The total external length of the evolving Kingman coalescent

Iulia Dahmer,† Götz Kersting†

December 23, 2014

Abstract

The evolving Kingman coalescent is the tree-valued process which records the time changes undergone by the genealogies of Moran populations. We consider the associated process of total external tree length of the evolving Kingman coalescent and its asymptotic behaviour when the number of leaves of the tree tends to infinity. We show that on the time scale of the Moran model slowed down by a factor equal to the population size, the (centred and rescaled) external length process converges to a stationary Gaussian process with almost surely continuous paths and covariance function \( c(s, t) = \left( \frac{s + t}{2} \right)^2 \). A key role in the evolution of the external length is played by the internal lengths of finite orders in the coalescent at a fixed time which behave asymptotically in a multivariate Gaussian manner (see [DK13]). A natural coupling of the Moran model to a critical branching process is used.

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

Key words and phrases. evolving Kingman coalescent, external length process, Gaussian process, coupling, critical branching process

1 Introduction and main result

In mathematical population genetics the Kingman coalescent is a classical model for describing the genealogies of populations. If their population size is equal to \( n \), the \( n \)-Kingman coalescent can be graphically represented as a binary tree which starts with \( n \) leaves and spends an exponential time \( X_k \) with parameter \( \left( \frac{k}{2} \right) \) having \( k \) branches. The inter-coalescence times are independent. When labelling the leaves of the tree by \( 1, \ldots, n \) one can define the \( n \)-Kingman coalescent as a partition-valued process started in the partition \( \pi_n = \{ \{1\}, \ldots, \{n\} \} \) of \( \{1, \ldots, n\} \) into singletons with the property that, when it is in a state \( \pi_k \), it jumps after the exponential time \( X_k \) to a state obtained by merging two randomly chosen blocks from \( \pi_k \).

As time runs the population evolves and its genealogy changes, giving rise to a tree-valued process known as the evolving Kingman coalescent. We consider in this paper populations started at remote past, so to say at time \( -\infty \), and driven by the Moran model. This is a stationary continuous time evolution model in which each pair of individuals from the population is picked at rate 1, at which moment one of the individuals dies and the other one gives birth to an offspring.

The evolving coalescent discloses features of the Kingman coalescent which are hardly visible in the static model. It is worth noting that some aspects of the tree arise from the recent past and others from the more distant past and that for certain functionals of the tree it may not be clear which one of these contributions dominates. As we shall see, the influence of the recent and distant past is reflected among others in different time-scales, namely the evolutionary time-scale and the generations time-scale. We come back to this in more detail below.

*Work partially supported by the DFG Priority Programme SPP 1590 “Probabilistic Structures in Evolution”. I. D. was partially supported by the German Academic Exchange Service (DAAD).
†Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany.
dahmer@math.uni-frankfurt.de, kersting@math.uni-frankfurt.de
Particular functionals of coalescent trees such as the total branch length (the sum of the lengths of all the branches of the tree) and the total external length (the sum of the lengths of the external branches) have been extensively investigated in the literature due to their statistical relevance in population models including mutations. These are modelled as points of a Poisson process with constant rate on the branches of the coalescent tree. The total tree length controls the total number of mutations seen in the population whereas the total external length controls the number of mutations that affect only single individuals.

In this paper we investigate the asymptotic behaviour of the total external length process of the evolving Kingman-coalescent. The next theorem states our main result.

**Theorem 1.** Let $L^n_t$ be the external length of the evolving Kingman $n$-coalescent at time $t \in \mathbb{R}$. Then, as $n \to \infty$

$$\left(\frac{n}{4 \log n} (L^n_t - 2)\right)_{t \in \mathbb{R}} \xrightarrow{f.d.d.} \mathcal{L},$$

where the limiting process $\mathcal{L} = (\mathcal{L}_h)_{h \in \mathbb{R}}$ is stationary, Gaussian, a.s. continuous, with covariance function

$$\text{COV}(\mathcal{L}_0, \mathcal{L}_h) = \left(\frac{2}{2 + h}\right)^2, h \geq 0.$$ 

For the one-dimensional distribution this was obtained by Janson and Kersting [JK11].

Let us contrast this result with the theorem of Pfaffelhuber, Wakolbinger and Weissnau [PWW11], who investigated the process $(L^n_t)_{t \in \mathbb{R}}$ of total tree length of the evolving Kingman $n$-coalescent:

**Theorem 2.** There exists a stationary process $L = (L_t)_{t \in \mathbb{R}}$ with paths in $\mathcal{D}$, the space of càdlàg functions equipped with the Skorokhod topology, such that

$$L^n - 2 \log n \xrightarrow{d} L \quad \text{as } n \to \infty.$$ 

$L_t$ has a Gumbel distribution for all $t$. Moreover

$$\frac{1}{t |\log t|} \mathbb{E}((L_t - L_0)^2) \to 4 \quad \text{as } t \to 0.$$ 

Let us compare these two theorems. Theorem 2 deals with the original time-scale from the Moran model, which we call the evolutionary time-scale. On the contrary, the time in Theorem 1 runs on the evolutionary time-scale \textit{slowed down} by a factor of $n$. This we call the generations time-scale. Note that on the evolutionary time scale reproduction events happen in the population with a rate of order $n^2$. Therefore, for large populations, after a time of order $\frac{1}{n}$ each individual will have taken part in a number of reproduction events of order 1. It is not surprising that the external length has to be considered on this time scale. However, it is less obvious that the evolutionary time-scale is the appropriate choice for the total length.
We use the notation $t$ for the time on the evolutionary time-scale and

$$h = h(n) := \frac{t}{n},$$

for the time points of the generations time-scale.

Note also that the limiting process $L$ is almost surely continuous, whereas $L$ is almost surely made up of jumps. This reflects the fact that the total length experiences big jumps (at the times when old families become extinct). For the external length such extremal events do not come into play. This is also reflected in the type of limiting distributions. Let us point out that the process $L$ is not a semimartingale ([DKW14b]), which so far is an open problem for the process $L$.

The Kingman coalescent is one of a big class of coalescent models which have been studied in the literature. The total length and the external length have been extensively studied for the Beta-coalescents. For the asymptotics in the static case we refer to [DIMR07], [IM07], [Moe10], [BBS08, BBS07], [BBL12], [DY12], [Ke12] and [DKW14a]. In the dynamic case, the evolving Beta$(2 - \alpha, \alpha)$-coalescent was introduced and investigated for $\alpha = 1$ (the Bolthausen-Sznitman coalescent) in [Sch12] and for $1 < \alpha < 2$ in [KSW14]. In the latter case the processes of (centred and rescaled) total tree length (for $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$) and of total external length (for $1 < \alpha < 2$) converge in the sense of the finite dimensional distributions to stationary moving average processes with stable distributions. In contrast to the Kingman case, the time-scale is the generations time-scale for both the total length process and the external length process, which here is obtained by slowing down time by a factor of $n^\alpha - 1$. Evolving coalescents have also been investigated from a different point of view, namely as evolving metric spaces in (e.g.) [DGPT12], [GPW09], [GPW13], [Gu14].

In the rest of this introduction we overview the proof of our theorem.

For describing the dynamics of the total external length of the Kingman coalescent we recall the notion of branches of order $i$ introduced in [DK13]. In the coalescent tree each branch is situated above a subtree. If this subtree has $i$ leaves, one says that the branch is of order $i$. If $i = 1$ the branches are external and if $i \geq 2$ they are internal. We denote by $L_i^n$ the total length of order $i$ (the sum of the lengths of branches of order $i$) of the coalescent of the population alive at time $t$.

Let us now look at the dynamics of the total external length. On the generation time scale each pair of individuals takes part in a reproduction event at rate $\frac{1}{n}$. As long as no reproduction event takes place, the external branches of the coalescent tree grow linearly and hence the external length grows at rate 1. At the time $\tau$ of a reproducing event, the total external length process has a jump. The length of the branch corresponding to the individual that dies and the length of the branch corresponding to the individual that reproduces are subtracted from the total external length. The first branch is removed from the tree, whereas the latter becomes an internal branch of order 2. Typically, by the removal of the branch corresponding to the individual that dies, a branch that was internal (of order 2) at time $\tau$ becomes part of an external branch at time $\tau$. We say that the internal branch was freed. Its length is then added to the total external length. An example is given in Figure 2.

Observe that the external length process is stationary, a property that it inherits from the stationarity of the Moran model. It thus suffices to analyse the dynamics of the external length on a time interval $(0, h)$ for $h > 0$.

From the tree at time 0 not only internal branches of order 2 may be freed by time $h$, but also internal branches of orders greater than 2. The lengths of the freed branches become part of the external length at time $h$ and hence $L^n_h = L^n_{h,1}$ can be written as

$$L^n_h = \sum_{i=1}^{n-1} \Lambda^{n,i}_h L^n_{i,0} + I^n_{0,h},$$

where by $\Lambda^{n,i}_h = \Lambda^{n,i}_{0,h}$ we denote the (random) proportion of the length $L^n_{i,0}$ of order $i$ that is freed by time $h$ and by $I^n_{0,h}$ the length added up in the time interval $(0, h)$. 

3
Figure 2: In Figure a) the external branches of the coalescent at time $h_1$ are marked in (dashed) blue. At time $\tau$ a reproduction event happens and the tree changes. The external branches at time $h_2$ are shown in Figure b) with the following code: the (dashed) blue pieces make up the part of the external length from time $h_1$ that is still external at time $h_2$, the red (dash and dot) piece is a branch that was internal at time $h_1$ and was freed at time $\tau$, while the orange (dotted) pieces represent the external length $I_{h_1, h_2}^r$ gathered in the time interval $(h_1, h_2)$.

A key role in the analysis of the external length process is played by the main result of [DK13] which states that for any $r \in \mathbb{N}$, as $n \to \infty$

$$\sqrt{\frac{n}{4 \log n}} \left( L_{h_1}^{n,1} - \eta_1, \ldots, L_{h_1}^{n,r} - \eta_r \right) \xrightarrow{d} N(0, I_r),$$

where $I_r$ denotes the $r \times r$-identity matrix and

$$\eta_i = \mathbb{E}(L_{h_1}^{n,i}) = \frac{2}{i} \quad \text{for every } i \geq 1. \quad (2)$$

Let us label the individuals alive at time 0 by $1, \ldots, n$ and denote by $\mathcal{P}_n$ the power set of $\{1, \ldots, n\}$. For $A \in \mathcal{P}_n$ denote

$L_A :=$the length of the branch supporting the leaves with labels in $A$ in the coalescent at time 0, with the convention $L_A = 0$ if there is no such branch in the coalescent and

$\mathcal{F}_A := \{\text{the branch supporting the leaves with labels in } A \text{ at time 0 is free at time } h\}$.

Observe that with this notation the $i$-th summand on the right-hand side of (1) can be rewritten as

$$\Lambda^{n,i} L_0^{n,i} = \sum_{A \in \mathcal{P}_n \atop |A|=i} \mathbf{1}_{\mathcal{F}_A} \cdot L_A$$

and that for any fixed $r \in \mathbb{N}$ we can write (1) as

$$L_h^n = \sum_{A \in \mathcal{P}_n \atop |A| \leq r} \mathbf{1}_{\mathcal{F}_A} \cdot L_A + R_h^{n,r} + I_{0, h}^r, \quad (3)$$

where

$$R_h^{n,r} := \sum_{A \in \mathcal{P}_n \atop |A| > r} \mathbf{1}_{\mathcal{F}_A} \cdot L_A \quad (4)$$
records the contribution of the lengths of the branches of orders larger than \( r \) to the external length at time \( h \).

The proof of Theorem 1 will show that for large \( r \) it is only the sum on the right-hand side of (3) that makes a non-negligible contribution to the external length at time \( h \). Note moreover that this sum is obtained by a random thinning (by means of the indicators \( \mathbf{1}_{A_{A_k}} \)) of the lengths \( L_{n,i}^{r} \) of order at most \( r \). In Section 2 we prove a general result, which assures that if a sum of (possibly dependent) random variables which converges in distribution to a normal-distributed random variable, is complemented with independent random coefficients, the sum obtained through this procedure continues to be asymptotically Gaussian. The result of [13] provides the asymptotic normality assumption on the initial sums \( L_{n,i}^{r} \).

Another key idea of the paper is that the free-mechanism described above can be coupled in a natural way to a birth and death process started with \(|A|\) individuals at time 0. This allows to approximate the thinning procedure by an independent one. The coupling to the birth and death process is presented in Section 3 whereas the proof of Theorem 1 is given in Section 4 after further preparatory work in Sections 2 and 3.

Notation. We will use the Vinogradov notation:

\[ a_n \ll b_n \text{ if there exists a finite constant } c \text{ independent of } n \text{ such that } |a_n| \leq cb_n. \]

2 Asymptotic normality of sums with random coefficients

The following Proposition contains a general statement of the form: If the sum of not necessarily independent random variables is asymptotically normal then this property persists if the summands are supplied with independent random coefficients.

