The Annals of Applied Probability
2019, Vol. 29, No. 2, 1188–1232
https://doi.org/10.1214/18-AAP1433

ON POISSON APPROXIMATIONS FOR THE EWENS SAMPLING FORMULA WHEN THE MUTATION PARAMETER GROWS WITH THE SAMPLE SIZE

BY KOJI TSUKUDA

The University of Tokyo

The Ewens sampling formula was first introduced in the context of population genetics by Warren John Ewens in 1972, and has appeared in a lot of other scientific fields. There are abundant approximation results associated with the Ewens sampling formula especially when one of the parameters, the sample size n or the mutation parameter θ which denotes the scaled mutation rate, tends to infinity while the other is fixed. By contrast, the case that θ grows with n has been considered in a relatively small number of works, although this asymptotic setup is also natural. In this paper, when θ grows with n, we advance the study concerning the asymptotic properties of the total number of alleles and of the component counts in the allelic partition assuming the Ewens sampling formula, from the viewpoint of Poisson approximations. Specifically, the main contributions of this paper are deriving Poisson approximations of the total number of alleles, an independent process approximation of small component counts, and functional central limit theorems, under the asymptotic regime that both n and θ tend to infinity.

1. Introduction. For a positive integer n, consider a sequence $\{C_j^n\}_{j=1}^{\infty}$ of nonnegative integer-valued random variables satisfying $\sum_{j=1}^n jC_j^n = n$ and $C_j^n = 0$ for $j > n$. For $b = 1, \dots, n$, let us denote $\mathbf{C}_b^n = (C_1^n, \dots, C_b^n)$ and $\mathbf{a}_b = (a_1, \dots, a_b)$, where $a_1, \dots, a_b$ are nonnegative integers. The random vector $\mathbf{C}_n^n$ denotes the component counts in a random combinatorial structure of size n. In the context of population genetics, Ewens (1972) introduced what is called *the Ewens sampling formula*

$$\mathsf{P}(\mathbf{C}_n^n = \mathbf{a}_n) = \frac{n!}{(\theta)_n} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{a_j} \frac{1}{a_j!} 1\left\{\sum_{j=1}^n ja_j = n\right\} \tag{1.1}$$

as the distribution of the allelic partition in a sample of size n from a random population following the stationary distribution of the infinitely-many neutral allele model with scaled mutation rate $\theta > 0$, where $(\theta)_n$ is the rising factorial $\theta(\theta+1)\cdots(\theta+n-1)$. The distribution of the descending order population frequency is referred to as the Poisson–Dirichlet distribution; see, for instance, Section 2.5 of Feng (2010) for the derivation of (1.1) and basic properties. Hereafter,

Received August 2017; revised August 2018.
MSC2010 subject classifications. Primary 60F05; secondary 60B12, 62E20, 92D10.
Key words and phrases. Ewens sampling formula, the Feller coupling, functional central limit theorem, Poisson approximation.

we consider (1.1) as a model of $\{C_j^n\}_{j=1}^n$. A random partition whose component counts follow (1.1) is called a *Ewens partition*. When a permutation of n letters is chosen equally likely, the numbers of cyclic permutations whose lengths are $1, \dots, n$ follow (1.1) with $\theta = 1$. The unsigned Stirling number of the first kind $\bar{s}(n, k)$ $(k = 1, \dots, n)$ is the coefficient of θ^k in $(\theta)_n$, and is in conformity with the number of permutations of n letters with k disjoint cycles. Hence, if (1.1) is assumed, the total number $K_n = \sum_{j=1}^n C_j^n$ of alleles included in a sample of size n, in other words the total number of distinct cycles in a random permutation, follows *the falling factorial distribution* (Watterson (1974a))

$$\mathsf{P}(K_n = k) = \bar{s}(n, k) \frac{\theta^k}{(\theta)_n} \tag{1.2}$$

for $k = 1, \dots, n$. In this paper, we will present asymptotic properties, especially Poisson approximations, of $\mathbf{C}_b^n$ and K_n under the asymptotic regime that both n and θ increase. Specifically, there are three major goals: under the asymptotic regime that θ grows with n, we will derive Poisson approximations of K_n; an independent process approximation of $\mathbf{C}_b^n$; and functional central limit theorems. As it will be seen in Section 2, these topics have been studied so far with fixed θ setting in many studies including Arratia, Barbour and Tavaré (1992), Arratia, Stark and Tavaré (1995), Arratia and Tavaré (1992a), Hansen (1990), Tsukuda (2018). As for the first goal, we consider Poisson approximations of K_n and $n - K_n$. As for the second goal, we show a condition to derive an independent process approximation of $\mathbf{C}_b^n$ for $b = 1, 2, \dots$ and its total variation asymptotics. As for the third goal, we extends functional central limit theorems for the Ewens sampling formula.

Let us explain motivations to study (1.1) and (1.2). A uniform random permutation is a classic combinatorial probabilistic model traceable back to Pierre-Rémond de Montmort's essay published in 1708. The Montmort problem is calculating $\mathsf{P}(C_1^n \neq 0)$ when $\theta = 1$. The uniform random permutation has been studied in Goncharov (1944), Shepp and Lloyd (1966), DeLaurentis and Pittel (1985), Arratia and Tavaré (1992b) and a lot of other works. See also Chapters 4 and 10 of Barbour, Holst and Janson (1992). Studies on (1.1) and (1.2) can be regarded as one on random permutations where a permutation is not chosen equally likely. Moreover, although (1.1) was derived firstly from the concrete model in population genetics, it has been widely applied to other fields. For instance, to conduct a Bayesian procedure for nonparametric problems in statistics, a favorable prior is the Dirichlet process introduced by Ferguson (1973), and a random sample partition from the Dirichlet process follows (1.1) (Antoniak (1974)). We refer to Favaro and James (2016) and Teh (2016) for more details about the literature in the Bayesian nonparametrics. In addition, distributions of large component counts in a lot of random combinatorial structures can be approximated by using the Ewens sampling formula (Arratia, Barbour and Tavaré (2000)). Crane (2016) and Chapter 41 of Johnson, Kotz and Balakrishnan (1997), whose write-up was provided by S. Tavaré and W. J. Ewens, are general review articles on the Ewens sampling formula. They introduce several results including applications to other fields such as ecology, physics and so on.

For (1.1), (1.2) and related probabilistic models, a lot of works have discussed asymptotic properties under the situations $n \to \infty$ with fixed θ or $\theta \to \infty$ with fixed n; see, for instance, Feng (2016). On the other hand, it is natural to consider the asymptotic regime that both n and θ tend to infinity. There are several motivations to consider this regime. A motivation in genetics is that θ is proportional to the population size in the original infinitely-many neutral allele model. Another motivation in genetics is that the large θ setting corresponds to the small homozygosity $\sum_{j=1}^{\infty} f_j^2$, where $\{f_j\}_{j=1}^{\infty}$ is the population frequency. The asymptotic regime that the population frequency decreases as the sample size increases is commonly discussed, and the expected homozygosity is given by $1/(1+\theta)$ under the infinitely-many neutral allele model (Ewens (1972)). When the Poisson–Dirichlet random population is considered, a corresponding assumption is that θ and n simultaneously increase. Moreover, when (1.1) is used as a statistical model of random partitions, the sample size n may be small relative to θ in some actual cases. When one of the parameters is not quite larger than the other, asymptotic properties established under the regime that only one parameter tends to infinity do not provide good approximations. A recipe for addressing this issue is considering the asymptotic regime that both parameters simultaneously tend to infinity. Furthermore, in Bayesian nonparametrics, there is a methodology using a prior depending on the sample size, so the parameter θ of the Dirichlet process may depend on n.

Finally, let us introduce some previous works in which the asymptotic regime that both n and θ tend to infinity is considered. Section 4 of Feng (2007) and Tsukuda (2017b) discussed the asymptotic behavior of K_n. Varron (2014) proved the nonparametric Bernstein–von Mises phenomenon for the Dirichlet process prior when $\theta^2/n \to 0$. Along the lines of these works, this paper provides novel results under the regime.

1.1. *Notation.* Consider sequences $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$. If $x_n/y_n \to 1$, then we write $x_n \sim y_n$. Let $c < \infty$ be a constant. If $x_n/y_n \to 0$, then we write $x_n = o(y_n)$, if $x_n/y_n \to c$ then we write $x_n = O(y_n)$, and if $x_n/y_n \to c \neq 0$ then we write $x_n = \Theta(y_n)$. Let $\sum_{j=1}^{0} z_j = 0$ and $\prod_{j=1}^{0} z_j = 1$ for any sequence $\{z_n\}_{n=-\infty}^{\infty}$, and let $(z)_0 = 1$ for any value z. When we consider the limits of n and θ simultaneously, we use the notation $\lim_{n,\theta}$.

Let $[x^k]f(x)$ denote the coefficient of x^k in the power series expansion of $f(x)$. Let $f^{(i)}(\cdot)$ denote the ith derivative of function $f(\cdot)$. Let $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor function and the ceiling function, respectively. Let $\Gamma(\cdot)$ be the gamma function and $\psi(\cdot) = (\log\Gamma(\cdot))'$ the digamma function. For real x, let x^+ denote the positive part of x.

The space $D[0,1]$ is the set of càdlàg functions on $[0,1]$ endowed with the Skorokhod topology. The space $L^2(0,1)$ is equivalence classes of real valued functions on $(0,1)$ which are square integrable with respect to the Lebesgue measure endowed with the L^2 topology.

Let $\mathcal{Z}_+$ be the set of all nonnegative integers. For $\mathcal{Z}_+^d$-valued random vectors X and Y (d is a positive integer), $d_{\mathrm{TV}}(X,Y)$ denotes the total variation distance between the distributions which X and Y follow, where $d_{\mathrm{TV}}(X,Y)$ is defined by $\sup_{A\subset\mathcal{Z}_+^d}|\mathsf{P}(X\in A)-\mathsf{P}(Y\in A)|$. Note that the definition is equivalent to $d_{\mathrm{TV}}(X,Y)=\frac{1}{2}\sum_{\mathbf{a}\in\mathcal{Z}_+^d}|\mathsf{P}(X=\mathbf{a})-\mathsf{P}(Y=\mathbf{a})|$.

The convergence of a sequence of random variables $\{X_i\}_{i\geq 1}$ to a random variable Y in probability and the weak convergence of $\{X_i\}_{i\geq 1}$ to Y are denoted by $X_i\to^p Y$ and $X_i\Rightarrow Y$, respectively.

Throughout the paper, Poisson variables frequently appear. To simplify explanations, let us denote a Poisson variable with mean $\Lambda(>0)$ by P_Λ.

1.2. *Asymptotic regimes.* Letting c be a finite constant, we study the following asymptotic regimes in this paper:

- Case A: $n/\theta\to\infty$;
- Case B: $n/\theta\to c$ with $0<c<\infty$;
- Case C: $n/\theta\to 0$;
- Case C1: $n/\theta\to 0$ and $n^2/\theta\to\infty$;
- Case C2: $n^2/\theta\to c$ with $0<c<\infty$;
- Case C3: $n^2/\theta\to 0$.

This division was introduced in Tsukuda (2017b). It should be noted that in Section 4 of Feng (2007), when θ does not converge to 0, the relation between n and θ are divided into Cases A, B, C above and $\theta\to\infty$ with fixed n. Moreover, throughout this paper, we assume that θ does not decrease as n increase.

REMARK 1.1. In Case C3, it holds that $K_n-n\to^p 0$. Note that when $\theta=o(1/\log n)$ in which we are not interested since $\theta\to 0$, it holds that $K_n-1\to^p 0$. These convergences can be checked through showing the convergence in first mean.

1.3. *Organization.* In Section 2, we review asymptotic results associated with the Ewens sampling formula in the literature which will be discussed in the later part of this paper, and introduce our contributions. Section 3 is devoted to show Poisson approximations of K_n and $n-K_n$ in Case A and Case C, respectively. Section 4 is devoted to discuss independent process approximations of $\mathbf{C}_b^n$ in a Ewens partition. Section 5 is devoted to show the functional central limit theorems for the Ewens sampling formula. In addition, the Appendix includes some auxiliary results used in proofs.

2. Results in the literature and this paper.

2.1. *Normal and Poisson approximations of K_n.* In the combinatorial context, it is worthwhile to know when typical distributions such as Normal, Poisson or other distributions asymptotically appear; see, for instance, Flajolet and Soria (1990). For the total number K_n of alleles which follows (1.2), Watterson (1974b) proved the following central limit theorem (CLT for short): For fixed $\theta > 0$,

$$\frac{K_n - \theta \log n}{\sqrt{\theta \log n}} \Rightarrow N(0, 1) \tag{2.1}$$

as $n \to \infty$, where $N(0, 1)$ is a standard normal variable. A stronger result, the Poisson approximation of K_n, was stated by Arratia and Tavaré (1992a): For fixed $\theta > 0$,

$$d_{\mathrm{TV}}(K_n, P_{\mathsf{E}[K_n]}) = \Theta\left(\frac{1}{\log n}\right) \tag{2.2}$$

as $n \to \infty$, where

$$\mathsf{E}[K_n] = \sum_{j=1}^{n} \frac{\theta}{\theta + j - 1}.$$

To improve the approximation accuracy, Yamato (2013) provided the following CLT which adopts another standardization: For fixed $\theta > 0$,

$$\frac{K_n - \theta(\log n - \psi(\theta))}{\sqrt{\theta(\log n - \psi(\theta))}} \Rightarrow N(0, 1) \tag{2.3}$$

as $n \to \infty$. Moreover, Yamato (2013) showed the approximation of K_n by a Poisson variable with the approximate mean: For fixed $\theta > 0$,

$$d_{\mathrm{TV}}(K_n, P_{\theta(\log n - \psi(\theta))}) = O\left(\frac{1}{\log n}\right) \tag{2.4}$$

as $n \to \infty$.

When θ grows with n, the standardization should be changed in many cases. Let $\mu = \theta \log(1 + n/\theta)$ and $\sigma^2 = \theta\{\log(1 + n/\theta) - n/(n + \theta)\}$. Tsukuda (2017b) showed that

$$\frac{K_n - \mu}{\sigma} \Rightarrow \begin{cases} N(0, 1) & \text{(Case A, B, C1)}, \\ (c/2 - P_{c/2})/\sqrt{c/2} & \text{(Case C2)}, \\ 0 & \text{(Case C3)}, \end{cases} \tag{2.5}$$

where $c = \lim_{n,\theta}(n^2/\theta)$ in Case C2.

