Adv. Appl. Prob. **41,** 1082–1101 (2009)
Printed in Northern Ireland

ON THE NUMBER OF ALLELIC TYPES FOR SAMPLES TAKEN FROM EXCHANGEABLE COALESCENTS WITH MUTATION

F. FREUND * ** AND
M. MÖHLE,* *** *Eberhard Karls Universität Tübingen*

Abstract

Let K_n denote the number of types of a sample of size n taken from an exchangeable coalescent process (Ξ-coalescent) with mutation. A distributional recursion for the sequence $(K_n)_{n\in\mathbb{N}}$ is derived. If the coalescent does not have proper frequencies, i.e. if the characterizing measure Ξ on the infinite simplex Δ does not have mass at 0 and satisfies $\int_\Delta |x|\Xi(\mathrm{d}x)/(x,x) < \infty$, where $|x| := \sum_{i=1}^\infty x_i$ and $(x,x) := \sum_{i=1}^\infty x_i^2$ for $x = (x_1, x_2, \ldots) \in \Delta$, then K_n/n converges weakly as $n \to \infty$ to a limiting variable K that is characterized by an exponential integral of the subordinator associated with the coalescent process. For so-called simple measures Ξ satisfying $\int_\Delta \Xi(\mathrm{d}x)/(x,x) < \infty$, we characterize the distribution of K via a fixed-point equation.

Keywords: Coalescent; distributional recursion; number of types; simultaneous multiple collisions; fixed point; subordinator

2000 Mathematics Subject Classification: Primary 60C05; 05C05
Secondary 60F05; 92D15

1. Introduction and main results

Exchangeable coalescents are Markovian processes with state space $\mathcal{E}$, the set of equivalence relations (partitions) on $\mathbb{N} := \{1, 2, \ldots\}$ with a block-merging mechanism. The class of exchangeable coalescents with multiple collisions has been independently introduced in [23] and [24]. These processes can be characterized by a finite measure Λ on the unit interval $[0, 1]$ and are hence also called Λ-coalescents. The best known example is the Kingman coalescent, for which $\Lambda = \delta_0$ is the Dirac measure at 0. This coalescent allows for only binary mergers of ancestral lineages. Another well-studied coalescent is the Bolthausen–Sznitman coalescent [5], for which Λ is uniformly distributed on $[0, 1]$. The full class of exchangeable coalescents allowing for simultaneous multiple collisions of ancestral lineages was discovered by Möhle and Sagitov [22] and Schweinsberg [26]. Schweinsberg [26] characterized exchangeable coalescents via a finite measure Ξ on the infinite simplex

$$\Delta := \left\{ x = (x_1, x_2, \ldots) \colon x_1 \geq x_2 \geq \cdots \geq 0,\ \sum_{i=1}^{\infty} x_i \leq 1 \right\}.$$

In the following, for convenience, we decompose $\Xi = a\delta_0 + \Xi_0$, where $a := \Xi(\{0\}) \in [0, \infty)$ and Ξ_0 does not have an atom at 0. Suppose that the coalescent is in a state with n blocks.

Received 15 August 2008.
* Postal address: Mathematisches Institut, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany.
** Email address: fabian.freund@uni-tuebingen.de
*** Email address: martin.moehle@uni-tuebingen.de

Then each $(k_1, \dots, k_j)$-collision ($k_1, \dots, k_j \in \mathbb{N}$ with $k_1 + \cdots + k_j = n$, $k_1 \geq \cdots \geq k_j$ and $k_1 \geq 2$) occurs at rate (see [26, Equation (11)])

$$\phi_j(k_1, \dots, k_j) = a\,\mathbf{1}_{\{r=1,\, k_1=2\}} + \int_\Delta \sum_{l=0}^{s} \binom{s}{l} (1-|x|)^{s-l} \sum_{\substack{i_1,\dots,i_{r+l} \in \mathbb{N} \\ \text{all distinct}}} x_{i_1}^{k_1} \cdots x_{i_{r+l}}^{k_{r+l}} \frac{\Xi_0(\mathrm{d}x)}{(x,x)},$$

where $s := |\{1 \leq i \leq j : k_i = 1\}|$, $r := j - s$, $|x| := \sum_{i=1}^\infty x_i$, and $(x,x) := \sum_{i=1}^\infty x_i^2$ for $x = (x_1, x_2, \dots) \in \Delta$. Note that $\phi_1(2) = \Xi(\Delta)$. For $n \in \mathbb{N}$, let $\varrho_n \colon \mathcal{E} \to \mathcal{E}_n$ denote the natural restriction to the set $\mathcal{E}_n$ of all equivalence relations on $\{1, \dots, n\}$. Let $R = (R_t)_{t \geq 0}$ be a coalescent process with simultaneous multiple collisions. The restricted coalescent process $(\varrho_n R_t)_{t \geq 0}$ is usually interpreted as a genealogical tree of a sample of n individuals. In the biological context it is natural to introduce mutations into this model as follows. Assume that each individual has a certain type. Independently of the genealogical tree, mutations occur along each branch of the tree according to a homogeneous Poisson process with rate $r > 0$. The infinitely many alleles model is assumed, i.e. each mutation leads to a new type never seen before in the sample. Recently, there has been much interest in the study of functionals of the restricted coalescent process $(\varrho_n R_t)_{t \geq 0}$; for example, the number of collisions has been studied in [10], [14], [15], and [17], the time back to the most recent common ancestor and the lengths of external branches have been studied in [6], [8], and [12], the total branch length has been studied in [9], and the number of segregating sites has been studied in [21]. Further typical quantities of interest are $K_i(n)$, the number of types which appear exactly i times in a sample of size n, and the summary statistics $K_n := \sum_{i=1}^n K_i(n)$, the total number of types in the sample. The most celebrated result in this context is Ewens' sampling formula [11] for the distribution of the allele frequency spectrum $(K_1(n), \dots, K_n(n))$ under the Kingman coalescent. Recently, asymptotic results for the allele frequency spectrum have been obtained by Berestycki *et al.* [2], [3] for beta$(2-\alpha, \alpha)$-coalescents with parameter $1 < \alpha < 2$ and by Basdevant and Goldschmidt [1] for the Bolthausen–Sznitman coalescent [5]. Here we are interested in the total number, K_n, of types of a sample of size $n \in \mathbb{N}$ taken from a Ξ-coalescent with mutation rate $r > 0$. The motivation for our interest in K_n is manifold. It is an observable quantity and, hence, important for biological and statistical applications. In combination with the results of [20] on the allele frequency spectrum and of [21] on the number of segregating sites, our study of K_n gives additional insight into the structure of exchangeable coalescent trees. Our first result (Theorem 1.1, below) provides a distributional recursion for the sequence $(K_n)_{n \in \mathbb{N}}$. In order to state the result, we need to introduce the rates

$$g_{nk} := \lim_{t \searrow 0} \frac{\mathrm{P}(|\varrho_n R_t| = k)}{t}, \qquad n, k \in \mathbb{N},\ k < n, \tag{1.1}$$

and the total rates

$$g_n := \lim_{t \searrow 0} \frac{\mathrm{P}(|\varrho_n R_t| < n)}{t} = \sum_{k=1}^{n-1} g_{nk}, \qquad n \in \mathbb{N}.$$

The total rates g_n, $n \in \mathbb{N}$, can be expressed in terms of the measure $\Xi = a\delta_0 + \Xi_0$ as (see [26, p. 36, Equation (70)])

$$g_n = a\binom{n}{2} + \int_\Delta \left(1 - (1-|x|)^n - \sum_{j=1}^{n} \binom{n}{j} (1-|x|)^{n-j} \sum_{\substack{i_1,\dots,i_j \in \mathbb{N} \\ \text{all distinct}}} x_{i_1} \cdots x_{i_j}\right) \frac{\Xi_0(\mathrm{d}x)}{(x,x)}. \tag{1.2}$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

A similar argument (see the remark after Lemma A.2 in Appendix A) shows that the rates (1.1) are given as

$$g_{nk} = a\binom{n}{2}\mathbf{1}_{\{k=n-1\}} + \int_\Delta \sum_{j=1}^{k} f_{nkj}(x)\frac{\Xi_0(\mathrm{d}x)}{(x,x)}, \qquad n,k\in\mathbb{N},\ k<n, \tag{1.3}$$

with

$$f_{nkj}(x) := \sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1<\cdots<i_j}} \sum_{\substack{n_1,\dots,n_j\in\mathbb{N}\\ n_1+\cdots+n_j=n-k+j}} \frac{n!}{(k-j)!\,n_1!\cdots n_j!}(1-|x|)^{k-j}x_{i_1}^{n_1}\cdots x_{i_j}^{n_j}$$

for $n,k\in\mathbb{N}$, $k<n$, and $j\in\{1,\dots,k\}$. The Λ-coalescent occurs if the measure Ξ is concentrated on the points $x=(u,0,0,\dots)\in\Delta$ with $u\in[0,1]$ and, hence, can be considered as a measure Λ on the unit interval $[0,1]$. In this case only the index $j=1$ contributes to the sum under the integral in (1.3) and, from $f_{nk1}(u,0,0,\dots)=\binom{n}{k-1}(1-u)^{k-1}u^{n-k+1}$, it follows that (1.3) takes the form

$$g_{nk} = \binom{n}{k-1}\int_{[0,1]} u^{n-k-1}(1-u)^{k-1}\Lambda(\mathrm{d}u), \qquad n,k\in\mathbb{N},\ k<n.$$

Similarly, for the Λ-coalescent, the total rates (1.2) are given as

$$g_n = \int_{[0,1]} \frac{1-(1-u)^n-nu(1-u)^{n-1}}{u^2}\Lambda(\mathrm{d}u), \qquad n\in\mathbb{N}.$$

Our first main result is the following distributional recursion for the number of types K_n.

