $(n - 1, n]$  is at least  $1/2$ . It follows that

$$U_z((n - 1, n]) \leq 2\alpha_K. \tag{45}$$

Let  $0 < H_1 < H_2 < \dots$  denote the points of a rate one Poisson process, independent of  $(S_t)_{t \geq 0}$ . Note that the process  $(Z_{H_n}^z)_{n=1}^\infty$  has the same potential measure as  $(Z_t^z)_{t \geq 0}$ , in the sense that for all Borel subsets  $A$  of  $\mathbb{R}$ ,

$$U^z(A) = \sum_{n=1}^\infty P(Z_{H_n}^z \in A).$$

We can choose an increasing sequence  $(z_m)_{m=1}^\infty$  tending to infinity such that

$$\lim_{m \rightarrow \infty} \beta_K^{z_m} = \beta_K. \tag{46}$$

It follows from (43) and the monotonicity of  $g$  that

$$z_m - S_{H_1} \leq Z_{H_1}^{z_m} \leq z_m + g(z_m - S_{H_1})H_1. \tag{47}$$

Let  $\varepsilon > 0$ . Choose a positive integer  $L$  large enough that  $P(S_{H_1} \geq 2LK) < \varepsilon$ . By (35) we can choose a positive integer  $m_0$  large enough that for all  $m \geq m_0$

$$P(g(z_m - S_{H_1})H_1 \geq 2K) < \varepsilon.$$

This together with (47) implies for all

$$P(z_m - 2LK \leq Z_{H_1}^{z_m} \leq z_m + 2K) \geq 1 - 2\varepsilon.$$

For the following we also require that  $z_{m_0} - 2LK > K$ .

Let  $\mu^z$  denote the distribution of  $Z_{H_1}^z$ . By applying the strong Markov property at time  $H_1$ , we get for  $m \geq m_0$ ,

$$\beta_K^{z_m} \leq \sum_{\ell=0}^L \int_{[z_m-2\ell K, z_m-2(\ell-1)K)} \beta_K^x \mu^{z_m}(dx) + 2\varepsilon. \tag{48}$$

Write

$$a_{m,\ell} = \int_{[z_m-2\ell K, z_m-2(\ell-1)K)} \beta_K^x \mu^{z_m}(dx). \tag{49}$$

It follows from (46) and (48) that

$$\beta_K - 2\varepsilon \leq \liminf_{m \rightarrow \infty} \sum_{\ell=0}^L a_{m,\ell} \leq \limsup_{m \rightarrow \infty} \sum_{\ell=0}^L a_{m,\ell} \leq \beta_K. \tag{50}$$

By (35), for all  $\ell \in \{0, 1, \dots, L\}$  we have

$$\lim_{m \rightarrow \infty} P(Z_{H_1}^{z_m} \in [z_m - 2\ell K, z_m - 2(\ell - 1)K]) = P(S_{H_1} \in (2(\ell - 1)K, 2\ell K]). \tag{51}$$

It follows from (38) and (51) that for  $\ell \in \{0, 1, \dots, L\}$ , we have

$$\limsup_{m \rightarrow \infty} a_{m,\ell} \leq \beta_K P(S_{H_1} \in (2(\ell - 1)K, 2\ell K]),$$

and then (50) yields

$$\liminf_{m \rightarrow \infty} a_{m,\ell} \geq \beta_K P(S_{H_1} \in (2(\ell - 1)K, 2\ell K]) - 2\varepsilon.$$

By taking  $\varepsilon \rightarrow 0$ , we see that for any fixed nonnegative integer  $\ell$ , we have

$$\lim_{m \rightarrow \infty} a_{m,\ell} = \beta_K P(S_{H_1} \in (2(\ell - 1)K, 2\ell K]). \tag{52}$$

Now we also see from (49) and (51) that

$$\liminf_{m \rightarrow \infty} a_{m,\ell} \leq \left( \liminf_{m \rightarrow \infty} \sup_{x \in [z_m-2\ell K, z_m-2(\ell-1)K)} \beta_K^x \right) P(S_{H_1} \in (2(\ell - 1)K, 2\ell K]).$$

In view of (39) and (52), it follows that for all  $\ell \in \{1, \dots, L\}$  and therefore for all positive integers  $\ell$ , we have

$$\liminf_{m \rightarrow \infty} \sup_{x \in [z_m-2\ell K, z_m-2(\ell-1)K)} \beta_K^x = \beta_K. \tag{53}$$