Proposition 1. Let \( r \in \mathbb{N} \) be fixed and let \( m_1(n), \ldots, m_r(n) \in \mathbb{N} \) be such that \( \lim_{n \to \infty} m_i(n) = \infty \) for each \( 1 \leq i \leq r \). Let \( Y_{i,j}, 1 \leq i \leq r, 1 \leq j \leq m_i(n) \) be random variables such that in \( \mathbb{R}^r \)

\[
\left( \sum_{j=1}^{m_i(n)} Y_{i,j} \right)_{i=1, \ldots, r} \overset{d}{\to} N(0, I_r).
\]

Also let \( U_{i,j}, 1 \leq i \leq r, 1 \leq j \leq m_i(n) \) be independent of one another and independent of the collection \( \{Y_{i,j}\}_{1 \leq i \leq r, 1 \leq j \leq m_i(n)} \), with the property that for fixed index \( i \) the variables \( U_{i,j}, 1 \leq j \leq m_i(n) \) are identically distributed with second moment \( \lambda_i \). If the following two conditions hold: for each \( 1 \leq i \leq r \)

\[
\sum_{j=1}^{m_i(n)} Y_{i,j}^2 \overset{p}{\to} 1 \quad \text{and} \quad \max_{1 \leq j \leq m_i(n)} |Y_{i,j}| \overset{p}{\to} 0
\]

then

\[
\sum_{i=1}^{r} \sum_{j=1}^{m_i(n)} U_{i,j} Y_{i,j} \overset{d}{\to} N(0, \sum_{i=1}^{r} \sum_{m_i(n)}).
\]

Proof. The characteristic function of the sum \( \sum_{i=1}^{r} \sum_{j=1}^{m_i(n)} U_{i,j} Y_{i,j} \) is

\[
\varphi_n(\lambda) = \mathbb{E}\left( \mathbb{E}\left( \exp \left( i \lambda \sum_{i=1}^{r} \sum_{j=1}^{m_i(n)} U_{i,j} Y_{i,j} \right) \mid Y_{i,j}, i \leq r, j \leq m_i(n) \right) \right) = \mathbb{E}\left( \prod_{i=1}^{r} \prod_{j=1}^{m_i(n)} \varphi_{U_{i,j}}(\lambda Y_{i,j}) \right). \quad (5)
\]

Using the fact that for complex numbers \( \{a_k\}_{1 \leq k \leq n} \) and \( \{b_k\}_{1 \leq k \leq n} \) with \( |a_k| \leq 1 \) and \( |b_k| \leq 1 \) for all \( k \) it holds that

\[
\left| \prod_{k=1}^{n} a_k - \prod_{k=1}^{n} b_k \right| \leq \sum_{k=1}^{n} |a_k - b_k|
\]
we obtain by denoting the expectation and variance of \( U_{i,j} \) by \( \mu_i \) and \( \sigma_i \) respectively that
\[
\left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp \left( i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 \right) \right| \leq \sum_{i,j} \left| \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \exp \left( i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 \right) \right|
\]
and by using a Taylor expansion that
\[
\left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp \left( i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 \right) \right| \leq \sum_{i,j} \left| 1 + i \mu_i \lambda Y_{i,j} - \frac{1}{2} m_i \lambda^2 Y_{i,j}^2 + o(|Y_{i,j}|^2) \right| - \left( 1 + i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 + \frac{h_i^2}{2} \lambda^2 Y_{i,j}^2 + O(|Y_{i,j}|^3) \right) \right| = \sum_{i,j} o(|Y_{i,j}|^2).
\]

By assumption \( \sum_{i,j} o(|Y_{i,j}|^2) \) converges in probability to 0 as \( n \to \infty \). It follows by dominated convergence that
\[
\mathbb{E} \left( \left| \prod_{i,j} \varphi_{U_{i,j}}(\lambda Y_{i,j}) - \prod_{i,j} \exp \left( i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2 \right) \right| \right) \to 0. \quad (6)
\]

From the assumptions on the collection of random variables \( Y_{i,j} \) \( 1 \leq i \leq r, 1 \leq j \leq m_i(n) \), it follows that \( \left( \sum_j Y_{i,j}, \sum_j Y_{i,j}^2 \right) \) converges in distribution and we obtain from (5) and (6) that
\[
\lim_{n \to \infty} \varphi^n(\lambda) = \lim_{n \to \infty} \mathbb{E} \left( \prod_{i=1}^r \prod_{j=1}^{m_i(n)} \exp(i \mu_i \lambda Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 Y_{i,j}^2) \right) \]
\[
= \lim_{n \to \infty} \mathbb{E} \left( \exp(i \lambda \sum_{i=1}^r \mu_i \sum_{j=1}^{m_i(n)} Y_{i,j} - \frac{\sigma_i^2}{2} \lambda^2 \sum_{i=1}^r \sum_{j=1}^{m_i(n)} Y_{i,j}^2) \right) \]
\[
= e^{-\frac{\lambda^2}{2} \sum_{i=1}^r \mu_i^2} \cdot e^{-\frac{\lambda^2}{2} \sum_{i=1}^r \sum_{j=1}^{m_i(n)} \sigma_i^2} \]
\[
= e^{-\frac{\lambda^2}{2} \sum_{i=1}^r m_i}.
\]

Since the right hand side is the characteristic function of the normal distribution with expectation 0 and variance \( \sum_{i=1}^r m_i \) the proof of the proposition is finished. \( \square \)

## 3 Coupling the Moran model to a critical branching process

In the time interval \( (0, \hbar) \) on the generations time-scale the individuals in the population are subject to binary reproduction events and death events according to the Moran dynamics with rate \( \frac{1}{n} \) per pair. In this section we give the description of the evolution of the number of descendants of the individuals with labels in a set \( A \in \mathcal{P}_n \) at time 0 in terms of a linear birth and death process. Similar ideas have been exploited e.g. in [PP13]. We call the descendants of the individuals with labels in \( A \) at time 0, the \( A \)-family and we are interested in the size of the family (number of individuals at each time point). We will then see how this description can then be extended to the pair of family sizes for two sets \( A \) and \( A' \) with \( A \cap A' = \emptyset \).

We start by fixing \( A \in \mathcal{P}_n \) with \( |A| = i \), \( 1 \leq i \leq n-1 \), and define
\[
Z := Z(A) = (Z_h)_{h \geq 0}
\]
to be the process recording the size of the \( A \)-family as time evolves. Then \( Z_0 = i \) and according to the Moran dynamics of the population, the process \( Z \) is a continuous time Markov process with jump rates
\[
r_{k,k'} = \begin{cases} \frac{1}{2} k (1 - \frac{k}{n}) & \text{if } k' = k + 1 \\ \frac{1}{2} k (1 - \frac{k}{n}) & \text{if } k' = k - 1. \end{cases}
\]
Indeed, it is only the reproduction events that happen between a member of the $A$-family and a non-$A$-family member, that affect the size $Z$ of the $A$-family. There are $k(n-k)$ such pairs and the reproduction events arise at rate $\frac{1}{n}$. The family size may either decrease by one or increase by one at the time of such a reproduction event, depending on whether the individual from the $A$-family dies or reproduces. The reproduction events between two $A$-family members or between two non-$A$-family members leave $Z$ unchanged. Note that $n$ and 0 are absorbing states for the process $Z$.

Let now $B = (B_h)_{h \geq 0}$ be a birth and death process started in $B_0 = i$ and having birth and death rates both equal to $\frac{1}{2}k$ when the process is in state $k$. Then the process $Z$ can be obtained from $B$ by the random time change (see [EK86] Section 6.1):

$$Z_h = B_{\theta_h}, \quad \text{with } \theta_h \text{ defined by } \begin{cases} h = \int_0^{\theta_h} \frac{1}{1-Bu} du, & \text{if a solution of this equation exists} \\ \theta_h = \theta, & \text{otherwise}, \end{cases}$$

where $\theta := \inf\{u \geq 0 : B_u = n\}$. This may also be expressed as

$$\theta_h = \int_0^h \frac{n-B_{\theta_s}}{n} ds \quad (7)$$

for all $h \geq 0$, as can be seen by taking derivatives. Note that $\theta_h < h$ for $h > 0$, thus $Z$ can be seen as a (random) slowed-down version of the birth and death process $B$.

Before further analysing the relation between the processes $B$ and $Z$, we compute a few basic quantities connected to $B$, that we will use in the sequel. Let

$$F(s, h) = \sum_{k=0}^{\infty} s^k \mathbb{P}(B_h = k \mid B_0 = 1)$$

denote the generating function of the process $B$. Then (see [AN72], Chapter III) $F$ is the unique solution of the backward equation

$$\frac{\partial}{\partial h} F(s, h) = \frac{1 + F^2(s, h)}{2} - F(s, h) = \frac{1}{2} (1 - F(s, h))^2$$

with boundary condition $F(s, 0) = s$. Integrating the equation gives

$$F(s, h) = 1 - \frac{1 - s}{1 + \frac{h}{2} \cdot s}, \quad (8)$$

which can be rewritten as

$$F(s, h) = 1 - \frac{1 - s}{1 + \frac{h}{2}} \cdot \sum_{j=0}^{\infty} \left( \frac{h}{h+2} \cdot s \right)^j.$$

This gives

$$\mathbb{P}(B_h = 0 \mid B_0 = 1) = \frac{h}{h+2} \quad \text{and} \quad \mathbb{P}(B_h = 1 \mid B_0 = 1) = \left( \frac{2}{h+2} \right)^2. \quad (9)$$

For our process $B$ started in $B_0 = i$ it follows that

$$p_{i,h} := \mathbb{P}(B_h = 1 \mid B_0 = i) = i \cdot \left( \frac{h}{h+2} \right)^{i-1} \left( \frac{2}{h+2} \right)^2. \quad (10)$$

Observe that for fixed $h$ the generating function $F$ given in (8) and all its derivatives exist for all $s < \frac{h^2}{h}$ and therefore all moments of $B_h$ exist and are finite functions of $h$. If the process $B$ starts in $B_0 = i$, then for $d \in \mathbb{N}$

$$\mathbb{E}_i \left( (B_h)^d \right) = \mathbb{E}_1 \left( \left( \sum_{j=1}^{i} (B_h(j)) \right)^d \right).$$

7
where by $B_h(j)$ we denoted the number of descendants at time $h$ of the $j$-th individual from the population at time 0. Expanding the sum it follows that

$$
E_i \left( (B_h)^d \right) \leq c_h \cdot i^d,
$$

(11)

for a constant $c_h$ depending on $h$. Moreover for $B_0 = 1$ it holds that

$$
E_1(B_h) = 1 \quad \text{and} \quad E_1(B_h^2) = h + 1.
$$

Now we come back to the connection with the process $Z$. Given the event that in the coalescent at time 0 there exists a branch that supports the leaves with labels in $A$, the event $\mathcal{F}_A$ that this branch is free at time $h$ can be written as

$$
\mathcal{F}_A = \{ Z_h = 1 \} = \{ B_{\theta_h} = 1 \}.
$$

We will use the notation

$$
p^n_{i,h} := P(Z_h = 1 \mid Z_0 = i) = P(\mathcal{F}_A).
$$

**Lemma 1.** The following bounds hold

$$
P_i(Z_h = 1) \leq 2i \cdot \left( \frac{h}{h+2} \right)^{i-1}
$$

(12)

and

$$
P_i(B_h = 1, Z_h \neq 1) + P_i(Z_h = 1, B_h \neq 1) \leq c \cdot \frac{1}{n} \cdot \gamma^i
$$

(13)

for some $\gamma < 1$ and $c < \infty$ depending only on $h$.

**Proof.** Recall $|A| = i$ and define the random variable

$$
N(B, h) := \text{number of individuals in A at time 0 that evolve according to B and have 0 descendants at time } h.
$$

Observe that since $B_0 = i$ and $\theta_h \leq h$

$$
\{ Z_h = 1 \} = \{ B_{\theta_h} = 1 \} \subset \{ N(B, \theta_h) = i - 1 \} \subset \{ N(B, h) \geq i - 1 \}.
$$

Using (9) we thus obtain that

$$
P_i(Z_h = 1) \leq i \cdot \left( \frac{h}{h+2} \right)^{i-1} + \left( \frac{h}{h+2} \right)^{i} \leq 2i \cdot \left( \frac{h}{h+2} \right)^{i-1},
$$

which gives the first claim of the lemma.

For the second claim we start with the first term on the left-hand side of (13). We define the $\sigma$-algebra

$$
\mathcal{F}_{\theta_h} := \sigma(B_s, s \leq \theta_h).
$$

Then by (10) and Hölder’s inequality

$$
P_i(B_h = 1, Z_h \neq 1) = E_i \left( P_{Z_h}(B_h - \theta_h = 1 \mid \mathcal{F}_{\theta_h}); Z_h \neq 1 \right)
$$

$$
\leq E_i \left( Z_h \left( \frac{h - \theta_h}{h - \theta_h + 2} \right)^{Z_h - 1}; Z_h \geq 2 \right)
$$

$$
\leq E_i \left( (h - \theta_h)Z_h \left( \frac{h - \theta_h + 2}{h + 2} \right)^{Z_h - 2} \right)
$$

$$
\leq E_i \left( (h - \theta_h)^3 \right)^{1/3} \cdot E_i \left( Z_h^3 \right)^{1/3} \cdot E_i \left( \left( \frac{h}{h + 2} \right)^{3(Z_h) - 2} \right)^{1/3}.
$$
Note from (7) that for all \( h \geq 0 \)
\[
h - \theta_h = \int_0^h \frac{B_{\theta_s}}{n} ds.
\]
By using Jensen’s inequality and Fubini’s theorem we obtain that
\[
E_i \left( (h - \theta_h)^3 \right) = E_i \left( \left( \int_0^h \frac{B_{\theta_s}}{n} ds \right)^3 \right) \leq \frac{1}{n^3} E_i \left( h^3 \int_0^h \frac{B_{\theta_s}^3}{h} ds \right) = \frac{h^2}{n^3} \int_0^h E_i (B_{\theta_s}^3) ds.
\]
Since the function \( x \mapsto x^3 \) is convex for \( x \geq 0 \) and the process \( B \) is a martingale, it follows that the process \( B^3 \) is a submartingale. Now \( \theta_h \) is a stopping time, \( \theta_h \leq h \) and therefore by (11) it holds that
\[
E_i (B_{\theta_h}^3) \leq c \cdot i^3
\]
and thus
\[
E_i \left( (h - \theta_h)^3 \right) \leq c \cdot \frac{1}{n^3} \cdot i^3.
\]
for \( c \) a finite constant depending on \( h \).

Using the same argument as above for the convex function \( x \mapsto x^3 \) we obtain that
\[
E_i \left( \left( \frac{h}{h+2} \right)^3 (Z_h - 2) \right) = E_i \left( \left( \frac{h}{h+2} \right)^3 (B_{\theta_h} - 2) \right) \leq E_i \left( \left( \frac{h}{h+2} \right)^3 (B_{\theta_h} - 2) \right)
\]
and moreover
\[
E_i \left( \left( \frac{h}{h+2} \right)^3 (Z_h - 2) \right) \leq \left( \frac{h+2}{h} \right)^6 \cdot F \left( \left( \frac{h}{h+2} \right)^3, h \right) i
\]
where \( F \) is the generating function of the process \( B \). Now using the fact that \( F(s,h) < 1 \) for \( s < 1 \) we obtain by plugging (15), (16) and (17) in (14) that
\[
\mathbb{P}_i (B_h = 1, Z_h \neq 1) \leq c \cdot \frac{1}{n} \cdot i^2 \cdot \gamma^i
\]
for some \( \gamma < 1 \) and \( c \leq \infty \) depending on \( h \). By further enlarging \( \gamma \) we can drop the factor \( i^2 \) and thus obtain that
\[
\mathbb{P}_i (B_h = 1, Z_h \neq 1) \leq c \cdot \frac{1}{n} \cdot \gamma^i
\]
(18)
For the second term on the left-hand side of (13) observe that
\[
\mathbb{P}_i (Z_h = 1, B_h \neq 1) \leq \mathbb{P}_i (Z_h = 1, \exists \text{ at least one birth or death event in } (\theta_h, h))
\]
\[
= \mathbb{E}_i \left( \mathbb{P}_i \left( \exists \text{ at least one birth or death event in } (\theta_h, h) \mid \mathcal{F}_{\theta_h}; Z_h = 1 \right) \right)
\]
\[
= \mathbb{E}_i \left( 1 - e^{-(h-\theta_h) \cdot Z_h = 1} \right)
\]
\[
\leq \mathbb{E}_i \left( h - \theta_h; Z_h = 1 \right)
\]
\[
\leq \mathbb{E}_i ((h - \theta_h)^2)^{1/2} \cdot \mathbb{P}_i (Z_h = 1)^{1/2}.
\]
Using (12), (16) and Jensen’s inequality we obtain that
\[
\mathbb{P}_i (Z_h = 1, B_h \neq 1) \leq c \cdot \frac{1}{n} \cdot \gamma^i
\]
for some \( \gamma < 1 \) and \( c \leq \infty \), which finishes the proof. \( \square \)
A direct consequence of the lemma is that

$$|p_{i, h}^0 - p_{i, h}| = \left| \mathbb{P}_i(Z_h = 1) - \mathbb{P}_i(B_h = 1) \right| = \left| \mathbb{P}_i(Z_h = 1, B_h \neq 1) - \mathbb{P}_i(B_h = 1, Z_h \neq 1) \right| \leq \mathbb{P}_i(Z_h = 1, B_h \neq 1) + \mathbb{P}_i(B_h = 1, Z_h \neq 1) \leq c \cdot \frac{1}{n} \cdot \gamma^i$$

for some $\gamma < 1$ and $c < \infty$.