Theorem 1.1. *The sequence $(K_n)_{n\in\mathbb{N}}$ satisfies the distributional recursion*

$$K_1=1 \quad \textit{and} \quad K_n \stackrel{\mathrm{D}}{=} B_n(K_{n-1}+1)+(1-B_n)K_{I_n}, \qquad n\in\{2,3,\dots\} \tag{1.4}$$

(by '$\stackrel{\mathrm{D}}{=}$' we denote equality in distribution), where B_n is a Bernoulli variable independent of $(K_2,\dots,K_{n-1},I_n)$ with distribution

$$\mathrm{P}(B_n=1)=1-\mathrm{P}(B_n=0)=\frac{nr}{g_n+nr}, \qquad n\in\mathbb{N},$$

and I_n is a random variable independent of $(K_2,\dots,K_{n-1})$ with distribution

$$r_{nk} := \mathrm{P}(I_n=k)=\frac{g_{nk}}{g_n}, \qquad n,k\in\mathbb{N},\ k<n. \tag{1.5}$$

Note that I_n is the number of equivalence classes (blocks) of the restricted coalescent process $(\varrho_n R_t)_{t\geq 0}$ after its first jump. The proof of Theorem 1.1 given in Section 2 involves a combination of what Kingman [18] called *natural coupling* and *temporal coupling*. The main argument of the proof is the same as that used in [20] and [21] for deriving similar recursions for the allele frequency spectrum and the number of segregating sites. The recursion for the summary statistics K_n is simpler than that for the allele frequency spectrum presented in [20]. It is therefore more useful to compute the distribution and other related functionals of the distribution of K_n for moderate values of n in reasonable time. Moreover, Theorem 1.1 is valid

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

for any arbitrary Ξ-coalescent. Our second result (Theorem 1.2, below) concerns measures Ξ satisfying

$$\Xi(\{0\}) = 0 \quad \text{and} \quad \int_{\Delta\setminus\{0\}} \frac{|x|}{(x,x)}\,\Xi(\mathrm{d}x) < \infty. \tag{1.6}$$

Recall that $|x| := \sum_{i=1}^{\infty} x_i$ and that $(x,x) := \sum_{i=1}^{\infty} x_i^2$ for $x = (x_1, x_2, \ldots) \in \Delta$. Note that (1.6) prevents Ξ from having too much mass near 0. Schweinsberg [26, Proposition 30] showed that the Ξ-coalescent does not have proper frequencies if and only if (1.6) holds. Not having proper frequencies is equivalent to having a positive fraction of singleton blocks with positive probability, which is particularly important for our convergence result presented in Theorem 1.2, below. For the special class of coalescent processes with multiple collisions (Λ-coalescents), (1.6) takes the form

$$\Lambda(\{0\}) = 0 \quad \text{and} \quad \int_{(0,1]} u^{-1}\Lambda(\mathrm{d}u) < \infty. \tag{1.7}$$

Pitman [23, Theorem 8] showed that the Λ-coalescent does not have proper frequencies if and only if (1.7) holds. Condition (1.7) excludes important examples such as the Kingman coalescent and the Bolthausen–Sznitman coalescent [5]. However, it includes, for example, all beta(a,b)-coalescents with parameters $a > 1$ and $b > 0$, which are studied in more detail in Section 5. Note that Theorem 1.2 covers a substantial class of Ξ-coalescents.

Theorem 1.2. *Suppose that the characterizing measure Ξ of the exchangeable coalescent $(R_t)_{t\ge 0}$ satisfies (1.6). Then K_n/n converges weakly as $n \to \infty$ to $K := r\int_0^\infty \mathrm{e}^{-rt}\mathrm{e}^{-X_t}\,\mathrm{d}t$, where $X = (X_t)_{t\ge 0}$ is a subordinator with Laplace exponent*

$$\Phi(\eta) = \int_{\Delta\setminus\{0\}} (1-(1-|x|)^{\eta})\frac{\Xi(\mathrm{d}x)}{(x,x)}, \qquad \eta \ge 0.$$

The limiting variable K has moments

$$\mathrm{E}(K^j) = \frac{r^j j!}{(r+\Phi(1))(2r+\Phi(2))\cdots(jr+\Phi(j))}, \qquad j \in \mathbb{N}. \tag{1.8}$$

We will see that the subordinator X appearing in Theorem 1.2 is related to the frequency of singletons, S_t, of R_t via $X_t = -\log S_t$, $t \ge 0$. Our proof of Theorem 1.2 is not based on the recursion presented in Theorem 1.1. It is rather a consequence of the chain of inequalities

$$M_n \le K_n \le N_n + 1,$$

where M_n denotes the number of mutated external branches and N_n denotes the total number of mutated branches of the restricted coalescent tree $(\varrho_n R_t)_{t\ge 0}$. Here we call a branch mutated if it is affected by at least one mutation. In a first step we show in Section 3 that Theorem 1.2 is valid with K_n replaced by the lower bound M_n. Then in Section 4 we verify that $(N_n - M_n)/n \to 0$ in probability (even in L^1), which completes the proof of Theorem 1.2 and, in addition, shows that Theorem 1.2 remains valid with K_n replaced by N_n. Note that if $K_1(n)$ denotes the number of types which appear exactly once in the sample of size n then $M_n \le K_1(n) \le K_n$, and, consequently, Theorem 1.2 also remains valid with K_n replaced by $K_1(n)$. Theorem 1.2 leaves open the question about the asymptotic behavior of K_n for the important class of Ξ-coalescents which do not satisfy condition (1.6). As mentioned before, some results for particular Λ-coalescents are known (see [1]–[3], [11], [20]); however, the problem concerning the asymptotic behavior of K_n for the full class of Ξ-coalescents remains open.

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

2. A recursion for the number of types

The proof of Theorem 1.1 is based on two fundamental properties of coalescent processes, which Kingman [18] called *natural coupling* and *temporal coupling*. We define natural coupling as follows. Suppose that the genealogy of a sample of size $n \in \mathbb{N}$ governed by a Ξ-coalescent is given. If a sub-sample of size $m \in \{1, \dots, n-1\}$ of this sample is taken, i.e. if $n-m$ individuals are removed from the sample, then the genealogical tree of the remaining sample of size m is governed by the same Ξ-coalescent. This consistency relation between different sample sizes is one of the fundamental properties of exchangeable coalescents. In fact, it is needed to prove the existence of exchangeable coalescent processes with state space $\mathcal{E}$ via Kolmogoroff's extension theorem. We define temporal coupling as follows. Consider a restricted coalescent process $(R_t^{(n)})_{t\geq 0} := (\varrho_n R_t)_{t\geq 0}$, and let $T_n := \inf\{t > 0 \colon R_t^{(n)} \neq R_0^{(n)}\}$ denote the time of its first jump. If after the first jump individuals belonging to the same equivalence class are identified, then the process started at time T_n is distributed as a coalescent with sample size $|R_{T_n}^{(n)}|$. Mathematically, this property essentially boils down to the strong Markov property. We will now verify Theorem 1.1.

Proof of Theorem 1.1. Recursion (1.4) is equivalent to $\mathrm{P}(K_1 = 1) = 1$ and

$$\mathrm{P}(K_n = k) = \frac{nr}{g_n + nr}\,\mathrm{P}(K_{n-1} = k-1) + \frac{g_n}{g_n + nr}\sum_{i=k}^{n-1} r_{ni}\,\mathrm{P}(K_i = k) \tag{2.1}$$

for $n \in \{2, 3, \dots\}$ and $k \in \{1, \dots, n\}$. We verify (2.1) analogously to the proofs presented in [20], by looking at the first event (either a coalescence or a mutation) that happens backward in time. The time W_n back to the first mutation is exponentially distributed with parameter nr. The time T_n back to the first coalescence is independent of W_n and exponentially distributed with parameter g_n. Thus, the first event backward in time is a mutation with probability $\mathrm{P}(W_n < T_n) = nr/(g_n + nr)$, and a coalescence with the complementary probability $\mathrm{P}(T_n < W_n) = g_n/(g_n + nr)$. Note that these two probabilities appear on the right-hand side of (2.1). Assume that the first event backward in time is a mutation. If we disregard the individual which is affected by this mutation, the number of types decreases by 1. Moreover, from the natural coupling property, it follows that the remaining tree is distributed as a coalescent restricted to the set $\{1, \dots, n-1\}$. This argument explains the appearance of the probability $\mathrm{P}(K_{n-1} = k-1)$ on the right-hand side of (2.1). If the first event backward in time is a coalescence then at the time of that coalescence event the coalescent process jumps to a partition with i blocks, $i \in \{1, \dots, n-1\}$, with probability $r_{ni} = g_{ni}/g_n$. By the temporal coupling property, the coalescent process stopped at that time is distributed as a coalescent restricted to the set $\{1, \dots, i\}$. As the number of types is not affected by a coalescence, the appearance of the sum on the right-hand side of (2.1) is explained. Note that it suffices to run the sum from k to $n-1$ as $\mathrm{P}(K_i = k) = 0$ for $i < k$.

Remarks. 1. In terms of the generating function $f_n(s) := \mathrm{E}(s^{K_n})$, $n \in \mathbb{N}$, $s \in \mathbb{C}$, recursion (1.4) (or (2.1)) is equivalent to $f_1(s) = s$ and

$$(g_n + nr)\,f_n(s) = nrs f_{n-1}(s) + \sum_{k=1}^{n-1} g_{nk} f_k(s), \qquad n \in \{2, 3, \dots\},\ s \in \mathbb{C}, \tag{2.2}$$

a formula which also follows (at least for coalescent processes with multiple collisions) by taking $s_1 = \cdots = s_n =: s$ in Equation (4) of [19].

