Asymptotic expansions and precise deviations in the Kingman coalescent

Fuqing Gao[*]   Yujing Wang[†]   Youzhou Zhou[‡]

Abstract

In this paper, we study the small-time asymptotic behavior of the Kingman coalescent. We obtain the Berry-Esseen bound and the Edgeworth expansion in the central limit theorem. Moreover, by the method of mod-ϕ convergence, we also obtain the precise large deviations and the precise moderate deviations. Last, we also obtain a non-asymptotic deviation inequality for the Kingman coalescent.

Keywords: Berry-Esseen bound; Edgeworth expansion; precise deviations; mod-ϕ convergence; Kingman coalescent.
MSC2020 subject classifications: 60F10; 60C05.
Submitted to ECP on May 7, 2020, final version accepted on January 25, 2021.

1 Introduction

The Kingman coalescent was introduced in 1982 by Kingman in [5]. It was used to describe the genealogy of a sample from a population. Ever since its appearance, the Kingman coalescent has found many applications in biology. It also serves as the dual process of the Fleming-Viot process with parent-independent mutation. As a dual process, the Kingman coalescent usually describes the backward movement in time in population genetics. More generally, coalescent processes, the Lambda coalescence in particular, are dual processes of some generalized Fleming-Viot processes.

Apart from describing sample genealogy, the Kingman coalescent also has other interpretations, such as the number of surviving ancient families. Due to genetic drift and mutation, the ancient families are lost as time moves forward. Let D_t be the number of surviving ancient families up to time t. Then D_t is a pure-death Markov chain with transition rates $q_{n,n-1} = \binom{n}{2}, q_{nk} = 0, k < n-1$, starting at ∞. Once an ancient family disappears, the Kingman coalescent experiences a coalescing event. These coalescing events arrive independently, and waiting times between coalescing events follow exponential distributions. Let T_n be the arriving time of the coalescing event, where n is the number of surviving ancient families right after T_n. Then $T_n = \sum_{k=n+1}^{\infty} \xi_k / \binom{k}{2}$ where $\{\xi_k, k \geq 1\}$ are $i.i.d.$ exponential random variables with mean 1. Like Poisson process, D_t and T_n have a dual relationship $\mathbb{P}(T_n \geq t) = \mathbb{P}(D_t \geq n)$. When

*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China. E-mail: `fqgao@whu.edu.cn`
†School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China.
 E-mail: `yujingwang@whu.edu.cn`
‡Department of Pure Mathematics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China.
 E-mail: `youzhou.zhou@xjtlu.edu.cn`

we study the small-time behavior of D_t, this dual relation facilitates our computation a lot.

As a standard coalescent model in population genetics, the Kingman coalescent has been thoroughly studied. For instance, the law of large numbers and the central limit theorem were summarized in [2](see also [7] for some discussion on the central limit theorem). Some functional central limit theorem was also obtained in [9]. Large deviation principle has been discussed in [3]. However, large deviation principle and moderate deviation principle only aim at seeking a leading term of the logarithm of tail probability. Therefore, polynomial correction terms will be missing if we only consider large deviations and moderate deviations. Especially, these correction terms play indispensable roles in statistical inference. The precise deviations, however, not only provide the leading terms but also polynomial correction terms of the tail probability. As far as we know, the precise large deviations and precise moderate deviations for the Kingman coalescent are still unknown. Similarly, we have not seen any asymptotic result like the Berry-Esseen bounds, the Edgeworth expansions and deviation inequalities for the Kingman coalescent in the existing literatures either. Similar problems for the Lambda-coalescents also deserve study.

In this paper, we are going to find the Berry-Esseen bound, establish the Edgeworth expansions in the central limit theorem(CLT for short) and the local central limit theorem. Some precise large deviations, precise moderate deviations, and deviation inequalities are also obtained for the Kingman coalescent. Due to the dual relation $\mathbb{P}(T_n \geq t) = \mathbb{P}(D_t \geq n)$, we will only focus on T_n. The asymptotic behavior of D_t is only a simple corollary of the asymptotic behavior of T_n. The main methodology is an asymptotic analysis of the generating function in the complex domain and the mod-ϕ convergence theory [4].

Our paper is organized as follows. In Section 2, will state the main results. In section 3, we show the Berry-Esseen bound and the Edgeworth expansions in the CLT. The Edgeworth expansions in local CLT will be presented in Section 4. Precise large deviations and precise moderate deviations are shown in Section 5. In Section 6, we will prove the deviation inequality.

2 Main results

Note that $T_n = \sum_{k=n+1}^{\infty} \xi_k / \binom{k}{2}$, a simple computation yields

$$\mathbb{E}(nT_n) = \sum_{k=n+1}^{\infty} \left\{ n / \binom{k}{2} \right\} = 2, \quad \mathrm{Var}(nT_n) = \sum_{k=n+1}^{\infty} \left\{ n^2 / \binom{k}{2}^2 \right\} \sim \frac{4}{3n}.$$

We will rescale nT_n as $Z_n = \frac{\sqrt{3n}}{2}(nT_n - 2)$. In the following, we are going to provide the main results on the asymptotic behavior of the rescaled quantity Z_n. We will mainly discuss the Edgeworth expansion in the CLT. The expansion involves Hermite polynomials $H_l(x)$ which are defined as

$$H_l(x) = (-1)^l \frac{1}{\sqrt{2\pi}} e^{x^2/2} \frac{d^l}{dx^l} \left[e^{-\frac{x^2}{2}} \right].$$

A notation $\mathcal{S}_k$ will be repeatedly used in the Edgeworth expansion. It is a set of k-tuple $(m_1, \cdots, m_k)$, where $m_1, \ldots, m_k$ are non-negative integers and

$$1 \cdot m_1 + 2 \cdot m_2 + \cdots + k \cdot m_k = k. \tag{2.1}$$

Now we are ready to state the main results.