Following (2.2) and (2.4), Theorem 3.1 and Proposition 3.1 give Poisson approximations in large θ setup. These results show that in Case A the distribution of K_n is approximately Poisson and in Case C the distribution of $n - K_n$ is approximately Poisson. Our results give a background of (2.5) in Cases A and C.

REMARK 2.1. Professor Shuhei Mano pointed out that the proof of Theorem 2 in Tsukuda (2017b), which asserts (2.5) above in Case A, B and C1, is incorrect in Case C1, even though the result holds true. The correction note will appear in Journal of Applied Probability.

REMARK 2.2. As a corollary to the large deviation principle for K_n when $\theta \to \infty$, Feng (2007) provided the following weak laws of large numbers in Corollary 4.1:

$$\frac{K_n}{\theta \log(n/\theta)} \to^p 1 \qquad \text{(Case A)},$$

$$\frac{K_n}{n} \to^p \begin{cases} \log\left(1+\dfrac{1}{c}\right)^c & \text{(Case B)}, \\ 1 & \text{(Case C)} \end{cases} \tag{2.6}$$

and $K_n \to^p n$ as $\theta \to \infty$ with fixed n. These laws of large numbers in Cases A, B and C can be obtained directly from the calculation of $\mathsf{E}[|K_n/\mathsf{E}[K_n] - 1|^2]$; see Proposition 2 of Tsukuda (2017b).

2.2. *Independent process approximations of* $\mathbf{C}_b^n$. Consider a sequence $\{Z_j\}_{j=1}^\infty$ of independent Poisson variables with $\mathsf{E}[Z_j] = \theta/j$ for $j = 1, 2, \dots$ and denote $\mathbf{Z}_b = (Z_1, \dots, Z_b)$ for a positive integer b. Then it is well known that (1.1) can be derived from *the conditioning relation*

$$\mathsf{P}(\mathbf{C}_n^n = \mathbf{a}_n) = \mathsf{P}\left(\mathbf{Z}_n = \mathbf{a}_n \,\middle|\, \sum_{j=1}^n jZ_j = n\right); \tag{2.7}$$

see, for instance, Watterson (1974a). It means that the dependence in $\{C_j^n\}_{j=1}^n$ is given by the condition $\sum_{j=1}^n jZ_j = n$. It is of interest to discuss whether the effect of this dependence asymptotically vanishes or not. It was answered by Arratia, Barbour and Tavaré (1992) who showed the small components can be approximated by independent Poisson variables: For any fixed positive integer b, it holds that

$$(C_1^n, \dots, C_b^n) \Rightarrow (Z_1, \dots, Z_b) \tag{2.8}$$

as $n \to \infty$. Note that (2.8) is equivalent to $\lim_{n\to\infty} d_{\mathrm{TV}}(\mathbf{C}_b^n, \mathbf{Z}_b) = 0$ because both $\mathbf{C}_b^n$ and $\mathbf{Z}_b$ are discrete. Our Corollary 4.1, which summarizes the results in Propositions 4.1 and 4.2, shows that in Case A the distribution of $\mathbf{C}_b^n$ is approximately one of $\mathbf{Z}_b$ for fixed b if, and only if, $\theta^2/n \to 0$.

It is more interesting to consider the case that b grows with n. To describe such results, let us begin with preparing some notions. For positive integer b, let us denote the total variation distance between $\mathbf{C}_b^n = (C_1^n, \dots, C_b^n)$ and $\mathbf{Z}_b = (Z_1, \dots, Z_b)$ by $d_b(n)$, that is, $d_b(n) = d_{\mathrm{TV}}(\mathbf{C}_b^n, \mathbf{Z}_b)$. A *coupling* of two random variables is constructed as a joint random variable which have the same marginal distributions as the original random variables. If $\mathbf{C}_b^n$ and $\mathbf{Z}_b$ are coupled, then it generally holds that $d_b(n) \le \mathsf{P}(\mathbf{C}_b^n \neq \mathbf{Z}_b)$. However, there exists a maximal coupling which attains the equality, so

$$d_b(n) = \inf_{\text{couplings}} \mathsf{P}\big(\mathbf{C}_b^n \neq \mathbf{Z}_b\big),$$

where the infimum in the above display is taken over all couplings of $\mathbf{C}_b^n$ and $\mathbf{Z}_b$ on a common probability space; see Section A.1 of Barbour, Holst and Janson (1992). Moreover, for positive integer b, let us denote the distance between the distributions of $\mathbf{C}_b^n = (C_1^n, \dots, C_b^n)$ and $\mathbf{Z}_b = (Z_1, \dots, Z_b)$ in the Wasserstein ℓ^1 metric by $d_b^W(n)$ which satisfies

$$d_b^W(n) = \inf_{\text{couplings}} \sum_{j=1}^{b} \mathsf{E}\big[|C_j^n - Z_j|\big].$$

For the definition of this distance, see Section A.1 of Barbour, Holst and Janson (1992). Since $d_b(n) \le \mathsf{P}(\mathbf{C}_b^n \neq \mathbf{Z}_b) = \mathsf{P}(\sum_{j=1}^b |C_j^n - Z_j| \ge 1) \le E[\sum_{j=1}^b |C_j^n - Z_j|]$ for any coupling of $\mathbf{C}_b^n$ and $\mathbf{Z}_b$, it holds that $d_b(n) \le d_b^W(n)$. As for the Ewens sampling formula, $d_b^W(n)$ is a convenient measure of approximations because a concrete construction, *the Feller coupling*, can be given. See Arratia, Barbour and Tavaré (1992, 2016). The Feller coupling is as follows: Let $\{\xi_j\}_{j=1}^\infty$ be a sequence of independent Bernoulli variables with $\mathsf{P}(\xi_j = 1) = p_j = \theta/(\theta + j - 1)$ for any $j = 1, 2, \dots$. Then the Ewens sampling formula (1.1) is given as the joint distribution of

$$C_1^n = \sum_{i=1}^{n-1} \xi_i \xi_{i+1} + \xi_n$$

and

$$C_j^n = \sum_{i=1}^{n-j} \xi_i (1 - \xi_{i+1}) \cdots (1 - \xi_{i+j-1}) \xi_{i+j} + \xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n)$$

for $j = 2, \dots, n$. Moreover, define

$$C_j^\infty = \sum_{i=1}^{\infty} \xi_i (1 - \xi_{i+1}) \cdots (1 - \xi_{i+j-1}) \xi_{i+j}$$

for $j = 1, 2, \dots$, then $\{C_j^\infty\}_{j=1}^\infty$ is a sequence of independent Poisson variables with $\mathsf{E}[C_j^\infty] = \theta/j$ $(j = 1, 2, \dots)$. That is because the convergences in probability $\xi_n \to^p 0$ and $\xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n) \to^p 0$ for any $j = 2, 3, \dots$ yield that $C_j^n \Rightarrow C_j^\infty$, and so (2.8) yields that $C_j^\infty =^d Z_j$ for any $j = 1, 2, \dots$. By using the

Feller coupling, Arratia, Barbour and Tavaré (1992) proved the Poisson process approximation when b grows with n:

$$d_b(n) \to 0 \quad \Leftrightarrow \quad b = o(n); \tag{2.9}$$

$$d_b(n) \le \frac{b\theta}{\theta+n}\left(\theta + \frac{n}{\theta+n-b}\right); \tag{2.10}$$

$$d_b^W(n) \le \frac{b\theta}{\theta+n-b}\left(\theta + \frac{n}{\theta+n}\right); \tag{2.11}$$

$$d_n^W(n) = O(1); \tag{2.12}$$

if $\theta \ge 1$ then

$$\frac{\theta(\theta-1)b}{\theta+n-1}\left\{1 - \frac{(\theta-1)(b+1)}{4(\theta+n-1)}\right\} \le d_b^W(n) \le \frac{b\theta(\theta+1)}{\theta+n}. \tag{2.13}$$

Note that (2.10), (2.11) and (2.13) are not asymptotic results. As for lower bound results for the total variation distance, which complement (2.10), Arratia, Barbour and Tavaré (1992) showed that

$$\liminf_{n\to\infty}(nd_b(n)) \ge \left(\frac{b\theta|\theta-1|}{2}\right)\exp\left(-\theta\sum_{j=1}^{b}\frac{1}{j}\right),$$

and Barbour (1992) showed that if $\theta \ne 1$ then $d_b(n) \ge c_3 b/n$ for some positive constant c_3 which depends on θ.

Another fascinating result for evaluating $d_b(n)$ is deriving the leading term of $d_b(n)$, which were given by Arratia, Stark and Tavaré (1995) for general logarithmic assemblies. If the Ewens sampling formula is considered, the statement is as follows: If $b = o(n/\log n)$, then

$$d_b(n) = \frac{|1-\theta|}{2n}\mathsf{E}\big[|T_{0b} - \theta b|\big] + o\left(\frac{b}{n}\right), \tag{2.14}$$

where $T_{0b} = \sum_{j=1}^{b} jZ_j$. As it is stated in Corollary 4 of their paper, if $\theta \ne 1$ and if $b = o(n/\log n)$ then the leading term of $d_b(n)$ is given by the first term in the right-hand side of (2.14).

An important application of the Poisson approximation for $\mathbf{C}_b^n$ is deriving the asymptotic property of short cycle lengths. The kth shortest cycle lengths $\{S_n^k\}_{k=1}^{\infty}$ is defined by

$$S_n^k = \inf\big(j : C_1^n + \cdots + C_j^n \ge k\big) \tag{2.15}$$

for $k = 1, 2, \ldots$ and $S_n^k = \infty$ when there is no such j. It was studied in, for instance, Shepp and Lloyd (1966) and Arratia and Tavaré (1992a).

In Case A, we will show independent process approximations in large θ setup. Theorem 4.1 provides the total variation asymptotics corresponding to

(2.14). Moreover, Proposition 4.3, which directly follows from (2.13), provides the asymptotics for $d_b^W(n)$. These results show that if $\theta \to \infty$ then the asymptotic decay rates of $d_b(n)$ and $d_b^W(n)$ are different. The convergence of $d_b^W(n)$ is applied to see an asymptotic property of the kth shortest cycle lengths in Corollary 4.2. Furthermore, following (2.12), Proposition 4.4 gives the bound of $d_n^W(n)$.

On the other hand, in Case C, such independent process approximations seem difficult. Indeed, our Theorem 4.2 shows that in Case C2 only C_1^n and C_2^n have nondegenerate limit, and in Case C3 all components of $\mathbf{C}_n^n$ become degenerate. This result implies asymptotic properties of the shortest and longest cycle length (Corollaries 4.3 and 4.4).

REMARK 2.3. Developing a method to generate combinatorial structures randomly is an important problem in combinatorics. The Feller coupling can be applied to generate not only component counts which follow (1.1) but also ones of other logarithmic combinatorial structures (Arratia et al. (2018)). Moreover, a recently proposed algorithm, probabilistic divide-and-conquer, is a efficient way to generate component counts of random structures whose laws are given by independent random variables with conditioning (like (2.7)) (Arratia and DeSalvo (2016), DeSalvo (2018)).

2.3. *Functional central limit theorems.* The results in Arratia, Barbour and Tavaré (1992) provide an elegant way to derive asymptotic properties. Among others, by using (2.12), Arratia and Tavaré (1992a) provided an alternative proof of the functional central limit theorem for the Ewens sampling formula which was originally proved by Hansen (1990): The random process

$$X_n^1(\cdot) = \left(\frac{\sum_{i=1}^{\lfloor n^u \rfloor} C_j^n - u\theta \log n}{\sqrt{\theta \log n}} \right)_{0 \le u \le 1} \tag{2.16}$$

converges weakly to $(B(u))_{0 \le u \le 1}$ in $D[0, 1]$ as $n \to \infty$, where $B(\cdot)$ is a standard Brownian motion. Note that DeLaurentis and Pittel (1985) demonstrated the weak convergence of $X_n^1(\cdot)$ when $\theta = 1$. The approach of Arratia and Tavaré (1992a) is generalized for broader logarithmic structures; see Arratia, Stark and Tavaré (1995) and Arratia, Barbour and Tavaré (2000). Moreover, by using the Poisson process approximation, Tsukuda (2018) provided a weighted version in $L^2(0, 1)$: Both of the random processes

$$X_n^2(\cdot) = \left(\frac{\sum_{i=1}^{\lfloor n^u \rfloor} C_j^n - \theta \sum_{j=1}^{\lfloor n^u \rfloor} 1/j}{\sqrt{\theta \sum_{j=1}^{\lfloor n^u \rfloor} 1/j}} \right)_{0 < u < 1} \tag{2.17}$$

and

$$X_n^3(\cdot) = \left(\frac{\sum_{i=1}^{\lfloor n^u \rfloor} C_j^n - u\theta \log n}{\sqrt{u\theta \log n}} 1\left\{ u > \frac{\varepsilon}{\log n} \right\} \right)_{0 < u < 1} \tag{2.18}$$

converge weakly to $(B(u)/\sqrt{u})_{0<u<1}$ in $L^2(0,1)$ as $n \to \infty$, where ε is a positive constant.

In Theorem 5.1, the weak convergence results for $X_n^1(\cdot)$, $X_n^2(\cdot)$ and $X_n^3(\cdot)$ are extended to the case where θ slightly increase as n increase. The meaning of slightly is the assumptions, (5.1) and (5.2), require that θ increases very slowly compared with n.

