ON THE EXTERNAL BRANCHES OF COALESCENT PROCESSES WITH MULTIPLE COLLISIONS WITH AN EMPHASIS ON THE BOLTHAUSEN–SZNITMAN COALESCENT

J.-S. Dhersin[1] and M. Möhle[2] August 30, 2012

Abstract

A recursion for the joint moments of the external branch lengths for coalescents with multiple collisions (Λ-coalescents) is provided. This recursion is used to derive asymptotic expansions as the sample size n tends to infinity for the moments of the total external branch length of the Bolthausen–Sznitman coalescent. The proof is based on an elementary difference method. An alternative differential equation method is developed which can be used to obtain exact solutions for the joint moments of the external branch lengths for the Bolthausen–Sznitman coalescent. The results for example show that the lengths of two randomly chosen external branches are positively correlated for the Bolthausen–Sznitman coalescent, whereas they are negatively correlated for the Kingman coalescent provided that $n \geq 4$.

Keywords: Asymptotic expansions; Bolthausen–Sznitman coalescent; external branches; joint moments; Kingman coalescent; multiple collisions

2010 Mathematics Subject Classification: Primary 60J25; 34E05; 60C05 Secondary 60J85; 92D15; 92D25

1 Introduction and main results

Let $\Pi = (\Pi_t)_{t\geq 0}$ be a coalescent process with multiple collisions (Λ-coalescent). For fundamental information on Λ-coalescents we refer the reader to [23] and [24]. For $n \in \mathbb{N} := \{1, 2, \ldots\}$ we denote with $\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0}$ the coalescent process restricted to $[n] := \{1, \ldots, n\}$. Note that $\Pi^{(n)}$ is Markovian with state space $\mathcal{E}_n$, the set of all equivalence relations (partitions) on $[n]$. For $\xi \in \mathcal{E}_n$ we write $|\xi|$ for the number of equivalence classes (blocks) of ξ. For $m \in \{1, \ldots, n-1\}$ let g_{nm} be the rate at which the block counting process $N^{(n)} := (N_t^{(n)})_{t\geq 0} := (|\Pi_t^{(n)}|)_{t\geq 0}$ jumps at its first jump time from n to m. It is well known (see, for example, [20, Eq. (13)]) that

$$g_{nm} = \binom{n}{m-1} \int_{[0,1]} x^{n-m-1}(1-x)^{m-1}\, \Lambda(dx) \tag{1}$$

for all $n, m \in \mathbb{N}$ with $m < n$. We furthermore introduce the total rates

$$g_n := \sum_{m=1}^{n-1} g_{nm} = \int_{[0,1]} \frac{1-(1-x)^n - nx(1-x)^{n-1}}{x^2}\, \Lambda(dx), \qquad n \in \mathbb{N}. \tag{2}$$

[1]Department de Mathematique, Insitut Galilée, Université Paris 13, Avenue Jean Baptiste Clément, 93430 Villetaneuse, France, E-mail: dhersin@math.univ-paris13.fr

[2]Mathematisches Institut, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany, E-mail address: martin.moehle@uni-tuebingen.de

We are interested in the external branches of the restricted coalescent process $\Pi^{(n)}$. More precisely, for $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ let

$$\tau_{n,i} \;:=\; \inf\{t > 0 : \{i\} \text{ is a singleton block of } \Pi_t^{(n)}\} \tag{3}$$

denote the length of the ith external branch of the restricted coalescent $\Pi^{(n)}$. Our first main result (Theorem 1.1) provides a general recursion for the joint moments

$$\mu_n(k_1, \ldots, k_j) \;:=\; \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j}), \qquad j \in \{1, \ldots, n\}, k_1, \ldots, k_j \in \mathbb{N}_0 := \{0, 1, \ldots\}, \tag{4}$$

of the external branch lengths. The proof of Theorem 1.1 is provided in Section 2.

Theorem 1.1 (Recursion for the joint moments of the external branch lengths) *For all $n \in \{2, 3, \ldots\}$, $j \in \{1, \ldots, n\}$ and $k_1, \ldots, k_j \in \mathbb{N}$ the joint moments $\mu_n(k_1, \ldots, k_j) := \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})$ of the lengths $\tau_{n,1}, \ldots, \tau_{n,n}$ of the external branches of a Λ-coalescent $\Pi^{(n)}$ satisfy the recursion*

$$\begin{aligned}
&\mu_n(k_1, \ldots, k_j) \\
&\quad = \quad \frac{1}{g_n} \sum_{i=1}^{j} k_i \, \mu_n(k_1, \ldots, k_{i-1}, k_i - 1, k_{i+1}, \ldots, k_j) + \sum_{m=2}^{n-1} p_{nm} \frac{(m-1)_j}{(n)_j} \, \mu_m(k_1, \ldots, k_j),
\end{aligned} \tag{5}$$

where $p_{nm} := g_{nm}/g_n$ and g_{nm} and g_n are defined via (1) and (2).

Remarks.

1. The recursion (5) works as follows. Let us call $k := k_1 + \cdots + k_j$ the order (or degree) of the moment $\mu_n(k_1, \ldots, k_j)$. Provided that all the moments of order $k-1$ are already computed, (5) is a recursion on n for the joint moments of order k, which can be solved iteratively. So one starts with $k = 1$ (and hence $j = 1$), in which case the recursion (5) reduces to $\mu_n(1) = 1/g_n + \sum_{m=2}^{n-1} p_{nm}((m-1)/n)\mu_m(1)$, $n \in \{2, 3, \ldots\}$. Since $\mu_2(1) = \mathbb{E}(\tau_{2,1}) = 1/g_2 = 1/\Lambda([0,1])$, this recursion determines the moments of order 1 completely. Now choose $k = 2$ in (5) which leads to a recursion for the second order moments. Iteratively, one can move to larger values of k.

2. For $j = 2$ and $k_1 = k_2 = 1$ the recursion (5) reduces to

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \frac{2}{g_n}\mathbb{E}(\tau_{n,1}) + \sum_{m=2}^{n-1} p_{nm} \frac{(m-1)_2}{(n)_2} \mathbb{E}(\tau_{m,1}\tau_{m,2}), \quad n \in \{2, 3, \ldots\}, \tag{6}$$

   with initial value $\mathbb{E}(\tau_{2,1}\tau_{2,2}) = 2/g_2^2$. Provided that $\mathbb{E}(\tau_{n,1})$ is known, (6) is a recursion on n for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$.

Note that Theorem 1.1 holds for arbitrary Λ-coalescents. For particular Λ-coalescents the recursion (5) can be used to derive exact solutions and asymptotic expansions for the joint moments of the lengths of the external branches. In the following we briefly discuss the star-shaped coalescent and the Kingman coalescent. Afterwards we intensively study the external branch lengths of the Bolthausen–Sznitman coalescent. For related results on external branches for beta coalescents we refer the reader to [7] and [8].

Example. (Star-shaped coalescent) For the star-shaped coalescent, where Λ is the Dirac measure at 1, the time T_n of the first jump of $\Pi^{(n)}$ is exponentially distributed with parameter $g_n = 1$, $n \in \{2,3,\ldots\}$. Furthermore, $p_{n1} = 1$ and $p_{nm} = 0$ for $n, m \in \mathbb{N}$ with $1 < m < n$. Thus, (5) reduces to $\mu_n(k_1,\ldots,k_j) = \sum_{i=1}^{j} k_i\, \mu_n(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)$ with solution $\mu_n(k_1,\ldots,k_j) = (k_1+\cdots+k_j)!$, which is obviously correct, since $\tau_{n,i} = T_n$ for all $i \in \{1,\ldots,n\}$ and, therefore, $\mu_n(k_1,\ldots,k_j) = \mathbb{E}(T_n^{k_1+\cdots+k_j}) = (k_1+\cdots+k_j)!$, $n \in \{2,3,\ldots\}$, $j \in \{1,\ldots,n\}$, $k_1,\ldots,k_j \in \mathbb{N}$.

Example. (Kingman coalescent) For the Kingman coalescent [19], where Λ is the Dirac measure at 0, the time T_n of the first jump of $\Pi^{(n)}$ is exponentially distributed with parameter $g_n = n(n-1)/2$, $n \in \{2,3,\ldots\}$. Furthermore, $p_{n,n-1} = 1$, $n \in \{2,3,\ldots\}$ and $p_{nm} = 0$ for all $m, n \in \mathbb{N}$ with $m < n-1$. Caliebe et al. [5, Theorem 1] verified that $n\tau_{n,1} \to Z$ in distribution as $n \to \infty$, where Z has density $x \mapsto 8/(2+x)^3$, $x \geq 0$. Janson and Kersting [18, Theorem 1] showed that the total external branch length $L_n^{external} := \sum_{i=1}^{n} \tau_{n,i}$ satisfies $(1/2)\sqrt{n/(\log n)}(L_n^{external} - 2) \to N(0,1)$ in distribution as $n \to \infty$. We are instead interested here in the moments of $\tau_{n,1}$. The recursion (5) for $j = 1$ reduces to

$$\mu_n(k) \;=\; \frac{2k}{n(n-1)}\,\mu_n(k-1) + \frac{n-2}{n}\,\mu_{n-1}(k), \qquad n \in \{2,3,\ldots\}, k \in \mathbb{N}.$$

Rewriting this recursion in terms of $a_n(k) := n(n-1)\mu_n(k)$ yields $a_n(k) = 2k\,\mu_n(k-1) + a_{n-1}(k)$, $n \in \{2,3,\ldots\}$, $k \in \mathbb{N}$, with solution $a_n(k) = 2k\sum_{m=2}^{n} \mu_m(k-1)$. Thus,

$$\mu_n(k) \;=\; \frac{2k}{n(n-1)} \sum_{m=2}^{n} \mu_m(k-1), \qquad n \in \{2,3,\ldots\}, k \in \mathbb{N}.$$

The first two moments are therefore $\mathbb{E}(\tau_{n,1}) = \mu_n(1) = 2/(n(n-1))\sum_{m=2}^{n} 1 = 2/n$ and

$$\mathbb{E}(\tau_{n,1}^2) \;=\; \mu_n(2) \;=\; \frac{4}{n(n-1)} \sum_{m=2}^{n} \frac{2}{m} \;=\; \frac{8(h_n-1)}{n(n-1)} \;=\; 8\frac{\log n}{n^2} + \frac{8(\gamma-1)}{n^2} + O\!\left(\frac{\log n}{n^3}\right),$$

where $\gamma \approx 0.577216$ denotes the Euler constant and $h_n := \sum_{i=1}^{n} 1/i$ the n-th harmonic number, $n \in \mathbb{N}$. Note that these results are in agreement with those of Caliebe et al. [5, Eq. (2)] and Janson and Kersting [18, p. 2205]. For the third moment we obtain

$$\mu_n(3) \;=\; \frac{6}{n(n-1)} \sum_{m=2}^{n} \frac{8(h_m-1)}{m(m-1)} \;=\; \frac{48}{n(n-1)} \sum_{m=2}^{n} \frac{h_m-1}{m(m-1)}.$$

The last sum simplifies considerably to

$$\begin{aligned}
\sum_{m=2}^{n} \frac{h_m-1}{m(m-1)} &= \sum_{m=2}^{n} \left(\frac{h_m}{m-1} - \frac{h_m}{m} - \frac{1}{m(m-1)} \right) \\
&= \sum_{m=1}^{n-1} \frac{h_{m+1}}{m} - \sum_{m=2}^{n} \frac{h_m}{m} - \left(1 - \frac{1}{n}\right) \\
&= h_2 + \sum_{m=2}^{n-1} \frac{h_m + \frac{1}{m+1}}{m} - \sum_{m=2}^{n-1} \frac{h_m}{m} - \frac{h_n}{n} - 1 + \frac{1}{n}
\end{aligned}$$

$$= \frac{3}{2} + \sum_{m=2}^{n-1} \frac{1}{m(m+1)} - \frac{h_n}{n} - 1 + \frac{1}{n}$$
$$= \frac{3}{2} + \frac{1}{2} - \frac{1}{n} - \frac{h_n}{n} - 1 + \frac{1}{n} = 1 - \frac{h_n}{n},$$