2.1 The Berry-Esseen bound and the Edgeworth expansion in the CLT

Theorem 2.1. *(Berry-Esseen bound). There exists a constant $C > 0$ such that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \leq x \right) - \Phi(x) \right| \leq \frac{C}{\sqrt{n}}, \quad n > 0. \tag{2.2}$$

Theorem 2.2. *(Edgeworth expansions in the CLT). For any $m \geq 2$, there exists a positive constant C, independent of n, such that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \leq x \right) - \Phi(x) - \sum_{k=1}^{m-2} Q_{k,n}(x) \right| \leq \frac{C}{n^{(m-1)/2}}, \quad n > 0, \tag{2.3}$$

where

$$Q_{k,n}(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(\sqrt{3/n} \right)^k \sum_{\mathcal{S}_k} H_{k+2(m_1+\cdots+m_k)-1}(x) \prod_{l=1}^{k} \frac{(a_l(n)/l!)^{m_l}}{m_l!}, \tag{2.4}$$

and

$$a_l(n) = \begin{cases} \frac{3A_{n,(l+2)} l!}{(l+2)}, & l \neq 2 \\ \frac{3A_{n,4}}{2} - \frac{n\epsilon_{n,2}}{2}, & l = 2 \end{cases}, \quad A_{n,l} = n^{2l-1} \sum_{k=n}^{\infty} \frac{1}{[k(k+1)]^l}, \epsilon_{n,2} = A_{n,2} - \frac{1}{3}.$$

Remark 2.3. Note that we have $A_{n,l} \to \frac{1}{2l-1}$ and $n\epsilon_{n,2} \to \frac{1}{2}$ as $n \to \infty$, then

$$a_l(n) \to \begin{cases} \frac{3l!}{(2l+3)(l+2)}, & l \neq 2 \\ -\frac{1}{28}, & l = 2 \end{cases}.$$

2.2 Local limit theorem

Let $p_n(x)$ be the density of Z_n. By local limit theorem, we mean the asymptotic behavior of $p_n(x)$ in uniform topology as $n \to \infty$. First, we establish the Edgeworth expansion in uniform topology, then the local limit theorem will be a simple corollary.

When we consider the Edgeworth expansion, the cumulative distribution $\mathbb{P}(Z_n \leq x)$ is approximated by $G_{m,n}(x) = \Phi(x) + \sum_{k=1}^{m-2} Q_{k,n}(x)$. It is natural to use $G'_{m,n}(x) = \phi(x) + \sum_{k=1}^{m-2} q_{k,n}(x)$, where $q_{k,n}(x) = Q'_{k,n}(x)$, to approximate the density $p_n(x)$ of Z_n.

Theorem 2.4. *(Local limit theorem). There exists a constant $C > 0$ such that*

$$\sup_{x \in \mathbb{R}} \left| p_n(x) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \leq \frac{C}{\sqrt{n}}, \quad n > 0. \tag{2.5}$$

Theorem 2.5. *(Edgeworth expansions in the local limit theorem). For any $m \geq 2$, there exists a positive constant C, independent of n, such that*

$$\sup_{x \in \mathbb{R}} \left| p_n(x) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - \sum_{k=1}^{m-2} q_{k,n}(x) \right| \leq \frac{C}{n^{(m-1)/2}}, \quad n > 0. \tag{2.6}$$

Remark 2.6. When $n = [\frac{2+\sqrt{tx}}{t}]$, by dual relation $\mathbb{P}(D_t \geq n) = \mathbb{P}(T_n \geq t)$, one can also obtain the CLT the Edgeworth expansion of D_t in the CLT and its Edgeworth expansions in the local central limit theorem.

2.3 Precise deviations

The strong limit of nT_n is 2 as $n \to \infty$(refer to [2]). Thus, for $\epsilon > 0$, both $\{nT_n > 2+\epsilon\}$ and $\{nT_n < 2-\epsilon\}$ are rare events. Precise deviation principle is about the estimations of $\mathbb{P}(nT_n > 2 + \epsilon)$ and $\mathbb{P}(nT_n < 2 - \epsilon)$, where ϵ indicates the level of deviations from the strong limit point 2. Large deviations correspond to the case $\epsilon = O(1)$, and moderate deviations corresponds to another case $\epsilon = o(1)$.

Precise large deviations and moderate deviations can be obtained by mod-ϕ convergence [4], which depends on carefully analysis of the moment generating function $\mathbb{E}e^{zn^2 T_n}$ of nT_n in complex domain $\mathcal{S}_{(-\infty,1/2)} := \{z \in \mathbb{C}; \operatorname{Re}(z) < 1/2\}$. In section 4, we will show that $\psi_n(z) = e^{-n\eta(z)}\mathbb{E}e^{zn^2 T_n}$ has a uniform limit on $\mathcal{S}_{(-\infty,1/2)}$, where

$$\eta(z) = -\int_1^\infty \log(1 - \frac{2z}{x^2})dx. \tag{2.7}$$

Here $\eta(z)$ can be regarded as the analytic continuation of the limiting log-Laplace transformation $\eta(\theta)$, which has been obtained in [3]. The rate function $I(x)$ is the Legendre transformation of $\eta(\theta)$, $\theta \in \mathbb{R}$, and

$$I(x) = \theta_x x - \eta(\theta_x) = \begin{cases} \theta_x x + \int_1^\infty \log(1 - \frac{2\theta_x}{u^2})du, & \text{if } x > 0 \\ \\ +\infty, & \text{if } x \leq 0 \end{cases} \tag{2.8}$$

where $\theta_x < \frac{1}{2}$ is the unique solution of the equation $x = \eta'(\theta)$. Moreover, $\eta'(0) = 2$ and $\eta''(0) = \frac{4}{3}$. One can also show that $I(2) = I'(2) = 0$, and $I''(2) = \frac{1}{\eta''(0)} = \frac{3}{4}$.