Let R_j be the jth cycle length in a random permutation of n which has K_n disjoint cycles, and the *loglength* of jth cycle is defined by $\log_n R_j$. Consider its empirical distribution function $F_n(\cdot)$ defined as

$$F_n(u) = \frac{\sum_{j=1}^{K_n} 1\{\log_n R_j \le u\}}{K_n} = \frac{\sum_{j=1}^{\lfloor n^u \rfloor} C_j^n}{K_n} \qquad (0 \le u \le 1).$$

Define the random processes

$$X_n^4(\cdot) = \bigl(\sqrt{\theta \log n}\bigl(F_n(u) - u\bigr)\bigr)_{0 \le u \le 1}, \tag{2.19}$$

and

$$X_n^5(\cdot) = \left(\sqrt{\theta \log n}\frac{(F_n(u) - u)}{\sqrt{u(1-u)}} 1\left\{\frac{\varepsilon}{\log n} < u < 1 - \frac{\varepsilon}{\log n}\right\}\right)_{0<u<1}, \tag{2.20}$$

where ε is a positive constant. When $\theta = 1$, the weak convergence of $X_n^4(\cdot)$ to a standard Brownian bridge $(B^\circ(u))_{0\le u\le 1}$ in $D[0,1]$ was shown by DeLaurentis and Pittel (1985); see the notes (2) after the theorem in their paper. Its extension to the Ewens sampling formula may have not appeared in the literature. We will present an extended version in Theorem 5.2 under the same assumptions as Theorem 5.1.

2.4. *Auxiliary results in the literature.* In this subsection, let us set out some auxiliary results concerning Poisson approximations which will be used in the proofs of our statements.

Consider a sequence of independent Bernoulli variables $\{\xi_j\}_{j=1}^\infty$ and its partial sum $S_n = \sum_{j=1}^n \xi_j$, where $\mathsf{P}(\xi_j = 1) = p_j$ for any $j = 1, 2, \ldots$. Then, by using the Chen–Stein method, Theorems 1 and 2 of Barbour and Hall (1984) gave a sharp bound for the Poisson approximation of a partial sum of Bernoulli variables: For a Poisson variable P_λ with mean $\lambda = \sum_{j=1}^n p_j$, it holds that

$$\frac{1 \wedge \lambda^{-1}}{32} \sum_{j=1}^n p_j^2 \le d_{\mathrm{TV}}(S_n, P_\lambda) \le \frac{1 - e^{-\lambda}}{\lambda} \sum_{j=1}^n p_j^2. \tag{2.21}$$

Moreover, from a property of the Hellinger integral, a bound for the total variation distance between two Poisson distributions were given in Theorem 2.1 of Yannaros (1991): For Poisson variables P_{λ_1} and P_{λ_2} with respective means λ_1 and λ_2, it holds that

$$d_{\mathrm{TV}}(P_{\lambda_1}, P_{\lambda_2}) \le \min\bigl(|\sqrt{\lambda_1} - \sqrt{\lambda_2}|, |\lambda_1 - \lambda_2|\bigr). \tag{2.22}$$

3. Poisson approximations of K_n. In this section, we establish Poisson approximations of K_n in Cases A and C. Define

$$p_j = \frac{\theta}{\theta + j - 1}, \qquad q_j = 1 - p_j = \frac{j-1}{\theta + j - 1} \qquad (j = 1, 2, \dots) \tag{3.1}$$

and let $\lambda_A = \sum_{j=1}^n p_j$ and $\lambda_C = \sum_{j=1}^n q_j$. Some asymptotic evaluations associated with $\{p_j\}_{j=1}^\infty$ and $\{q_j\}_{j=1}^\infty$ are presented in Section A.1. Introduce two Poisson variables P_{λ_A} and P_{λ_C}. Then $\mathsf{E}[P_{\lambda_A}] = \mathsf{E}[K_n]$ and $\mathsf{E}[P_{\lambda_C}] = n - \mathsf{E}[K_n]$. We first show Poisson approximations corresponding to (2.2).

THEOREM 3.1. (i) *In Case A*,

$$d_{\mathrm{TV}}(K_n, P_{\lambda_A}) \le \frac{n\theta + n + \theta}{\theta(n+\theta)\log(1+n/\theta) + n/2}, \tag{3.2}$$

and

$$d_{\mathrm{TV}}(K_n, P_{\lambda_A}) = \Theta\left(\frac{1}{\log(n/\theta)}\right).$$

(ii) *In Case C*,

$$d_{\mathrm{TV}}(n - K_n, P_{\lambda_C}) \le \frac{2n(n+\theta)}{3\theta^2}\left(1 - e^{-n^2/2\theta}\right), \tag{3.3}$$

and

$$d_{\mathrm{TV}}(n - K_n, P_{\lambda_C}) = \Theta\left(\frac{n}{\theta}\left(1 \wedge \frac{n^2}{\theta}\right)\right).$$

PROOF. Let $\{\xi_j\}_{j=1}^\infty$ and $\{\zeta_j\}_{j=1}^\infty$ be sequences of independent Bernoulli variables with respective parameters $\mathsf{P}(\xi_j = 1) = p_j$ and $\mathsf{P}(\zeta_j = 1) = q_j$ for $j = 1, 2, \dots$. Then it holds that $K_n =^d \sum_{i=1}^n \xi_i$ and that $n - K_n =^d \sum_{i=1}^n \zeta_i$ (see, for instance, (41.12) of Johnson, Kotz and Balakrishnan (1997)). To prove the desired results, we will use (2.21) and Proposition A.1.

(i) The result (3.2) follows from

$$d_{\mathrm{TV}}(K_n, P_{\lambda_A}) \le \frac{\sum_{j=1}^n p_j^2}{\sum_{j=1}^n p_j} \le \frac{n\theta/(n+\theta) + 1}{\theta\log(1+n/\theta) + n/\{2(\theta+n)\}}.$$

Since $\sum_{j=1}^n p_j \to \infty$, it holds for enough large n that

$$d_{\mathrm{TV}}(K_n, P_{\lambda_A}) \ge \frac{\sum_{j=1}^n p_j^2}{32\sum_{j=1}^n p_j} \ge \frac{n\theta/(n+\theta)}{32\theta\log(1+n/\theta)}.$$

The above two displays yield $d_{\mathrm{TV}}(K_n, P_{\lambda_A}) = \Theta(1/\log(n/\theta))$.

(ii) The result (3.3) follows from

$$
\begin{aligned}
d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) &\le \bigl(1-e^{-\sum_{j=1}^n q_j}\bigr)\frac{\sum_{j=1}^n q_j^2}{\sum_{j=1}^n q_j} \\
&\le \bigl(1-e^{-n^2/2\theta}\bigr)\frac{n(n-1)(2n-1)/(6\theta^2)}{n(n-1)/\{2(\theta+n)\}}.
\end{aligned}
$$

In Case C1, since $\sum_{j=1}^n q_j \to \infty$, it holds for enough large n that

$$
d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) \ge \frac{\sum_{j=1}^n q_j^2}{32\sum_{j=1}^n q_j} \ge \frac{n(n-1)(2n-1)/\{6(\theta+n)^2\}}{32n(n-1)/2\theta}.
$$

The above two displays yield $d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) = \Theta(n/\theta)$. In Case C2, since $1-e^{-\sum_{j=1}^n q_j} \le 1$ and since $(1/\sum_{j=1}^n q_j)$ is bounded by some constant for enough large n, the same evaluation provides $d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) = \Theta(n/\theta)$. In Case C3, since $\sum_{j=1}^n q_j \to 0$, it holds that $1-e^{-\sum_{j=1}^n q_j} \sim n^2/(2\theta)$ and that

$$
d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) \ge \frac{1}{32}\sum_{j=1}^n q_j^2 \ge \frac{n(n-1)(2n-1)}{192(\theta+n)^2}
$$

for enough large n. We thus have $d_{\mathrm{TV}}(n-K_n, P_{\lambda_C}) = \Theta(n^3/\theta^2)$. This completes the proof. $\square$

REMARK 3.1. From asymptotic properties of the Poisson distribution and Theorem 3.1, the result of (2.5) in Cases A and C can be derived.

In Theorem 3.1, we have considered Poisson variables with rigorous means $\mathsf{E}[K_n]$ and $n-\mathsf{E}[K_n]$. Next, let us discuss centerings by approximate means presented by Yamato (2013) and Tsukuda (2017b) from the viewpoint of Poisson approximation. Introduce three Poisson variables P_{μ_A}, P_{μ_a} and P_{μ_C}, where $\mu_A = \theta\log(1+n/\theta)$, $\mu_a = \theta(\log n - \psi(\theta))$ and $\mu_C = n - \theta\log(1+n/\theta)$. Next, Proposition 3.1 corresponds to (2.4). It follows from Lemma 3.1 presented after the proposition together with the triangle inequality.

PROPOSITION 3.1. (i) *In Case A, if* $(\log(n/\theta))/\theta \to \infty$ *then*

$$
d_{\mathrm{TV}}(K_n, P_{\mu_A}) = O\biggl(\frac{1}{\sqrt{\theta\log(n/\theta)}}\biggr),
$$

and if $(\log(n/\theta))/\theta = O(1)$ *then*

$$
d_{\mathrm{TV}}(K_n, P_{\mu_A}) = O\biggl(\frac{1}{\log(n/\theta)}\biggr).
$$

Moreover, in Case A, if $(\theta^3 \log(n/\theta))/n^2 = O(1)$ *then*

$$d_{\mathrm{TV}}(K_n, P_{\mu_a}) = O\left(\frac{1}{\log(n/\theta)}\right),$$

and if $(\theta^3 \log(n/\theta))/n^2 \to \infty$ *and* $\theta^3/(n^2 \log(n/\theta)) = O(1)$ *then*

$$d_{\mathrm{TV}}(K_n, P_{\mu_a}) = O\left(\frac{\theta^{3/2}}{n\sqrt{\log(n/\theta)}}\right).$$

(ii) *In Case C, it holds that*

$$d_{\mathrm{TV}}(n - K_n, P_{\mu_C}) = O\left(\frac{n}{\theta}\right).$$

LEMMA 3.1. (i) *In Case A, it holds that*

$$d_{\mathrm{TV}}(P_{\lambda_A}, P_{\mu_A}) = O\left(\frac{1}{\sqrt{\theta \log(n/\theta)}}\right) \tag{3.4}$$

and that

$$d_{\mathrm{TV}}(P_{\lambda_A}, P_{\mu_a}) = O\left(\frac{\theta^{3/2}}{n\sqrt{\log(n/\theta)}}\right). \tag{3.5}$$

(ii) *In Case C, it holds that*

$$d_{\mathrm{TV}}(P_{\lambda_C}, P_{\mu_C}) = O\left(\frac{1}{\sqrt{\theta}}\left(1 \wedge \frac{n}{\sqrt{\theta}}\right)\right). \tag{3.6}$$

PROOF. We will use (2.22). (i) First, we see (3.4). Since λ_A and μ_A tend to infinity in Case A, $|\sqrt{\lambda_A} - \sqrt{\mu_A}| \le |\lambda_A - \mu_A|$ for enough large n. Moreover, by using Proposition A.1, $|\sqrt{\lambda_A} - \sqrt{\mu_A}| = |\lambda_A - \mu_A|/(\sqrt{\lambda_A} + \sqrt{\mu_A})$ is

$$\frac{|\sum_{j=1}^n p_j - \theta \log(1 + n/\theta)|}{\sqrt{\sum_{j=1}^n p_j} + \sqrt{\theta \log(1 + n/\theta)}} = O\left(\frac{n/(n+\theta)}{\sqrt{\theta \log(1 + n/\theta)}}\right),$$

and hence (3.4). Next, we see (3.5). By using Propositions A.1 and A.2, $|\sqrt{\lambda_A} - \sqrt{\mu_a}| = |\lambda_A - \mu_a|/(\sqrt{\lambda_A} + \sqrt{\mu_a})$ is

$$\frac{|\sum_{j=1}^n p_j - \theta(\log n - \psi(\theta))|}{\sqrt{\sum_{j=1}^n p_j} + \sqrt{\theta(\log n - \psi(\theta))}} = O\left(\frac{\theta^2/n}{\sqrt{\theta \log(1 + n/\theta)}}\right),$$

and hence (3.5).

(ii) First, consider Case C1. Since λ_C and μ_C tend to infinity in Case C1, $|\sqrt{\lambda_C} - \sqrt{\mu_C}| \le |\lambda_C - \mu_C|$ for enough large n. By using Proposition A.1, $|\sqrt{\lambda_C} - \sqrt{\mu_C}| = |\lambda_C - \mu_C|/(\sqrt{\lambda_C} + \sqrt{\mu_C})$ is

$$\frac{|\sum_{j=1}^n q_j - (n - \theta \log(1 + n/\theta))|}{\sqrt{\sum_{j=1}^n q_j} + \sqrt{n - \theta \log(1 + n/\theta)}} = O\left(\frac{n/(\theta + n)}{\sqrt{n^2/\theta}}\right) = O\left(\frac{1}{\sqrt{\theta}}\right),$$

and hence (3.6) holds as $d_{\mathrm{TV}}(P_{\lambda_C}, P_{\mu_C}) = O(1/\sqrt{\theta})$.

Next, consider Case C2. The magnitude relationship of $|\sqrt{\lambda_C}-\sqrt{\mu_C}|$ and $|\lambda_C-\mu_C|$ is not determined, but they have the same bound $O(1/\sqrt{\theta})$ because $1/\sqrt{\theta}=\Theta(1/n)$. Hence (3.6) holds as $d_{\mathrm{TV}}(P_{\lambda_C},P_{\mu_C})=O(1/\sqrt{\theta})$.

Finally, consider Case C3. Since λ_C and μ_C tend to 0, $|\sqrt{\lambda_C}-\sqrt{\mu_C}|\ge|\lambda_C-\mu_C|$ for enough large n. By using Proposition A.1, it holds that

$$|\lambda_C-\mu_C|=\left|\sum_{j=1}^{n}q_j-\left(n-\theta\log\left(1+\frac{n}{\theta}\right)\right)\right|=O\left(\frac{n}{\theta}\right),$$

and hence (3.6) holds as $d_{\mathrm{TV}}(P_{\lambda_C},P_{\mu_C})=O(n/\theta)$. This completes the proof. $\square$

4. Independent process approximation of $\mathbf{C}_b^n$. In this section, we discuss Poisson approximations of $\mathbf{C}_b^n$. As it is stated in Section 1.2, we assume that θ does not decrease as n increase. Additionally, in several results, we will suppose that there exists a positive integer n_0 such that $\theta\ge 1$ for all $n\ge n_0$. The other case, $\theta<1$ for all n, is not discussed because we are interested in large θ. Moreover, in the proofs of Lemma 4.1 and Proposition 4.2, the case will be further divided.