Let us now consider another set $A' \in \mathcal{P}_n$ with cardinality $i'$ and such that $A \cap A' = \emptyset$. We denote by $Z = (Z_h)_{h \geq 0}$ and $Z' = (Z'_h)_{h \geq 0}$ the processes recording the sizes of the $A$-family and the $A'$-family respectively. We will be interested in the process

$$Z := (Z(A), Z'(A')) = (Z_h, Z'_h)_{h \geq 0}.$$

The process $Z$ is Markov, it starts in the state $Z_0 = (i, i')$ and evolves as follows: jumps happen at the times of reproduction events that involve an individual from the $(A \cup A')$-family and an individual from the $(\{1, \ldots, n\} \setminus (A \cup A'))$-family (we call these events $\alpha$-events) and at the times of the reproduction events involving one individual from the $A$-family and one individual from the $A'$-family (we call these events $\beta$-events). Therefore the process $Z$ jumps at the following rates:

$$\rho(k, k', (j, j')) = \begin{cases} \frac{1}{2} k (1 - \frac{k + k'}{n}) & \text{if } j = k + 1, j' = k' \\ \frac{1}{2} k (1 - \frac{k + k'}{n}) & \text{if } j = k - 1, j' = k' \\ \frac{1}{2} k' (1 - \frac{k + k'}{n}) & \text{if } j = k, j' = k' + 1 \\ \frac{1}{2} k' (1 - \frac{k + k'}{n}) & \text{if } j = k, j' = k' - 1 \\ \frac{1}{2} \frac{kk'}{n} & \text{if } j = k + 1, j' = k' - 1 \\ \frac{1}{2} \frac{kk'}{n} & \text{if } j = k - 1, j' = k' + 1. \end{cases}$$

The first four expressions refer to the $\alpha$-events and the last two to the $\beta$-events.

Let now $B = (B_h)_{h \geq 0}$ and $B' = (B'_h)_{h \geq 0}$ be two independent birth and death processes started in $B_0 = i$ and $B'_0 = i'$ respectively, each having birth and death rates both equal to $\frac{1}{2} k$ when the process is in state $k$. The process $Z$ can be constructed starting from the processes $B$ and $B'$ in a similar way as we did before, but one additional step will be needed in the construction in order to include the $\beta$-events.

Let first $Y = (Y, Y')$ be the process obtained by the random time change of $B := (B, B')$ defined by:

$$\mathcal{Y}_h = B_{\tau_h}, \quad \text{with } \tau_h \text{ defined by} \begin{cases} h = \int_0^{\tau_h} \frac{1}{1 - 2u - \rho} \, du & \text{if a solution of this equation exists} \\ \tau_h = \tau & \text{otherwise,} \end{cases}$$

(20)

where $\tau := \inf\{u \geq 0 : B_u + B'_u = n\}$. Again by taking derivatives we have that

$$\tau_h = \int_0^{\tau_h} \frac{n - B_{\tau_u} - B'_{\tau_u}}{n} \, du.$$ 

Since $\tau_h < h$, $\mathcal{Y}$ can be seen as a slowed-down version of $B$. Observe that the process $\mathcal{Y}$ has the same distribution as the process $Z$ in which only the $\alpha$-events are considered, the $\beta$-events being neglected. In order to further obtain $Z$ from $\mathcal{Y}$, one then needs to add to $\mathcal{Y}$ a process of $\beta$-events, such that the first $\beta$-event occurs at rate $\frac{kk'}{n}$ if the process $\mathcal{Y}$ is in the state $(k, k')$. The further evolution of the process then has to be Markovian with the above rates $\rho$. For the sequel it is not necessary to specify this evolution in more detail.
We now use the coupling of the processes \( Z \) and \( B \) described above in order to prove the following lemma.

**Lemma 2.** If \( A \cap A' = \emptyset \), then

\[
P_{i,i'}(B_h = 1, Z_h \neq 1, Z'_h = 1) + P_{i,i'}(Z_h = 1, B_h \neq 1, Z'_h = 1) \leq c \cdot \frac{1}{n} \cdot \gamma^{i + i'}.
\]

for some \( \gamma < 1 \) and \( c < \infty \) depending on \( h \).

**Proof.** Let \( E_{\beta,h} \) denote the event

\[
E_{\beta,h} := \{ \text{the first } \beta \text{-event takes place in } (0, h) \}
\]

and observe that

\[
\{B_h = 1, Z_h \neq 1, Z'_h = 1\} \subset \{B_h = 1, Z'_h = 1, E_{\beta,h}\} \cup \{B_h = 1, Y'_h = 1, Y_h \neq B_h\}
\]

and similarly for the other terms in (21). Therefore the left-hand side of (21) is bounded from above by

\[
P_{i,i'}(B_h = 1, Z_h = 1, E_{\beta,h}) + P(B_h = 1, Y'_h = 1, Y_h \neq B_h) \leq c \cdot \frac{1}{n} \cdot \gamma^{i + i'}.
\]

We start by evaluating the first term on the right hand side. Let us define \( J_h := \sigma(B_s, B'_s, s \leq h) \) and remember that given the event that \( \mathcal{Y}_t \) is in the state \((y, y')\), the first \( \beta \)-event happens at rate \( \frac{2\gamma}{n} \). Using Fubini’s theorem we obtain that

\[
P_{i,i'}(B_h = 1, Z'_h = 1, E_{\beta,h}) = E_{i,i'} \left( P(E_{\beta,h} \mid J_h); B_h = 1, Z'_h = 1 \right)
\]

\[
= E_{i,i'} \left( 1 - e^{-\int_0^h \frac{Y_s \cdot Y'_s}{n} ds}; B_h = 1, Z'_h = 1 \right)
\]

\[
\leq E_{i,i'} \left( \int_0^h (Y_s + Y'_s)^2 ds; B_h = 1, Z'_h = 1 \right)
\]

\[
= \frac{1}{n} \int_0^h E_{i,i'}((Y_s + Y'_s)^2; B_h = 1, Z'_h = 1) ds.
\]

Now, by Hölder’s inequality and using the definition of \( \mathcal{Y} \) given in (20) we obtain that

\[
P_{i,i'}(B_h = 1, Z'_h = 1, E_{\beta,h}) \leq \frac{1}{n} \int_0^h E_{i,i'}((Y_s + Y'_s)^6)^{1/3} P_i(B_h = 1)^{1/3} P_i(Z'_h = 1)^{1/3} ds
\]

\[
= \frac{1}{n} \int_0^h E_{i,i'}((B_{\tau_s} + B'_{\tau_s})^6)^{1/3} P_i(B_h = 1)^{1/3} P_i(Z'_h = 1)^{1/3} ds.
\]

Since the function \( x \mapsto x^6 \) is convex and \( B + B' \) is a martingale, it follows that the process \((B + B')^6\) is a submartingale. Now \( \tau_h \) is a stopping time, \( \tau_s \leq h \) for \( s \leq h \) and therefore it holds that \( \mathbb{E}(\tau_s) + B'_s) \leq \mathbb{E}(B_s + B'_s) \) for all \( s \leq h \). Thus

\[
P_{i,i'}(B_h = 1, Z'_h = 1, E_{\beta,h}) \leq \frac{1}{n} \int_0^h E_{i,i'}((B_h + B'_h)6)^{1/3} P_i(B_h = 1)^{1/3} P_i(Z'_h = 1)^{1/3} ds.
\]
We now use (11), (10) and Lemma 1 and obtain further that
\[
P_{i,i'}(B_h = 1, Z_h = 1, E_{i,h}) \ll \frac{1}{n} \cdot (i + i')^2 \cdot \left( i \cdot \left( \frac{h}{h+2} \right)^{i-1} \left( \frac{2}{h+2} \right)^2 \right)^{i/3} \cdot \left( 2i' \cdot \left( \frac{h}{h+2} \right)^{i'-1} \right)^{i/3}
\]

The same argument works for evaluating the third, fifth and seventh term on the right hand side of (22) and gives the same bound.

We now turn to the second term on the right-hand side of (22). We follow an idea from the proof of Lemma 1. Let \( H_{\tau_h} \) be the \( \sigma \)-algebra
\[
H_{\tau_h} := \sigma \left( B_s, s \leq \tau_h, B' \right).
\]

Then similarly to (14) (remember that here, in the two-dimensional context, we use the notation \((Y, Y')\) for the "slowed-down" birth-death process, whereas in the one-dimensional case this was denoted by \(Z\)) by means of (10)
\[
P_{i,i'}(B_h = 1, Y_h' = 1, Y_h \neq B_h) = E_{i,i'} \left( P_{i,h}(B_{h-\tau_h} = 1 \mid H_{\tau_h}); Y_h \neq 1, Y_h' = 1 \right)
\]
\[
\leq E_{i,i'} \left( (h - \tau_h)Y_h \left( \frac{h}{h+2} \right)^{Y_h-2}; Y_h' = 1 \right)
\]
\[
\leq E_{i,i'} \left( (h - \tau_h)^4 \cdot \mathbb{E}_i(Y_h^4)^{1/4} \cdot \mathbb{E}_i \left( \left( \frac{h}{h+2} \right)^{4(Y_h-2)} \right)^{1/4} \cdot P_{i'}(Y_h' = 1)^{1/4} \right).
\]

The arguments from above work for evaluating the fourth, sixth and eighth term on the right hand side of (22) by similarly incorporating the event \( \{B'_h = 1\} \) and gives the same bound. This finishes the proof. 

As a direct consequence of the previous two lemmas we obtain that for \( A \cap A' = \emptyset \) with \( |A| = i \) and \( |A'| = i' \)
\[
\left| \text{COV} \left( 1_{\mathcal{A}}, 1_{\mathcal{A}'} \right) \right| \leq c \cdot \frac{1}{n} \cdot \gamma^{i+i'},
\]
for some \( \gamma < 1 \) and \( c < \infty \) depending on \( h \).

To see this, observe that since the birth-death processes \( B \) and \( B' \) are independent, it holds that
\[
\text{COV} \left( 1_{\mathcal{A}}, 1_{\mathcal{A}'} \right) = \text{COV} \left( 1_{\{ Z_h = 1 \}}, 1_{\{ Z'_h = 1 \}} \right)
\]
\[
= \text{COV} \left( 1_{\{ B_h = 1 \}}, 1_{\{ B'_h = 1 \}} \right)
\]
\[
+ \text{COV} \left( 1_{\{ B_h = 1 \}}, 1_{\{ Z'_h = 1 \}} - 1_{\{ B'_h = 1 \}} \right) + \text{COV} \left( 1_{\{ Z_h = 1 \}} - 1_{\{ B_h = 1 \}}, 1_{\{ Z'_h = 1 \}} \right)
\]
\[
= E_{i,i'} \left( 1_{\{ B_h = 1 \}} \left( 1_{\{ Z'_h = 1 \}} - 1_{\{ B'_h = 1 \}} \right) \right) - P_{i'}(B_h = 1) E_{i'}(1_{\{ Z'_h = 1 \}} - 1_{\{ B'_h = 1 \}})
\]
\[
+ E_{i,i'} \left( 1_{\{ Z'_h = 1 \}} \left( 1_{\{ Z_h = 1 \}} - 1_{\{ B_h = 1 \}} \right) \right) - P_{i'}(Z'_h = 1) E_{i'}(1_{\{ Z_h = 1 \}} - 1_{\{ B_h = 1 \}})
\]
We start by making some notation. We fix a set \( A \in \mathcal{P}_n \) and view the coalescent tree of the population alive at time 0 from the leaves towards the root. The branch supporting the leaves with labels in \( A \) is formed at the level \( K_A \) and ends at level \( J_A \), thus
\[
K_A := \max\{1 \leq k \leq n : A \in \pi_k\} \quad \text{and} \quad J_A := \max\{1 \leq j < K_A : A \notin \pi_j\}.
\]
These are the lower level and the upper level of the branch supporting the leaves with labels in \( A \). Recall that \( \pi_k \) denotes the state of the coalescent after \( n-k \) coalescing events. For a set \( A \) of leaves which is not supported by some branch (which means that \( A \notin \pi_k \) for all \( k \)) we set \( K_A = J_A = n \).

The length of branch supporting the leaves with labels in \( A \) is then
\[
L_A := \sum_{j=J_A+1}^{K_A} X_j,
\]
where the \( X_j \sim \text{Exp}(\frac{1}{j}) \) are the inter-coalescence times. For sets \( A \) of leaves that are not supported by any branch, we have that \( L_A = 0 \).

The following lemma gives bounds on the distribution weights of the lower and upper levels of branches in the coalescent of the population alive at time 0. For the remainder of this section we consider the sets \( A, A' \in \mathcal{P}_n \) with \( |A| = i \) and \( |A'| = i' \) fixed.

**Lemma 3.** Let \( K := K_A, J := J_A \) and \( J' := J_{A'} \). Then, for \( 2 \leq j < k < n \) and \( 2 \leq j' < n \) the following bounds hold:
\[
\begin{align*}
\mathbb{P}(K = k) &\leq \frac{ik^2}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}, \\
\mathbb{P}(K = k, J = j) &\leq \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i}}, \\
\mathbb{P}(J = j) &\leq \frac{2j}{\binom{n}{i}} \cdot \frac{(n-j-1) \cdots (n-j-i+1)}{(n-1) \cdots (n-i)} \\
&\leq \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}}
\end{align*}
\]
and for \( A \neq A' \)
\[
\mathbb{P}(J = j, J' = j') \leq \begin{cases} 
\frac{4j(i')}{(n-1)(n-2)} & \text{if } A \cap A' = \emptyset, \\
\frac{4j(i'-i)}{(n-1)(n-2)} & \text{if } A' \subset A \text{ and } j' > j.
\end{cases}
\]

Otherwise, \( \mathbb{P}(J = j, J' = j') = 0 \).

