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

Jean-Stephane Dhersin, Martin Moehle

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

J.-S. Dhersin[1] and M. Möhle[2] January 24, 2013

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 joint moments of the external branch lengths and for the moments of the total external branch length of the Bolthausen–Sznitman coalescent. These expansions are based on a differential equation approach, which is as well useful 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 [22] and [23]. 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, [19, Eq. (13)]) that

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

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

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

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

[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

length of the ith external branch of the restricted coalescent $\Pi^{(n)}$. Note that $\tau_{1,1} = 0$. Our first main result (Theorem 1.1) provides a general recursion for the joint moments

$$\mu_n(k_1, \ldots, k_j) \;:=\; \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j}), \qquad j \in \{1, \ldots, n\}, k_1, \ldots, k_j \in \mathbb{N}_0 := \{0, 1, \ldots\}, \tag{3}$$

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

$$\mu_n(k) \;=\; \frac{1}{g_n} \sum_{i=1}^{j} k_i \, \mu_n(k - e_i) + \sum_{m=j+1}^{n-1} p_{nm} \frac{(m-1)_j}{(n)_j} \, \mu_m(k), \tag{4}$$

where e_i, $i \in \{1, \ldots, j\}$, denotes the ith unit vector in $\mathbb{R}^j$, $p_{nm} := g_{nm}/g_n$ and g_{nm} and g_n are defined via (1) and (2).

Remarks. The recursion (4) works as follows. Let us call $d := k_1 + \cdots + k_j$ the order (or degree) of the moment $\mu_n(k_1, \ldots, k_j)$. Provided that all the moments of order $d-1$ are already known, (4) is a recursion on n for the joint moments of order d, which can be solved iteratively. So one starts with $d = 1$ (and hence $j = 1$), in which case (4) reduces to $\mu_n(1) = 1/g_n + \sum_{m=2}^{n-1} p_{nm}((m-1)/n)\mu_m(1)$, $n \geq 2$. 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 $d = 2$ in (4) leading to a recursion for the second order moments. Iteratively, one can move to higher orders. Note that for $j = 2$ and $k_1 = k_2 = 1$ the recursion (4) reduces to

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \frac{2}{g_n}\mathbb{E}(\tau_{n,1}) + \sum_{m=2}^{n-1} p_{nm} \frac{(m-1)_2}{(n)_2} \mathbb{E}(\tau_{m,1}\tau_{m,2}), \quad n \in \{2, 3, \ldots\}. \tag{5}$$

Note that Theorem 1.1 holds for arbitrary Λ-coalescents. For particular Λ-coalescents the recursion (4) 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 Bolthausen–Sznitman coalescent. For related results on external branches for beta-coalescents we refer the reader to [7], [8] and [17].

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, $p_{nm} = \delta_{m1}$ for $n, m \in \mathbb{N}$ with $m < n$. Thus, (4) reduces to $\mu_n(k) = \sum_{i=1}^{j} k_i \, \mu_n(k - e_i)$ with solution $\mu_n(k) = (k_1 + \cdots + k_j)!$, which is obviously correct, since $\tau_{n,i} = T_n$ for all $i \in \{1, \ldots, n\}$ and, therefore, $\mu_n(k) = \mathbb{E}(T_n^{k_1 + \cdots + k_j}) = (k_1 + \cdots + k_j)!$, $n \geq 2$, $j \in \{1, \ldots, n\}$, $k_1, \ldots, k_j \in \mathbb{N}$.

Example. (Kingman coalescent) For the Kingman coalescent [18], 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, $p_{nm} = \delta_{m,n-1}$ for $m, n \in \mathbb{N}$ with $m < n$. 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 [16, 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 (4) 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 [16, p. 2205]. For the third moment we obtain

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

Since $h_{m+1} - h_m = 1/(m+1)$, the last sum simplifies considerably to

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

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

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

so these generating functions satisfy the recursion

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

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 $\mathrm{Li}_2(t) := -\int_0^t(\log(1-x))/x\,\mathrm{d}x = \sum_{k=1}^{\infty} t^k/k^2$ 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 (4) for $j = 2$ and $k_1 = k_2 = 1$ reduces to (see (5))

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

It is readily checked by induction on n that this recursion is solved 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, $\mu_n(1,1) = 4/n^2 - 4/n^3 + O((\log n)/n^4)$, $n \to \infty$. 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	0.666667	0.444444	0
4	0.5	0.240741	−0.009259
5	0.4	0.152222	−0.007778
10	0.2	0.038096	−0.001904
100	0.02	0.000396	−0.000004
$n \to \infty$	$\frac{2}{n}$	$\frac{4}{n^2} - \frac{4}{n^3} + O(\frac{\log n}{n^4})$	$-\frac{4}{n^3} + O(\frac{\log n}{n^4})$