4.1. *Asymptotic independence of* $\mathbf{C}_b^n$ (*Case A*). First, we see the asymptotic independence of small components $\mathbf{C}_b^n=(C_1^n,\dots,C_b^n)$ in this setting. We begin with Proposition 4.1 which shows that if $\theta^2/n\to 0$ then the previously established Poisson approximation results (2.8) is still valid even for large θ when b is fixed. Recall $\mathbf{a}_b$ defined in Section 1 and $\{Z_j\}_{j=1}^{\infty}$ and $\mathbf{Z}_b=(Z_1,\dots,Z_b)$ defined in Section 2.2.

PROPOSITION 4.1. *Suppose that there exists a positive integer n_0 such that $\theta\ge 1$ for all $n\ge n_0$. In Case A, if $\theta^2/n\to 0$ then $\mathsf{P}(\mathbf{C}_b^n=\mathbf{a}_b)\sim\mathsf{P}(\mathbf{Z}_b=\mathbf{a}_b)$ for any $\boldsymbol{a}_b$ with any fixed positive integer b.*

Before proving Proposition 4.1, let us introduce a useful expression of $\mathsf{P}(\mathbf{C}_b^n=\mathbf{a}_b)$. Define $T_{lm}=\sum_{j=l+1}^{m}jZ_j$ for $l=0,1,\dots,n-1$ and $m=l+1,\dots,n$, then it follows from (2.7) that

$$\mathsf{P}(\mathbf{C}_b^n=\mathbf{a}_b)=\mathsf{P}(\mathbf{Z}_b=\mathbf{a}_b)\frac{\mathsf{P}(T_{bn}=n-a)}{\mathsf{P}(T_{0n}=n)},\tag{4.1}$$

where $\mathbf{a}_b=(a_1,\dots,a_b)$ and $a=\sum_{j=1}^{b}ja_j$.

Moreover, let us prepare two lemmas which will be used several times.

LEMMA 4.1. *Let $f(x)=\exp(-\theta\sum_{j=1}^{b}x^j/j)$. For $\theta>1$, $k=1,\dots,\lceil\theta\rceil-1$ and any positive integers $a<n$ and b, it holds that*

$$\frac{1}{k!}\left|\frac{f^{(k)}(1)}{f(1)}\right|\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}}\le\left(\frac{b\theta^2}{n-a}\right)^k.\tag{4.2}$$

Moreover, suppose that there exists a positive integer n_0 such that $\theta \geq 1$ for all $n \geq n_0$, then in Case A, it holds that

$$\sum_{k=0}^{\lceil\theta\rceil-1} \frac{1}{k!}\left|\frac{f^{(k)}(1)}{f(1)}\right|\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} = 1 + \frac{b\theta(\theta-1)}{\theta+n-a} + O\left(\frac{b^2\theta^4}{n^2}\right) \tag{4.3}$$

if $a = o(n)$, $b = o(n/\theta^2)$ and $\theta^2/n \to 0$.

PROOF. Let $g(x)$ be $-\theta\sum_{j=1}^{b} x^j/j$. It holds that

$$g^{(i)}(x) = -\theta\sum_{j=i}^{b}\frac{j\cdots(j-i+1)}{j}x^{j-i} = -\theta\sum_{j=1}^{b-i+1}(j)_{i-1}x^{j-1}$$

for $1 \leq i \leq b$. Thus, for $1 \leq i \leq b$,

$$0 \geq g^{(i)}(1) = -\theta\sum_{j=1}^{b-i+1}(j)_{i-1} \geq -\theta(b-i+1)_{i-1}(b-i+1) \geq -\theta b^i.$$

For $i > b$, $g^{(i)}(1) = 0 \geq -\theta b^i$. The Faà di Bruno formula yields that

$$f^{(k)}(x) = \exp\bigl(g(x)\bigr)\sum_{j=1}^{k} B_{k,j}\bigl(\bigl(g^{(1)}(x), \dots, g^{(k-j+1)}(x)\bigr)\bigr),$$

where $B_{k,j}(\cdot)$ is the partial Bell polynomial, so

$$\frac{f^{(k)}(1)}{f(1)} = \sum_{j=1}^{k} B_{k,j}\bigl(\bigl(g^{(1)}(1), \dots, g^{(k-j+1)}(1)\bigr)\bigr)$$

for any $k = 1, 2, \dots$. By using the triangle inequality,

$$\begin{aligned}
&\bigl|B_{k,j}\bigl(\bigl(g^{(1)}(1), \dots, g^{(k-j+1)}(1)\bigr)\bigr)\bigr| \\
&\quad\leq n!\sum_{\{s.:\sum s_i=j,\sum is_i=k\}}\prod_{i=1}^{n}\left(\frac{|g^{(i)}(1)|}{i!}\right)^{s_i}\frac{1}{s_i!} \\
&\quad\leq \theta^j b^k \mathcal{S}(k,j),
\end{aligned}$$

where $\mathcal{S}(k,j)$ is the Stirling number of the second kind. The above two displays and the triangle inequality imply that

$$\begin{aligned}
\left|\frac{f^{(k)}(1)}{f(1)}\right| &\leq \sum_{j=1}^{k}\bigl|B_{k,j}\bigl(\bigl(g^{(1)}(1), \dots, g^{(k-j+1)}(1)\bigr)\bigr)\bigr| \\
&\leq b^k\sum_{j=1}^{k}\theta^j\mathcal{S}(k,j)
\end{aligned}$$

$$\le b^k \sum_{j=1}^{k} \theta^j \bar{s}(k, j) = b^k (\theta)_k.$$

For $k \le \lceil \theta \rceil - 1$, the Stirling formula yields that

$$(\theta)_k \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}} = \frac{\Gamma(\theta + k)\Gamma(\theta - k + n - a)\Gamma(\theta)}{\Gamma(\theta)\Gamma(\theta - k)\Gamma(\theta + n - a)} \le \frac{\theta^{2k}}{(n-a)^k},$$

where we have used $\Gamma(\theta + k)/\Gamma(\theta - k) = (\theta - k)(\theta^2 - (k-1)^2) \cdots (\theta^2 - 1^2)\theta \le \theta^{2k}$ and $\Gamma(\theta - k + n - a)/\Gamma(\theta + n - a) = 1/((\theta - 1 + n - a) \cdots (\theta - k + n - a)) \le 1/(n-a)^k$. We thus have

$$\left| \frac{f^{(k)}(1)}{f(1)} \right| \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}} \le \left(\frac{b\theta^2}{n-a} \right)^k$$

for $k \le \lceil \theta \rceil - 1$, which is (4.2).

Next, we prove (4.3). If $\theta \le 1$ for all n, the result is obvious because the left-hand side of (4.3) is 1. Otherwise, there exists a positive integer $\tilde{n}_0$ such that $\theta > 1$ for all $n \ge \tilde{n}_0$. Let n be an positive integer such that $n \ge \tilde{n}_0$, then the desired result follows from

$$\frac{1}{1!} \frac{f^{(1)}(1)}{f(1)} \frac{(\theta - 1)_{n-a}}{(\theta)_{n-a}} = b\theta \frac{\theta - 1}{\theta + n - a}$$

and from

$$\sum_{k=2}^{\lceil \theta \rceil - 1} \frac{1}{k!} \left| \frac{f^{(k)}(1)}{f(1)} \right| \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}} \le \sum_{k=2}^{\lceil \theta \rceil - 1} \left(\frac{b\theta^2}{n-a} \right)^k = O\left(\left(\frac{b\theta^2}{n} \right)^2 \right).$$

This completes the proof. $\square$

LEMMA 4.2. *Let $f(x) = \exp(-\theta \sum_{j=1}^{b} x^j / j)$ and let*

$$h(x) = (1-x)^{-\theta} \sum_{k=\lceil \theta \rceil}^{\infty} \frac{f^{(k)}(1)}{k!} (-1)^k (1-x)^k.$$

Then, for any positive integers $a < n$ and b, it holds that

$$\left| [x^{n-a}] h(x) \right| \le \frac{1}{r_1^n} \left(\frac{b e^{(r_2 - 1)^b}}{r_2} \right)^\theta \frac{1}{r_2 - 1 - r_1} \{ r_1^a (1 + r_1) r_2 \},$$

where $r_1 = 1 + c_1 r$, $r_2 = 2 + c_2 r$, $1 < c_1 < c_2$, and r is an arbitrary positive constant.

PROOF. Consider a complex variable $\mathbf{z} \in \mathcal{C}$. Since $h(\mathbf{z})$ and $f(\mathbf{z})$ are analytic in $\mathcal{C}$, by using the Cauchy inequality for coefficients, we have

$$\begin{aligned}
\sup_{|\mathbf{z}|=r_1} |h(\mathbf{z})| &\le \sup_{|\mathbf{z}|=r_1} \sum_{k=\lceil\theta\rceil}^{\infty} \left|\frac{f^{(k)}(1)}{k!}\right| |(1-\mathbf{z})^{k-\theta}| \\
&\le \sum_{k=\lceil\theta\rceil}^{\infty} \left|\frac{f^{(k)}(1)}{k!}\right| \sup_{|\mathbf{z}|=r_1} |(1-\mathbf{z})^{k-\theta}| \\
&\le \sum_{k=\lceil\theta\rceil}^{\infty} \frac{\sup_{|1-\mathbf{z}|=r_2} |f(\mathbf{z})|}{r_2^k} (1+r_1)^{k-\theta} \\
&= \frac{\sup_{|1-\mathbf{z}|=r_2} |f(\mathbf{z})|}{(1+r_1)^{\theta}} \sum_{k=\lceil\theta\rceil}^{\infty} \left(\frac{1+r_1}{r_2}\right)^k \\
&= \frac{\sup_{|1-\mathbf{z}|=r_2} |f(\mathbf{z})|}{(1+r_1)^{\theta}} \frac{\{(1+r_1)/r_2\}^{\lceil\theta\rceil}}{1-(1+r_1)/r_2}.
\end{aligned}$$

The right-hand side is

$$\frac{\sup_{|1-\mathbf{z}|=r_2} |f(\mathbf{z})|}{(1+r_1)^{\theta-\lceil\theta\rceil} r_2^{\lceil\theta\rceil} \{1-(1+r_1)/r_2\}} \le \left(\frac{be^{(r_2-1)^b}}{r_2}\right)^{\theta} \frac{(1+r_1)r_2}{r_2-1-r_1}, \tag{4.4}$$

because

$$\begin{aligned}
\sup_{|1-\mathbf{z}|=r_2} |f(\mathbf{z})| &= \sup_{|1-\mathbf{z}|=r_2} \left|\exp\left(-\theta \sum_{j=1}^{b} \frac{\mathbf{z}^j}{j}\right)\right| \\
&\le \exp\left(\theta \sum_{j=1}^{b} \frac{(r_2-1)^j}{j}\right) \\
&\le \exp(\theta\{\log b + (r_2-1)^b\}) \\
&= \left(be^{(r_2-1)^b}\right)^{\theta},
\end{aligned}$$

where we have used Lemma A.3 for the second inequality. Hence, it follows from the Cauchy inequality again that

$$\begin{aligned}
|[x^{n-a}]h(x)| &\le \frac{\sup_{|\mathbf{z}|=r_1} |h(\mathbf{z})|}{r_1^{n-a}} \\
&\le \frac{1}{r_1^n} \left(\frac{be^{(r_2-1)^b}}{r_2}\right)^{\theta} \frac{1}{r_2-1-r_1} \{r_1^a(1+r_1)r_2\}.
\end{aligned}$$

This completes the proof. $\square$

PROOF OF PROPOSITION 4.1. From (4.1), in order to prove the desired result, it suffices to show that

$$\frac{\mathsf{P}(T_{bn} = n - a)}{\mathsf{P}(T_{0n} = n)} \to 1.$$

We first calculate $g_{n-a} = \exp(\theta \sum_{j=b+1}^{n} 1/j)\mathsf{P}(T_{bn} = n - a)$. Letting $f(x) = \exp(-\theta \sum_{j=1}^{b} x^j/j)$, we have

$$g_{n-a} = [x^{n-a}](1-x)^{-\theta} f(x),$$

see equation (5) of Arratia, Barbour and Tavaré (1992).

Let n be a positive integer such that $n \geq n_0$. It holds that

$$\begin{aligned}
&[x^{n-a}](1-x)^{-\theta} f(x) \\
&\quad = [x^{n-a}](1-x)^{-\theta} \left\{ \sum_{k=0}^{\lceil\theta\rceil-1} \frac{f^{(k)}(1)}{k!}(-1)^k(1-x)^k \right. \\
&\qquad \left. + \sum_{k=\lceil\theta\rceil}^{\infty} \frac{f^{(k)}(1)}{k!}(-1)^k(1-x)^k \right\} \\
&\quad = [x^{n-a}](1-x)^{-\theta} \left\{ \sum_{k=0}^{\lceil\theta\rceil-1} \frac{f^{(k)}(1)}{k!}(-1)^k(1-x)^k \right\} + [x^{n-a}]h(x),
\end{aligned} \tag{4.5}$$

where

$$h(x) = (1-x)^{-\theta} \sum_{k=\lceil\theta\rceil}^{\infty} \frac{f^{(k)}(1)}{k!}(-1)^k(1-x)^k.$$

Since the right-hand side of (4.5) is

$$\frac{f(1)(\theta)_{n-a}}{(n-a)!} \left\{ 1 + \sum_{k=1}^{\lceil\theta\rceil-1} \frac{(-1)^k}{k!} \frac{f^{(k)}(1)}{f(1)} \frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} \right\} + [x^{n-a}]h(x), \tag{4.6}$$

the first term and the second term is evaluated in Lemmas 4.1 and 4.2, respectively. From Lemma 4.1, the elements in the bracket of the first term is $1 + O(\theta^2/n)$. Next, we see $[x^{n-a}]h(x)$. It follows from Lemma 4.2 that

$$\left|[x^{n-a}]h(x)\right| \leq \frac{1}{r_1^n} \left(\frac{b e^{(r_2-1)^b}}{r_2} \right)^{\theta} \frac{1}{r_2 - 1 - r_1} \{r_1^a (1 + r_1) r_2\},$$

where $r_1 = 1 + c_1 r$ and $r_2 = 2 + c_2 r$ with constants c_1, c_2 such that $1 < c_1 < c_2$. By letting r be a positive constant, the right-hand side is $o(1/n^k)$ for any positive k since b is fixed and since $\theta^2/n \to 0$. Thence $[x^{n-a}]h(x) = o(1/n)$.

