

## ON ASYMPTOTICS OF EXCHANGEABLE COALESCENTS WITH MULTIPLE COLLISIONS

ALEXANDER GNEDIN,\* *Utrecht University*

ALEX IKSANOV,\*\* *National Taras Shevchenko University of Kiev*

MARTIN MÖHLE,\*\*\* *University of Düsseldorf*

### Abstract

We study the number of collisions,  $X_n$ , of an exchangeable coalescent with multiple collisions ( $\Lambda$ -coalescent) which starts with  $n$  particles and is driven by rates determined by a finite characteristic measure  $\nu(dx) = x^{-2}\Lambda(dx)$ . Via a coupling technique, we derive limiting laws of  $X_n$ , using previous results on regenerative compositions derived from stick-breaking partitions of the unit interval. The possible limiting laws of  $X_n$  include normal, stable with index  $1 \leq \alpha < 2$ , and Mittag-Leffler distributions. The results apply, in particular, to the case when  $\nu$  is a beta( $a - 2, b$ ) distribution with parameters  $a > 2$  and  $b > 0$ . The approach taken allows us to derive asymptotics of three other functionals of the coalescent: the absorption time, the length of an external branch chosen at random from the  $n$  external branches, and the number of collision events that occur before the randomly selected external branch coalesces with one of its neighbours.

*Keywords:* Absorption time; coupling; exchangeable coalescent; external branch length; multiple collisions; number of collisions; regenerative composition

2000 Mathematics Subject Classification: Primary 60G09

Secondary 60C05; 60K05

### 1. Introduction

Let  $\Pi_n = (\Pi_n(t), t \geq 0)$  be an exchangeable coalescent with multiple collisions, as introduced in [22] and [25]. This is a càdlàg (right continuous with left limits) Markov process that starts at  $t = 0$  with  $n$  particles and evolves according to the following dynamics. When  $m$  particles are present, each  $k$ -tuple of them collides and merges to form a single particle at rate

$$\lambda_{m,k} = \int_{[0,1]} x^k(1-x)^{m-k} \nu(dx), \quad 2 \leq k \leq m. \quad (1.1)$$

The definition of the rates requires that the *characteristic measure*  $\nu$  has a finite second moment. These processes are sometimes called ‘lambda-coalescents’, following Pitman [22], who introduced the rates by means of the finite measure  $\Lambda(dx) = x^2\nu(dx)$ .

The coalescent starting with  $n$  particles is usually understood as a process with values in the set of partitions of  $[n] = \{1, \dots, n\}$ . To this end, we need to label the initial particles by

---

Received 18 July 2008; revision received 13 October 2008.

\* Postal address: Department of Mathematics, Utrecht University, Postbus 80010, 3508 TA Utrecht, The Netherlands.  
Email address: a.v.gnedin@uu.nl

\*\* Postal address: Faculty of Cybernetics, National Taras Shevchenko University of Kiev, 01033 Kiev, Ukraine.  
Email address: iksan@unicyb.kiev.ua

\*\*\* Postal address: Mathematical Institute, University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany.  
Email address: moehle@math.uni-duesseldorf.de

integers and, tracking their evolution, to record which of the initial particles comprise a merger. In this context, the exchangeability means that the restriction of the process to  $[m]$  (or any other subset of  $[n]$  with  $m < n$  elements) is a stochastic copy of the process starting with  $m$  particles; see [3] and [23] for a survey and details. For most of our purposes, however, it is sufficient to identify the coalescent with a pure-death process of counting the particles, with state space  $[n]$  and transition rates

$$g_{n,m} = \binom{n}{m-1} \lambda_{n,n-m+1}, \quad 1 \leq m \leq n-1,$$

so that the total transition rate from state  $n$  is

$$g_n = \sum_{m=1}^{n-1} g_{n,m} = \int_{[0,1]} (1 - (1-x)^n - nx(1-x)^{n-1}) \nu(dx).$$

In terms of the richer partition-valued process, the event  $\Pi_n(t) = m$  occurs when at time  $t$  the partition of  $[n]$  has  $m$  blocks, and a merge of some  $k \geq 2$  blocks into a single block corresponds to a transition from  $m$  to  $m - k + 1$ .

The coalescent starts in the state  $\Pi_n(0) = n$  and terminates in the state 1, as all particles eventually merge into one particle. The *number of collisions*  $X_n$  is the number of transitions which the process needs to proceed from  $n$  to the terminal state 1. The *absorption time*  $\tau_n = \inf\{t \geq 0: \Pi_n(t) = 1\}$  is the time when the process enters the terminal state. Note that in biological applications of the coalescent theory the absorption time  $\tau_n$  has the meaning of the time back to the most recent common ancestor. The main focus of this paper is on the weak asymptotic behaviour of  $X_n$  and  $\tau_n$  as  $n$  tends to  $\infty$ . In particular, the main result (Theorem 3.1) is a criterion describing all possible limit laws of, properly normalised and centred,  $X_n$ : these can be normal, stable with index  $1 \leq \alpha < 2$ , and Mittag–Leffler distributions. Another result (Theorem 4.1) provides sufficient conditions under which  $\tau_n$  possesses the same collection of limit laws. The limiting results for both  $X_n$  and  $\tau_n$  will be obtained by comparison with a simpler process which was partly investigated in connection with regenerative composition structures [9], [16].