Thus, the third moment of $\tau_{n,1}$ is

$$\mathbb{E}(\tau_{n,1}^3) = \mu_n(3) = \frac{48}{n(n-1)}\left(1 - \frac{h_n}{n}\right) = \frac{48}{n^2} - 48\frac{\log n}{n^3} + O\left(\frac{1}{n^3}\right).$$

For the fourth moment we obtain

$$\mathbb{E}(\tau_{n,1}^4) = \mu_n(4) = \frac{8}{n(n-1)} \sum_{m=2}^{n} \mu_m(3) = \frac{384}{n(n-1)} \sum_{m=2}^{n} \frac{1 - h_m/m}{m(m-1)},$$

a formula which does not seem to simplify much further. One may also introduce the generating functions $g_k(t) := \sum_{n=2}^{\infty} \mu_n(k) t^n$, $k \in \mathbb{N}$, $|t| < 1$. For all $k \geq 2$ we have

$$t^2 g_k''(t) = \sum_{n=2}^{\infty} n(n-1)\mu_n(k) t^n = \sum_{n=2}^{\infty} 2k \sum_{m=2}^{n} \mu_m(k-1) t^n$$
$$= 2k \sum_{m=2}^{\infty} \mu_m(k-1) t^m \sum_{n=m}^{\infty} t^{n-m} = \frac{2k}{1-t} g_{k-1}(t),$$

so these generating functions satisfy the recursion

$$g_k(t) = 2k \int_0^t \int_0^s \frac{g_{k-1}(u)}{u^2(1-u)}\, du\, ds, \qquad k \geq 2, 0 \leq t < 1,$$

with initial function $g_1(t) = \sum_{n=2}^{\infty} (2/n) t^n = -2t - 2\log(1-t)$. Using this recursion, $g_k(t)$ can be computed iteratively, however, the expressions become quite involved with increasing k. For example, $g_2(t) = 8t - 4(1-t)\log^2(1-t) - 8(1-t)\mathrm{Li}_2(t)$, $|t| < 1$, where $\mathrm{Li}_2(t) := -\int_0^t (\log(1-x))/x\, dx = \sum_{k=1}^{\infty} t^k/k^2$ denotes the dilogarithm function. In principle higher order moments and as well joint moments can be calculated analogously, however the expressions become more and more nasty with increasing order. In the following we exemplary derive an exact formula for $\mu_n(1,1) = \mathbb{E}(\tau_{n,1}\tau_{n,2})$. The recursion (5) for $j = 2$ and $k_1 = k_2 = 1$ reduces to (see (6))

$$\mu_n(1,1) = \frac{2}{g_n}\mu_n(1) + \frac{(n-2)_2}{(n)_2}\, \mu_{n-1}(1,1) = \frac{8}{n^2(n-1)} + \frac{(n-2)(n-3)}{n(n-1)}\, \mu_{n-1}(1,1), \qquad n \geq 2.$$

It is readily checked by induction on n that the solution of this recursion is given by $\mu_2(1,1) = 2$ and

$$\mu_n(1,1) = \frac{4(n^2 - 5n + 4h_n)}{n(n-1)^2(n-2)}, \qquad n \in \{3, 4, \ldots\}.$$

In particular, the asymptotic expansion $\mu_n(1,1) = 4/n^2 - 4/n^3 + O((\log n)/n^4)$, $n \to \infty$, holds. Moreover, $\mathrm{Cov}(\tau_{n,1}, \tau_{n,2}) = \mu_n(1,1) - (\mu_n(1))^2 = 4(n^2 - 5n + 4h_n)/(n(n-1)^2(n-2)) - 4/n^2 < 0$ for all $n \geq 4$. Thus, for the Kingman coalescent, the lengths of two randomly chosen external branches are (slightly) negatively correlated for all $n \geq 4$. We have used the derived formulas to compute the following table.

n	$\mu_n(1)=\mathbb{E}(\tau_{n,1})$	$\mu_n(1,1)=\mathbb{E}(\tau_{n,1}\tau_{n,2})$	$\mathrm{Cov}(\tau_{n,1},\tau_{n,2})$
2	1	2	1
3	$\frac{2}{3}\approx 0.666667$	$\frac{4}{9}\approx 0.444444$	0
4	$\frac{1}{2}=0.5$	$\frac{13}{54}\approx 0.240741$	$-\frac{1}{108}\approx -0.009259$
5	$\frac{2}{5}=0.4$	$\frac{137}{900}\approx 0.152222$	$-\frac{7}{900}\approx -0.007778$
10	$\frac{1}{5}=0.2$	$\frac{38881}{1020600}\approx 0.038096$	$-\frac{1943}{1020600}\approx -0.001904$
100	$\frac{1}{50}=0.02$	≈ 0.000396	≈ -0.000004
∞	$\frac{2}{n}$	$\frac{4}{n^2}-\frac{4}{n^3}+O(\frac{\log n}{n^4})$	$-\frac{4}{n^3}+O(\frac{\log n}{n^4})$

Table 1: Covariance of $\tau_{n,1}$ and $\tau_{n,2}$ for the Kingman coalescent

In the following we focus on the Bolthausen–Sznitman coalescent [4], where Λ is the uniform distribution on $[0,1]$. Our second main result (Theorem 1.2) provides asymptotic expansions for all the joint moments of the external branch lengths for the Bolthausen–Sznitman coalescent.

Theorem 1.2 (Expansion for the joint moments of the external branch lengths) *For the Bolthausen–Sznitman coalescent, the joint moments $\mu_n(k_1,\dots,k_j):=\mathbb{E}(\tau_{n,1}^{k_1}\cdots\tau_{n,j}^{k_j})$ of the lengths $\tau_{n,1},\dots,\tau_{n,n}$ of the external branches satisfy the asymptotic expansion*

$$\mu_n(k_1,\dots,k_j) \;=\; \frac{k_1!\cdots k_j!}{\log^{k_1+\cdots+k_j} n}\biggl(1+\frac{\kappa_j(k_1,\dots,k_j)}{\log n}+O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr),\quad j\in\mathbb{N},k_1,\dots,k_j\in\mathbb{N}_0, \tag{7}$$

where the coefficients $\kappa_j(k_1,\dots,k_j)$, $j\in\mathbb{N}$, $k_1,\dots,k_j\in\mathbb{N}_0$, are recursively defined via $\kappa_1(0):=0$, $\kappa_j(k_1,\dots,k_{i-1},0,k_{i+1},\dots,k_j):=\kappa_{j-1}(k_1,\dots,k_{i-1},k_{i+1},\dots,k_j)$, $j\in\mathbb{N}\setminus\{1\}$, $i\in\{1,\dots,j\}$, $k_1,\dots,k_{i-1},k_{i+1},\dots,k_j\in\mathbb{N}_0$, and

$$\kappa_j(k_1,\dots,k_j) \;:=\; \frac{1}{j}\biggl(\sum_{i=1}^{j}k_i+j\Psi(j)-j+1+\sum_{i=1}^{j}\kappa_j(k_1,\dots,k_{i-1},k_i-1,k_{i+1},\dots,k_j)\biggr)$$

for $j,k_1,\dots,k_j\in\mathbb{N}$. Here $\Psi:=\Gamma'/\Gamma$ denotes the derivative of $\log\Gamma$ (also called the digamma function).

Remarks.

1. For $j=1$ the recursion for $\kappa_j(k_1,\dots,k_j)$ reduces to $\kappa_1(k)=k-\gamma+\kappa_1(k-1)$, $k\in\mathbb{N}$, where $\gamma=-\Psi(1)\approx 0.577216$ denotes the Euler constant. By induction on k it follows that $\kappa_1(k)=k(k+1)/2-k\gamma$, $k\in\mathbb{N}_0$, in agreement with [11, Theorem 1.2].

2. For $k_1=\cdots=k_j=1$ the recursion for $\kappa_j(k_1,\dots,k_j)$ reduces to $\kappa_1(1)=1+\Psi(1)+\kappa_1(0)=1-\gamma$ and $\kappa_j(1,\dots,1)=\Psi(j)+1/j+\kappa_{j-1}(1,\dots,1)$ for $j\in\{2,3,\dots\}$. An induction on j it follows that $\kappa_j(1,\dots,1)=(j+1)h_j-j-j\gamma$, $j\in\mathbb{N}$, where $h_j:=\sum_{i=1}^{j}1/i$ denotes the jth harmonic number, $j\in\mathbb{N}$.

3. For $j=2$ and $k_1=k_2=1$ the expansion (7) has the form

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \mu_n(1,1) \;=\; \frac{1}{\log^2 n}+\frac{\kappa_2(1,1)}{\log^3 n}+O\Bigl(\frac{1}{\log^4 n}\Bigr),\qquad n\to\infty, \tag{8}$$

with $\kappa_2(1,1)=5/2-2\gamma\approx 1.345569$. In particular, $\mathrm{Cov}(\tau_{n,1},\tau_{n,2})=\mu_n(1,1)-(\mu_n(1))^2=1/(2\log^3 n)+O(1/\log^4 n)$, $n\to\infty$. Thus, for the Bolthausen–Sznitman coalescent, $\tau_{n,1}$ and

$\tau_{n,2}$ are asymptotically positively correlated. With some more effort (see Corollary 3.3 and the remark thereafter) exact solutions for $\mathbb{E}(\tau_{n,1})$ and $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ are obtained and it follows that $\tau_{n,1}$ and $\tau_{n,2}$ are positively correlated for all $n \geq 2$. Note that this positive correlation differs substantially from the situation for the Kingman coalescent, where $\tau_{n,1}$ and $\tau_{n,2}$ are slightly negatively correlated for all $n \geq 4$.

The following convergence result is a direct consequence of Theorem 1.2.

Corollary 1.3 (Weak limiting behavior of the external branch lengths)
For the Bolthausen–Sznitman coalescent, $(\log n)(\tau_{n,1}, \ldots, \tau_{n,n}, 0, 0, \ldots) \to (\tau_1, \tau_2, \ldots)$ *in distribution as* $n \to \infty$*, where* $\tau_1, \tau_2, \ldots$ *are independent and all exponentially distributed with parameter* 1*.*

The following result is as well a direct consequence of Theorem 1.2 and concerns the asymptotics of the moments of the total external branch length $L_n^{external} := \sum_{i=1}^n \tau_{n,i}$ of the Bolthausen–Sznitman coalescent.

Corollary 1.4 (Asymptotics of the moments of the total external branch length)
Fix $k \in \mathbb{N}$*. For the Bolthausen–Sznitman coalescent, the* k*th moment of* $L_n^{external}$ *has the asymptotic expansion*

$$\mathbb{E}((L_n^{external})^k) \;=\; \frac{n^k}{\log^k n}\biggl(1 + \frac{c_k}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr), \qquad n \to \infty, \tag{9}$$

where $c_k := (k+1)h_k - k - k\gamma$*.*

Remarks.

1. Note that $c_1 = 1 - \gamma$, $c_2 = 5/2 - 2\gamma$, $c_3 = 13/3 - 3\gamma$, and $c_4 = 77/12 - 4\gamma$. The expansion (9) quasi coincides with the expansion of the kth moment of the total branch length L_n (see, for example, [9, Corollary 4.3]), only the coefficient c_k is of an additive term k smaller than the corresponding coefficient $m_k = (k+1)h_k - k\gamma$ (see [9]) in the asymptotic expansion of $\mathbb{E}(L_n^k)$.