Now we have

$$g_{n-a} = f(1)\frac{(\theta)_{n-a}}{(n-a)!}\left(1 + O\left(\frac{\theta^2}{n}\right)\right) + o\left(\frac{1}{n}\right) \tag{4.7}$$

and, as a result,

$$\begin{aligned}
&\mathsf{P}(T_{bn} = n-a) \\
&\quad = \exp\left(-\theta\sum_{j=1}^{n}\frac{1}{j}\right)\frac{(\theta)_{n-a}}{(n-a)!}\left(1 + O\left(\frac{\theta^2}{n}\right)\right) + \exp\left(-\theta\sum_{j=b+1}^{n}\frac{1}{j}\right)o\left(\frac{1}{n}\right).
\end{aligned}$$

On the other hand,

$$\mathsf{P}(T_{0n} = n) = \exp\left(-\theta\sum_{j=1}^{n}\frac{1}{j}\right)[x^n](1-x)^{-\theta} = \exp\left(-\theta\sum_{j=1}^{n}\frac{1}{j}\right)\frac{(\theta)_n}{n!}. \tag{4.8}$$

If $\theta \to c < \infty$, $(\theta)_n/n! \sim n^{\theta-1}/\Gamma(\theta)$ and so

$$\exp\left(\theta\sum_{j=1}^{b}\frac{1}{j}\right)\frac{n!}{(\theta)_n} \le \exp(\theta\log b)\frac{e^{\theta}n!}{(\theta)_n} \sim n\Gamma(\theta)\left(\frac{be}{n}\right)^{\theta} = o(n). \tag{4.9}$$

If $\theta \to \infty$, Lemma A.2 and the Stirling formula yield that

$$\frac{(\theta)_n}{n!} \sim \frac{n^{\theta-1}}{\Gamma(\theta)} \sim \frac{n^{\theta-1}\theta^{1/2}e^{\theta}}{\sqrt{2\pi}\theta^{\theta}},$$

and hence

$$\begin{aligned}
\exp\left(\theta\sum_{j=1}^{b}\frac{1}{j}\right)\frac{n!}{(\theta)_n} &\sim \left(\frac{\theta e^{(\sum_{j=1}^{b}\frac{1}{j}-1)}}{n}\right)^{\theta}\frac{\sqrt{2\pi}n}{\theta^{1/2}} \\
&\le \sqrt{2\pi}\frac{n}{\theta^{1/2}}\left(\frac{\theta b}{n}\right)^{\theta} = o(n).
\end{aligned} \tag{4.10}$$

From what has already been proved, we obtain

$$\begin{aligned}
\frac{\mathsf{P}(T_{bn} = n-a)}{\mathsf{P}(T_{0n} = n)} &= \frac{n!}{(\theta)_n}\frac{(\theta)_{n-a}}{(n-a)!}\left(1 + O\left(\frac{\theta^2}{n}\right)\right) + o(n)o\left(\frac{1}{n}\right) \\
&\sim \frac{\theta^{\theta}}{n^{\theta-1}\theta^{1/2}e^{\theta}}\frac{(n-a)^{\theta-1}\theta^{1/2}e^{\theta}}{\theta^{\theta}} \to 1.
\end{aligned}$$

This completes the proof. □

Let us provide some remarks on Proposition 4.1.

REMARK 4.1. Proposition 4.1 indicates that when $\theta^2/n \to 0$ the components of $(C_1^n, \dots, C_b^n)$ are asymptotically independent, and C_j^n asymptotically follows the Poisson distribution with mean θ/j for $j = 1, \dots, b$. As a consequence, for any fixed b, if $\theta \to c < \infty$ then

$$(C_1^n, \dots, C_b^n) \Rightarrow (Z_1^*, \dots, Z_b^*),$$

where $\{Z_j^*\}_{j=1}^{\infty}$ is a sequence of independent Poisson variables with $\mathsf{E}[Z_j^*] = c/j$ for $j = 1, 2, \dots$, and if $\theta \to \infty$ then

$$\frac{1}{\sqrt{\theta}}\left(C_1^n - \theta, \sqrt{2}\left(C_2^n - \frac{\theta}{2}\right), \dots, \sqrt{b}\left(C_b^n - \frac{\theta}{b}\right)\right) \Rightarrow N_b(0, I),$$

where $N_b(0, I)$ is a b-dimensional standard normal variable with independent coordinates.

REMARK 4.2. Proposition 4.3 presented in the next subsection is stronger than Proposition 4.1, but the proof is included. That is because some evaluations are different from the proof of Theorem 1 of Arratia, Barbour and Tavaré (1992) who used the Darboux lemma (see Theorem of Knuth and Wilf (1989)) and because Lemmas 4.1 and 4.2 will appear also in the proof of Theorem 4.1.

In Proposition 4.1, $\theta^2/n \to 0$ is assumed. Our second result in this subsection, Proposition 4.2, shows that this assumption is necessary for the approximation of $\{C_j^n\}_{j=1}^b$ by $\{Z_j\}_{j=1}^b$.

PROPOSITION 4.2. *Suppose that there exists a positive integer n_0 such that $\theta \geq 1$ for all $n \geq n_0$. In Case A, $\mathsf{P}(\mathbf{C}_b^n = \mathbf{a}_b) \sim \mathsf{P}(\mathbf{Z}_b = \mathbf{a}_b)$ for any $\boldsymbol{a}_b$ with any fixed positive integer b only if $\theta^2/n \to 0$.*

PROOF. To prove the assertion, we see the case that $b = 1$. Let $f(x) = \exp(-\theta x)$, then we have $f^{(k)}(x) = (-\theta)^k f(x)$. From (4.6), $g_{n-a} = [x^{n-a}](1 - x)^{-\theta} f(x)$ equals

$$\frac{f(1)(\theta)_{n-a}}{(n-a)!} \sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!} \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}} + [x^{n-a}]h(x).$$

Since

$$\begin{aligned}
\frac{\mathsf{P}(T_{1n} = n - a)}{\mathsf{P}(T_{0n} = n)} &= \frac{\exp(-\theta \sum_{j=2}^n 1/j) g_{n-a}}{\exp(-\theta \sum_{j=1}^n 1/j)(\theta)_n/n!} \\
&= \frac{n!}{(\theta)_n} \frac{(\theta)_{n-a}}{(n-a)!} \sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!} \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}} + o(1) \\
&\sim \sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!} \frac{(\theta - k)_{n-a}}{(\theta)_{n-a}}
\end{aligned}$$

from the proof of Proposition 4.1, it is enough to show that

$$\sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!}\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} \to 1$$

only if $\theta^2/n \to 0$.

Since θ is assumed not to decrease as n increases, we study the following three cases: (i) there exists a positive integer n_1 such that $\theta \geq 2$ for all $n \geq n_1$; (ii) $\theta < 2$ for all n and there exists a positive integer n_2 such that $\theta > 1$ for all $n \geq n_2$; (iii) $\theta \leq 1$ for all n. First, consider (i). Let n be a positive integer such that $n \geq n_1$. Then it holds that

$$\begin{aligned}\sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!}\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} &\geq \sum_{k=0}^{\lceil\theta\rceil-2} \frac{\theta^k}{k!}\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} \\ &\geq \sum_{k=0}^{\lceil\theta\rceil-2} \frac{\theta^k}{k!}\frac{(\lceil\theta\rceil-1-k)_{n-a}}{(\lceil\theta\rceil-1)_{n-a}},\end{aligned}$$

where we have used Lemma A.4 for the second inequality. The right-hand side equals

$$\begin{aligned}&\sum_{k=0}^{\lceil\theta\rceil-2} \frac{\theta^k}{k!}\frac{(\lceil\theta\rceil-2)!}{(\lceil\theta\rceil-k-2)!}\frac{(\lceil\theta\rceil-2-k+n-a)!}{(\lceil\theta\rceil-2+n-a)!} \\ &\quad= \sum_{k=0}^{\lceil\theta\rceil-2} \binom{\lceil\theta\rceil-2}{k}\theta^k \frac{1}{(\lceil\theta\rceil-2-k+1+n-a)\cdots(\lceil\theta\rceil-2+n-a)} \\ &\quad= \sum_{k=0}^{\lceil\theta\rceil-2} \binom{\lceil\theta\rceil-2}{k}\left(\frac{\theta}{n}\right)^k \frac{1}{(1+\frac{\lceil\theta\rceil-2-k+1-a}{n})\cdots(1+\frac{\lceil\theta\rceil-2-a}{n})} \\ &\quad\geq \sum_{k=0}^{\lceil\theta\rceil-2} \binom{\lceil\theta\rceil-2}{k}\left\{\frac{\theta}{n(1+\frac{\lceil\theta\rceil-2-a}{n})}\right\}^k.\end{aligned}$$

From the binomial theorem, the right-hand side is equal to

$$\left\{1+\frac{\theta}{n(1+\frac{\lceil\theta\rceil-2-a}{n})}\right\}^{\lceil\theta\rceil-2} = \left[\left\{1+\frac{\theta}{n(1+\frac{\lceil\theta\rceil-2-a}{n})}\right\}^{n/\theta}\right]^{\theta(\lceil\theta\rceil-2)/n}.$$

The above display is not less than 1 and converges to 1 only if $\theta^2/n \to 0$. Second, consider (ii). Let n be a positive integer such that $n \geq n_2$. Then it holds that

$$\sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!}\frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} = 1 + \frac{\theta(\theta-1)}{\theta+n-a-1},$$

which converges to 1 only if $\theta^2/n \to 0$. Finally, consider (iii). Let n be a positive integer such that $\theta = 1$. Then it holds that

$$\sum_{k=0}^{\lceil\theta\rceil-1} \frac{\theta^k}{k!} \frac{(\theta-k)_{n-a}}{(\theta)_{n-a}} = 1.$$

This completes the proof. $\square$

Thence, we have the following corollary to Propositions 4.1 and 4.2.

COROLLARY 4.1. *Suppose that there exists a positive integer n_0 such that $\theta \geq 1$ for all $n \geq n_0$. In Case A, $\mathsf{P}(\mathbf{C}_b^n = \mathbf{a}_b) \sim \mathsf{P}(\mathbf{Z}_b = \mathbf{a}_b)$ for any $\boldsymbol{a}_b$ with any fixed positive integer b if, and only if, $\theta^2/n \to 0$.*

4.2. *Total variation asymptotics* (*Case A*). Subsequently, let us derive the result corresponding to (2.14) following a similar program to Arratia, Stark and Tavaré (1995). Our result in this subsection, Theorem 4.1, shows that if $\theta^2/n \to 0$ then the total variation asymptotics obtained by Arratia, Stark and Tavaré (1995) is still valid.

THEOREM 4.1. *Suppose that there exists a positive integer n_0 such that $\theta \geq 1$ for all $n \geq n_0$. In Case A, if $\theta^2/n \to 0$, then*

$$d_b(n) = \frac{\theta-1}{2n}\mathsf{E}\big[|T_{0b} - \theta b|\big] + o\left(\frac{b\theta^2}{n}\right) \tag{4.11}$$

for

$$b = o\left(\frac{n}{\theta^2 \log n}\right). \tag{4.12}$$

In addition, when $\theta \to \infty$, it holds that $d_b(n) = o(b\theta^2/n)$.

To prove Theorem 4.1, let us prepare some notions and lemmas. It follows from (2.7) that

$$d_b(n) = \sum_{a=0}^{\infty} \mathsf{P}(T_{0b} = a)\left(1 - \frac{\mathsf{P}(T_{bn} = n-a)}{\mathsf{P}(T_{0n} = n)}\right)^+;$$

see (50) of Arratia, Stark and Tavaré (1995). First, in Lemma 4.3 via the evaluation of the large deviation probability for T_{0b}, we see that $d_b(n)$ can be approximated by

$$\sum_{a=0}^{\lfloor J_n \rfloor} \mathsf{P}(T_{0b} = a)\left(1 - \frac{\mathsf{P}(T_{bn} = n-a)}{\mathsf{P}(T_{0n} = n)}\right)^+$$

with

$$J_n = \min\bigl(b\theta\log n, b^{2/3}(\theta n)^{1/3}\bigr).$$

From the definition, if $1 \le b \le n\theta^{-2}(\log n)^{-3}$ then $J_n = b\theta\log n$ and otherwise $J_n = b^{2/3}(\theta n)^{1/3} = b(\theta n/b)^{1/3}$. In contrast to Arratia, Stark and Tavaré (1995), J_n includes θ since we consider $\theta \to \infty$, but a similar treatment performs well.

LEMMA 4.3. *In Case A, with $b = o(n/\theta^2)$, it holds that*

$$\sum_{a>J_n} \mathsf{P}(T_{0b} = a)\biggl(1 - \frac{\mathsf{P}(T_{bn} = n-a)}{\mathsf{P}(T_{0n} = n)}\biggr)^+ \le \mathsf{P}(T_{0b} > J_n) = o\biggl(\biggl(\frac{b}{n}\biggr)^k\biggr)$$

for any positive k.

PROOF. The first inequality is obvious, so we see the latter one. From Lemma 8 of Arratia, Stark and Tavaré (1995), for any $b \ge 1$, $w > 0$, it holds that

$$\log \mathsf{P}(T_{0b} \ge bw) \le \log(\theta e/w)^w. \tag{4.13}$$

If $1 \le b \le n\theta^{-2}(\log n)^{-3}$ then, by putting $w = \theta\log n$, the right-hand side of (4.13) is

$$(\theta\log n)(1 - \log\log n) \sim -\theta(\log\log n)\log n$$

which tends to minus infinity faster than $-k\log n$ for any positive k. If $b \ge n\theta^{-2}(\log n)^{-3}$ then, by putting $w = (\theta n/b)^{1/3}$, the right-hand side of (4.13) is

$$\biggl(\frac{\theta n}{b}\biggr)^{1/3}\biggl(1 - \frac{1}{3}\log\biggl(\frac{n}{b}\biggr)\biggr) \sim -\frac{1}{3}\biggl(\frac{\theta n}{b}\biggr)^{1/3}\log\biggl(\frac{n}{b}\biggr)$$

which tends to minus infinity faster than $-k\log(n/b)$ for any positive k. This completes the proof. □

The next lemma shows that $(|1-\theta|/n)\mathsf{E}[(T_{0b} - \theta b)^+ 1\{T_{0b} \le J_n\}]$ is approximately $(|1-\theta|/n)\mathsf{E}[(T_{0b} - \theta b)^+]$.