Intuitively, a high concentration of the measure  $\nu$  near 0 increases the intensity of small jumps of  $\Pi_n$  and, hence, increases the number of collisions,  $X_n$ . This is well exemplified by the different kinds of asymptotic behaviour of  $X_n$  for measures  $\nu$  with densities  $\nu(dx) = c x^{a-3}(1-x)^{b-1} dx$ , where  $a, b, c > 0$ . For  $0 < a < 1$  and  $b > 0$ , the order of growth of  $X_n$  is linear [5], [11], [18]; for  $a = 1$  and  $b = 1$ , the order of growth of  $X_n$  is  $n/\log n$  [6], [7], [18], [21]; for  $1 < a < 2$  and  $b = 1$ , the order of growth of  $X_n$  is  $n^{2-a}$  [18]; while, for  $a = 2$  and  $b > 0$ , the order of growth of  $X_n$  is  $\log^2 n$  [20]. The class of beta distributions covered by the present paper corresponds to the range  $a > 2$ , when  $X_n$  exhibits a logarithmic growth. See [2] for other aspects of beta coalescents.

We shall assume throughout that the characteristic measure  $\nu$  is a probability measure on the open interval  $(0, 1)$  such that

- (i) the support of  $\nu$  is not contained in a sequence of the form  $(1 - \delta\gamma^n, n = 0, 1, \dots)$  for some  $\delta > 0$  and  $\gamma \in (0, 1)$ ,

- (ii)  $\nu$  satisfies

$$\int_{(0,1)} |\log x| \nu(dx) < \infty. \tag{1.2}$$

Note that the more general case of a finite measure  $\nu$  can be reduced to the case of a probability measure by a linear time change of the coalescent.

### 2. Annihilator and a coupling

Adopting Pitman’s [22, p. 1884] approach, the coalescent process  $\Pi_n$  can be constructed as follows. Let  $0 < s_1 < s_2 < \dots$  be the arrival times of a unit Poisson process, and let  $1 - \eta_1, 1 - \eta_2, \dots$  be an independent, identically distributed (i.i.d.) sample from  $\nu$ , also independent of the Poisson process. At time  $s_1$ , each of the  $n$  particles is marked ‘head’ with probability  $1 - \eta_1$  or ‘tail’ with probability  $\eta_1$ . Those particles marked ‘head’ are merged into one particle, provided that there are at least two ‘heads’. Thus,  $k \geq 2$  ‘heads’ imply that  $n - k + 1$  particles remain after merging, i.e.  $\Pi_n(s_1) = n - k + 1$ , and  $k = 1$  or  $k = 0$  ‘heads’ imply that  $\Pi_n(s_1) = n$ . The procedure is then iterated at time  $s_2$  with the remaining particles and the new probability for ‘heads’ being  $1 - \eta_2$ , and so on. It is immediately checked that this description agrees with the rates given in (1.1). The rules define the  $\Pi_n$ s consistently for all values of  $n$ , a property used by Pitman [22] to introduce a partition-valued coalescent process with infinitely many particles.

There is a simpler process on  $\{0\} \cup [n]$ , whose transition rate from  $m$  to  $m - k$  for  $k = 1, \dots, m$  is  $\binom{n}{m} \lambda_{n,k}$  (where  $\lambda_{n,1}$  is defined by (1.1) with  $k = 1$ ). Let  $\mathbb{K}_n = (\mathbb{K}_n(t), t \geq 0)$  denote the process starting at  $n$  and terminating at 0. We shall call this process *annihilator* in view of the following interpretation: when there are  $m$  particles, any  $k$ -tuple of them collides and annihilates at rate  $\lambda_{m,k}$ , so that after the collision,  $m - k$  particles remain (we stress that the case  $k = 1$  of ‘singular’ collision is not excluded). A sequential construction of the annihilator is almost the same as that of the coalescent. At the generic time  $s_j$  of a unit Poisson process, a random value  $1 - \eta_j$  is sampled from  $\nu$ , and each of the remaining particles (if any) is marked ‘head’ with probability  $1 - \eta_j$  or ‘tail’ with probability  $\eta_j$ , then the particles marked ‘heads’ are removed.