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

2. Recursion (2.1) for the distribution of K_n is useful to compute the probabilities $\mathrm{P}(K_n = k)$ successively for $k = n, n-1, \dots, 1$. For example, for $k = n$, it follows that

$$(g_n + nr)\,\mathrm{P}(K_n = n) = nr\,\mathrm{P}(K_{n-1} = n-1)$$

and, therefore,

$$\mathrm{P}(K_n = n) = \prod_{i=2}^{n} \frac{ir}{g_i + ir} = \frac{r^{n-1} n!}{\prod_{i=2}^{n}(g_i + ir)}, \qquad n \in \mathbb{N}.$$

Note that $\mathrm{P}(K_n = n)$ is the probability of having only singletons in a sample of size n.

Example. (*Kingman coalescent.*) For the Kingman coalescent ($\Lambda = \delta_0$), we have $I_n \equiv n-1$, $g_n = g_{n,n-1} = n(n-1)/2$, and $g_{ni} = 0$ for $i \in \{1, \dots, n-2\}$. Recursion (1.4) reduces to $K_n \stackrel{\mathrm{D}}{=} B_n + K_{n-1}$. Therefore, $K_n \stackrel{\mathrm{D}}{=} \sum_{i=1}^{n} B_i$, $n \in \mathbb{N}$, where $B_1, B_2, \dots$ are independent Bernoulli variables with $\mathrm{P}(B_n = 1) = nr/(g_n + nr) = \theta/(\theta + n - 1)$, $n \in \mathbb{N}$ and $\theta := 2r$. It follows easily that $\mathrm{P}(K_n = k) = \theta^k s(n,k)/[\theta]_n$, where $[\theta]_n := \theta(\theta+1)\cdots(\theta+n-1)$ and the $s(n,k)$ denote the absolute Stirling numbers of the first kind. Moreover, $\mathrm{E}(K_n) = \theta \sum_{i=0}^{n-1} 1/(\theta+i) \sim \theta \log n$ and $\mathrm{var}(K_n) = \theta \sum_{i=1}^{n-1} i/(\theta+i)^2 \sim \theta \log n$. By the Lindeberg–Feller central limit theorem, $(K_n - \theta \log n)/\sqrt{\theta \log n}$ is asymptotically standard normal distributed. All these results are of course well known and go at least back to the seminal work of Ewens [11].

Example. (*Star-shaped coalescent.*) For the star-shaped coalescent ($\Lambda = \delta_1$), we have $I_n \equiv 1$, $g_{n1} = g_n = 1$, and $g_{ni} = 0$ for $i \in \{2, \dots, n-1\}$. Therefore, (2.2) reduces to $(1 + nr) f_n(s) = nrs f_{n-1}(s) + s$, $n \in \{2, 3, \dots\}$, $s \in \mathbb{C}$. We refer the reader to [20, Section 4] for more details. In particular, in [20] it was shown that K_n/n converges almost surely to a limiting random variable K, beta distributed with parameter 1 and $1/r$, that is, $\mathrm{P}(K > x) = (1-x)^{1/r}$, $0 < x < 1$.

Remark. (*Recursion for the factorial moments of K_n.*) Taking the jth derivative with respect to s in (2.2) and applying the Leibniz rule yields

$$(g_n + nr) f_n^{(j)}(s) = nr(s f_{n-1}^{(j)}(s) + j f_{n-1}^{(j-1)}(s)) + \sum_{k=1}^{n-1} g_{nk} f_k^{(j)}(s)$$

for $n \in \{2, 3, \dots\}$, $j \in \mathbb{N}$, and $s \in \mathbb{C}$. For $n \in \mathbb{N}$ and $j \in \mathbb{N}_0$, let $\mu_n^{(j)} := \mathrm{E}((K_n)_j) = \mathrm{E}(K_n(K_n-1)\cdots(K_n-j+1))$ denote the jth descending factorial moment of K_n. Taking the limit $s \to 1$, it follows that

$$(g_n + nr)\mu_n^{(j)} = nr(\mu_{n-1}^{(j)} + j\mu_{n-1}^{(j-1)}) + \sum_{k=1}^{n-1} g_{nk}\mu_k^{(j)}, \qquad n \in \{2, 3, \dots\},\ j \in \mathbb{N}.$$

This recursion with initial condition $\mu_1^{(j)} = \delta_{j1}$ (Kronecker symbol) is useful to compute the factorial moments of K_n. For example, for $j = n$, we have $(g_n + nr)\mu_n^{(n)} = n^2 r \mu_{n-1}^{(n-1)}$ and, therefore,

$$\mu_n^{(n)} = \prod_{i=2}^{n} \frac{i^2 r}{g_i + ir} = \frac{r^{n-1}(n!)^2}{\prod_{i=2}^{n}(g_i + ir)}, \qquad n \in \mathbb{N},$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

a result which also follows from $\mu_n^{(n)} = n!\,\mathrm{P}(K_n = n)$. In particular, the first moment $\mu_n := \mu_n^{(1)} = \mathrm{E}(K_n)$ follows the recursion $\mu_1 = 1$ and

$$(g_n + nr)\mu_n = nr(\mu_{n-1} + 1) + \sum_{k=1}^{n-1} g_{nk}\mu_k, \qquad n \in \{2, 3, \ldots\}.$$

It seems to be nontrivial to solve any of these recursions except for the Kingman coalescent ($\Lambda = \delta_0$) and the star-shaped coalescent ($\Lambda = \delta_1$). We therefore focus on asymptotic results for K_n as the sample size n tends to ∞.

3. The number of mutated external branches

We say that a branch of the restricted coalescent tree $(\varrho_n R_t)_{t\geq 0}$ is mutated if it is affected by at least one mutation. In this section we study the asymptotics of the number, M_n, of mutated external branches of $(\varrho_n R_t)_{t\geq 0}$ under the assumption that the measure Ξ satisfies condition (1.6).

Lemma 3.1. *Suppose that the characterizing measure Ξ of the exchangeable coalescent process $R = (R_t)_{t\geq 0}$ satisfies (1.6). Then, $M_n/n \xrightarrow{\mathrm{D}} M$ as $n \to \infty$, where M is a random variable uniquely determined by its moments*

$$\mathrm{E}(M^k) = \mathrm{E}\biggl(\prod_{i=1}^{k}(1 - \mathrm{e}^{-rL_i})\biggr), \qquad k \in \mathbb{N},$$

with $L_i := \sup\{t > 0\colon \{i\}$ is a block of $R_t\}$, $i \in \mathbb{N}$. (By '$\xrightarrow{\mathrm{D}}$' we denote convergence in distribution.)

Proof. For $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$, let

$$L_{n,i} := \sup\{t > 0\colon \{i\} \text{ is a block of } \varrho_n R_t\}$$

denote the length of the ith external branch of the restricted coalescent tree $(\varrho_n R_t)_{t\geq 0}$. Fix $k \in \mathbb{N}$ and $t_1, \ldots, t_k \in [0, \infty)$. For $n \geq k$, we have

$$\begin{aligned}
&\mathrm{P}(L_{n,1} > t_1, \ldots, L_{n,k} > t_k)\\
&\quad = \mathrm{P}(\{1\} \text{ is a block of } \varrho_n R_{t_1}, \ldots, \{k\} \text{ is a block of } \varrho_n R_{t_k})\\
&\quad \to \mathrm{P}\biggl(\bigcap_{n\in\mathbb{N}} \{\{1\} \text{ is a block of } \varrho_n R_{t_1}, \ldots, \{k\} \text{ is a block of } \varrho_n R_{t_k}\}\biggr)\\
&\quad = \mathrm{P}(\{1\} \text{ is a block of } R_{t_1}, \ldots, \{k\} \text{ is a block of } R_{t_k})\\
&\quad = \mathrm{P}(L_1 > t_1, \ldots, L_k > t_k).
\end{aligned}$$

Thus, for all $k \in \mathbb{N}$, $(L_{n,1}, \ldots, L_{n,k}) \xrightarrow{\mathrm{D}} (L_1, \ldots, L_k)$ as $n \to \infty$. For $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$, let $E_{n,i}$ denote the event that the ith external branch of the restricted tree $(\varrho_n R_t)_{t\geq 0}$ is affected by at least one mutation. Conditional on the lengths $L_{n,1}, \ldots, L_{n,n}$ of the external branches, the mutation Poisson process with parameter $r > 0$ acts independently on all these branches.

