The internal branch lengths of the Kingman coalescent

Götz Kersting† Iulia Stanciu†

February 6, 2014

Abstract

In the Kingman coalescent tree the length of order $r$ is defined as the sum of the lengths of all branches that support $r$ leaves. For $r = 1$ these branches are external, while for $r \geq 2$ they are internal and carry a subtree with $r$ leaves. In this paper we prove that for any $s \in \mathbb{N}$ the vector of rescaled lengths of orders $1 \leq r \leq s$ converges to the multivariate standard normal distribution as the number of leaves of the Kingman coalescent tends to infinity. To this end we use a coupling argument which shows that for any $r \geq 2$ the (internal) length of order $r$ behaves asymptotically in the same way as the length of order 1 (that is the external length).

MSC 2000 subject classifications. 60K35, 60F05, 60J10

Key words and phrases. coalescent, internal branch length, asymptotic distribution, coupling, Markov chain

—Work partially supported by the DFG Priority Programme SPP 1590 “Probabilistic Structures in Evolution”. I. S. is supported by the German Academic Exchange Service (DAAD).

†Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany.

kersting@math.uni-frankfurt.de, stanciu@math.uni-frankfurt.de
1 Introduction and main result

The Kingman coalescent was introduced in [Ki82] as a model for describing the genealogical relationships between the individuals for a wide class of population models (see [Wa08] for details). The state space of the Kingman $n$-coalescent, $n \in \mathbb{N}$, is the set $P_n$ of partitions of the set $\{1, 2, \ldots, n\}$. The process starts in the partition into singletons $\pi_n = \{\{1\}, \ldots, \{n\}\}$ and has the following dynamics: given that the process is in the state $\pi_k$, it jumps after a random time $X_k$ to a state $\pi_{k-1}$ which is obtained by merging two randomly chosen elements from $\pi_k$. The random inter-coalescence times $X_k$ are independent, exponentially distributed random variables with parameters $(k/2)$. The process can be viewed graphically as a rooted tree that starts from $n$ leaves labelled from 1 to $n$ and whose any two branches coalesce independently at rate 1. Each branch of this tree is situated above a subtree. If this subtree has $r$ leaves, we say that the branch is of order $r$. The branches of order $r \geq 2$ are the internal branches, while those of order 1 are the external ones (they support subtrees consisting of just one node).

![Diagram](https://example.com/diagram.png)

Figure 1: The dashed (red) branch is an internal branch of order 4, it supports the leaves 1, 2, 3 and 4. It is formed at level $\sigma(1, \ldots, 4) = 5$ and ends at level $\rho(1, \ldots, 4) = 3$. Its length is $S_{1,2,3,4} = X_4 + X_5$. The dotted (green) branches are the branches of order three. The numbers of branches of orders 1 to 10 at level 5 are $W_5(1) = 3$, $W_5(2) = 0$, $W_5(3) = 1$, $W_5(4) = 1$ and $W_5(i) = 0$ for $i \geq 5$.

Let us look at the tree from the leaves towards the root. Then the branch of order $r$ supporting the leaves $i_1, \ldots, i_r$ is formed at the level $\sigma(i_1, \ldots, i_r)$ and ends at level $\rho(i_1, \ldots, i_r)$, where

$$\sigma(i_1, \ldots, i_r) = \max\{1 \leq k \leq n : \{i_1, \ldots, i_r\} \in \pi_k\} \quad \text{and}$$

$$\rho(i_1, \ldots, i_r) = \max\{1 \leq k < \sigma(i_1, \ldots, i_r) : \{i_1, \ldots, i_r\} \notin \pi_k\}.$$ 

For a subset $\{i_1, \ldots, i_r\}$ of leaves, which is not supported by some branch (which means that $\{i_1, \ldots, i_r\} \notin \pi_k$ for all $k$) we set $\sigma(i_1, \ldots, i_r) = \rho(i_1, \ldots, i_r) = n$. 

2
Let \( S_{i_1, \ldots, i_r} \) denote the length of the branch of order \( r \) that supports the leaves \( i_1, \ldots, i_r \) and write \( L_{n,r} \) for the total length of order \( r \). Then,

\[
S_{i_1, \ldots, i_r} = \sum_{l=\rho(i_1, \ldots, i_r) + 1}^{\sigma(i_1, \ldots, i_r)} X_l
\]

and

\[
L_{n,r} = \sum_{1 \leq i_1 < \cdots < i_r \leq n} S_{i_1, \ldots, i_r}.
\]

The length of a randomly chosen external branch in the coalescent tree has been studied by Freund and Möhle [FM09] for the Bolthausen-Sznitman coalescent and by Gnedin et al. [GIM08] for the \( \Lambda \)-coalescent. Asymptotic results concerning the total external length of Beta-coalescents were given by Möhle [Moe10] for the case \( 0 < \alpha < 1 \), by Kersting et al. [KPSJ13] for the case \( \alpha = 1 \) and by Kersting et al. [KSW12] for the case \( 1 < \alpha < 2 \). For the case \( 1 < \alpha < 2 \) a weak law of large numbers result concerning \( L_{n,r} \) can be easily deduced from Theorem 9 of Berestycki et al. [BBS07] and also from Dhersin and Yuan [DY12].

Fu and Li [FL93] computed the expectation and variance of the total external branch length of the Kingman \( n \)-coalescent and Caliebe et al. [CNKR07] derived the asymptotic distribution of a randomly chosen external branch. In [JJK11] Janson and Kersting obtained the asymptotic normality of the total external branch length. Our main result states that the same kind of asymptotics holds for the lengths of order \( r \geq 1 \). Moreover, these lengths are asymptotically independent.

**Theorem.** For any \( s \in \mathbb{N} \), as \( n \to \infty \)

\[
\sqrt{\frac{n}{4 \log n}} \left( L_{n,1} - \mu_1, \ldots, L_{n,s} - \mu_s \right) \xrightarrow{d} N \left( 0, I_s \right),
\]

where \( I_s \) denotes the \( s \times s \)-identity matrix and \( \mu_r = \mathbb{E}(L_{n,r}) = \frac{2}{r} \) for every \( r \geq 1 \).

In a forthcoming paper our theorem will be a main building block for proving a functional limit theorem for the total external length of the evolving Kingman-coalescent.

![Figure 2: External length versus internal length of order 2. The plot is based on 1000 coalescent realisations with \( n = 100 \).](image-url)
The scatterplot for the lengths of orders 1 and 2 in Figure 2 confirms the theorem. The bulk of the points are located around the mean \((\mu_1, \mu_2) = (2, 1)\). Also, in this region hardly any correlation between the two lengths is visible. The outliers are due to exceptionally long branches whose occurrence has been explained in detail in [JK11] for the external case. The simulation shows that this phenomenon appears similarly in the case of internal lengths, as one would expect.

As to the proof of the theorem, for the case \(s = 1\) a hidden symmetry within the Kingman coalescent is used in [JK11]. Here we substantially build on this result, however the proof is rather different. It consists of a coupling device for Markov chains, which connects the total length of order \(r\) to the total external length: For \(1 \leq k \leq n\) let \(W_k(r)\) denote the number of order \(r\) at level \(k\), the number of branches of order \(r\) among the \(k\) branches present in the coalescent tree after the \((n - k)\)-th coalescing event. (Note that here and elsewhere we are suppressing the \(n\) in the notation.) That is,

\[
W_k(r) := \left| \left\{ \{i_1, \ldots, i_r\} \subset \{1, \ldots, n\} : i_1 < \cdots < i_r, \sigma(i_1, \cdots, i_r) \geq k > \rho(i_1, \cdots, i_r) \right\} \right|.
\]

In particular \(W_n(r) = 0, W_{n-1}(r) = 0, \ldots, W_{n-r+2}(r) = 0\) and \(W_1(r) = 0\) for \(r < n\). For an example see Figure 1.

It is important to notice that for any \(s \in \mathbb{N}\) the random vectors \((W_k(1), W_k(2), \ldots W_k(s))\) form a Markov chain if \(k\) runs from \(n\) to 1 (a property which facilitated our simulations). The transition probabilities of the Markov chain are given explicitly in Section 3. The idea of our proof is to couple \((W_k(r))_{n \geq k \geq 1}\) for \(1 \leq r \leq s\) jointly with \(s\) independent copies of the Markov chain of external numbers \((W_k(1))_{n \geq k \geq 1}\). Since in addition the length of order \(r\) is essentially specified by the chain \((W_k(r))_{n \geq k \geq 1}\), it consequently gets the asymptotical behaviour of the external length.