Fix a positive integer  $M$ . By (46) and (53), we can choose  $m$  sufficiently large that  $\beta_K^{z_m} > 2\beta_K/3$  and for  $\ell \in \{1, \dots, 3M\}$ , there exists a point  $x_\ell \in [z_m - 2\ell K, z_m - 2(\ell - 1)K)$  such that  $\beta_K^{x_\ell} > 2\beta_K/3$ . Set  $x_0 = z_m$ . We now consider the processes  $Z^{x_0}, Z^{x_3}, Z^{x_6}, \dots, Z^{x_{3M}}$ , which satisfy the stochastic differential equation (43) with the same driving subordinator but different initial values. For  $1 \leq \ell \leq M$ , we have

$$4K \leq Z_0^{x_{3(\ell-1)}} - Z_0^{x_{3\ell}} \leq 8K. \tag{54}$$

Because  $g$  is nonincreasing, the processes  $Z^{x_{3(\ell-1)}}$  and  $Z^{x_{3\ell}}$  get closer together over time but do not cross, which means

$$0 \leq Z_t^{x_{3(\ell-1)}} - Z_t^{x_{3\ell}} \leq 8K \tag{55}$$

for all  $t \in [0, \tau_K^{x_{3\ell}}]$ . Thus,

$$\begin{aligned} \int_0^{\tau_K^{x_{3\ell}}} |g(Z_t^{x_{3\ell}}) - g(Z_t^{x_{3(\ell-1)}})| dt &\leq \sum_{n=0}^{\infty} \int_0^{\tau_K^{x_{3\ell}}} |g(Z_t^{x_{3\ell}}) - g(Z_t^{x_{3(\ell-1)}})| \mathbb{1}_{\{Z_t^{x_{3\ell}} \in (K+n, K+n+1]\}} dt \\ &\leq \sum_{n=0}^{\infty} \int_0^{\tau_K^{x_{3\ell}}} |g(K+n) - g(K+n+1+8K)| \mathbb{1}_{\{Z_t^{x_{3\ell}} \in (K+n, K+n+1]\}} dt. \end{aligned}$$

In view of (45), we get a telescoping sum, and

$$\begin{aligned} E \left[ \int_0^{\tau_K^{x_{3\ell}}} |g(Z_t^{x_{3(\ell-1)}}) - g(Z_t^{x_{3\ell}})| dt \right] &\leq 2\alpha_K \sum_{n=0}^{\infty} (g(K+n) - g(K+n+1+8K)) \\ &\leq 2\alpha_K \sum_{n=0}^{8K} g(K+n) \\ &\leq 2\alpha_K (8K+1)g(K). \end{aligned} \tag{56}$$

Let  $D_\ell$  be the event that

$$\int_0^{\tau_K^{x_{3\ell}}} |g(Z_t^{x_{3(\ell-1)}}) - g(Z_t^{x_{3\ell}})| dt \leq K.$$

By Markov's Inequality and (56),

$$P(D_\ell^c) \leq \frac{2\alpha_K(8K+1)g(K)}{K}. \tag{57}$$

It follows from (54) that on the event  $D_\ell$ , we have  $Z_t^{x_{3(\ell-1)}} - Z_t^{x_{3\ell}} \geq 3K$  for all  $t \in [0, \tau_K^{x_{3\ell}}]$ . Furthermore, after time  $\tau_K^{x_{3\ell}}$ , the process  $Z^{x_{3\ell}}$  is no longer affected by the drift term involving  $g$ , and thus it decreases at least as fast as  $Z^{x_{3(\ell-1)}}$ . It follows that on  $D_\ell$ , we have  $Z_t^{x_{3(\ell-1)}} - Z_t^{x_{3\ell}} \geq 3K$  for all  $t \geq 0$ , and thus the process  $Z^{x_{3\ell}}$  can not be in the interval  $[-(K+1), K]$  at the same time as  $Z^{x_{3(\ell-1)}}$  or any other process  $Z^{x_{3j}}$  with  $j < \ell$ . Let

$$I_\ell = \begin{cases} \{t \geq 0 : -(K+1) \leq Z_t^{x_{3\ell}} \leq K \text{ and } \tau_K^{x_{3\ell}} \leq t \leq \tau_K^{x_{3\ell}} + 1\} & \text{on } D_\ell, \\ \emptyset & \text{on } D_\ell^c. \end{cases}$$

