Some large deviations in Kingman’s coalescent

Andrej Depperschmidt, Peter Pfaffelhuber
and Annika Scheuringer

June 24, 2014

Abstract

Kingman’s coalescent is a random tree that arises from classical population genetic models such as the Moran model. The individuals alive in these models correspond to the leaves in the tree and the following two laws of large numbers concerning the structure of the tree-top are well-known: (i) The (shortest) distance, denoted by $T_n$, from the tree-top to the level when there are $n$ lines in the tree satisfies $nT_n \xrightarrow{\text{a.s.}} 2$ almost surely; (ii) At time $T_n$, the population is naturally partitioned in exactly $n$ families where individuals belong to the same family if they have a common ancestor at time $T_n$ in the past. If $F_{i,n}$ denotes the size of the $i$th family, then $n\left(F_{1,n}^2 + \cdots + F_{n,n}^2\right) \xrightarrow{\text{a.s.}} 2$ almost surely. For both laws of large numbers we prove corresponding large deviations results. For (i), the rate of the large deviations is $n$ and we can give the rate function explicitly. For (ii), the rate is $n$ for downwards deviations and $\sqrt{n}$ for upwards deviations. For both cases we give the exact rate function.

1 Introduction

Kingman’s coalescent is a random tree introduced by [Kingman (1982)] as the genealogy arising in large population genetic models. It has infinitely many leaves and is usually constructed from leaves to the root as follows: given that there are $k$ lines in the tree, after some exponential time with rate $\binom{k}{2}$, two lines are chosen uniformly and merged to one line, leaving the tree with $k-1$ lines. Due to the quadratic rate $\binom{k}{2}$ the tree immediately comes down from infinity to finitely many leaves ([Donnelly (1991)]). Since the seminal paper by [Pitman (1999)] this random tree has been generalized to other infinite trees arising in population genetics models.

For the Kingman coalescent some laws of large numbers and central limit theorems have been proved. They are nicely summarized in [Aldous (1999)], Chapter 4.2; see also Proposition 2.1 below. For $\varepsilon > 0$ let $N_\varepsilon$ denote the number of lines time $\varepsilon$ in the past. Then, since the Kingman coalescent immediately comes down from infinity, $N_\varepsilon$ is finite. Furthermore it is approximately $2/\varepsilon$. Equivalently, the time $T_\varepsilon$ it takes the coalescent to go from infinitely many lines to $n$ lines is approximately $2/n$ for large $n$. Going to the fine structure, at time $T_n$ the infinite population is decomposed in $n$ families (whose joint distribution is exchangeable) and every leaf in the tree belongs to exactly one of the $n$ families whose frequencies are denoted by $F_{1,n}, \ldots, F_{n,n}$. It is known that for large $n$ a randomly chosen $F_{i,n}$ is approximately exponentially distributed with mean $1/n$. This translates into several laws of large numbers; see e.g. (35) in [Aldous (1999)].

AMS 2010 subject classification. 60F10, 60G09

Keywords and phrases. Kingman coalescent, large deviations, uniform spacings
In particular, the probability of picking (from the initial infinite population) two leaves that belong to the same family, given by $F^2_{1,n} + \cdots + F^2_{n,n}$, is approximately $2/n$.

The main goal of the present paper is to study the corresponding large deviations results. To the best of our knowledge, except for [Angel et al. (2012), cf. Remark 2.6], results in this direction are not present in the literature. We formulate our results in the next section. Theorem 1 gives a full large deviation principle for the distributions of $nT_n$. The proof, given in Section 3, is an application of the Gärtner-Ellis Theorem. As a byproduct, we derive a large deviation principle for the distributions of $\varepsilon N_\varepsilon$ in Corollary 2.4. Large deviations of $n(F^2_{1,n} + \cdots + F^2_{n,n})$ are considered in Theorem 2, and exact rate functions for downwards and upwards deviations are given. The proof is given in Section 4.2. For the upward deviations we use a variant of Cramér’s theorem for heavy-tailed random variables; see e.g. [Gantert et al. (2014)]. For the downward deviations we use a connection to self-normalized large deviations; see [Shao (1997)]. This connection was pointed out to us by Alain Rouault and Nina Gantert. Since the rate function for downward deviations is hard to treat analytically we provide in Theorem 3 a simple lower bound. The proof of that bound is given in Section 4.3.

## 2 Main results

The Kingman coalescent can be seen as a discrete graph, more precisely a discrete tree with infinitely many leaves. Let $S_2, S_3, \ldots$ be independent exponentially distributed random variables with mean 1. Then the Kingman coalescent tree can be constructed from the root to the leaves as follows.

1. Start the tree with two lines from the root.
2. For $k \geq 2$ the tree stays with $k$ lines for the amount of time $S_k/(\binom{k}{2})$. After that time, one of the $k$ lines is randomly chosen. This line splits in two so that the number of lines jumps from $k$ to $k + 1$.
3. Stop upon reaching infinitely many lines, which happens after (the almost surely finite) time $T_1 := \sum_{k=2}^{\infty} S_k/(\binom{k}{2})$.

The random variable $T_1$ is the total tree height. Alternatively, $T_1$ is the time to the most recent common ancestor (MRCA) of the infinite population (of leaves). Counted from the top of the tree at time $\varepsilon > 0$ a random number $N_\varepsilon$ of active lines in the Kingman tree is present, i.e.

$$
N_\varepsilon := \inf\{n : T_n < \varepsilon\} \quad \text{for} \quad T_n := \sum_{k=n+1}^{\infty} \frac{S_k}{\binom{k}{2}}.
$$

At time $T_n$ every leaf belongs to one of $n$ disjoint families and all members of each such family stem from the same line at time $T_n$. Let us denote the frequencies of these families (which exist due to exchangeability by deFinetti’s Theorem) by $F_{1,n}, \ldots, F_{n,n}$.

The following results are well known (see [Aldous (1999) for (2.2) and (2.3) and [Evans (2000) for (2.4)]; proofs can also be found in [Depperschmidt et al. (2013)].)