Moreover, if \( A \cap A' = \emptyset \) then for a finite constant \( c \) depending on \( i \) and \( i' \)
\[
\mathbb{P}(J = j, J' = j') \leq \mathbb{P}(J = j)\mathbb{P}(J' = j')\left(1 + \frac{c}{n}\right).
\]

**Remark 1.** Note that the case \( i = 1, k = n \) is not covered by the Lemma. In this situation \( K = n \) and evidently \( \mathbb{P}(K = n, J = j) = \mathbb{P}(J = j) \).
Proof. To begin with let $i \geq 2$. Using the fact that the event $\{K = k\}$ is the disjoint union (over $n > k_1 > \cdots > k_{i-2} > k$) of the events (the branch with leaves in $A$ is formed through $i-1$ mergers happening at the levels $k_1, \ldots, k_{i-2}$ and $k$) we obtain that

$$
\mathbb{P}(K = k) = \sum_{n > k_1 > \cdots > k_{i-2} > k} \binom{n-i}{2} \binom{n-1-i}{2} \cdots \frac{(k_i+2-i-1)}{2} \cdot \frac{(k_i+1-i)}{2} \cdot \frac{(k_i-i-1)}{2} \cdots \frac{(k_i+2)}{2} \cdot \frac{(k_i+1)}{2} \cdot \frac{(k_i)}{2} \cdots \frac{(k)}{2}.
$$

$$
= \sum_{n > k_1 > \cdots > k_{i-2} > k} \binom{n}{i} \cdot \binom{n-1}{i} \cdots \binom{n-i+1}{i} \cdot \frac{k(k-1)}{2}.
$$

$$
= \frac{(n-k-1)(n-k-i+2)!}{(i-2)!} \cdot \binom{n}{i} \cdot \binom{n-1}{i} \cdots \binom{n-i+1}{i} \cdot \frac{2!(i-1)!}{n \cdots (n-i+1)(n-i+2) \cdots (n-i)} \cdot \frac{1}{i}.
$$

$$
= \frac{i k^2}{(n-1)(n-2) \cdot \binom{n}{i}}.
$$

Using the fact that the event $\{J = j\}$ means that the branch ends at level $j$ by a merger with one of the other $j$ lines present in the tree, we obtain in a similar way using the equality (25) for the joint distribution weights of $K$ and $J$ that

$$
\mathbb{P}(K = k, J = j) = \mathbb{P}(K = k) \cdot \frac{(j-i-1) \cdots (j+1)}{(i-2) \cdots (j+1)} \cdot \frac{j}{(i+1)}.
$$

$$
= \binom{n-k-1}{i-2} \cdot \binom{i}{2} \cdots \binom{n-i+1}{i} \cdots \binom{j}{2} \cdots \frac{j}{(i+1)} \cdot \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{(i)}.
$$

The following holds for the distribution weights of $J$: By summing up over all $j + 1 \leq k \leq n - i + 1$ we obtain that

$$
\mathbb{P}(J = j) = \sum_{k = j+1}^{n-i+1} \mathbb{P}(K = k, J = j) = \binom{i}{2} \cdots \frac{j}{(i+1)} \cdot \frac{1}{(i)} \cdot \sum_{k = j+1}^{n-i+1} \binom{n-k-1}{i-2} \cdot \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{(i)}.
$$

$$
= \binom{i}{2} \cdots \frac{j}{(i+1)} \cdot \frac{1}{(i)} \cdot \binom{n-j-1}{i-1} \cdot \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{(i)}.
$$

where the last equality follows from the fact that $\sum_{k = j+1}^{n-i+1} \binom{n-k-1}{i-2}$ gives for fixed $j$ the number of ways in which one can choose $i-1$ integers $n > k_1 > \cdots > k_{i-2} > k > j$ first by choosing the smallest integer $n-i+1 > k > j$ and then the other $i-2$ ones $n > k_1 > \cdots > k_{i-2} > k$. We further obtain for the distribution weights of $J$ that

$$
\mathbb{P}(J = j) = \frac{(n-j-1) \cdots (n-j-i+1)}{(i-1)!} \cdot \frac{2ij}{(n-i+1)(n-1) \cdots (n-i)} \cdot \frac{1}{(i)}.
$$

$$
= \frac{2j}{(i)} \cdot \frac{(n-j-1) \cdots (n-j-i+1)}{(n-1) \cdots (n-i)}
$$

$$
\leq \frac{2j}{(i)} \cdot \frac{(n-2) \cdots (n-i)}{(n-1) \cdots (n-i)} = \frac{2j}{(n-1)} \cdot \frac{1}{(i)}.
$$
This continues to hold also for $i = 1$, as seen by the following quick calculation:

$$
\mathbb{P}(J = j) = \frac{(n-1)}{2} \cdot \frac{(j+1)}{2} \cdot \frac{j}{n} = \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{1}}.
$$

Let us now consider $i \geq 1$ and look at the joint distribution of $J$ and $J'$ and consider first the case where $A \cap A' = \emptyset$. First assume that $j' < j$. Note that the $(i' - 1)$ levels at which the mergers through which the branch with leaves in $A'$ is formed are all different from the $(i - 1)$ levels at which the mergers through which the branch with leaves in $A$ is formed and they are also different from $j$. At these $(i - 1)$ (and respectively $(i' - 1)$) levels pairs of branches with leaves in $A$ (respectively $A'$) are chosen to merge, whereas at the other levels (excepting $j$ and $j'$) branches with leaves in $\{1, \ldots, n\} \setminus (A \cup A')$. At level $j$ the branch with leaves in $A$ ends through a merger with one of the existing branches with leaves in $\{1, \ldots, n\} \setminus (A \cup A')$. At level $j'$ the branch with leaves in $A'$ ends by merging with one of the $j'$ branches with leaves in $\{1, \ldots, n\} \setminus (A \cup A')$. Therefore, similar as above, we obtain that

$$
\mathbb{P}(J = j, J' = j') = \sum_{n > k_i > \cdots > k_{i-1} > j} \sum_{\{k'_i, \ldots, k'_{i-1}, j\} \cap \{k_i, \ldots, k_{i-1}, j\} = \emptyset} \frac{(j)}{2} \cdot \frac{(j')}{2} \cdot \frac{(j - j')}{2} \cdot (j - q) \cdot j',
$$

where $1 \leq q \leq i' - 1$ is the number of branches with leaves in $A'$ extant at level $j$. Indeed, the coalescence event bringing the coalescent tree from $j + 1$ down to $j$ branches involves the branch with leaves in $A$ and one other branch. The $q$ current branches that carry the leaves in the set $A'$ cannot take part in this coalescence and therefore there are $j - q$ branches with whom the branch with leaves in $A$ can coalesce. Then

$$
\mathbb{P}(J = j, J' = j') \leq \binom{n - j - 1}{i - 1} \binom{n - j' - i - 1}{i' - 1} \frac{(j)}{2} \cdot \frac{(j')}{2} \cdot \frac{(j - j')}{2} \cdot (j - q) \cdot j' \leq 4jj' \cdot \frac{(n - 3) \cdots (n - i - 1) \cdot (n - i - 3) \cdots (n - (i + i') - 1)}{(n - 1) \cdots (n - (i + i'))} \cdot \frac{i'i'}{n \cdots (n - (i + i') + 1)} \leq \frac{4jj'}{(n - 1)(n - 2)} \cdot \frac{1}{\binom{n}{i, i', n-(i+i')}}. \tag{27}
$$

Observe also that by putting together (27) and (26) one obtains

$$
\mathbb{P}(J = j, J' = j') \leq \binom{n - j - 1}{i - 1} \binom{n - j' - i - 1}{i' - 1} \frac{(j)}{2} \cdot \frac{(j')}{2} \cdot \frac{(j - j')}{2} \cdot (j - q) \cdot j' = \mathbb{P}(J = j) \mathbb{P}(J' = j') \cdot \frac{(n)}{2} \cdot \frac{(n-i)}{2} \cdot \frac{(n-i')}{2} \cdot \frac{(n-i-i'+1)}{2} \leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \left(1 + \frac{c}{n}\right)
$$

for a finite constant $c$ depending on $i$ and $i'$.

The case $j < j'$ is analogue. In the case $j = j'$, the two branches we consider end at level $j$ by coalescing with one another and therefore we obtain with a similar argument as above that
\[ \begin{align*}
\mathbb{P}(J = j, J' = j') &= \sum_{n > k_1 > \cdots > k_{i-1} > j} \sum_{n' > k'_{i-1} > j} \binom{n}{2} \cdots \binom{n-(i+i')}{2} \cdot \binom{j}{2} \cdot 1 \\
&= \left( i - 1, i' - 1, n - j - i - i' + 1 \right) \binom{j}{2} \cdots \binom{j}{2} \cdot \binom{j}{2} \cdot 1 \\
&\leq \frac{2j^2}{(n-1)(n-2)} \cdot \frac{1}{(i, i', n - (i+i')')}.
\end{align*} \]

Note that the factor \(\binom{j}{2}\) in the numerator is not cancelled by a factor in the denominator.

Also in this case by putting together (28) and (26) we get

\[ \mathbb{P}(J = j, J' = j') = \left( \binom{n}{2} \cdots \binom{n-(i+i')}{2} \cdot j \cdot \binom{j}{2} \cdots \binom{j}{2} \cdot \frac{j}{2} \right) \frac{(n-j-i)!(n-j-i')!}{(n-j-1)!(n-j-i-i' + 1)!} \]

\[ \leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \cdot \frac{(n-1)!}{(n-2)\cdots(n-i+i')!} \cdot \frac{(n-j-i)!(n-j-i')!(n-j-i-i')}{(n-j-1)\cdots(n-j-i+1)!} \]

\[ \leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \left( 1 + \frac{c}{n} \right), \]

for a finite constant \(c\).

Let now \(A' \subset A\), \(A \neq A'\). Then the branch with leaves in \(A'\) can only end at a level \(j' > j\) and if the mergers through which the branch with leaves in \(A\) is formed are \(k_1, \ldots, k_{i-1}\), then \(j' \in \{k_1, \ldots, k_{i-1}\}\). If the branches with leaves in \(A'\) coalesce at the levels \(\{k'_1, \ldots, k'_{i-1}\}\), we write \(\{k_1, \ldots, k_{i-1}\} = \{k'_1, \ldots, k'_{i-1}, j', l_1, \ldots, l_{i-i'-1}\}\) and have that

\[ \begin{align*}
\mathbb{P}(J = j, J' = j') &= \sum_{n > k'_1 > \cdots > k'_{i-1} > j'} \sum_{n' > l_1 > \cdots > l_{i-i'-1} > j'} \binom{n}{2} \cdots \binom{n-(i+i')}{2} \cdot \binom{j}{2} \cdots \binom{j}{2} \\
&\leq \mathbb{P}(J = j) \mathbb{P}(J' = j') \left( 1 + \frac{c}{n} \right),
\end{align*} \]

where \(1 \leq q \leq i-i'\) is the number of branches with leaves in \(A'\) extant at level \(j'\). Note that most of the binomials cancel and that the summands depend on the sequences \(k'_1, \ldots, k'_{i-1}\) and \(l_1, \ldots, l_{i-i'-1}\) only through the factor \(q\). We obtain

\[ \mathbb{P}(J = j, J' = j') \leq \left( \binom{n-j'-1}{i-1} \right) \left( \binom{n-j-i'-1}{i-1} \right) \frac{(i-i')!}{(n-1)(n-2) \cdots (n-i-i' + 1)!} \cdot j \cdot (i-i'). \]

Replacing \(j\) by 1 and \(j'\) by 2 in the binomials we obtain

\[ \mathbb{P}(J = j, J' = j') \leq 4j(i-i') \cdot \frac{1}{(n-1)(n-2)} \cdot \frac{(i-i')!}{n \cdots (n-i+1)} \cdot \frac{1}{(i-i'i', n-i)}, \]

This finishes the proof of the lemma. \( \Box \)
Let us now introduce some more notation. For \( i \in \mathbb{N} \) let
\[
G_i := \sigma(J_A, K_A : A \in \mathcal{P}_n, |A| = i)
\]
be the \( \sigma \)-algebra containing the information about the upper and lower levels of the branches of order \( i \) in the coalescent at time 0 and let
\[
G := \sigma \left( \bigcup_{i=1}^{r} G_i \right).
\]
Then the conditional expectation of the length of the branch supporting the leaves with labels in \( A \) is
\[
E(L_A | G) = E \left( \sum_{j=J_A+1}^{K_A} X_j | G \right) = \sum_{j=J_A+1}^{K_A} E(X_j) = \frac{2}{J_A} - \frac{2}{K_A}.
\]
Note also that by (2)
\[
E(L_A) = \frac{1}{(i)} \cdot E(L^{n,i}) = \frac{1}{(i)} \cdot \frac{2}{r}.
\]
The following lemma employs Lemma 3 extensively and will be of use in the proof of Theorem 1.

**Lemma 4.** For fixed \( i \geq 1 \) as \( n \to \infty \)
\[
\frac{n}{4 \log n} \sum_{A \in \mathcal{P}_n, |A| = i} (E(L_A | G) - E(L_A))^2 \xrightarrow{p} 1
\]
and
\[
\max_{A \in \mathcal{P}_n, |A| = i} \sqrt{\frac{n}{4 \log n}} |E(L_A | G) - E(L_A)| \xrightarrow{p} 0.
\]

**Proof.** Remember that if \( J_A = n \) then \( K_A = n \), thus by (30) it holds that
\[
\frac{n}{4 \log n} \sum_{A \in \mathcal{P}_n, |A| = i} E(L_A | G)^2 = \frac{n}{4 \log n} \sum_{A \in \mathcal{P}_n, |A| = i} E(L_A | G)^2 \cdot 1_{\{J_A < n\}}
\]
\[
= \frac{n}{\log n} \sum_{A \in \mathcal{P}_n, |A| = i} \frac{1}{J_A} \cdot 1_{\{J_A < n\}} - \frac{n}{\log n} \sum_{A \in \mathcal{P}_n, |A| = i} \frac{2}{J_A K_A} \cdot 1_{\{J_A < n\}}
\]
\[
+ \frac{n}{\log n} \sum_{A \in \mathcal{P}_n, |A| = i} \frac{1}{K_A^2} \cdot 1_{\{J_A < n\}}.
\]
In what follows we show that
\[
\frac{n}{\log n} \sum_{|A| = i} \frac{1}{J_A} \cdot 1_{\{J_A < n\}} \xrightarrow{p} 1,
\]
whereas the other two sums on the right-hand side of (34) converge to 0 in probability.