The sequence of positive decrements of  $\mathbb{K}_n$  is a random *composition* (ordered partition) of the integer  $n$ . For instance, the path  $10, 7, 2, 1, 0, \dots$  of  $\mathbb{K}_{10}$  corresponds to the composition  $(3, 5, 1, 1)$  of the integer 10. As  $n$  varies, the random compositions are consistent with respect to a sampling operation, i.e. they comprise a *composition structure* (see [10] and [12]). Two features differ the transition rules of  $\Pi_n$  and  $\mathbb{K}_n$ :

- (i) if at time  $s_j$  exactly  $k \geq 2$  of the remaining particles are marked ‘heads’ then  $\Pi_n$  decrements by  $k - 1$ , while  $\mathbb{K}_n$  decrements by  $k$ ,
- (ii) if  $k = 1$  then  $\Pi_n$  has no transition at  $s_j$ , while  $\mathbb{K}_n$  decrements by 1.

Note that neither  $\Pi_n$  nor  $\mathbb{K}_n$  jump at time  $s_j$  if none of the remaining particles are marked ‘heads’.

To couple the coalescent and the annihilator, we just define them using the same unit Poisson process with arrival times  $s_1, s_2, \dots$  and the same sample  $1 - \eta_1, 1 - \eta_2, \dots$  from  $\nu$ . We shall call the initial  $n$  particles *primary* and their followers resulting from mergers *secondary*. The primary particles are labelled  $1, 2, \dots, n$ . If at some time exactly  $k \geq 1$  particles are marked ‘heads’, these  $k$  particles are replaced by one secondary particle. The variable  $\mathbb{K}_n(t)$  is identified with the number of primary particles alive at time  $t$ , and  $\Pi_n(t)$  is identified with the total number of primary and secondary particles at time  $t$  (thus, by this coupling we always have  $\Pi_n(t) \geq \mathbb{K}_n(t)$ ). For instance, if at time  $s_1$  the particles 2, 3, and 6 are marked ‘heads’, the number of primary particles decreases by three and one secondary particle is born.

Let  $K_n$  be the number of transitions of  $\mathbb{K}_n$  as the process proceeds from  $n$  to 0, and let  $K_{n,1}$  be the number of jumps of size 1 (singular collisions). Introduce the absorption time of the annihilator  $\sigma_n = \inf\{t \geq 0 : \mathbb{K}_n(t) = 0\}$ . We further define  $K_{n,0}$  to be the number of epochs  $s_j < \sigma_n$  when none of the primary particles are marked ‘heads’. We define  $U_n$  to be the number of particles that remain at time  $\sigma_n$ ; all these particles are secondary, so  $U_n = \Pi_n(\sigma_n)$ . For two real-valued random variables  $X$  and  $Y$ , we say that  $X$  is stochastically smaller than  $Y$  and write  $X \leq_D Y$  if  $P\{X \leq x\} \geq P\{Y \leq x\}$  for every  $x \in \mathbb{R}$ .

**Lemma 2.1.** *The following stochastic order relations are true:*

$$K_n - K_{n,1} \leq_D X_n \leq_D K_n + K_{n,0} + X_{U_n}, \tag{2.1}$$

where, on the right-hand side,  $(K_n, K_{n,0}, U_n)$  is assumed to be independent of  $\{X_j : j \in \mathbb{N}\}$ .

*Proof.* We use the above coupling. By collision of at least two primary particles, both processes  $\Pi_n$  and  $K_n$  decay; hence,  $K_n - K_{n,1} \leq X_n$ . The number of jumps of  $\Pi_n$  before and including time  $\sigma_n$  does not exceed the number of Poisson epochs, which is  $K_n + K_{n,0}$ , and after time  $\sigma_n$ , the coalescent evolves with  $U_n$  particles, independently of the history (given  $U_n$ ). This yields the upper bound for  $X_n$ .

**Lemma 2.2.** *The sequence of distributions of  $X_{U_n}$  ( $n \in \mathbb{N}$ ) is tight.*

*Proof.* First of all,  $1 \leq X_{U_n} < U_n$ , because the number of collisions after  $\sigma_n$  does not exceed the number of particles that remain. To bound  $U_n$ , we focus on the behaviour of the secondary particles. At any time  $s_j \leq \sigma_n$ , each of the secondary particles is removed with probability  $1 - \eta_j$  and at most one secondary particle is born (possibly by collision of only primary particles). We can only increase the number of secondary particles if we assume that a secondary particle is always born at every time  $s_j$ .

For  $j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ , denote by  $Q_j$  the number of secondary particles at time  $s_j$  ( $s_0 := 0$ ), and let  $\mathcal{F}$  be the  $\sigma$ -field generated by all the variables  $\eta_k$ ,  $k \in \mathbb{N}$ . Note that  $Q_0 = 0$ . With the above convention, given  $\mathcal{F}$ , the sequence  $\{Q_j, j \in \mathbb{N}_0\}$  becomes a time-inhomogeneous Markov chain on the set of nonnegative integers with transition probabilities

$$P\{Q_j = k \mid Q_{j-1} = m\} = \binom{m}{k-1} \eta_j^{k-1} (1 - \eta_j)^{m-k+1}, \quad k \in \{1, \dots, m+1\}.$$