The discussion above implies that the sets  $I_\ell$  are disjoint. Let

$$\kappa = E[1 \wedge \inf\{t : S_t > 1\}].$$

Given the event  $D_\ell \cap \{Z_{\tau_K^{x_{3\ell}}} \in [-K, K]\}$ , the expected Lebesgue measure of  $I_\ell$  is at least  $\kappa$ . Therefore, using (57) and the fact that  $\beta_K^{x_{3\ell}} > 2\beta_K/3$  followed by (42), we get

$$E \left[ \int_0^\infty \mathbb{1}_{\{t \in I_\ell\}} dt \right] \geq \kappa \left( \frac{2\beta_K}{3} - \frac{2\alpha_K(8K+1)g(K)}{K} \right) \geq \frac{\kappa\beta_K}{3}.$$

On the event that  $Z_{\tau_K^{x_{3\ell}}}^{x_{3\ell}} \in [-K, K]$ , because of (55), we have  $Z_{\tau_K^{x_{3\ell}}}^{z_m} \leq (8\ell + 1)K$ . During the next time unit, the process  $Z^{z_m}$  can increase by at most  $g(K)$ , so if  $t \in I_\ell$ , then using that  $g(K) < K$ , we get

$$Z_t^{z_m} \leq (8\ell + 1)K + g(K) \leq 10\ell K.$$

We next note that if  $t \in I_\ell$  then  $Z_t^{z_m} \geq K$  because  $Z_t^{z_m} - Z_t^{x_{3\ell}} \geq 3K$  as described above. It follows that

$$U^{z_m}([K, 10\ell K]) = E \left[ \int_0^\infty \mathbb{1}_{\{K \leq Z_t^{z_m} \leq 10\ell K\}} dt \right] \geq \sum_{j=1}^\ell E \left[ \int_0^\infty \mathbb{1}_{\{t \in I_j\}} dt \right] \geq \frac{\kappa\beta_K \ell}{3},$$

and therefore if  $y \geq 10K$ , then

$$U^{z_m}([K, y]) \geq \frac{\kappa\beta_K y}{60K}. \tag{58}$$

Because the process  $(Z_{H_n}^{z_m})_{n=0}^\infty$  is decreasing after it drops below the level  $K$ , it can only jump below zero one time. In particular, the expected number of times the process jumps below zero is bounded above by one. Therefore, letting  $\nu_x$  denote the conditional distribution of  $Z_{H_n}^{z_m} - Z_{H_{n+1}}^{z_m}$  given  $Z_{H_n} = x$ , we have

$$1 \geq \int_K^\infty \nu_x([x, \infty)) U^{z_m}(dx) \geq \int_K^{3M} \nu_x([x, \infty)) U^{z_m}(dx).$$

Let  $\mu$  denote the distribution of the random variable  $S_{H_1} - H_1 g(K)$ . Because  $g$  is decreasing, we have  $\nu_x([x, \infty)) \geq \mu([x, \infty))$  for all  $x \geq K$ . Therefore,

$$1 \geq \int_K^{3M} \mu([x, \infty)) U^{z_m}(dx) = \int_K^\infty \int_K^{y \wedge 3M} U^{z_m}(dx) \mu(dy) \geq \int_K^{3M} U^{z_m}([K, y]) \mu(dy).$$

Combining this result with (58) gives

$$1 \geq \frac{\kappa\beta_K}{60K} \int_{10K}^{3M} y \mu(dy).$$

Because  $E[S_1] = \infty$ , we have  $E[S_{H_1} - H_1 g(K)] = \infty$ , so the right-hand side is bigger than one for sufficiently large positive integers  $M$ , a contradiction.  $\square$

**Proof of Theorem 3.** Let  $K \geq 2$  be a positive integer. If  $2 \leq N_n(T_n-) \leq K$  and the event in (15) holds, then

$$-L + \log 2 \leq \log n - \left( S(T_n-) - \int_0^{T_n} f(Y_n(s)) ds \right) \leq L + \log K, \tag{59}$$

with  $f$  defined in (11), and the left inequality holds with  $T_n-$  replaced by any  $t \in [0, T_n)$ . In particular, putting  $K' := L + \log K$ , we have

$$-K' \leq Y_n(t) \quad \text{for all } t \in [0, T_n). \tag{60}$$

The right inequality in (59) says that  $Y_n(T_n-) \leq K'$ . With  $z := \log n$  we have  $Y_n(t) = Y_t^z$  in the notation of Proposition 17, hence  $\tau_{K'}^z < T_n$ . Thus  $-K' \leq Y_{\tau_{K'}^z}^z$  by (60). On the other hand we have  $Y_{\tau_{K'}^z}^z \leq K'$  by definition, and consequently  $Y_{\tau_{K'}^z}^z \in [-K', K']$ . Note that  $E[S_1] = \infty$  by (5). Therefore, combining (37) and (15) we see that  $P(N_n(T_n-) \leq K) \rightarrow 0$  as  $n \rightarrow \infty$ , which proves Theorem 3.  $\square$

**7. Proof of Theorems 4 and 5**

We prepare the proof of Theorem 4 by a few lemmas.