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 $\mu_n(k) := \mathbb{E}(\tau_{n,1}^{k_1}\cdots\tau_{n,j}^{k_j})$ of the lengths $\tau_{n,1},\ldots,\tau_{n,n}$ of the external branches satisfy the asymptotic expansion

$$\mu_n(k) \;=\; \frac{k_1!\cdots k_j!}{\log^{k_1+\cdots+k_j} n}\left(1 + \frac{\kappa_j(k)}{\log n} + O\Big(\frac{1}{\log^2 n}\Big)\right), \quad j \in \mathbb{N}, k = (k_1,\ldots,k_j) \in \mathbb{N}_0^j, \qquad (6)$$

where the coefficients $\kappa_j(k)$, $j \in \mathbb{N}$, $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$, *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*

$$\kappa_j(k) \;:=\; \Psi(j+1) - 1 + \frac{1}{j}\sum_{i=1}^{j} k_i + \frac{1}{j}\sum_{i=1}^{j} \kappa_j(k - e_i), \qquad j \in \mathbb{N}, k = (k_1, \dots, k_j) \in \mathbb{N}^j.$$

Here $\Psi := \Gamma'/\Gamma$ *denotes the digamma function (derivative of* $\log\Gamma$*). Note that* $\Psi(t) = -\gamma + \int_0^1 (1 - (1-x)^{t-1})/x \,\mathrm{d}x$ *for all* $t > 0$, *where* $\gamma = -\Psi(1) \approx 0.577216$ *denotes Euler's constant.*

Remarks.

1. The coefficient $\kappa_j(k_1, \dots, k_j)$ is symmetric with respect to the j coordinates $k_1, \dots, k_j$ and strictly positive whenever $d := k_1 + \cdots + k_j > 0$.

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

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

   It does not seem to be straightforward to derive a simple formula for $\kappa_j(k)$ for general $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$.

2. For $j = 2$ and $k_1 = k_2 = 1$ the expansion (6) has the form

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \mu_n(1,1) \;=\; \frac{1}{\log^2 n} + \frac{\kappa_2(1,1)}{\log^3 n} + O\biggl(\frac{1}{\log^4 n}\biggr), \qquad n \to \infty, \tag{7}$$

   with $\kappa_2(1,1) = \Psi(2) + \Psi(3) = 5/2 - 2\gamma \approx 1.345569$. 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 3.2 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 $n \geq 2$, in contrast to 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 two corollaries are a direct consequence of Theorem 1.2.

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

The following result 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 expansion*

$$\mathbb{E}((L_n^{external})^k) \;=\; \frac{n^k}{\log^k n}\biggl(1 + \frac{c_k}{\log n} + O\biggl(\frac{1}{\log^2 n}\biggr)\biggr), \qquad n \to \infty, \tag{8}$$

where $c_k := \sum_{i=2}^{k+1} \Psi(i) = (k+1)h_k - k - k\gamma$.

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 (8) 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 $\mathbb{E}(L_n^k)$. In particular, the total internal branch length $L_n^{internal} := L_n - L_n^{external}$ has expectation $\mathbb{E}(L_n^{internal}) = \mathbb{E}(L_n - L_n^{external}) = (m_1 - c_1)n/\log^2 n + O(n/\log^3 n) = n/\log^2 n + O(n/\log^3 n)$.

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 [21, p. 277] yields the asymptotic expansion

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

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,

$$\mathrm{Var}(L_n^{external}) \;=\; \frac{1}{2}\frac{n^2}{\log^3 n} + O\biggl(\frac{n^2}{\log^4 n}\biggr), \qquad n \to \infty. \tag{9}$$

The moments of $L_n^{external}$ do not provide much information on the distributional limiting behavior of $L_n^{external}$ as $n \to \infty$. Nevertheless, the expansions of the centered moments of L_n and $L_n^{external}$ coincide, which supports (or at least does not contradict) the intuition that the distributional limiting behavior of $L_n^{external}$ should essentially coincide with that of L_n (see, for example, [9, Theorem 5.2]). Note that the total internal branch length $L_n^{internal} = L_n - L_n^{external}$ satisfies

$$\frac{\log^2 n}{n}\mathbb{E}(L_n^{internal}) \;=\; 1 + O\biggl(\frac{1}{\log n}\biggr) \;\to\; 1$$

as $n \to \infty$. We conjecture that, for the Bolthausen–Sznitman coalescent,

$$\lim_{n\to\infty} \frac{\log^4 n}{n^2}\mathrm{Var}(L_n^{internal}) \;=\; 0 \tag{10}$$

and show at the end of Section 4 that (10) implies the following result.

Conjecture 1.5 (Weak convergence of the total external branch length)
For the Bolthausen–Sznitman coalescent, if the conjectured formula (10) holds, then

$$\frac{\log^2 n}{n}L_n^{internal} \;\to\; 1$$

in probability as $n \to \infty$ and

$$\frac{\log^2 n}{n}L_n^{external} - \log n - \log\log n \;\to\; L - 1 \tag{11}$$

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, except for the additional shift -1 on the right hand side in (11), the same limiting law as in (11) is known for the number of cuts needed to isolate the root of a random recursive tree ([10], [15]). Essentially the same scaling and convergence result has been obtained for random records and cuttings in binary search trees by Holmgren [13, Theorem 1.1] and more generally in split trees (Holmgren [12, Theorem 1.1] and [14, 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 1-stable limiting laws.
To the best of the authors knowledge the weak limiting behavior of $L_n^{internal}$ is unknown for the Bolthausen–Sznitman coalescent. We conjecture that $L_n^{internal}$, properly centered and scaled, is asymptotically normal.