2. As in the proof of [9, Corollary 4.4] it follows that $n^{-1}(\log n)L_n^{external} \to 1$ in probability as $n \to \infty$. The same argument as given in [22, p. 277] yields the asymptotic expansion

$$\mathbb{E}((L_n^{external} - \mathbb{E}(L_n^{external}))^k) \;=\; \frac{(-1)^k}{k(k-1)}\frac{n^k}{\log^{k+1} n} + O\Bigl(\frac{n^k}{\log^{k+2} n}\Bigr), \qquad k \geq 2,$$

of the centered moments of $L_n^{external}$, which coincide with those (see [9, Eq. (28)]) of the centered moments of the total branch length L_n. In particular,

$$\mathrm{Var}(L_n^{external}) \;=\; \frac{1}{2}\frac{n^2}{\log^3 n} + O\Bigl(\frac{n^2}{\log^4 n}\Bigr), \qquad n \to \infty. \tag{10}$$

The moments of $L_n^{external}$ unfortunately do not provide much information on the distributional limiting behavior of $L_n^{external}$ as $n \to \infty$. Nevertheless, the asymptotic expansions of the centered moments of L_n and $L_n^{external}$ coincide, which supports (or at least does not contradict) our intuition that the distributional limiting behavior of $L_n^{external}$ coincides with that of L_n. Based on the weak convergence result [9, Theorem 5.2] for L_n we therefore state the same convergence result for $L_n^{external}$ as a conjecture.

Conjecture 1.5 (Weak convergence of the total external branch length)
For the Bolthausen–Sznitman coalescent,

$$\frac{\log^2 n}{n} L_n^{external} - \log n - \log\log n \;\to\; L \tag{11}$$

in distribution as $n \to \infty$, where L is a 1-stable random variable with characteristic function $t \mapsto \exp(it \log|t| - \pi|t|/2)$, $t \in \mathbb{R}$.

Remark. The same scaling and limiting behavior as in (11) is known for the number of cuts to isolate the root of a random recursive tree ([10], [16]). Essentially the same scaling and convergence result has been obtained for random records and cuttings in binary search trees by Holmgren [14, Theorem 1.1] and more generally in split trees (Holmgren [13, Theorem 1.1] and [15, Theorem 1.1]) introduced by Devroye [6]. The logarithmic height of the involved trees seems to be one of the main sources for the occurrence of such scalings and of the 1-stable limiting law.

2 Proofs

Proof. (of Theorem 1.1) Let T_n denote the time of the first jump of the block counting process $N^{(n)}$ and let I_n denote the state of the block counting process $N^{(n)}$ after its first jump. Note that T_n is exponentially distributed with parameter g_n and that $\mathbb{P}(I_n = m) = p_{nm} := g_{nm}/g_n$ for all $m \in \{1, \ldots, n-1\}$. Moreover, T_n and I_n are independent. We verify (5) by induction on the number j of involved external branches.

Step 1. The case $j = 1$:
We have to verify that

$$\mathbb{E}(\tau_{n,1}^k) \;=\; \frac{k}{g_n}\mathbb{E}(\tau_{n,1}^{k-1}) + \sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\,\mathbb{E}(\tau_{m,1}^k), \qquad n \in \{2,3,\ldots\}, k \in \mathbb{N}. \tag{12}$$

Note that (12) was already verified in [11], however we provide a proof which turns out to be generalizable to an arbitrary number j of external branches. We verify (12) by induction on $k \in \mathbb{N}$. Conditional on $I_n = m$, the external branch 1 is involved in the first collision event with probability $q_1 := (n-m+1)/n$ and it is not involved in the first collision event with complementary probability $q_0 := (m-1)/n$. Thus,

$$\mathbb{E}(\tau_{n,1}^k \,|\, I_n = m) \;=\; q_1\,\mathbb{E}(T_n^k) + q_0\,\mathbb{E}((T_n + \tau_{m,1})^k),$$

where $\tau_{m,1}$ is independent of T_n. Binomial expansion yields

$$\mathbb{E}(\tau_{n,1}^k \,|\, I_n = m) \;=\; q_1\mathbb{E}(T_n^k) + q_0\sum_{i=0}^{k}\binom{k}{i}\mathbb{E}(T_n^i)\mathbb{E}(\tau_{m,1}^{k-i}) \;=\; \mathbb{E}(T_n^k) + q_0\sum_{i=0}^{k-1}\binom{k}{i}\mathbb{E}(T_n^i)\mathbb{E}(\tau_{m,1}^{k-i}).$$

Multiplication with $p_{nm} = \mathbb{P}(I_n = m)$ and summation over all $m \in \{1, \ldots, n-1\}$ leads to

$$\begin{aligned}\mathbb{E}(\tau_{n,1}^k) &= \mathbb{E}(T_n^k) + \sum_{i=0}^{k-1}\binom{k}{i}\mathbb{E}(T_n^i)\sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\mathbb{E}(\tau_{m,1}^{k-i}) \\ &= \mathbb{E}(T_n^k) + \sum_{i=1}^{k-1}\binom{k}{i}\mathbb{E}(T_n^i)\sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\mathbb{E}(\tau_{m,1}^{k-i}) + \sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\mathbb{E}(\tau_{m,1}^k). \end{aligned} \tag{13}$$

For $k = 1$ this equation reduces to

$$\mathbb{E}(\tau_{n,1}) \;=\; \frac{1}{g_n} + \sum_{m=2}^{n-1} p_{nm} \frac{m-1}{n} \mathbb{E}(\tau_{m,1}).$$

Thus, (12) holds for $k = 1$. The induction step from $\{1, \ldots, k-1\}$ to k (≥ 2) works as follows. By induction it is allowed to replace the first sum over m in (13) by $\mathbb{E}(\tau_{n,1}^{k-i}) - (k-i) g_n^{-1} \mathbb{E}(\tau_{n,1}^{k-i-1})$. We therefore obtain

$$\begin{aligned}
\mathbb{E}(\tau_{n,1}^k) &= \mathbb{E}(T_n^k) + \sum_{i=1}^{k-1} \binom{k}{i} \mathbb{E}(T_n^i) \Big(\mathbb{E}(\tau_{n,1}^{k-i}) - \frac{k-i}{g_n} \mathbb{E}(\tau_{n,1}^{k-i-1}) \Big) + \sum_{m=2}^{n-1} p_{nm} \frac{m-1}{n} \mathbb{E}(\tau_{m,1}^k) \\
&= \sum_{i=1}^{k} \binom{k}{i} \mathbb{E}(T_n^i) \mathbb{E}(\tau_{n,1}^{k-i}) - \sum_{i=1}^{k-1} \binom{k}{i} \mathbb{E}(T_n^i) \frac{k-i}{g_n} \mathbb{E}(\tau_{n,1}^{k-i-1}) + \sum_{m=2}^{n-1} p_{nm} \frac{m-1}{n} \mathbb{E}(\tau_{m,1}^k) \\
&= k \mathbb{E}(T_n) \mathbb{E}(\tau_{n,1}^{k-1}) + \sum_{i=2}^{k} \binom{k}{i} \mathbb{E}(T_n^i) \mathbb{E}(\tau_{n,1}^{k-i}) - \sum_{i=2}^{k} \binom{k}{i-1} \mathbb{E}(T_n^{i-1}) \frac{k-i+1}{g_n} \mathbb{E}(\tau_{n,1}^{k-i}) \\
&\quad + \sum_{m=2}^{n-1} p_{nm} \frac{m-1}{n} \mathbb{E}(\tau_{m,1}^k) \\
&= \frac{k}{g_n} \mathbb{E}(\tau_{n,1}^{k-1}) + \sum_{m=2}^{n-1} p_{nm} \frac{m-1}{n} \mathbb{E}(\tau_{m,1}^k),
\end{aligned}$$

since, for all $i \in \{2, \ldots, k\}$,

$$\binom{k}{i} \mathbb{E}(T_n^i) - \binom{k}{i-1} \mathbb{E}(T_n^{i-1}) \frac{k-i+1}{g_n} \;=\; \binom{k}{i} \frac{i!}{g_n^i} - \binom{k}{i-1} \frac{(i-1)!}{g_n^{i-1}} \frac{k-i+1}{g_n} \;=\; 0.$$

This finishes the induction and completes the proof of (12). Thus, (5) holds for $j = 1$.

The induction step from to $\{1, \ldots, j-1\}$ to j (≥ 2) for general number j of involved external branches is somewhat technical. We therefore consider next the case $j = 2$.

Step 2. The case $j = 2$.
The proof works again by induction on the order $k := k_1 + k_2$. We thus first have to verify (6), which is the particular case $j = 2$ and $k_1 = k_2 = 1$ of the general recursion (5).
Conditional on $I_n = m$, the two external branches 1 and 2 are both involved in the first collision event with probability $q_{11} := (n-m+1)_2/(n)_2$, the external branch 1 is involved and the external branch 2 is not involved in the first collision event with probability $q_{10} := (n-m+1)(m-1)/(n)_2$, and both external branches are not involved in the first collision event with probability $q_{00} := (m-1)_2/(n)_2$. Note that $q_{10} = q_{01}$, that $q_{11} + q_{10} + q_{01} + q_{00} = 1$ and that $q_{10} + q_{00} = (m-1)/n$. Thus,

$$\begin{aligned}
\mathbb{E}(\tau_{n,1}\tau_{n,2} \,|\, I_n = m) &= q_{11} \mathbb{E}(T_n^2) + q_{10} \mathbb{E}(T_n(T_n + \tau_{m,2})) \\
&\qquad + q_{01} \mathbb{E}((T_n + \tau_{m,1}) T_n) + q_{00} \mathbb{E}((T_n + \tau_{m,1})(T_n + \tau_{m,2})) \\
&= \mathbb{E}(T_n^2) + 2 \mathbb{E}(T_n) \mathbb{E}(\tau_{m,1})(q_{10} + q_{00}) + q_{00} \mathbb{E}(\tau_{m,1}\tau_{m,2}) \\
&= \frac{2}{g_n^2} + \frac{2}{g_n} \mathbb{E}(\tau_{m,1}) \frac{m-1}{n} + \mathbb{E}(\tau_{m,1}\tau_{m,2}) \frac{(m-1)_2}{(n)_2}.
\end{aligned}$$

Multiplication of both sides with $p_{nm} = \mathbb{P}(I_n = m)$ and summation over all $m \in \{1, \ldots, n-1\}$ yields

$$\begin{aligned}\mathbb{E}(\tau_{n,1}\tau_{n,2}) &= \frac{2}{g_n}\left(\frac{1}{g_n} + \sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\,\mathbb{E}(\tau_{m,1})\right) + \sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_2}{(n)_2}\,\mathbb{E}(\tau_{m,1}\tau_{m,2}) \\ &= \frac{2}{g_n}\mathbb{E}(\tau_{n,1}) + \sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_2}{(n)_2}\mathbb{E}(\tau_{m,1}\tau_{m,2}),\end{aligned}$$

which is (6). Concerning the induction step from $\{1, \ldots, k_1 + k_2 - 1\}$ to $k := k_1 + k_2 \in \{3, 4, \ldots\}$ we have to verify that

$$\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2}) = \frac{k_1}{g_n}\mathbb{E}(\tau_{n,1}^{k_1-1}\tau_{n,2}^{k_2}) + \frac{k_2}{g_n}\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2-1}) + \sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_2}{(n)_2}\mathbb{E}(\tau_{m,1}^{k_1}\tau_{m,2}^{k_2}).$$