Let us denote by \( A = A(n, i, a) \) the event that at least one branch of order \( i \) ends between level 2 and level \( a\sqrt{n} \) in the coalescent, where \( a \) is a positive constant:
\[
A = A(n, i, a) := \{ J_A < a\sqrt{n} \text{ for some } A \in \mathcal{P}_n, |A| = i \}.
\]
Note by Lemma 3 that
\[
P(A) \leq \sum_{|A| = i} P(J_A < a\sqrt{n}) \leq \sum_{|A| = i} \sum_{2j < a\sqrt{n}} \frac{2j}{(n-1)} \cdot \frac{1}{(i)} \leq 2a^2.
\]
Since $\sum_{|A|=i} \frac{1}{n} \cdot 1_{\{J_A < n\}} \neq \sum_{|A|=i} \frac{1}{n} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}$ only on the event $\mathcal{A}$, which for large $a$ has arbitrary small probability, in order to prove (35) it suffices to show that for all $a > 0$

$$\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}} \xrightarrow{\mathbb{P}} 1 \text{ as } n \to \infty.$$ Using Lemma 3 we obtain that for all $a > 0$

$$\mathbb{E}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}\right) = \frac{n}{\log n} \sum_{|A|=i} \sum_{a^{\sqrt{n}} \leq j < n} \frac{1}{j^2} \cdot \binom{n}{i} \cdot \frac{(n-j-1) \cdots (n-j+i)}{(n-1) \cdots (n-i)}$$

$$= \frac{n}{(n-1) \log n} \sum_{a^{\sqrt{n}} \leq j < n} \frac{2}{j} \cdot \binom{n}{i} \cdot \frac{(n-2-(j-1)) \cdots (n-i-(j-1))}{(n-2) \cdots (n-i)}$$

$$\to 1$$

as $n \to \infty$. For the proof divide the sum into the parts $a^{\sqrt{n}} \leq j < \frac{n}{\log n}$ and $\frac{n}{\log n} \leq j < n$.

For the variance we obtain

$$\mathbb{V}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}\right) \leq \frac{n^2}{\log^2 n} \sum_{A} \mathbb{E}\left(\frac{1}{J_A^4} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}\right)$$

$$+ \frac{n^2}{\log^2 n} \sum_{A \cap A' = \emptyset} \mathbb{C}O\mathbb{V}\left(\frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}, \frac{1}{J_{A'}^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_{A'} < n\}}\right).$$

Note that for $A \cap A' \notin \{A, A', \emptyset\}$ either $J_A$ or $J_{A'} = n$ and thus the corresponding covariances are non-positive.

By Lemma 3 it holds for $A \cap A' = \emptyset$ that

$$\mathbb{C}O\mathbb{V}\left(\frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}, \frac{1}{J_{A'}^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_{A'} < n\}}\right)$$

$$= \sum_{a^{\sqrt{n}} \leq j, j' < n} \frac{1}{J_A^2} \cdot \left(\mathbb{P}(J_A = j, J_{A'} = j') - \mathbb{P}(J_A = j)\mathbb{P}(J_{A'} = j')\right)$$

$$\leq c \cdot \sum_{a^{\sqrt{n}} \leq j, j' < n} \frac{1}{J_A^2} \cdot \mathbb{P}(J_A = j)\mathbb{P}(J_{A'} = j') \cdot \frac{1}{n}.$$ for a constant $c$ depending on $i$ and $i'$. Plugging this in (35) and using again Lemma 3 we obtain that

$$\mathbb{V}\left(\frac{n}{\log n} \sum_{|A|=i} \frac{1}{J_A^2} \cdot 1_{\{a^{\sqrt{n}} \leq J_A < n\}}\right) \ll \frac{n^2}{\log^2 n} \sum_{A} \sum_{a^{\sqrt{n}} \leq j < n} \frac{1}{j^4} \cdot \frac{2}{n-1} \cdot \frac{1}{\binom{n}{i}}$$

$$+ \frac{n^2}{\log^2 n} \sum_{A \cap A' = \emptyset} c(\log n)^2 \cdot \frac{1}{n^3} \cdot \frac{1}{\binom{n}{i}^2}$$

$$\ll \frac{n^2}{\log^2 n} \left(\frac{1}{a^2 n^2} + \frac{c(\log n)^2}{n^3}\right)$$

$$\ll \frac{1}{n}.$$
and thus

\[ \mathbb{V}(\frac{n}{\log n} \sum_{|A| = i} \frac{1}{J_A^2} \cdot 1_{(a \sqrt{n} \leq J_A < n)}) \to 0 \quad \text{as } n \to \infty. \]

This together with \(54\) and \(56\) proves \(55\).

We now look at the other two sums on the right-hand side of \(34\). Note by Lemma \(3\) that

\[ \mathbb{E}\left( \frac{n}{\log n} \sum_{|A| = i} \frac{1}{J_A} \cdot 1_{(J_A < n)} \right) \leq \frac{n}{\log n} \sum_{A} \sum_{k \geq 3} \frac{1}{k^2} \cdot \frac{ik^2}{(n-1)(n-2)} \cdot \frac{1}{(\binom{n}{i})} \]

\[ \ll \frac{n}{\log n} \sum_{A} \frac{1}{n} \cdot \frac{i}{(\binom{n}{i})} \]

\[ \ll \frac{i}{\log n} \rightarrow 0 \]

as \(n \to \infty\). By the Cauchy-Schwarz inequality we further obtain using \(34\) that

\[ \frac{n}{\log n} \sum_{|A| = i} \frac{1}{J_A K_A} \cdot 1_{(J_A < n)} \leq \frac{n}{\log n} \sum_{A} \frac{1}{J_A^2} \cdot 1_{(J_A < n)} \cdot \sqrt{\sum_{A} \frac{1}{K_A^2} \cdot 1_{(J_A < n)}} \rightarrow 0 \]

as \(n \to \infty\) and thus by \(34\) we obtain that

\[ \frac{n}{4 \log n} \sum_{|A| = i} \mathbb{E}(L_A | G)^2 \overset{p}{\rightarrow} 1 \quad (39) \]

as \(n \to \infty\).

Now note by \(51\) that for \(i \geq 1\)

\[ \frac{n}{4 \log n} \sum_{|A| = i} \mathbb{E}(L_A)^2 = \frac{n}{4 \log n} \sum_{A} \left( \frac{1}{(\binom{n}{i})} \cdot \frac{2}{i} \right)^2 = \frac{n}{4 \log n} \cdot \frac{1}{(\binom{n}{i})} \cdot \frac{4}{i^2} \to 0 \]

as \(n \to \infty\). Because of \(39\) it follows by the Cauchy-Schwarz inequality that

\[ \frac{n}{4 \log n} \sum_{|A| = i} 2\mathbb{E}(L_A | G) \cdot \mathbb{E}(L_A) \leq \frac{n}{2 \log n} \sqrt{\sum_{A} \mathbb{E}(L_A | G)^2} \cdot \sqrt{\sum_{A} \mathbb{E}(L_A)^2} \overset{p}{\rightarrow} 0 \]

as \(n \to \infty\) and thus \(32\) holds.

Finally observe that by \(30\) and \(31\)

\[ |\mathbb{E}(L_A | G) - \mathbb{E}(L_A)| \leq \frac{2}{J_A} + \frac{2}{n} \leq \frac{1}{a \sqrt{n}} + \frac{2}{n} \]

on the event \(A^c\). Thus the claim \(33\) is a direct consequence of \(36\).

5 Contribution of big families

In this section we show that the branches of big orders from the coalescent at time 0 make only a negligible contribution to the external length at a later time \(h\) when the number of individuals in the population gets large.
Recall the definition of $R_{n,h}^{n,r}$ given in (4) and let 

$$a_{n,h} := \left( \log \left( \frac{h + 2}{h} \right) \right)^{-1} \log n. \quad (40)$$

Then we can write

$$R_{h}^{n,r} = \tilde{R}_{h}^{n,r} + \bar{R}_{h}^{n,r},$$

where

$$\tilde{R}_{h}^{n,r} := \sum_{A \in P_n \atop r < |A| \leq a_{n,h}} 1_{X_A} \cdot L_A, \quad \bar{R}_{h}^{n,r} := \sum_{A \in P_n \atop a_{n,h} < |A| < n} 1_{X_A} \cdot L_A.$$

**Proposition 2.** For each $h \geq 0$ there exists an $\varepsilon(r) = o(1)$ when $r \to \infty$, such that

$$\forall \left( \tilde{R}_{h}^{n,r} \right) \leq \varepsilon(r) \frac{\log n}{n}$$

holds. Moreover $E(\bar{R}_{h}^{n,r}) = O\left( \frac{1}{n} \right)$.

**Proof.** First we prove the second claim. Since the collection of random variables $\{1_{X_A}\}_{A \in P_n}$ is independent of the collection of random variables $\{L_A\}_{A \in P_n}$ due to the Poissonian structure embedded in the Moran model, it holds that

$$E\left( \sum_{|A| > a_{n,h}} 1_{X_A} \cdot L_A \right) = \sum_{i > a_{n,h}} \sum_{|A| = i} E(1_{X_A}) \cdot E(L_A).$$

Since $E(1_{X_A}) \leq 2i \cdot \left( \frac{h}{h + 2} \right)^{i-1}$ (see (12)), using the fact that $E(L_{n,i}^h) = 2i$ and the relation

$$E(L_{n,i}^h) = E\left( \sum_{|A| = i} L_A \right) = \binom{n}{i} E(L_A),$$

we obtain that

$$E(\tilde{R}_{h}^{n,r}) = E\left( \sum_{|A| > a_{n,h}} 1_{X_A} \cdot L_A \right) \leq \sum_{i > a_{n,h}} \binom{n}{i} \cdot 2i \cdot \left( \frac{h}{h + 2} \right)^{i-1} \cdot \frac{2}{i} \cdot \frac{1}{\binom{n}{i}} \ll \left( \frac{h}{h + 2} \right)^{\log \left( \frac{h + 2}{h} \right)} \log n \leq \frac{1}{n}.$$

Thus $E(\tilde{R}_{h}^{n,r}) = O\left( \frac{1}{n} \right)$.

We now proceed to proving the first claim of the Proposition. For the variance of $\tilde{R}_{h}^{n,r}$ it holds that

$$\forall \left( \tilde{R}_{h}^{n,r} \right) = \forall\left( \sum_{r < |A| \leq a_{n,h}} 1_{X_A} \cdot L_A \right)$$

$$= \forall\left( \sum_{r < |A| \leq a_{n,h}} (1_{X_A} - E(1_{X_A})) \cdot L_A + \sum_{r < |A| \leq a_{n,h}} E(1_{X_A}) \cdot L_A \right)$$

$$= \forall\left( \sum_{r < |A| \leq a_{n,h}} (1_{X_A} - E(1_{X_A})) \cdot L_A \right) + \forall\left( \sum_{r < |A| \leq a_{n,h}} E(1_{X_A}) \cdot L_A \right)$$

$$+ \text{COV}\left( \sum_{r < |A| \leq a_{n,h}} (1_{X_A} - E(1_{X_A})) \cdot L_A, \sum_{r < |A| \leq a_{n,h}} E(1_{X_A}) \cdot L_A \right).$$
Taking into account the independence of the collections of random variables \( \{1_{\mathcal{G}_A}\}_{A \in \mathcal{P}_n} \) and \( \{L_A\}_{A \in \mathcal{P}_n} \), the last term on the right-hand side is equal to zero. Therefore, using the fact that \( \mathbb{E}(1_{\mathcal{G}_A}) \) depends on the set \( A \) only through its cardinality and again from independence, we have that

\[
\mathbb{V}(\tilde{R}_h^{n,r}) = \mathbb{E}\left(\sum_{r < |A| \leq s_{n,h}} (1_{\mathcal{G}_A} - \mathbb{E}(1_{\mathcal{G}_A})) \cdot L_A \right)^2 + \mathbb{V}\left(\sum_{r < |A| \leq s_{n,h}} \mathbb{E}(1_{\mathcal{G}_A}) \cdot L_A \right)
\]

\[
= \sum_{r < |A| \leq s_{n,h}} \mathbb{E}\left((1_{\mathcal{G}_A} - \mathbb{E}(1_{\mathcal{G}_A}))\cdot (1_{\mathcal{G}_{A'}} - \mathbb{E}(1_{\mathcal{G}_{A'}}))\right) \cdot \mathbb{E}(L_AL_{A'}) + \mathbb{V}\left(\sum_{r < i \leq s_{n,h}} p_{i,h}^{n} \cdot \mathcal{L}_{n,i}\right).
\]

Remember that \( p_{i,h}^{n} = \mathbb{E}(1_{\mathcal{G}_A}) \). We start by evaluating the first term on the right-hand side. To this aim note that if \( A \) and \( A' \) are such that \( A \cap A' \neq \emptyset, \{0, A, A'\} \), then there cannot be two branches in the coalescent tree, one supporting the leaves in \( A \) and the other one supporting the leaves in \( A' \) and therefore in this case the product \( L_AL_{A'} \) is equal to zero. Thus, omitting in the notation the restrictions on the cardinalities of the sets, it holds that

\[
\sum_{A,A' \in \mathcal{P}_n} \mathbb{E}\left((1_{\mathcal{G}_A} - \mathbb{E}(1_{\mathcal{G}_A}))\cdot (1_{\mathcal{G}_{A'}} - \mathbb{E}(1_{\mathcal{G}_{A'}}))\right) \cdot \mathbb{E}(L_AL_{A'})
\]

\[
\leq 2 \sum_{A' \subseteq A} \mathbb{E}(1_{\mathcal{G}_A}1_{\mathcal{G}_{A'}}) \cdot \mathbb{E}(L_AL_{A'}) + \sum_{A \cap A' = \emptyset} \text{COV}\left(1_{\mathcal{G}_A}, 1_{\mathcal{G}_{A'}}\right) \cdot \mathbb{E}(L_AL_{A'}) + \sum_{A} \mathbb{V}(1_{\mathcal{G}_A}) \cdot \mathbb{E}(L_AL_{A'})
\]

Using the definition \([24]\) of \( L_A \) we obtain that

\[
\mathbb{E}(L_AL_{A'}) = \mathbb{E}\left(\mathbb{E}(L_AL_{A'} \mid J_A, J_{A'}, K_A, K_{A'})\right)
\]

\[
= \mathbb{E}\left(\sum_{j = J_A + 1}^{K_A} \sum_{l = J_{A'} + 1}^{K_{A'}} X_j \cdot X_l \mid J_A, J_{A'}, K_A, K_{A'}\right)
\]

\[
= \mathbb{E}\left(\sum_{j = J_A + 1}^{K_A} \sum_{l = J_{A'} + 1}^{K_{A'}} \mathbb{E}(X_jX_l)\right).
\]

The exponential inter-coalescence times \( X_j \) and \( X_l \) are independent if \( j \neq l \) and in this case \( \mathbb{E}(X_jX_l) = \frac{2}{j(j-1)l(l-1)} \), whereas if \( j = l \) it holds that \( \mathbb{E}(X_jX_l) = \mathbb{E}(X_j^2) = 2\left(\frac{2}{j(j-1)}\right)^2 = 2\frac{2}{j(j-1)^2} \). We thus obtain that

\[
\mathbb{E}(L_AL_{A'}) \leq 2 \mathbb{E}\left(\sum_{j = J_A + 1}^{K_A} \sum_{l = J_{A'} + 1}^{K_{A'}} \frac{2}{j(j-1)l(l-1)}\right)
\]

\[
= 2 \mathbb{E}\left(\sum_{j = J_A + 1}^{K_A} \frac{2}{j(j-1)} \cdot \sum_{l = J_{A'} + 1}^{K_{A'}} \frac{2}{l(l-1)}\right)
\]

\[
= 2 \mathbb{E}\left(\left(\frac{2}{J_A} - \frac{2}{K_A}\right) \left(\frac{2}{J_{A'}} - \frac{2}{K_{A'}}\right)\right)
\]

\[
\leq 8 \mathbb{E}\left(\frac{1}{J_A J_{A'}}\right).
\]