2 Proof of Theorem 1.1

Let $T = T_n$ denote the time of the first jump of the block counting process $N^{(n)}$ and let $I = I_n$ denote the state of $N^{(n)}$ at its first jump. Note that T and I are independent, T is exponentially distributed with parameter g_n and $p_{nm} := \mathbb{P}(I = m) = g_{nm}/g_n$, $m \in \{1, \dots, n-1\}$. For $i \in \{1, \dots, n\}$ and $h > 0$ define $\tau_i' := \tau_{n,i} - h \wedge T$. By the Markov property, for $h \to 0$,

$$\begin{aligned}\mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j} 1_{\{T>h\}}) &= \mathbb{E}((\tau_1' + h)^{k_1} \cdots (\tau_j' + h)^{k_j} 1_{\{T>h\}}) \\ &= \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})\mathbb{P}(T > h) + h \sum_{i=1}^{j} k_i \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,i-1}^{k_{i-1}} \tau_{n,i}^{k_i - 1} \tau_{n,i+1}^{k_{i+1}} \cdots \tau_{n,j}^{k_j}) + o(h).\end{aligned}$$

Also for $h \to 0$,

$$\mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j} 1_{\{T \le h\}}) = \mathbb{E}((\tau_1' + T)^{k_1} \cdots (\tau_j' + T)^{k_j} 1_{\{T \le h\}}) = \mathbb{E}((\tau_1')^{k_1} \cdots (\tau_j')^{k_j} 1_{\{T \le h\}}) + o(h).$$

Now at time T either the event $A := \{\text{one of the individuals 1 to } j \text{ is involved in the first collision}\}$ occurs, in which case $\tau_i' = 0$ for some $i \in \{1, \dots, j\}$, and the above expectation vanishes since $k_1, \dots, k_j > 0$, or none of these j individuals is involved in the first collision. Then, by the strong Markov property,

$$\mathbb{E}((\tau_1')^{k_1} \cdots (\tau_j')^{k_j} 1_{\{T \le h, I = m, A^c\}}) = \mathbb{E}(\tau_{m,1}^{k_1} \cdots \tau_{m,j}^{k_j})\mathbb{P}(T \le h, I = m, A^c),$$

where A^c denotes the complement of A. Adding both expectations yields

$$\begin{aligned}\mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j}) &= \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})\mathbb{P}(T > h) + h \sum_{i=1}^{j} k_i \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,i-1}^{k_{i-1}} \tau_{n,i}^{k_i - 1} \tau_{n,i+1}^{k_{i+1}} \cdots \tau_{n,j}^{k_j}) \\ &\quad + \sum_{m=j+1}^{n-1} \mathbb{E}(\tau_{m,1}^{k_1} \cdots \tau_{m,j}^{k_j})\mathbb{P}(T \le h)\mathbb{P}(I = m)\frac{(m-1)_j}{(n)_j} + o(h).\end{aligned}$$

Collecting both terms involving $\mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})$ on the left hand side and letting $h \to 0$ gives the claim, since $\mathbb{P}(T \le h) = 1 - e^{-g_n h} \sim g_n h$ as $h \to 0$. □

3 Differential equations approach

A differential equations approach is provided, which is used in the proof of Theorem 1.2 given in the following Section 4. This approach furthermore yields for example an exact expression for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ in terms of Stirling numbers (see Corollary 3.2). Let $D := \{z \in \mathbb{C} : |z| < 1\}$ denote the open unit disc in the complex plane. For $j \in \mathbb{N}$ and $k = (k_1, \ldots, k_j) \in \mathbb{N}_0^j$ define the generating function

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

where, for $n \geq j$, we use the abbreviation $a_n := \mu_n(k) := \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})$ for convenience. Note that, due to the natural coupling property of n-coalescents, the sequence $(a_n)_{n \geq j}$ is non-increasing. Thus, f_k and all its derivatives $f_k', f_k'', \ldots$ are analytic functions on D. In order to state the following result it is convenient to introduce $L(z) := -\log(1-z)$, $z \in D$, and to define the functions $g_k : D \to \mathbb{C}$, $k = (k_1, \ldots, k_j) \in \mathbb{N}^j$, via $g_1(z) := z/(1-z)$ and

$$g_k(z) \;:=\; \sum_{i=1}^{j} k_i f_{k-e_i}^{(j-1)}(z) \tag{12}$$

for all $z \in D$ and all $k = (k_1, \ldots, k_j) \in \mathbb{N}^j$ satisfying $k_1 + \cdots + k_j > 1$, where e_i, $i \in \{1, \ldots, j\}$, denotes the ith unit vector in $\mathbb{R}^j$.