Set

$$N_n := \inf\{k \geq 1 : n\eta_1 \cdots \eta_k \leq 1\}, \quad n \in \mathbb{N}.$$

Almost surely (a.s.) on the event  $\{N_n \geq 2\}$ ,

$$E(Q_{N_n} \mid \mathcal{F}) = 1 + \sum_{j=2}^{N_n} \eta_j \eta_{j+1} \cdots \eta_{N_n} =: 1 + \Theta_n, \quad n \in \mathbb{N}.$$

Even though the subsequent argument is the same as in the proof of [19, Theorem 1.1], we include it for completeness. For  $k = 2, 3, \dots$ , we have

$$\begin{aligned} \Theta_n \mathbf{1}_{\{N_n=k\}} &= \eta_1 \cdots \eta_{N_n} \left( \frac{1}{\eta_1} + \frac{1}{\eta_1 \eta_2} + \cdots + \frac{1}{\eta_1 \cdots \eta_{N_n-1}} \right) \mathbf{1}_{\{N_n=k\}} \\ &\leq \frac{1}{n} \left( \frac{1}{\eta_1} + \frac{1}{\eta_1 \eta_2} + \cdots + \frac{1}{\eta_1 \cdots \eta_{k-1}} \right) \mathbf{1}_{\{N_n=k\}} \quad \text{a.s.,} \end{aligned}$$

which implies that

$$\Theta_n \mathbf{1}_{\{N_n \geq 2\}} \leq \frac{1}{n} \sum_{k=1}^{N_n-1} \frac{1}{\eta_1 \cdots \eta_k} \quad \text{a.s.}$$

By the key renewal theorem,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left( \sum_{k=0}^{N_n-1} \frac{1}{\eta_1 \cdots \eta_k} \right) \\ &= \lim_{n \rightarrow \infty} \int_0^{\log n} e^{-(\log n - x)} \left( \sum_{k=0}^{\infty} \mathbb{P}\{-\log \eta_1 - \cdots - \log \eta_k \in dx\} \right) \\ &= \frac{1}{\mathbb{E}(-\log \eta)}, \end{aligned}$$

no matter whether  $M_1 := \mathbb{E}(-\log \eta)$  is finite or infinite. Since  $\mathbb{E}(Q_{N_n} \mathbf{1}_{\{N_n=1\}}) = o(1)$ , we have proved that

$$\limsup_{n \rightarrow \infty} \mathbb{E}(Q_{N_n}) \leq \frac{1}{M_1} + 1,$$

which implies that the sequence of distributions of  $Q_{N_n}$  ( $n \in \mathbb{N}$ ) is tight. Note that  $N_n$  is the number of steps that the multiplicative renewal process  $(\eta_1 \cdots \eta_j, j = 0, 1, \dots)$  needs to drop below level  $1/n$ . Now we wish to replace the fixed level  $1/n$  by a random level associated with the last primary particle disappearing at time  $\sigma_n$ ; this random level can be identified with the minimal order statistic  $R_{1,n}$  of the sample of size  $n$  from the uniform  $[0, 1]$  distribution (see Remark 3.1, below, for this connection). From the subadditivity of the first passage time processes for random walks, and from the convergence in distribution of  $-\log R_{1,n} - \log n$  (which is a standard fact from the extreme-value theory), it follows that the number of steps,  $\rho_n$  say, the renewal process needs to cross the interval  $(\min(R_{1,n}, 1/n), \max(R_{1,n}, 1/n))$  remains stochastically bounded as  $n \rightarrow \infty$ . On the event  $A_n := \{R_{1,n} \leq 1/n\}$ , as the process proceeds from  $1/n$  to  $R_{1,n}$ , the number of secondary particles may increase at most by  $\rho_n$ . On the complementary event,  $U_n - Q_{N_n} \leq \chi_n U_n$  a.s., where  $\chi_n \in (0, 1)$  is a stochastic sequence (depending on  $\rho_n$ ) which is bounded away from 1. The latter inequality follows from the observation that in the coalescent process under consideration the number of particles involved in the generic collision is proportional to the number of existing particles. In conclusion,  $|Q_{N_n} - U_n|$  is stochastically bounded on both events, which completes the proof.

### 3. Number of collisions

With these preliminaries we are ready to derive the asymptotic behaviour of the number of collisions,  $X_n$ , from the previously known results on  $K_n$  [16]. Introduce the logarithmic moments

$$M_1 := \mathbb{E}(-\log \eta) \quad \text{and} \quad M_2 := \text{var}(\log \eta)$$

for a random variable  $1 - \eta$  with distribution  $\nu$ , and recall the standing assumptions of the paper given at the end of Section 1. It is worth mentioning explicitly that in parts (A)–(C) of the following theorem  $M_1 < \infty$ .