![Figure 3](image)

**Figure 3:** a) Simulations of the external numbers \(W_k(1)\) (in orange) and internal numbers \(W_k(2)\) of order 2 (in blue) for a coalescent with \(n = 100\) for \(1 \leq k \leq n\). The black dashed curves represent the expectations as given in Lemma 1. Figure b) gives the representations in double logarithmic scale for a coalescent with \(n = 10^4\).

The simulations in Figure 3 give an impression of the behaviour of the lengths of different orders. In the range between the levels \(n\) and \(n^{1-\epsilon}\) for small \(\epsilon > 0\) (closer to the leaves) they differ substantially, as seen in Figure 3b). This deviation is only due to expectations and does not appear at the level of fluctuations. Indeed it is known from [JK11] that for the external length the
fluctuations are induced just by the $W_k(1)$ with $n^{1-\varepsilon} \geq k \geq \sqrt{n}$. As suggested by Figure 3b) in this region the evolution of the chains is similar for orders $r \geq 2$. The difference in expectation is negligible in our construction, as we couple the jumps of the chains and afterwards consider the lengths of different orders centred at expectation.

The interest in the quantities $L^{n,r}$ arose from models where the population is subject to mutation, the mutations being modelled as points of a Poisson process with constant rate $\frac{\theta}{2}$ on the branches of the coalescent tree. In the infinitely many sites mutation model, in which every new mutation occurs at a new locus on the DNA, mutations that are located on the external branches of the coalescent tree affect only single individuals, whereas mutations located on an internal branch of order $r \geq 2$ affect all $r$ individuals sitting at the leaves supported by that particular branch. In a population of size $n$ let $M_r(n)$ denote the number of mutations carried by exactly $r$ individuals. The vector $(M_1(n), \ldots, M_{n-1}(n))$, called the site frequency spectrum, and the total number $S_n := \sum_{r=1}^{n-1} M_r(n)$ of mutations that affect the population, called the number of segregating sites, are quantities of statistical importance. Berestycki et al. [BBS07] obtained a weak law of large numbers for $M_r(n)$, $r \geq 1$, in the case of Beta-coalescents with $1 < \alpha < 2$.

For the Kingman coalescent it is know that the number of segregating sites $S_n$, when rescaled by $\log n$, converges almost surely as $n \to \infty$ to $\theta$ (see for example [Be09] Theorem 2.11). The expectation of $M_r(n)$ (which is equal to $\frac{\theta_r}{2}$), as well as the variances and the covariances of the numbers of mutations $M_r(n)$, were computed by Fu [Fu95] and Durrett [Du08]. We obtain the following result as a direct consequence of our Theorem.

**Corollary.** For any $s \in \mathbb{N}$, as $n \to \infty$

$$ (M_1(n), \ldots, M_s(n)) \xrightarrow{d} (M_1, \ldots, M_s), $$

where $M_1, \ldots, M_s$ are independent Poisson-distributed random variables with parameters $\theta \mu_1, \ldots, \theta \mu_s$.

For the proof of the corollary note from the Poissonian structure of the mutation process that the characteristic function of $(M_1(n), \ldots, M_s(n))$ is

$$ \varphi(M_1(n), \ldots, M_s(n))(\lambda_1, \ldots, \lambda_s) = \mathbb{E} \left[ e^{i(\lambda_1 M_1(n) + \cdots + \lambda_s M_s(n))} \right] = \mathbb{E} \left[ e^{\theta L^{n,1}(e^{i\lambda_1} - 1) \cdot \cdots \cdot e^{\theta L^{n,s}(e^{i\lambda_s} - 1)}} \right], $$

where $\mathcal{T}$ denotes the $\sigma$-algebra containing the whole information about the coalescent tree. From our Theorem it follows that $L^{n,r} \overset{p}{\to} \mu_r$ as $n \to \infty$ (which can also be deduced from the results of Fu [Fu95]) and therefore

$$ \varphi(M_1(n), \ldots, M_s(n))(\lambda_1, \ldots, \lambda_s) \to e^{\theta \mu_1(e^{i\lambda_1} - 1) \cdot \cdots \cdot e^{\theta \mu_s(e^{i\lambda_s} - 1)}}, $$

as $n \to \infty$.

**Notation.** Throughout $c$ denotes a finite constant whose value is not important and may change from line to line.
2 Moment computations

Lemma 1. For the expectation and variance of $W_k(r)$ the following is true. For $n > r$
\[
\mathbb{E}(W_k(r)) = \frac{(n-k)\ldots(n-k-r+2)}{(n-1)\ldots(n-r)} \cdot k(k-1) \quad \text{and} \quad \mathbb{V}(W_k(r)) \leq c \frac{k^2}{n},
\]
where $c < \infty$ is a constant depending on $r$. In particular
\[
\mathbb{E}(W_k(r)) = \left( \frac{n-k}{n} \right)^{r-1} \cdot \frac{k^2}{n} + O\left( \frac{k}{n} \right) = \frac{k^2}{n} + O\left( \frac{k^3}{n^2} + \frac{k}{n} \right)
\]
uniformly in $k \leq n$. Also, for $\alpha \geq 2$
\[
\mathbb{E}(W_k(r)^\alpha) = \left( \left( \frac{n-k}{n} \right)^{r-1} \cdot \frac{k^2}{n} \right)^\alpha + O\left( \left( \frac{k}{n} \right)^{\alpha-1} + \frac{k^2}{n} \right) = O\left( \frac{k^{2\alpha}}{n^{\alpha}} + \frac{k^2}{n} \right).
\]

Proof. In order to compute the moments of $W_k(r)$, let us again label the leaves of the coalescent tree from 1 to $n$ and note that $W_k(r)$ can be written as
\[
W_k(r) = \sum_{1 \leq i_1 < \cdots < i_r \leq n} 1_{\{i_1, \ldots, i_r \} \in \pi_k},
\]
where $\pi_k$ is the state of the coalescent process at time $k$. Then for $n > r$, using the fact that the event $\{1, \ldots, r\} \in \pi_k$ is the disjoint union (over $n > l_1 > l_2 > \cdots > l_{r-1} \geq k$) of the events (the branch supporting leaves $1, \ldots, r$ is formed by $r-1$ coalescing events happening at levels $l_1, l_2, \ldots, l_{r-1}$), we have from exchangeability that
\[
\mathbb{E}(W_k(r)) = \mathbb{E}\left( \sum_{1 \leq i_1 < \cdots < i_r \leq n} 1_{\{i_1, \ldots, i_r \} \in \pi_k} \right) = \binom{n}{r} \mathbb{P}(\{1, \ldots, r\} \in \pi_k) = \binom{n}{r} \prod_{j=1}^r \frac{\binom{n-j}{2}}{\binom{n}{2}} = \frac{\binom{n}{r}}{\binom{n}{2}} \prod_{j=1}^r \frac{\binom{n-j}{2}}{\binom{n-j+1}{2}}.
\]