Lemma 3.1 *For the Bolthausen–Sznitman coalescent, the function f_k, $k = (k_1, \ldots, k_j) \in \mathbb{N}^j$, satisfies the differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}z}\big((L(z))^{j-1} f_k^{(j-1)}(z)\big) \;=\; \frac{(L(z))^{j-2}}{1-z} g_k(z), \quad z \in D \setminus \{0\}, \tag{13}$$

with solution

$$f_k^{(j-1)}(z) \;=\; \frac{1}{(L(z))^{j-1}} \int_0^z \frac{(L(t))^{j-2}}{1-t} g_k(t)\,\mathrm{d}t, \quad z \in D \setminus \{0\}. \tag{14}$$

In particular,

$$f_1(z) \;=\; \int_0^z \frac{t}{(1-t)^2 L(t)}\,\mathrm{d}t \quad \textit{and} \quad f_{(1,1)}'(z) \;=\; \frac{2}{L(z)} \int_0^z \frac{t}{(1-t)^3 L(t)}\,\mathrm{d}t, \quad z \in D \setminus \{0\}. \tag{15}$$

Proof. 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, $p_{nm} := g_{nm}/g_n = n/((n-1)(n-m)(n-m+1))$, $m, n \in \mathbb{N}$ with $m < n$. Fix $j \in \mathbb{N}$ and $k = (k_1, \ldots, k_j) \in \mathbb{N}^j$ and, for $n \in \mathbb{N}$, define $a_n := \mu_n(k)$ for convenience. For $n \geq \max(2, j)$ the recursion (4) reads

$$a_n \;=\; q_n + \sum_{m=j+1}^{n-1} p_{nm} \frac{(m-1)_j}{(n)_j}\, a_m \;=\; q_n + \frac{n}{(n-1)(n)_j} \sum_{m=j+1}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)}\, a_m,$$

where $q_n := g_n^{-1} \sum_{i=1}^{j} k_i \mu_n(k - e_i)$ for all $n \geq \max(2, j)$. Thus,

$$(n-1)(n-1)_{j-1} a_n \;=\; (n-1)(n-1)_{j-1} q_n + \sum_{m=j+1}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)} a_m. \tag{16}$$

Before we come back to the recursion (16) let us first verify that

$$\sum_{n=\max(2,j)}^{\infty} (n-1)(n-1)_{j-1} q_n z^{n-j} \;=\; g_k(z), \qquad z \in D. \tag{17}$$

Obviously (17) holds for $j = 1$ and $k_1 = 1$, since in this case $q_n = 1/g_n = 1/(n-1)$ and $g_1(z) = z/(1-z)$ by definition. For $k = (k_1, \ldots, k_j) \in \mathbb{N}^j$ with $k_1 + \cdots + k_j > 1$ we have

$$\begin{aligned}\sum_{n=\max(2,j)}^{\infty} (n-1)(n-1)_{j-1} q_n z^{n-j} \;&=\; \sum_{n=\max(2,j)}^{\infty} (n-1)_{j-1} \sum_{i=1}^{j} k_i \mu_n(k-e_i) z^{n-j} \\ &=\; \Big(\frac{\mathrm{d}}{\mathrm{d}z}\Big)^{j-1} \sum_{i=1}^{j} k_i \sum_{n=\max(2,j)}^{\infty} \mu_n(k-e_i) z^{n-1} \;=\; \sum_{i=1}^{j} k_i f_{k-e_i}^{(j-1)}(z) \;=\; g_k(z).\end{aligned}$$

Thus, (17) is established. In view of $(n-1)(n-1)_{j-1} = (n-1)_j + (j-1)(n-1)_{j-1}$ and (17), by multiplying both sides in (16) with z^{n-j} and summing over all $n \geq \max(2,j)$, the recursion (16) translates to

$$\begin{aligned} z f_k^{(j)}(z) + (j-1) f_k^{(j-1)}(z) \;&=\; g_k(z) + \sum_{n=\max(2,j)}^{\infty} \sum_{m=j+1}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)} a_m z^{n-j} \\ &=\; g_k(z) + \sum_{m=j+1}^{\infty} (m-1)_j a_m z^{m-j} \sum_{n=m+1}^{\infty} \frac{1}{(n-m)(n-m+1)} z^{n-m} \\ &=\; g_k(z) + z a(z) \Big(\frac{\mathrm{d}}{\mathrm{d}z}\Big)^{j} \sum_{m=j}^{\infty} a_m z^{m-1} \\ &=\; g_k(z) + z a(z) f_k^{(j)}(z), \end{aligned} \tag{18}$$

where $a(z) := \sum_{n=1}^{\infty} z^n/(n(n+1))$ for $z \in D$. Since $z(1-a(z)) = (1-z)L(z)$, the differential equation (18) can be rewritten in the form (13). For $j > 1$ the only solution of (13) being continuous at 0 (and for $j = 1$ the only solution of (13) with $f_k(0) = 0$) is given by (14). Since $g_1(z) = z/(1-z)$, (14) reduces for $j := k_1 := 1$ to the first equation in (15), in agreement with [11, Lemma 3.1, Eq. (3.3)]). Noting that $g_{(1,1)}(z) = f'_{(0,1)}(z) + f'_{(1,0)}(z) = 2f'_1(z) = 2z/((1-z)^2 L(z))$, the formula for $f'_{(1,1)}(z)$ in (15) follows by choosing $j := 2$ and $k_1 := k_2 := 1$ in (14). □