**Theorem 3.1.** *The following two assertions are equivalent.*

- (i) *There exist constants  $\{a_n, b_n : n \in \mathbb{N}\}$  with  $a_n > 0$  and  $b_n \in \mathbb{R}$  such that, as  $n \rightarrow \infty$ ,  $(X_n - b_n)/a_n$  converges weakly to some nondegenerate and proper distribution.*
- (ii) *The distribution of  $(-\log \eta)$  either belongs to the domain of attraction of a stable law or the function  $x \mapsto P\{-\log \eta > x\}$  slowly varies at  $\infty$ .*

Furthermore, there are five different regimes of convergence.

(A) *If  $M_2 < \infty$  then, with  $b_n := M_1^{-1} \log n$  and  $a_n := (M_1^{-3} M_2 \log n)^{1/2}$ , the limiting law of  $(X_n - b_n)/a_n$  is standard normal.*

(B) *If  $M_2 = \infty$  and*

$$\int_{(x,1)} (\log y)^2 \nu(dy) \sim L(-\log x) \quad \text{as } x \rightarrow 0$$

*for some  $L$  slowly varying at  $\infty$ , then, with  $b_n := M_1^{-1} \log n$  and  $a_n := M_1^{-3/2} c_{[\log n]}$ , where  $c_n$  is any sequence satisfying  $\lim_{n \rightarrow \infty} nL(c_n)/c_n^2 = 1$ , the limiting law of  $(X_n - b_n)/a_n$  is standard normal.*

(C) *Assume that, for some function  $L$  slowly varying at  $\infty$ , the relation*

$$P\{\eta \leq x\} \sim (-\log x)^{-\alpha} L(-\log x) \quad \text{as } x \rightarrow 0 \tag{3.1}$$

*holds with  $\alpha \in [1, 2)$  and that  $M_1 < \infty$  if  $\alpha = 1$ . Then, with  $b_n := M_1^{-1} \log n$  and  $a_n := M_1^{-(\alpha+1)/\alpha} c_{[\log n]}$ , where  $c_n$  is any sequence satisfying  $\lim_{n \rightarrow \infty} nL(c_n)/c_n^\alpha = 1$ , the limiting law of  $(X_n - b_n)/a_n$  is  $\alpha$ -stable with characteristic function*

$$t \mapsto \exp \left\{ -|t|^\alpha \Gamma(1 - \alpha) \left( \cos\left(\frac{\pi\alpha}{2}\right) + i \sin\left(\frac{\pi\alpha}{2}\right) \operatorname{sgn}(t) \right) \right\}, \quad t \in \mathbb{R}.$$

(D) *Assume that  $M_1 = \infty$  and that (3.1) holds with  $\alpha = 1$ . Let  $c$  be any positive function satisfying  $\lim_{x \rightarrow \infty} xL(c(x))/c(x) = 1$ , and set  $\psi(x) := x \int_{\exp\{-c(x)\}}^1 P\{\eta \leq y\} / y \, dy$ . Let  $b(x)$  be any positive function satisfying  $b(\psi(x)) \sim \psi(b(x)) \sim x$ . Then, with  $b_n := b(\log n)$  and  $a_n := b(\log n)c(b(\log n))/\log n$ , the limiting law of  $(X_n - b_n)/a_n$  is 1-stable with characteristic function*

$$t \mapsto \exp \left\{ -|t| \left( \frac{\pi}{2} - i \log |t| \operatorname{sgn}(t) \right) \right\}, \quad t \in \mathbb{R}.$$

(E) *If (3.1) holds with  $\alpha \in [0, 1)$  then, with  $a_n := \log^\alpha n / L(\log n)$ , the limiting law of  $X_n/a_n$  is scaled Mittag-Leffler  $\theta_\alpha$  (exponential if  $\alpha = 0$ ) with moments*

$$\int_{[0,\infty)} x^n \theta_\alpha(dx) = \frac{n!}{\Gamma^n(1 - \alpha) \Gamma(1 + n\alpha)}, \quad n \in \mathbb{N}.$$

*Proof.* Following the same line of argument as in the proof of Theorem 2.1 of [16], we can show that a necessary condition for the convergence of  $(X_n - b_n)/a_n$  is that the scaling satisfies  $a_n \rightarrow \infty$ , which will be assumed in the sequel. To derive the result from Theorem 2.3 of [16], it is enough to show that  $(X_n - b_n)/a_n$  weakly converges to a proper and nondegenerate distribution if and only if  $(K_n - b_n)/a_n$  weakly converges to the same distribution. This equivalence is implied by (2.1). Indeed, by Lemma 2.2,  $X_{U_n}/a_n$  converges in probability to 0. Under assumption (1.2),  $K_{n,0}$  converges in distribution by [16, Theorem 2.2] and  $K_{n,1}$  converges in distribution by [16, Proposition 5.2] (see also [13] for the case in which  $M_1 < \infty$ ).