LEMMA 4.4. *In Case A, if $\theta^2/n \to 0$ then it holds that*

$$\frac{|1-\theta|}{n}\mathsf{E}\bigl[|T_{0b} - \theta b| 1\{T_{0b} > J_n\}\bigr] = o\biggl(\biggl(\frac{b}{n}\biggr)^k\biggr)$$

for $b = o(n/\theta^2)$ and for any positive k.

PROOF. From the Schwartz inequality, it follows that

$$
\begin{aligned}
&\frac{|1-\theta|}{n}\mathsf{E}\bigl[|T_{0b}-b\theta|1\{T_{0b}>J_n\}\bigr]\\
&\quad\le\left(\frac{|1-\theta|^2}{n^2}\mathsf{E}\bigl[(T_{0b}-\theta b)^2\bigr]\mathsf{P}(T_{0b}>J_n)\right)^{1/2}\\
&\quad\le\left(\frac{|1-\theta|^2}{n^2}\theta b^2\mathsf{P}(T_{0b}>J_n)\right)^{1/2}\\
&\quad=\left[\left\{\frac{|1-\theta|\theta^{1/2}b}{n}\right\}^2\mathsf{P}(T_{0b}>J_n)\right]^{1/2}\\
&\quad\le\left[\left\{\frac{(1+\theta)^2 b}{n}\right\}^2\mathsf{P}(T_{0b}>J_n)\right]^{1/2},
\end{aligned}
$$

where we have used $\mathsf{E}[(T_{0b}-\mathsf{E}[T_{0b}])^2]=\mathrm{var}(T_{0b})=\sum_{j=1}^{b} j^2(\theta/j)=\theta\sum_{j=1}^{b} j\le\theta b^2$ for the second inequality. Lemma 4.3 yields that $\mathsf{P}(T_{0b}>J_n)=o((b/n)^{2k})$ for any positive k. This completes the proof. $\square$

PROOF OF THEOREM 4.1. Let n be a positive integer such that $n\ge n_0$. Since it follows from Lemma 4.3 that

$$
d_b(n)=\sum_{a=0}^{\lfloor J_n\rfloor}\mathsf{P}(T_{0b}=a)\left(1-\frac{\mathsf{P}(T_{bn}=n-a)}{\mathsf{P}(T_{0n}=n)}\right)^{+}+o\left(\left(\frac{b}{n}\right)^{k}\right)
$$

for any positive k, we see the first term.

As same as the proof of Proposition 4.1, let $g_{n-a}=\exp(\theta\sum_{j=b+1}^{n}1/j)\mathsf{P}(T_{bn}=n-a)$. For $a\le\lfloor J_n\rfloor$, (4.6) and (4.3) yield

$$
\begin{aligned}
g_{n-a}&=\frac{f(1)(\theta)_{n-a}}{(n-a)!}\left\{1+\frac{b\theta(\theta-1)}{\theta+n-a}+O\left(\frac{b^2\theta^4}{n^2}\right)\right\}+[x^{n-a}]h(x)\\
&=\frac{f(1)(\theta)_{n-a}}{(n-a)!}\left\{1+\frac{b\theta(\theta-1)}{n}\left(1+O\left(\frac{a+\theta}{n}\right)\right)+O\left(\frac{b^2\theta^4}{n^2}\right)\right\}\\
&\quad+\bigl[x^{n-a}\bigr]h(x).
\end{aligned}
$$

Here, we should evaluate the last term in the right-hand side for a and b growing with n.

If b does not diverge, as it is seen in the proof of Proposition 4.1, $[x^{n-a}]h(x)=o(1/n^2)$ since $a/n\le J_n/n\to 0$. Thence, we consider the case that $b\to\infty$. Using Lemma 4.2 with $r=1/b$, we have

$$
\bigl|[x^{n-a}]h(x)\bigr|\le\frac{b}{(1+c_1/b)^{n-a}}\left(\frac{be^{(1+c_2/b)^b}}{2+c_2/b}\right)^{\theta}\frac{(2+c_1/b)(2+c_2/b)}{(c_2-c_1)}.
$$

Since $(1+c_1/b)^{n-a} \sim e^{(n-a)c_1/b}$ and since $(1+c_2/b)^b \sim e^{c_2}$, the right-hand side is asymptotically equal to

$$\begin{aligned}
&\frac{b^{\theta+1}A_1^{\theta}}{\exp((n-a)c_1/b)}A_2\\
&\quad = A_2\exp\biggl((\theta+1)\log b + \theta\log A_1 - \frac{(n-a)c_1}{b}\biggr)\\
&\quad = A_2\exp\biggl(\frac{n}{b}\biggl\{\frac{b(\theta+1)\log n}{n}\frac{\log b}{\log n} + \frac{b\theta}{n}\log A_1 - \biggl(1-\frac{a}{n}\biggr)c_1\biggr\}\biggr),
\end{aligned}$$

where $A_1 = (\exp(e^{c_2}))/2$ and $A_2 = 4/(c_2-c_1)$. From (4.12), the right-hand side is

$$\begin{aligned}
A_2\exp\biggl(-\frac{n}{b}(c_1+o(1))\biggr) &= A_2\exp\biggl(-\frac{n(c_1+o(1))}{b\log n}\log n\biggr)\\
&= A_2 n^{-\frac{n(c_1+o(1))}{b\log n}},
\end{aligned} \tag{4.14}$$

where we have used $a/n \le J_n/n \to 0$. The right-hand side is $o(1/n^k)$ for any positive constant k. After all, we have $|[x^{n-a}]h(x)| = o(b/n^2)$ even when $b\to\infty$.

Now g_{n-a} is expanded as

$$\frac{f(1)(\theta)_{n-a}}{(n-a)!}\biggl\{1+\frac{b\theta(\theta-1)}{n}\biggl(1+O\biggl(\frac{a+\theta}{n}\biggr)\biggr)+O\biggl(\frac{b^2\theta^4}{n^2}\biggr)\biggr\}+o\biggl(\frac{b}{n^2}\biggr).$$

This expansion, $f(1) = \exp(-\sum_{j=1}^{b} 1/j)$,

$$\mathsf{P}(T_{bn} = n-a) = \exp\biggl(-\theta\sum_{j=b+1}^{n}\frac{1}{j}\biggr)g_{n-a}$$

and (4.8)–(4.10) yield that

$$\frac{\mathsf{P}(T_{bn}=n-a)}{\mathsf{P}(T_{0n}=n)} = \frac{n!(\theta)_{n-a}}{(n-a)!(\theta)_n}\biggl\{1+\frac{b\theta(\theta-1)}{n}+o\biggl(\frac{b\theta^2}{n}\biggr)\biggr\}+o(n)o\biggl(\frac{b}{n^2}\biggr).$$

Since $a\theta/n \le aJ_n/n \to 0$ and $a^2/(nb) \le J_n^2/(nb) \to 0$ which follow from $\theta/J_n \to 0$ and (4.12), the binomial expansion and Lemma A.2 yield that

$$\begin{aligned}
\frac{n!(\theta)_{n-a}}{(n-a)!(\theta)_n} &= \frac{(\theta)_{n-a}/(n-a)!}{(\theta)_n/n!}\\
&= \frac{(n-a)^{\theta-1}\{1+\frac{\theta(\theta-1)}{2(n-a)}+O(\frac{\theta^4}{n^2})\}/\Gamma(\theta)}{n^{\theta-1}\{1+\frac{\theta(\theta-1)}{2n}+O(\frac{\theta^4}{n^2})\}/\Gamma(\theta)}\\
&= \biggl(1-\frac{a}{n}\biggr)^{\theta-1}\frac{\{1+\frac{\theta(\theta-1)}{2n}+O(\frac{\theta^2(a+\theta^2)}{n^2})\}}{\{1+\frac{\theta(\theta-1)}{2n}+O(\frac{\theta^4}{n^2})\}}
\end{aligned}$$

$$= \left\{1 - \frac{a(\theta-1)}{n} + O\left(\frac{a^2\theta^2}{n^2}\right)\right\}\left(1 + O\left(\frac{\theta^2(a+\theta^2))}{n^2}\right)\right)$$
$$= 1 - \frac{(\theta-1)a}{n} + O\left(\frac{\theta^2(a^2+\theta^2)}{n^2}\right)$$
$$= 1 - \frac{(\theta-1)a}{n} + o\left(\frac{b\theta^2}{n}\right).$$

Therefore, it holds that

$$\frac{\mathsf{P}(T_{bn} = n-a)}{\mathsf{P}(T_{0n} = n)}$$
$$= \left\{1 - \frac{(\theta-1)a}{n} + o\left(\frac{b\theta^2}{n}\right)\right\}\left\{1 + \frac{b\theta(\theta-1)}{n} + o\left(\frac{b\theta^2}{n}\right)\right\} + o\left(\frac{b}{n}\right)$$
$$= 1 - \left\{\frac{(\theta-1)a}{n} - \frac{b\theta(\theta-1)}{n}\right\} + o\left(\frac{b\theta^2}{n}\right).$$

From what has already been proved, it holds that

$$d_b(n) = \sum_{a=0}^{\lfloor J_n \rfloor} \mathsf{P}(T_{0b} = a)\left(\frac{(\theta-1)a}{n} - \frac{b\theta(\theta-1)}{n}\right)^+ + o\left(\frac{b\theta^2}{n}\right)$$
$$= \frac{1}{n}\sum_{a=0}^{\infty} \mathsf{P}(T_{0b} = a)\big((\theta-1)(a-b\theta)\big)^+ 1\{a \le J_n\} + o\left(\frac{b\theta^2}{n}\right)$$
$$= \frac{1}{n}\mathsf{E}\big[\big((\theta-1)(T_{0b}-b\theta)\big)^+ 1\{T_{0b} \le J_n\}\big] + o\left(\frac{b\theta^2}{n}\right)$$
$$= \frac{1}{n}\mathsf{E}\big[\big((\theta-1)(T_{0b}-b\theta)\big)^+\big] + o\left(\frac{b\theta^2}{n}\right)$$
$$= \frac{\theta-1}{2n}\mathsf{E}\big[|T_{0b}-b\theta|\big] + o\left(\frac{b\theta^2}{n}\right),$$

where we have used Lemma 4.4 in the fourth equality and the relation $\mathsf{E}[(T_{0b} - b\theta)^+] = \mathsf{E}[|T_{0b} - b\theta|]/2$, which follows from $\mathsf{E}[T_{0b} - b\theta] = 0$, in the fifth equality.

Finally, consider the case that $\theta \to \infty$. It follows from the Jensen inequality that

$$\mathsf{E}\big[|T_{0b} - b\theta|\big] \le \sqrt{\mathsf{E}\big[|T_{0b} - b\theta|^2\big]} = O\big(\theta^{1/2}b\big),$$

which implies $d_b(n) = o(b\theta^2/n)$. This completes the proof. $\square$

4.3. *Poisson process approximations via the Feller coupling* (*Case A*). Next, result is the Poisson process approximation via the Feller coupling (see Section 2.2) to obtain the leading term of $d_b^W(n)$. The following result follows directly from (2.13).

PROPOSITION 4.3. *Suppose that there exists a positive integer n_0 such that $\theta \geq 1$ for all $n \geq n_0$ and $\theta^2/n \to 0$. In Case A, $d_b^W(n) \to 0$ if, and only if, $b = o(n/\theta^2)$. In addition, when $\theta \to \infty$, it holds that $d_b^W(n) \sim b\theta^2/n$.*

REMARK 4.3. When $\theta \to \infty$, Theorem 4.1 and Proposition 4.3 lead

$$d_b(n) = o\left(\frac{b\theta^2}{n}\right), \qquad d_b^W(n) = \Theta\left(\frac{b\theta^2}{n}\right)$$

for $b = o(n/(\theta^2 \log n))$, which shows that the asymptotic decay rates of $d_b(n)$ and $d_b^W(n)$ are different.

As an application of Proposition 4.3, let us show the asymptotic property of the kth shortest cycle lengths S_n^k defined in (2.15) in a Ewens partition in large θ setup.

COROLLARY 4.2. *Let r be a positive integer such that $r = o(n/\theta^2)$ and let $\delta_r = \sum_{j=1}^r \theta/j$. Under the assumption of Proposition 4.3,*

$$\mathsf{P}(S_n^k \leq r) \sim \sum_{x=0}^{k-1} e^{-\delta_r} \frac{\delta_r^x}{x!}.$$

PROOF. Proposition 4.3 yields that

$$\mathsf{P}(S_n^k \leq r) = \mathsf{P}\left(\sum_{j=1}^r C_j^n < k\right) \sim \mathsf{P}\left(\sum_{j=1}^r Z_j < k\right) = \sum_{x=0}^{k-1} e^{-\delta_r} \frac{\delta_r^x}{x!}.$$

This completes the proof. $\square$

REMARK 4.4. Corollary 4.2 yields that, under the assumption of Proposition 4.3, $\mathsf{P}(S_n^1 = 1) \sim e^{-\theta}$, so if $\theta \to \infty$ then $\mathsf{P}(S_n^1 = 1) \to 1$ and if $\theta \to c < \infty$ then $\mathsf{P}(S_n^1 = 1) \to e^{-c} < 1$. Note that when the *Pitman sampling formula* which is a generalization of (1.1) and is defined by (17) in Pitman (1995) is considered, the shortest cycle length converges to 1 in probability except the cases of the symmetric Dirichlet–multinomial distribution and the Ewens sampling formula (see Mano (2017)).

Moreover, we obtain a bound for $d_n^W(n)$, which gives an extension of (2.12) to large θ setup. Its applications to functional central limit theorems will be presented in Section 5.