We have

$$\begin{aligned}\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2} \,|\, I_n = m) &= q_{11}\mathbb{E}(T_n^{k_1+k_2}) + q_{01}\mathbb{E}((T_n + \tau_{m,1})^{k_1}T_n^{k_2}) \\ &\quad + q_{10}\mathbb{E}(T_n^{k_1}(T_n + \tau_{m,2})^{k_2}) + q_{00}\mathbb{E}((T_n + \tau_{m,1})^{k_1}(T_n + \tau_{m,2})^{k_2}).\end{aligned}$$

Binomial expansion together with the fact that T_n is independent of $(\tau_{m,1}, \tau_{m,2})$ leads to

$$\begin{aligned}\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2} \,|\, I_n = m) &= q_{11}\mathbb{E}(T_n^{k_1+k_2}) + q_{01}\sum_{i_1=0}^{k_1}\binom{k_1}{i_1}\mathbb{E}(T_n^{i_1+k_2})\mathbb{E}(\tau_{m,1}^{k_1-i_1}) \\ &\quad + q_{10}\sum_{i_2=0}^{k_2}\binom{k_2}{i_2}\mathbb{E}(T_n^{k_1+i_2})\mathbb{E}(\tau_{m,2}^{k_2-i_2}) \\ &\quad + q_{00}\sum_{i_1=0}^{k_1}\sum_{i_2=0}^{k_2}\binom{k_1}{i_1}\binom{k_2}{i_2}\mathbb{E}(T_n^{i_1+i_2})\mathbb{E}(\tau_{m,1}^{k_1-i_1}\tau_{m,2}^{k_2-i_2}) \\ &= \mathbb{E}(T_n^{k_1+k_2}) + (q_{01}+q_{00})\sum_{i_1=0}^{k_1-1}\binom{k_1}{i_1}\mathbb{E}(T_n^{i_1+k_2})\mathbb{E}(\tau_{m,1}^{k_1-i_1}) \\ &\quad + (q_{10}+q_{00})\sum_{i_2=0}^{k_2-1}\binom{k_2}{i_2}\mathbb{E}(T_n^{k_1+i_2})\mathbb{E}(\tau_{m,2}^{k_2-i_2}) \\ &\quad + q_{00}\sum_{i_1=0}^{k_1-1}\sum_{i_2=0}^{k_2-1}\binom{k_1}{i_1}\binom{k_2}{i_2}\mathbb{E}(T_n^{i_1+i_2})\mathbb{E}(\tau_{m,1}^{k_1-i_1}\tau_{m,2}^{k_2-i_2}).\end{aligned}$$

Note that $q_{10} = q_{01}$ and that $q_{01} + p_{00} = (m-1)/n$. Multiplication with $p_{nm} = \mathbb{P}(I_n = m)$ and summation over all $m \in \{1, \ldots, n-1\}$ leads to

$$\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2}) = \mathbb{E}(T_n^{k_1+k_2}) + \sum_{i_1=0}^{k_1-1}\binom{k_1}{i_1}\mathbb{E}(T_n^{i_1+k_2})\sum_{m=2}^{n-1} p_{nm}\frac{m-1}{n}\mathbb{E}(\tau_{m,1}^{k_1-i_1})$$

$$+\sum_{i_2=0}^{k_2-1}\binom{k_2}{i_2}\mathbb{E}(T_n^{k_1+i_2})\sum_{m=2}^{n-1}p_{nm}\frac{m-1}{n}\mathbb{E}(\tau_{m,2}^{k_2-i_2})$$
$$+\sum_{i_1=0}^{k_1-1}\sum_{i_2=0}^{k_2-1}\binom{k_1}{i_1}\binom{k_2}{i_2}\mathbb{E}(T_n^{i_1+i_2})\sum_{m=2}^{n-1}p_{nm}\frac{(m-1)_2}{(n)_2}\mathbb{E}(\tau_{m,1}^{k_1-i_1}\tau_{m,2}^{k_2-i_2}).$$

By (12), the first two sums over m are equal to $\mathbb{E}(\tau_{n,1}^{k_1-i_1})-(k_1-i_1)g_n^{-1}\mathbb{E}(\tau_{n,1}^{k_1-i_1-1})$ and $\mathbb{E}(\tau_{n,2}^{k_2-i_2})-(k_2-i_2)g_n^{-1}\mathbb{E}(\tau_{n,2}^{k_2-i_2-1})$ respectively. Moreover, by induction it is allowed to replace the third (last) sum over m (except for the case $i_1=i_2=0$) by $\mathbb{E}(\tau_{n,1}^{k_1-i_1}\tau_{n,2}^{k_2-i_2})-(k_1-i_1)g_n^{-1}\mathbb{E}(\tau_{n,1}^{k_1-i_1-1}\tau_{n,2}^{k_2-i_2})-(k_2-i_2)g_n^{-1}\mathbb{E}(\tau_{n,1}^{k_1-i_1}\tau_{n,2}^{k_2-i_2-1})$. We therefore obtain

$$\begin{aligned}
\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2}) \;=\; & \mathbb{E}(T_n^{k_1+k_2})+\sum_{i_1=0}^{k_1-1}\binom{k_1}{i_1}\mathbb{E}(T_n^{i_1+k_2})\Big(\mathbb{E}(\tau_{n,1}^{k_1-i_1})-\frac{k_1-i_1}{g_n}\mathbb{E}(\tau_{n,1}^{k_1-i_1-1})\Big)\\
&+\sum_{i_2=0}^{k_2-1}\binom{k_2}{i_2}\mathbb{E}(T_n^{k_1+i_2})\Big(\mathbb{E}(\tau_{n,2}^{k_2-i_2})-\frac{k_2-i_2}{g_n}\mathbb{E}(\tau_{n,2}^{k_2-i_2-1})\Big)\\
&+\sum_{\substack{i_1<k_1,i_2<k_2\\ i_1+i_2>0}}\binom{k_1}{i_1}\binom{k_2}{i_2}\mathbb{E}(T_n^{i_1+i_2})\times\\
&\qquad\times\Big(\mathbb{E}(\tau_{n,1}^{k_1-i_1}\tau_{n,2}^{k_2-i_2})-\frac{k_1-i_1}{g_n}\mathbb{E}(\tau_{n,1}^{k_1-i_1-1}\tau_{n,2}^{k_2-i_2})-\frac{k_2-i_2}{g_n}\mathbb{E}(\tau_{n,1}^{k_1-i_1}\tau_{n,2}^{k_2-i_2-1})\Big)\\
&+\sum_{m=2}^{n-1}p_{nm}\frac{(m-1)_2}{(n)_2}\mathbb{E}(\tau_{m,1}^{k_1}\tau_{m,2}^{k_2}).
\end{aligned}$$

The last sum is already what we need. For the other sums the main task is to reorder all the summands in terms of the moments of $\tau_{n,1}$ and $\tau_{n,2}$. The coefficient in front of $\mathbb{E}(\tau_{n,1}^{k_1-1}\tau_{n,2}^{k_2})$ is $\binom{k_1}{1}\binom{k_2}{0}\mathbb{E}(T_n^{1+0})=k_1/g_n$ and the coefficient in front of $\mathbb{E}(\tau_{n,1}^{k_1}\tau_{n,2}^{k_2-1})$ is $\binom{k_1}{0}\binom{k_2}{1}\mathbb{E}(T_n^{0+1})=k_2/g_n$ as required. It remains to verify that all the other coefficients are equal to zero. Careful investigation shows that the coefficient in front of $\mathbb{E}(\tau_{n,1}^{k_1-i_1})$ (with $i_1<k_1$) is

$$\begin{aligned}
&\binom{k_1}{i_1}\mathbb{E}(T_n^{i_1+k_2})\\
&\qquad-\binom{k_1}{i_1-1}\mathbb{E}(T_n^{(i_1-1)+k_2})\frac{k_1-(i_1-1)}{g_n}-\binom{k_1}{i_1}\binom{k_2}{k_2-1}\mathbb{E}(T_n^{i_1+(k_2-1)})\frac{k_2-(k_2-1)}{g_n}\\
=\;&\binom{k_1}{i_1}\frac{(i_1+k_2)!}{g_n^{i_1+k_2}}-\binom{k_1}{i_1-1}\frac{(i_1+k_2-1)!}{g_n^{i_1+k_2-1}}\frac{k_1-i_1+1}{g_n}-\binom{k_1}{i_1}k_2\frac{(i_1+k_2-1)!}{g_n^{i_1+k_2-1}}\frac{1}{g_n}\\
=\;&\binom{k_1}{i_1}\frac{(i_1+k_2-1)!}{g_n^{i_1+k_2}}\big((i_1+k_2)-i_1-k_2\big)\;=\;0.
\end{aligned}$$

In the same manner it follows that the coefficient in front of $\mathbb{E}(\tau_{n,2}^{k_2-i_2})$ (with $i_2<k_2$) vanishes. The coefficient in front of $\mathbb{E}(\tau_{n,1}^{k_1-i_1}\tau_{n,2}^{k_2-i_2})$ (with $1\le i_1<k_1$ and $1\le i_2<k_2$) is

$$\binom{k_1}{i_1}\binom{k_2}{i_2}\mathbb{E}(T_n^{i_1+i_2})$$

$$
\begin{aligned}
& -\binom{k_1}{i_1-1}\binom{k_2}{i_2}\mathbb{E}(T_n^{(i_1-1)+i_2})\frac{k_1-(i_1-1)}{g_n} - \binom{k_1}{i_1}\binom{k_2}{i_2-1}\mathbb{E}(T_n^{i_1+(i_2-1)})\frac{k_2-(i_2-1)}{g_n} \\
= \;& \binom{k_1}{i_1}\binom{k_2}{i_2}\frac{(i_1+i_2)!}{g_n^{i_1+i_2}} \\
& -\binom{k_1}{i_1-1}\binom{k_2}{i_2}\frac{(i_1+i_2-1)!}{g_n^{i_1+i_2-1}}\frac{k_1+i_1+1}{g_n} - \binom{k_1}{i_1}\binom{k_2}{i_2-1}\frac{(i_1+i_2-1)!}{g_n^{i_1+i_2-1}}\frac{k_2-i_2+1}{g_n} \\
= \;& \binom{k_1}{i_1}\binom{k_2}{i_2}\frac{(i_1+i_2-1)!}{g_n^{i_1+i_2}}\big((i_1+i_2)-i_1-i_2\big) \;=\; 0.
\end{aligned}
$$

The last coefficient to consider is that in front of $\mathbb{E}(\tau_{n,1}^0\tau_{n,2}^0)$ $(=1)$, which is

$$
\begin{aligned}
& \mathbb{E}(T_n^{k_1+k_2}) - \binom{k_1}{k_1-1}\mathbb{E}(T_n^{(k_1-1)+k_2})\frac{k_1-(k_1-1)}{g_n} - \binom{k_2}{k_2-1}\mathbb{E}(T_n^{k_1+(k_2-1)})\frac{k_2-(k_2-1)}{g_n} \\
= \;& \frac{(k_1+k_2)!}{g_n^{k_1+k_2}} - k_1\frac{(k_1+k_2-1)!}{g_n^{k_1+k_2-1}}\frac{1}{g_n} - k_2\frac{(k_1+k_2-1)!}{g_n^{k_1+k_2-1}}\frac{1}{g_n} \;=\; 0.
\end{aligned}
$$

This finishes the induction and the proof of (5) for $j=2$ is complete. Note that we have used (12), so in order to verify (5) for $j=2$, we have used (5) for $j=1$.