Most binomials in the nominator and the denominator cancel. The summands turn out to be equal such that
\[
\mathbb{E}(W_k(r)) = \binom{n}{r} \prod_{n > l_1 > l_2 > \cdots > l_{r-1} \geq k} \frac{\binom{r}{2} \cdots \binom{2}{2}}{\binom{n-r+1}{2}} \cdot \frac{k}{2} = \binom{n}{r} \binom{n-k}{2} \binom{2}{2} \cdots \binom{2}{2} \cdot \frac{k}{2} = \frac{(n-k)\ldots(n-k-r+2)}{(n-1)\ldots(n-r)} \cdot k(k-1).
\]
The computation of the second moment of $W_k(r)$ follows in a similar way. Note that the event 
\{\{i_1, \ldots, i_r\}, \{j_1, \ldots, j_r\} \in \pi_k\} is nonempty only if the sets \{i_1, \ldots, i_r\} and \{j_1, \ldots, j_r\} are identical or disjoint. Thus, for $n > 2r$

$$
\mathbb{E}(W_k^2(r)) = \binom{n}{r} \mathbb{E}(1_{\{1, \ldots, r\} \in \pi_k}) \mathbb{E}(1_{\{1, \ldots, r\} \cap \{1, \ldots, r\} = \emptyset}) + \binom{n}{r, n - 2r} \mathbb{E}(1_{\{1, \ldots, r\} \cap \{1, \ldots, r\} = \emptyset})
$$

$$
= \binom{n}{r} \mathbb{P}(\{1, \ldots, r\} \in \pi_k) + \binom{n}{r, r, n - 2r} \mathbb{P}(\{1, \ldots, r\}, \{r + 1, \ldots, 2r\} \in \pi_k)
$$

$$
= \mathbb{E}(W_k(r)) + \sum_{\Delta \leq k} \binom{n - k}{r, r, n - 2r} \binom{\ell_1'}{\ell_1'' - 2r} \cdot \binom{\ell_2'}{\ell_2'' + 2} \cdot \binom{\ell_2'' - 2r}{\ell_2'' + 2}
$$

where the sum is taken over all $n > l_1 > l_2 > \cdots > l_{r-1} \geq k$ and all $n > \ell_1' > \ell_2' > \cdots > \ell_{r-1}' \geq k$ such that $\{l_1, \ldots, l_{r-1}\} \cap \{\ell_1', \ldots, \ell_{r-1}'\} = \emptyset$. The sequences $(l_j)_{1 \leq j \leq r-1}$ and $(\ell_j')_{1 \leq j \leq r-1}$ denote the coalescence times of the branches supporting leaves from the sets \{1, \ldots, r\} and \{r + 1, \ldots, 2r\} respectively. The sequence $(\ell_j')_{1 \leq j \leq 2r-2}$ is the reordering of $l_1, \ldots, l_{r-1}, \ell_1', \ldots, \ell_{r-1}'$ in decreasing order. Thus

$$
\mathbb{E}(W_k^2(r)) = \mathbb{E}(W_k(r)) + \sum_{\Delta \leq k} \binom{n - k}{r, r, n - 2r} \binom{\ell_1'}{\ell_1'' - 2r} \cdot \binom{\ell_2'}{\ell_2'' + 2} \cdot \binom{\ell_2'' - 2r}{\ell_2'' + 2}
$$

$$
= \mathbb{E}(W_k(r)) + \frac{(n - k) \cdots (n - k - 2r + 3)}{(n - 1) \cdots (n - 2r)} k(k - 1)^2(k - 2).
$$

The variance of $W_k$ is then for $k \leq n - 1$

$$
\mathbb{V}(W_k(r)) = \mathbb{E}(W_k(r)) \left(1 + \frac{(n - k - r + 1) \cdots (n - k - 2r + 3)}{(n - r - 1) \cdots (n - 2r)} k(k - 1)ight)
$$

$$
= \frac{1}{(n - r - 1) \cdots (n - 2r)} \cdot \frac{1}{(n - 1) \cdots (n - r)}
$$

$$
\leq \frac{c k^2}{n}
$$

for $c < \infty$.

For the other claims we use the same type of argument as above. We have that

$$
\binom{n}{r, r, n - 2r} \mathbb{P}(\{1, \ldots, r\}, \{(\alpha - 1)r + 1, \ldots, \alpha r\} \in \pi_k)
$$

$$
= \binom{n}{r, r, n - 2r} \mathbb{P}(\{1, \ldots, r\}, \{(\alpha - 1)r + 1, \ldots, \alpha r\} \cap \{1, \ldots, r\} = \emptyset)
$$

$$
= \binom{n - k}{r, r, n - 2r} \binom{\ell_1'}{\ell_1'' - 2r} \cdot \binom{\ell_2'}{\ell_2'' + 2} \cdot \binom{\ell_2'' - 2r}{\ell_2'' + 2}
$$

$$
= \frac{(n - k) \cdots (n - k - 2r + 1)}{(n - 1) \cdots (n - 2r + 1) \cdot k(k - 1)^2 \cdots (k - \alpha + 1)}
$$

$$
\leq \frac{c k^2}{n}
$$

for $c < \infty$.
\[
\frac{(n - k)^{\alpha(r-1)}}{n^{\alpha r}} \cdot k^{2\alpha} + O\left(\frac{(n - k)^{\alpha(r-1)}}{n^{\alpha r}} \cdot k^{2\alpha} - k^{2\alpha-1}\right) + O\left(\frac{(n - k)^{\alpha(r-1)}}{n^{\alpha r+1}} \cdot k^{2\alpha}\right)
\]

\[
= \frac{(n - k)^{\alpha(r-1)}}{n^{\alpha r}} \cdot k^{2\alpha} + O\left(\frac{k^{2\alpha-1}}{n^{\alpha}}\right).
\]

(2)

In particular this gives the asymptotic expansion of \(\mathbb{E}(W_k(r))\). Also

\[
\binom{n}{r, \ldots, r, n - \alpha r} \mathbb{P}\left(\{1, \ldots, r\}, \ldots, \{(\alpha - 1)r + 1, \ldots, \alpha r\} \in \pi_k\right) = O\left(\frac{k^{2\alpha}}{n^{\alpha}}\right).
\]

(3)

Moreover, by expanding \(\left(\sum_{1 \leq i_1 < \ldots < i_r \leq n} 1\{i_1, \ldots, i_r \in \pi_k\}\right)^\alpha\)

\[
\mathbb{E}(W_k^\alpha(r)) = \binom{n}{r, \ldots, r, n - \alpha r} \mathbb{P}\left(\{1, \ldots, r\}, \ldots, \{(\alpha - 1)r + 1, \ldots, \alpha r\} \in \pi_k\right)
\]

\[
+ O\left(\sum_{\beta=1}^{\alpha-1} \binom{n}{r, \ldots, r, n - \beta r} \mathbb{P}\left(\{1, \ldots, r\}, \ldots, \{(\beta - 1)r + 1, \ldots, \beta r\} \in \pi_k\right)\right).
\]

(4)

The last claim now follows from (2), (3), (4) and the fact that \(\sum_{\beta=1}^{\alpha-1} \binom{k^2}{\beta} = O\left(\frac{k^2}{n^{\alpha-1}} + \frac{k^2}{n}\right)\).

\[\square\]

**Remark.** It can be read from the computation in (1) that in the case \(r = 2\), given the event \(\{\rho(1,2) < k\}\), the random variable \(\sigma(1,2)\) is uniformly distributed on the set of levels \(\{k, \ldots, n - 1\}\). A similar observation can be made for the case \(r > 2\).

Using the numbers \(W_k(r)\) we have the following simplified expression for the length of order \(r\):

\[
\mathcal{L}^{n,r} = \sum_{2 \leq k \leq n} W_k(r) \cdot X_k.
\]

(5)

We note from this representation that there are two sources of randomness in the length of order \(r\), one coming from the numbers \(W_k(r)\) and one coming from the exponential inter-coalescence times. It is easy to see that taking out the randomness introduced by the exponential times (i.e. replacing them by their expectations) leads to an error that is asymptotically \(O_P\left(n^{-1/2}\right)\) and therefore converges to 0 after the rescaling by \(\sqrt{\frac{n}{\log n}}\). Indeed, by using the independence between the \(X_k\)'s and the \(W_k(r)\)'s and Lemma 1 we have that for a constant \(c < \infty\)

\[
\mathbb{V}\left(\sum_{2 \leq k \leq n} W_k(r) \cdot (X_k - \mathbb{E}(X_k))\right) = \sum_{2 \leq k \leq n} \mathbb{V}\left(W_k(r) \cdot (X_k - \mathbb{E}(X_k))\right)
\]

\[
= \sum_{2 \leq k \leq n} \mathbb{E}(W_k^2(r)) \cdot \mathbb{E}((X_k - \mathbb{E}(X_k))^2)
\]

\[
\leq c \sum_{2 \leq k \leq n} \left(\frac{k^4}{n^2} + \frac{k^2}{n}\right) \cdot \frac{1}{k} \leq c \frac{1}{n}
\]
and therefore
\[ L^{n,r} = L^{n,r} + O_P(n^{-1/2}), \]
where
\[ L^{n,r} := \sum_{2 \leq k \leq n} W_k(r) \cdot E(X_k) = \sum_{2 \leq k \leq n} W_k(r) \cdot \frac{2}{k(k-1)}. \] (6)

As a consequence, in the proof of our Theorem we need only focus on the length \( L^{n,r} \) which we will, for convenience, still call the length of order \( r \).