**Proposition 2.1 (Laws of large numbers).**

Let $(T_n)_{n=1,2,\ldots}$ and $(N_\varepsilon)_{\varepsilon>0}$ and $(F_{1,n}, \ldots, F_{n,n})_{n=1,2,\ldots}$ be as above. Then

$$
nT_n \xrightarrow{n \to \infty} 2 \quad \text{almost surely,} \quad (2.2)
$$

$$
\varepsilon N_\varepsilon \xrightarrow{\varepsilon \to 0} 2 \quad \text{almost surely,} \quad (2.3)
$$

2
\[
\sum_{k=1}^{n} F_{k,n}^{2} \xrightarrow{n \to \infty} 2 \quad \text{almost surely.} \tag{2.4}
\]

**Remark 2.2** (Interpretation of (2.4)). We note that the left hand side of (2.4) has the interpretation of a **homozygosity by descent** in the following sense: when picking two leaves from the tree at time 0, the probability that both share a common ancestor at time \(T_n\) is \(\sum_{k=1}^{n} F_{k,n}^{2}\). Then, the law of large number states that the homozygosity by descent at time \(T_n\) is approximately \(2/n\) for large \(n\).

In the present paper we are interested in large deviations results corresponding to the statements of Proposition 2.1. We start with large deviations connected with (2.2). First we introduce some notation. For \(n = 1, 2, \ldots\), let \(\mu_n\) denote the distribution of \(nT_n\), i.e. \(\mu_n(\cdot) = P(nT_n \in \cdot)\). Furthermore we denote by \(B(\mathbb{R})\) the Borel \(\sigma\)-algebra on \(\mathbb{R}\) and for \(\Gamma \in B(\mathbb{R})\) we denote by \(\Gamma^{\text{int}}\) the interior and by \(\overline{\Gamma}\) the closure of \(\Gamma\). For \(x > 0\), let \(t_x < 1\) be the unique solution of the equation \(x = f(t)\), where the continuous and increasing function \(f: (-\infty, 1) \to (0, \infty)\) is defined by (see Figure 1 for a plot)

\[
f(t) := \begin{cases} 
\frac{1}{\sqrt{t}} \log \left(1 + \frac{\sqrt{t}}{1 - \sqrt{t}}\right) & : 0 < t < 1, \\
2 & : t = 0, \\
\frac{2}{\sqrt{|t|}} \arctan \sqrt{|t|} & : t < 0.
\end{cases} \tag{2.5}
\]

The proof of the following theorem is given in Section 3.1.

**Theorem 1** (LDP for \((\mu_n)_{n=1,2,\ldots}\)). The sequence \((\mu_n)_{n=1,2,\ldots}\) satisfies a large deviation principle with scale \(n\) and good rate function \(I\) given by

\[
I(x) := \begin{cases} 
\frac{tx}{2} + \int_{1}^{\infty} \log \left(1 - \frac{tx}{y^2}\right) dy & : x > 0, \\
\infty & : x \leq 0.
\end{cases} \tag{2.6}
\]

In other words, for any \(\Gamma \in B(\mathbb{R})\) we have

\[-\inf_{x \in \Gamma} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \Gamma} I(x) .\]

**Remark 2.3** (Interpretation). Both, the function \(f\) from (2.5) and \(I\) from (2.6) are plotted in Figure 1. The minimum of the rate function is attained at \(x = 2\). This fact is clear from the law of large numbers, (2.2). In addition, \(I(x) = \infty\) for \(x \leq 0\) because \(nT_n > 0\) almost surely.

Let us now have a closer look at the behaviour of \(I(x)\) for \(x\) near 0 and for large \(x\). Since \(2 \arctan(t) \xrightarrow{t \to \infty} \pi\), we have that \(\sqrt{tx|x|x^{-1}0} \xrightarrow{t \to \infty} \pi\), and hence, \(tx^{-1}0 \xrightarrow{t \to \infty} \pi^2/2\). In this case,

\[
xI(x) \xrightarrow{x \to 0} -\frac{x^2}{2} + x \int_{1}^{\infty} \log \left(1 + \frac{\pi^2}{x^2y^2}\right) dy \quad \text{and} \quad \frac{x^2}{2} + x \int_{0}^{\infty} \log \left(1 + \frac{1}{y^2}\right) dz = \frac{\pi^2}{2} . \tag{2.7}
\]
Figure 1: This figure displays the functions $f$ and $I$ from (2.5) and (2.6), respectively.

where the last equality follows from $\int_{-1}^{0} \log(1 + 1/z^2)dz = \pi$. To understand the behaviour for large $x$, note that since

$$f(t) \approx \log 2 - \log(1 - \sqrt{t}),$$

for $x \to \infty$ we have $2/(1 - \sqrt{t_x}) \approx e^x$ and in particular $t_x \approx (1 - 2e^{-t})^2 \approx 1$. It follows

$$I(x) \approx \frac{1}{2} + x \int_{0}^{\infty} \log(1 - 1/y^2)dy \approx \frac{1}{2}.$$

Note that (2.2) and (2.3) are equivalent. Indeed, $\{T_n \geq \varepsilon\} = \{N_{\varepsilon} \geq n\}$ (this also holds with $\geq$ replaced by $\leq$) by construction, and $T_n \downarrow 0$ as $n \to \infty$ and $N_{\varepsilon} \uparrow \infty$ as $\varepsilon \to 0$. Hence, Theorem 1 translates into a large deviation principle for $\varepsilon N_{\varepsilon}$. In the following we denote by $\nu_{\varepsilon}$ the distribution of $\varepsilon N_{\varepsilon}$, i.e.

$$\nu_{\varepsilon}(\cdot) = \mathbb{P}(\varepsilon N_{\varepsilon} \in \cdot).$$

The proof of the next result is given in Section 3.2; see Figure 2 for a plot of the rate function $\hat{I}$.