Theorem 2.7. *(Precise large deviations). For $x > 2$, the upper tail probability satisfies*

$$\mathbb{P}(nT_n \geq x) = e^{-nI(x)}\frac{1}{\theta_x \sqrt{2\pi n \eta''(\theta_x)}}\Big(1 + o(1)\Big) \text{ as } n \to \infty. \tag{2.9}$$

For $0 < x < 2$, the lower tail probability satisfies

$$\mathbb{P}(nT_n \leq x) = e^{-nI(x)}\frac{1}{\theta_x \sqrt{2\pi n \eta''(\theta_x)}}\Big(1 + o(1)\Big) \text{ as } n \to \infty. \tag{2.10}$$

For moderate deviations, we specify the deviation $s = x - 2$ as $s = \sqrt{\frac{\eta''(0)}{n}}\beta_n y = o(1)$, where $0 < \beta_n \ll \sqrt{n}$. Since θ_x is the solution of $\eta(\theta) = x$, then $\theta_x \sim s = \frac{2}{\sqrt{3n}}\beta_n y$ as $n \to \infty$.

Theorem 2.8. *(Precise moderate deviations).*

(i) If $x = 2 + \sqrt{\frac{\eta''(0)}{n}}\beta_n y$, $y > 0$, and $\beta_n \sim n^{1/2-1/m}, m \geq 3$, then

$$\mathbb{P}\Big(nT_n \geq x\Big)$$
$$= \frac{3}{4\sqrt{2\pi}y\beta_n}\exp\left\{-\beta_n^2\frac{y^2}{2} - \beta_n^2\sum_{i=3}^{m+1}\frac{(2/\sqrt{3})^i I^{(i)}(2)}{i!}y^i\left(\frac{\beta_n}{\sqrt{n}}\right)^{i-2}\right\}(1 + o(1)) \tag{2.11}$$

(ii) If $x = 2 + \sqrt{\frac{\eta''(0)}{n}}\beta_n y$, $y < 0$, and $\beta_n \sim n^{1/2-1/m}, m \geq 3$, then

$$\mathbb{P}\Big(nT_n \geq x\Big)$$
$$= \frac{3}{-4\sqrt{2\pi}y\beta_n}\exp\left\{-\beta_n^2\frac{y^2}{2} - \beta_n^2\sum_{i=3}^{m+1}\frac{(2/\sqrt{3})^i I^{(i)}(2)}{i!}(-y)^i\left(\frac{\beta_n}{\sqrt{n}}\right)^{i-2}\right\}(1 + o(1)) \tag{2.12}$$

(iii) If $x = 2 + \sqrt{\frac{\eta''(0)}{n}}\beta_n y$ and $\beta_n = o(n^{1/6})$, *then*

$$\mathbb{P}\Big(nT_n \geq x\Big) = \bar{\Phi}(\beta_n y)\Big(1 + o(1)\Big), \tag{2.13}$$

where $\bar{\Phi}(y) := \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

2.4 Deviation inequality

Theorem 2.9. *(Deviation inequality). For any* $x \geq 0$, $n \geq 1$, *and* $r \in (0,1)$, *both* $\mathbb{P}\Big(Z_n \geq x\Big)$ *and* $\mathbb{P}\Big(Z_n \leq -x\Big)$ *are bounded as follows:*

$$\mathbb{P}\Big(Z_n \geq x\Big), \mathbb{P}\Big(Z_n \leq -x\Big) \leq \begin{cases} e^{-\frac{(1-r)}{2n}x^2}, & 0 \leq x \leq \frac{\sqrt{nr}}{\sqrt{3}(1-r)}; \\[2mm] e^{-\frac{\sqrt{n}xr}{2\sqrt{3}}}, & x > \frac{\sqrt{nr}}{\sqrt{3}(1-r)}. \end{cases} \tag{2.14}$$

Remark 2.10.

The rescaled quantity Z_n has either positive or negative deviation. When the absolute deviation is small, the decay is of Gaussian type; when the absolute deviation is large, the decay behaves like a Poisson tail.

3 Berry-Esseen bounds and asymptotic expansions of CLT

In this section, we are going to prove the Berry-Esseen bound and the Edgeworth expansions in the CLT. Note that Berry-Esseen bound is only a special case of the Edgeworth expansions in the CLT. Thus, it suffices to show the Edgeworth expansions in the CLT. But first we need the following basic lemma (refer to [8]).

Lemma 3.1. *Let* F *be a cumulative distribution function. Assume that* G *is a differentiable function, satisfying* $G(-\infty) = 0$ *and* $G(+\infty) = 1$, *and* G' *is bounded. Then for any* $r > 0$,

$$\sup_{x \in \mathbb{R}} |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-r}^r \left| \frac{\phi_F(s) - \phi_G(s)}{s} \right| ds + \frac{24}{r\pi} \sup_{x \in \mathbb{R}} |G'(x)|,$$

where $\phi_F(s) := \int_{\mathbb{R}} e^{isx} dF(x)$ *is the Fourier transform of function* F.

By Lemma 3.1, we can establish the Edgeworth expansions of cumulative distribution $\mathbb{P}(Z_n \leq x)$ of Z_n through expansion of its characteristic function $\phi_n(\theta) = \mathbb{E}e^{i\theta Z_n}$. Note that $\int_{\mathbb{R}} e^{i\theta x} d\Phi(x) = e^{-\frac{\theta^2}{2}}$ and $\Phi^{(l)}(x) = (-1)^{l-1} H_{l-1}(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and $H_{l-1}(x)$ are Hermite polynomials. By integration by parts, we have

$$(-i\theta)^l e^{-\frac{\theta^2}{2}} = \int_{\mathbb{R}} e^{i\theta x} d\Phi^{(l)}(x), \tag{3.1}$$

which builds up a correspondence between expansions of $\mathbb{P}(Z_n \leq x)$ and its characteristic function $\phi_n(\theta)$.