3 The coupling

Our proof follows a coupling argument which substantially relies on the observation that for every \( s \in \mathbb{N} \) the vector \( V_k := (W_k(1), W_k(2), \ldots, W_k(s)) \) follows for \( n \geq k \geq 1 \) the dynamics of an inhomogenous Markov chain with state space \( \mathcal{X}_{n,s} := \{0, 1, \ldots, n\}^s \). We let time run in coalescent direction (from the leaves to the root of the tree) and for convenience we consider the evolution of the chain \((V_k)_{n \geq k \geq 1}\), running in the same direction, namely from level \( n \) to level 1.

For every \( 1 \leq r \leq s \) we denote by \( \Delta W_{n-1}(r), \ldots, \Delta W_1(r) \) the sizes of the jumps of the chain \((W_k(r))_{n \geq k \geq 1}\),
\[ \Delta W_k(r) := W_k(r) - W_{k+1}(r), \quad n - 1 \geq k \geq 1 \]
and observe that \( \Delta W_k(r) \in \{-2, -1, 0, 1\} \) for all \( k \). The jumps of size 1 correspond to the levels at which a new branch of order \( r \) is formed (by the coalescence of two other branches), whereas the jumps of sizes -1 and -2 happen at the levels at which one or respectively two branches of order \( r \) end (by coalescence of one of them with some other branch or by mutual coalescence).

For \( 1 \leq k \leq n \) and \( v, v' \in \mathcal{X}_{n,s} \) let
\[ P^k_v(v') := \mathbb{P} (V_{k-1} = v' \mid V_k = v) \]
denote the transition probabilities of \((V_k)_{n \geq k \geq 1}\). They are given for \( v = (w_1, \ldots, w_s) \), \( w_1 + \cdots + w_s \leq k \) by
\[ P^k_v(v') = \begin{cases} \binom{k-w_1-w_2-\cdots-w_s}{2} \binom{k}{2} & \text{if } v = v', \\ \frac{w_i(k-w_1-w_2-\cdots-w_s)}{k} & \text{if } v = v' - e_i \text{ for some } i, \\ \frac{w_i}{k} & \text{if } v = v' - 2e_i + e_{2i} \text{ for some } i, \\ \frac{w_iw_j}{k} & \text{if } v = v' - e_i - e_j + e_{i+j} \text{ for some } i \neq j, \\ 0 & \text{else}, \end{cases} \] (7)
where \( e_i = (\delta_{i,j})_{1 \leq i \leq s} \) (note that \( e_i = (0, \ldots, 0) \) for \( i > s \)). For all other \( v \in \mathcal{X}_{n,s} \) we set for definiteness \( P^k_v(v') = \delta_{v,v'} \).
In particular for \( s = 2 \) the transition probabilities in (7) are given for \( w_1 + w_2 \leq k \) by

\[
\begin{align*}
P^k_{(w_1, w_2)}(w_1, w_2) &= \frac{(k-w_1-w_2)}{(k^2)} \\
&= \frac{w_1(k-w_1-w_2)}{(k^2)} \quad P^k_{(w_1, w_2)}(w_1, w_2-1) = \frac{w_2(k-w_1-w_2)}{(k^2)}
\end{align*}
\]

and for \( s = 1 \) for \( w \leq k \) by

\[
P^k_w(w) = \frac{(k-w)}{(k^2)} , \quad P^k_w(w - 1) = \frac{w(k-w)}{(k^2)} , \quad P^k_w(w - 2) = \frac{(w^2)}{(k^2)} . \quad (8)
\]

Let us now describe the coupling in detail. Let \( 1 < a_n \leq n \) and \( s \in \mathbb{N} \) be fixed. Starting at \( a_n \) we couple the Markov chain \( (V_k)_{a_n \leq k \geq 1} \), \( V_k = (W_k(1), \ldots, W_k(s)) \), with another chain \( (\tilde{V}_k)_{a_n \leq k \geq 1} \), \( \tilde{V}_k = (\tilde{W}_k(1), \ldots, \tilde{W}_k(s)) \), defined on the same probability space as \( (V_k)_{a_n \geq k \geq 1} \) and having the same state space \( \mathcal{X}_{n,s} \). The components of \( (\tilde{V}_k)_{a_n \leq k \geq 1} \) evolve as independent copies of \( (W_k(1))_{a_n \geq k \geq 1} \), the Markov chain of external numbers. Therefore its transition probabilities \( \tilde{P}^k_v(\cdot) \) for \( v \in \mathcal{X}_{n,s} \) and \( 1 < k \leq a_n \) are given by the product of the transition probabilities of its components, given in (8). The process \( (V_k, \tilde{V}_k)_{a_n \geq k \geq 1} \) is constructed as a Markov chain, where the jumps are coupled.

Thus let \( Q^k_v \) and \( \tilde{Q}^k_v \) denote the conditional distributions of the jumps \( \Delta V_k \) and \( \Delta \tilde{V}_k \) of the two Markov chains, given the current states \( v \) and \( \tilde{v} \) respectively. (The notations \( \Delta V_k \) and \( \Delta \tilde{V}_k \) refer to component-wise differences). For the sequel it is important that the leading terms of \( Q^k_v \) and \( \tilde{Q}^k_v \) agree. More precisely, from (7) and (8), under the constrain that \( w_1 + \cdots + w_s \leq k \)

\[
Q^k_v(z), \tilde{Q}^k_v(z) = \begin{cases} 
1 - \frac{2w_1}{k} - \cdots - \frac{2w_s}{k} + O\left(\frac{w_1^2 + \cdots + w_s^2}{k^2}\right) & \text{ if } z = (0, \ldots, 0), \\
\frac{2w_i}{k} + O\left(\frac{w_1^2 + \cdots + w_s^2}{k^2}\right) & \text{ if } z = -e_i \text{ for some } i, \\
O\left(\frac{w_1^2 + \cdots + w_s^2}{k^2}\right) & \text{ else,}
\end{cases} \quad (9)
\]

where \( z \in \{-2, -1, 0, 1\}^s \) and \( e_i = (\delta_{i,j})_{1 \leq j \leq s} \). Here we use that \( w_i w_j \leq w_i^2 + w_j^2 \) and \( w_i \leq k \).

As it is well-known (see for example [LPW09]), an optimal coupling of the two distributions is specified as follows. Let \( \| \cdot \|_{TV} \) denote the total variation distance between two distributions and define

\[
p = p_{v, \tilde{v}} := 1 - \| Q^k_v - \tilde{Q}^k_v \|_{TV} .
\]
Then, with probability $p$ choose $\Delta V_k = \Delta \tilde{V}_k = Z$, where the random variable $Z$ has distribution $\gamma_I$, given by its weights
\[
\gamma_I(z) = \frac{Q_k^v(z) \land \tilde{Q}_k^v(z)}{p},
\]
z $\in \{-2, -1, 0, 1\}$ and with probability $1 - p$ choose $\Delta V_k$ according to the probability distribution weights
\[
\gamma_{III}(z) = \frac{(Q_k^v(z) - \tilde{Q}_k^v(z))^+}{1 - p},
\]
and independently choose $\Delta \tilde{V}_k$ according to the probability distribution weights
\[
\gamma_{III}(z) = \frac{(\tilde{Q}_k^v(z) - Q_k^v(z))^+}{1 - p},
\]
z $\in \{-2, -1, 0, 1\}$.

This coupling is optimal in the sense that the probability $\mathbb{P}\left(\Delta V_k \neq \Delta \tilde{V}_k \mid V_k = v, \tilde{V}_k = \tilde{v}\right)$ is minimal among the corresponding probabilities for couplings of the two distributions $Q_v^k$ and $\tilde{Q}_v^k$, and therefore it is equal to $\left\|Q_v^k - \tilde{Q}_v^k\right\|_{TV}$. As starting distributions of the coupled chains $(V_k)_{a_n \geq k \geq 1}$ and $(\tilde{V}_k)_{a_n \geq k \geq 1}$ we allow any distribution of $(V_{a_n}, \tilde{V}_{a_n})$ such that the marginals are the distributions of $V_{a_n}$ and $\tilde{V}_{a_n}$ respectively.

The next two lemmas give essential properties of the coupling.