**Lemma 18.** *Let  $i \geq 2$  and  $\varepsilon > 0$ . Then there is a  $k > i$  such that for all  $n$*

$$P(L_n = i, N_n(t) \notin [i + 1, k] \text{ for all } t \geq 0) \leq \varepsilon.$$

**Proof.** Without loss of generality  $\Lambda(\{0\}) = 0$ , because otherwise the coalescent comes down from infinity, and the claim is immediate.

Recall the definition of  $\tau_{k,n}$  in (13). We have

$$P(L_n = i, N_n(t) \notin [i + 1, k] \text{ for all } t \geq 0) \leq P(N_n(\tau_{\ell,n}) = i)$$

with  $\ell = k + 1$ . As before, let  $(t_m, p_m)$ ,  $m \geq 1$ , be the first two coordinates of the points of  $\Psi$  in an arbitrary order. Denote  $\tilde{p} := p_m$  if  $t_m = \tau_{\ell,n}$ . Define for  $\kappa = 6i/\varepsilon$  the events

$$A := \left\{ \exists i \geq 1 : t_m \leq \tau_{\ell,n}, \frac{1}{\kappa N_n(t_m-)} < 1 - p_i \leq \frac{\kappa}{N_n(t_m-)} \right\},$$

$$B := \left\{ 1 - \tilde{p} \leq \frac{1}{\kappa N_n(\tau_{\ell,n}-)} \right\},$$

$$C := \left\{ 1 - \tilde{p} > \frac{\kappa}{N_n(\tau_{\ell,n}-)} \right\}.$$

Then

$$P(N_n(\tau_{\ell,n}) = i) \leq P(A) + P(\{N_n(\tau_{\ell,n}) = i\} \cap B) + P(\{N_n(\tau_{\ell,n}) = i\} \cap C). \tag{61}$$

We estimate these probabilities.

First denote

$$\sigma_j = \tau_{n/\kappa^j, n}, \quad j = 0, 1, \dots,$$

and let  $r$  be the smallest integer such that  $n/\kappa^r \leq \ell$ . Then

$$P(A) \leq \sum_{j=0}^{r-1} P\left( \exists i : \sigma_j < t_m \leq \sigma_{j+1}, \frac{1}{\kappa N_n(t_m-)} < 1 - p_m \leq \frac{\kappa}{N_n(t_i-)} \right)$$

$$\leq \sum_{j=0}^{r-1} P\left( \exists i : \sigma_j < t_m \leq \sigma_{j+1}, \frac{\kappa^{j-1}}{n} < 1 - p_i \leq \frac{\kappa^{j+2}}{n} \right).$$

From Lemma 7 we have  $E[\sigma_{j+1} - \sigma_j] \leq C_\kappa$  for a suitable constant  $C_\kappa$  depending on  $\kappa$ , thus

$$P(A) \leq \sum_{j=0}^{r-1} E[\sigma_{j+1} - \sigma_j] \int_{[1-\kappa^{j+2}/n, 1-\kappa^{j-1}/n]} \frac{\Lambda(dp)}{p^2}$$

$$\leq 3C_\kappa \int_{[1-\kappa^{r+1}/n, 1]} \frac{\Lambda(dp)}{p^2}$$

$$\leq 3C_\kappa \int_{[1-\kappa/\ell, 1]} \frac{\Lambda(dp)}{p^2}.$$



Thus, if we choose  $\ell$  sufficiently large we obtain

$$P(A) \leq \frac{\varepsilon}{3}. \quad (62)$$

Second we have on the event  $B$  with  $b = N_n(\tau_{\ell,n-})$

$$\frac{\tilde{p}}{(b-i+2)(1-\tilde{p})} \geq \frac{1}{b} \left( \frac{1}{1-\tilde{p}} - 1 \right) \geq \frac{b\kappa - 1}{b} \geq \frac{\kappa}{2}$$

