



**HAL**  
open science

# On the external branches of coalescents with multiple collisions

Jean-Stephane Dhersin, Martin Moehle

► **To cite this version:**

Jean-Stephane Dhersin, Martin Moehle. On the external branches of coalescents with multiple collisions . 2012. hal-00732541v3

**HAL Id: hal-00732541**

**<https://hal.science/hal-00732541v3>**

Preprint submitted on 24 Feb 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ON THE EXTERNAL BRANCHES OF COALESCENTS WITH MULTIPLE COLLISIONS

J.-S. DHERSIN<sup>1</sup> and M. MÖHLE<sup>2</sup>

February 24, 2013

## Abstract

A recursion for the joint moments of the external branch lengths for coalescents with multiple collisions ( $\Lambda$ -coalescents) is provided. This recursion is used to derive asymptotic results 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 asymptotic results 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 ( $\Lambda$ -coalescent). For fundamental information on  $\Lambda$ -coalescents we refer the reader to [23] and [24]. For  $n \in \mathbb{N} := \{1, 2, \dots\}$  we denote with  $\Pi^{(n)} = (\Pi_t^{(n)})_{t \geq 0}$  the coalescent process restricted to  $[n] := \{1, \dots, n\}$ . 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  $\xi$ . For  $m \in \{1, \dots, n-1\}$  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, [21, Eq. (13)]) that

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

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

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

We are interested in the external branches of the restricted coalescent process  $\Pi^{(n)}$ . More precisely, for  $n \in \mathbb{N}$  and  $i \in \{1, \dots, n\}$  let  $\tau_{n,i} := \inf\{t > 0 : \{i\} \text{ is a singleton block of } \Pi_t^{(n)}\}$  denote the length of the  $i$ th 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

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

<sup>1</sup>Department de Mathématique, Institut Galilée, Université Paris 13, Avenue Jean Baptiste Clément, 93430 Villetaneuse, France, E-mail: dherstin@math.univ-paris13.fr

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

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

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

where  $e_i$ ,  $i \in \{1, \dots, j\}$ , denotes the  $i$ th 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  $d := k_1 + \dots + k_j$  the order (or degree) of the moment  $\mu_n(k_1, \dots, k_j)$ . 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

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

Note that Theorem 1.1 holds for arbitrary  $\Lambda$ -coalescents. For particular  $\Lambda$ -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 [8], [9] and [19].

**Example.** (Star-shaped coalescent) For the star-shaped coalescent, where  $\Lambda$  is the Dirac measure at 1, the time  $T_n$  of the first jump of  $\Pi^{(n)}$  is exponentially distributed with parameter  $g_n = 1$ ,  $n \in \{2, 3, \dots\}$ . 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  $\mu_n(k) = (k_1 + \dots + k_j)!$ , which is obviously correct, since  $\tau_{n,i} = T_n$  for all  $i \in \{1, \dots, n\}$  and, therefore,  $\mu_n(k) = \mathbb{E}(T_n^{k_1 + \dots + k_j}) = (k_1 + \dots + k_j)!$ ,  $n \geq 2$ ,  $j \in \{1, \dots, n\}$ ,  $k_1, \dots, k_j \in \mathbb{N}$ .

**Example.** (Kingman coalescent) For the Kingman coalescent [20], where  $\Lambda$  is the Dirac measure at 0, the time  $T_n$  of the first jump of  $\Pi^{(n)}$  is exponentially distributed with parameter  $g_n = n(n-1)/2$ ,  $n \in \{2, 3, \dots\}$ . Furthermore,  $p_{nm} = \delta_{m,n-1}$  for  $m, n \in \mathbb{N}$  with  $m < n$ . Caliebe et al. [6, Theorem 1] verified that  $n\tau_{n,1} \rightarrow Z$  in distribution as  $n \rightarrow \infty$ , where  $Z$  has density  $x \mapsto 8/(2+x)^3$ ,  $x \geq 0$ . Janson and Kersting [17, Theorem 1] showed that the total external branch length  $L_n^{external} := \sum_{i=1}^n \tau_{n,i}$  satisfies  $(1/2)\sqrt{n/(\log n)}(L_n^{external} - 2) \rightarrow N(0, 1)$  in distribution as  $n \rightarrow \infty$ . We are instead interested here in the moments of  $\tau_{n,1}$ . The recursion (4) for  $j = 1$  reduces to

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

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

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

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. [6, Eq. (2)] and Janson and Kersting [17, 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

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

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,  $g_2(t) = 8t - 4(1-t)\log^2(1-t) - 8(1-t)\text{Li}_2(t)$ ,  $|t| < 1$ , where  $\text{Li}_2(t) := -\int_0^t (\log(1-x))/x dx = \sum_{k=1}^{\infty} t^k/k^2$  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 exemplarily 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))

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

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

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

In particular,  $\mu_n(1,1) = 4/n^2 - 4/n^3 + O((\log n)/n^4)$ ,  $n \rightarrow \infty$ . Moreover,  $\text{Cov}(\tau_{n,1}, \tau_{n,2}) = \mu_n(1,1) - (\mu_n(1))^2 = 4(n^2 - 5n + 4h_n)/(n(n-1)^2(n-2)) - 4/n^2 < 0$  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.