**Lemma 2.** There is a $c < \infty$ such that the above defined coupling satisfies for $r \leq s$
\[
\mathbb{P}\left(\Delta W_k(r) \neq \Delta \tilde{W}_k(r)\right) \leq c \left(\frac{k}{a_n \sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k}\right)
\]
and
\[
\mathbb{E}\left(\left|W_k(r) - \tilde{W}_k(r)\right|\right) \leq c \left(\frac{k^2}{a_n \sqrt{n}} + \frac{a_n k^2}{n^2} + 1\right)
\]
for all $1 \leq k < a_n$.

**Proof.** In the proof we write shortly $W_k$ instead of $W_k(r)$ and similarly $\tilde{W}_k$, $\Delta W_k$ and $\Delta \tilde{W}_k$ instead of $W_k(r)$, $\Delta W_k(r)$ and $\Delta \tilde{W}_k(r)$ respectively.

From (9) it follows that for both the chains $(V_k)_{a_n \geq k \geq 1}$ and $(\tilde{V}_k)_{a_n \geq k \geq 1}$ jumps of sizes $(0, \ldots, 0)$ and $-e_i$ with $1 \leq i \leq r$ occur with probabilities of larger order than jumps of other sizes. It follows from the coupling that
\[
\{\Delta W_k \neq \Delta \tilde{W}_k\} \subset \{\Delta W_k = -1, \Delta V_k \neq \Delta \tilde{V}_k\}
\]
\[
\cup \{\Delta \tilde{W}_k = -1, \Delta V_k \neq \Delta \tilde{V}_k\}
\]
\[
\cup \{\Delta W_k \in \{1, -2\}\} \cup \{\Delta \tilde{W}_k = -2\}
\]
Thus, writing shortly $\mathbb{P}^k_{v, \tilde{v}}(\cdot)$ for the conditional probability given the event $\{V_k = v, \tilde{V}_k = \tilde{v}\}$, we obtain
\[ \mathbb{P}_{v, \tilde{v}}^{k+1} (\Delta W_k \neq \Delta \tilde{W}_k) \leq (1 - p) \gamma_{II} (\Delta W_k = -1) + (1 - p) \gamma_{III} (\Delta \tilde{W}_k = -1) \]
\[ + \mathbb{P}_{v, \tilde{v}}^{k+1} (\Delta W_k \in \{1, -2\}) + \mathbb{P}_{v, \tilde{v}}^{k+1} (\Delta \tilde{W}_k = -2) \]
\[ \leq (1 - p) \gamma_{II} (\Delta V_k = -e_r) + (1 - p) \gamma_{III} (\Delta \tilde{V}_k = -e_r) \]
\[ + c \cdot \sum_{i=1}^{r} \left( \frac{w_i^2 + \tilde{w}_i^2}{k^2} \right). \]

Using now the definitions of \( \gamma_{II} \) and \( \gamma_{III} \) we get that
\[ \mathbb{P}_{v, \tilde{v}}^{k+1} (\Delta W_k \neq \Delta \tilde{W}_k) \leq \left| Q_{v, \tilde{v}}^{k+1} (-e_r) - \tilde{Q}_{v, \tilde{v}}^{k+1} (-e_r) \right| + c \cdot \sum_{i=1}^{r} \left( \frac{w_i^2 + \tilde{w}_i^2}{k^2} \right). \]  

(10)

Let us introduce the filtration \( \mathbb{F} = (\mathcal{F}_k)_{1 \leq k \leq a_n} \) with \( \mathcal{F}_{a_n} \subset \mathcal{F}_{a_n - 1} \subset \cdots \subset \mathcal{F}_1 \) defined by
\[ \mathcal{F}_k = \sigma \left( (V_j)_{k \leq j \leq a_n}, (\tilde{V}_j)_{k \leq j \leq a_n} \right). \]

Then (10) in view of (9) may be written as
\[ \mathbb{P} \left( \Delta W_k \neq \Delta \tilde{W}_k \mid \mathcal{F}_{k+1} \right) \leq \frac{2}{k} \cdot |W_{k+1} - \tilde{W}_{k+1}| + c \cdot \sum_{i=1}^{r} \left( \frac{W_{k+1}^2(i) + \tilde{W}_{k+1}^2(i)}{k^2} \right) \]
\[ \text{for a constant } c < \infty. \]  

(11)

Taking expectation in the equality above we obtain using Lemma 1
\[ \mathbb{P} \left( \Delta W_k \neq \Delta \tilde{W}_k \right) \leq \frac{2}{k} \cdot \mathbb{E} \left( |W_{k+1} - \tilde{W}_{k+1}| \right) + c \left( \frac{k^2}{n^2} + \frac{1}{n} \right), \]

(12)

We now proceed to finding a bound for \( \mathbb{E} \left( |W_k - \tilde{W}_k| \right) \) for \( 2 \leq k \leq a_n \). From the transition probabilities (7) we get that
\[ \mathbb{E}[\Delta W_k \mid \mathcal{F}_{k+1}] \]
\[ = (-1) \cdot \frac{W_{k+1}(r) (k + 1 - W_{k+1}(1) - \cdots - W_{k+1}(r))}{\binom{k+1}{2}} \]
\[ + (-1) \cdot \frac{W_{k+1}(r) W_{k+1}(1) + \cdots + W_{k+1}(r) W_{k+1}(r - 1)}{\binom{k+1}{2}} + (-2) \cdot \frac{W_{k+1}^2(r)}{\binom{k+1}{2}} + 1 \cdot \frac{Z_{k+1}}{\binom{k+1}{2}} \]
\[ = -\frac{2}{k+1} W_{k+1} + \frac{Z_{k+1}}{\binom{k+1}{2}} \]

where, letting \( d_r = 1 \) if \( r \) is even and 0 otherwise,
\[ Z_k = Z_k(r) := \sum_{1 \leq i \leq r-1, \ i \neq r-i} W_k(i) W_k(r - i) + d_r \cdot \binom{W_k(r/2)}{2}. \]  

(13)
Using (14) and (15) we obtain
\[ E[\Delta W_k \mid F_{k+1}] = -\frac{2}{k+1}W_{k+1} + \frac{Z_{k+1}}{(k+1)^2} \]
(14)
and also
\[ E[\Delta \tilde{W}_k \mid F_{k+1}] = -\frac{2}{k+1}\tilde{W}_{k+1}. \]
(15)

Now note that the absolute value of the difference between the jumps of \( W_{k+1} \) and \( \tilde{W}_{k+1} \) is at most 3. Thus
\[
E[|W_k - \tilde{W}_k| \mid F_{k+1}] = E[|W_{k+1} - \tilde{W}_{k+1} + \Delta W_k - \Delta \tilde{W}_k| \mid F_{k+1}]
\leq E[|W_{k+1} - \tilde{W}_{k+1} + \Delta W_k - \Delta \tilde{W}_k| \mid F_{k+1}] \cdot 1_{\{W_{k+1} - \tilde{W}_{k+1} \geq 3\}}
+ E[|W_{k+1} - \tilde{W}_{k+1} + \Delta \tilde{W}_k - \Delta W_k| \mid F_{k+1}] \cdot 1_{\{W_{k+1} - \tilde{W}_{k+1} \leq -3\}}
+ \left( |W_{k+1} - \tilde{W}_{k+1}| + E[|\Delta W_k - \Delta \tilde{W}_k| \mid F_{k+1}] \right) \cdot 1_{\{|W_{k+1} - \tilde{W}_{k+1}| \leq 2\}}.
\]

Using (14) and (15) we obtain
\[
E[|W_k - \tilde{W}_k| \mid F_{k+1}] \leq \left( W_{k+1} - \tilde{W}_{k+1} - \frac{2}{k+1}\left(W_{k+1} - \tilde{W}_{k+1}\right) + \frac{Z_{k+1}}{(k+1)^2} \right) \cdot 1_{\{W_{k+1} - \tilde{W}_{k+1} \geq 3\}}
+ \left( \tilde{W}_{k+1} - W_{k+1} - \frac{2}{k+1}\left(\tilde{W}_{k+1} - W_{k+1}\right) - \frac{Z_{k+1}}{(k+1)^2} \right) \cdot 1_{\{W_{k+1} - \tilde{W}_{k+1} \leq -3\}}
+ \left( |W_{k+1} - \tilde{W}_{k+1}| + 3 \cdot P(\Delta W_k \neq \Delta \tilde{W}_k \mid F_{k+1}) \right) \cdot 1_{\{|W_{k+1} - \tilde{W}_{k+1}| \leq 2\}}.
\]