**Corollary 2.4** (LDP for $(\nu_{\varepsilon})_{\varepsilon > 0}$). For $\varepsilon \downarrow 0$ the family $(\nu_{\varepsilon})_{\varepsilon > 0}$ satisfies a large deviation principle with scale $1/\varepsilon$ and good rate function $\hat{I}$ given by

$$\hat{I}(x) = \begin{cases} xI(x) & : x > 0, \\ \frac{x^2}{2} & : x = 0, \\ \infty & : x < 0, \end{cases}$$

with $I$ from (2.6). In particular, for $\Gamma \in \mathcal{B}(\mathbb{R})$ we have

$$-\inf_{x \in \Gamma} \hat{I}(x) \leq \lim_{\varepsilon \to 0} \inf_{\varepsilon > 0} \varepsilon \log \nu_{\varepsilon}(\Gamma) \leq \lim_{\varepsilon \to 0} \sup_{\varepsilon > 0} \varepsilon \log \nu_{\varepsilon}(\Gamma) \leq -\inf_{x \in \Gamma} \hat{I}(x).$$

**Remark 2.5** (The full distribution of $N_{\varepsilon}$). The distributions $\nu_{\varepsilon}, \varepsilon > 0$ (as well as $\mu_n, n = 1, 2, \ldots$) have been described explicitly in the literature. [Tavaré 1984, Section 6, gives]

$$\mathbb{P}(N_{\varepsilon} = n) = \sum_{k=n}^{\infty} e^{-\varepsilon} \frac{(-1)^{k-n} (2k-1) \cdots (n + k - 2)}{n!(k-n)!}.$$

In principle, this formula must also give the large deviations for the measures $\nu_{\varepsilon}$, but this does not seem straight-forward.
Figure 2: The figure on the left shows the rate function $\hat{I}$ from Corollary 2.4. The figure on the right is a comparison of $\hat{I}$ with the lower bound obtained from Angel et al. (2012).

**Remark 2.6** (The rate function $\hat{I}$ and comparison with Angel et al. (2012)). Although the main goal of Angel et al. (2012) was the analysis of spatial $\Lambda$-coalescents, they also provide some large deviations bounds on Kingman’s coalescent. These bounds are mainly based on Markov inequality. Precisely, in Lemma 2.2 in Angel et al. (2012) it is shown that for $0 < x < \frac{1}{2}$

$$P(|\varepsilon N_\varepsilon - 2| > x) < e^{-\frac{x^2}{4}}$$

and therefore

$$\limsup_{\varepsilon \to 0} \varepsilon \log P(|\varepsilon N_\varepsilon - 2| > x) \leq -\frac{x^2}{4}.$$ 

In the neighbourhood of 2 the last inequality translates easily into a bound for the rate function $\hat{I}$ from (2.8); see Figure 2. Namely, for $x \in (1.5, 2.5)$ we have

$$\hat{I}(x) \geq \frac{(x - 2)^2}{4}.$$ 

Next, we state some large deviations results connected to (2.4). For

$$W_n := n \sum_{k=1}^{n} F_{k,n}^2$$

we know from (2.4) that $W_n \xrightarrow{n \to \infty} 2$ holds almost surely. The proof of this result is based on the well-known fact (see e.g. Section 5 in Kingman (1982)) that the distribution of $W_n$ can be derived using uniform order statistics: Let $U_1, \ldots, U_{n-1}$ be independent and uniformly distributed on $[0, 1]$, and $0 < U_{(1)} < \cdots < U_{(n-1)} < 1$ be their order statistics. Additionally, let $R_1, \ldots, R_n$ be independent exponentially distributed random variables with mean 1. Then,

$$(F_{1,n}, \ldots, F_{n,n}) \xrightarrow{d} (U_{(1)}, U_{(2)} - U_{(1)}, \ldots, U_{(n-1)} - U_{(n-2)}, 1 - U_{(n-1)})$$

(2.9)

Here the second equality in distribution is one of the well known representations of uniform spacings; see e.g. Section 4.1 in Pyke (1965). It follows

$$W_n \xrightarrow{d} n \frac{1}{n} \sum_{k=1}^{n} R_k^2 \left( \frac{1}{n} \sum_{j=1}^{n} R_j \right)^2 = \frac{1}{n} \sum_{k=1}^{n} R_k^2 \left( \frac{1}{n} \sum_{j=1}^{n} R_j \right)^2.$$ 

(2.10)
We will use this representation to obtain large deviations results for $W_n$. In particular we show that upwards large deviations of $W_n$ are on the scale $\sqrt{n}$ while downwards large deviations are on the scale $n$. The proof is given in Section 4.2.

**Theorem 2** (Large deviations of $W_n$). For each $x \geq 2$, we have

$$
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \log P \left( W_n \geq x \right) = -\sqrt{x-2}.
$$

(2.11)

Furthermore $P(W_n < 1) = 0$ and for each $1 < x < 2$, we have

$$
\lim_{n \to \infty} \frac{1}{n} \log P \left( W_n \leq x \right) = -\tilde{I}(x).
$$

(2.12)

The function $\tilde{I}(x)$ is positive for $1 < x < 2$ and is given by

$$
\tilde{I}(x) := \sup_{c \geq 0} \inf_{t \geq 0} M(x, c, t).
$$

(2.13)

**Theorem 3** (Lower bound on downwards large deviations of $W_n$). For $1 < x < 2$ we have

$$
\liminf_{n \to \infty} \frac{1}{n} \log P \left( W_n \leq x \right) \geq 1 - \frac{1}{\sqrt{x-1}}.
$$

(2.15)

**Remark 2.8** (Rationale and use of the representation in Lemma 2.7). The main point in the proof of Lemma 2.7 is that $W_n$ does not depend on the order of the $R_k$ and hence we can as well order them according to their size.

Let us briefly explain how we will use (2.14) in the proof of (2.15). Since $W_n$ is minimal if $R_1 = \cdots = R_n$ (whence $W_n = 1$), we have to look for possibilities that all $R_k$’s are of about the same size in order to obtain a large deviations result for $W_n$. Let $R_{(1)}, \ldots, R_{(n)}$ denote the above exponential random variables ordered in increasing order, i.e. $R_{(i)}$ is the $i$th smallest value. Using “competing exponential clocks” arguments (see also the proof of the lemma) one can see that $R_{(i)} - R_{(i-1)}$ is exponentially distributed with mean $1/(n-i+1)$. Hence, one way of obtaining similar values for all $R_k$’s arises if $R_{(1)}$ is particularly large, which then leads to a large deviations result for $W_n$. 