Lemma 3.2. *There exists positive constants* δ, N, M, *such that*

$$\phi_n(\theta) = \Psi_n(\theta) + R(\theta) \text{ for } n \geq N \text{ and } \frac{\sqrt{3}|\theta|}{\sqrt{n}} < \delta,$$

where $\Psi_n(\theta) = e^{-\frac{\theta^2}{2}} + \sum_{k=1}^{m-2} \Psi_{k,n}(\theta)$,

$$\Psi_{k,n}(\theta) = (\sqrt{3/n})^k \sum_{\mathcal{S}_k} \prod_{i=1}^k \frac{[a_i(n)/i!]^{m_i}}{m_i!} (i\theta)^{\sum_{i=1}^k (i+2)m_i} e^{-\frac{\theta^2}{2}},$$

and $|R(\theta)| \leq \frac{M}{(\sqrt{n})^{m-1}}(\sum_{s=0}^{3(m-1)}|\theta|^s)e^{-\frac{\theta^2}{4}}$.

Due to the correspondence relation (3.1), we know

$$\Psi_n(\theta) = \int_{\mathbb{R}} e^{i\theta x} dG_{m,n}(x) = e^{-\frac{\theta^2}{2}} + \sum_{k=1}^{m-2} \Psi_{k,n}(\theta),$$

where $G_{m,n}(x) = \Phi(x) + \sum_{k=1}^{m-2} Q_{k,n}(x)$. In the following, we will show that $\mathbb{P}(Z_n \leq x)$ has an expansion $G_{m,n}(x) = \Phi(x) + \sum_{k=1}^{m-2} Q_{k,n}(x)$.

Proof of Theorem 2.2. To apply Lemma 3.1, we take $r = Cn^{(m-1)/2}$, where $0 < C < \min\{\sqrt{14/9}, \delta\}$. Then

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \leq x \right) - G_{m,n}(x) \right| \leq \frac{2}{\pi} \int_0^{Cn^{(m-1)/2}} \left| \frac{\phi_n(\theta) - \Psi_n(\theta)}{\theta} \right| d\theta + \frac{24}{Cn^{(m-1)/2}\pi} \sup_{x \in \mathbb{R}} |G'_{m,n}(x)|$$

$$= K_1 + K_2 + \frac{24}{Cn^{(m-1)/2}\pi} \sup_{x \in \mathbb{R}} |G'_{m,n}(x)|,$$

where

$$K_1 = \frac{2}{\pi} \int_0^{Cn^{1/2}} \left| \frac{\phi_n(\theta) - \Psi_n(\theta)}{\theta} \right| d\theta, \ K_2 = \frac{2}{\pi} \int_{Cn^{1/2}}^{Cn^{(m-1)/2}} \left| \frac{\phi_n(\theta) - \Psi_n(\theta)}{\theta} \right| d\theta.$$

There exists a positive constant M, independent of n, such that $\frac{24}{Cn^{(m-1)/2}\pi} \sup_{x \in \mathbb{R}} |G'_{m,n}(x)| < M$, then

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \leq x \right) - G_{m,n}(x) \right| \leq K_1 + K_2 + \frac{M}{n^{(m-1)/2}}.$$

It remains to show that both K_1 and K_2 are bounded by $\frac{M}{n^{(m-1)/2}}$. Note that

$$|K_2| \leq \frac{2n^{(m-1)/2}}{\pi\sqrt{n}} \sup_{\theta > C\sqrt{n}} |\phi_n(\theta)| + \frac{2}{C\pi\sqrt{n}} \int_{C\sqrt{n}}^{\infty} |\Psi_n(\theta)| d\theta.$$

It is easy to see that $|\Psi_n(\theta)| \leq M \sum_{l=0}^{2(m-2)} |\theta|^l e^{-\frac{\theta^2}{2}}$. Then by Gaussian probability estimates, we know there exist two positive constants C_1, C_2 such that

$$\frac{2}{C\pi\sqrt{n}} \int_{C\sqrt{n}}^{\infty} |\Psi_n(\theta)| d\theta \leq C_1 e^{-C_2 n}.$$

Moreover,

$$|\phi_n(\theta)| \leq \exp\left\{ -\sum_{k=n}^{\infty} \log \left| 1 - \frac{\sqrt{3}i\theta n^{3/2}}{k(k+1)} \right| \right\} = \exp\left\{ -\frac{1}{2} \sum_{k=n}^{\infty} \log \left| 1 + \frac{3\theta^2 n^3}{k^2(k+1)^2} \right| \right\}.$$

So

$$\sup_{\theta > C\sqrt{n}} |\phi_n(\theta)| \leq \exp\left\{ -\frac{1}{2} \sum_{k=n}^{\infty} \log \left| 1 + \frac{3C^2 n^4}{k^2(k+1)^2} \right| \right\}.$$

By inequality $\log(1+x) > x - \frac{x^2}{2}, x > 0$, we have $\sup_{\theta > C\sqrt{n}} |\phi_n(\theta)| \leq C_1 e^{-C_2 n}$.