Step 3. The general case $j \geq 2$:
The proof of the induction step for general number j (≥ 2) of involved external branches works essentially the same (induction over the order $k := k_1 + \cdots + k_j$ of the moments), however, the details become more involved and a bit nasty to write down. We leave the details to the interested reader. □

Proof. (of Theorem 1.2) For the Bolthausen–Sznitman coalescent, $g_{nm} = n/((n-m)(n-m+1))$, $m \in \{1,\ldots,n-1\}$ and $g_n = n-1$, $n \in \mathbb{N}$. Thus, the random state I_n of the block counting process after its first jump has distribution

$$
p_{nm} \;=\; \mathbb{P}(I_n = m) \;=\; \frac{n}{(n-1)(n-m)(n-m+1)}, \qquad n \in \mathbb{N}, m \in \{1,\ldots,n-1\}.
$$

We verify (7) by induction on the sum $k := k_1 + \cdots + k_j \in \mathbb{N}_0$. Obviously (7) holds for $k=0$. In order to verify the induction step from $k-1$ to $k \in \mathbb{N}$ fix $k_1,\ldots,k_j \in \mathbb{N}_0$ with $k_1+\cdots+k_j = k$ and for $n \in \mathbb{N}$ define $a_n := \mu_n(k_1,\ldots,k_j)$ for convenience. We apply the difference method used in [17] to the recursion (5), which is of the form

$$
\begin{aligned}
a_n &= q_n + \sum_{m=1}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}\,a_m \\
&= q_n + \sum_{m=1}^{n-1} \frac{n}{(n-1)(n-m)(n-m+1)}\frac{(m-1)_j}{(n)_j}\,a_m \\
&= q_n + \frac{n}{(n-1)(n)_j}\sum_{m=1}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)}\,a_m,
\end{aligned}
$$

where

$$\begin{aligned} q_n &:= \frac{1}{g_n}\sum_{i=1}^{j} k_i \mu_n(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j) \\ &= \frac{1}{g_n}\sum_{i=1}^{j}\frac{k_1!\cdots k_j!}{\log^{k-1} n}\left(1+\frac{\kappa_j(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\right) \\ &= \frac{1}{g_n}\frac{k_1!\cdots k_j!}{\log^{k-1} n}\sum_{i=1}^{j}\left(1+\frac{\kappa_j(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\right). \end{aligned}$$

From $1/g_n = 1/(n-1) = \sum_{i=1}^{\infty}(1/n)^i = 1/n + O(1/n^2)$ it follows that

$$q_n = \frac{jk_1!\cdots k_j!}{n\log^{k-1} n} + \frac{k_1!\cdots k_j!\sum_{i=1}^{j}\kappa_j(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)}{n\log^k n} + O\Big(\frac{1}{n\log^{k+1} n}\Big).$$

The difference $b_n := a_n - k_1!\cdots k_j!/\log^k n$ satisfies the recursion

$$\begin{aligned} b_n &= a_n - \frac{k_1!\cdots k_j!}{\log^k n} \\ &= q_n + \frac{n}{(n-1)(n)_j}\sum_{m=1}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)}\left(b_m + \frac{k_1!\cdots k_j!}{\log^k m}\right) - \frac{k_1!\cdots k_j!}{\log^k n} \\ &= q_n' + \frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)}\,b_m, \end{aligned}$$

with

$$q_n' := q_n + k_1!\cdots k_j!\frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} - \frac{k_1!\cdots k_j!}{\log^k n}.$$

Corollary 4.2 in the appendix yields

$$\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} = \frac{n^j}{\log^k n} - \frac{jn^{j-1}}{\log^{k-1} n} + \left(k + j\Psi(j) - \frac{j(j+1)}{2}\right)\frac{n^{j-1}}{\log^k n} + O\Big(\frac{n^{j-1}}{\log^{k+1} n}\Big)$$

and hence

$$\begin{aligned} &\frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} \\ &= \left(\frac{1}{n^j} + \frac{\binom{j}{2}+1}{n^{j+1}} + O\Big(\frac{1}{n^{j+2}}\Big)\right)\times \\ &\qquad \times\left(\frac{n^j}{\log^k n} - j\frac{n^{j-1}}{\log^{k-1} n} + (k + j\Psi(j) - j(j+1)/2)\frac{n^{j-1}}{\log^k n} + O\Big(\frac{n^{j-1}}{\log^{k+1} n}\Big)\right) \\ &= \frac{1}{\log^k n} - \frac{j}{n\log^{k-1} n} + \left(k + j\Psi(j) - j + 1\right)\frac{1}{n\log^k n} + O\Big(\frac{1}{n\log^{k+1} n}\Big). \end{aligned}$$

We therefore obtain

$$\begin{aligned} q'_n &= k_1!\cdots k_j!\Bigg(\sum_{i=1}^{j}\kappa_j(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)+k+j\Psi(j)-j+1\Bigg)\frac{1}{n\log^k n} \\ &\quad +O\Bigg(\frac{1}{n\log^{k+1} n}\Bigg) \;=\; jK\frac{1}{n\log^k n}+O\Bigg(\frac{1}{n\log^{k+1} n}\Bigg) \end{aligned}$$

with

$$\begin{aligned} K &:= k_1!\cdots k_j!\frac{1}{j}\Bigg(k+j\Psi(j)-j+1+\sum_{i=1}^{j}\kappa_j(k_1,\ldots,k_{i-1},k_i-1,k_{i+1},\ldots,k_j)\Bigg) \\ &= k_1!\cdots k_j!\kappa_j(k_1,\ldots,k_j) \end{aligned}$$

by the recursive definition of the coefficients $\kappa_j(k_1,\ldots,k_j)$. We now proceed with the recursion for b_n in the same manner. The difference $c_n := b_n - K/\log^{k+1} n$ satisfies the recursion

$$\begin{aligned} c_n &= b_n-\frac{K}{\log^{k+1} n} \\ &= q'_n+\frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)}\Bigg(c_m+\frac{K}{\log^{k+1} m}\Bigg)-\frac{K}{\log^{k+1} n} \\ &= q''_n+\frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)}\,c_m \end{aligned}$$

with

$$q''_n := q'_n+K\frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^{k+1} m}-\frac{K}{\log^{k+1} n}.$$

Corollary 4.2 yields

$$\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^{k+1} m} \;=\; \frac{n^j}{\log^{k+1} n}-j\frac{n^{j-1}}{\log^k n}+O\Bigg(\frac{n^{j-1}}{\log^{k+1} n}\Bigg)$$

such that we obtain

$$q''_n \;=\; (jK-jK)\frac{1}{n\log^k n}+O\Bigg(\frac{1}{n\log^{k+1} n}\Bigg) \;=\; O\Bigg(\frac{1}{n\log^{k+1} n}\Bigg).$$

By Lemma 4.3, applied with k replaced by $k+1$, it follows that $c_n = O(1/\log^{k+2} n)$, and the induction step is established. □

Proof. (of Corollary 1.3) Theorem 1.2 clearly implies that, for $j \in \mathbb{N}$ and $k_1,\ldots,k_j \in \mathbb{N}_0$, $\mathbb{E}((\tau_{n,1}\log n)^{k_1}\cdots(\tau_{n,j}\log n)^{k_j}) = (\log n)^{k_1+\cdots+k_j}\mu_n(k_1,\ldots,k_j) \to k_1!\cdots k_j! = \mathbb{E}(\tau_1^{k_1}\cdots\tau_j^{k_j})$ as $n\to\infty$. For all $i\in\{1,\ldots,j\}$ and all $0\le\theta<1$ we have $\sum_{r=0}^{\infty}(\theta^r/r!)\mathbb{E}(\tau_i^r) = \sum_{r=0}^{\infty}\theta^r = 1/(1-\theta) < \infty$. Therefore (see [2], Theorems 30.1 and 30.2 for the one-dimensional case and Problem 30.6 on

p. 398 for the multi-dimensional case) the above convergence of moments implies the convergence $(\log n)(\tau_{n,1},\dots,\tau_{n,j}) \to (\tau_1,\dots,\tau_j)$ in distribution as $n\to\infty$ for each $j\in\mathbb{N}$. The convergence of all these j-dimensional distributions is already equivalent (see Billingsley [3, p. 19]) to the convergence of the full processes $(\log n)(\tau_{n,1},\dots,\tau_{n,n},0,0,\dots) \to (\tau_1,\tau_2,\dots)$ in distribution as $n\to\infty$. □

Proof. (of Corollary 1.4) The total external branch length $L_n^{external}$ satisfies (see, for example, [21, p. 2165]

$$\mathbb{E}((L_n^{external})^k) \;=\; \sum_{j=1}^{k}\binom{n}{j}\sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}}\frac{k!}{k_1!\cdots k_j!}\,\mu_n(k_1,\dots,k_j),\qquad n\in\{2,3,\dots\},k\in\mathbb{N}.$$

By Theorem 1.2 it follows that

$$\mathbb{E}((L_n^{external})^k) \;=\; \sum_{j=1}^{k}\binom{n}{j}\sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}}\frac{k!}{k_1!\cdots k_j!}\frac{k_1!\cdots k_j!}{\log^k n}\left(1+\frac{\kappa_j(k_1,\dots,k_j)}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right).$$

From this representation it follows immediately that asymptotically all the summands with indices $j<k$ can be disregarded, so only the summand with index $j=k$ is asymptotically of any importance. Thus, we obtain

$$\begin{aligned}\mathbb{E}((L_n^{external})^k) &= \binom{n}{k}\frac{k!}{\log^k n}\left(1+\frac{\kappa_k(1,\dots,1)}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right)\\ &= \frac{n^k}{\log^k n}\left(1+\frac{\kappa_k(1,\dots,1)}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right),\end{aligned}$$

where $\kappa_k(1,\dots,1)=(k+1)h_k-k-k\gamma$ by the second remark after Theorem 1.2. □

3 Differential equations approach

The expansions of Theorem 1.2 in Section 1 are essentially based on the difference method used in the proof of Theorem 1.2. An alternative method based on generating functions is now provided. This approach yields for example exact expressions for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ in terms of Stirling numbers. For $j\in\mathbb{N}$ and $k=(k_1,\dots,k_j)\in\mathbb{N}^j$ define the generating function

$$f_k(z) \;:=\; \sum_{n=j}^{\infty}\mathbb{E}(\tau_{n,1}^{k_1}\cdots\tau_{n,j}^{k_j})z^{n-1}.$$

Note that $f_k^{(l)}(0)=0$ for $l\in\{1,\dots,j-2\}$ and that $f_k^{(j-1)}(0)=(j-1)!\mathbb{E}(\tau_{j,1}^{k_1}\cdots\tau_{j,j}^{k_j})$. We shall see soon, that the functions f_k are analytic in the domain $D:=\{s\in\mathbb{C}:|s|<1\}$. The functions f_k even have analytic continuations on $\mathbb{C}\setminus[1,\infty)$, but we will not use these continuations in our approach. We start with the following lemma, which is well known (see [11, Lemma 3.1, Eq. (3.3)]).