$n$	$\mu_n(1) = \mathbb{E}(\tau_{n,1})$	$\mu_n(1,1) = \mathbb{E}(\tau_{n,1}\tau_{n,2})$	$\text{Cov}(\tau_{n,1}, \tau_{n,2})$
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
$n \rightarrow \infty$	$\frac{2}{n}$	$\frac{4}{n^2} - \frac{4}{n^3} + O(\frac{\log n}{n^4})$	$-\frac{4}{n^3} + O(\frac{\log n}{n^4})$

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 [5], where  $\Lambda$  is the uniform distribution on  $[0, 1]$ . Our second main result (Theorem 1.2) provides the asymptotics of all the joint moments of the external branch lengths for the Bolthausen–Sznitman coalescent.

**Theorem 1.2 (Asymptotics of the joint moments of the external branch lengths)**

For the Bolthausen–Sznitman coalescent, the joint moments  $\mu_n(k) := \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j})$ ,  $j \in \mathbb{N}$ ,  $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$ , of the lengths  $\tau_{n,1}, \dots, \tau_{n,n}$  of the external branches satisfy

$$\mu_n(k) \sim \frac{k_1! \cdots k_j!}{\log^{k_1 + \cdots + k_j} n}, \quad n \rightarrow \infty. \quad (6)$$

**Remark.** For  $j = 2$  and  $k_1 = k_2 = 1$  Eq. (6) implies that  $\mathbb{E}(\tau_{n,1}\tau_{n,2}) = \mu_n(1,1) \sim 1/\log^2 n \sim (\mu_n(1))^2$  as  $n \rightarrow \infty$ , which does not provide much information on the covariance  $\text{Cov}(\tau_{n,1}, \tau_{n,2}) = \mu_n(1,1) - (\mu_n(1))^2$ . 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,  $(\log n)(\tau_{n,1}, \dots, \tau_{n,n}, 0, 0, \dots) \rightarrow (\tau_1, \tau_2, \dots)$  in distribution as  $n \rightarrow \infty$ , where  $\tau_1, \tau_2, \dots$  are independent and all exponentially distributed with parameter 1.

The following result concerns the asymptotics 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 total external branch length)**

Fix  $k \in \mathbb{N}$ . For the Bolthausen–Sznitman coalescent, the  $k$ th moment of  $L_n^{external}$  satisfies

$$\mathbb{E}((L_n^{external})^k) \sim \frac{n^k}{\log^k n}, \quad n \rightarrow \infty. \tag{7}$$

In particular,  $\frac{\log n}{n} L_n^{external} \rightarrow 1$  in probability as  $n \rightarrow \infty$ .

The moments of  $L_n^{external}$  do not provide much information on the distributional limiting behavior of  $L_n^{external}$  as  $n \rightarrow \infty$ . Let  $L_n$  denote the total branch length (the sum of the lengths of all branches) of the Bolthausen–Sznitman  $n$ -coalescent. Kersting et al. [18, Theorem 1.1] recently showed that the internal branch length  $L_n^{internal} := L_n - L_n^{external}$  satisfies

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

in probability. Combining this result with [10, Theorem 5.2] it follows that (see [18, Corollary 1.2])

$$\frac{\log^2 n}{n} L_n^{external} - \log n - \log \log n \rightarrow L - 1 \tag{8}$$

in distribution as  $n \rightarrow \infty$ , 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 (8), the same limiting law as in (8) is known for the number of cuts needed to isolate the root of a random recursive tree ([11], [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 [7]. 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 distributional limiting behavior of  $L_n^{internal}$ , properly centered and scaled, is so far unknown for the Bolthausen–Sznitman coalescent.