**Remark 3.1.** We briefly describe a more general construction with possibly infinite characteristic measure  $\nu$  satisfying  $\int_{(0,1)} x \nu(dx) < \infty$ . Let  $\tilde{\nu}$  be the image of  $\nu$  under the transformation  $x \mapsto -\log x$ , let  $(\tilde{S}_t)$  be a subordinator with Lévy measure  $\tilde{\nu}$ , and let  $S_t = \exp\{-\tilde{S}_t\}$ . Then  $(S_t)$  is a nonincreasing positive pure-jump process with  $S_0 = 1$ . Primary particles are labelled by independent uniform $[0, 1]$  points  $R_1, \dots, R_n$ , also independent of  $(S_t)$ . The annihilator process  $\mathbb{K}_n = (\mathbb{K}_n(t), t \geq 0)$  is defined by setting  $\mathbb{K}_n(0) = n$  and by decrementing the process at time  $t$  by  $k$  if the interval  $[S_t, S_{t-}]$  contains exactly  $k$  of the uniform points. Thus,  $\mathbb{K}_n$  may have a jump at  $t$  only if  $S_{t-} > S_t$ . The coalescent  $\Pi_n$  is defined in a similar way, except  $R_1, \dots, R_n$  represent primary particles and new points representing secondary particles appear. Set  $\Pi_n(0) = n$ . Conditionally given that  $[S_t, S_{t-}]$  contains  $k \geq 1$  points, primary and secondary, let  $\Pi_n$  decrement by  $k - 1$  and let one point with uniform distribution on  $[0, S_t]$  appear to represent a secondary particle. The inequality  $X_n \geq K_n - K_{n,1}$  is still true, but another bound in (2.1) makes no sense for infinite  $\nu$ , since a variable analogous to  $K_{n,0}$  is identically  $\infty$ . It seems plausible that in this more general setting  $X_n$  and  $K_n - K_{n,1}$  still have the same limit laws and require the same scaling. Asymptotics of  $K_n$  and  $K_{n,1}$  were studied for infinite measures  $\nu$  satisfying a condition of regular variation in [14] and [15].

**Example 3.1.** Application of Theorem 3.1(A) to coalescents governed by beta measures  $\nu(dx) = cx^{a-3}(1-x)^{b-1}dx$  with parameters  $a > 2$  and  $b, c > 0$  shows that the scaled and centred number of collisions

$$\frac{X_n - M_1^{-1} \log n}{(M_2 M_1^{-3} \log n)^{1/2}}$$

converges in distribution to the standard normal law, where  $M_1 = \Psi(a - 2 + b) - \Psi(b)$ ,  $M_2 = \Psi'(b) - \Psi'(a - 2 + b)$ , and  $\Psi$  is the logarithmic derivative of the gamma function.

For the special case  $\nu(dx) = dx$  (uniform distribution), we have  $g_{n,m}/g_n = 1/(n - 1)$ . Thus,  $\Pi_n$  jumps from  $n$  to every  $m \in \{1, \dots, n - 1\}$  with the same probability. In this case, Theorem 3.1 implies that  $(X_n - \log n)/\sqrt{\log n}$  is asymptotically normal, which agrees with a classical result known in the context of records or cyclic partition of permutations. We refer the reader to survey [1] and the references therein for some further information.

More generally, when  $\nu$  is a beta(1,  $b$ ) distribution, a straightforward calculation shows that  $\Pi_n$  jumps from  $n \geq 2$  to any state  $m \in \{1, \dots, n - 1\}$  with probability  $w_m/(w_1 + \dots + w_{n-1})$ , where  $w_k := \Gamma(k + b - 1)/\Gamma(k)$ . By [24, Proposition 2],  $X_n$  is representable as the sum of  $n - 1$  independent indicators  $I_k$  with

$$P\{I_k = 1\} = \frac{w_k}{w_1 + \dots + w_k}, \quad k \in \{1, \dots, n - 1\}.$$

It can be shown that  $P\{I_k = 1\} \sim b/k$  as  $k \rightarrow \infty$ . From this, as  $n \rightarrow \infty$ , the expectation and the variance of  $X_n$  are both asymptotically equal to  $b \log n$ . By the central limit theorem, the distribution of  $(X_n - b \log n)/\sqrt{b \log n}$  converges to the standard normal law, which again agrees with Theorem 3.1(A). The latter result is classic and dates back to [17].

The merging evolution of  $n$  particles in the coalescent is often represented by a rooted tree with  $n$  leaves, and with the internal nodes corresponding to collisions [3], [23]. One interpretation of the similarity of the asymptotic behaviour of  $X_n$  and  $K_n$  is that almost all internal nodes of the tree are directly connected to one of the leaves.

Some other properties of the coalescent follow easily in the same way. Let  $Z_n$  be the time of the first collision of a tagged particle (say, labelled 1). In the tree representation,

$Z_n$  is the length of an external branch, which connects a leaf to its closest internal node. The asymptotic behaviour of  $Z_n$  for the Kingman coalescent and the Bolthausen–Sznitman coalescent ( $\nu(dx) = x^{-2} dx$ ) was studied in [4] and [8], respectively. In our setting we have the following proposition.