Figure 3: Numerical comparison of the exact rate function for downwards large deviations of $W_n$ from (2.12) in Theorem 2 and the lower bound from (2.15) in Theorem 3.

Remark 2.9 (Interpretation, uniform spacings and the Poisson-Dirichlet distribution).
1. Let us give some heuristics about the rates arising in Theorem 2. For (2.11), we have to ask ourselves about the easiest way $W_n$ becomes too large. From (2.9), we see that this is the case if one of the $R_k$’s is too large, making this kind of deviations a local property in the sense that only a single of the $R_k$’s has to show some untypical behavior. This is different when looking at (2.12), i.e., too small values of $W_n$. First, observe that $W_n$ is small only if all (or many) families have about equal sizes (extreme case $F_{1,n} = \cdots = F_{n,n} = \frac{1}{n}$ gives the minimal value $W_n = 1$). Hence, such downward deviations require to study a global property of the random variable $W_n$, which is significantly harder. For the proof of (2.12) we will interpret $W_n$ as a self-normalised sum and use from Shao (1997) a result on large deviations result for such sums.

2. From (2.9), we see that in fact $W_n$ is a function of uniform order statistics, which, for instance, have been studied in detail (although no large deviations results were given) in Pyke (1965). Hence, Theorem 2 may as well be interpreted as a large deviations result for uniform order statistics.

3. As stated in Remark 2.2, $W_n/n$ can be interpreted as homozygosity at time $T_n$. Using a Poisson process along the tree with intensity $\theta/2$, we can ask for the probability of picking two leaves from the tree which are not separated by a Poisson mark, denoted by homozygosity in state, abbreviated by $H_{\theta}/\theta$. This quantity is closely related to the Poisson-Dirichlet distribution and some large deviations (in the limit of large $\theta$) were derived in Dawson and Feng (2006). It is shown there in Theorem 5.1 that $H_{\theta}/\theta \xrightarrow{\theta \to \infty} 0$ and that

$$P(H_{\theta} > \theta x) = e^{-\theta (I(\sqrt{\theta}) + o(1))}$$

for $I(x) = -\log(1 - x)$. However, a large deviation principle for the quantity $H_{\theta}$ (noting that $H_{\theta} \xrightarrow{\theta \to \infty} 1$), which corresponds to the results from Theorem 2, could not be obtained by Dawson and Feng (2006). At least, it was shown that its scale cannot be larger than $\sqrt{\theta}$.
3 Proof of Theorem 1 and Corollary 2.4

3.1 Proof of Theorem 1

The proof of Theorem 1 is an application of the Gärtner-Ellis theorem; see for instance Section 2.3 in [Dembo and Zeitouni (2010)].

Let $\Lambda_n(t) := \log \mathbb{E}[e^{\frac{t}{n}T_n}]$ and $\mu_n(\cdot) = P(nT_n \in \cdot)$. To show that the sequence $\mu_1, \mu_2, \ldots$ satisfies a large deviation principle with scale $n$ and a good rate function we need to check the following three conditions.

GE1 $\Lambda(t) := \lim_{n \to \infty} \frac{1}{n} \Lambda_n(nt)$ exists for all $t$ as a limit in $\mathbb{R} = \mathbb{R} \cup \{\pm \infty\}$. Furthermore $t \to \Lambda(t)$ is lower-semicontinuous, $0 \in D^\circ_\Lambda$, where $D_\Lambda := \{t : \Lambda(t) < \infty\}$.

GE2 $\Lambda$ is differentiable on $D^\circ_\Lambda$.

GE3 $\Lambda$ is steep, i.e. $\Lambda'(t_n) \xrightarrow{n \to \infty} \infty$ whenever $t_1, t_2, \ldots \in D^\circ_\Lambda$ and $t_n \xrightarrow{n \to \infty} t \in \partial D_\Lambda$.

Then the good rate function is given by

$$x \mapsto I(x) = \sup_{t \in \mathbb{R}} (tx - \Lambda(t)). \quad (3.1)$$

We proceed in three steps. First, we compute $\Lambda(t) := \lim_{n \to \infty} \frac{1}{n} \Lambda_n(nt)$. Second, we check the further assumptions of the Gärtner-Ellis theorem and obtain $I$ as the Fenchel-Legendre transform of $\Lambda$. In the third step, for the rate function $I$ from (3.1) we obtain its simplified form given in Theorem 1

**Step 1. The limit of $\frac{1}{n} \Lambda_n(nt)$:** We will show that

$$\Lambda(t) = \lim_{n \to \infty} \frac{1}{n} \Lambda_n(nt) = \begin{cases} -\int_1^\infty \log \left(1 - \frac{2t}{x}\right) \, dx & : t \leq \frac{1}{2}, \\ \infty & : t > \frac{1}{2}. \end{cases} \quad (3.2)$$

For this, recall from (2.7) that $T_n = \sum_{k=n+1}^\infty S_k / \binom{k}{2}$ where $S_k / \binom{k}{2}$ is exponentially distributed with rate $\binom{k}{2}$ as well as independent of $S_\ell$ for all $\ell \neq k$. Furthermore recall that the moment generating function of an exponentially distributed random variable $R$ with rate $\lambda > 0$ is given by

$$\mathbb{E}[e^{\lambda R}] = \begin{cases} \frac{1}{1-\lambda t} & : t < \lambda, \\ \infty & : t \geq \lambda. \end{cases} \quad (3.3)$$

Hence, for each $n \in \mathbb{N}$ and $t \in \mathbb{R}$ we obtain by the monotone convergence theorem

$$\varphi_n(nt) := \mathbb{E}[e^{nt^2T_n}] = \mathbb{E}[e^{nt^2 \sum_{k=n+1}^\infty S_k / \binom{k}{2}}] = \prod_{k=n+1}^\infty \mathbb{E}[e^{nt^2 S_k / \binom{k}{2}}]. \quad (3.4)$$