Corollary 3.2 (Exact formula for $\mathbb{E}(\tau_{n,1}\tau_{n,2})$)
Fix $n \in \{2, 3, \ldots\}$. For the Bolthausen–Sznitman coalescent,

$$\mathbb{E}(\tau_{n,1}\tau_{n,2}) \;=\; \frac{2}{(n-1)!} \sum_{k=1}^{n-1} \frac{2^k - 1}{k^2} s(n-2, k-1), \tag{19}$$

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	0.75	0.75	0.1875
4	0.638889	0.509259	0.101080
5	0.572917	0.397569	0.069336
10	0.431647	0.215119	0.028800
100	0.228368	0.057067	0.004915
$n\to\infty$	$\frac{1}{\log n}+\frac{1-\gamma}{\log^2 n}+O(\frac{1}{\log^3 n})$	$\frac{1}{\log^2 n}+\frac{5/2-2\gamma}{\log^3 n}+O(\frac{1}{\log^4 n})$	$\frac{1}{2\log^3 n}+O(\frac{1}{\log^4 n})$

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

Proof. (of Corollary 3.2) We write $f := f_{(1,1)}$ for convenience. The substitution $u = L(t) = -\log(1-t)$ below the second integral in (15) yields

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

From (see [1, p. 824]) $(L(z))^k/k! = \sum_{i=k}^{\infty} z^i/i!s(i,k)$ we conclude that

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

For a power series $g(z)=\sum_{n=0}^{\infty} g_n z^n$ we denote in the following with $[z^n]g(z) := g_n$ the coefficient in front of z^n in the series expansion of g. Using this notation we obtain

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

It remains to divide by $i+1$ and to substitute $n=i+2$. □

4 Proofs of Theorem 1.2, Corollary 1.3, and Corollary 1.4

For the proof of Theorem 1.2 the following technical lemma is needed. Recall that Ψ denotes the digamma function.

Lemma 4.1 *For all $\alpha \in (0,\infty)$ and all $k \in \mathbb{N}_0$, as $n \to \infty$,*

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

Proof. Let us first verify that

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

We verify (20) via a dissection method. Define the sequence $(a_n)_{n\in\mathbb{N}}$ via $a_1 := 1$ and $a_n := n - \lfloor n/\log^{k+1} n \rfloor$ for $n \geq 2$. Furthermore, put $c_k := 1/\log^k 2$, $p \geq 0$. The sequence $(a_n)_{n\in\mathbb{N}}$ is chosen such that

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

Moreover,

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

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

$$S_1 \;:=\; \sum_{l=1}^{n-1} \frac{\log^k n - \log^k(n-l)}{l} \;=\; O(\log^{k-1} n). \tag{21}$$

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

$$S_1 \;=\; -\sum_{l=1}^{n-1} \frac{1}{l} \sum_{j=1}^{k} \binom{k}{j} \log^{k-j} n \log^j(1-l/n) \;=\; \sum_{j=1}^{k} \binom{k}{j} \log^{k-j} n \sum_{l=1}^{n-1} \frac{-\log^j(1-l/n)}{l}.$$

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

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

(see 23.2.7 of [1]), where ζ denotes the zeta function, (21) holds and (20) is established.
The rest of the proof is now straightforward. Since $0 \leq (n-l)^\alpha \leq n^\alpha$ for all $0 \leq l \leq n$, it follows from (20) that

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

Thus, it suffices to show that $\sum_{l=1}^{n-2}(n-l)^\alpha/l = n^\alpha \log n - \Psi(\alpha+1)n^\alpha + O(n^\alpha/\log n)$. In the following it is even show that $\sum_{l=1}^{n-2}(n-l)^\alpha/l = n^\alpha \log n - \Psi(\alpha+1)n^\alpha + O(n^{\alpha-1})$ or, equivalently, that $\sum_{l=1}^{n-2}(1-l/n)^\alpha/l = \log n - \Psi(\alpha+1) + O(1/n)$. We have

$$\sum_{l=1}^{n-2}\frac{(1-l/n)^\alpha}{l} \;=\; \sum_{l=1}^{n-2}\frac{1}{l} - \sum_{l=1}^{n-2}\frac{1-(1-l/n)^\alpha}{l} \;=\; \log n - \Psi(\alpha+1) + O(1/n),$$

since $\sum_{l=1}^{n-2} 1/l = \log n + \gamma + O(1/n)$ and the last sum is a Riemann sum and hence equals $\int_0^1 (1-(1-x)^\alpha)/x \,\mathrm{d}x + O(1/n) = \Psi(\alpha+1) + \gamma + O(1/n)$. □

Lemma 4.1 can be restated in terms of generating functions as follows.

Corollary 4.2 *Let $\alpha \in (0,\infty)$ and $k \in \mathbb{N}_0$. The effect to the growth of the coefficients when multiplying the generating function $F(z) := \sum_{n=2}^\infty z^n n^\alpha/\log^k n$ with $L(z) := -\log(1-z)$ is*

$$[z^n](L(z)F(z)) \;=\; \frac{n^\alpha}{\log^{k-1} n}\left(1 - \frac{\Psi(\alpha+1)}{\log n} + O\left(\frac{1}{\log^2 n}\right)\right), \qquad n\to\infty. \tag{22}$$

Proof. Clearly, $[z^n](L(z)F(z)) = \sum_{l=1}^{n-2}[z^l]L(z)\,[z^{n-l}]F(z) = \sum_{l=1}^{n-2}(n-l)^\alpha/(l\log^k(n-l))$. Now apply Lemma 4.1. □

Let us now turn to the converse of Corollary 4.2.