**Proposition 3.1.** *As  $n \rightarrow \infty$ ,  $Z_n$  converges weakly to an exponential variable with parameter  $p := E(1 - \eta) = \int_{(0,1)} x \nu(dx)$ . Moreover, the number of collisions in the coalescent before and including time  $Z_n$  converges weakly to a geometric variable with parameter  $p$ .*

*Proof.* In the process with infinitely many particles, a tagged particle is marked ‘head’ at rate  $p$ , and when this occurs, the particle collides with infinitely many other particles, because the assumption that  $\nu$  is a probability measure implies that the coalescent does not come down from  $\infty$  [22] (which is also obvious from the construction). This proves the first claim. The second claim follows similarly, by noting that the time of the first collision of a given primary particle is  $s_1$  with probability  $p$ .

#### 4. Absorption times

The coupling allows us to derive the asymptotics of  $\tau_n$ .

**Lemma 4.1.** *The following stochastic order relations hold:*

$$\sigma_n \leq_D \tau_n \leq_D \sigma_n + \tau'_{U_n},$$

where, on the right-hand side,  $\tau'_j \stackrel{D}{=} \tau_j$ ,  $j \in \mathbb{N}$  (where ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution), and  $\sigma_n$ ,  $(\tau'_j)$ , and  $U_n$  are assumed to be independent.

*Proof.* These relations are obvious from the coupling. When all the primary particles disappear at time  $\sigma_n$ , there are  $U_n$  secondary particles left.

From the results obtained in [16], it follows that  $\sigma_n$  has the same limiting distribution as the first passage time through the level  $\log n$  for a compound Poisson process  $\{R_t : t \geq 0\}$  with the generic jump  $(-\log \eta)$  and intensity 1. By Lemma 2.2 and Lemma 4.1, the same asymptotics are valid for  $\tau_n$ . This argument leads to the following result.

**Theorem 4.1.** *Assume that Theorem 3.1(ii) holds. Then, with  $X_n$  replaced by  $\tau_n$ , condition (i) and parts (B)–(E) of Theorem 3.1 hold without changes, while part (A) holds with  $a_n = (M_1^{-3}(M_2 + M_1^2) \log n)^{1/2}$ .*

*Proof.* We focus on the asymptotics of the first passage time process  $\{U_x : x \geq 0\}$  defined as

$$U_x := \inf\{t \geq 0 : R_t \geq x\}, \quad x \geq 0.$$

Let  $N_x$  be the number of renewals on  $[0, x]$  of the renewal process with  $N_0 = 0$  and steps distributed as  $(-\log \eta)$ . Obviously,  $U_x = s_{N_x}$ , where, as above,  $s_1, s_2, \dots$  are the arrival times of an independent unit Poisson process. Assume that Theorem 3.1(ii) holds, and let  $a(x)$  and  $b(x)$  be the same as in [16, Proposition 8.1] (setting  $a_n := a(\log n)$  and  $b_n := b(\log n)$  gives the normalizing constants for Theorem 3.1(B)–(E)). We will exploit the following decomposition:

$$\frac{s_{N_x} - b(x)}{a(x)} = \frac{s_{N_x} - N_x \sqrt{N_x}}{\sqrt{N_x} a(x)} + \frac{N_x - b(x)}{a(x)}. \tag{4.1}$$

In view of the independence and since  $N_x \rightarrow \infty$  almost surely as  $x \rightarrow \infty$ , by the central limit theorem, as  $x \rightarrow \infty$ ,  $(s_{N_x} - N_x)/\sqrt{N_x}$  converges in distribution to a random variable  $\xi_1$  with the standard normal distribution. By [16, Proposition 8.1], the right-most term in (4.1) converges in distribution to a random variable  $\xi_2$  with a proper and nondegenerate law. Note that  $\xi_1$  and  $\xi_2$  are independent.

*Case (A).* By [16, Proposition 8.1],  $b(x) = M_1^{-1}x$ ,  $a(x) = (M_2M_1^{-3}x)^{1/2}$ , and  $\xi_2$  has the standard normal law. Since, by the strong law of large numbers, as  $x \rightarrow \infty$ ,  $N_x/x \rightarrow M_1^{-1}$  almost surely, we conclude that the left-hand side of (4.1) converges in distribution to a random variable  $M_1M_2^{-1/2}\xi_1 + \xi_2$  which has the normal law with zero mean and variance  $M_1^2M_2^{-1} + 1$ . Therefore,  $(U_x - M_1^{-1}x)/(M_1^{-3}(M_2 + M_1^2)x)^{1/2}$  weakly converges to the standard normal law, as desired.

*Cases (B)–(E).* By [16, Proposition 8.1], as  $x \rightarrow \infty$ ,  $\sqrt{N_x}/a(x)$  converges to 0 in probability. Therefore, the left-hand side of (4.1) weakly converges to the same law as the right-most term in (4.1). We again refer the reader to [16, Proposition 8.1] for the identification of the normalizing constants and the relevant limiting laws.