We have to consider two cases $t > \frac{1}{2}$ and $t \leq \frac{1}{2}$ separately. First suppose that $t > \frac{1}{2}$. Then there exists $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$ we have

$$2t \geq \frac{n+1}{n} \quad \text{i.e.} \quad \frac{tn^2}{n+1} = 2t \cdot \frac{n}{n+1} > 1.$$ 

Consequently, using (3.3), we obtain $\mathbb{E}[e^{nt^2T_n}] = \infty$ for each $n \geq n_0$. Hence, $\varphi_n(nt) = \infty$ and $\Lambda_n(nt) = \log \varphi_n(nt) = \infty$ for $n$ large enough. Thus, we have

$$\Lambda(t) = \lim_{n \to \infty} \frac{1}{n} \Lambda_n(nt) = \infty \quad \text{for all } t > \frac{1}{2}.$$
Now suppose that $t \leq \frac{1}{2}$. For $n \in \mathbb{N}$ and $k \geq n + 1$ we have $\binom{k}{n} \geq tn^2$. Furthermore using (3.4) and (3.3) we can write

$$\varphi_n(nt) = \prod_{k=n+1}^{\infty} \frac{1}{1 - \frac{2tn^2}{k(k-1)}}.$$  

Using this we can rewrite $\frac{1}{n} A_n(nt)$ for $t \leq \frac{1}{2}$ as

$$\frac{1}{n} A_n(nt) = \frac{1}{n} \log \left( \prod_{k=n+1}^{\infty} \frac{1}{1 - \frac{2tn^2}{k(k-1)}} \right) = - \frac{1}{n} \sum_{k=n}^{\infty} \log \left( 1 - \frac{2t}{k} \right)$$

Using (3.4) and (3.3) we can write

$$\varphi_n(nt) = \prod_{k=n+1}^{\infty} \frac{1}{1 - \frac{2tn^2}{k(k-1)}}.$$  

Using this we can rewrite $\frac{1}{n} A_n(nt)$ for $t \leq \frac{1}{2}$ as

$$\frac{1}{n} A_n(nt) = \frac{1}{n} \log \left( \prod_{k=n+1}^{\infty} \frac{1}{1 - \frac{2tn^2}{k(k-1)}} \right) = - \frac{1}{n} \sum_{k=n}^{\infty} \log \left( 1 - \frac{2t}{k} \right)$$

and by the dominated convergence theorem we obtain

$$\frac{1}{n} A_n(nt) \xrightarrow{n \to \infty} - \int_1^{\infty} \log \left( 1 - \frac{2t}{x^2} \right) \, dx.$$  

Hence, GE1 is shown with $\Lambda$ as in (3.2). Moreover, we have $D_\Lambda = (-\infty, \frac{1}{2}]$, $\Lambda(\frac{1}{2}) = \int_1^{\infty} \log(1 - \frac{1}{x^2}) \, dx = \pi$ and $\Lambda$ is lower-semi-continuous.

**Step 2. Further assumptions of the Gärtner-Ellis theorem**: We proceed by checking the assumptions GE2 and GE3. For differentiability of $\Lambda$ for $t < \frac{1}{2}$ consider for $-\infty < r < 0 < s < \frac{1}{2}$ the function

$$f : (1, \infty) \times (r, s) \to \mathbb{R}$$

$$(x, t) \mapsto - \log \left( 1 - \frac{2t}{x^2} \right).$$

We have $\int_1^{\infty} |f(x, t)| \, dx < \infty$ for $(x, t)$ and the derivative

$$\frac{d}{dt} f(x, t) = \frac{2}{x^2 - 2t}$$

exists for each $x \in (1, \infty)$ and is continuous in $t$. Hence, we can interchange differentiation and integration and obtain

$$\Lambda'(t) = \int_1^{\infty} \frac{2}{x^2 - 2t} \, dx.$$  

Furthermore, for a sequence $t_1, t_2, \ldots$ with $t_n \uparrow \frac{1}{2}$ we obtain

$$\lim_{n \to \infty} \Lambda'(t_n) = \lim_{n \to \infty} \int_1^{\infty} \frac{2}{x^2 - 2t_n} \, dx = \lim_{n \to \infty} \frac{1}{\sqrt{2t_n}} \left( \log \left( 1 + \sqrt{2t_n} \right) - \log \left( 1 - \sqrt{2t_n} \right) \right) = \infty,$$

i.e. condition GE3 is also satisfied.

**Step 3. Properties of $I$**: Applying the Gärtner-Ellis theorem reveals that the sequence of distributions of $nT_n, n = 1, 2, \ldots$ satisfies a large deviation principle with good rate function

$$I(x) = \sup_{t \leq \frac{1}{2}} \left[ tx + \int_1^{\infty} \log \left( 1 - \frac{2t}{y^2} \right) \, dy \right] = \sup_{t \leq 1} \left[ \frac{t}{2} x + \int_1^{\infty} \log \left( 1 - \frac{t}{y^2} \right) \, dy \right].$$
In order to compute that supremum, we write for \( t \geq 0 \)
\[
2 \frac{\partial}{\partial t} \left[ \frac{t}{2} x + \int_1^{\infty} \log \left( 1 - \frac{t}{y^2} \right) dy \right] = x - 2 \int_1^{\infty} \frac{1}{y^2 - t} dy
\]
\[
= x + \frac{1}{\sqrt{t}} \int_1^{\infty} \frac{1}{y + \sqrt{t}} - \frac{1}{y - \sqrt{t}} dy
\]
\[
= x - \frac{1}{\sqrt{t}} \log \frac{1 + \sqrt{t}}{1 - \sqrt{t}}
\]
while for \( t \leq 0 \)
\[
2 \frac{\partial}{\partial t} \left[ \frac{t}{2} x + \int_1^{\infty} \log \left( 1 - \frac{t}{y^2} \right) dy \right] = x - 2 \int_1^{\infty} \frac{1}{y^2 + |t|} dy
\]
\[
= x - \frac{2}{\sqrt{|t|}} \arctan \sqrt{|t|}.
\]
It is easy to see that the second derivative is negative throughout, such that the supremum is attained at \( t = t_x \) given by the solution of \( f(t_x) = x \) for \( f \) as in (2.5). Finally we note that for \( t \in [0, 1) \) the range of \( t \mapsto \log \frac{1 + \sqrt{t}}{1 - \sqrt{t}} \) is \([2, \infty)\) and for \( t \in (-\infty, 0] \) the range of \( t \mapsto \frac{2}{\sqrt{|t|}} \arctan \sqrt{|t|} \) is \((0, 2]\). Hence, the scale function \( I \) is of the form given in (2.6).