## 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  $h \rightarrow 0$ ,

$$\begin{aligned} \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j} 1_{\{T > h\}}) &= \mathbb{E}((\tau'_1 + h)^{k_1} \cdots (\tau'_j + h)^{k_j} 1_{\{T > h\}}) \\ &= \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j}) \mathbb{P}(T > h) + h \sum_{i=1}^j k_i \mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,i-1}^{k_{i-1}} \tau_{n,i}^{k_i-1} \tau_{n,i+1}^{k_{i+1}} \cdots \tau_{n,j}^{k_j}) + o(h). \end{aligned}$$

Also for  $h \rightarrow 0$ ,

$$\mathbb{E}(\tau_{n,1}^{k_1} \cdots \tau_{n,j}^{k_j} 1_{\{T \leq h\}}) = \mathbb{E}((\tau'_1 + T)^{k_1} \cdots (\tau'_j + T)^{k_j} 1_{\{T \leq h\}}) = \mathbb{E}((\tau'_1)^{k_1} \cdots (\tau'_j)^{k_j} 1_{\{T \leq h\}}) + o(h).$$

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,

$$\mathbb{E}((\tau'_1)^{k_1} \cdots (\tau'_j)^{k_j} 1_{\{T \leq h, I=m, A^c\}}) = \mathbb{E}(\tau_{m,1}^{k_1} \cdots \tau_{m,j}^{k_j}) \mathbb{P}(T \leq h, I = m, A^c),$$

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

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

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

### 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  $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$  define the generating function

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

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  $n$ -coascent, the sequence  $(a_n)_{n \geq j}$  is non-increasing. Thus,  $f_k$  and all its derivatives  $f'_k, f''_k, \dots$  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  $g_k : D \rightarrow \mathbb{C}$ ,  $k = (k_1, \dots, k_j) \in \mathbb{N}^j$ , via  $g_1(z) := z/(1-z)$  and

$$g_k(z) := \sum_{i=1}^j k_i f_{k-e_i}^{(j-1)}(z) \tag{9}$$

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

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

$$\frac{d}{dz}((L(z))^{j-1} f_k^{(j-1)}(z)) = \frac{(L(z))^{j-2}}{1-z} g_k(z), \quad z \in D \setminus \{0\}, \quad (10)$$

with solution

$$f_k^{(j-1)}(z) = \frac{1}{(L(z))^{j-1}} \int_0^z \frac{(L(t))^{j-2}}{1-t} g_k(t) dt, \quad z \in D \setminus \{0\}. \quad (11)$$

In particular,

$$f_1(z) = \int_0^z \frac{t}{(1-t)^2 L(t)} dt \quad \text{and} \quad f'_{(1,1)}(z) = \frac{2}{L(z)} \int_0^z \frac{t}{(1-t)^3 L(t)} dt, \quad z \in D \setminus \{0\}. \quad (12)$$

**Proof.** For the Bolthausen–Sznitman coalescent,  $g_{nm} = n/((n-m)(n-m+1))$ ,  $m \in \{1, \dots, n-1\}$  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  $k = (k_1, \dots, k_j) \in \mathbb{N}^j$  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,

$$(n-1)(n-1)_{j-1} a_n = (n-1)(n-1)_{j-1} q_n + \sum_{m=j+1}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)} a_m. \quad (13)$$

Before we come back to the recursion (13) let us first verify that

$$\sum_{n=\max(2,j)}^{\infty} (n-1)(n-1)_{j-1} q_n z^{n-j} = g_k(z), \quad z \in D. \quad (14)$$

Obviously (14) 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  $k = (k_1, \dots, k_j) \in \mathbb{N}^j$  with  $k_1 + \dots + k_j > 1$  we have

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



Thus, (14) is established. In view of  $(n-1)(n-1)_{j-1} = (n-1)_j + (j-1)(n-1)_{j-1}$  and (14), by multiplying both sides in (13) with  $z^{n-j}$  and summing over all  $n \geq \max(2, j)$ , the recursion (13) translates to

$$\begin{aligned}
zf_k^{(j)}(z) + (j-1)f_k^{(j-1)}(z) &= g_k(z) + \sum_{n=\max(2,j)}^{\infty} \sum_{m=j+1}^{n-1} \frac{(m-1)_j}{(n-m)(n-m+1)} a_m z^{n-j} \\
&= g_k(z) + \sum_{m=j+1}^{\infty} (m-1)_j a_m z^{m-j} \sum_{n=m+1}^{\infty} \frac{1}{(n-m)(n-m+1)} z^{n-m} \\
&= g_k(z) + za(z) \left( \frac{d}{dz} \right)^j \sum_{m=j}^{\infty} a_m z^{m-1} \\
&= g_k(z) + za(z) f_k^{(j)}(z), \tag{15}
\end{aligned}$$