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

Thus, for fixed $j \in \mathbb{N}$, we have

$$\begin{aligned}\mathrm{P}(E_{n,1} \cap \cdots \cap E_{n,j}) &= \mathrm{E}(\mathrm{P}(E_{n,1} \cap \cdots \cap E_{n,j} \mid L_{n,1}, \ldots, L_{n,j})) \\ &= \mathrm{E}(\mathrm{P}(E_{n,1} \mid L_{n,1}) \cdots \mathrm{P}(E_{n,j} \mid L_{n,j})) \\ &= \mathrm{E}((1 - \mathrm{e}^{-rL_{n,1}}) \cdots (1 - \mathrm{e}^{-rL_{n,j}})) \\ &\to \mathrm{E}((1 - \mathrm{e}^{-rL_1}) \cdots (1 - \mathrm{e}^{-rL_j})).\end{aligned}$$

From $M_n = \sum_{i=1}^n \mathbf{1}_{E_{n,i}}$, it follows that

$$\mathrm{E}(M_n^k) = \mathrm{E}\left(\left(\sum_{i=1}^n \mathbf{1}_{E_{n,i}}\right)^k\right) = \sum_{i_1,\ldots,i_k=1}^n \mathrm{E}(\mathbf{1}_{E_{n,i_1}} \cdots \mathbf{1}_{E_{n,i_k}}).$$

For each fixed n, the events $E_{n,i},\ i \in \{1, \ldots, n\}$, are exchangeable. Therefore,

$$\mathrm{E}(M_n^k) = \sum_{j=1}^k S(k, j)(n)_j \,\mathrm{P}(E_{n,1} \cap \cdots \cap E_{n,j}),$$

where $S(k, j)$ denotes the Stirling number of the second kind, i.e. the number of ways to partition a set with k elements into j nonempty subsets. Division by n^k and taking the limit $n \to \infty$ yields, for all $k \in \mathbb{N}_0$,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}\left(\left(\frac{M_n}{n}\right)^k\right) &= \sum_{j=1}^k S(k, j) \lim_{n\to\infty} \frac{(n)_j}{n^k} \mathrm{P}(E_{n,1} \cap \cdots \cap E_{n,j}) \\ &= \lim_{n\to\infty} \mathrm{P}(E_{n,1} \cap \cdots \cap E_{n,k}) \\ &= \mathrm{E}((1 - \mathrm{e}^{-rL_1}) \cdots (1 - \mathrm{e}^{-rL_k})) \\ &=: \mu_k.\end{aligned}$$

For all $m, k \in \mathbb{N}_0$,

$$\begin{aligned}\sum_{j=0}^m \binom{m}{j} (-1)^j \mu_{k+j} &= \lim_{n\to\infty} \mathrm{E}\left(\sum_{j=0}^m \binom{m}{j} (-1)^j \left(\frac{M_n}{n}\right)^{k+j}\right) \\ &= \lim_{n\to\infty} \mathrm{E}\left(\left(\frac{M_n}{n}\right)^k \left(1 - \frac{M_n}{n}\right)^m\right) \\ &\geq 0.\end{aligned}$$

Thus (Hausdorff moment problem), the sequence $(\mu_k)_{k\in\mathbb{N}_0}$ is a moment sequence of some random variable M taking values in the unit interval $[0, 1]$. The convergence of moments implies the convergence $M_n/n \xrightarrow{\mathrm{D}} M$.

Remark. The interpretation of the distribution of the limiting external branch lengths $L_i,\ i \in \mathbb{N}$, in terms of the frequency spectrum of the coalescent is as follows. Let S_t denote the frequency of singletons of R_t. Conditional on $S_{t_1}, \ldots, S_{t_k}$, the probability that i is still a singleton at time $t_i,\ i \in \{1, \ldots, k\}$, is $S_{t_1} \cdots S_{t_k}$. Therefore, for $t_1, \ldots, t_k \in [0, \infty)$,

$$\mathrm{P}(L_1 > t_1, \ldots, L_k > t_k) = \mathrm{E}(S_{t_1} \cdots S_{t_k}),$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

or, equivalently (in agreement with the principle of inclusion and exclusion),

$$\mathrm{P}(L_1 \le t_1, \dots, L_k \le t_k) = \mathrm{E}((1 - S_{t_1}) \cdots (1 - S_{t_k})).$$

Thus, the distribution function of $(L_1, \dots, L_k)$ can be expressed in terms of the process $S = (S_t)_{t \ge 0}$.

Corollary 3.1, below, expresses the distribution of the limiting random variable M appearing in Lemma 3.1 in terms of the process $(S_t)_{t \ge 0}$. There is the following rough intuition for the form of the integral in Corollary 3.1. A contribution to M_n occurs every time a lineage that has not yet coalesced experiences its first mutation. The time of a first mutation is exponentially distributed with parameter r, so at each time t the infinitesimal growth of M_n due to a not yet coalesced lineage is $r\mathrm{e}^{-rt}$. Since S_t is the fraction of singletons at time t, the infinitesimal growth of M_n at time t is approximately $r\mathrm{e}^{-rt} n S_t$. In [21], in which the number of segregating sites is the quantity of interest, any mutation contributes to the count rather than just the first one, so we get r in Proposition 5.1 of [21] in place of the $r\mathrm{e}^{-rt}$ in Corollary 3.1.

Corollary 3.1. *The limiting variable M appearing in Lemma 3.1 satisfies*

$$M \overset{\mathrm{D}}{=} r \int_0^\infty \mathrm{e}^{-rt} S_t \, \mathrm{d}t.$$

Proof. Fix $k \in \mathbb{N}$, and define $g \colon \mathbb{R}^k \to \mathbb{R}$ via

$$g(t) := (-1)^k \exp(-r(t_1 + \cdots + t_k)) \quad \text{for } t = (t_1, \dots, t_k) \in \mathbb{R}^k.$$

Note that

$$h(t) := \frac{\partial^k}{\partial t_1 \cdots \partial t_k} g(t) = r^k \exp(-r(t_1 + \cdots + t_k)).$$

For $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k) \in \mathbb{R}^k$ with $x_i \le y_i$ for all $1 \le i \le k$, define

$$\Delta_x^y g := \sum_{\varepsilon_1, \dots, \varepsilon_k \in \{0,1\}} (-1)^{\varepsilon_1 + \cdots + \varepsilon_k} g(\varepsilon_1 x_1 + (1 - \varepsilon_1) y_1, \dots, \varepsilon_k x_k + (1 - \varepsilon_k) y_k).$$

Define $L := (L_1, \dots, L_k)$, and let P_L the distribution of L. Then we have

$$\begin{aligned} \mathrm{E}(M^k) &= \mathrm{E}((1 - \mathrm{e}^{-rL_1}) \cdots (1 - \mathrm{e}^{-rL_k})) \\ &= \mathrm{E}(\Delta_0^L g) \\ &= \int_{\mathbb{R}_+^k} \Delta_0^y g \, \mathrm{P}_L(\mathrm{d}y) \\ &= \int_{\mathbb{R}_+^k} \int_{\mathbb{R}_+^k} \mathbf{1}_{[0,y)}(t) h(t) \lambda^k(\mathrm{d}t) \, \mathrm{P}_L(\mathrm{d}y). \end{aligned}$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

An application of Fubini's theorem yields

$$\begin{aligned}\mathrm{E}(M^k) &= \int_{\mathbb{R}_+^k} h(t) \int_{\mathbb{R}_+^k} \mathbf{1}_{(t,\infty)}(y)\, \mathrm{P}_L(\mathrm{d}y)\lambda^k(\mathrm{d}t) \\ &= \int_{\mathbb{R}_+^k} h(t)\, \mathrm{P}(L > t)\lambda^k(\mathrm{d}t) \\ &= \int_{\mathbb{R}_+^k} r^k \exp(-r(t_1 + \cdots + t_k))\, \mathrm{E}(S_{t_1} \cdots S_{t_k})\lambda^k(\mathrm{d}t_1, \ldots, \mathrm{d}t_k) \\ &= \mathrm{E}\biggl(\biggl(\int_0^\infty r\mathrm{e}^{-rt} S_t\, \mathrm{d}t\biggr)^k\biggr).\end{aligned}$$

Thus, the moments of the random variables M and $\int_0^\infty r\mathrm{e}^{-rt} S_t\, \mathrm{d}t$ coincide. As both random variables almost surely take values in the unit interval $[0, 1]$, they are equal in distribution.

The moments of M can be expressed in terms of the measure Ξ as follows.

Remark. Assume that the measure Ξ of the exchangeable coalescent $(R_t)_{t\geq 0}$ satisfies (1.6). From the Poisson construction of the Ξ-coalescent (see [26]), it follows that the process $X = (X_t)_{t\geq 0}$, defined via $X_t := -\log S_t$ for $t \geq 0$, is a drift-free subordinator with Laplace exponent

$$\Phi(\eta) = \int_{\Delta\setminus\{0\}} \frac{1 - (1 - |x|)^\eta}{(x, x)}\, \Xi(\mathrm{d}x), \qquad \eta \geq 0.$$

Note that, for $\eta \in \mathbb{N}$, $\mathrm{e}^{-t\Phi(\eta)} = \mathrm{E}(\mathrm{e}^{-\eta X_t}) = \mathrm{E}(S_t^\eta)$ is the probability that $\{1\}, \ldots, \{\eta\}$ are (singleton) blocks of R_t. The Lévy measure ϱ on $(0, \infty]$ of the subordinator X is hence the image of the measure $\nu(\mathrm{d}x) := \Xi(\mathrm{d}x)/(x, x)$ via the transformation $T(x) := -\log(1 - |x|)$, i.e. $\varrho(A) = \int_{T^{-1}(A)} (x, x)^{-1}\, \Xi(\mathrm{d}x)$ for all Borel subsets A of $(0, \infty]$. This result is in agreement with Proposition 26 of [23] for the special situation when the coalescent allows for only multiple collisions (Λ-coalescent). From

$$\int_{\Delta\setminus\{0\}} \frac{|x|}{(x, x)}\, \Xi(\mathrm{d}x) = \int_{\Delta\setminus\{0\}} |x|\nu(\mathrm{d}x) = \int_{(0,\infty]} (1 - \mathrm{e}^{-y})\varrho(\mathrm{d}y)$$

and $(1 - \mathrm{e}^{-1}) \min(y, 1) \leq 1 - \mathrm{e}^{-y} \leq \min(y, 1),\ y \geq 0$, it follows that (1.6) is equivalent to

$$\varrho(\{0\}) = 0 \quad \text{and} \quad \int_{(0,\infty]} \min(y, 1)\varrho(\mathrm{d}y) < \infty.$$