In order to bound the first term on the right-hand side of \((42)\) observe first that by \((12)\)

\[
\mathbb{E}\left(1_{\mathcal{G}_A}1_{\mathcal{G}_{A'}}\right) \leq \mathbb{E}\left(1_{\mathcal{G}_A}\right) \leq 2t\left(\frac{h}{h+2}\right)^{t-1}.
\]

Now, using \((43)\) together with the fact that \( A' \subseteq A \) implies that \( J_{A'} > J_A \) and Lemma 3 we obtain
\[
\sum_{A' \subseteq A} \mathbb{E}\left(1_{\mathcal{F}_A} 1_{\mathcal{F}_{A'}}\right) \cdot \mathbb{E}(L_A L_{A'}) 
\]

(44)

\[
\leq 16 \sum_{i', i < l} \left(1 - \frac{n}{h + 2}\right) \cdot i \cdot \left(\frac{h}{h + 2}\right)^{-1} \cdot \sum_{2 \leq j, j' \leq l} \frac{1}{jj'} \mathbb{P}(J_A = j, J_{A'} = j') 
\]

\[
\leq 16 \sum_{i', i < l} \left(1 - \frac{n}{h + 2}\right) \cdot i \cdot \left(\frac{h}{h + 2}\right)^{-1} \cdot \sum_{2 \leq j, j' \leq l} \frac{1}{jj'} \cdot \frac{4j(i - i')}{(n - 1)(n - 2)} \cdot \frac{1}{(i - i', i', n - i)} 
\]

\[
\leq 64 \sum_{i', i < l} \left(1 - \frac{n}{h + 2}\right) \cdot i \cdot \left(\frac{h}{h + 2}\right)^{-1} \cdot \sum_{2 \leq j, j' \leq l} \frac{1}{jj'} 
\]

\[
\ll \frac{1}{n} \cdot \sum_{i', i < l} i(i - i') \cdot \left(\frac{h}{h + 2}\right)^{-1} 
\]

\[
\ll \frac{1}{n} \cdot \sum_{r > r} i^3 \cdot \left(\frac{h}{h + 2}\right)^{-1} 
\]

\[
\ll \frac{1}{n}. 
\]

For bounding the second term on the right-hand side of (44) observe that for \(A \cap A' = \emptyset\) by (43) and Lemma 3 it holds that

\[
\mathbb{E}(L_A L_{A'}) \leq 8 \sum_{j, j' \geq 2} \frac{1}{jj'} \mathbb{P}(J_A = j, J_{A'} = j') 
\]

\[
\leq 32 \sum_{j, j' \geq 2} \frac{1}{jj'} \frac{j}{(n - 1)(n - 2)} \cdot \frac{1}{(i, i', n - (i + i'))} 
\]

\[
\ll \frac{1}{n} \cdot \sum_{r > r} i(i - i') \cdot \left(\frac{h}{h + 2}\right)^{-1} 
\]

and therefore by (23) we obtain that for some \(\gamma < 1\) depending on \(h\) it holds

\[
\sum_{A \cap A' = \emptyset} \mathbb{E}(1_{\mathcal{F}_A} 1_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A L_{A'}) \ll \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot \gamma^{i(i + i')} \cdot \frac{1}{(i, i', n - (i + i'))} \ll \frac{1}{n} \cdot \sum_{r > r} \gamma^{i(i + i')} \ll \frac{1}{n}. 
\]

We now turn to the last term on the right-hand side of (12), namely \(\sum_A \mathbb{V}(1_{\mathcal{F}_A}) \cdot \mathbb{E}(L_{A}^2)\). From (12) it holds that

\[
\mathbb{V}(1_{\mathcal{F}_A}) \leq \mathbb{E}\left(1_{\mathcal{F}_A}\right) \leq 2 \left(\frac{h}{h + 2}\right)^{-1} 
\]

and by (43) and Lemma 3

\[
\mathbb{E}(L_{A}^2) \leq 8 \sum_{j \geq 2} \frac{1}{j^2} \mathbb{P}(J_A = j) \leq 16 \sum_{j \geq 2} \frac{1}{j^2} \cdot \frac{j}{(n - 1)} \cdot \frac{1}{(i, i', i') \cdot \mathbb{E}(L_{A}^2) \leq 2} 
\]

This leads to

\[
\sum_A \mathbb{V}(1_{\mathcal{F}_A}) \cdot \mathbb{E}(L_{A}^2) \ll \sum_{r < i \leq n} \left(\begin{array}{c} n \\ i \end{array}\right) \cdot \left(\frac{h}{h + 2}\right)^{-1} \cdot \frac{1}{(i, i', i') \cdot \mathbb{V}(L_{A}^2) \leq 2} 
\]

\[
\leq \varepsilon_1(r) \frac{\log n}{n} 
\]

(45)

where \(\varepsilon_1(r) := \sum_{i = r+1}^{\infty} \left(\frac{h}{h + 2}\right)^{-1} \cdot \frac{1}{(i, i', i') \cdot \mathbb{V}(L_{A}^2) \leq 2} \) has the property that

\[
\varepsilon_1(r) = o(1) \quad \text{when } r \to \infty. 
\]

(46)
Putting together (42) and (44) - (45) we obtain for the first term on the right-hand side of (41) that
\[
\sum_{r < |A|, |A'| \leq a_{n,h}} E \left( \left( 1 \mathcal{F}_A - E(1 \mathcal{F}_A) \right) \left( 1 \mathcal{F}_{A'} - E(1 \mathcal{F}_{A'}) \right) \right) E(L_A L_{A'}) \ll \left( \varepsilon_1 \frac{\log n}{n} + \frac{1}{n} \right).
\] (47)

In order to obtain the claim of the Proposition we are left to bound the second term on the right-hand side of (41). For this term it holds that
\[
\mathbb{V} \left( \sum_{r < i \leq a_{n,h}} p_{n,i,h} \cdot L_{n,i} \right) = \sum_{r < i \leq a_{n,h}} (p_{n,i,h})^2 \cdot \mathbb{V} \left( L_{n,i} \right) + \sum_{r < i', i \leq a_{n,h}, i \neq i'} p_{n,i,h} p_{n,i',h} \cdot \text{COV} \left( L_{n,i}, L_{n,i'} \right).
\] (48)

The variances and covariances of the internal lengths of different orders can be easily obtained from the results of Fu [Fu95] on the variances and covariances of the numbers \(M_i(n)\) of mutations carried by exactly \(i\) individuals in a population of size \(n\) evolving according to the Moran model under the infinitely many sites mutation model. In this setting mutations are modelled as points of a Poisson process with constant rate \(\phi/2\) per unit length on the branches of the coalescent tree. Therefore it holds that
\[
\left( M_i(n) \mid L_{n,i} \right) \sim \text{Poisson} \left( \frac{\phi}{2} L_{n,i} \right).
\]

We thus obtain by the law of total variance and of total covariance respectively that
\[
\mathbb{V}(M_i(n)) = \mathbb{V} \left( \mathbb{E}(M_i(n) \mid L_{n,i}) \right) + \mathbb{E} \left( \mathbb{V}(M_i(n) \mid L_{n,i}) \right)
= \mathbb{V} \left( \frac{\phi}{2} L_{n,i} \right) + \mathbb{E} \left( \frac{\phi}{2} L_{n,i} \right)
= \frac{\phi^2}{4} \mathbb{V}(L_{n,i}) + \frac{\phi}{i}
\] (49)

and due to the independence ensured by the Poisson structure
\[
\text{COV}(M_i(n), M_{i'}(n)) = \text{COV} \left( \mathbb{E}(M_i(n) \mid L_{n,i}, L_{n,i'}), \mathbb{E}(M_{i'}(n) \mid L_{n,i}, L_{n,i'}) \right)
+ \mathbb{E} \left( \text{COV}(M_i(n), M_{i'}(n) \mid L_{n,i}, L_{n,i'}) \right)
= \text{COV} \left( \frac{\phi}{2} L_{n,i}, \frac{\phi}{2} L_{n,i'} \right) + 0
= \frac{\phi^2}{4} \text{COV}(L_{n,i}, L_{n,i'}).
\] (50)

The results of [Fu95] say that
\[
\mathbb{V}(M_i(n)) = \phi^2 \sigma_{ii} + \frac{\phi}{i} \quad \text{and} \quad \text{COV}(M_i(n), M_{i'}(n)) = \phi^2 \sigma_{ii'},
\] (51)

where in particular for \(i < n/2\)
\[
\sigma_{ii} = \beta_n(i + 1)
\]
and for \(i > i', i + i' < n/2\)
\[
\sigma_{ii'} = \frac{\beta_n(i + 1) - \beta_n(i)}{2}
\]
with \(h_n = \sum_{j=1}^{n-1} \frac{1}{j}\) and
\[
\beta_n(i) = \frac{2n}{(n - i + 1)(n - i)}(h_{n+1} - h_i) - \frac{2}{n - i}.
\]
Therefore from (49), (50) and (51) it follows that
\[ \mathbb{V}(L^{n,i}) = 4\sigma_{ii} \quad \text{and} \quad \text{COV}(L^{n,i}, L^{n,i'}) = 4\sigma_{ii'}, \]

Turning now to (48) note from the definition of (3) (when centred) the following holds
\[ \mathbb{E}(p_{n,h}^{i}) = \frac{h}{h+2} i \cdot c \cdot \log n, \]

for \( c \) a finite constant independent of \( i \). Also \( \sigma_{ii'} \leq 0 \) for \( i > i', i + i' < n/2 \) (see (36) in [Fu95]). Hence using (12), (48) becomes
\[ \mathbb{V} \left( \sum_{r < i \leq a_{n,h}} p_{n,h}^{i} \cdot L^{n,i} \right) \leq \sum_{r < i \leq a_{n,h}} i^2 \left( \frac{h}{h+2} \right)^{2i} \cdot c \cdot \log n \cdot \frac{\log n}{n}, \]

where \( \varepsilon_2(r) := c \sum_{i=r+1}^{\infty} i^2 \left( \frac{h}{h+2} \right)^{2i} \) has the property that
\[ \varepsilon_2(r) = o(1) \quad \text{when} \quad r \to \infty. \]

Putting now together (41), (46), (47), (52) and (53) we obtain the claim.

6 Proof of Theorem 1

6.1 Preliminaries

We start by making the observation that for the first term of the decomposition on the right-hand side of (3) (when centred) the following holds
\[ \sum_{|A| \leq r} (1_{\mathcal{F}_A} \cdot L_A - \mathbb{E}(1_{\mathcal{F}_A} \cdot L_A)) = \sum_{|A| \leq r} 1_{\mathcal{F}_A} \cdot (L_A - \mathbb{E}(L_A)) + O_P(n^{-1/2}). \]

Indeed, since the collections of random variables \( \{1_{\mathcal{F}_A}\}_{A \in \mathcal{P}_n} \) and \( \{L_A\}_{A \in \mathcal{P}_n} \) are independent we have that
\[ \sum_{|A| \leq r} 1_{\mathcal{F}_A} \cdot (L_A - \mathbb{E}(L_A)) = \sum_{|A| \leq r} 1_{\mathcal{F}_A} \cdot (L_A - \mathbb{E}(L_A)) + \sum_{|A| \leq r} (1_{\mathcal{F}_A} - \mathbb{E}(1_{\mathcal{F}_A})) \cdot \mathbb{E}(L_A) \]

and using (23) and (31) we obtain
\[ \mathbb{V} \left( \sum_{|A| \leq r} (1_{\mathcal{F}_A} - \mathbb{E}(1_{\mathcal{F}_A})) \cdot \mathbb{E}(L_A) \right) = \sum_{|A| \leq r} \text{COV}(1_{\mathcal{F}_A}, 1_{\mathcal{F}_{A'}}) \cdot \mathbb{E}(L_A) \mathbb{E}(L_{A'}) \]
\[ \leq \sum_{i=1}^{r} \sum_{|A| = i, |A'| = i'} \frac{1}{n} \cdot \left( \frac{n}{i} \right) \left( \frac{n}{i'} \right) + \sum_{i=1}^{r} \sum_{|A| = i, |A| = i'} \frac{1}{n} \cdot \left( \frac{n}{i} \right) \left( \frac{n}{i'} \right) \]
\[ \leq \frac{1}{n}. \]
For the second inequality we used the fact that the number of non-disjoint sets with cardinalities \( i \) and \( i' \) is

\[
\left| \{ A, A' \in \mathcal{P}_n, |A| = i, |A'| = i \text{ such that } A \cap A' \neq \emptyset \} \right| \leq \binom{n}{i} \cdot i \binom{n}{i'} (n / i') - 1 \leq \frac{1}{n} \binom{n}{i} \binom{n}{i'} .
\]

(55)

Thus (53) holds.