Thus, there exist two positive constants M_1, M_2 such that $|K_2| \leq M_1 e^{-nM_2}$. By Lemma 3.2, we know

$$|K_1| \leq \frac{2M}{\pi} \frac{1}{n^{(m-1)/2}} \int_0^{C\sqrt{n}} \sum_{s=0}^{3(m-1)} \theta^s e^{-\frac{\theta^2}{4}} d\theta \leq \frac{2DM}{\pi} \frac{1}{n^{(m-1)/2}}$$

where $D = \int_0^{\infty} \sum_{s=0}^{3(m-1)} \theta^s e^{-\frac{\theta^2}{4}} d\theta < \infty$. So the expansion of $\mathbb{P}(Z_n \leq x)$ is $G_{m,n}(x)$, and its uniform approximation error $\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \leq x \right) - G_{m,n}(x) \right|$ is bounded by $\frac{M}{n^{(m-1)/2}}$. □

Proof of Lemma 3.2. Note that $\phi_n(\theta) = \mathbb{E}e^{\frac{\sqrt{3}}{2}i\theta n^{\frac{3}{2}}T_n} \cdot e^{-\sqrt{3}i\theta n^{\frac{1}{2}}}$.

By the definition of T_n, we have $\mathbb{E}e^{\frac{\sqrt{3}}{2}i\theta n^{\frac{3}{2}}T_n} = e^{-\sum_{k=n}^{\infty} \log(1-\frac{\sqrt{3}i\theta n^{\frac{3}{2}}}{k(k+1)})}$. Because $\log(1-x) = -\sum_{l=1}^{\infty}\frac{x^l}{l}$, switching summation order yields $\phi_n(\theta) = \exp\left\{\sum_{l=2}^{\infty}\frac{(\sqrt{3}i\theta)^l}{l}\frac{A_{n,l}}{(\sqrt{n})^{l-2}}\right\}$, where $A_{n,l} = n^{2l-1}\sum_{k=n}^{\infty}\frac{1}{[k(k+1)]^l}$, and

$$A_{n,l} = \sum_{k=n}^{\infty}\frac{1}{(\frac{k}{n})^{2l}(1+\frac{1}{k})^l}\frac{1}{n} \leq 1 + \int_1^{\infty}\frac{dx}{x^{2l}} = \frac{2l}{2l-1}.$$

Introducing $\epsilon_{n,2} = A_{n,2} - \frac{1}{3}$ and $w = \frac{i\sqrt{3}\theta}{\sqrt{n}}$, we can rewrite $\phi_n(\theta)$ as $\phi_n(\theta) = e^{-\frac{\theta^2}{2}}e^{-g(w)}e^{R_1(\theta)}$, where $R_1(\theta) = \sum_{l=m+1}^{\infty}\frac{(\sqrt{3}i\theta)^l}{l}\frac{A_{n,l}}{(\sqrt{n})^{l-2}}$, $g(w) = \sum_{l=1}^{m-2}\frac{w^l}{l!}\theta^2 a_l(n)$, and

$$a_l(n) = \begin{cases} \frac{3l!A_{n,l+2}}{l+2} & l \neq 2 \\ \frac{3A_{n,4}}{2} - n\epsilon_{n,2} & l = 2. \end{cases}$$

One can also show that there exists constants $N, 0 < \delta < \frac{1}{8}$, such that $a_l(n) \leq 2l! + 1, \forall n \geq N$, and $\forall |w| < \delta$,

$$|R_1(\theta)| \leq \max\left\{\frac{\sqrt{3}^{m-1}}{n^{(m-1)/2}}|\theta|^{m+1}, |\theta|^2/8\right\} \tag{3.2}$$

and

$$|g(w)| + |R_1(\theta)| \leq \max\left\{\frac{M}{\sqrt{n}}|\theta|^3, |\theta|^2/8\right\}. \tag{3.3}$$

Next we will expand $f(w) = e^{-g(w)}$ to get the expansion of $\phi_n(\theta)$. We claim that we can expand $f(w)$ as

$$e^{-g(w)} = \sum_{k=0}^{m-2}\frac{b_k(n)}{k!}w^k + R_2(\theta), \tag{3.4}$$

where $|R_2(\theta)| \leq \frac{M}{n^{(m+1)/2}}\sum_{s=1}^{m-1}|\theta|^s e^{\frac{\theta^2}{8}}$. Indeed, the error bound of $R_2(\theta)$ can be obtained through composition of power series. By the uniqueness of Taylor expansion, we know $b_k(n) = f^{(k)}(0)$. To specify $f^{(k)}(0)$, we turn to Faá di Bruno formula(see [1] and [6]). To this end, we define $h(z) = e^{-z}$, then $f(w) = h(g(w))$. Applying Faá di Bruno formula to $f(w)$, we have

$$f^{(k)}(0) = \sum_{\mathcal{S}_k}\frac{k!}{1!^{m_1}\cdots k!^{m_k}m_1!\cdots m_k!}h^{(\sum_{i=1}^{k}m_i)}(g(0))\prod_{i=1}^{k}[g^{(i)}(0)]^{m_i}.$$

Note that $g(0) = 0$, so $h^{(\sum_{i=1}^{k}m_i)}(g(0)) = (-1)^{\sum_{i=1}^{k}m_i}$ and $g^{(i)}(0) = a_i(n)$. Thus,

$$b_0(n) = 1, \quad b_k(n) = k!\sum_{\mathcal{S}_k}(-\theta^2)^{\sum_{i=1}^{k}m_i}\prod_{i=1}^{k}\frac{[a_i(n)/i!]^{m_i}}{m_i!}, k \geq 1.$$