By (11) we have that
\[
E[|W_k - \tilde{W}_k| \mid F_{k+1}]
\leq \left( |W_{k+1} - \tilde{W}_{k+1}| - \frac{2}{k+1}|W_{k+1} - \tilde{W}_{k+1}| + \frac{Z_{k+1}}{(k+1)^2} \right) \cdot 1_{\{|W_{k+1} - \tilde{W}_{k+1}| \geq 3\}}
+ \left( |W_{k+1} - \tilde{W}_{k+1}| + \frac{6}{k}|W_{k+1} - \tilde{W}_{k+1}| + c \sum_{i=1}^{r} \left( \frac{W_{k+1}^2(i) + \tilde{W}_{k+1}^2(i)}{k^2} \right) \right) \cdot 1_{\{|W_{k+1} - \tilde{W}_{k+1}| \leq 2\}}
\leq |W_{k+1} - \tilde{W}_{k+1}| \left( 1 - \frac{2}{k+1} \right) + \frac{16}{k} + c \sum_{i=1}^{r} \left( \frac{W_{k+1}^2(i) + \tilde{W}_{k+1}^2(i)}{k^2} \right) + \frac{Z_{k+1}}{(k+1)^2}.
\]

Taking expectation and using Lemma 1 we obtain that
\[
E(|W_k - \tilde{W}_k|) \leq \left( 1 - \frac{2}{k+1} \right) E[|W_{k+1} - \tilde{W}_{k+1}|] + c \left( \frac{k^2}{n^2} + \frac{1}{k} \right).
\]

Dividing the previous inequality by \( k(k-1) \) we obtain a recurrence formula that we iterate from \( k \) up to \( a_n - 1 \):
\[
\frac{1}{k(k-1)} \mathbb{E}(|W_k - \tilde{W}_k|) \leq \frac{1}{k(k+1)} \left( \mathbb{E}(|W_{k+1} - \tilde{W}_{k+1}|) + c \left( \frac{k^2}{n^2} + \frac{1}{k} \right) \right)
\]
\[
\leq \frac{1}{a_n(a_n-1)} \mathbb{E}(|W_{a_n} - \tilde{W}_{a_n}|) + c \sum_{j=k}^{a_n-1} \left( \frac{1}{n^2} + \frac{1}{j^3} \right)
\]
\[
\leq \frac{1}{a_n(a_n-1)} \left( \mathbb{E}(|W_{a_n} - \mathbb{E}(W_{a_n})|) + \mathbb{E}(W_{a_n} - \mathbb{E}(\tilde{W}_{a_n})) + \mathbb{E}\left( \mathbb{E}(W_{a_n}) - \mathbb{E}(\tilde{W}_{a_n}) \right) + c \left( \frac{a_n}{n^2} + \frac{1}{k^2} \right) \right).
\]

Finally by Lemma 1

\[
\frac{1}{k(k-1)} \mathbb{E}(|W_k - \tilde{W}_k|) \leq c \left( \frac{1}{a_n(a_n-1)} \left( \frac{a_n}{\sqrt{n}} \left( \frac{a_n^3}{n^2} + \frac{a_n}{n} \right) + \frac{a_n}{n^2} + \frac{1}{k^2} \right) \right) \tag{16}
\]
\[
\leq c \left( \frac{1}{a_n\sqrt{n}} + \frac{a_n}{n^2} + \frac{1}{k^2} \right).
\]

This gives the second claim of the Lemma. Using now (16) in (12) yields the first claim. \( \square \)

**Lemma 3.** There is a constant \( c < \infty \) such that for \( r \leq s \) it holds that

\[
\mathbb{V}(W_k(r) - \tilde{W}_k(r)) \leq c \cdot \left( \frac{k^2}{a_n\sqrt{n}} + \frac{a_n k^2}{n^2} + \frac{k^3}{a_n n} + 1 \right)
\]

for all \( 1 \leq k < a_n \).

**Proof.** We again write here shortly \( W_k, \tilde{W}_k, \Delta W_k \) and \( \Delta \tilde{W}_k \) instead of \( W_k(r), \tilde{W}_k(r), \Delta W_k(r) \) and \( \Delta \tilde{W}_k(r) \) respectively.

Using (14) and (15) together with the fact that \( |\Delta W_{k-1} - \Delta \tilde{W}_{k-1}| \leq 3 \) we obtain

\[
\mathbb{V}\left(W_{k-1} - \tilde{W}_{k-1}\right) = \mathbb{V}\left(W_k - \tilde{W}_k + \Delta W_{k-1} - \Delta \tilde{W}_{k-1}\right)
\]
\[
\leq \mathbb{V}\left(W_k - \tilde{W}_k\right)
\]
\[
+ 2 \mathbb{E}\left( \mathbb{E}\left[ \left(W_k - \tilde{W}_k - \mathbb{E}\left(W_k - \tilde{W}_k\right)\right) \left(\Delta W_{k-1} - \Delta \tilde{W}_{k-1} - \mathbb{E}\left(\Delta W_{k-1} - \Delta \tilde{W}_{k-1}\right)\right) | \mathcal{F}_k \right] \right)
\]
\[
+ \mathbb{E}\left(\Delta W_{k-1} - \Delta \tilde{W}_{k-1}\right)^2
\]
\[
= \left(1 - \frac{4}{k}\right) \mathbb{V}\left(W_k - \tilde{W}_k\right) + 2 \mathbb{E}\left(\left(W_k - \tilde{W}_k - \mathbb{E}\left(W_k - \tilde{W}_k\right)\right) \cdot \frac{Z_k - \mathbb{E}(Z_k)}{\binom{k}{2}}\right)
\]
\[
+ 9 \mathbb{P}(W_{k-1} \neq \Delta \tilde{W}_{k-1}).
\]

Applying now the Cauchy-Schwarz inequality for the second term and Lemma 2 for the third term on the right-hand side of the inequality above we obtain that for a constant \( c < \infty \)

\[
\mathbb{V}\left(W_{k-1} - \tilde{W}_{k-1}\right) \leq \left(1 - \frac{4}{k}\right) \mathbb{V}\left(W_k - \tilde{W}_k\right) + \frac{4}{k(k-1)} \mathbb{V}\left(W_k - \tilde{W}_k\right)^{1/2} \cdot \mathbb{V}(Z_k)^{1/2}
\]
\[
+ c \left( \frac{k}{a_n \sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k} \right). \tag{17}
\]
Let us now look closer at the variance of $Z_k$. In order to bound it from above, it is sufficient to bound the terms of the form $\mathbb{V}(W_k(i)W_k(j))$, $1 \leq i, j \leq r$ (see the definition of $Z_k$ in (13)). Writing shortly $W_k'$ and $W_k''$ for $W_k(i)$ and $W_k(j)$ respectively we have that

$$
\mathbb{V}(W_k(i)W_k(j)) \leq \mathbb{E}\left(W_k'' \mathbb{E}(W_k') - \mathbb{E}(W_k')\mathbb{E}(W_k'')\right)^2
$$

$$
= \mathbb{E}\left((W_k' - \mathbb{E}(W_k'))W_k'' + \mathbb{E}(W_k')(W_k'' - \mathbb{E}(W_k''))\right)^2
$$

$$
\leq 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))^2\right) + 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))(W_k'' - \mathbb{E}(W_k''))\mathbb{E}(W_k'')\right)
$$

Using the fact that $W_k'' \leq k$ and then Lemma II we obtain

$$
\mathbb{V}(W_k(i)W_k(j)) \leq 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))^2\right) + 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))\mathbb{E}(W_k'')\right)
$$

$$
\leq 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))^2\right) + 2k \cdot \mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2\mathbb{E}(W_k'')\right) + 2\mathbb{E}(W_k')^2\mathbb{V}(W_k'')
$$

$$
\leq 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))^2\right) + 2k \cdot \mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2\mathbb{E}(W_k'')\right) + 2\mathbb{E}(W_k')^2\mathbb{V}(W_k'')
$$

$$
\leq 2\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2(W_k'' - \mathbb{E}(W_k''))^2\right) + \frac{k^5}{n^2}
$$

for a constant $c < \infty$. Moreover

$$
\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2\right)
$$

$$
= \mathbb{E}((W_k')^4) - 4\mathbb{E}((W_k')^3)\mathbb{E}(W_k') + 6\mathbb{E}((W_k')^2)(\mathbb{E}(W_k'))^2 - 4\mathbb{E}(W_k')^2(\mathbb{E}(W_k'))^3 + (\mathbb{E}(W_k'))^4.
$$