Lemma 4.3 *Let $\alpha \in (0,\infty)$ and $k \in \mathbb{N}_0$. The effect on the growth of the coefficients when the generating function $F(z) := \sum_{n=2}^\infty z^n n^\alpha/\log^k n$ is divided by $L(z) := -\log(1-z)$ is*

$$[z^n]\frac{F(z)}{L(z)} \;=\; \frac{n^\alpha}{\log^{k+1} n}\left(1 + \frac{\Psi(\alpha+1)}{\log n} + O\left(\frac{1}{\log^2 n}\right)\right), \qquad n\to\infty. \tag{23}$$

Proof. Note first that

$$[z^n]\frac{F(z)}{L(z)} \;=\; \sum_{l=0}^{n}[z^l]\frac{(1-z)F(z)}{z}\,[z^{n-l}]\frac{z}{(1-z)L(z)}. \tag{24}$$

Using that

$$\begin{aligned}[z^n]\frac{(1-z)F(z)}{z} \;&=\; [z^n]\frac{F(z)}{z} - [z^n]F(z) \;=\; \frac{(n+1)^\alpha}{\log^k(n+1)} - \frac{n^\alpha}{\log^k n}\\ &=\; \frac{\alpha n^{\alpha-1}}{\log^k n} - k\frac{n^{\alpha-1}}{\log^{k+1} n} + O\left(\frac{n^{\alpha-2}}{\log^k n}\right)\end{aligned}$$

and that (see, for example, [21, p. 274, Eq. (19)])

$$[z^n]\frac{z}{(1-z)L(z)} \;=\; \frac{1}{\log n} + \frac{\Psi(1)}{\log^2 n} + O\left(\frac{1}{\log^3 n}\right)$$

it follows from (24) that

$$\begin{aligned}
[z^n]\frac{F(z)}{L(z)} &= \sum_{l=2}^{n-2}\left(\frac{\alpha l^{\alpha-1}}{\log^k l} - k\frac{l^{\alpha-1}}{\log^{k+1} l} + O\left(\frac{l^{\alpha-2}}{\log^k l}\right)\right)\times \\
&\qquad\qquad \times\left(\frac{1}{\log(n-l)} + \frac{\Psi(1)}{\log^2(n-l)} + O\left(\frac{1}{\log^3(n-l)}\right)\right) \\
&= \sum_{l=2}^{n-2}\frac{\alpha l^{\alpha-1}}{\log^k l\log(n-l)} + \Psi(1)\sum_{l=2}^{n-2}\frac{\alpha l^{\alpha-1}}{\log^k l\log^2(n-l)} \\
&\quad -k\sum_{l=2}^{n-2}\frac{l^{\alpha-1}}{\log^{k+1} l\log(n-l)} - k\Psi(1)\sum_{l=2}^{n-2}\frac{l^{\alpha-1}}{\log^{k+1} l\log^2(n-l)} + O\left(\frac{n^\alpha}{\log^{k+3} n}\right).
\end{aligned}$$

Applying the formula

$$\sum_{l=2}^{n-2}\frac{\alpha l^{\alpha-1}}{\log^p l\log^q(n-l)} = \frac{n^\alpha}{\log^{p+q} n}\left(1 - \frac{p\Psi(\alpha)+q\Psi(1)-(p+q)\Psi(\alpha+1)}{\log n} + O\left(\frac{1}{\log^2 n}\right)\right),$$

$\alpha\in(0,\infty)$, $p,q\in[0,\infty)$, which is a particular case of Panholzer's summation formula [21, p. 273, Eq. (16)], to all the four sums above yields

$$\begin{aligned}
[z^n]\frac{F(z)}{L(z)} &= \frac{n^\alpha}{\log^{k+1} n}\left(1 - \frac{k\Psi(\alpha)+\Psi(1)-(k+1)\Psi(\alpha+1)}{\log n} + O\left(\frac{1}{\log^2 n}\right)\right) \\
&\quad +\Psi(1)\frac{n^\alpha}{\log^{k+2} n}\left(1+O\left(\frac{1}{\log n}\right)\right) - \frac{k}{\alpha}\frac{n^\alpha}{\log^{k+2} n}\left(1+O\left(\frac{1}{\log n}\right)\right) \\
&= \frac{n^\alpha}{\log^{k+1} n} + \left((k+1)\Psi(\alpha+1) - k\Psi(\alpha) - \frac{k}{\alpha}\right)\frac{n^\alpha}{\log^{k+2} n} + O\left(\frac{n^\alpha}{\log^{k+3} n}\right),
\end{aligned}$$

which is (23), since $(k+1)\Psi(\alpha+1) - k\Psi(\alpha) - k/\alpha = \Psi(\alpha+1)$. □