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 (15) can be rewritten in the form (10). For  $j > 1$  the only solution of (10) being continuous at 0 (and for  $j = 1$  the only solution of (10) with  $f_k(0) = 0$ ) is given by (11). Since  $g_1(z) = z/(1-z)$ , (11) reduces for  $j := k_1 := 1$  to the first equation in (12), in agreement with [12, 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 (12) follows by choosing  $j := 2$  and  $k_1 := k_2 := 1$  in (11).  $\square$

**Corollary 3.2 (Exact formula for  $\mathbb{E}(\tau_{n,1}\tau_{n,2})$ )**

Fix  $n \in \{2, 3, \dots\}$ . 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 [12]) it can be checked that  $\text{Cov}(\tau_{n,1}, \tau_{n,2}) = \mathbb{E}(\tau_{n,1}\tau_{n,2}) - (\mathbb{E}(\tau_{n,1}))^2 > 0$  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.

$n$	$\mathbb{E}(\tau_{n,1})$	$\mathbb{E}(\tau_{n,1}\tau_{n,2})$	$\text{Cov}(\tau_{n,1}, \tau_{n,2})$
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

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 (12) yields

$$f'(z) = \frac{2}{L(z)} \int_0^{L(z)} \frac{e^{2u} - e^u}{u} du$$

$$\begin{aligned}
&= \frac{2}{L(z)} \int_0^{L(z)} \frac{1}{u} \left( \sum_{k=0}^{\infty} \frac{(2u)^k}{k!} - \sum_{k=0}^{\infty} \frac{u^k}{k!} \right) du \\
&= \frac{2}{L(z)} \sum_{k=1}^{\infty} \frac{2^k - 1}{k!} \int_0^{L(z)} u^{k-1} du \\
&= \frac{2}{L(z)} \sum_{k=1}^{\infty} \frac{2^k - 1}{k!} \frac{(L(z))^k}{k} \\
&= 2 \sum_{k=1}^{\infty} \frac{2^k - 1}{kk!} (L(z))^{k-1}.
\end{aligned}$$

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

**Proof.** (of Theorem 1.2) Let us verify (6) by induction on the degree  $d := k_1 + \dots + k_j$ . Obviously (6) holds for  $d = 0$ , i.e. for all  $j \in \mathbb{N}$  and  $k_1 = \dots = k_j = 0$ . In order to verify (6) for  $d = 1$  it suffices to show that  $a_n := \mu_n(1) \sim 1/\log n$  as  $n \rightarrow \infty$  since  $\mu_n(k) = \mu_n(1)$  for all  $k = (k_1, \dots, k_j) \in \mathbb{N}_0^j$  satisfying  $k_1 + \dots + k_j = 1$ . By (12) and de l'Hospital's rule

$$f_1(z) = \int_0^z \frac{t}{(1-t)^2 L(t)} dt \sim \frac{1}{(1-z)L(z)}, \quad z \nearrow 1.$$

Since  $a_n = \mathbb{E}(\tau_{n,1})$  is non-increasing in  $n$ , Karamata's Tauberian theorem for power series [4, p. 40, Corollary 1.7.3], applied with  $c := \rho := 1$  and  $l(x) := 1/\log x$ , yields  $a_n \sim l(n) = 1/\log n$ . Thus, (6) holds for  $d = 1$ .

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  $d := k_1 + \dots + k_j > 1$ . We can and do assume without loss of generality that  $k = (k_1, \dots, k_j) \in \mathbb{N}^j$ . By the induction hypothesis

$$b_n := \sum_{i=1}^j k_i \mu_n(k - e_i) \sim \sum_{i=1}^j k_i \frac{k_1! \dots (k_i - 1)! \dots k_j!}{\log^{d-1} n} \sim \frac{jk_1! \dots k_j!}{\log^{d-1} n}, \quad n \rightarrow \infty.$$

Since  $b_n$  is non-increasing in  $n$ , the same Tauberian theorem as used above for  $d = 1$ , but now applied with  $c := jk_1! \cdots k_j!$ ,  $\rho := 1$  and  $l(x) := 1/\log^{d-1} x$ , yields

$$b(z) := \sum_{n=\max(2,j)}^{\infty} b_n z^{n-1} \sim \frac{jk_1! \cdots k_j!}{(1-z)(L(z))^{d-1}}, \quad z \nearrow 1.$$

Note that  $b(z) = \sum_{i=1}^j k_i f_{k-e_i}(z)$ . Applying de l'Hospital's rule  $(j-1)$ -times yields

$$g_k(z) = \sum_{i=1}^j k_i f_{k-e_i}^{(j-1)}(z) = b^{(j-1)}(z) \sim \frac{j!k_1! \cdots k_j!}{(1-z)^j (L(z))^{d-1}}, \quad z \nearrow 1.$$