Note that the finiteness of the last integral is the typical condition for a measure ϱ to be a Lévy measure of some subordinator. From Proposition 3.1 of [7], it follows that $M \stackrel{\mathrm{D}}{=} r\int_0^\infty \mathrm{e}^{-rt-X_t}\, \mathrm{d}t$ has moments

$$\mathrm{E}(M^k) = \frac{r^k k!}{(r + \Phi(1))(2r + \Phi(2)) \cdots (kr + \Phi(k))}, \qquad k \in \mathbb{N}. \tag{3.1}$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

In particular,

$$\begin{aligned}\operatorname{var}(M) &= \mathrm{E}(M^2) - (\mathrm{E}(M))^2 \\ &= \frac{2r^2}{(r+\Phi(1))(2r+\Phi(2))} - \frac{r^2}{(r+\Phi(1))^2} \\ &= \frac{2r^2(r+\Phi(1)) - r^2(2r+\Phi(2))}{(r+\Phi(1))^2(2r+\Phi(2))} \\ &= \frac{r^2}{(r+\Phi(1))^2(2r+\Phi(2))} \int_{\Delta\setminus\{0\}} \frac{|x|^2}{(x,x)}\,\Xi(\mathrm{d}x), \end{aligned} \tag{3.2}$$

as $2\Phi(1) - \Phi(2) = \int_{\Delta\setminus\{0\}} |x|^2/(x,x)\,\Xi(\mathrm{d}x)$.

In the final remark of this section a distributional fixed-point equation for M is derived for Ξ-coalescents satisfying

$$\Xi(\{0\}) = 0 \quad \text{and} \quad \int_{\Delta\setminus\{0\}} \frac{\Xi(\mathrm{d}x)}{(x,x)} < \infty. \tag{3.3}$$

In the spirit of Bertoin and Le Gall [4] we call measures Ξ satisfying (3.3) simple measures. Note that (3.3) implies (1.6).

Remark. If (3.3) holds then the Lévy measure ϱ of the subordinator $X = (X_t)_{t\geq 0}$ is finite ($m_0 := \varrho((0,\infty]) = \nu(\Delta\setminus\{0\}) < \infty$), which means that X is a compound Poisson process $X_t = \sum_{i=1}^{N_t} \eta_i$, where $N := (N_t)_{t\geq 0}$ is a homogeneous Poisson process with parameter m_0 and the η_i, $i \in \mathbb{N}$, are random variables, independent of each other and of N, with common distribution function $y \mapsto \mathrm{P}(\eta_i \leq y) = m_0^{-1}\varrho((0,y])$. Let $T_1 < T_2 < T_3 < \cdots$ denote the jump times of the Poisson process N. Note that $T_{i+1} - T_i$ is exponentially distributed with parameter m_0. We have

$$\begin{aligned} M &\overset{\mathrm{D}}{=} \int_0^\infty r\mathrm{e}^{-rt} S_t\,\mathrm{d}t \\ &= \sum_{i=0}^{\infty} \int_{T_i}^{T_{i+1}} r\mathrm{e}^{-rt} S_t\,\mathrm{d}t \\ &= \int_0^{T_1} r\mathrm{e}^{-rt}\,\mathrm{d}t + \mathrm{e}^{-\eta_1}\int_{T_1}^{T_2} r\mathrm{e}^{-rt}\,\mathrm{d}t + \mathrm{e}^{-\eta_1-\eta_2}\int_{T_2}^{T_3} r\mathrm{e}^{-rt}\,\mathrm{d}t + \cdots \\ &= (1-\mathrm{e}^{-rT_1}) + \mathrm{e}^{-\eta_1}(\mathrm{e}^{-rT_1} - \mathrm{e}^{-rT_2}) + \mathrm{e}^{-\eta_1-\eta_2}(\mathrm{e}^{-rT_2} - \mathrm{e}^{-rT_3}) + \cdots \\ &= (1-\mathrm{e}^{-rT_1}) + \mathrm{e}^{-\eta_1}\mathrm{e}^{-rT_1}((1-\mathrm{e}^{-r(T_2-T_1)}) + \mathrm{e}^{-\eta_2}(\mathrm{e}^{-r(T_2-T_1)} - \mathrm{e}^{-r(T_3-T_1)}) + \cdots) \\ &= B + A(1-B)M_1, \end{aligned}$$

with $A := \mathrm{e}^{-\eta_1}$, $B := 1-\mathrm{e}^{-rT_1}$, and $M_1 \overset{\mathrm{D}}{=} M$. Thus, M satisfies the distributional fixed-point equation

$$M \overset{\mathrm{D}}{=} B + A(1-B)M, \tag{3.4}$$

where A and B are independent (and independent of M), B is beta distributed with parameters 1 and m_0/r, i.e. $\mathrm{P}(B > x) = (1-x)^{m_0/r}$, $x \in (0,1)$, and the distribution of $1-A$ is the image of the measure $\nu_0 := \nu/m_0$ under the transformation $|\cdot|\colon \Delta\setminus\{0\} \to (0,1]$, $x \mapsto |x|$. Using an argument similar to that of Vervaat [27], it can be shown that the distribution of

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

M is uniquely determined by the fixed-point equation (3.4). The distribution of M coincides with the stationary distribution of the process $(Y_n)_{n\in\mathbb{N}_0}$ recursively defined by $Y_0 := 0$ and $Y_{n+1} := A_n(1 - B_n)Y_n + B_n$, where $((A_n, B_n))_{n\in\mathbb{N}_0}$ is a sequence of independent, identically distributed random variables with $(A_n, B_n) \stackrel{\mathrm{D}}{=} (A, B)$. Note that

$$Y_n = \sum_{i=0}^{n-1} B_{n-i-1} \prod_{j=n-i}^{n-1} A_j(1 - B_j) \stackrel{\mathrm{D}}{=} \sum_{i=0}^{n-1} B_i \prod_{j=0}^{i-1} A_j(1 - B_j), \qquad n \in \mathbb{N}_0,$$

and, hence, that $M \stackrel{\mathrm{D}}{=} \sum_{i=0}^{\infty} B_i \prod_{j=0}^{i-1} A_j(1 - B_j)$.

4. The total number of mutated branches

In order to analyze the total number, N_n, of mutated branches, we need to study C_n, the number of collision events that take place in the restricted coalescent process $(\varrho_n R_t)_{t\geq 0}$ until there is just a single block. Note that, in general, $C_n \geq X_n$, the number of jumps. For Λ-coalescents, we have $C_n = X_n$.

Lemma 4.1. *Let R be a Ξ-coalescent. If (1.6) holds then $C_n/n \to 0$ in L^1.*

Proof. For $n \in \mathbb{N}$, define $a_n := \mathrm{E}(C_n)$ for convenience. Note that the sequence $(a_n)_{n\in\mathbb{N}}$ satisfies the recursion $a_1 = 0$ and $a_n = v_n + \sum_{k=1}^{n-1} r_{nk} a_k$ for $n \in \{2, 3, \ldots\}$ with $r_{nk} := \mathrm{P}(I_n = k)$, $n, k \in \mathbb{N}$, $k < n$, and $v_n := \mathrm{E}(V_n)$, where V_n denotes the number of collision events of the restricted coalescent process $(\varrho_n R_t)_{t\geq 0}$ that take place at the time of the first jump of $(\varrho_n R_t)_{t\geq 0}$. Note that $V_n \equiv 1$ for coalescents with only multiple (no simultaneous multiple) collisions. We verify the convergence $C_n/n \to 0$ in L^1 by contradiction, in analogy to Gnedin's proof of Proposition 3 of [13]. Note that a similar argument is used on p. 219 of [16]. Assume that there exists $\varepsilon > 0$ such that $a_n > n\varepsilon$ for infinitely many values of n. Selecting ε smaller, for any fixed c, we can obtain the inequality $a_n > \varepsilon n + c$ for infinitely many values of n. Let n_c be the minimum such n. Then $n_c \to \infty$ as $c \to \infty$. For $k < n_c$, we have $a_k \leq \varepsilon k + c$, which implies that

$$\begin{aligned} \varepsilon n_c + c < a_{n_c} &= v_{n_c} + \sum_{k=1}^{n_c-1} r_{n_c,k} a_k \\ &\leq v_{n_c} + c + \varepsilon \sum_{k=1}^{n_c-1} k r_{n_c,k} \\ &= v_{n_c} + c + \varepsilon\, \mathrm{E}(I_{n_c}). \end{aligned}$$

The constant c cancels and it follows that $\varepsilon\, \mathrm{E}(n_c - I_{n_c}) < v_{n_c}$. For $c \to \infty$, we obtain the promised contradiction, since $\mathrm{E}(n - I_n)/v_n \to \infty$ as $n \to \infty$ by Corollary A.2 in Appendix A. Thus, for all $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that $a_n/n \leq \varepsilon$ for all $n \geq n_0$. In other words, $a_n/n \to 0$ as $n \to \infty$.

We are now able to show that if (1.6) holds then the total number, N_n, of mutated branches and the number, K_n, of types both have the same asymptotic behavior as M_n as $n \to \infty$.