Last, we can rewrite $\phi_n(\theta)$ as

$$\phi_n(\theta) = e^{-\frac{\theta^2}{2}}e^{-g(w)}e^{R_1(\theta)} = e^{-\frac{\theta^2}{2}}\left[\sum_{k=0}^{m-2}\frac{b_k(n)}{k!}w^k + R_2(\theta)\right]e^{R_1(\theta)}$$

$$= e^{-\frac{\theta^2}{2}}\sum_{k=0}^{m-2}\frac{b_k(n)}{k!}w^k + R_3(\theta),$$

where

$$R_3(\theta) = e^{-\frac{\theta^2}{2}} \left[\sum_{k=0}^{m-2} \frac{b_k(n)}{k!} w^k \right] (e^{R_1(\theta)} - 1) + e^{-\frac{\theta^2}{2}} e^{R_1(\theta)} R_2(\theta) + e^{-\frac{\theta^2}{2}} R_2(\theta)(e^{R_1(\theta)} - 1)$$

and $\left| \sum_{k=0}^{m-2} \frac{b_k(n)}{k!} w^k \right| \leq |e^{-g(w)}| + |R_2(\theta)|$. A careful computation shows that the error term $R_3(\theta)$ satisfies $|R_3(\theta)| \leq \frac{M}{n^{(m-1)/2}} (\sum_{s=0}^{2m} |\theta|^s) e^{-\frac{\theta^2}{4}}$.

Now we see that Lemma 3.2 is true because $w = \frac{i\sqrt{3}\theta}{\sqrt{n}}$, and

$$\Psi_{k,n}(\theta) = \frac{b_k(n)}{k!} w^k = (\sqrt{3/n})^k \sum_{\mathcal{S}_k} (i\theta)^{\sum_{i=1}^{k}(i+2)m_i} \prod_{i=1}^{k} \frac{[a_i(n)/i!]^{m_i}}{m_i!}. \qquad \square$$

4 Local limit theorems

We only need to prove Theorem 2.5 because Theorem 2.4 is a simple corollary of Theorem 2.5. By the correspondence (3.1), one can deduce that

$$\Psi_{k,n}(\theta) = \int_{\mathbb{R}} e^{i\theta x} q_{k,n}(x) dx = \int_{\mathbb{R}} e^{i\theta x} dQ_{k,n}(x).$$

Applying inverse Fourier transform, we know

$$\left| p_n(x) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - \sum_{k=1}^{m-2} q_{k,n}(x) \right| = \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} e^{-i\theta x} \phi_n(\theta) d\theta - \int_{-\infty}^{\infty} e^{-i\theta x} \Psi_n(\theta) d\theta \right|$$

$$\leq I_1 + I_2,$$

where

$$I_1 = \frac{1}{2\pi} \int_{|\theta| \leq C\sqrt{n}} |\phi_n(\theta) - \Psi_n(\theta)| d\theta, \quad I_2 = \frac{1}{2\pi} \int_{|\theta| \geq C\sqrt{n}} |\phi_n(\theta) - \Psi_n(\theta)| d\theta, \ C > 0.$$

By Lemma 3.2, there exist positive constants C_1 and C_2 such that

$$|I_1| \leq \frac{1}{2\pi} \int_{|\theta| \leq C\sqrt{n}} \frac{C_1 \sum_{s=0}^{3(m-1)} |\theta|^s}{n^{(m-1)/2}} e^{-\theta^2/6} d\theta \leq \frac{C_2}{n^{(m-1)/2}}$$

where $C_2 = \frac{C_1}{2\pi} \int_{\mathbb{R}} \sum_{s=0}^{3(m-1)} |\theta|^s e^{-\theta^2/6} d\theta < \infty$.

Meanwhile, in the proof of Theorem 2.2, we have already shown that there exist positive constants D_1, D_2 such that $|I_2| \leq D_1 e^{-D_2 n}$. Thus we have proved Theorem 2.5.

5 Mod-ϕ convergence and precise deviations

In this section, we will show precise deviation principle. The key tool is the mod-ϕ convergence theory. First we certify a Lemma on mod-ϕ convergence for Kingman coalescent.

Lemma 5.1. *(Mod-ϕ convergence). Let $\eta(z)$ be the function defined in (2.7). Then $\eta(z)$ is the cumulant of a non-constant infinitely divisible distribution ϕ. Moreover, on any compact subset of $\mathcal{S}_{(-\infty, 1/2)} = \{z \in \mathbb{C} \mid \text{Re}(z) < \frac{1}{2}\}$, we have*

$$\psi_n(z) := e^{-n\eta(z)} E(e^{zn^2 T_n}) \xrightarrow{n \to \infty} \psi(z) = 1 \ uniformly. \tag{5.1}$$

Furthermore, there exists positive constant M such that

$$\sup_{|z| \leq \frac{1}{8}} \left| \psi_n(z) - \psi(z) \right| \leq \frac{M}{n}, \ \ n \geq 3. \tag{5.2}$$

Proof. By the definition of T_n and $\eta(z)$, we know

$$\psi_n(z) = \exp\left\{ n \int_1^\infty \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx \right\}. \tag{5.3}$$

where $\log(1 - \frac{2z}{x^2})$ is the principal branch. It is not difficult to see that $\psi_n(z)$ is analytic on $\mathcal{S}_{(-\infty,1/2)}$ because $\mathrm{Re}(1 - \frac{2z}{x^2}) > 1 - 2\mathrm{Re}(z) > 0$. Next, for given $z \in \mathcal{S}_{(-\infty,1/2)}$, we will show that

$$\log \psi_n(z) = n \int_1^\infty \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx \xrightarrow{n\to\infty} 0$$

uniformly. Indeed,

$$n \int_1^\infty \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx$$

$$= \sum_{k=n}^\infty n \int_{\frac{k}{n}}^{\frac{k}{n}+\frac{1}{n}} \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx$$

$$= \int_0^1 \sum_{k=n}^\infty \left(\log(1 - \frac{2z}{(\frac{k}{n} + \frac{u}{n})^2}) - \log(1 - \frac{2z}{(\frac{k}{n})^2}) \right) du$$

$$+ \int_0^1 \sum_{k=n}^\infty \left(\log(1 - \frac{2z}{(\frac{k}{n})^2}) - \log(1 - \frac{2z}{\frac{k}{n}\frac{k+1}{n}}) \right) du$$