Thus, by (11) and by one further application of de l'Hospital's rule,

$$f_k^{(j-1)}(z) = \frac{1}{(L(z))^{j-1}} \int_0^z \frac{(L(t))^{j-2}}{1-t} g_k(t) dt \sim \frac{(j-1)!k_1! \cdots k_j!}{(1-z)^j (L(z))^d}, \quad z \nearrow 1.$$

Using again de l'Hospital's rule  $(j-1)$ -times it follows that

$$f_k(z) \sim \frac{k_1! \cdots k_j!}{(1-z)(L(z))^d}, \quad z \nearrow 1.$$

Since  $a_n := \mu_n(k)$  is non-increasing in  $n$ , again Karamata's Tauberian theorem for power series, now applied with  $c := k_1! \cdots k_j!$ ,  $\rho := 1$  and  $l(x) := 1/\log^d x$ , yields  $a_n \sim k_1! \cdots k_j! / \log^d n$ .  $\square$

**Proof.** (of Corollary 1.3) Theorem 1.2 clearly implies that, for  $j \in \mathbb{N}$  and  $k_1, \dots, k_j \in \mathbb{N}_0$ ,  $\mathbb{E}((\tau_{n,1} \log n)^{k_1} \cdots (\tau_{n,j} \log n)^{k_j}) = (\log n)^{k_1 + \cdots + k_j} \mu_n(k_1, \dots, k_j) \rightarrow k_1! \cdots k_j! = \mathbb{E}(\tau_1^{k_1} \cdots \tau_j^{k_j})$  as  $n \rightarrow \infty$ . For all  $i \in \{1, \dots, j\}$  and all  $0 \leq \theta < 1$  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  $(\log n)(\tau_{n,1}, \dots, \tau_{n,j}) \rightarrow (\tau_1, \dots, \tau_j)$  in distribution as  $n \rightarrow \infty$  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  $(\log n)(\tau_{n,1}, \dots, \tau_{n,n}, 0, 0, \dots) \rightarrow (\tau_1, \tau_2, \dots)$  in distribution as  $n \rightarrow \infty$ .  $\square$

**Proof.** (of Corollary 1.4) The external branch length  $L_n^{external}$  satisfies (see [22, p. 2165])

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

By Theorem 1.2,  $\mu_n(k_1, \dots, k_j) \sim k_1! \cdots k_j! / \log^k n$  as  $n \rightarrow \infty$ . Therefore, asymptotically the summand with index  $j = k$  dominates the others, so asymptotically all the summands with indices  $j < k$  can be disregarded. Thus,  $\mathbb{E}((L_n^{external})^k) \sim \binom{n}{k} k! / \log^k n \sim n^k / \log^k n$ . This convergence of all moments  $\mathbb{E}((\frac{\log n}{n} L_n^{external})^k) \rightarrow 1$  as  $n \rightarrow \infty$  implies the convergence  $\frac{\log n}{n} L_n^{external} \rightarrow 1$  in distribution (and hence in probability) as  $n \rightarrow \infty$ .  $\square$

**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, Géométrie 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] BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. (1987) *Regular Variation*. Cambridge University Press, Cambridge. MR0898871
- [5] BOLTHAUSEN, E. AND SZNITMAN, A.-S. (1998) On Ruelle’s probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197**, 247–276. MR1652734
- [6] 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
- [7] DEVROYE, L. (1999) Universal limit laws for depths in random trees. *SIAM J. Comput.* **28**, 409–432. MR1634354
- [8] 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.* **123**, 1691–1715. MR number not yet available
- [9] DHERSIN, J.-S. AND YUAN, L. (2012) On the total length of external branches for beta coalescents. Preprint.
- [10] 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
- [11] 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
- [12] FREUND, F. AND MÖHLE, M. (2009) On the time back to the most recent common ancestor and the external branch length of the Bolthausen–Sznitman coalescent. *Markov Processes Relat. Fields* **15**, 387–416. MR2554368

- [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] JANSON, S. AND KERSTING, G. (2011) On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16**, 2203–2218. MR2861672
- [18] KERSTING, G., PARDO, J. C., AND SIRI-JÉGOUSSE, A. (2013) Total internal and external lengths of the Bolthausen–Sznitman coalescent. Preprint.
- [19] KERSTING, G., STANCIU, I., AND WAKOLBINGER, A. (2012) The total external branch length of beta-coalescents. Preprint.
- [20] KINGMAN, J. F. C. (1982) The coalescent. *Stoch. Process. Appl.* **13**, 235–248. MR0671034
- [21] MÖHLE, M. (2006) On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, 35–53. MR2202319
- [22] 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
- [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