In the following we will furthermore use the trivial effect to the growth of the coefficients when differentiating and integrating $F(z) := \sum_{n=2}^\infty z^n n^\alpha/\log^k n$, $\alpha\in[0,\infty)$, $k\in\mathbb{N}_0$:

$$[z^n]\frac{\mathrm{d}}{\mathrm{d}z}F(z) = \frac{(n+1)^{\alpha+1}}{\log^k(n+1)} = \frac{n^{\alpha+1}}{\log^k n}\left(1+O\left(\frac{1}{n}\right)\right) = n[z^n]F(z)\left(1+O\left(\frac{1}{n}\right)\right) \tag{25}$$

and

$$[z^n]\int_0^z F(t)\,\mathrm{d}t = \frac{(n-1)^\alpha}{n\log^k(n-1)} = \frac{n^{\alpha-1}}{\log^k n}\left(1+O\left(\frac{1}{n}\right)\right) = \frac{1}{n}[z^n]F(z)\left(1+O\left(\frac{1}{n}\right)\right). \tag{26}$$

We are now able to provide a proof of Theorem 1.2 based on generating functions.

Proof. (of Theorem 1.2) Let us verify (6) by induction on the degree $d := k_1+\cdots+k_j$. Obviously (6) holds for $d=0$, i.e. for all $j\in\mathbb{N}$ and $k_1=\cdots=k_j=0$. It also holds for $d=1$, since (see, for example, [21, p. 274, Eq. (19)])

$$[t^n]\frac{t}{(1-t)^2L(t)} = \frac{n}{\log n}\left(1+\frac{\Psi(2)}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right) \tag{27}$$

and, therefore, for all $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$ satisfying $k_1 + \cdots + k_j = 1$,

$$\mu_n(k) \;=\; \mu_n(1) \;=\; [z^n] f_1(z) \;=\; [z^n] \int_0^z \frac{t}{(1-t)^2 \log(1-t)}\,\mathrm{d}t \;=\; \frac{1}{\log n}\biggl(1 + \frac{\Psi(2)}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr),$$

where the last equality follows by applying the integration rule (26) to (27). In order to verify the induction step from $d-1$ to $d > 1$ fix $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$ with $d := k_1 + \cdots + k_j > 1$. Due to the definition of the coefficients $\kappa_j(k)$ in Theorem 1.2 for the case when $k_i = 0$ for some i we can and do assume without loss of generality that $k = (k_1, \dots, k_j) \in \mathbb{N}^j$. Then, by (12), $g_k(t) = \sum_{i=1}^j k_i f_{k-e_i}^{(j-1)}(t)$ and furthermore, by induction,

$$[t^n] f_{k-e_i}(t) \;=\; \frac{k_1! \cdots k_j!}{\log^{d-1} n}\biggl(1 + \frac{\kappa_j(k-e_i)}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr), \qquad i \in \{1, \dots, j\}.$$

Applying $(j-1)$-times the differentiation rule (25) yields

$$[t^n] f_{k-e_i}^{(j-1)}(t) \;=\; \frac{k_1! \cdots k_j! n^{j-1}}{\log^{d-1} n}\biggl(1 + \frac{\kappa_j(k-e_i)}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr), \qquad i \in \{1, \dots, j\}$$

and, hence,

$$\begin{aligned}
[t^n] g_k(t) \;&=\; \sum_{i=1}^j k_i [t^n] f_{k-e_i}^{(j-1)}(t) \;=\; \sum_{i=1}^j \frac{k_1! \cdots k_j! n^{j-1}}{\log^{d-1} n}\biggl(1 + \frac{\kappa_j(k-e_i)}{\log n} + \biggl(\frac{1}{\log^2 n}\biggr)\biggr) \\
&=\; k_1! \cdots k_j! \frac{j n^{j-1}}{\log^{d-1} n} + k_1! \cdots k_j! S \frac{n^{j-1}}{\log^d n} + O\biggl(\frac{n^{j-1}}{\log^{d+1} n}\biggr),
\end{aligned}$$

where the abbreviation $S := \sum_{i=1}^j \kappa_j(k-e_i)$ is used for convenience. Consequently,

$$\begin{aligned}
[t^n] \frac{g_k(t)}{1-t} \;&=\; \sum_{l=0}^n [t^l] g_k(t) [t^{n-l}] \frac{1}{1-t} \;=\; \sum_{l=0}^n [t^l] g_k(t) \\
&=\; k_1! \cdots k_j! \sum_{l=2}^{n-2} \frac{j l^{j-1}}{\log^{d-1} l} + k_1! \cdots k_j! S \sum_{l=2}^{n-2} \frac{l^{j-1}}{\log^d l} + O\biggl(\frac{n^j}{\log^{d+1} n}\biggr) \\
&=\; \frac{k_1! \cdots k_j! n^j}{\log^{d-1} n}\biggl(1 + \frac{C_1}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr)
\end{aligned}$$

with $C_1 := (d - 1 + S)/j$, where we have applied the summation formula

$$\sum_{l=2}^{n-2} \frac{l^{j-1}}{\log^q l} \;=\; \frac{n^j}{j \log^q n}\biggl(1 + \frac{q/j}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr), \quad j \in \mathbb{N}, q \geq 0,$$

which is a particular case of Panholzer's summation formula [21, p. 273, Eq. (16)]. Applying (23) for $j = 1$ and, for $j > 1$, (22) $(j-2)$-times gives

$$[t^n] \frac{(L(t))^{j-2} g_k(t)}{1-t} \;=\; \frac{k_1! \cdots k_j! n^j}{\log^{d-j+1} n}\biggl(1 + \frac{C_2}{\log n} + O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr),$$

where $C_2 := C_1 - (j-2)\Psi(j+1)$. The integration rule (26) yields

$$[z^n]\int_0^z \frac{(L(t))^{j-2}}{1-t} g_k(t)\,\mathrm{d}t \;=\; \frac{k_1!\cdots k_j! n^{j-1}}{\log^{d-j+1} n}\biggl(1+\frac{C_2}{\log n}+O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr).$$