Inserting now the formulas from Lemma II the leading terms cancel and we obtain that

$$
\mathbb{E}\left((W_k' - \mathbb{E}(W_k'))^2\right) = O\left(\frac{k^6}{n^3} + \frac{k^2}{n}\right).
$$

Using the Cauchy-Schwarz inequality and (19) in (18) we get that

$$
\mathbb{V}(W_k(i)W_k(j)) \leq c \left(\frac{k^6}{n^3} + \frac{k^5}{n^2} + \frac{k^2}{n}\right)
$$

and therefore

$$
\mathbb{V}(Z_k) \leq c \left(\frac{k^5}{n^2} + \frac{k^2}{n}\right)
$$

for some constant $c < \infty$. Plugging this into (17) we obtain that

$$
\mathbb{V}\left(W_{k-1} - \tilde{W}_{k-1}\right) \leq \left(1 - \frac{4}{k}\right)\mathbb{V}\left(W_k - \tilde{W}_k\right)
$$

$$
+ c \left(\mathbb{V}(W_k - \tilde{W}_k)\right)^{1/2} \left(\frac{\sqrt{k}}{n} + \frac{1}{k\sqrt{n}}\right) + c \left(\frac{k}{a_n\sqrt{n}} + \frac{a_nk}{n^2} + \frac{1}{k}\right).
$$
Observe that
\[
\left( \frac{\sqrt{k}}{n} + \frac{1}{n^{1/2}} \right)
\leq \begin{cases} 
\frac{\sqrt{k}}{n} & \text{if } \left( \frac{\sqrt{k}}{n} \right)^{1/2} \geq c \left( \frac{k^{3/2}}{n} + \frac{1}{\sqrt{n}} \right), \\
2c^2 \left( \frac{k^2}{n^2} + \frac{1}{kn} \right) & \text{else}
\end{cases}
\]
and therefore \(\eqref{Eq1} \) becomes
\[
\mathbb{V} \left( W_{k-1} - \tilde{W}_{k-1} \right) \leq \left( 1 - \frac{3}{k} \right) \mathbb{V} \left( W_k - \tilde{W}_k \right) + c \left( \frac{k}{a_n \sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k} \right).
\]
We now divide both sides by \(\binom{k-1}{3} \) and iterate up to \(a_n\). Since \(\mathbb{V} \left( W_{a_n} - \tilde{W}_{a_n} \right) \leq c \frac{a_n^2}{n} \), we obtain
\[
\frac{1}{\binom{k-1}{3}} \mathbb{V} \left( W_{k-1} - \tilde{W}_{k-1} \right) \leq c \cdot \sum_{j=k}^{a_n} \frac{1}{\binom{j}{3}} \left( \frac{1}{a_n \sqrt{n}} + \frac{a_n j}{n^2} + \frac{1}{j} \right) + c \cdot \frac{1}{\binom{a_n}{3}} \cdot \frac{a_n^2}{n}
\]
and therefore
\[
\mathbb{V} \left( W_{k-1} - \tilde{W}_{k-1} \right) \leq c \cdot k^3 \left( \frac{1}{ka_n \sqrt{n}} + \frac{a_n}{kn^2} + \frac{1}{k^3} + \frac{1}{a_n n} \right)
\]
\[
\leq c \cdot \left( \frac{k^2}{a_n \sqrt{n}} + \frac{a_n k^2}{n^2} + k^3 \frac{1}{a_n n} + 1 \right).
\]
This is the claim. \(\square\)

4 Proof of the Theorem

The proof that \(\mu_r\), the expected length of order \(r\), is equal to \(\frac{2}{r}\) for every \(r \geq 1\) can be found in [Be09], Theorem 2.11 or in [Du08], Theorem 2.1. Another quick way to see this is by using Lemma [H]

\[
\mathbb{E}(L^{n,r}) = \mathbb{E} \left( \sum_{k=2}^{n} W_k(r) X_k \right) = \sum_{k=2}^{n} \mathbb{E}(W_k(r)) \mathbb{E}(X_k) = \sum_{k=2}^{n} \frac{(n-k) \cdots (n-k-r+2)}{(n-1) \cdots (n-r)} \cdot k(k-1) \cdot \frac{1}{\binom{k}{2}}
\]
\[
= \frac{2}{(n-1) \cdots (n-r)} \cdot \sum_{j=1}^{n-r} j(j+1) \cdots (j+r-2).
\]

The claim follows now from the fact that \(\sum_{j=1}^{n} j(j+1) \cdots (j+i) = \frac{1}{i+2} n(n+1) \cdots (n+i+1)\).

The asymptotic normality of the total external branch length of the Kingman coalescent (case \(s = 1\)) was proved in [JK11]. We will prove the theorem for \(s \geq 2\).

For \(1 \leq r \leq s\) we divide \(L^{n,r}\) and the corresponding coupled quantity into parts. For \(1 \leq b_n < a_n \leq n\) let
\[
L_{a_n,b_n}^{n,r} := \sum_{b_n < k \leq a_n} \frac{2}{k(k-1)} \cdot W_k \quad \text{and} \quad \tilde{L}_{a_n,b_n}^{n,r} := \sum_{b_n < k \leq a_n} \frac{2}{k(k-1)} \cdot \tilde{W}_k
\]
be the length of order \( r \) collected between the levels \( b_n \) and \( a_n \) in the coalescent tree and the corresponding quantity obtained from the coupling. Note that \( L_{n,1}^{n,r} = L_{n,1}^{n,r} \) with \( L_{n,1}^{n,r} \) defined in (6) and let similarly

\[
\overline{L}_{n,1}^{n,r} := \overline{L}_{n,1}^{n,r}.
\] (22)

Using the coupling we will show that for \( \epsilon > 0 \)

\[
P\left( \sqrt{\frac{n}{\log n}} \cdot \left\| L_{n,1}^{n,r} - \mathbb{E}(L_{n,1}^{n,r}) , \ldots , L_{n,s}^{n,r} - \mathbb{E}(L_{n,s}^{n,r}) \right\| \geq \epsilon \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\] (23)

Once (23) has been proved, the claim of the theorem follows since the components of the second vector above are from construction independent and identically distributed and they converge weakly to the standard normal distribution as \( n \rightarrow \infty \), as follows from the case \( s = 1 \) proved in [JK11].

The convergence in (23) is a direct consequence of the following result.

**Proposition 1.** For \( L_{n,r}^{n,r} \) and \( \overline{L}_{n,r}^{n,r} \) defined in (6) and (22) respectively one has for all \( 1 \leq r \leq s \) and \( \epsilon > 0 \)

\[
P\left( \left\| \sqrt{\frac{n}{\log n}} \cdot \left( L_{n,r}^{n,r} - \mathbb{E}(L_{n,r}^{n,r}) \right) - \left( \overline{L}_{n,r}^{n,r} - \mathbb{E}(\overline{L}_{n,r}^{n,r}) \right) \right\| \geq \epsilon \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\] (24)

**Proof.** We have by the Cauchy-Schwarz inequality that

\[
\forall \left( \sum_{b_n < k \leq a_n} \frac{2}{k(k-1)} \cdot \left( W_k - \widetilde{W}_k - (\mathbb{E}(W_k) - \mathbb{E}(\widetilde{W}_k)) \right) \right)
\]

\[
= \sum_{b_n < k \leq a_n} \sum_{b_n < l \leq a_n} \frac{2}{k(k-1)} \cdot \frac{2}{l(l-1)} \cdot \mathbb{C} \mathbb{O} \mathbb{V} \left( W_k - \widetilde{W}_k , W_l - \widetilde{W}_l \right)
\]

\[
\leq \sum_{b_n < k \leq a_n} \sum_{b_n < l \leq a_n} \frac{2}{k(k-1)} \cdot \frac{2}{l(l-1)} \cdot \mathbb{V} \left( W_k - \widetilde{W}_k \right)^{1/2} \mathbb{V} \left( W_l - \widetilde{W}_l \right)^{1/2}
\]

\[
= \left( \sum_{b_n < k \leq a_n} \frac{2}{k(k-1)} \cdot \mathbb{V} \left( W_k - \widetilde{W}_k \right)^{1/2} \right)^2.
\]