### 3.2 Proof of Corollary 2.4

The proof is based on the fact that \( \{ T_n \geq \varepsilon \} = \{ N_{\varepsilon} \geq n \} \). Thus, for \( x \geq 2 \) we have
\[
xI(x) = x \lim_{n \to \infty} \frac{1}{n} \log P(nT_n \geq x) = \lim_{n \to \infty} \frac{x}{n} \log P\left( \frac{x}{n} N_{x/n} \geq x \right) = \lim_{\varepsilon \to 0} \varepsilon \log P(\varepsilon N_{\varepsilon} \geq x)
\]
and for \( 0 < x \leq 2 \)
\[
xI(x) = x \lim_{n \to \infty} \frac{1}{n} \log P(nT_n \leq x) = \lim_{n \to \infty} \frac{x}{n} \log P\left( \frac{x}{n} N_{x/n} \leq x \right) = \lim_{\varepsilon \to 0} \varepsilon \log P(\varepsilon N_{\varepsilon} \leq x).
\]
The value \( \hat{I}(0) \) follows from (2.7). Since the rate function \( I \) attains its minimum at \( x = 2 \), is decreasing below and increasing above \( 2 \), the result follows.

### 4 Proof of Lemma 2.7, Theorem 2 and Theorem 3

#### 4.1 Proof of Lemma 2.7

When looking at (2.10), note that \( W_n \) does not depend on the order of the \( R_k \)'s. Therefore, it is possible to order them according to their size. Precisely, let \( 0 < R_{(1)} < \cdots < R_{(n)} \) be their order statistics. Then it is well-known that
\[
(R_{(n)}, R_{(n-1)}, \ldots, R_{(1)}) \overset{d}{=} \left( \sum_{k=1}^{n} \frac{R_k}{n}, \sum_{k=2}^{n} \frac{R_k}{n}, \ldots, \frac{R_n}{n} \right), \quad \text{i.e.} \quad R_{(n-k+1)} \overset{d}{=} \sum_{i=k}^{n} \frac{R_i}{n}.
\]
Indeed, the smallest of \( n \) independent exponentially distributed mean 1 random variables is exponentially distributed with mean \( \frac{1}{n} \) (as does \( \frac{R_n}{n} \)), and the second smallest
then has the same distribution as \( \frac{R_n}{n} + \frac{R_{n+1}}{n+1} \) etc. Now, we obtain (2.14) as follows

\[
W_n \overset{d}{=} \frac{1}{n} \sum_{k=1}^{n} R_k^2 \overset{d}{=} \frac{1}{n} \sum_{k=1}^{n} \left( \sum_{i=k}^{n} \frac{R_i}{i} \right)^2 \overset{d}{=} \frac{1}{n} \left( 2 \sum_{j=1}^{n} \sum_{i=1}^{j} \sum_{k=1}^{n} \frac{R_i R_j}{ij} - \sum_{i=1}^{n} \sum_{k=1}^{n} R_i^2 \right) \overset{d}{=} \frac{1}{n} \left( 2 \sum_{j=1}^{n} \sum_{i=1}^{j} \frac{R_i R_j}{j} - \sum_{i=1}^{n} R_i^2 \right) \overset{d}{=} \frac{1}{n} \left( 2 \sum_{j=1}^{n} \sum_{i=1}^{j} \frac{R_i R_j}{j} - \sum_{i=1}^{n} R_i^2 \right).
\]

**4.2 Proof of Theorem 2**

We start by proving (2.11). Let \( x \geq 2 \) and let \( R_1, R_2, \ldots \) be independent exponential random variables with mean 1. In what follows we set

\[
X_n := \frac{1}{n} \sum_{k=1}^{n} R_k \quad \text{and} \quad Z_n := \frac{1}{n} \sum_{k=1}^{n} R_k^2.
\]

According to (2.10), it suffices to show that

\[
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \log P \left( \frac{Z_n}{X_n^2} \geq x \right) = -\sqrt{x-2}.
\]

(4.2)

To this end we will show that for all \( 0 < \varepsilon < 1 \),

\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \log P \left( \frac{Z_n}{X_n^2} \geq \frac{x}{1 - \varepsilon} \right) \leq -\sqrt{x-2}
\]

(4.3)

as well as

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n}} \log P \left( \frac{Z_n}{X_n^2} \geq x \right) \geq -\sqrt{(1+\varepsilon)x-2}
\]

(4.4)

and obtain (4.2) by letting \( \varepsilon \to 0 \). For (4.3) we have

\[
P \left( \frac{Z_n}{X_n^2} \geq \frac{x}{1 - \varepsilon} \right) \leq P \left( Z_n \geq x \right) + P \left( X_n \leq \sqrt{1 - \varepsilon} \right).
\]

We consider the two terms on the right hand side of the last display separately and start with the first one. Observe that \( E[e^{\lambda R_1^2}] = \infty \) for \( \lambda > 0 \), \( E[R_1^2] = 2 \) and \( P \left( R_1^2 \geq t \right) = e^{-\sqrt{t}} \) for \( t \geq 0 \). We use a variant of Cramér’s theorem for heavy-tailed random variables from Gantert et al. (2014). In particular, we refer to the statement around equation (1.2) there (the assumption there is fulfilled with \( X_1 \) replaced by \( R_1^2 \) and \( r = \frac{1}{2} \), \( m = 2 \) and \( c = 1 \)). We obtain

\[
P \left( Z_n \geq x \right) = e^{-\sqrt{\pi(\sqrt{x-2} + o(1))}} \quad \text{as } n \to \infty.
\]