Applying (23) $(j-1)$-times and recalling (14) we obtain

$$[z^n]f_k^{(j-1)}(z) \;=\; [z^n]\frac{1}{(L(z))^{j-1}}\int_0^z \frac{(L(t))^{j-2}}{1-t} g_k(t)\,\mathrm{d}t \;=\; \frac{k_1!\cdots k_j! n^{j-1}}{\log^{d} n}\biggl(1+\frac{C_3}{\log n}+O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr)$$

with $C_3 := (j-1)\Psi(j) + C_2$. Finally, applying $(j-1)$-times the integration rule (26) yields

$$\mu_n(k) \;=\; [z^n]f_k(z) \;=\; \frac{k_1!\cdots k_j!}{\log^{d} n}\biggl(1+\frac{C_3}{\log n}+O\Bigl(\frac{1}{\log^2 n}\Bigr)\biggr).$$

The induction step is complete, if we can show that the coefficient C_3 coincides with $\kappa_j(k)$. But this is clear, since

$$\begin{aligned}
C_3 &= (j-1)\Psi(j) - (j-2)\Psi(j+1) + C_1\\
&= (j-1)\biggl(\Psi(j+1) - \frac{1}{j}\biggr) - (j-2)\Psi(j+1) + C_1\\
&= \Psi(j+1) - 1 + \frac{1}{j} + C_1 \;=\; \Psi(j+1) - 1 + \frac{1}{j} + \frac{d-1+S}{j}\\
&= \Psi(j+1) - 1 + \frac{d+S}{j} \;=\; \kappa_j(k)
\end{aligned}$$

by the recursive definition of $\kappa_j(k)$ given in Theorem 1.2. □

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 $(\log n)(\tau_{n,1},\ldots,\tau_{n,j}) \to (\tau_1,\ldots,\tau_j)$ 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 $(\log n)(\tau_{n,1},\ldots,\tau_{n,n},0,0,\ldots) \to (\tau_1,\tau_2,\ldots)$ in distribution as $n\to\infty$. □

Proof. (of Corollary 1.4) The external branch length $L_n^{external}$ satisfies (see [20, p. 2165])

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

By Theorem 1.2 it follows that

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

This representation shows that asymptotically the summand with index $j = k$ dominates the others, so asymptotically all the summands with indices $j < k$ can be disregarded. Thus, we obtain

$$\mathbb{E}((L_n^{external})^k) \;=\; \binom{n}{k}\frac{k!}{\log^k n}\left(1+\frac{c_k}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right) \;=\; \frac{n^k}{\log^k n}\left(1+\frac{c_k}{\log n}+O\left(\frac{1}{\log^2 n}\right)\right),$$

where $c_k := \kappa_k(1,\dots,1) = \sum_{i=2}^{k+1}\Psi(i) = (k+1)h_k - k - k\gamma$ by the first remark after Theorem 1.2. □

We finally provide a simple argument which shows that if the conjectured asymptotics (10) for the variance of $L_n^{internal}$ holds, then Conjecture 1.5 must be correct.

Proof. (that (10) implies the correctness of Conjecture 1.5) Suppose that (10) is true, i.e. $\mathrm{Var}(L_n^{internal}) = o(n^2/\log^4 n)$. Since $\mathbb{E}(L_n^{internal}) \sim n/\log^2 n$, an application of the Tschebyscheff inequality yields $L_n^{internal}/\mathbb{E}(L_n^{internal}) \to 1$ in probability or, equivalently,

$$\frac{\log^2 n}{n} L_n^{internal} \;\to\; 1 \tag{28}$$

in probability as $n \to \infty$. By [9, Theorem 5.2],

$$\frac{\log^2 n}{n} L_n - \log n - \log\log n \;\to\; L \tag{29}$$

in distribution as $n \to \infty$, where L is a 1-stable random variable with characteristic function $t \mapsto \exp(it\log|t| - \pi|t|/2)$, $t \in \mathbb{R}$. Subtracting (28) from (29) and noting that $L_n - L_n^{internal} = L_n^{external}$ yields the result by an application of Slutzky's lemma. □

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. The authors are indebted to an anonymous referee for his profound suggestions leading to a significant improvement of the article.

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. (2013) On the length of an external branch in the beta-coalescent. *Stoch. Process. Appl.*, to appear

[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] 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

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

[14] 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

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

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

[17] Kersting, G., Stanciu, I., and Wakolbinger, A. (2012) The total external branch length of beta-coalescents. Preprint.

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

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

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

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

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

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