We next show that by replacing the exponential times \( X_j \) in the lengths \( L_A \) appearing on the right-hand side of (53) leads to a negligible error as the total population size \( n \) tends to infinity. In other words, the randomness brought in by the inter-coalescent times can be neglected for big population sizes. The information on the tree structure is contained in the \( \sigma \)-algebra \( \mathcal{G} \) (defined in (29)) and therefore the observation we just made amounts to saying that for \( i \geq 1 \)

\[
\sum_{|A| = i} 1_{\mathcal{A}_i} L_A = \sum_{|A| = i} 1_{\mathcal{A}_i} \mathbb{E}(L_A \mid \mathcal{G}) + O_P(n^{-1/2}).
\]

(56)

Indeed since the collection of random variables \( \{ 1_{\mathcal{A}_i} \}_{A \in \mathcal{P}_n} \) is independent of \( \{ L_A \}_{A \in \mathcal{P}_n} \) and independent of the \( \sigma \)-algebra \( \mathcal{G} \), it holds that

\[
\mathbb{E}\left( \left( \sum_{|A| = i} 1_{\mathcal{A}_i} (L_A - \mathbb{E}(L_A \mid \mathcal{G}_i)) \right)^2 \mid \mathcal{G} \right) = \sum_{|A| = i} 1_{\mathcal{A}_i} \mathbb{E}(1_{\mathcal{A}_i} (L_A - \mathbb{E}(L_A \mid \mathcal{G})) (L_A' - \mathbb{E}(L_A' \mid \mathcal{G})) \mid \mathcal{G}) = \sum_{|A| = i} 1_{\mathcal{A}_i} \mathbb{E}(1_{\mathcal{A}_i}) \mathbb{C}_\mathcal{V}(L_A, L_A' \mid \mathcal{G})
\]

Note that if \( A \cap A' \notin \{ A, A', \emptyset \} \) then there cannot be two branches in the coalescent, one supporting the leaves with labels in \( A \) and one the leaves with labels in \( A' \) and therefore in such case \( L_A \) or \( L_A' \) is (by definition) equal to 0 and thus \( \mathbb{E}(L_A, L_A' \mid \mathcal{G}) = 0 \) and \( \mathbb{C}_\mathcal{V}(L_A, L_A' \mid \mathcal{G}) \leq 0 \). Using now (24) we obtain that

\[
\mathbb{E}\left( \left( \sum_{|A| = i} 1_{\mathcal{A}_i} (L_A - \mathbb{E}(L_A \mid \mathcal{G}_i)) \right)^2 \mid \mathcal{G} \right) \leq \sum_A \mathbb{V}(\sum_{j = J_A + 1}^{K_A} X_j \mid \mathcal{G}) + \sum_{A \cap A' = \emptyset} \mathbb{C}_\mathcal{V}(\sum_{j = J_A + 1}^{K_A} X_j, \sum_{j' = J_{A'} + 1}^{K_{A'}} X_{j'} \mid \mathcal{G}) = \sum_A \sum_{j = J_A + 1}^{K_A} \mathbb{V}(X_j) + \sum_{A \cap A' = \emptyset} \sum_{j = J_A + 1}^{K_A} \sum_{j' = J_{A'} + 1}^{K_{A'}} \mathbb{C}_\mathcal{V}(X_j, X_{j'}).
\]

Recall that the inter-coalescent times \( X_j \) are independent and exponentially distributed with parameter \( (\lambda_j) \). Therefore

\[
\mathbb{E}\left( \left( \sum_{|A| = i} 1_{\mathcal{A}_i} (L_A - \mathbb{E}(L_A \mid \mathcal{G}_i)) \right)^2 \mid \mathcal{G} \right) \leq \sum_A \sum_{j = J_A + 1}^{K_A} \mathbb{V}(X_j) + \sum_{A \cap A' = \emptyset} \sum_{j = J_A \vee J_{A'} + 1}^{K_A \wedge K_{A'}} \frac{1}{(\lambda_j)^2} = \sum_A \sum_{j = J_A + 1}^{K_A} \frac{1}{(\lambda_j)^2} + \sum_{A \cap A' = \emptyset} \sum_{j = J_A \vee J_{A'} + 1}^{K_A \wedge K_{A'}} \frac{1}{(\lambda_j)^2} \]

\[
\ll \sum_A \frac{1}{J_A^2} + \sum_{A \cap A' = \emptyset} \frac{1}{J_A \vee J_{A'}^2}.
\]

25
Taking expectation and using Lemma 3 we obtain that
\[
E\left( \left( \sum_{|A| = i} 1_{\mathbb{A}_A} (L_A - E(L_A | \mathcal{G}_i)) \right)^2 \right) \leq \sum_A \sum_{j \geq 1} \frac{1}{j^3} \cdot \frac{2j}{(n-1)} \cdot \frac{1}{\binom{n}{i}} 
+ \sum_{A \cap A' = \emptyset} \sum_{j \geq 2} \sum_{j' = 2} \frac{1}{j^3} \cdot \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i,i,n-2i}} 
+ \sum_{A \cap A' = \emptyset} \sum_{j \geq 2} \sum_{j' = j+1} \frac{1}{j^2} \cdot \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{\binom{n}{i,i,n-2i}} \ll \frac{1}{n}.
\]

This yields (56).

The calculation above remains valid if we replace \( 1_{\mathbb{A}_A} \) by 1. Thus, in the same manner, we obtain that
\[
\sum_{|A| = i} L_A = \sum_{|A| = i} E(L_A | \mathcal{G}) + O_P(n^{-1/2}). \tag{57}
\]

### 6.2 From dependent to independent coefficients

This subsection contains a key building block of the proof of Theorem 1, namely we show that the dependent random coefficients \( \{1_{\mathbb{A}_A}\}_{A \in \mathcal{P}_n} \), which appear in the sum on the right-hand side of (56), can be replaced for big population sizes by random coefficients that are independent of one another and of the coalescent at time 0.

We group the branches of the coalescent at time 0 according to whether their upper and lower levels are below or above level \( \log n \) with
\[
b_n := \frac{n}{\log n}.
\]

and for each group we perform the replacement as follows: if the branch supporting the leaves with labels in the set \( A \) with \( |A| = i \) is formed above level \( b_n \), i.e. \( J_A < b_n \), or ends below or at level \( b_n \), i.e. \( J_A \geq b_n \), we replace the coefficient \( 1_{\mathbb{A}_A} \) by \( \widetilde{W}_{A,h} := 1_{\{B_h(A) = 1\}} \), where the processes \( B(A) \) are independent (for different \( A \)) birth and death processes started with \( i \) individuals at time 0 with birth and death rates equal to \( \frac{k}{2} \) when the process is in state \( k \). Moreover, we take the collections of birth and death processes to be independent of one another for different values of \( i \). For the branches that are formed below level \( b_n \) and end above this level, i.e. \( J_A \geq b_n \), we replace the coefficients \( 1_{\mathbb{A}_A} = 1_{\{Z_0(A) = 1\}} \) by \( W_{A,h} := 1_{\{B_h(A) = 1\}} \), where the processes \( B(A) \) are the coupled birth and death processes introduced in Section 3. The basic fact here is that the branches for which \( J_A < b_n \) have leaves in sets \( A \) which are disjoint and we do the coupling for all such sets at once.

We next show that the following holds:
\[
\sum_{|A| \leq r} 1_{\mathbb{A}_A} \cdot (E(L_A | \mathcal{G}) - E(L_A)) = \sum_{|A| \leq r} V_{A,h} \cdot (E(L_A | \mathcal{G}) - E(L_A)) + O_P\left( \left( \frac{\log n}{n} \right)^{1/2} \right), \tag{58}
\]
where we used the compact notation
\[
V_{A,h} := \widetilde{W}_{A,h} \cdot 1_{\{K_A < b_n\} \cup \{J_A \geq b_n\}} + W_{A,h} \cdot 1_{\{J_A < b_n \leq K_A\}}. \tag{59}
\]

Note that from the construction described above the random variables \( V_{A,h} \) are all independent and are Bernoulli(\( p_{n,h} \))-distributed if the set \( A \) has cardinality \( i \). Also, the random variables \( V_{A,h} \) are independent of the tree at time 0 (even though \( J_A \) and \( K_A \) are used in their definition).
Observe first that for \(|A| = |A'| = i, A \cap A' = \emptyset\)

\[
\left| E((1_{\mathcal{F}_A} - V_{A,h})(1_{\mathcal{F}_{A'} - V_{A',h}})) \right| \\
\leq \left| \text{COV}(1_{\mathcal{F}_A} - V_{A,h}, 1_{\mathcal{F}_{A'} - V_{A',h}}) + E(1_{\mathcal{F}_A} - V_{A,h})E(1_{\mathcal{F}_{A'} - V_{A',h}}) \right| \\
\leq \left| \text{COV}(1_{\mathcal{F}_A}, 1_{\mathcal{F}_{A'}}) \right| + |p_i^n - p_i|^2 \\
\ll \frac{1}{n},
\]

where the last inequality follows from (23) and (19).

We consider the three groups of branches separately. For each \(1 \leq i \leq r\) fixed we have due to independence that

\[
E\left( \left( \sum_{|A|=i} (1_{\mathcal{F}_A} - V_{A,h}) \cdot (E(L_A | \mathcal{G}) - E(L_A)) \cdot 1_{\left\{ K_A < b_n \right\}} \right)^2 \right) \\
= \sum_{|A|=|A'|=i} E\left( (1_{\mathcal{F}_A} - V_{A,h})(1_{\mathcal{F}_{A'} - V_{A',h}}) \right) \\
\cdot E\left( (E(L_A | \mathcal{G}) - E(L_A))(E(L_{A'} | \mathcal{G}) - E(L_{A'})); K_A, K_{A'} < b_n \right).
\]

Remember that if \(A \cap A' \notin \{A, A', \emptyset\}\) then at least one of the branches with leaves in \(A\), respectively \(A'\), does not exist and by definition \(K_A = n\) or \(K_{A'} = n\). Therefore the event \(\{K_A, K_{A'} < b_n\}\) is empty and thus in the sum above the summands for \(A \cap A' \notin \{A, A', \emptyset\}\) are equal to 0. Using (60) and \(|A| = |A'|\) we obtain that

\[
E\left( \left( \sum_{|A|=i} (1_{\mathcal{F}_A} - V_{A,h}) \cdot (E(L_A | \mathcal{G}) - E(L_A)) \cdot 1_{\left\{ K_A < b_n \right\}} \right)^2 \right) \\
\ll \sum_A E\left( (E(L_A | \mathcal{G})^2 + E(L_A)^2); K_A < b_n \right) + \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot E\left( (E(L_A | \mathcal{G})E(L_{A'} | \mathcal{G}) + E(L_A)^2) \right) \\
\ll \sum_A E\left( \frac{1}{J_A} + E(L_A)^2; K_A < b_n \right) + \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot E\left( \frac{1}{J_AJ_{A'}} + E(L_A)^2) \right),
\]

where the last inequality follows from (30). Let us write for short \(K := K_{\{1, \ldots, i\}}, J := J_{\{1, \ldots, i\}}\) and \(J' := J'_{\{1+1, \ldots, 2i\}}\). Using (31) and Lemma 8 we obtain that

\[
E\left( \left( \sum_{|A|=i} (1_{\mathcal{F}_A} - V_{A,h}) \cdot (E(L_A | \mathcal{G}) - E(L_A)) \cdot 1_{\left\{ K_A < b_n \right\}} \right)^2 \right) \\
\ll \left( \begin{array}{l} n \end{array} \right) \sum_{k \leq b_n} \sum_{j=2}^{k-1} \left( \frac{1}{j^2} - \frac{1}{n} \right) \left( \frac{2}{i} \right)^2 \cdot \mathbb{P}(K = k, J = j) \\
\quad + \left( \begin{array}{l} n \end{array} \right) \sum_{i, i, n - 2i} \left( \frac{1}{j} \right) \left( \frac{1}{i} \right) \sum_{j, j' \geq 2} \left( \frac{1}{j^2} + \frac{1}{n} \right) \left( \frac{2}{i} \right)^2 \cdot \mathbb{P}(J = j, J' = j') \\
\ll \left( \begin{array}{l} n \end{array} \right) \sum_{k \leq b_n} \sum_{j=2}^{k-1} \frac{2ij}{(n-1)(n-2)} \cdot \frac{1}{n} \left( \begin{array}{l} n \end{array} \right) \\
\quad + \left( \begin{array}{l} n \end{array} \right) \sum_{i, i, n - 2i} \left( \frac{1}{i} \right) \sum_{j, j' \geq 2} \frac{1}{j} \cdot \frac{4jj'}{(n-1)(n-2)} \cdot \frac{1}{i, i, n - 2i} \\
\ll \sum_{k \leq b_n} \frac{i \log k}{n^2} \cdot \frac{1}{n} \ll \frac{1}{n}. \quad (61)
\]

27
For the branches for which the upper level \(J\) is greater than or equal to \(b_n\) we use a similar argument. Again if \(A \cap A' \notin \{A, A', \emptyset\}\) then by definition \(J_A = K_A = n\) or \(J_{A'} = K_{A'} = n\) and thus by (30) it holds that \(\mathbb{E}(L_A \mid \mathcal{G}) = 0\) or \(\mathbb{E}(L_{A'} \mid \mathcal{G}) = 0\). Thus by (60) we obtain for \(1 \leq i \leq r\) fixed that

\[
\mathbb{E}
\left(\sum_{|A|=i} \left(1_{\mathcal{F}_A} - V_{A,h}\right) \cdot \left(\mathbb{E}(L_A \mid \mathcal{G}) - \mathbb{E}(L_A)\right) \cdot 1_{\{J_A \geq b_n\}}\right)^2
\leq \sum_A \mathbb{E}\left(\mathbb{E}(L_A \mid \mathcal{G})^2 + \mathbb{E}(L_A)^2; J_A \geq b_n\right)
+ \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot \mathbb{E}\left(\mathbb{E}(L_A \mid \mathcal{G})\mathbb{E}(L_{A'} \mid \mathcal{G}) + \mathbb{E}(L_A)^2\right) + \sum_{A \cap A' \notin \{A, A', \emptyset\}} \mathbb{E}(L_A)^2
\leq \sum_A \mathbb{E}\left(\frac{1}{J_A} + \mathbb{E}(L_A)^2; J_A \geq b_n\right) + \frac{1}{n},
\]

where we bounded the second sum by \(c \cdot \frac{1}{n}\) as above, for a finite constant \(c\) and for the third sum we used (55) and (51). Writing again for short \(J := J_{\{1, \ldots, i\}}\) and using (51) and Lemma 3 we obtain that

\[
\mathbb{E}
\left(\sum_{|A|=i} \left(1_{\mathcal{F}_A} - V_{A,h}\right) \cdot \left(\mathbb{E}(L_A \mid \mathcal{G}) - \mathbb{E}(L_A)\right) \cdot 1_{\{J_A \geq b_n\}}\right)^2
\leq \left(\frac{n}{i}\right) \sum_{j \geq b_n} \left(\frac{n}{j^2} + \frac{1}{\binom{n}{i}^2}\left(\frac{2}{j}\right)^2\right) \cdot \mathbb{P}(J = j) + \frac{1}{n}
\leq \left(\frac{n}{i}\right) \sum_{j \geq b_n} \frac{1}{j^2} \cdot \frac{2j}{n - 1} \cdot \frac{1}{\binom{n}{i}} + \frac{1}{n}
\leq \frac{1}{n} \cdot \log \frac{n}{b_n} + \frac{1}{n}
\leq \frac{1}{n} \cdot \log \log n.
\]

We are now left to consider the branches that are formed below and end above level \(b_n\) in the coalescent. Again, using the same argument as above and the fact that for \(A \cap A' \notin \{A, A', \emptyset\}\) the event \(\{J_A < b_n \leq K_A, J_{A'} < b_n \leq K_{A'}\}\) is empty, we have that

\[
\mathbb{E}
\left(\sum_{|A|=i} \left(1_{\mathcal{F}_A} - V_{A,h}\right) \cdot \left(\mathbb{E}(L_A \mid \mathcal{G}) - \mathbb{E}(L_A)\right) \cdot 1_{\{J_A < b_n \leq K_A\}}\right)^2
\leq \sum_A \mathbb{E}\left((1_{\mathcal{F}_A} - W_{A,h})^2\right) \cdot \mathbb{E}\left(\mathbb{E}(L_A \mid \mathcal{G})^2 + \mathbb{E}(L_A)^2; J_A < b_n \leq K_A\right)
+ \sum_{A \cap A' = \emptyset} \frac{1}{n} \cdot \mathbb{E}\left(\mathbb{E}(L_A \mid \mathcal{G})\mathbb{E}(L_{A'} \mid \mathcal{G}) + \mathbb{E}(L_A)^2\right)
\leq \sum_A \mathbb{P}(1_{\mathcal{F}_A} \neq 1_{\{B_{h,A}(A) = 1\}}) \cdot \mathbb{E}\left(\frac{1}{J_A} + \mathbb{E}(L_A)^2; J_A < b_n \leq K_A\right) + \frac{1}{n}.
\]

From Lemma 4 it holds that

\[
\mathbb{P}(1_{\mathcal{F}_A} \neq 1_{\{B_{h,A}(A) = 1\}}) \leq \frac{1}{n},
\]

28
which plugged in (63) leads to

\[
E\left(\left(\sum_{|A|=i} (1_{A} - V_{A,h}) \cdot (E(L_{A} | G) - E(L_{A})) \cdot 1_{(J_{A} < b_{n} \leq K_{A})}\right)^2\right)
\]

\[
\ll \frac{1}{n} \sum_{A} E\left(\frac{1}{|A|} + E(L_{A})^2; J_{A} < b_{n} \leq K_{A}\right) + \frac{1}{n}
\]

\[
\ll \frac{1}{n} \binom{n}{i} \sum_{k \geq b_{n}, j \leq b_{n}} \left(\frac{1}{j^2} + \frac{1}{(j-i)^2}\right) \cdot \mathbb{P}(K = k, J = j) + \frac{1}{n}
\]

\[
\ll \frac{1}{n} \binom{n}{i} \sum_{k \geq b_{n}, j \leq b_{n}} \left(\frac{1}{j^2} \cdot \left(\frac{2i}{n}\right)^2\right) \cdot \frac{1}{\binom{n}{i}} + \frac{1}{n}
\]

\[
\ll \frac{1}{n} \sum_{k \geq b_{n}} \log b_{n} \cdot \frac{1}{n^2} + \frac{1}{n}
\]

\[
\ll \frac{1}{n}
\]

Putting now together (61), (62) and (64) we obtain (58).