Corollary 4.1. *Let $(R_t)_{t\geq 0}$ be a Ξ-coalescent with mutation rate $r > 0$ satisfying (1.6). Then, $N_n/n \stackrel{\mathrm{D}}{\to} M$, and $K_n/n \stackrel{\mathrm{D}}{\to} M$ as well, where M is the random variable defined in Corollary 3.1 with moments (3.1).*

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

Proof. We have $M_n \le K_n \le N_n + 1$. Thus, by Lemma 3.1, it suffices to verify that $(N_n + 1 - M_n)/n \to 0$ in probability. We even show that $(N_n + 1 - M_n)/n \to 0$ in L^1. We have

$$\begin{aligned} 0 &\le K_n - M_n \\ &\le N_n + 1 - M_n \\ &= \text{number of nonexternal mutated branches} + 1 \\ &\le \text{number of nonexternal branches} + 1 \\ &= C_n. \end{aligned}$$

It remains to note that $C_n/n \to 0$ in L^1 by Lemma 4.1.

Note that Corollary 4.1 in particular completes the proof of Theorem 1.2.

5. Examples

In this section we apply Theorem 1.2 to some concrete examples.

Example 5.1. (*Dirac coalescents.*) Fix a point $c \in \Delta \setminus \{0\}$, and suppose that $\Xi = \delta_c$ is the Dirac measure in c. Then, condition (1.6) holds, as $\int_{\Delta\setminus\{0\}} (|x|/(x, x))\Xi(\mathrm{d}x) = |c|/(c, c) < \infty$. By Theorem 1.2, all three random variables, M_n/n, K_n/n, and N_n/n, converge in distribution to $M := r\int_0^\infty \mathrm{e}^{-rt-X_t}\,\mathrm{d}t$, where $X = (X_t)_{t\ge 0}$ is a subordinator with Laplace exponent $\Phi(\eta) = (1 - (1 - |c|)^\eta)/(c, c)$, $\eta \ge 0$. The Lévy measure $\varrho = (1/(c, c))\delta_{-\log(1-|c|)}$ is hence the Dirac measure in $-\log(1 - |c|)$ scaled by the factor $1/(c, c)$. We have $\Phi(1) = |c|/(c, c)$ and $\Phi(2) = |c|(2 - |c|)/(c, c)$, and, therefore, by (3.1) and (3.2),

$$\mathrm{E}(M) = \frac{r}{r + |c|/(c, c)}$$

and

$$\mathrm{var}(M) = \frac{r^2|c|^2/(c, c)}{(r + |c|/(c, c))^2(2r + |c|(2 - |c|)/(c, c))}.$$

Note that $m_0 := \int_{\Delta\setminus\{0\}} (1/(x, x))\Xi(\mathrm{d}x) = 1/(c, c) < \infty$, i.e. (3.3) holds as well. Thus, by (3.4), M satisfies the distributional fixed-point equation $M \stackrel{\mathrm{D}}{=} B + (1 - |c|)(1 - B)M$, where B is a random variable independent of M and beta distributed with parameters 1 and $m_0/r = 1/((c, c)r)$. Even for this quite simple situation of Dirac coalescents, it does not seem to be straightforward to find simpler characterizations for the distribution of M.

Example 5.2. (*Beta coalescents.*) Let Λ be beta distributed with parameters $a > 1$ and $b > 0$, i.e. Λ has density $u \mapsto (B(a, b))^{-1}u^{a-1}(1 - u)^{b-1}$, $u \in (0, 1)$, with respect to the Lebesgue measure on $(0, 1)$, where $B(\cdot, \cdot)$ denotes the beta function. In this situation we have

$$\int_{[0,1]} u^{-1}\Lambda(\mathrm{d}u) = \frac{B(a - 1, b)}{B(a, b)} = \frac{a + b - 1}{a - 1} < \infty.$$

Thus, Theorem 1.2 is applicable and all three random variables, M_n/n, K_n/n, and N_n/n, converge in distribution to $M := r\int_0^\infty \mathrm{e}^{-rt-X_t}\,\mathrm{d}t$ as $n \to \infty$, where $X = (X_t)_{t\ge 0}$ is a subordinator with Laplace exponent

$$\Phi(\eta) = \frac{1}{B(a, b)}\int_0^1 \frac{1 - (1 - u)^\eta}{u^2}u^{a-1}(1 - u)^{b-1}\,\mathrm{d}u, \qquad \eta \ge 0.$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

The expansion $1-(1-u)^\eta = \sum_{i=1}^\infty \binom{\eta}{i}(-1)^{i+1}u^i$ yields

$$\Phi(\eta) = \frac{1}{B(a,b)}\sum_{i=1}^\infty \binom{\eta}{i}(-1)^{i+1}B(a+i-2,b)$$
$$= \frac{a+b-1}{a-1}\sum_{i=1}^\infty \binom{\eta}{i}(-1)^{i+1}\prod_{j=1}^{i-1}\frac{a-2+j}{a+b-2+j}, \qquad \eta \geq 0.$$

Note that $\Phi(1) = (a+b-1)/(a-1)$ and $\Phi(2) = (a+2b-1)/(a-1)$. The mean and the variance of M can be easily deduced from (3.1) and (3.2). From $\varrho((0,y]) = \nu((0,1-\mathrm{e}^{-y}]) = \int_{(0,1-\mathrm{e}^{-y}]} u^{-2}\Lambda(\mathrm{d}u)$, it follows that the Lévy measure ϱ of the subordinator X has density $y \mapsto (B(a,b))^{-1}(1-\mathrm{e}^{-y})^{a-3}(\mathrm{e}^{-y})^b$, $y \in (0,\infty)$, with respect to the Lebesgue measure on $(0,\infty)$. If $a > 2$ then

$$m_0 := \int_{[0,1]} u^{-2}\Lambda(\mathrm{d}u) = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)} < \infty.$$

In this case, by (3.4), M satisfies the distributional fixed-point equation $M \stackrel{\mathrm{D}}{=} B + A(1-B)M$, where A and B are independent (and independent of M), $1-A$ is beta distributed with parameters $a-2$ and b, and B is beta distributed with parameters 1 and m_0/r. For special parameter values of a and b, the Laplace exponent Φ can be further simplified. For example, for the $\beta(2-\alpha,\alpha)$-coalescent with $0 < \alpha < 1$,

$$\Phi(\eta) = \frac{1}{1-\alpha}\sum_{i=1}^\infty \binom{\eta}{i}\binom{\alpha-1}{i-1} = \frac{\eta\Gamma(\eta+\alpha)}{(1-\alpha)\Gamma(\alpha+1)\Gamma(\eta+1)}, \qquad \eta \geq 0.$$

Note that if the conjecture on p. 495 of [1] is correct then we have identified (in the notation of [1]) the distribution of the random variable C_1, namely $C_1 \stackrel{\mathrm{D}}{=} M$.

Example 5.3. Suppose that the measure Ξ is concentrated on the subset Δ^* of all points $x \in \Delta$ satisfying $|x| = 1$ and that $m_0 := \int_{\Delta\setminus\{0\}}(1/(x,x))\Xi(\mathrm{d}x) < \infty$. Concrete examples are the star-shaped coalescent, where Ξ is the Dirac measure in $(1,0,0,\ldots)$, or the Poisson–Dirichlet coalescent with parameter $\theta > 0$, where Ξ is assumed to have density $x \mapsto (x,x)$ with respect to the Poisson–Dirichlet distribution with parameter $\theta > 0$. Then, (1.6) and (3.3) coincide and are both satisfied. Thus, Theorem 1.2 is applicable, i.e. all three random variables, M_n/n, K_n/n, and N_n/n, converge in distribution to a limiting variable K with moments (1.8). As the measure Ξ is concentrated on Δ^*, the Laplace exponent $\Phi(\eta) \equiv m_0$ is constant. Therefore, K has moments $\mathrm{E}(K^j) = r^j j!/((r+m_0)\cdots(jr+m_0))$, $j \in \mathbb{N}$. It follows that K is beta distributed with parameters 1 and m_0/r.

Appendix A

In this appendix basic results for Ξ-coalescents $R = (R_t)_{t\geq 0}$ are derived. We first restrict our attention to coalescents with (only) multiple collisions, as the proofs are in this case less technical. We then extend the results to Ξ-coalescents. Our first result (Lemma A.1) concerns the number of blocks I_n of the restricted coalescent process $(\varrho_n R_t)_{t\geq 0}$ after its first jump. Note that I_n has distribution (1.5) and that we define $I_1 := 0$ for convenience. Lemma A.1 is well known from the literature (see, for example, [25, Lemma 3]); however, we provide a proof which can be extended to the full class of coalescents with simultaneous multiple collisions (see Lemma A.2, below).

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

Lemma A.1. *Let R be a Λ-coalescent. Then, for all $n \in \mathbb{N}$,*

$$g_n \,\mathrm{E}(n - I_n) = \int_{[0,1]} \frac{(1-u)^n - 1 + nu}{u^2} \Lambda(\mathrm{d}u)$$

with continuous extension of the function below the integral for $u \searrow 0$.