Lemma 3.1 *For all $z\in D$, $f_1(z) \;=\; \displaystyle\int_0^z \frac{t}{(1-t)^2(-\log(1-t))}\,dt.$*

Proof. Define the auxiliary function

$$a(z) \;:=\; \sum_{n=1}^{\infty} \frac{1}{n(n+1)} z^n \;=\; 1 - \log(1-z) + \frac{\log(1-z)}{z}, \qquad z \in D.$$

For $z \in D$ we have

$$\begin{aligned}
z f_1'(z) &= \sum_{n=2}^{\infty} (n-1)\mathbb{E}(\tau_{n,1}) z^{n-1} \\
&= \sum_{n=2}^{\infty} (n-1)\Big(\frac{1}{n-1} + \sum_{m=1}^{n-1} \frac{m-1}{(n-1)(n-m)(n-m+1)} \mathbb{E}(\tau_{m,1})\Big) z^{n-1} \\
&= \sum_{n=2}^{\infty} z^{n-1} + \sum_{n=2}^{\infty} \sum_{m=1}^{n-1} \frac{m-1}{(n-m)(n-m+1)} \mathbb{E}(\tau_{m,1}) z^{n-1} \\
&= \frac{z}{1-z} + \sum_{n=2}^{\infty} \sum_{l=1}^{n-1} \frac{n-l-1}{l(l+1)} \mathbb{E}(\tau_{n-l,1}) z^{n-1} \\
&= \frac{z}{1-z} + \sum_{l=1}^{\infty} \frac{z^l}{l(l+1)} \sum_{n=l+1}^{\infty} (n-l-1)\mathbb{E}(\tau_{n-l,1}) z^{n-l-1} \\
&= \frac{z}{1-z} + \sum_{l=1}^{\infty} \frac{z^l}{l(l+1)} \sum_{m=1}^{\infty} (m-1)\mathbb{E}(\tau_{m,1}) z^{m-1} \\
&= \frac{z}{1-z} + a(z) z f_1'(z).
\end{aligned}$$

Thus $z(1-a(z))f_1'(z) = z/(1-z)$ or, equivalently,

$$f_1'(z) \;=\; \frac{1}{(1-z)(1-a(z))} \;=\; \frac{z}{(1-z)^2(-\log(1-z))}, \qquad z \in D,$$

since $1 - a(z) = (1-z)(-\log(1-z))/z$. The result follows by integration, since $f_1(0) = 0$. □

In the following a solution for $f_{(1,1)}$ is presented. As a corollary, an exact formula for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ is provided.

Lemma 3.2 *The function $f := f_{(1,1)}$ satisfies the differential equation*

$$(1-z)(-\log(1-z))f''(z) + f'(z) \;=\; \frac{2z}{(1-z)^2(-\log(1-z))}, \qquad z \in D. \tag{14}$$

with initial conditions $f(0) = 0$ and $f'(0) = 2$. The solution f satisfies

$$f'(z) \;=\; \frac{2}{-\log(1-z)} \int_0^z \frac{t}{(1-t)^3(-\log(1-t))}\, dt, \qquad z \in D \setminus \{0\}. \tag{15}$$

Proof. For $n \in \{2,3,\ldots\}$ we write $a_n := \mathbb{E}(\tau_{n,1}\tau_{n,2})$ for convenience. Note that $f(z) = \sum_{n=2}^{\infty} a_n z^{n-1}$. In particular, $f(0) = 0$ and $f'(0) = a_2 = \mathbb{E}(\tau_{2,1}\tau_{2,2}) = \mathbb{E}(T_2^2) = 2/g_2^2 = 2$. For $z \in D$ we have

$$\begin{aligned}
z^2 f''(z) + z f'(z) &= \sum_{n=2}^{\infty} (n-1)^2 a_n z^{n-1} \\
&= \sum_{n=2}^{\infty} (n-1)^2 \Big(\frac{2}{g_n} \mathbb{E}(\tau_{n,1}) + \sum_{m=1}^{n-1} p_{nm} \frac{(m-1)_2}{(n)_2} a_m \Big) z^{n-1} \\
&= 2 \sum_{n=2}^{\infty} (n-1) \mathbb{E}(\tau_{n,1}) z^{n-1} + \sum_{n=2}^{\infty} \sum_{m=1}^{n-1} \frac{(m-1)(m-2)}{(n-m)(n-m+1)} a_m z^{n-1} \\
&= 2 z f_1'(z) + \sum_{n=2}^{\infty} \sum_{l=1}^{n-1} \frac{(n-l-1)(n-l-2)}{l(l+1)} a_{n-l} z^{n-1} \\
&= 2 z f_1'(z) + \sum_{l=1}^{\infty} \frac{z^l}{l(l+1)} \sum_{n=l+1}^{\infty} (n-l-1)(n-l-2) a_{n-l} z^{n-l-1} \\
&= 2 z f_1'(z) + \sum_{l=1}^{\infty} \frac{z^l}{l(l+1)} \sum_{m=1}^{\infty} (m-1)(m-2) a_m z^{m-1} \\
&= 2 z f_1'(z) + a(z) z^2 f''(z),
\end{aligned}$$

or, equivalently,

$$z^2(1-a(z)) f''(z) + z f'(z) \;=\; 2 z f_1'(z) \;=\; \frac{2z^2}{(1-z)^2(-\log(1-z))}, \qquad z \in D,$$

which yields the differential equation (14), since $1 - a(z) = (1-z)(-\log(1-z))/z$. Division of (14) by $1-z$ shows that

$$\frac{2z}{(1-z)^3(-\log(1-z))} \;=\; -\log(1-z) f''(z) + \frac{1}{1-z} f'(z) \;=\; \frac{d}{dz}(-\log(1-z) f'(z))$$

and integration together with the initial condition $f'(0) = 2$ yields (15). □

Corollary 3.3 (Exact formula for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$)
Fix $n \in \{2,3,\ldots\}$. For the Bolthausen–Sznitman coalescent,

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \frac{2}{(n-1)!} \sum_{k=1}^{n-1} \frac{2^k - 1}{k^2} s(n-2, k-1), \tag{16}$$

where the $s(n,k)$ denote the absolute Stirling numbers of the first kind.

Remark. Together with the exact formula $\mathbb{E}(\tau_{n,1}) = ((n-1)!)^{-1} \sum_{k=1}^{n-1} s(n-1,k)/k$ for the mean of $\tau_{n,1}$ (see, for example, Proposition 1.2 of [11]) it can be checked that $\mathrm{Cov}(\tau_{n,1}, \tau_{n,2}) = \mathbb{E}(\tau_{n,1}\tau_{n,2}) - (\mathbb{E}(\tau_{n,1}))^2 > 0$ for all $n \geq 2$. Thus, for all $n \geq 2$, $\tau_{n,1}$ and $\tau_{n,2}$ are positively correlated. We have used the exact formulas for $\mathbb{E}(\tau_{n,1})$ and $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ to compute the entries of the following table.

n	$\mathbb{E}(\tau_{n,1})$	$\mathbb{E}(\tau_{n,1}\tau_{n,2})$	$\mathrm{Cov}(\tau_{n,1},\tau_{n,2})$
2	1	2	1
3	$\frac{3}{4} = 0.75$	$\frac{3}{4} = 0.75$	$\frac{3}{16} = 0.1875$
4	$\frac{23}{36} \approx 0.638889$	$\frac{55}{108} \approx 0.509259$	$\frac{131}{1296} \approx 0.101080$
5	$\frac{55}{96} \approx 0.572917$	$\frac{229}{576} \approx 0.397569$	$\frac{71}{1024} \approx 0.069336$
10	≈ 0.431647	≈ 0.215119	≈ 0.028800
100	≈ 0.228368	≈ 0.057067	≈ 0.004915
200	≈ 0.198537	≈ 0.042758	≈ 0.003341
300	≈ 0.184283	≈ 0.036676	≈ 0.002716
400	≈ 0.175300	≈ 0.033092	≈ 0.002362
500	≈ 0.168891	≈ 0.030652	≈ 0.002128
1000	≈ 0.151582	≈ 0.024546	≈ 0.001568
∞	$= \frac{1}{\log n} + \frac{1-\gamma}{\log^2 n} + O(\frac{1}{\log^3 n})$	$= \frac{1}{\log^2 n} + \frac{5/2-2\gamma}{\log^3 n} + O(\frac{1}{\log^4 n})$	$= \frac{1}{2\log^3 n} + O(\frac{1}{\log^4 n})$

Table 2: Covariance of $\tau_{n,1}$ and $\tau_{n,2}$ for the Bolthausen–Sznitman coalescent

Proof. (of Corollary 3.3) The substitution $u = -\log(1-t)$ in the integral on the right hand side in (15) yields

$$\begin{aligned}
f'(z) &= \frac{2}{-\log(1-z)} \int_0^{-\log(1-z)} \frac{e^{2u} - e^u}{u}\, du \\
&= \frac{2}{-\log(1-z)} \int_0^{-\log(1-z)} \frac{1}{u}\Big(\sum_{k=0}^{\infty} \frac{(2u)^k}{k!} - \sum_{k=0}^{\infty} \frac{u^k}{k!} \Big)\, du \\
&= \frac{2}{-\log(1-z)} \sum_{k=1}^{\infty} \frac{2^k - 1}{k!} \int_0^{-\log(1-z)} u^{k-1}\, du \\
&= \frac{2}{-\log(1-z)} \sum_{k=1}^{\infty} \frac{2^k - 1}{k!} \frac{(-\log(1-z))^k}{k} \\
&= 2\sum_{k=1}^{\infty} \frac{2^k - 1}{kk!} (-\log(1-z))^{k-1}.
\end{aligned}$$

This series expansion of f' in particular shows that f has an analytic continuation on $\mathbb{C} \setminus [1,\infty)$. From (see [1, p. 824]) $(-\log(1-z))^k/k! = \sum_{i=k}^{\infty} z^i/i! s(i,k)$ we conclude that

$$f'(z) \;=\; 2\sum_{k=1}^{\infty} \frac{2^k - 1}{k^2} \sum_{i=k-1}^{\infty} \frac{z^i}{i!} s(i,k-1) \;=\; 2\sum_{i=0}^{\infty} \frac{z^i}{i!} \sum_{k=1}^{i+1} \frac{2^k - 1}{k^2} s(i,k-1).$$

Thus, using the notation $[z^i]f'(z)$ for the coefficient in front of z^i in the series expansion of $f'(z)$,

$$(i+1)a_{i+2} \;=\; [z^i]f'(z) \;=\; \frac{2}{i!} \sum_{k=1}^{i+1} \frac{2^k - 1}{k^2} s(i,k-1)$$

or, equivalently,

$$a_{i+2} \;=\; \frac{2}{(i+1)!} \sum_{k=1}^{i+1} \frac{2^k - 1}{k^2} s(i,k-1).$$

It remains to substitute $n = i + 2$. □

Final remarks and open problems.

1. Based on the solution (15) for $f_{(1,1)}$ it is possible to derive the asymptotic expansion (8) of $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ alternatively via expansions of the functions and integrals occurring in (15). We leave the details to the reader.

2. Without to much effort it should be possible to derive differential equations in analogy to (14) for f_k for arbitrary $k = (k_1, \ldots, k_j)$ and to gain information on (the coefficients of the series expansion of) f_k from these differential equations (exact solutions and asymptotic expansions of the coefficients via singularity analysis).

4 Appendix

Basic expansions (Lemma 4.1 and Corollary 4.2) are provided which are needed in the proof of Theorem 1.2. We furthermore provide a result (see Lemma 4.3), which may serve as a stopping rule for difference procedures of the form used in the proof of Theorem 1.2.