(4.6)

For the second term on the right hand side of (4.5) by the (classical) Cramér theorem we obtain

\[
P \left( X_n \leq \sqrt{1 - \varepsilon} \right) = e^{-nI_{exp}(\sqrt{1-\varepsilon})(1+o(1))}, \quad \text{as } n \to \infty,
\]

(4.7)
where

\[ I_{\text{exp}}(y) := y - 1 - \log(y) \] (4.8)

is the Fenchel-Legendre transform of the function \( t \mapsto \log \mathbb{E}[e^{\lambda R_1}] \). Now, using (4.5), (4.6) and (4.7) we obtain

\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \log \mathbb{P}\left( \frac{Z_n}{X_n^2} \geq \frac{x}{1 - \varepsilon} \right) \leq \limsup_{n \to \infty} \frac{1}{\sqrt{n}} \log \left( e^{-\sqrt{n(\sqrt{x} - x + o(1))}} + e^{-n I_{\text{exp}}(\sqrt{1 - \varepsilon}(1 + o(1)))} \right) = -\sqrt{x - 2},
\]

which shows (4.3). For the proof of (4.4) we write

\[
\mathbb{P}(Z_n x_n \geq x) \geq \mathbb{P}(Z_n x_n \geq x, X_n^2 \leq 1 + \varepsilon) \geq \mathbb{P}(Z_n x_n \geq x(1 + \varepsilon)) - \mathbb{P}(X_n \geq 1 + \varepsilon).
\] (4.9)

Again we consider both terms in the last line separately. For the first term, as in (4.6) we obtain

\[
\mathbb{P}(Z_n x_n \geq x(1 + \varepsilon)) = e^{-\sqrt{n(\sqrt{x(1 + \varepsilon)} - x + o(1))}}, \quad \text{as } n \to \infty.
\] (4.10)

For the second term, we use the same argument as for (4.7) and get

\[
\mathbb{P}(X_n \geq 1 + \varepsilon) \leq e^{-n I_{\text{exp}}(\sqrt{1 + \varepsilon})(1 + o(1))}, \quad \text{as } n \to \infty.
\] (4.11)

Combining (4.10) and (4.11) with (4.9) now gives (4.4) which proves (2.11).

Since the minimum of \( W_n \) is 1 (when \( F_{k,n} = 1/n \) for all \( k \)) the assertion \( \mathbb{P}(W_n < 1) = 0 \) is clear. It remains to prove (2.12), show that the rate function is of the form (2.13) and justify the positivity of \( \tilde{I}(x) \) for \( x \in (1, 2) \).

For \( x \in (1, 2) \) using (2.10) we obtain

\[
\mathbb{P}(W_n \leq x) = \mathbb{P} \left( \frac{\sum_{j=1}^{n} R_j}{\sqrt{n} \sqrt{\sum_{k=1}^{n} R_k}} \geq \frac{1}{\sqrt{x}} \right).
\] (4.12)

Furthermore, for \( x \in (1, 2) \) we have \( 1/\sqrt{x} > 1/\sqrt{2} = \mathbb{E}[R_1]/\sqrt{\mathbb{E}[R_1]} \). Thus, we can use Theorem 1.1 from Shao (1997) and obtain

\[
\mathbb{P}(W_n \leq x)^{1/n} = \sup_{c \geq 0} \inf_{t \geq 0} \mathbb{E} \left[ \exp \left( t(cR_1 - \frac{1}{2\sqrt{x}} (R_1^2 + c^2)) \right) \right].
\] (4.13)

Now we have

\[
E \left[ \exp \left( t(cR_1 - \frac{1}{2\sqrt{x}} (R_1^2 + c^2)) \right) \right] = \int_{0}^{\infty} \exp \left( -y + t(cy - \frac{1}{2\sqrt{x}} (y^2 + c^2)) \right) dy
\]

and elementary integration yields

\[
= \frac{\sqrt{2\pi x^{1/4}}}{\sqrt{t}} \exp \left( \frac{(tc - 1/2) x - t^2 c^2}{2\sqrt{x}} \right) \Phi \left( \frac{(tc - 1)x^{1/4}}{\sqrt{t}} \right),
\]

where \( \Phi \) denotes the distribution function of the one dimensional standard Gaussian distribution. Taking \log of the last term we obtain (2.13).
Now we fix $x \in (1, 2)$ and show that $\tilde{I}(x)$ is positive. In the sequel we write
\[ h(r, c) := cr - \frac{1}{2\sqrt{x}}(r^2 + c^2). \]
We have
\[
\inf_{t \geq 0} E \left[ \exp \left( th(R_1, c) \right) \right] \\
\geq E \left[ \inf_{t \geq 0} \exp \left( th(R_1, c) \right) \right] \\
= E \left[ \mathbb{1}_{\{h(R_1, c) < 0\}} \inf_{t \geq 0} \exp \left( th(R_1, c) \right) \right] + E \left[ \mathbb{1}_{\{h(R_1, c) \geq 0\}} \inf_{t \geq 0} \exp \left( th(R_1, c) \right) \right] \\
= P(h(R_1, c) \geq 0).
\]
The function $r \mapsto h(r, c)$ is non-negative on the interval $[r_1, r_2]$ where $r_{1/2} = r_{1/2}(c) := c(\sqrt{x} \pm \sqrt{x - 1})$ are the zeros of the function. It follows
\[ E \left[ \inf_{t \geq 0} \exp \left( th(R_1, c) \right) \right] = P(r_1 \leq R_1 \leq r_2) = e^{-c(\sqrt{x} - \sqrt{x - 1})} - e^{-c(\sqrt{x} + \sqrt{x - 1})}. \]
Finally, by elementary calculation we obtain
\[ \sup_{c \geq 0} (e^{-c(\sqrt{x} - \sqrt{x - 1})} - e^{-c(\sqrt{x} + \sqrt{x - 1})}) = \frac{2\sqrt{x - 1}}{\sqrt{x} + \sqrt{x - 1}} \left( \frac{\sqrt{x} - \sqrt{x - 1}}{\sqrt{x} + \sqrt{x - 1}} \right)^{\frac{\sqrt{x} - \sqrt{x - 1}}{\sqrt{x} + \sqrt{x - 1}}}. \]
This expression (and therefore also $\tilde{I}(x)$) is positive for $x \in (1, 2)$. Thus, the proof of Theorem 3 is concluded.