6.3 Proof of the Theorem

For \( h > 0 \) remember from (1) that, considering the coalescent tree of the population alive at time \( h \), \( I_{0,h}^{n} \) denotes the amount of external length in the coalescent at time \( h \) that is gathered in the region near the leaves of the tree, more precisely in the time interval \((0, h)\). We start by showing that

\[
I_{0,h}^{n} - E(I_{0,h}^{n}) = O_{P}(n^{-1/2}).
\]

Remember that \( h = \frac{t}{n} \), where \( t \) is measured on the evolutionary time scale, the scale on which also the lengths in the coalescent are measured.

Let \( \bar{X}_{j}, 2 \leq j \leq n \), denote the inter-coalescence times in the coalescent of the population alive at time \( h \) and consider the random time

\[
T := \sum_{j \geq \frac{t}{n}} \bar{X}_{j}.
\]

Since \( \bar{X}_{j} \sim \text{Exp}\left(\frac{1}{j}\right) \), it holds that \( E(T) = \frac{t}{n} + O\left(\frac{1}{n^{2}}\right) \) and \( V(T) = O\left(\frac{1}{n^{2}}\right) \). Observe that the quantity \( I_{0,T}^{n} - I_{0,T}^{n} \) is the external length gathered in an interval of length \( \left|\frac{t}{n} - T\right| \). Noticing that in this interval some of the \( n \) external branches may become internal, it holds that

\[
\left|I_{0,T}^{n} - I_{0,T}^{n}\right| \leq n \cdot \left|\frac{t}{n} - T\right|
\]

and thus

\[
I_{0,T}^{n} - E(I_{0,T}^{n}) = I_{0,T}^{n} - E(I_{0,T}^{n}) + O_{P}(n^{-1/2}).
\]

In order to estimate \( I_{0,T}^{n} \) we use Proposition 3 in [JK11]. The arguments used in its proof are valid also for \( \alpha \) and \( \beta \) depending on \( n \). Choosing \( \alpha \) such that \( n^{\alpha} = \frac{2}{2+\beta} n \) and \( \beta = 1 \) the proof gives

\[
V(I_{0,T}^{n}) = 8 \left(1 - \frac{\log \left(\frac{2}{2+\beta} n\right)}{\log n}\right) \cdot \frac{\log n}{n} + O(n^{-1}) = O(n^{-1}),
\]

and therefore (65) holds.
For ease of notation let us now consider two times $0 < h_1 < h_2$. We show that for any $\alpha_1, \alpha_2 \in \mathbb{R}$ the linear combination
\[
\sqrt{\frac{n}{4 \log n}} \left( \alpha_1 (L_{h_1}^n - \mathbb{E}(L_{h_1}^n)) + \alpha_2 (L_{h_2}^n - \mathbb{E}(L_{h_2}^n)) \right)
\]
of the (centred and rescaled) lengths at the two times converges in distribution to a normal distributed random variable. The argument can be immediately extended for linear combinations of the lengths at times $0 < h_1 < \cdots < h_k, k \in \mathbb{N}$.

Recall that for each $r \in \mathbb{N}$
\[
L_{h_i}^n = \sum_{|A| \leq r} 1_{\mathcal{F}_A(h_i)} \cdot L_A + \tilde{R}_{h_i}^{n,r} + \tilde{R}_h^{n,r} + I_{h_i}^n,
\]
where we used the notation $\mathcal{F}_A(h_i)$ for the event that the branch with leaves in the set $A$ from the coalescent at time 0 is free at time $h_i, i \in \{1,2\}$. The second and third term have been treated in Proposition 2 and we now focus on the first term. From (59) and (60)
\[
\sum_{|A| \leq r} 1_{\mathcal{F}_A(h_i)} \cdot L_A - \mathbb{E}\left( \sum_{|A| \leq r} 1_{\mathcal{F}_A(h_i)} \cdot L_A \right) = \sum_{|A| \leq r} 1_{\mathcal{F}_A(h_i)} (\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A)) + O_P(n^{-1/2}).
\]
Equation (58) allows to replace the dependent random coefficients $1_{\mathcal{F}_A(h_1)}$ by the independent coefficients $V_{A,h_1}$ defined in (59). In order to obtain the corresponding independent random coefficients for the time $h_2$ we let the birth and death processes $\tilde{W}_{A,h_1}$ and $W_{A,h}$ used to define $V_{A,h_1}$ run further until time $h_2$ and define the coefficients $V_{A,h_2}$ as in (59) by using the states of the processes $\tilde{W}_{A,h_2}$ and $W_{A,h_2}$ at time $h_2$. Note that the argument from the previous subsection used to prove (58) continues to hold also for the time $h_2$ and the coefficients defined as above. We thus obtain using (68), (69) and Proposition 2 that for $r \in \mathbb{N}$
\[
\sqrt{\frac{n}{4 \log n}} \left( \alpha_1 (L_{h_1}^n - \mathbb{E}(L_{h_1}^n)) + \alpha_2 (L_{h_2}^n - \mathbb{E}(L_{h_2}^n)) \right)
= \sqrt{\frac{n}{4 \log n}} \sum_{|A| \leq r} \left( \alpha_1 V_{A,h_1} + \alpha_2 V_{A,h_2} \right) \cdot (\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A))
+ \sqrt{\frac{n}{4 \log n}} \left( \alpha_1 (\tilde{R}_{h_1}^{n,r} - \mathbb{E}(\tilde{R}_{h_1}^{n,r})) + \alpha_2 (\tilde{R}_{h_2}^{n,r} - \mathbb{E}(\tilde{R}_{h_2}^{n,r})) \right)
+ O_P\left( \left( \log \log n \right) / \log n \right)^{1/2}.
\]

We now argue that the collection of random variables $\{ \sqrt{\frac{n}{4 \log n}}(\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A)) \}, 1 \leq i \leq r, |A| = i$, fulfils the assumptions of Proposition 1. First we make use of the main result of [DK13] which says that the (centred and rescaled) internal lengths of different orders are asymptotically normal and independent. Together with (57) it follows that
\[
\left( \sqrt{\frac{n}{4 \log n}} \sum_{|A| = i} (\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A)) \right)_{i=1,\ldots,r} \xrightarrow{d} N(0, I_r)
\]
where $I_r$ is the $r \times r$ identity matrix. The fact that the other two assumptions of Proposition 1 hold has been proved in Lemma 3. Therefore, we can apply Proposition 1 for the collection of random variables $\{ \sqrt{\frac{n}{4 \log n}}(\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A)) \}, 1 \leq i \leq r, |A| = i$, and the independent random coefficients $\{ \alpha_1 V_{A,h_1} + \alpha_2 V_{A,h_2} \}, 1 \leq i \leq r, |A| = i$.

Letting now $n \to \infty$ and using again Proposition 2 we obtain that the linear combination of the external lengths at the times $h_1$ and $h_2$ is asymptotically normal distributed. Note that we may interchange the limits.
in $r$ and $n$ since Proposition\textsuperscript{2} gives an uniform estimate in $r$. Therefore the stationary process $\mathcal{L}^n$ converges in finite distributions to a stationary Gaussian process. It remains to compute its covariance function. 

Observe that for $h_1 = 0$ and $h_2 = h$

$$\alpha_1 V_{A,h_1} + \alpha_2 V_{A,h_2} = \alpha_1 \delta_{1,|A|} + \alpha_2 V_{A,h}$$

with second moment from\textsuperscript{30}

$$m_A = \begin{cases} 
\alpha_1^2 + 2\alpha_1\alpha_2 \left( \frac{2}{h+2} \right)^2 + \alpha_2^2 \left( \frac{2}{h+2} \right)^2, & \text{if } |A| = 1 \\
\alpha_2^2 \cdot \left( \frac{h}{h+2} \right)^{i-1} \left( \frac{2}{h+2} \right)^2, & \text{if } |A| = i > 1.
\end{cases}$$

Therefore from Proposition\textsuperscript{1} \[ \sqrt{\frac{n}{4 \log n}} \sum_{|A| \leq r} \left( \alpha_1 V_{A,0} + \alpha_2 V_{A,h} \right) \cdot (\mathbb{E}(L_A | \mathcal{G}) - \mathbb{E}(L_A)) \]

\[ \overset{d}{\rightarrow} N \left( 0, \alpha_1^2 + 2\alpha_1\alpha_2 \left( \frac{2}{h+2} \right)^2 + \alpha_2^2 \left( \frac{2}{h+2} \right)^2 \sum_{i=0}^{r} i \cdot \left( \frac{h}{h+2} \right)^{i-1} \right). \]

Taking the limit $r \to \infty$ we obtain that

$$\sqrt{\frac{n}{4 \log n}} \left( \alpha_1 (L^0_n - \mathbb{E}(L^0_n)) + \alpha_2 (L^h_n - \mathbb{E}(L^h_n)) \right) \overset{d}{\rightarrow} N \left( 0, \alpha_1^2 + 2\alpha_1\alpha_2 \left( \frac{2}{h+2} \right)^2 + \alpha_2^2 \right).$$

This gives the covariance function.

The almost sure continuity of the paths follows from Lemma 6.4.6 in [MR06] using stationarity. There one has to check that there exists a $\delta > 0$ such that

$$\int_0^\delta \frac{\sigma^+(u)}{u \left( \log \frac{1}{u} \right)^{1/2}} du < \infty,$$

where

$$\sigma^+(u) := \sup_{|t-s| \leq u, t,s \in [-1/2, 1/2]} (\mathbb{E}(L^1_t - L^1_s)^2)^{1/2}.$$

This criterion is easily verified in our case, which finishes the proof.

References

[AN72] Athreya, K.B., Ney, P.E. (1972) Branching Processes, Springer

[BBL12] Berestycki, J., Berestycki, N., Limic, V. (2012) Asymptotic sampling formulae for Lambda-coalescents, to appear in Ann. INST. H. Poincaré, arXiv: 1201.6512

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

[BBS08] Berestycki, J., Berestycki, N., Schweinsberg, J. (2008) Small time properties of Beta-coalescents, Ann. INST. H. Poincaré 44, 214 - 238

[DK13] Dahmer, I., Kersting G., (2013) The internal branch lengths of the Kingman coalescent, to appear in Ann. Appl. Probab., arXiv: 1303.4562
[DKW14a] Dahmer, I., Kersting G., Wakolbinger, A., (2014) The total external branch length of Beta-coalescents, Combinatorics, Probability and Computing, Special Issue on Analysis of Algorithms, 23, Special Issue 06, 1010-1027

[DKW14b] Dahmer, I., Knobloch, R., Wakolbinger, A., (2014) The Kingman tree length process has infinite quadratic variation, arXiv:1402.2113

[DGP12] Depperschmidt, A., Greven, A., Pfaffelhuber, P. (2012) Tree-valued Fleming-Viot dynamics with mutation and selection, Ann. Appl. Probab. 22(6), 2560-2615

[DY12] Dherin, J.-S., Yuan, L. (2012) Asymptotic behavior of the total length of external branches for Beta-coalescents, arXiv: 1202.5859

[DIMR07] Drmota, M., Iksanov, A., Möhle, M., Rösl, U. (2007) Asymptotic results about the total branch length of the Bolthausen-Sznitman coalescent, Stoch. Proc. Appl. 117, 1404 - 1421

[EK86] Ethier, S.N., Kurtz, T.G. (1986) Markov processes. Characterization and convergence, John Wiley and sons

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

[GPW13] Greven, A., Pfaffelhuber, P., Winter, A. (2013) Tree-valued resampling dynamics. Martingale problems and applications, Prob. Theo. Rel. Fields 155, 789-838

[GPW09] Greven, A., Pfaffelhuber, P., Winter, A. (2009) Convergence in distribution of random metric measure spaces: (Lambda-coalescent measure trees), Prob. Theo. Rel. Fields 145(1), 285-322

[Gu14] Gufler, S. (2014) Lookdown representation for tree-valued Fleming-Viot processes, arXiv:1404.3682

[IM07] Iksanov, A., 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

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

[Ke12] Kersting, G. (2012) The asymptotic distribution of the length of Beta-coalescent trees, Ann. Appl. Probab. 22, 2086 - 2107

[KSW14] Kersting, G., Schweinsberg, J., Wakolbinger, A. (2014) The evolving beta coalescent, Electron. J. Probab. 19, 1-27

[Moe10] 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

[MR06] Marcus, M.B., Rosen, J. (2006) Markov Processes, Gaussian Processes and Local Times, Cambridge University Press

[PWW11] Pfaffelhuber, P., Wakolbinger. A., Weisshaft, H. (2011) The tree length of an evolving coalescent, Prob. Theo. Rel. Fields, 151, 529 - 557

[PP13] Pokalyuk, C., Pfaffelhuber, P. (2013) The ancestral selection graph under strong directional selection, Theor. Pop. Biol., 87, 25 - 33

[Sch12] Schweinsberg, J. (2012) Dynamics of the evolving Bolthausen-Sznitman coalescent, Electron. J. Probab., 91, 150