Lemma 4.1 *For all $k \in \mathbb{N}$, as $n \to \infty$,*

$$\sum_{l=1}^{n-2} \frac{1}{\log^k(n-l)} \;=\; \frac{n}{\log^k n} + O\Big(\frac{n}{\log^{k+1} n}\Big), \tag{17}$$

$$\sum_{l=1}^{n-2} \frac{1}{l}\frac{1}{\log^k(n-l)} \;=\; \frac{1}{\log^{k-1} n} + \frac{\gamma}{\log^k n} + O\Big(\frac{1}{\log^{k+1} n}\Big), \tag{18}$$

$$\sum_{l=1}^{n-2} \frac{1}{l+1}\frac{1}{\log^k(n-l)} \;=\; \frac{1}{\log^{k-1} n} + \frac{\gamma-1}{\log^k n} + O\Big(\frac{1}{\log^{k+1} n}\Big), \tag{19}$$

and

$$\sum_{l=1}^{n-2} \frac{1}{l(l+1)}\frac{1}{\log^k(n-l)} \;=\; \frac{1}{\log^k n} + \frac{k-1}{n\log^k n} + O\Big(\frac{1}{n\log^{k+1} n}\Big). \tag{20}$$

Proof. Eq. (17) is a special case of Panholzer's summation formula (see [22, Lemma 4.1, Eq. (16)]) and probably as well a consequence of Euler's summation formula (see, for example, [12, p. 469]). We prefer here to verify (17) via a direct decomposition method, which will turn out to work as well for (18), (19), and (20). We basically split sums of the form $\sum_{l=1}^{n-2} \ldots$ into two parts $\sum_{l=1}^{a_n} \ldots$ and $\sum_{n=a_n+1}^{n-2} \ldots$, and handle these two parts separately. We will work with the sequence $(a_n)_{n\in\mathbb{N}}$ defined via $a_1 := 1$ and $a_n := n - \lfloor n/\log^{k+1} n \rfloor$ for $n \geq 2$. We will furthermore need the constants $c_k := 1/\log^k 2 > 0$, $k \in \mathbb{N}$. In order to verify (17) it is sufficient to show that

$$\sum_{l=1}^{n-2} \Big(\frac{1}{\log^k(n-l)} - \frac{1}{\log^k n}\Big) \;=\; O\Big(\frac{n}{\log^{k+1} n}\Big).$$

The sequence $(a_n)_{n\in\mathbb{N}}$ is chosen such that

$$\begin{aligned} 0 \;&\le\; \sum_{l=a_n+1}^{n-2}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}\Big) \;\le\; \sum_{l=a_n+1}^{n-2}\frac{1}{\log^k(n-l)} \\ &\le\; \sum_{l=a_n+1}^{n-2} c_k \;\le\; c_k(n-a_n) \;=\; O\Big(\frac{n}{\log^{k+1} n}\Big). \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{l=1}^{a_n}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}\Big) \;&=\; \sum_{l=1}^{a_n}\frac{\log^k n-\log^k(n-l)}{\log^k n\log^k(n-l)} \\ &\le\; \frac{1}{\log^k n\log^k(n-a_n)}\sum_{l=1}^{n-1}(\log^k n-\log^k(n-l)) \\ &\sim\; \frac{1}{\log^{2k} n}\sum_{l=1}^{n-1}(\log^k n-\log^k(n-l)), \end{aligned}$$

since $\log(n-a_n)\sim\log n$ as $n\to\infty$. Thus, in order to verify (17), it remains to show that

$$\sum_{l=1}^{n-1}(\log^k n-\log^k(n-l)) \;=\; O(n\log^{k-1} n). \tag{21}$$

Applying the formula $b^k-a^k=b^k-(a-b+b)^k=-\sum_{j=1}^{k}\binom{k}{j}(a-b)^j b^{k-j}$ with $a:=\log(n-l)$ and $b:=\log n$ yields

$$\begin{aligned} \sum_{l=1}^{n-1}(\log^k n-\log^k(n-l)) \;&=\; -\sum_{l=1}^{n-1}\sum_{j=1}^{k}\binom{k}{j}\log^{k-j} n\log^j(1-l/n) \\ &=\; \sum_{j=1}^{k}\binom{k}{j}\log^{k-j} n\sum_{l=1}^{n-1}(-\log^j(1-l/n)). \end{aligned}$$

Since for each $j\in\{1,\dots,k\}$,

$$\frac{1}{n}\sum_{l=1}^{n-1}(-\log^j(1-l/n)) \;\to\; \int_0^1(-\log^j(1-x))\,dx \;=\; (-1)^{j+1}j!,$$

it follows that (21) holds, and (17) is established.
Let us now verify (18). We essentially apply the same decomposition method. Note that $\sum_{l=1}^{n-2}1/l=\log n+\gamma+O(1/n)$. Thus, (18) holds if we can verify that

$$\sum_{l=1}^{n-2}\frac{1}{l}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}\Big) \;=\; O\Big(\frac{1}{\log^{k+1} n}\Big).$$

The sequence $(a_n)_{n\in\mathbb{N}}$ is chosen such that

$$0 \;\le\; \sum_{l=a_n+1}^{n-2} \frac{1}{l}\Big(\frac{1}{\log^k(n-l)} - \frac{1}{\log^k n}\Big) \;\le\; c_k \sum_{l=a_n+1}^{n-2} \frac{1}{l} \;=\; O\Big(\frac{1}{\log^{k+1} n}\Big).$$

Moreover,

$$\begin{aligned}
\sum_{l=1}^{a_n} \frac{1}{l}\Big(\frac{1}{\log^k(n-l)} - \frac{1}{\log^k n}\Big) &= \sum_{l=1}^{a_n} \frac{1}{l} \frac{\log^k n - \log^k(n-l)}{\log^k n \log^k(n-l)} \\
&\le \frac{1}{\log^k n \log^k(n-a_n)} \sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l)}{l} \\
&\sim \frac{1}{\log^{2k} n} \sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l)}{l},
\end{aligned}$$

since $\log(n-a_n) \sim \log n$ as $n \to \infty$. Thus, it remains to verify that

$$\sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l)}{l} \;=\; O(\log^{k-1} n). \tag{22}$$

Applying the formula $b^k - a^k = b^k - (a-b+b)^k = -\sum_{j=1}^k \binom{k}{j}(a-b)^j b^{k-j}$ to $a := \log(n-l)$ and $b := \log n$ yields

$$\begin{aligned}
\sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l)}{l} &= -\sum_{l=1}^{n-1} \frac{1}{l} \sum_{j=1}^{k} \binom{k}{j} \log^{k-j} n \log^j(1-l/n) \\
&= \sum_{j=1}^{k} \binom{k}{j} \log^{k-j} n \sum_{l=1}^{n-1} \frac{-\log^j(1-l/n)}{l}.
\end{aligned}$$

Since for each $j \in \{1,\dots,k\}$,

$$\sum_{l=1}^{n-1} \frac{-\log^j(1-l/n)}{l} \;\to\; \int_0^1 \frac{-\log^j(1-x)}{x}\,dx \;=\; (-1)^{j+1} j!\zeta(j+1) \;\in\; \mathbb{R},$$

it follows that (22) holds and (18) is established. Eq. (19) is shown similarly. Let us finally come to (20). Note first that

$$\begin{aligned}
n\sum_{l=1}^{n-2} \frac{-\log(1-l/n)}{l(l+1)} &= \frac{1}{n}\sum_{l=1}^{n-2} \frac{-\log(1-l/n)}{(l/n)(l+1)/n} \;=\; \int_{1/n}^{1-1/n} \frac{-\log(1-x)}{x^2}\,dx + O(1) \\
&= \Big[\log x + \frac{(1-x)\log(1-x)}{x}\Big]_{1/n}^{1-1/n} + O(1) \;=\; \log n + O(1)
\end{aligned}$$

and, therefore,

$$\sum_{l=1}^{n-2} \frac{1}{l(l+1)}\Big(\frac{1}{\log^k n} + k\frac{-\log(1-l/n)}{\log^{k+1} n}\Big)$$

$$
\begin{aligned}
&= \frac{1}{\log^k n}\sum_{l=1}^{n-2}\Big(\frac{1}{l}-\frac{1}{l+1}\Big)+\frac{k}{\log^{k+1} n}\sum_{l=1}^{n-2}\frac{-\log(1-l/n)}{l(l+1)}\\
&= \frac{1}{\log^k n}\Big(1-\frac{1}{n-1}\Big)+\frac{k}{\log^{k+1} n}\sum_{l=1}^{n-2}\frac{-\log(1-l/n)}{l(l+1)}\\
&= \frac{1}{\log^k n}\Big(1-\frac{1}{n}+O\Big(\frac{1}{n^2}\Big)\Big)+\frac{k}{\log^{k+1} n}\Big(\frac{\log n}{n}+O\Big(\frac{1}{n}\Big)\Big)\\
&= \frac{1}{\log^k n}+\frac{k-1}{n\log^k n}+O\Big(\frac{1}{n\log^{k+1} n}\Big).
\end{aligned}
$$

Thus, in order to verify (20), it suffices to verify that

$$
\sum_{l=1}^{n-2}\frac{1}{l(l+1)}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}-k\frac{-\log(1-l/n)}{\log^{k+1} n}\Big) \;=\; O\Big(\frac{1}{n\log^{k+1} n}\Big).
$$

The function $f_{nk}:[0,n)\to\mathbb{R}$, defined via

$$
f_{nk}(x) \;:=\; \frac{1}{\log^k(n-x)}-\frac{1}{\log^k n}-k\frac{-\log(1-x/n)}{\log^{k+1} n}
$$

has derivative

$$
f'_{nk}(x) \;=\; \frac{k}{n-x}\Big(\frac{1}{\log^{k+1}(n-x)}-\frac{1}{\log^{k+1} n}\Big) \;\geq\; 0
$$

and satisfies $f_{nk}(0)=0$. Thus, $f_{nk}\geq 0$ and it follows that $(a_n)_{n\in\mathbb{N}}$ is chosen such that

$$
\begin{aligned}
0 \;&\leq\; \sum_{l=a_n+1}^{n-2}\frac{1}{l(l+1)}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}-k\frac{-\log(1-l/n)}{\log^{k+1} n}\Big)\\
&\leq\; \sum_{l=a_n+1}^{n-2}\frac{1}{l(l+1)}\frac{1}{\log^k(n-l)} \;\leq\; c_k\sum_{l=a_n+1}^{n-2}\Big(\frac{1}{l}-\frac{1}{l+1}\Big)\\
&=\; c_k\Big(\frac{1}{a_n+1}-\frac{1}{n-1}\Big) \;=\; c_k\frac{n-a_n-2}{(a_n+1)(n-1)}\\
&\sim\; c_k\frac{n/\log^{k+1} n}{n^2} \;=\; O\Big(\frac{1}{n\log^{k+1} n}\Big).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\sum_{l=1}^{a_n}\frac{1}{l(l+1)}\Big(\frac{1}{\log^k(n-l)}-\frac{1}{\log^k n}-k\frac{-\log(1-l/n)}{\log^{k+1} n}\Big)\\
&\quad= \sum_{l=1}^{a_n}\frac{1}{l(l+1)}\frac{\log^k n-\log^k(n-l)-k(-\log(1-l/n))\log^k(n-l)/\log n}{\log^k n\log^k(n-l)}\\
&\quad\leq \frac{1}{\log^k n\log^k(n-a_n)}\sum_{l=1}^{a_n}\frac{\log^k n-\log^k(n-l)+k\log(1-l/n)\log^k(n-l)/\log n}{l(l+1)}
\end{aligned}
$$

$$\sim \quad \frac{1}{\log^{2k} n} \sum_{l=1}^{a_n} \frac{\log^k n - \log^k(n-l) + k\log(1-l/n)\log^k(n-l)/\log n}{l(l+1)},$$

since $\log(n-a_n) \sim \log n$ as $n \to \infty$. Thus, it remains to verify that

$$\sum_{l=1}^{a_n} \frac{\log^k n - \log^k(n-l) + k\log(1-l/n)\log^k(n-l)/\log n}{l(l+1)} \;=\; O\Big(\frac{\log^{k-1} n}{n}\Big). \tag{23}$$