Consider two parametrizations $w_1(s) = \frac{2z}{(k/n)^2(1+\frac{s}{k})^2}, s \in [0, u]$ and $w_2(s) = \frac{2z}{(k/n)^2(1+\frac{s}{k})}, s \in [0, 1]$. We know

$$\sum_{k=n}^\infty \left[\log(1 - \frac{2z}{(\frac{k}{n} + \frac{u}{n})^2}) - \log(1 - \frac{2z}{(\frac{k}{n})^2}) \right] = \frac{1}{k} \int_0^u \sum_{k=n}^\infty \frac{\frac{4z}{(k/n)^2(1+\frac{s}{k})^3}}{1 - \frac{2z}{(\frac{k}{n}+\frac{s}{n})^2}} ds$$

$$= \int_0^u \sum_{p=0}^\infty 4z(2z)^p \sum_{k=n}^\infty \frac{1}{(\frac{k}{n} + \frac{s}{n})^{3+2p}} \frac{1}{n} ds, \tag{5.4}$$

and

$$\sum_{k=n}^\infty \left[\log(1 - \frac{2z}{(\frac{k}{n})^2}) - \log(1 - \frac{2z}{\frac{k}{n}\frac{k+1}{n}}) \right] = -\frac{1}{k} \int_0^1 \sum_{k=n}^\infty \frac{\frac{2z}{(k/n)^2(1+\frac{s}{k})^2}}{1 - \frac{2z}{(\frac{k}{n})^2(1+\frac{s}{k})}} ds$$

$$= -\int_0^1 \sum_{p=0}^\infty 2z(2z)^p \sum_{k=n}^\infty \frac{1}{\frac{k}{n}(\frac{k}{n} + \frac{s}{n})^{2+2p}} \frac{1}{n} ds. \tag{5.5}$$

As functions of s in $[0, 1]$, both $\sum_{k=n}^\infty \frac{1}{\frac{k}{n}(\frac{k}{n} + \frac{s}{n})^{2+2p}} \frac{1}{n}$ and $\sum_{k=n}^\infty \frac{1}{(\frac{k}{n} + \frac{s}{n})^{3+2p}} \frac{1}{n}$ are positive and increasing. It is also easy to show that both of them converge uniformly to a constant $\int_1^\infty \frac{dx}{x^{3+2p}} = \frac{1}{2p+2}$ as $n \to \infty$. Then we know both (5.4) and (5.5) are uniformly convergent in any compact subset of $\mathcal{S}_{(-\infty,\frac{1}{2})}$. Switching the order of summation and integration, we know

$$\lim_{n\to\infty} n \int_1^\infty \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx = 0$$

Now we proceed to show (5.2). Since $|z| < \frac{1}{8}$, we have $|\frac{2z}{x^2}| < 1$ and $|\frac{2z}{[k(k+1)]/n^2}| < 1$. Therefore, $\frac{1}{1 - \frac{2z}{(\frac{k}{n})^2(1+\frac{s}{k})}}, \frac{1}{1 - \frac{2z}{(\frac{k}{n})^2(1+\frac{s}{k})^2}}, \frac{1}{(1+\frac{s}{k})^3}$ and $\frac{1}{(1+\frac{s}{k})^2}$ in (5.4) and (5.5) can be expanded as geometric series and binomial series. After integration by terms, we have

$$n \int_1^\infty \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n}\frac{[nx]+1}{n}}) \right) dx$$

$$= \sum_{p=1}^{\infty} \frac{(-1)^p}{n^p} \sum_{l=1}^{\infty} (2z)^{l+1} \left[\sum_{k=n}^{\infty} \frac{1}{(k/n)^{2(l+1)+p+1}} \frac{1}{n} \right]$$
$$\left[\binom{2l+p+2}{p} \frac{2}{(p+1)(p+2)} - \binom{l+p+1}{p} \frac{1}{p+1} \right].$$

When $n \geq 3$, simple calculation shows that

$$\sup_{|z| < \frac{1}{8}} \left| n \int_1^{\infty} \left(\log(1 - \frac{2z}{x^2}) - \log(1 - \frac{2z}{\frac{[nx]}{n} \frac{[nx]+1}{n}}) \right) dx \right|$$
$$\leq M \sum_{l=1}^{\infty} \frac{1}{4^{l+1}} \sum_{p=1}^{\infty} \frac{1}{n^p} \binom{2l+p+2}{p} \leq \frac{M}{n}.$$

So we have $\sup_{|z| < \frac{1}{8}} \left| e^{-n\eta(z)} E(e^{zn^2 T_n}) - 1 \right| \leq \frac{M}{n}$. Last, one can easily verify that $\eta(z)$ is a cumulant of an infinitely divisible distribution because

$$e^{\eta(i\theta)} = \lim_{n \to \infty} \prod_{k=n}^{\infty} \left(1 - \frac{it}{\frac{k(k+1)}{2n^2}} \right)^{-1/n},$$

and $\left(1 - \frac{it}{\frac{k(k+1)}{2n^2}} \right)^{-1/n}$ is the characteristic function of Gamma distribution with parameters $\frac{k(k+1)}{2n^2}$ and $\frac{1}{n}$. $\qquad \square$

By Theorem 4.2.1, Theorem 4.3.1 and Corollary 4.3.5 in [4], Theorem 2.7 and Theorem 2.8 follow immediately from Lemma 5.1.

6 Deviation inequality

In this section, we prove Theorem 2.9. Different from precise deviations, deviation inequality is about the non-asymptotic estimations of tail probabilities $\mathbb{P}(Z_n \geq x)$ and $\mathbb{P}(Z_n \leq x)$. We need to manipulate the moment generating function of Z_n to establish deviation inequality.