and consequently on the event  $B$  with  $\alpha = \sum_{j < \ell} \binom{b}{j-1} (1-\tilde{p})^{j-1} \tilde{p}^{b-j+1}$

$$\begin{aligned} & P(\{N_n(\tau_{\ell,n}) = i\} \cap B | \tilde{p}, N_n(\tau_{\ell,n-}) = b, B) \\ &= \frac{1}{\alpha} \binom{b}{i-1} (1-\tilde{p})^{i-1} \tilde{p}^{b-i+1} \\ &= \frac{b-i+2}{i-1} \frac{1-\tilde{p}}{\tilde{p}} P(\{N_n(\tau_{\ell,n}) = i-1\} \cap B | \tilde{p}, N_n(\tau_{\ell,n-}) = b, B) \\ &\leq \frac{2}{\kappa(i-1)}. \end{aligned}$$

Thus, since  $\kappa \geq 6/\varepsilon$

$$P(\{N_n(\tau_{\ell,n}) = i\} \cap B) \leq \frac{\varepsilon}{3}. \quad (63)$$

Third we have on the event  $C$ , again with  $b = N_n(\tau_{\ell,n-})$  and with  $b \geq 2i$

$$\frac{\tilde{p}}{(b-i+1)(1-\tilde{p})} \leq \frac{2}{b} \frac{1}{1-\tilde{p}} \leq \frac{2}{\kappa}$$

and consequently for  $\ell \geq 2i$

$$\begin{aligned} & P(\{N_n(\tau_{\ell,n}) = i\} \cap C | \tilde{p}, N_n(\tau_{\ell,n-}) = b, C) \\ &= \frac{i}{b-i+1} \frac{\tilde{p}}{1-\tilde{p}} P(\{N_n(\tau_{\ell,n}) = i+1\} \cap C | \tilde{p}, N_n(\tau_{\ell,n-}) = b, C) \\ &\leq \frac{2i}{\kappa} \end{aligned}$$

implying

$$P(\{N_n(\tau_{\ell,n}) = i\} \cap C) \leq \frac{\varepsilon}{3} \quad (64)$$

for  $\kappa = 6i/\varepsilon$ . Now from (61), (62), (63) and (64) our claim follows.  $\square$

Recall that  $\rho_{ij}$  denotes the rate for a jump of  $N_n$  from state  $i$  to  $j$ , and  $\rho_i$  is the rate at which  $N_n$  leaves  $i$ . Next let for  $n \in \mathbb{N}$

$$\mu_i^{(n)} := \frac{1}{\rho_i} P(N_n(t) = i \text{ for some } t \geq 0).$$

Also let

$$P_{ij} := \frac{\rho_{ij}}{\rho_i}, \quad 1 \leq j < i,$$

be the transition probability from state  $i$  to  $j$  of the block-counting process of our  $\Lambda$ -coalescent.

**Lemma 19.** Suppose that there are numbers  $\mu_i$ ,  $i \geq 2$ , not all vanishing, such that for some increasing sequence  $(n_m)_{m \geq 1}$  of natural numbers, as  $m \rightarrow \infty$ ,

$$\mu_i^{(n_m)} \rightarrow \mu_i$$

for all  $i \geq 2$ . Then the measure  $\mu = (\mu_i)_{i \geq 2}$  is  $\rho$ -invariant, i.e. satisfies the first condition in (3).

**Proof.** First we have for  $i \geq 2$

$$\mu_i^{(n)} \rho_{i1} = P(N_n(t) = i \text{ for some } t \geq 0) P_{i1} = P(L_n = i)$$

and therefore in the limit (along the specified sequence) by Fatou's Lemma

$$\sum_{i \geq 2} \mu_i \rho_{i1} \leq 1.$$

Second for  $2 \leq i < k$

$$\begin{aligned} P(L_n = i, N_n(t) \notin [i+1, k] \text{ for all } t \geq 0) &= \sum_{j > k} P(N_n(t) = j \text{ for some } t \geq 0) P_{ji} P_{i1} \\ &= \sum_{j > k} \mu_j^{(n)} \rho_{ji} P_{i1}. \end{aligned}$$

Applying Lemma 18 to the left-hand term it follows that for any  $\varepsilon > 0$  there is a  $k$  such that for all  $n$

$$\sum_{j > k} \mu_j^{(n)} \rho_{ji} \leq \varepsilon.$$

Therefore we may proceed in the equation

$$\mu_i^{(n)} \rho_i = \sum_{j=i+1}^n \mu_j^{(n)} \rho_{ji}$$

along the given subsequence to the limit to obtain

$$\mu_i \rho_i = \sum_{j=i+1}^{\infty} \mu_j \rho_{ji}, \quad i \geq 2. \tag{65}$$