Proof. We have

$$g_n \,\mathrm{E}(I_n) = \sum_{k=1}^{n-1} k g_{nk} = \sum_{k=1}^{n-1} k \int_{[0,1]} \binom{n}{k-1} u^{n-k-1} (1-u)^{k-1} \Lambda(\mathrm{d}u).$$

Substituting $i = k - 1$ and interchanging the summation with the integral yields

$$\begin{aligned} g_n \,\mathrm{E}(I_n) &= \int_{[0,1]} \sum_{i=0}^{n-2} (i+1) \binom{n}{i} u^{n-i} (1-u)^i \frac{\Lambda(\mathrm{d}u)}{u^2} \\ &= \int_{[0,1]} \frac{n(1-u) + 1 - n^2 u (1-u)^{n-1} - (n+1)(1-u)^n}{u^2} \Lambda(\mathrm{d}u) \\ &= \int_{[0,1]} \frac{n + 1 - nu - n^2 u (1-u)^{n-1} - (n+1)(1-u)^n}{u^2} \Lambda(\mathrm{d}u). \end{aligned}$$

Now subtract this expression from

$$n g_n = \int_{[0,1]} \frac{n - n(1-u)^n - n^2 u (1-u)^{n-1}}{u^2} \Lambda(\mathrm{d}u).$$

Corollary A.1. *If (1.7) holds then $\mathrm{E}(n - I_n) \sim (n/g_n) \int_{[0,1]} u^{-1} \Lambda(\mathrm{d}u) \to \infty$ as $n \to \infty$.*

Proof. For $n \in \mathbb{N}$, define the auxiliary function $H(n) := \int_{[0,1]} (1 - (1-u)^n) u^{-2} \Lambda(\mathrm{d}u)$. Note that $1 - (1-u)^n \le nu$ for $n \in \mathbb{N}$ and $u \in [0,1]$, and, therefore, $H(n) \le n \int_{[0,1]} \Lambda(\mathrm{d}u)/u = nH(1) < \infty$ for all $n \in \mathbb{N}$. By Lemma A.1, $g_n \,\mathrm{E}(n - I_n) = nH(1) - H(n)$. If we can show that $g_n/n \to 0$ and that $H(n)/n \to 0$ as $n \to \infty$, then

$$\mathrm{E}(n - I_n) = \frac{n}{g_n} \left(H(1) - \frac{H(n)}{n} \right) \sim \frac{n}{g_n} H(1) \to \infty,$$

and we are done. Since

$$\begin{aligned} g_n &= \int_{[0,1]} \frac{1 - (1-u)^n - nu(1-u)^{n-1}}{u^2} \Lambda(\mathrm{d}u) \\ &\le \int_{[0,1]} \frac{1 - (1-u)^n}{u^2} \Lambda(\mathrm{d}u) \\ &= H(n), \end{aligned}$$

it remains to verify that $H(n)/n \to 0$ as $n \to \infty$. By assumption, the measure $\mu(\mathrm{d}u) := \Lambda(\mathrm{d}u)/u$ is finite and has no mass at 0. We have

$$\frac{H(n)}{n} = \int_{[0,1]} \frac{1 - (1-u)^n}{nu} \frac{\Lambda(\mathrm{d}u)}{u} = \int_{[0,1]} f_n(u) \mu(\mathrm{d}u),$$

where $f_n(u) := (1 - (1-u)^n)/(nu)$ for $n \in \mathbb{N}$ and $u \in [0,1]$. Obviously, $0 \le f_n \le 1$ for all $n \in \mathbb{N}$ and f_n converges pointwise to 0 on $(0,1]$ as $n \to \infty$. Thus, $H(n)/n \to 0$ as $n \to \infty$ by dominated convergence.

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

In the following, Lemma A.1 is extended to Ξ-coalescents.

Lemma A.2. *Let $\Xi = a\delta_0 + \Xi_0$ be a finite measure on the infinite simplex Δ, and let $(R_t)_{t\geq 0}$ be a Ξ-coalescent. For $n \in \mathbb{N}$, let I_n be the number of equivalence classes (blocks) of the restricted coalescent process $(\varrho_n R_t)_{t\geq 0}$ after its first jump ($I_1 := 0$). Then, for all $n \in \mathbb{N}$,*

$$g_n \,\mathrm{E}(n - I_n) = a\binom{n}{2} + \int_\Delta \left(n|x| - \sum_{i=1}^{\infty} (1 - (1 - x_i)^n) \right) \frac{\Xi_0(\mathrm{d}x)}{(x, x)}. \tag{A.1}$$

Proof. Fix $n \in \mathbb{N}$. The first summand on the right-hand side of (A.1) is obvious, because, with probability $a = \Xi(\{0\})$, the coalescent behaves as the Kingman coalescent, in which case we have $I_n = n - 1$ and $g_n = \binom{n}{2}$. Thus, without loss of generality, we can and do assume that $a = 0$. In the following we exploit Schweinsberg's [26] Poisson process construction of exchangeable coalescents. Note that this construction is essentially equivalent to Kingman's [18] paintbox construction and closely related to the Bernoulli sieve [13]. For given $x \in \Delta$, partition $[0, 1)$ into intervals $J_0, J_1, J_2, \ldots$ of lengths $x_0 := 1 - |x|, x_1, x_2, \ldots$, i.e. $J_0 := [0, x_0)$, $J_1 := [x_0, x_0 + x_1)$, $J_2 := [x_0 + x_1, x_0 + x_1 + x_2)$, and so on. Let $U_1, \ldots, U_n$ be independent random variables uniformly distributed on $[0, 1)$. For $i \in \mathbb{N}_0$, let

$$X_i := X_i(n) := \sum_{j=1}^{n} \mathbf{1}_{J_i}(U_j)$$

denote the number of $U_1, \ldots, U_n$ which fall into the interval J_i. Note that X_i is binomially distributed with parameters n and x_i, and that $\sum_{i=0}^{\infty} X_i = n$. Therefore,

$$\begin{aligned}
&\mathrm{P}\Big(\bigcap_{i\in\mathbb{N}} \{X_i \leq 1\} \Big) \\
&\quad = \mathrm{P}(X_0 = n) + \sum_{l=1}^{n} \sum_{\substack{i_1,\ldots,i_l\in\mathbb{N} \\ i_1<\cdots<i_l}} \mathrm{P}(X_0 = n - l,\ X_{i_1} = 1, \ldots, X_{i_l} = 1) \\
&\quad = x_0^n + \sum_{l=1}^{n} \binom{n}{l} x_0^{n-l} \sum_{\substack{i_1,\ldots,i_l\in\mathbb{N} \\ \text{all distinct}}} x_{i_1} \cdots x_{i_l}.
\end{aligned}$$

We have

$$\begin{aligned}
g_n \,\mathrm{E}(I_n) &= \sum_{k=1}^{n-1} k g_{nk} \\
&= \sum_{k=1}^{n-1} k \int_\Delta \mathrm{P}\Big(X_0 + \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i\geq 1\}} = k \Big) \frac{\Xi_0(\mathrm{d}x)}{(x, x)} \\
&= \int_\Delta \sum_{k=1}^{n-1} k \,\mathrm{P}\Big(X_0 + \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i\geq 1\}} = k \Big) \frac{\Xi_0(\mathrm{d}x)}{(x, x)} \\
&= \int_\Delta \Big(\mathrm{E}(X_0) + \sum_{i=1}^{\infty} \mathrm{P}(X_i \geq 1) - n \,\mathrm{P}\Big(\bigcap_{i=1}^{\infty} \{X_i \leq 1\} \Big) \Big) \frac{\Xi_0(\mathrm{d}x)}{(x, x)}.
\end{aligned} \tag{A.2}$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

Now subtract this expression from (see [26, p. 36, Equation (70)])

$$ng_n = n\int_\Delta \Bigg(1 - x_0^n - \sum_{l=1}^{n} \binom{n}{l} x_0^{n-l} \sum_{\substack{i_1,\dots,i_l\in\mathbb{N}\\ \text{all distinct}}} x_{i_1}\cdots x_{i_l}\Bigg)\frac{\Xi_0(\mathrm{d}x)}{(x,x)}$$
$$= \int_\Delta \Bigg(n - n\,\mathrm{P}\Bigg(\bigcap_{i=1}^{\infty}\{X_i \le 1\}\Bigg)\Bigg)\frac{\Xi_0(\mathrm{d}x)}{(x,x)},$$

and note that $\mathrm{E}(X_0) = n(1-|x|)$ and $\mathrm{P}(X_i \ge 1) = 1-(1-x_i)^n$.

Remark. The Poisson process construction used in the previous proof is particularly helpful in deriving (1.3) for the rates g_{nk}, $n,k\in\mathbb{N}$ with $k<n$, defined in (1.1). Note that, with the notation used in the previous proof, for $n,k\in\mathbb{N}$ with $k<n$,

$$g_{nk} = a\binom{n}{2}\mathbf{1}_{\{k=n-1\}} + \int_\Delta \mathrm{P}\Bigg(X_0 + \sum_{i=1}^{\infty}\mathbf{1}_{\{X_i\ge 1\}} = k\Bigg)\frac{\Xi_0(\mathrm{d}x)}{(x,x)}$$

with

$$\begin{aligned}
&\mathrm{P}\Bigg(X_0 + \sum_{i=1}^{\infty}\mathbf{1}_{\{X_i\ge1\}} = k\Bigg)\\
&\quad= \sum_{j=1}^{k}\mathrm{P}\Bigg(X_0 = k-j,\ \sum_{i=1}^{\infty}\mathbf{1}_{\{X_i\ge1\}} = j\Bigg)\\
&\quad= \sum_{j=1}^{k}\sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1<\cdots<i_j}} \mathrm{P}(X_0 = k-j,\ X_{i_1}\ge1,\dots,X_{i_j}\ge1,\ X_i = 0 \text{ for all } i\in\mathbb{N}\setminus\{i_1,\dots,i_j\})\\
&\quad= \sum_{j=1}^{k}\sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1<\cdots<i_j}}\sum_{\substack{n_1,\dots,n_j\in\mathbb{N}\\ n_1+\cdots+n_j=n-(k-j)}} \frac{n!}{(k-j)!\,n_1!\cdots n_j!}(1-|x|)^{k-j}x_{i_1}^{n_1}\cdots x_{i_j}^{n_j},
\end{aligned}$$

which proves (1.3).

For $n\in\mathbb{N}\setminus\{1\}$, we now study the number, V_n, of internal branches of the restricted coalescent process which start after the time T_n of the first jump of the restricted coalescent process $(\varrho_n R_t)_{t\ge0}$. Note that $V_n = I_n - S_n$, where S_n denotes the number of singleton blocks of the restricted coalescent process $(\varrho_n R_t)_{t\ge0}$ after its first jump.