**Example 4.1.** Theorem 4.1 is obviously applicable to the class of beta coalescents with parameters  $a > 2$  and  $b, c > 0$  considered in Example 3.1. Consequently, for such beta coalescents, the scaled and centred absorption time  $(\tau_n - M_1^{-1} \log n)/(M_1^{-3}(M_2 + M_1^2) \log n)^{1/2}$  converges in distribution to the standard normal law, with  $M_1$  and  $M_2$  as given in Example 3.1.

### Acknowledgement

A. Iksanov gratefully acknowledges the support by the German Scientific Foundation (project number 436UKR 113/93/0-1).

### References

- [1] BAI, Z. D., HWANG, H. K. AND LIANG, W. Q. (1998). Normal approximations of the number of records in geometrically distributed random variables. *Random Structures Algorithms* **13**, 319–334.
- [2] BERESTYCKI, J., BERESTYCKI, N. AND SCHWEINSBERG, J. (2008). Small time properties of beta coalescents. *Ann. Inst. H. Poincaré Prob. Statist.* **44**, 214–238.
- [3] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge University Press.
- [4] CALIEBE, A., NEININGER, R., KRAWCZAK, M. AND RÖSLER, U. (2007). On the length distribution of external branches in coalescence trees: genetic diversity within species. *Theoret. Pop. Biol.* **72**, 245–252.
- [5] DELMAS, J.-F., DHERSIN, J.-S. AND SIRI-JEGOUSSE, A. (2008). Asymptotic results on the length of coalescent trees. *Ann. Appl. Prob.* **18**, 997–1025.
- [6] DRMOTA, M., IKSANOV, A., MÖHLE, M. AND RÖSLER, U. (2007). Asymptotic results concerning the total branch length of the Bolthausen–Sznitman coalescent. *Stoch. Process. Appl.* **117**, 1404–1421.
- [7] DRMOTA, M., IKSANOV, A., MÖHLE, M. AND RÖSLER, U. (2009). A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. To appear in *Random Structures Algorithms*.
- [8] FREUND, F. AND MÖHLE, M. (2007). On the time back to the most recent common ancestor and the external branch length of the Bolthausen–Sznitman coalescent. Submitted.
- [9] GNEDIN, A. V. (2004). The Bernoulli sieve. *Bernoulli* **10**, 79–96.
- [10] GNEDIN, A. AND PITMAN, J. (2005). Regenerative composition structures. *Ann. Prob.* **33**, 445–479.
- [11] GNEDIN, A. AND YAKUBOVICH, Y. (2007). On the number of collisions in  $\Lambda$ -coalescents. *Electron. J. Prob.* **12**, 1547–1567.
- [12] GNEDIN, A., DONG, R. AND PITMAN, J. (2007). Exchangeable partitions derived from Markovian coalescents. *Ann. Appl. Prob.* **17**, 1172–1201.
- [13] GNEDIN, A., IKSANOV, A. AND ROESLER, U. (2008). Small parts in the Bernoulli sieve. *Discrete Math. Theoret. Computer Sci.* **A1**, 239–246.
- [14] GNEDIN, A., PITMAN, J. AND YOR, M. (2006). Asymptotic laws for compositions derived from transformed subordinators. *Ann. Prob.* **34**, 468–492.

- [15] GNEDIN, A., PITMAN, J. AND YOR, M. (2006). Asymptotic laws for regenerative compositions: gamma subordinators and the like. *Prob. Theory Relat. Fields* **135**, 576–602.
- [16] GNEDIN, A., IKSANOV, A., NEGADAJLOV, P. AND ROESLER, U. (2008). The Bernoulli sieve revisited. Preprint. Available at <http://arxiv.org/abs/0801.4725>.
- [17] GONCHAROV, V. L. (1962). On the field of combinatorial analysis. *Amer. Math. Soc. Trans. Ser. 2*, **19**, 1–46.
- [18] IKSANOV, A. AND MÖHLE, M. (2008). On the number of jumps of random walks with a barrier. *Adv. Appl. Prob.* **40**, 206–228.
- [19] IKSANOV, A. AND TERLETSKY, YU. (2008). On asymptotic behavior of certain recursions with random indices of linear growth. *Prob. Stat. Forum* **1**, 62–67.
- [20] IKSANOV, A., MARYNYCH, A. AND MÖHLE, M. (2007). On the number of collisions in beta(2,b)-coalescents. Submitted.
- [21] PANHOLZER, A. (2004). Destruction of recursive trees. In *Mathematics and Computer Science III*, Birkhäuser, Basel, pp. 267–280.
- [22] PITMAN, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27**, 1870–1902.
- [23] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.
- [24] ROSS, S. M. (1982). A simple heuristic approach to simplex efficiency. *Europ. J. Operat. Res.* **9**, 344–346.
- [25] SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36**, 1116–1125.