Thus  $\mu$  is  $\rho$ -invariant. □

**Lemma 20.** Let  $\nu = (\nu_i)_{i \geq 2}$  be a measure satisfying (3). Then for any integer  $a \geq 1$  there are probability measures  $\omega_a = (\omega_{i,a})_{1 \leq i \leq a}$  on  $\{1, \dots, a\}$  such that for any  $i \geq 2$  we have  $\omega_{i,a} \rightarrow 0$  as  $a \rightarrow \infty$ , and for  $1 \leq i \leq a$

$$\nu_i = \sum_{n=i}^a \mu_i^{(n)} \omega_{n,a}. \tag{66}$$

**Proof.** Denote for  $i, j \geq 1$

$$\tilde{P}_{ij} := \frac{\nu_j \rho_{ji}}{\nu_i \rho_i} = \frac{\nu_j \rho_j}{\nu_i \rho_i} P_{ji},$$

where we set the undefined quantity  $v_1 \rho_1$  equal to 1. Then the  $\rho$ -invariance and the norming of  $\nu$  (according to the second condition in (3)) implies  $\sum_j \tilde{P}_{ij} = 1$  for  $i \geq 1$ . Thus we may consider the Markov chain  $(\tilde{X}_r)_{r=0,1,\dots}$  on  $\mathbb{N}$  with initial state  $\tilde{X}_0 = 1$  and transition matrix  $(\tilde{P}_{ij})$ . We claim that it fulfils the equation

$$v_j \rho_j = P(\tilde{X}_r = j \text{ for some } r), \quad j \geq 1.$$

We show this claim by induction. For  $j = 1$  both terms are equal to 1. Suppose that it holds for  $1 \leq i \leq j - 1$ . Then

$$P(\tilde{X}_r = j \text{ for some } r) = \sum_{i=1}^{j-1} v_i \rho_i \tilde{P}_{ij} = v_j \rho_j \sum_{i=1}^{j-1} P_{ji} = v_j \rho_j.$$

Next define for an integer  $a > 1$  the random times

$$\xi_a := \max\{r \geq 0 : \tilde{X}_r \leq a\}$$

and for  $1 \leq i < a$

$$\eta_{ia} := P(\tilde{X}_1 > a | \tilde{X}_0 = i).$$

Then for  $a > 1$  and  $1 = i_0 < i_1 < i_2 < \dots < i_r \leq a$

$$\begin{aligned} P(\tilde{X}_0 = i_0, \tilde{X}_1 = i_1, \tilde{X}_2 = i_2, \dots, \tilde{X}_r = i_r, \xi_a = r) &= \tilde{P}_{1i_1} \tilde{P}_{i_1 i_2} \dots \tilde{P}_{i_{r-1} i_r} \eta_{i_r a} \\ &= \omega_{i_r, a} P_{i_r i_{r-1}} \dots P_{i_2 i_1} P_{i_1 1} \end{aligned}$$

with

$$\omega_{i,a} := v_i \rho_i \eta_{ia}, \quad 1 \leq i < a$$

and  $i_r = 1$  in the case  $r = 0$  (then both products of transition probabilities are set to be 1). For fixed  $i$ , summing over  $1 < i_1 < i_2 < \dots < i_r := i \leq a$  and  $r \geq 0$  we obtain the equality  $P(\tilde{X}_{\xi_a} = i) \eta_{ia} = \omega_{i,a}$ , and thus  $\sum_{1 \leq i \leq a} \omega_{i,a} = 1$ . Therefore we may view the time-reversed process  $Y_0 = \tilde{X}_{\xi_a}, Y_1 = \tilde{X}_{\xi_a - 1}, \dots, Y_{\xi_a} = \tilde{X}_0$  as a Markov chain on  $\{1, \dots, a\}$  with initial distribution  $\omega_a$ , transition probabilities  $P_{ij}$  and killed after reaching 1. This process coincides in distribution with the block-counting process of our original coalescent in discrete time, now with initial distribution  $\omega_a$ . This gives another way to express  $\nu_i$ : For  $1 \leq i < a$

$$\rho_i \nu_i = P(Y_r = i \text{ for some } r \leq \xi_a) = \sum_{n=i}^{a-1} \rho_i \mu_i^{(n)} \omega_{n,a},$$

which is (66). Also  $\eta_{ia} \rightarrow 0$  for  $a \rightarrow \infty$ , which implies  $\omega_{i,a} \rightarrow 0$ . Thus the proof is finished. □