Lemma A.3. *For all $n\in\mathbb{N}\setminus\{1\}$,*

$$g_n\,\mathrm{E}(V_n) = a\binom{n}{2} + \int_\Delta \sum_{i=1}^{\infty}\Big(1-(1-x_i)^n - nx_i(1-x_i)^{n-1}\Big)\frac{\Xi_0(\mathrm{d}x)}{(x,x)}. \tag{A.3}$$

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

Proof. Fix $n \in \mathbb{N} \setminus \{1\}$. Again, without loss of generality, we can and do assume that $a = 0$. Using the notation of the previous proof, it follows that

$$\begin{aligned} g_n \,\mathrm{E}(S_n) &= \sum_{s=0}^{n-1} s \int_\Delta \mathrm{P}\Bigg(X_0 + \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i=1\}} = s \Bigg) \frac{\Xi_0(\mathrm{d}x)}{(x,x)} \\ &= \int_\Delta \sum_{s=0}^{n-1} s \,\mathrm{P}\Bigg(X_0 + \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i=1\}} = s \Bigg) \frac{\Xi_0(\mathrm{d}x)}{(x,x)} \\ &= \int_\Delta \Bigg(\mathrm{E}(X_0) + \sum_{i=1}^{\infty} \mathrm{P}(X_i = 1) - n \,\mathrm{P}\Bigg(\bigcap_{i=1}^{\infty} \{X_i \le 1\} \Bigg) \Bigg) \frac{\Xi_0(\mathrm{d}x)}{(x,x)}. \end{aligned}$$

If we subtract this quantity from the expression already derived for $g_n \,\mathrm{E}(I_n)$, (A.2), we arrive at

$$g_n \,\mathrm{E}(V_n) = \int_\Delta \sum_{i=1}^{\infty} \mathrm{P}(X_i \ge 2) \frac{\Xi_0(\mathrm{d}x)}{(x,x)},$$

and the lemma follows from $\mathrm{P}(X_i \ge 2) = 1 - (1-x_i)^n - n x_i (1-x_i)^{n-1}$.

Remark. Fix $n \in \mathbb{N} \setminus \{1\}$. For Λ-coalescents, (A.3) reduces to

$$g_n \,\mathrm{E}(V_n) = \int_{[0,1]} (1 - (1-x)^n - nx(1-x)^{n-1}) \frac{\Lambda(\mathrm{d}x)}{x^2} = g_n.$$

Thus, $\mathrm{E}(V_n) = 1$, which is clear, as $V_n \equiv 1$ for coalescents with only multiple (no simultaneous multiple) collisions.

Corollary A.2. *If (1.6) holds then* $\lim_{n\to\infty} \mathrm{E}(n - I_n)/\,\mathrm{E}(V_n) = \infty$.

Proof. Define the auxiliary function $H : \mathbb{N} \to \mathbb{R}$ via

$$H(n) := \int_{\Delta\setminus\{0\}} \sum_{i=1}^{\infty} (1 - (1-x_i)^n) \frac{\Xi(\mathrm{d}x)}{(x,x)}, \qquad n \in \mathbb{N}.$$

Note that $1 - (1-x_i)^n \le n x_i$ for $n \in \mathbb{N}$ and $x_i \in [0,1]$, and, therefore,

$$0 < H(n) \le n \int_{\Delta\setminus\{0\}} |x| \frac{\Xi(\mathrm{d}x)}{(x,x)} = nH(1) < \infty.$$

We rewrite (A.1) in terms of the auxiliary function H as $g_n \,\mathrm{E}(n - I_n) = nH(1) - H(n)$. Moreover, from (A.3), it follows that $g_n \,\mathrm{E}(V_n) \le H(n)$. Thus,

$$\frac{\mathrm{E}(n - I_n)}{\mathrm{E}(V_n)} \ge \frac{nH(1) - H(n)}{H(n)} = \frac{nH(1)}{H(n)} - 1.$$

It remains to verify that $\lim_{n\to\infty} H(n)/n = 0$. By assumption, the measure $\mu(\mathrm{d}x) := (|x|/(x,x))\,\Xi(\mathrm{d}x)$ is finite and has no mass at 0. We have

$$H(n) = \int_{\Delta\setminus\{0\}} f_n(x) \mu(\mathrm{d}x),$$

where $f_n(x) := \sum_{i=1}^{\infty} (1 - (1-x_i)^n)/(n|x|)$ for $n \in \mathbb{N}$ and $x \in \Delta \setminus \{0\}$. From $1 - (1-x_i)^n \le n x_i$ for $x_i \in [0,1]$, it follows that $0 \le f_n \le 1$ for all $n \in \mathbb{N}$. It is shown below that f_n converges

https://doi.org/10.1239/aap/1261669587 Published online by Cambridge University Press

pointwise to 0 on $\Delta \setminus \{0\}$ as $n \to \infty$. Therefore, $H(n)/n \to 0$ as $n \to \infty$ by dominated convergence and the corollary is established. In order to verify the pointwise convergence of f_n to 0, fix $x \in \Delta \setminus \{0\}$ and let $\delta_{\mathbb{N}}$ denote the counting measure on $\mathbb{N}$. We have

$$|x| f_n(x) = \sum_{i=1}^{\infty} \frac{1-(1-x_i)^n}{n} = \int g_n \, \mathrm{d}\delta_{\mathbb{N}}$$

with $g_n \colon \mathbb{N} \to \mathbb{R}$ defined via $g_n(i) := (1-(1-x_i)^n)/n$. Obviously, $g_n \to 0$ pointwise as $n \to \infty$, as $0 \le g_n \le 1/n$ for all $n \in \mathbb{N}$. Moreover, $g_n(i) \le x_i =: g(i)$ for all $n \in \mathbb{N}$. The function g is integrable with respect to the counting measure $\varepsilon_{\mathbb{N}}$ ($\int g \, \mathrm{d}\delta_{\mathbb{N}} = \sum_{i=1}^{\infty} x_i \le 1$). Thus, $f_n(x) \to 0$ as $n \to \infty$ by dominated convergence.

Acknowledgements

The authors thank Alex Iksanov for fruitful comments concerning the remark before Corollary 3.1 and the final remark of Section 3. We furthermore thank Yuri Yakubovich for pointing out a summation error in (1.3).

References

[1] Basdevant, A.-L. and Goldschmidt, C. (2008). Asymptotics of the allele frequency spectrum associated with the Bolthausen–Sznitman coalescent. *Electron. J. Prob.* **13,** 486–512.
[2] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Prob.* **35,** 1835–1887.
[3] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2008). Small-time behavior of beta coalescents. *Ann. Inst. H. Poincaré Prob. Statist.* **44,** 214–238.
[4] Bertoin, J. and Le Gall, J.-F. (2003). Stochastic flows associated to coalescent processes. *Prob. Theory Relat. Fields* **126,** 261–288.
[5] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197,** 247–276.
[6] Caliebe, A., Neininger, R., Krawczak, M. and Rösler, U. (2007). On the length distribution of external branches in coalescence trees: genetic diversity within species. *Theoret. Pop. Biol.* **72,** 245–252.
[7] Carmona, P., Petit, F. and Yor, M. (1997). On the distribution and asymptotic results for exponential integrals of Lévy processes. In *Exponential Functionals and Principal Values Related to Brownian Motion*, ed. M. Yor, Biblioteca de la Revista Matematica Iberoamericana, Madrid, pp. 73–121.
[8] 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.
[9] 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.
[10] 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. *Random Structures Algorithms* **34,** 319–336.
[11] Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3,** 87–112.
[12] Freund, F. and Möhle, M. (2009). On the time back to the most recent common ancestor and the external branch length of the Bolthausen–Sznitman coalescent. *Markov Process. Relat. Fields* **15,** 387–416.
[13] Gnedin, A. V. (2004). The Bernoulli sieve. *Bernoulli* **10,** 79–96.
[14] Gnedin, A. and Yakubovich, Y. (2007). On the number of collisions in Λ-coalescents. *Electron. J. Prob.* **12,** 1547–1567.
[15] Iksanov, A. and Möhle, M. (2007). A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Prob.* **12,** 28–35.
[16] Iksanov, A. and Möhle, M. (2008). On the number of jumps of random walks with a barrier. *Adv. Appl. Prob.* **40,** 206–228.
[17] Iksanov, A., Marynych, A. and Möhle, M. (2009). On the number of collisions in beta(2,b)-coalescents. *Bernoulli* **15,** 829–845.
[18] Kingman, J. F. C. (1982). On the genealogy of large populations. In *Essays in Statistical Science* (J. Appl. Prob. Spec. Vol. **19A**), eds J. Gani and E. J. Hannan, Applied Probability Trust, Sheffield, pp. 27–43.

[19] Möhle, M. (2005). Coalescent theory – simultaneous multiple collisions and sampling distributions. *Oberwolfach Reports* **40,** 2279–2282.
[20] Möhle, M. (2006). On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12,** 35–53.
[21] Möhle, M. (2006). On the number of segregating sites for populations with large family sizes. *Adv. Appl. Prob.* **38,** 750–767.
[22] Möhle, M. and Sagitov, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Prob.* **29,** 1547–1562.
[23] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27,** 1870–1902.
[24] Sagitov, S. (2003). Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Prob.* **40,** 839–854.
[25] Schweinsberg, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Prob.* **5,** 1–11.
[26] Schweinsberg, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Prob.* **5,** 50 pp.
[27] Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Prob.* **11,** 750–783.