Proof of Theorem 2.9:

Note that $\mathbb{E} e^{\theta Z_n} = e^{\sum_{l=2}^{\infty} \frac{(\sqrt{3}\theta n^{\frac{3}{2}})^l}{l} \sum_{k=n}^{\infty} (\frac{1}{k(k+1)})^l}$. Then for any $0 < \theta < r\frac{\sqrt{n}}{\sqrt{3}}$ and $0 < r < 1$, we have

$$\sum_{l=2}^{\infty} \frac{(\sqrt{3}\theta n^{\frac{3}{2}})^l}{l} \sum_{k=n}^{\infty} (\frac{1}{k(k+1)})^l = \sum_{l=2}^{\infty} \frac{(\sqrt{3}\theta)^l}{l(\sqrt{n})^{l-2}} \sum_{k=n}^{\infty} \frac{1}{[k(k+1)/n^2]^l} \frac{1}{n}$$
$$\leq \frac{3\theta^2}{2} \sum_{l=2}^{\infty} \frac{(\sqrt{3}\theta)^{l-2}}{(\sqrt{n})^{l-2}} \int_1^{\infty} \frac{dx}{x^{2l}} \leq \frac{\theta^2}{2} \frac{1}{1 - \sqrt{3}\theta/\sqrt{n}} \leq \frac{\theta^2}{2(1-r)}.$$

Applying Chernoff's upper bound, we have

$$\mathbb{P}\left(Z_n \geq x \right) \leq \inf_{0 < \theta \leq r\frac{\sqrt{n}}{\sqrt{3}}} e^{-\theta x} \mathbb{E} e^{\theta Z_n} \leq \exp\left\{ - \sup_{0 < \theta \leq r\frac{\sqrt{n}}{\sqrt{3}}} \{\theta x - \frac{\theta^2}{2(1-r)}\} \right\}$$

If we use substitution $\theta' = \frac{\theta}{\sqrt{n}}$, then

$$\mathbb{P}\left(Z_n \geq x \right) \leq \exp\left\{ - n \sup_{0 < \theta' \leq \frac{r}{\sqrt{3}}} \left\{ \frac{x}{\sqrt{n}} \theta' - \frac{(\theta')^2}{2(1-r)} \right\} \right\}.$$

Because the parameter x may change the shape of the function $f(\theta^{'}) = \frac{x}{\sqrt{n}}\theta^{'} - \frac{(\theta^{'})^2}{2(1-r)}$, we need to compute the supremum separately when $0 < x \leq \frac{\sqrt{n}r}{\sqrt{3}(1-r)}$ and $x \geq \frac{\sqrt{n}r}{\sqrt{3}(1-r)}$. For $0 < x \leq \frac{\sqrt{n}r}{\sqrt{3}(1-r)}$, we can show that $\sup_{0 < \theta' \leq \frac{r}{\sqrt{3}}} f(\theta^{'}) = f((1-r)x/\sqrt{n}) = \frac{(1-r)x^2}{2n}$. When $x \geq \frac{\sqrt{n}r}{\sqrt{3}(1-r)}$, we have $\sup_{0 < \theta' \leq \frac{r}{\sqrt{3}}} f(\theta^{'}) = f(r/\sqrt{3}) > \frac{xr}{\sqrt{3n}} - \frac{r}{6(1-r)}\frac{x\sqrt{3}(1-r)}{\sqrt{n}} = \frac{xr}{2\sqrt{3n}}$. Hence, for $r \in (0,1)$, we have

$$\mathbb{P}\left(Z_n \geq x\right) \leq \begin{cases} \exp(-\frac{(1-r)}{2n}x^2) & 0 < x \leq \frac{\sqrt{n}r}{\sqrt{3}(1-r)} \\ \exp(-\frac{xr\sqrt{n}}{2\sqrt{3}}) & x \geq \frac{\sqrt{n}r}{\sqrt{3}(1-r)} \end{cases}.$$

Meanwhile, the estimation of $\mathbb{P}\left(Z_n \leq x\right), x < 0$, can be done similarly by applying the above procedure to $-Z_n$.

References

[1] Abramowitz,D.J. and Stegun, I.A.: Handbook of mathematical functions with formulas, graphs, and mathematical tables. *Dover, New York,* (1964). MR-0415956

[2] Aldous, D.J.: Deterministic and Stochastic Models for Coalescence (Aggregation and Coagulation): A Review of the Mean-Field Theory for Probabilists.*Bernoulli.* **5(1)** (1999). MR-1673235

[3] Depperschmidt,A., Pfaffelhuber,P., and Scheuringer,A.: Some large deviations in Kingman's coalescent.*Electronic Communications in Probability.* **20** (2015). MR-3304413

[4] Feray,V., Meliot,P., and Nikeghbali,A.: Mod-ϕ Convergence I: Normality zones and precise deviations.*Springer Briefs in Probability and Mathematical Statistics.* (2016). MR-3585777

[5] Kingman, J.F.C.: The coalescent. *Stochastic Process. Appl.* **13(3)** (1982). MR-0671034

[6] Gao,F.G., Zhu,L.: Precise deviations for Hawkes processes. *Bernoulli.* **27** (2021):221–248. MR-4177368

[7] Griffiths, R.C.: Asymptotic line-of-descent distributions. *J. Math. Biol.* **21(1)** (1984):67–75. MR-0770713

[8] Petrov,V.V.: Sums of Independent Random Variables. *Springer, New York* (1975). MR-0388499

[9] Limic,V., Talarczyk,A: Diffusion limits at small times for Λ-coalescents with a Kingman component. *Electron. J. Probab.* **20(45)** (2015):1–20. MR-3339865

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: http://www.imstat.org/publications/open.htm