**Proof of Theorem 4.** (i) Let  $i \geq 2$ . If  $L_n \rightarrow \infty$  in probability, then as  $n \rightarrow \infty$

$$\mu_i^{(n)} = \frac{P(L_n = i)}{\rho_{i1}} \rightarrow 0.$$

Now suppose that there is a measure  $\nu$  satisfying (3). Then we may apply Lemma 20. Let  $\varepsilon > 0$  and  $b > i$  such that  $\mu_i^{(n)} \leq \varepsilon$  for  $n > b$ . From (66) for  $a > b$

$$\nu_i \leq \sum_{n=i}^b \frac{1}{\rho_{i1}} \omega_{n,a} + \sum_{n=b+1}^a \varepsilon \omega_{n,a} \leq \sum_{n=i}^b \frac{1}{\rho_{i1}} \omega_{n,a} + \varepsilon.$$

In the limit  $a \rightarrow \infty$ , since  $\omega_{n,a} \rightarrow 0$  for fixed  $n$ , we obtain  $\nu_i \leq \varepsilon$ . Thus  $\nu_i = 0$  for all  $i \geq 2$ , which is a contradiction. Hence there is no solution to (3).

(ii) Now by assumption there is an increasing sequence of natural numbers  $n_m, m \geq 1$ , such that as  $m \rightarrow \infty$

$$\mu_i^{(n_m)} = \frac{P(L_{n_m} = i)}{\rho_{i1}} \rightarrow \alpha \frac{\pi_i}{\rho_{i1}}$$

for all  $i \geq 2$  and for some  $\alpha > 0$ . From Lemma 19 it follows that  $\mu_i := \pi_i/\rho_{i1}$  are the weights of a  $\rho$ -invariant measure  $\mu$ .

Now let  $\nu$  be any solution of (3). By assumption we have  $\mu_i^{(n)} \sim \frac{\mu_i}{\mu_2} \mu_2^{(n)}$  as  $n \rightarrow \infty$ . Therefore from Lemma 20 it follows by a similar argument as in the proof of (i) that, as  $a \rightarrow \infty$ ,

$$\nu_i = \sum_{n=i}^a \mu_i^{(n)} \omega_{n,a} \sim \sum_{n=i}^a \frac{\mu_i}{\mu_2} \mu_2^{(n)} \omega_{n,a} \sim \frac{\mu_i}{\mu_2} \sum_{n=2}^a \mu_2^{(n)} \omega_{n,a} = \frac{\mu_i}{\mu_2} \nu_2.$$

This shows that  $\nu$  is a multiple of  $\mu$ .

(iii) In the remaining situation by means of a diagonal argument there are two increasing sequences such that  $\mu_i^{(n)}$  converges along both sequences for all  $i \geq 2$ , but now the limiting measures are not multiples of each other. Thus another application of Lemma 19 gives the claim. This finishes the proof. □

**Proof of Theorem 5.** Let  $0 = \gamma_0 < \gamma_1 < \dots < \gamma_{\zeta_n} = T_n$  be the jump times of  $\hat{N}_n$  and let  $\Delta_i := \gamma_{i+1} - \gamma_i$  the interim times. For the proof it is now sufficient to show for fixed  $r \geq 1$  convergence in distribution of the random vectors  $(\hat{N}_n(0), \Delta_0, \dots, \hat{N}_n(\gamma_r), \Delta_r)$  to the corresponding limiting distribution. The event  $\{\zeta_n < r\}$  has vanishing probability as  $n \rightarrow \infty$ . In view of the strong Markov property of  $N_n$  as  $n \rightarrow \infty$  we have for  $2 \leq i_0 < i_1 < \dots < i_r < n$

$$\begin{aligned} &P(\hat{N}_n(0) = i_0, \Delta_0 \in dt_0, \dots, \hat{N}_n(\gamma_r) = i_r, \Delta_r \in dt_r) \\ &= P(N_n((T_n - \gamma_r)-) = i_r, \Delta_r \in dt_r, \dots, N_n(T_n-) = i_0, \Delta_0 \in dt_0) \\ &= \mu_{i_r}^{(n)} \rho_{i_r} \cdot e^{-\rho_{i_r} t_r} \cdot \rho_{i_r i_{r-1}} dt_r \dots e^{-\rho_{i_0} t_0} \cdot \rho_{i_0 1} dt_0. \end{aligned}$$