PROPOSITION 4.4. *In Case A, it holds that $d_n^W(n) = O(\theta \log(1+\theta))$.*

PROOF. By using the triangle inequality and (2.11), it holds that

$$
\begin{aligned}
d_n^W(n) &\le \sum_{j=1}^{n} \mathsf{E}[|C_j^n - C_j^\infty|] \\
&\le \sum_{j=1}^{b} \mathsf{E}[|C_j^n - C_j^\infty|] + \sum_{j=b+1}^{n} \mathsf{E}[C_j^n] + \sum_{j=b+1}^{n} \mathsf{E}[C_j^\infty] \qquad (4.15)\\
&\le \frac{b\theta(\theta+1)}{\theta+n-b} + 1 + 2\theta \log\left(\frac{n}{b}\right),
\end{aligned}
$$

for any $b = 1, 2, \ldots, n$; see the proof of Theorem 2 of Arratia, Barbour and Tavaré (1992). When $\theta \to \infty$, by setting $b = \lfloor n/\theta \rfloor$, the first and third terms in the right-hand side of (4.15) are $O(\theta)$ and $O(\theta \log \theta)$, respectively. Otherwise, by setting $b = \lfloor n/2 \rfloor$ the result holds with the bound $d_n^W(n) = O(1)$. Hence, $d_n^W(n) = O((\theta \log \theta) \vee 1) = O(\theta \log(1+\theta))$. This completes the proof. □

4.4. *On asymptotic independence of* $\mathbf{C}_b^n$ (*Case C*). In this subsection, we will see that the independent process approximation for component counts seems difficult in Case C.

The probability mass function (1.1) is obtained from the conditioning relation (2.7) with a sequence of independent Poisson variables with respective means θ/j. We also get (1.1) from (2.7) with Poisson variables with respective means $(\theta/j)(n/\theta)^j$ (see, for instance, Watterson (1974a)). The following lemma shows that in Case C $\mathsf{E}[Z_j] = (\theta/j)(n/\theta)^j$ rather fit.

LEMMA 4.5. *In Case C, it holds that* $\mathsf{E}[C_j^n] \sim \theta/j(n/\theta)^j$ *for* $j = 1, 2, \ldots$. *Therefore, for* $j = 2, 3, \ldots$, *if* $\theta(n/\theta)^j \to 0$, *then*

$$
C_j^n \to^p 0. \qquad (4.16)
$$

PROOF. It holds that

$$
\mathsf{E}[C_j^n] = \frac{\theta}{j} \frac{n!}{(n-j)!} \frac{\Gamma(n+\theta-j)}{\Gamma(n+\theta)}
$$

which is (2.18) of Watterson (1974a). Since the Stirling formula $\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} + O(x^{x-3/2}/e^x)$ as $x \to \infty$ yields that $\Gamma(x-c)/\Gamma(x) \sim x^{-c}$ as $x \to \infty$ for any $c < x$, it holds that

$$
\mathsf{E}[C_j^n] \sim \frac{\theta}{j} n^j \frac{1}{(n+\theta)^j} \sim \frac{\theta}{j}\left(\frac{n}{\theta}\right)^j.
$$

Hence, the result (4.16) follows from $C_j^n \ge 0$. This completes the proof. □

REMARK 4.5. Since

$$\frac{n!}{(n-j)!n^j} \le \frac{n!}{(n-j)!(n-j+1)_j} = 1$$

and

$$\frac{\Gamma(n-j+\theta)\theta^j}{\Gamma(n+\theta)} = \frac{\theta^j}{(n-j+\theta)_j} \le 1,$$

it holds that

$$\frac{\mathsf{E}[jC_j^n]}{\theta(n/\theta)^j} = \frac{n!}{(n-j)!n^j}\frac{\Gamma(n-j+\theta)\theta^j}{\Gamma(n+\theta)} \le 1 \tag{4.17}$$

for $j = 1, 2, \dots, n$.

According to Lemma 4.5, it may be natural to consider that the distributions of C_j^n and Poisson variable with mean $(\theta/j)(n/\theta)^j$ are asymptotically similar, but Proposition A.3 indicates that, except Case C3, an independent process approximation by Poisson variables with means $(\theta/j)(n/\theta)^j$ seems difficult in the sense of the joint distribution. Actually, the following theorem shows that in Case C2 the linear relation $n - (C_1^n + 2C_2^n) \Rightarrow 0$ between C_1^n and C_2^n asymptotically remains.

THEOREM 4.2. (i) *In Case* C2, *it holds that*

$$(C_1^n - n, C_2^n) \Rightarrow (-2P_{c/2}, P_{c/2}),$$

where $c = \lim_{n,\theta}(n^2/\theta)$, *and* $\sum_{j=3}^n |C_j^n| \Rightarrow 0$.

(ii) *In Case* C3, *it holds that* $\sum_{j=1}^n |C_j^n - n1\{j=1\}| \Rightarrow 0$.

PROOF. (i) It follows from (4.17) that

$$\mathsf{E}\left[\sum_{j=3}^n jC_j^n\right] \le \sum_{j=3}^n \theta\left(\frac{n}{\theta}\right)^j = \frac{n^3}{\theta^2}\frac{1-(n/\theta)^{n-2}}{1-n/\theta},$$

which implies that $n - (C_1^n + 2C_2^n) = \sum_{j=3}^n jC_j^n \to^p 0$. It yields that $\sum_{j=3}^n |C_j^n| \to^p 0$ and so $K_n - (C_1^n + C_2^n) \to^p 0$. Hence, $K_n - n + C_2^n \to^p 0$, which implies that

$$\begin{aligned}
&K_n - n - \left(\theta\log\left(1+\frac{n}{\theta}\right) - n\right) + C_2^n + \left(\theta\log\left(1+\frac{n}{\theta}\right) - n\right)\\
&\quad = K_n - \theta\log\left(1+\frac{n}{\theta}\right) + C_2^n - \left(n - \theta\log\left(1+\frac{n}{\theta}\right)\right)\\
&\quad \to^p 0.
\end{aligned}$$

We conclude from (2.5), which means $K_n - \theta \log(1 + n/\theta) \Rightarrow c/2 - P_{c/2}$, that

$$C_2^n - \left(n - \theta \log\left(1 + \frac{n}{\theta}\right)\right) = C_2^n - \frac{n^2}{2\theta} + o(1) \Rightarrow P_{c/2} - \frac{c}{2},$$

hence that $C_2^n \Rightarrow P_{c/2}$. Moreover, from what has already been proved, we obtain $n - C_1^n \Rightarrow 2P_{c/2}$.

(ii) Since

$$\mathsf{E}\big[|n - C_1^n|\big] = \mathsf{E}\big[n - C_1^n\big] = n - \frac{n}{1 + (n-1)/\theta} = \frac{n^2}{\theta} + O\left(\frac{n^3}{\theta^2}\right),$$

Lemma 4.5 yields the result. This completes the proof. □

As direct applications of Theorem 4.2, we show the following corollaries which represent properties of the shortest cycle length $S_n = S_n^1$ (recalling that S_n^k is defined in (2.15)) and the longest cycle length L_n in a Ewens partition, where $L_n = \sup(j : C_j^n \geq 1)$. These extreme sizes are of interest in the combinatorial context; see, for instance, Mano (2017).

COROLLARY 4.3. *In Case C2 or C3, it holds that* $\mathsf{P}(S_n = 1) \to 1$.

PROOF. In Case C3, the conclusion is obvious, so we see Case C2. It follows from Theorem 4.2 that

$$\mathsf{P}(S_n \geq 2) = \mathsf{P}\big(C_1^n = 0\big) = \mathsf{P}\big(n - C_1^n \geq n\big) \sim \mathsf{P}(2P_{c/2} \geq n) \to 0.$$

This completes the proof. □

COROLLARY 4.4. *In Case C2, for a positive integer r*

$$\mathsf{P}(L_n \leq r) \to \begin{cases} e^{-c/2} & (r = 1), \\ 1 & (r \geq 2). \end{cases}$$

PROOF. For $r = 1$, it follows from Theorem 4.2 that

$$\mathsf{P}(L_n \leq 1) = \mathsf{P}\big(C_1^n = n\big) = \mathsf{P}\big(n - C_1^n = 0\big) \to \mathsf{P}(P_{c/2} = 0) = e^{-c/2}.$$

For $r \geq 2$, it follows from Theorem 4.2 that

$$\mathsf{P}(L_n \leq r) \geq \mathsf{P}(L_n \leq 2) = \mathsf{P}\left(\sum_{j=3}^{n} C_j^n = 0\right) \to 1.$$

This completes the proof. □

REMARK 4.6. Since the marginal distribution of C_1^n is given by

$$\mathsf{P}(C_1^n = k) = \frac{\theta^k}{k!}\left\{\sum_{j=0}^{n-k}(-1)^j \frac{\theta^j}{j!}\frac{(n+1-k-j)_{k+j}}{(n+\theta-k-j)_{k+j}}\right\}$$

for $k = 0, 1, \dots, n$, Corollary 4.4 when $r = 1$ directly follows from

$$\begin{aligned}\mathsf{P}(C_1^n = n) &= \frac{\theta^n}{(\theta)_n} = 1\frac{1}{1+1/\theta}\cdots\frac{1}{1+(n-1)/\theta} = 1 - \frac{1}{\theta}\frac{n(n-1)}{2} + o(1) \\ &= e^{-n^2/(2\theta)} + o(1).\end{aligned}$$

5. Functional central limit theorems. In this section, our first result is Theorem 5.1 which slightly extend the functional central limit theorems for the Ewens sampling formula proved by Hansen (1990) and Tsukuda (2018) in which θ is assumed to be fixed.

THEOREM 5.1. (i) *In Case A, if*

$$\frac{\theta}{\log n}(\log\theta)^2 \to 0, \tag{5.1}$$

then the random process $X_n^1(\cdot)$ defined in (2.16) *converges weakly to a standard Brownian motion $(B(u))_{0\le u\le 1}$ in $D[0,1]$.*

(ii) *In Case A, if*

$$\frac{\theta\log\log n}{\log n}(\log\theta)^2 \to 0, \tag{5.2}$$

then both of the random processes $X_n^2(\cdot)$ and $X_n^3(\cdot)$, which are defined in (2.17) *and* (2.18), *respectively, converge weakly to $(B(u)/\sqrt{u})_{0<u<1}$ in $L^2(0,1)$.*

REMARK 5.1. It follows from Theorem 5.1 (i) that if (5.1) holds then (2.1) holds. But as it is stated in (2.5), the asymptotic normality of K_n holds for far larger θ.

The second result of this section is weak convergences of $X_n^4(\cdot)$ and $X_n^5(\cdot)$

THEOREM 5.2. (i) *In Case A, if* (5.1) *holds then the random process $X_n^4(\cdot)$ defined in* (2.19) *converges weakly to a standard Brownian bridge $(B^\circ(u))_{0\le u\le 1}$ in $D[0,1]$.*

(ii) *In Case A, if* (5.2) *holds then the random process $X_n^5(\cdot)$ defined in* (2.20) *converges weakly to $(B^\circ(u)/\sqrt{u(1-u)})_{0<u<1}$ in $L^2(0,1)$.*

Before proving these results, let us prepare the following lemma which will appear in the proof of Theorem 5.1. This lemma states the error bounds of Poisson process approximations in the sense of the expectation of the error in the supremum norm and in the L^2 norm.

LEMMA 5.1. *In Case A*,

$$\mathsf{E}\left[\sup_{u\in[0,1]}\left|\sum_{j=1}^{\lfloor n^u\rfloor}\frac{(C_j^n - C_j^\infty)}{\sqrt{\theta\log n}}\right|\right] = O\left(\sqrt{\frac{\theta}{\log n}}\log(1+\theta)\right) \tag{5.3}$$

and

$$\mathsf{E}\left[\left[\int_0^1\left\{\frac{\sum_{j=1}^{\lfloor n^u\rfloor}(C_j^n - C_j^\infty)}{\sqrt{\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}}\right\}^2 du\right]^{1/2}\right] = O\left(\sqrt{\frac{\theta\log\log n}{\log n}}\log(1+\theta)\right). \tag{5.4}$$

PROOF. The desired result (5.3) follows from

$$\begin{aligned}\mathsf{E}\left[\sup_{u\in[0,1]}\left|\sum_{j=1}^{\lfloor n^u\rfloor}\frac{(C_j^n - C_j^\infty)}{\sqrt{\theta\log n}}\right|\right] &\le \mathsf{E}\left[\sup_{u\in[0,1]}\sum_{j=1}^{\lfloor n^u\rfloor}\frac{|C_j^n - C_j^\infty|}{\sqrt{\theta\log n}}\right]\\ &= \frac{\sum_{j=1}^n \mathsf{E}[|C_j^n - C_j^\infty|]}{\sqrt{\theta\log n}},\end{aligned}$$

and the proof of Proposition 4.4. The other result (5.4) follows from

$$\begin{aligned}&\int_0^1\left|\frac{\sum_{j=1}^{\lfloor n^u\rfloor}(C_j^n - C_j^\infty)}{\sqrt{\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}}\right|^2 du\\ &\quad\le \frac{2(\frac{1}{\log 2} - \log\log 2 + \log\log n)}{\theta\log n}\left(\sum_{j=1}^n|C_j^n - C_j^\infty|\right)^2\end{aligned}$$

(for this evaluation see the proof of Lemma 3.1 of Tsukuda (2018)) and the proof of Proposition 4.4. This completes the proof. □

Next, we prove Theorems 5.1 and 5.2.