Using Lemma 3 we obtain

\[
\forall \left( \sum_{b_n < k \leq a_n} \frac{2}{k(k-1)} \cdot \left( W_k - \widetilde{W}_k - (\mathbb{E}(W_k) - \mathbb{E}(\widetilde{W}_k)) \right) \right)
\] (25)

\[
\leq c \left( \sum_{b_n < k \leq a_n} \frac{1}{k^2} \cdot \left( \frac{k}{\sqrt{a_n n^{1/4}}} + \frac{\sqrt{a_n}}{n} + \frac{k^{3/2}}{\sqrt{a_n n}} + 1 \right) \right)^2
\]

\[
\leq c \left( \frac{1}{\sqrt{a_n n^{1/4}}} \log \frac{a_n}{b_n} + \frac{\sqrt{a_n}}{n} \log \frac{a_n}{b_n} + \frac{1}{\sqrt{n}} + \frac{1}{b_n} \right)^2
\]

17
\[
\leq c \left( \left( \frac{1}{a_n \sqrt{n}} + \frac{a_n}{n^2} \right) \log \frac{a_n}{b_n} + \frac{1}{n} + \frac{1}{\sqrt{n}} \right).
\]

In order to show that the claim holds we consider three regions in the coalescent tree, namely between level \(n\) and level \(\frac{n}{(\log n)^2}\), between level \(\frac{n}{(\log n)^2}\) and level \(n^{1/2}\) and finally between level \(n^{1/2}\) and level 1, and write the lengths \(L_{n,1}^{n, r}\) and \(\tilde{L}_{n,1}^{n, r}\) as sums of the lengths gathered in these three regions.

For the first region let
\[
a_n = n \quad \text{and} \quad b_n = \frac{n}{(\log n)^2}.
\]

We obtain from (25) and Chebyshev’s inequality that
\[
L_{a_n, b_n}^{n, r} - \mathbb{E}(L_{a_n, b_n}^{n, r}) = \tilde{L}_{a_n, b_n}^{n, r} - \mathbb{E}(\tilde{L}_{a_n, b_n}^{n, r}) + \mathcal{O}_P \left( \frac{\log \log n}{\sqrt{n}} \right)
\]
and therefore
\[
\sqrt{\frac{n}{\log n}} \left( \left( L_{a_n, b_n}^{n, r} - \mathbb{E}(L_{a_n, b_n}^{n, r}) \right) - \left( \tilde{L}_{a_n, b_n}^{n, r} - \mathbb{E}(\tilde{L}_{a_n, b_n}^{n, r}) \right) \right) \to 0 \quad (26)
\]
in probability as \(n \to \infty\).

The second region we consider is the one between the levels \(a_n\) and \(b_n\) with
\[
a_n = \frac{n}{(\log n)^2} \quad \text{and} \quad b_n = n^{1/2}.
\]

We put together the coupling for the two regions by taking the starting distribution for the second region to be the distribution of the chain at the end of the first region. Again from (25) we get that
\[
L_{a_n, b_n}^{n, r} - \mathbb{E}(L_{a_n, b_n}^{n, r}) = \tilde{L}_{a_n, b_n}^{n, r} - \mathbb{E}(\tilde{L}_{a_n, b_n}^{n, r}) + \mathcal{O}_P \left( \frac{1}{\sqrt{n}} \right)
\]
and therefore
\[
\sqrt{\frac{n}{\log n}} \left( \left( L_{a_n, b_n}^{n, r} - \mathbb{E}(L_{a_n, b_n}^{n, r}) \right) - \left( \tilde{L}_{a_n, b_n}^{n, r} - \mathbb{E}(\tilde{L}_{a_n, b_n}^{n, r}) \right) \right) \to 0 \quad (27)
\]
in probability as \(n \to \infty\).

For the region in the coalescent between the levels \(n^{1/2}\) and 1 we claim that
\[
\mathbb{E} \left( \sqrt{\frac{n}{\log n}} \cdot L_{n^{1/2}, 1}^{n, r} \right) \to 0 \quad (28)
\]
and
\[
\mathbb{E} \left( \sqrt{\frac{n}{\log n}} \cdot \tilde{L}_{n^{1/2}, 1}^{n, r} \right) \to 0 \quad (29)
\]
as \(n \to \infty\). The second claim follows directly from Proposition 3 in [JK11], whereas for (28) we get similarly using Lemma 4 that
\[
\mathbb{E} (L_{a_n, b_n}^{n, r}) = \mathbb{E} \left( \sum_{b_n < k \leq a_n} W_k \cdot \frac{2}{k(k-1)} \right) \leq c \sum_{b_n < k \leq a_n} \frac{k^2}{n} \cdot \frac{1}{k(k-1)} \leq c \cdot \frac{a_n}{n},
\]

18
for some constant $c < \infty$. Therefore, setting $a_n = \frac{n}{2}$ and $b_n = 1$ in (30), we obtain our claim (28). Since both $\sqrt{\frac{n}{\log n}} \cdot L_{n,1}^{n,r}$ and $\sqrt{\frac{n}{\log n}} \cdot \tilde{L}_{n,1}^{n,r}$ are positive random variables, it follows from (28) and (29) respectively, that

$$\sqrt{\frac{n}{\log n}} \cdot L_{n,1}^{n,r} \to 0 \quad \text{and} \quad \sqrt{\frac{n}{\log n}} \cdot \tilde{L}_{n,1}^{n,r} \to 0$$

(31)

in probability as $n \to \infty$.

Writing

$$L_{n,1}^{n,r} = L_{n,1}^{n,r} + L_{n,1}^{n,r} + L_{n,1}^{n,r}$$

and

$$\tilde{L}_{n,1}^{n,r} = \tilde{L}_{n,1}^{n,r} + \tilde{L}_{n,1}^{n,r} + \tilde{L}_{n,1}^{n,r}$$

and using (26) - (29) and (31) we get the claim of the proposition and therefore our theorem is proved.

References

[Be09] Berestycki, N. (2009) Recent progress in Coalescent theory, *Ensaios Matematicos* Vol. 16, 1-193

[BBS07] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2007) Beta-coalescents and continuous stable random trees *Ann. Probab.* 35, 1835-1887

[CNKR07] Caliebe, A., Neininger, R., Krawczak, M. and Rösl, U. (2007) On the length distribution of external branches in coalescent trees, *Genetic diversity within species. Theor. Population Biology* 72, 245-252

[DY12] Dhersin, J.-S. and Yuan, L. (2012) Asymptotic behavior of the total length of external branches for Beta-coalescents, [arXiv:1202.5859](https://arxiv.org/abs/1202.5859)

[Du08] Durrett, R. (2008) Probability Models for DNA Sequence Evolution *Springer*

[FM09] 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 Proc. Rel. Fields.* 15, 387-416

[FL93] Fu, Y.X. and Li, W.H. (1993) Statistical tests of neutrality of mutations, *Genetics* 133, 693-709

[Fu95] Fu, X.Y. (1995) Statistical properties of segregating sites, *Theor. Pop. Biol.* 48, 172-197

[GIM08] Gnedin, A., Iksanov, A. and Möhle, M. (2008) On asymptotics of exchangeable coalescents with multiple collisions, *J. Appl. Probab.* 45, 1186-1195

[JK11] Janson, S., Kersting, G. (2011) On the total external length of the Kingman coalescent, *Electron. J. Probab.* 16, 2203-2218

[KPSJ13] Kersting, G., Pardo, J.C., Siri-Jégousse, A. (2013) Total internal and external lengths of the Bolthausen-Sznitman coalescent. [arXiv:1302.1463](https://arxiv.org/abs/1302.1463)
[KSW12] Kersting, G., Stanciu, I. and Wakolbinger, A. (2012) The total external branch length of Beta-coalescents, arXiv:1212.6070v1

[Ki82] Kingman, J. F. C. (1982) The coalescent, Stoch. Proc. Appl. 13, 235-248

[LPW09] Levin, D., Peres, Y., Wilmer, E. (2009) Markov Chains and mixing times, American Mathematical Society

[Moe10] M¨ohle, 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

[Wa08] Wakeley, J. (2008) Coalescent Theory: An Introduction, Roberts & Company.