Theorem 4(ii) implies

$$\begin{aligned} &P(\hat{N}_n(0) = i_0, \Delta_0 \in dt_0, \dots, \hat{N}_n(\gamma_r) = i_r, \Delta_r \in dt_r) \\ &\rightarrow \mu_{i_r} \rho_{i_r} \cdot e^{-\rho_{i_r} t_r} \cdot \rho_{i_r i_{r-1}} dt_r \dots e^{-\rho_{i_0} t_0} \cdot \rho_{i_0 1} dt_0. \end{aligned}$$

For  $i < j$  define rates  $\hat{\rho}_{ij}$  and  $\hat{\rho}_i$  by

$$\mu_i \hat{\rho}_{ij} = \mu_j \rho_{ji}, \quad \hat{\rho}_i := \sum_{j>i} \hat{\rho}_{ij}.$$

Since  $\mu$  is  $\rho$ -invariant,

$$\hat{\rho}_i = \frac{1}{\mu_i} \sum_{j>i} \mu_j \rho_{ji} = \rho_i.$$

With these terms the above convergence statement transforms into

$$\begin{aligned} &P(\hat{N}_n(0) = i_0, \Delta_0 \in dt_0, \dots, \hat{N}_n(\gamma_r) = i_r, \Delta_r \in dt_r) \\ &\rightarrow \mu_{i_0} \rho_{i_0 1} \cdot e^{-\hat{\rho}_{i_0} t_0} \cdot \hat{\rho}_{i_0 i_1} dt_0 \dots e^{-\hat{\rho}_{i_{r-1}} t_{r-1}} \cdot \hat{\rho}_{i_{r-1} i_r} dt_{r-1} \cdot e^{-\hat{\rho}_{i_r} t_r} \cdot \hat{\rho}_{i_r} dt_r \\ &= P(\hat{N}_\infty(0) = i_0, \Delta_0 \in dt_0, \dots, \hat{N}_\infty(\gamma_r) = i_r, \Delta_r \in dt_r). \end{aligned}$$

This is our claim. □

## Acknowledgement

We are grateful to a referee for valuable remarks which helped to improve the presentation.

## References

- [1] R. Abraham and J.-F. Delmas. A construction of a  $\beta$ -coalescent via the pruning of binary trees. *J. Appl. Probab.* **50** (2013) 772–790. [MR3102514](#)
- [2] R. Abraham and J.-F. Delmas.  $\beta$ -Coalescents and stable Galton-Watson trees. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** (2015) 451–476. [MR3368966](#)
- [3] R. Durrett. *Probability: Theory and Examples*, 4th edition. Cambridge University Press, Cambridge, 2010. [MR2722836](#)
- [4] B. Eldon and J. Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172** (2006) 2621–2633.
- [5] A. Gnedin, A. Iksanov and A. Marynych. On  $\Lambda$ -coalescents with dust component. *J. Appl. Probab.* **48** (2011) 1133–1151. [MR2896672](#)
- [6] C. Goldschmidt and J. B. Martin. Random recursive trees and the Bolthausen–Sznitman coalescent. *Electron. J. Probab.* **10** (2005) 718–745. [MR2164028](#)
- [7] R. Grübel and K. Hagemann. Leader election: A Markov chain approach. *Math. Appl.* **44** (2016) 113–143. [MR3557093](#)
- [8] O. Hénard. The fixation line in the  $\Lambda$ -coalescent. *Ann. Appl. Probab.* **25** (2015) 3007–3032. [MR3375893](#)
- [9] A. E. Kyprianou. *Fluctuations of Lévy Processes with Applications*, 2nd edition. Springer, Heidelberg, 2014. [MR3155252](#)
- [10] M. Möhle. Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent. *Stochastic Process. Appl.* **120** (2010) 2159–2173. [MR2684740](#)
- [11] M. Möhle. On hitting probabilities of beta coalescents and absorption times of coalescents that come down from infinity. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** (2014) 141–159. [MR3225970](#)
- [12] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.* **27** (1999) 1870–1902. [MR1742892](#)
- [13] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** (1999) 1116–1125. [MR1742154](#)
- [14] J. Schweinsberg. A necessary and sufficient condition for the  $\Lambda$ -coalescent to come down from infinity. *Electron. Commun. Probab.* **5** (2000) 1–11. [MR1736720](#)