### 4.3 Proof of Theorem 3

We prove the inequality (2.15) using Lemma 2.7. Let $1 < x < 2$ and set $y = \frac{1}{\sqrt{x-1}} - 1$. For $\varepsilon > 0$ we have
\[
\frac{1}{n} \log P(W_n \leq x + \varepsilon) \\
= \frac{1}{n} \log P \left( \frac{2 \sum_{i=1}^{n} \sum_{k=1}^{l} R_k R_i \sum_{k=1}^{n} R_k^2}{(\sum_{k=1}^{n} R_k)^2} \leq x + \varepsilon \right) \\
\geq \frac{1}{n} \log P \left( \frac{2 \sum_{i=1}^{n} \sum_{k=1}^{l} R_k R_i - \sum_{k=1}^{n} R_k^2}{(\sum_{k=1}^{n} R_k)^2} \leq x + \varepsilon, R_n > ny \right) \\
= \frac{1}{n} \log \left\{ P(R_n > ny) \left[ \frac{2 \sum_{i=1}^{n} \sum_{k=1}^{l} R_k R_i - \sum_{k=1}^{n} R_k^2}{(\sum_{k=1}^{n} R_k)^2} \leq x + \varepsilon \right| R_n > ny \right\}.
\]
Now $\frac{1}{n} \log P(R_n > ny) = -y$, and conditioning in the second factor in the curly braces can be removed by using the fact that conditioned on $R_n > ny$ the exponential random variable $R_n$ has the same distribution as $ny + R_n$. After some elementary calculations we see that the last line of the above display equals
\[ -y + \frac{1}{n} \log \left( \frac{1/2 \sum_{i=1}^{n} \sum_{k=1}^{l} R_k R_i - \sum_{k=1}^{n} R_k^2}{(1/2 \sum_{k=1}^{n} R_k)^2 + 2y \frac{1}{n} \sum_{k=1}^{n} R_k^2 + y^2} \leq x + \varepsilon \right). \]
From the strong law of large numbers and (2.4) with Lemma 2.7 we know that
\[ \frac{1}{n} \sum_{k=1}^{n} R_k \overset{n \rightarrow \infty}{\rightarrow} 1, \text{ and } \frac{1}{n} \left( 2 \sum_{i=1}^{n} \sum_{k=1}^{l} R_k R_i \sum_{k=1}^{n} R_k^2 \right) \overset{n \rightarrow \infty}{\rightarrow} 2 \text{ almost surely.} \]
It follows that almost surely

\[ \frac{1}{n} \left( 2 \sum_{l=1}^{n} \sum_{k=1}^{l} \frac{R_k R_l}{l} - \sum_{k=1}^{n} \frac{R_k^2}{k} \right) + 2y \frac{1}{n} \sum_{k=1}^{n} R_k + y^2 \rightarrow_{n \to \infty} \frac{2 + 2y + y^2}{1 + 2y + y^2} = x. \]

Thus,

\[ \liminf_{n \to \infty} \frac{1}{n} \log P \left( W_n \leq 2 - x + \varepsilon \right) \geq -y = 1 - \frac{1}{\sqrt{x - 1}}. \]

The rest follows by letting \( \varepsilon \downarrow 0. \)

Acknowledgments

We thank Shui Feng for pointing out connections to Dawson and Feng (2006) and Nina Gantert and Alain Rouault for pointing out the reference Shao (1997) and fruitful email discussion that led to the exact rate function in (2.12). This research was supported by the DFG through grants Pf-672/6-1 to AD and PP.

References

Aldous, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli 5(1), 3–48.

Angel, O., N. Berestycki, and V. Limic (2012). Global divergence of spatial coalescents. Probab. Theory Related Fields 152(3-4), 625–679.

Dawson, D. and S. Feng (2006). Asymptotic behavior of the Poisson-Dirichlet distribution for large mutation rate. Ann. Appl. Probab. 16(2), 562–582.

Dembo, A. and O. Zeitouni (2010). Large deviations techniques and applications, Volume 38 of Stochastic Modelling and Applied Probability. Berlin: Springer-Verlag. Corrected reprint of the second (1998) edition.

Depperschmidt, A., A. Greven, and P. Pfaffelhuber (2013). Path-properties of the tree-valued Fleming–Viot process. Electron. J. Probab. 18(84), 1–47.

Donnelly, P. (1991). Weak convergence to a markov chain with an entrance boundary: ancestral processes in population genetics. Ann. Probab. 19, 1102–1117.

Evans, S. (2000). Kingman’s coalescent as a random metric space. In Stochastic Models: Proceedings of the International Conference on Stochastic Models in Honour of Professor Donald A. Dawson, Ottawa, Canada, June 10-13, 1998 (L.G Gorostiza and B.G. Ivanoff eds.), Canad. Math. Soc.

Gantert, N., K. Ramanan, and F. Rembart (2014). Large deviations for weighted sums of stretched exponential random variables. preprint. arXiv:1401.4577.

Kingman, J. (1982). The coalescent. Stoch. Proc. Appl. 13(3), 235–248.

Pitman, J. (1999). Coalescents with multiple collisions. Ann. Probab. 27(4), 1870–1902.

Pyke, R. (1965). Spacings. (With discussion.). J. Roy. Statist. Soc. Ser. B 27, 395–449.

Shao, Q.-M. (1997). Self-normalized large deviations. Ann. Probab. 25(1), 285–328.

Tavaré, S. (1984). Line-of-descent and genealogical processes and their applications in population genetics models. Theor. Pop. Biol. 26, 119–164.