Applying the formula $b^k - a^k + k(a-b)a^k/b = (1-k(a-b)/b)(b^k-a^k) + k(a-b)b^{k-1} = (1-k(a-b)/b)(-\sum_{j=1}^{k}\binom{k}{j}(a-b)^j b^{k-j}) + k(a-b)b^{k-1} = \sum_{j=2}^{k+1}\big(k\binom{k}{j-1} - \binom{k}{j}\big)(a-b)^j b^{k-j}$ with $a := \log(n-l)$ and $b := \log n$ yields

$$\begin{aligned}
0 \;&\le\; \sum_{l=1}^{a_n} \frac{\log^k n - \log^k(n-l) + k\log(1-l/n)\log^k(n-l)/\log n}{l(l+1)} \\
&\le\; \sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l) + k\log(1-l/n)\log^k(n-l)/\log n}{l(l+1)} \\
&=\; \sum_{l=1}^{n-1} \frac{1}{l(l+1)} \sum_{j=2}^{k+1} \left(k\binom{k}{j-1} - \binom{k}{j}\right) \log^{k-j} n \log^j(1-l/n) \\
&=\; \sum_{j=2}^{k+1} \left(k\binom{k}{j-1} - \binom{k}{j}\right) \log^{k-j} n \sum_{l=1}^{n-1} \frac{\log^j(1-l/n)}{l(l+1)}.
\end{aligned}$$

Since for each $j \in \{2,\ldots,k+1\}$,

$$n\sum_{l=1}^{n-1} \frac{\log^j(1-l/n)}{l(l+1)} \;=\; \frac{1}{n}\sum_{l=1}^{n-1} \frac{\log^j(1-l/n)}{(l/n)((l+1)/n)} \;\to\; \int_0^1 \frac{\log^j(1-x)}{x^2}\,dx \;\in\; \mathbb{R},$$

it follows that the expression on the left hand side in (23) is even equal to $O((\log^{k-2} n)/n)$. In particular, (23) holds, and, hence, (20) is established. □

Corollary 4.2 *For all $j,k \in \mathbb{N}$, as $n \to \infty$,*

$$\begin{aligned}
&\sum_{m=2}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} \\
&\qquad =\quad \frac{n^j}{\log^k n} - j\frac{n^{j-1}}{\log^{k-1} n} + \Big(k + j\Psi(j) - \frac{j(j+1)}{2}\Big)\frac{n^{j-1}}{\log^k n} + O\Big(\frac{n^{j-1}}{\log^{k+1} n}\Big).
\end{aligned}$$

Proof. We have

$$\sum_{m=2}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} \;=\; \sum_{l=1}^{n-2} \frac{(n-l-1)_j}{l(l+1)\log^k(n-l)}.$$

Plugging in

$$\begin{aligned}(n-l-1)_j &= \prod_{i=1}^{j}(n-(l+i))\\ &= n^j - n^{j-1}\sum_{i=1}^{j}(l+i) + \sum_{r=2}^{j}(-1)^r n^{j-r}\sum_{1\le i_1<\cdots<i_r\le j}(l+i_1)\cdots(l+i_r)\\ &= n^j - n^{j-1}\Big(jl+\frac{j(j+1)}{2}\Big) + \sum_{r=2}^{j}(-1)^r n^{j-r}\bigg(\binom{j}{r}l^r + O(l^{r-1})\bigg)\end{aligned}$$

it follows that

$$\begin{aligned}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} &= n^j\sum_{l=1}^{n-2}\frac{1}{l(l+1)\log^k(n-l)}\\ &\quad -jn^{j-1}\sum_{l=1}^{n-2}\frac{1}{(l+1)\log^k(n-l)} - \frac{j(j+1)}{2}n^{j-1}\sum_{l=1}^{n-2}\frac{1}{l(l+1)\log^k(n-l)}\\ &\quad +\sum_{r=2}^{j}(-1)^r n^{j-r}\sum_{l=1}^{n-2}\frac{\binom{j}{r}l^r+O(l^{r-1})}{l(l+1)\log^k(n-l)}.\end{aligned}$$

Applying Lemma 4.1 to all the sums over l on the right hand side yields

$$\begin{aligned}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^k m} &= n^j\Big(\frac{1}{\log^k n}+\frac{k-1}{n\log^k n}+O\Big(\frac{1}{n\log^{k+1} n}\Big)\Big)\\ &\quad -jn^{j-1}\Big(\frac{1}{\log^{k-1} n}+\frac{\gamma-1}{\log^k n}+O\Big(\frac{1}{\log^{k+1} n}\Big)\Big) - \frac{j(j+1)}{2}n^{j-1}\Big(\frac{1}{\log^k n}+O\Big(\frac{1}{n\log^k n}\Big)\Big)\\ &\quad +\sum_{r=2}^{j}(-1)^r n^{j-r}\bigg(\binom{j}{r}\frac{1}{r-1}\frac{n^{r-1}}{\log^k n}+O\Big(\frac{n^{r-1}}{\log^{k+1} n}\Big)\bigg)\\ &= \frac{n^j}{\log^k n} - j\frac{n^{j-1}}{\log^{k-1} n}\\ &\quad +\bigg(k-1-j(\gamma-1)-\frac{j(j+1)}{2}+\sum_{r=2}^{j}(-1)^r\binom{j}{r}\frac{1}{r-1}\bigg)\frac{n^{j-1}}{\log^k n}+O\Big(\frac{n^{j-1}}{\log^{k+1} n}\Big).\end{aligned}$$

The result follows, since $\sum_{r=2}^{j}(-1)^r\binom{j}{r}\frac{1}{r-1} = jh_{j-1} - j + 1 = j\Psi(j) + j\gamma - j + 1$ for all $j \in \mathbb{N}$. □

The following Lemma 4.3 serves as a stopping rule for difference procedures of the form as for example used in the proof of Theorem 1.2. For similar stopping rules we refer the reader to [17, Lemma A.2].

Lemma 4.3 *Fix $j \in \mathbb{N}$ and let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers satisfying the recursion*

$$a_n = q_n + \sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}\,a_m, \qquad n \in \{2,3,\ldots\}$$

for some given sequence $(q_n)_{n\in\mathbb{N}}$. *If* $q_n = O(1/(n\log^k n))$ *for some fixed* $k \in \mathbb{N}$, *then* $a_n = O(1/\log^{k+1} n)$.

Proof. By assumption there exists a constant $C > 0$ such that $|q_n| \le C/(n\log^k n)$ for all $n \in \mathbb{N}\setminus\{1\}$. Moreover, by Corollary 4.2,

$$\begin{aligned}\sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}\frac{1}{\log^{k+1} m} &= \frac{n}{(n-1)(n)_j}\sum_{m=2}^{n-1}\frac{(m-1)_j}{(n-m)(n-m+1)\log^{k+1} m}\\ &= \Big(\frac{1}{n^j}+O\Big(\frac{1}{n^{j+1}}\Big)\Big)\Big(\frac{n^j}{\log^{k+1} n}-j\frac{n^{j-1}}{\log^k n}+O\Big(\frac{n^{j-1}}{\log^{k+1} n}\Big)\Big)\\ &= \frac{1}{\log^{k+1} n}-j\frac{1}{n\log^k n}+O\Big(\frac{1}{n\log^{k+1} n}\Big).\end{aligned}$$

Thus, there exists a constant $n_0 \in \mathbb{N}$ such that the inequality

$$\sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}\frac{1}{\log^{k+1} m} \le \frac{1}{\log^{k+1} n}-\frac{j}{2}\frac{1}{n\log^k n}$$

holds for all integers $n > n_0$. Choose a constant $D > 2C/j$ sufficiently large such that

$$|a_n| \le \frac{D}{\log^{k+1} n} \tag{24}$$

for all $n \in \{2,3,\dots,n_0\}$. In the following it is verified by induction on n that (24) holds for all $n \in \mathbb{N}\setminus\{1\}$. For $n \in \{2,3,\dots,n_0\}$, (24) holds by the choice of D. Suppose now that $n > n_0$ and that $|a_m| \le D/\log^{k+1} m$ is already verified for all $m \in \{2,3,\dots,n-1\}$. Then,

$$\begin{aligned}|a_n| &\le |q_n| + \sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}|a_m| \le \frac{C}{n\log^k n}+\sum_{m=2}^{n-1} p_{nm}\frac{(m-1)_j}{(n)_j}\frac{D}{\log^{k+1} m}\\ &\le \frac{C}{n\log^k n}+D\Big(\frac{1}{\log^{k+1} n}-\frac{j}{2}\frac{1}{n\log^k n}\Big) = \frac{D}{\log^{k+1} n}+\frac{C-jD/2}{n\log^k n} \le \frac{D}{\log^{k+1} n},\end{aligned}$$

since $C - jD/2 < 0$. Thus, the induction is completed and the lemma is established. □

Acknowledgement. The first author benefited from the support of the 'Agence Nationale de la Recherche': ANR MANEGE (ANR09-BLAN-0215). Research was partially completed while the second author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2011. The second author furthermore acknowledges the support of the Université Paris 13. He in particular thanks the Laboratoire Analyse, Géometrie et Applications at the Université Paris 13 for its kind hospitality during a one month visit in 2011, where parts of this work was done.

References

[1] Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* 9th printing. Dover, New York. MR1225604

[2] Billingsley, P. (1995) *Probability and Measure.* 3rd Edition, Wiley, New York. MR1324786

[3] Billingsley, P. (1999) *Convergence of Probability Measures.* 2nd Edition, Wiley, New York. MR1700749

[4] Bolthausen, E. and Sznitman, A.-S. (1998) On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197**, 247–276. MR1652734

[5] Caliebe, A., Neininger, R., Krawczak, M., and Rösler, U. (2007) On the length distribution of external branches in coalescent trees: genetic diversity within species. *Theor. Popul. Biol.* **72**, 245–252. MR number not available

[6] Devroye, L. (1999) Universal limit laws for depths in random trees. *SIAM J. Comput.* **28**, 409–432. MR1634354

[7] Dhersin, J.-S., Freund, F., Siri-Jégousse, A., and Yuan, L. (2012) On the length of an external branch in the beta-coalescent. Preprint.

[8] Dhersin, J.-S. and Yuan, L. (2012) On the total length of external branches for beta coalescents. Preprint.

[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. MR2353033

[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 Struct. Algorithms* **34**, 319–336. MR2504401

[11] 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 Processes Relat. Fields* **15**, 387–416. MR2554368

[12] Graham, R. L., Knuth, D. E., and Patashnik, O. (1994) *Concrete Mathematics. A foundation for computer science.* Second Edition. Addison-Wesley Publishing Company, Reading, MA. MR1397498

[13] Holmgren, C. (2008) Random records and cuttings in split trees. Extended abstract. *Discrete Math. Theor. Comput. Sci. (DMTCS) Proceedings Series Volume* **A1**, 273–286. MR2508793

[14] Holmgren, C. (2010) Random records and cuttings in binary search trees. *Combin. Probab. Comput.* **19**, 391–424. MR2607374

[15] Holmgren, C. (2011) A weakly 1-stable distribution for the number of random records and cuttings in split trees. *Adv. Appl. Probab.* **43**, 151–177. MR2761152

[16] Iksanov, A. and Möhle, M. (2007) A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.* **12**, 28–35. MR2407414

[17] Iksanov, A., Marynych, A., and Möhle, M. (2009) On the number of collisions in beta(2,b)-coalescents. *Bernoulli* **15**, 829–845. MR2555201

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

[19] Kingman, J.F.C. (1982) The coalescent. *Stoch. Process. Appl.* **13**, 235–248. MR0671034

[20] Möhle, M. (2006) On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, 35–53. MR2202319

[21] Möhle, M. (2010) Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent. *Stoch. Process. Appl.* **120**, 2159–2173. MR2684740

[22] Panholzer, A. (2004) Destruction of recursive trees. In: Mathematics and Computer Science III. Birkhäuser, Basel, 267–280. MR2090518

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

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