PROOF OF THEOREM 5.1. (i) From (5.3) and the assumption (5.1), it follows that

$$\begin{aligned}\sup_{u\in[0,1]}\left|\sum_{j=1}^{\lfloor n^u\rfloor}\frac{(C_j^n - C_j^\infty)}{\sqrt{\theta\log n}}\right| &\le \sup_{u\in[0,1]}\sum_{j=1}^{\lfloor n^u\rfloor}\frac{|C_j^n - C_j^\infty|}{\sqrt{\theta\log n}}\\ &= \sum_{j=1}^n\frac{|C_j^n - C_j^\infty|}{\sqrt{\theta\log n}} \to^p 0.\end{aligned} \tag{5.5}$$

By using the functional central limit theorem for Poisson processes in $D[0,1]$, the random process

$$\Bigl(\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - \sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}{\sqrt{\sum_{j=1}^{n}\theta/j}}\Bigr)_{0\le u\le 1}$$

converges weakly to a standard Brownian motion $(B(u))_{0\le u\le 1}$ in $D[0,1]$ (see the Proof of Theorem 5 of Arratia and Tavaré (1992a)). Since

$$\sup_{u\in[0,1]}\Bigl|\sum_{j=1}^{\lfloor n^u\rfloor}\frac{\theta}{j} - u\theta\log n\Bigr| = O(\theta),$$

the random process

$$\Bigl(\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - u\theta\log n}{\sqrt{\theta\log n}}\Bigr)_{0\le u\le 1} \tag{5.6}$$

converges weakly to $(B(u))_{0\le u\le 1}$ in $D[0,1]$ because of the assumption (5.1). From (5.5) and the weak convergence of (5.6), Theorem 2.7(iv) of van der Vaart (1998) yields the result.

(ii) First, we argue $X_n^2(\cdot)$. From (5.4) and (5.2), it follows that

$$\int_0^1\Bigl|\frac{\sum_{j=1}^{\lfloor n^u\rfloor}(C_j^n - C_j^\infty)}{\sqrt{\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}}\Bigr|^2 du \to^p 0.$$

It holds that

$$\begin{aligned}&\Bigl(\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - \sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}{\sqrt{\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}}\Bigr)_{0<u<1}\\ &\quad =^d \Bigl(\frac{N^1(\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j) - \sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}{\sqrt{\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j}}\Bigr)_{0<u<1},\end{aligned} \tag{5.7}$$

where $(N^1(t))_{t\ge 0}$ is a homogeneous Poisson process with unit intensity satisfying $N^1(\theta\sum_{j=1}^{\lfloor t\rfloor}1/j) = \sum_{j=1}^{\lfloor t\rfloor} C_j^\infty$ for all $t>0$. Since

$$\sup_{u\in(0,1)}\frac{|\sum_{j=1}^{\lfloor n^u\rfloor}\theta/j - u\theta\log n|}{\theta\log n}\to 0$$

and the other hypotheses hold with $\lambda = 1$, $s_n(u) = \sum_{j=1}^{\lfloor n^u\rfloor}\theta/j$ and $f(n) = \theta\log n$ (see Section 6.2 of Tsukuda (2018)) Lemma A.5 in the Appendix implies that (5.7) converges weakly to $(B(u)/\sqrt{u})_{0<u<1}$ in $L^2(0,1)$. From what has been already proved, Theorem 2.7(iv) of van der Vaart (1998) yields the result.

Next, we argue $X_n^3(\cdot)$. It follows that

$$\int_0^{\frac{\varepsilon}{\log n}} \frac{(N^1(u\theta\log n) - u\theta\log n)^2}{u\theta\log n}\,du \to^p 0,$$

$$\int_{\frac{\varepsilon}{\log n}}^1 \frac{\{\sum_{j=1}^{\lfloor n^u\rfloor}(C_j^n - C_j^\infty)\}^2}{u\theta\log n}\,du \to^p 0,$$

$$\int_{\frac{\varepsilon}{\log n}}^1 \frac{(\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - N^1(u\theta\log n))^2}{u\theta\log n}\,du \to^p 0,$$

from the almost same argument as the proof of Theorem 7.1 of Tsukuda (2018) by the assumption (5.2). So, we have

$$\int_0^1 \biggl(X_n^3(u) - \frac{N^1(u\theta\log n) - u\theta\log n}{\sqrt{u\theta\log n}}\biggr)^2\,du \to^p 0.$$

From Lemma A.5 in the Appendix with $\lambda = 1$, $s_n(u) = uf(n)$ and $f(n) = \theta\log n$, it holds that the random process

$$\biggl(\frac{N^1(u\theta\log n) - u\theta\log n}{\sqrt{u\theta\log n}}\biggr)_{0<u<1}$$

converges weakly to $(B(u)/\sqrt{u})_{0<u<1}$ in $L^2(0,1)$. Consequently, the desired result follows. This completes the proof. $\square$

PROOF OF THEOREM 5.2. (i) Since it holds that

$$\begin{aligned} X_n^4(u) &= \frac{\theta\log n}{K_n}\biggl(\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - uK_n}{\sqrt{\theta\log n}}\biggr) \\ &= \frac{\theta\log n}{K_n}\biggl\{\frac{(1-u)\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - u\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^n}{\sqrt{\theta\log n}}\biggr\} \end{aligned}$$

for any $u \in [0,1]$, it is sufficient to show

$$\frac{K_n}{\theta\log n} \to^p 1 \tag{5.8}$$

and

$$\biggl(\frac{(1-u)\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - u\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^n}{\sqrt{\theta\log n}}\biggr)_{0\le u\le 1} \Rightarrow \bigl(B^\circ(u)\bigr)_{0\le u\le 1} \tag{5.9}$$

in $D[0,1]$. First, (5.8) holds because the assumption (5.1) yields that $\log\theta/\log n \to 0$ and because it follows from (2.6) that

$$\frac{K_n}{\theta\log n} = \frac{K_n}{\theta\log(n/\theta)(1 + \frac{\log\theta}{\log(n/\theta)})} \to^p 1.$$

Next, we show (5.9). Since it follows from Proposition 4.4 and from the assumption (5.1) (see (5.5)) that

$$\sup_{u\in[0,1]}\left|\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - \sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty}{\sqrt{\theta \log n}}\right| \le \frac{\sum_{j=1}^{n} |C_j^n - C_j^\infty|}{\sqrt{\theta \log n}} \to^p 0$$

and that

$$\sup_{u\in[0,1]}\left|\frac{\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^n - \sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^\infty}{\sqrt{\theta \log n}}\right| \le \frac{\sum_{j=1}^{n} |C_j^n - C_j^\infty|}{\sqrt{\theta \log n}} \to^p 0,$$

the triangle inequality yields that

$$\sup_{u\in[0,1]}\left|\frac{(1-u)\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - u\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^n}{\sqrt{\theta \log n}} - P_4^\circ(u)\right| \to^p 0,$$

where

$$\left(P_4^\circ(u)\right)_{0\le u\le 1} = \left(\frac{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - u\sum_{j=1}^{n} C_j^\infty}{\sqrt{\theta \log n}}\right)_{0\le u\le 1}.$$

By using the functional central limit theorem for Poisson processes in $D[0,1]$, $P_4^\circ(\cdot)$ converges weakly to $(B^\circ(u))_{0\le u\le 1}$ in $D[0,1]$.

(ii) By the same reason as (i), it is sufficient to show (5.8) and

$$\begin{aligned}&\left(\frac{(1-u)\sum_{j=1}^{\lfloor n^u\rfloor} C_j^n - u\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^n}{\sqrt{u(1-u)\theta \log n}} 1\left\{\frac{\varepsilon}{\log n} < u < 1-\frac{\varepsilon}{\log n}\right\}\right)_{0<u<1}\\ &\quad\Rightarrow \left(\frac{B^\circ(u)}{\sqrt{u(1-u)}}\right)_{0<u<1}\end{aligned} \tag{5.10}$$

in $L^2(0,1)$. Here, we show (5.10). First, it holds that

$$\begin{aligned}&\int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{(1-u)^2\{\sum_{j=1}^{\lfloor n^u\rfloor}(C_j^n - C_j^\infty)\}^2}{u(1-u)\theta \log n}\,du\\ &\quad\le \int_{\frac{\varepsilon}{\log n}}^{1} \frac{(\sum_{j=1}^{\lfloor n^u\rfloor}|C_j^n - C_j^\infty|)^2}{u\theta \log n}\,du\\ &\quad\le \int_{\frac{\varepsilon}{\log n}}^{1} \frac{(\sum_{j=1}^{n}|C_j^n - C_j^\infty|)^2}{u\theta \log n}\,du\\ &\quad\le \frac{\log\log n - \log\varepsilon}{\theta \log n}\left(\sum_{j=1}^{n}\left|C_j^n - C_j^\infty\right|\right)^2,\end{aligned} \tag{5.11}$$

and that

$$\int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{u^2\{\sum_{j=\lfloor n^u\rfloor+1}^{n}(C_j^n - C_j^\infty)\}^2}{u(1-u)\theta\log n}\,du$$

$$(5.12)\qquad \le \int_0^{1-\frac{\varepsilon}{\log n}} \frac{(\sum_{j=\lfloor n^u\rfloor+1}^{n}|C_j^n - C_j^\infty|)^2}{(1-u)\theta\log n}\,du$$

$$\le \frac{\log\log n - \log\varepsilon}{\theta\log n}\left(\sum_{j=1}^{n}\left|C_j^n - C_j^\infty\right|\right)^2.$$

These right-hand sides of (5.11) and (5.12) converge to 0 in probability because the expectations of their square root converge to 0 by the assumption (5.2). Second, it follows from

$$\begin{aligned}\mathsf{E}\left[\left(N^1(u\theta\log n) - uN^1(\theta\log n)\right)^2\right] &= \operatorname{var}\left(N^1(u\theta\log n) - uN^1(\theta\log n)\right)\\ &= (1-u)^2 u\theta\log n + u^2(1-u)\theta\log n\\ &= u(1-u)\theta\log n\end{aligned}$$

that

$$\int_0^{\frac{\varepsilon}{\log n}} \frac{(N^1(u\theta\log n) - uN^1(\theta\log n))^2}{u(1-u)\theta\log n}\,du \to^p 0,$$

and that

$$\int_{1-\frac{\varepsilon}{\log n}}^{1} \frac{(N^1(u\theta\log n) - uN^1(\theta\log n))^2}{u(1-u)\theta\log n}\,du \to^p 0,$$

where $(N^1(t))_{t\ge 0}$ is defined in the proof of Theorem 5.1. Third, it holds that

$$\begin{aligned}&\int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{1}{u(1-u)\theta\log n}\\ &\quad\times\left\{\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - u\sum_{j=1}^{n} C_j^\infty - \left(N^1(u\theta\log n) - uN^1(\theta\log n)\right)\right\}^2 du\\ &\quad= \int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{1}{u(1-u)\theta\log n}\left[(1-u)\left(\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - N^1(u\theta\log n)\right)\right.\\ &\qquad\left. - u\left\{\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^\infty - \left(N^1(\theta\log n) - N^1(u\theta\log n)\right)\right\}\right]^2 du\\ &\quad\le \int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{2(1-u)^2}{u(1-u)\theta\log n}\left(\sum_{j=1}^{\lfloor n^u\rfloor} C_j^\infty - N^1(u\theta\log n)\right)^2 du\end{aligned}$$

$$
\begin{aligned}
&+\int_{\frac{\varepsilon}{\log n}}^{1-\frac{\varepsilon}{\log n}} \frac{2u^2}{u(1-u)\theta\log n} \\
&\quad\times\Bigg\{\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^{\infty}-\big(N^1(\theta\log n)-N^1(u\theta\log n)\big)\Bigg\}^2\,du \\
\le{}& \int_{\frac{\varepsilon}{\log n}}^{1} \frac{2(\sum_{j=1}^{\lfloor n^u\rfloor} C_j^{\infty}-N^1(u\theta\log n))^2}{u\theta\log n}\,du \\
&+\int_0^{1-\frac{\varepsilon}{\log n}} \frac{2\{\sum_{j=\lfloor n^u\rfloor+1}^{n} C_j^{\infty}-(N^1(\theta\log n)-N^1(u\theta\log n))\}^2}{(1-u)\theta\log n}\,du.
\end{aligned}
$$

The distributions of the first term and second term in the right-hand side are equal to

$$
\int_{\frac{\varepsilon}{\log n}}^{1} \frac{2(N^1(\theta(\sum_{j=1}^{\lfloor n^u\rfloor} 1/j-u\log n)))^2}{u\theta\log n}\,du
$$

and

$$
\int_0^{1-\frac{\varepsilon}{\log n}} \frac{2(N^1(\theta(\sum_{j=1}^{n} 1/j-\log n))-N^1(\theta(\sum_{j=1}^{\lfloor n^u\rfloor} 1/j-u\log n))))^2}{(1-u)\theta\log n}\,du,
$$

respectively. Both of them converge to 0 in probability because their expectations tend to 0 from the assumption (5.2). Thus, the triangle inequality yields that

$$
\int_0^1 \left|X_n^5(u)-P_5^{\circ}(u)\right|^2 du \to^p 0,
$$

where

$$
\big(P_5^{\circ}(u)\big)_{0<u<1}=\left(\frac{N^1(\theta u\log n)-uN^1(\theta\log n)}{\sqrt{u(1-u)\theta\log n}}\right)_{0<u<1}.
$$

Since

$$
\begin{aligned}
\big(P_5^{\circ}(u)\big)_{0<u<1}&=\left(\int_0^{\theta\log n} \frac{1\{t\le u\theta\log n\}-u}{\sqrt{u(1-u)\theta\log n}}\,dN^1(t)\right)_{0<u<1} \\
&=\left(\int_0^{\theta\log n} \frac{1\{t\le u\theta\log n\}-u}{\sqrt{u(1-u)\theta\log n}}\big(dN^1(t)-dt\big)\right)_{0<u<1},
\end{aligned}
$$

Theorem 4 of Tsukuda (2017a) yields that $(P_5^{\circ}(u))_{0<u<1}\Rightarrow(B^{\circ}(u)/\sqrt{u})_{0<u<1}$ by setting $H_s=1$ with $d=1$, $\lambda_s=1$ and $T=\theta\log n$. Consequently, we have (5.10). This completes the proof. $\square$

APPENDIX. AUXILIARY RESULTS

A.1. Asymptotic evaluations associated with $\{p_j\}$ and $\{q_j\}$. Let us show some asymptotic evaluations associated with $\{p_j\}_{j=1}^{\infty}$ and $\{q_j\}_{j=1}^{\infty}$ defined in (3.1).

PROPOSITION A.1. (i) *It holds that*

$$\frac{n}{2(\theta+n)} \le \sum_{j=1}^{n} p_j - \theta \log\left(1+\frac{n}{\theta}\right) \le \frac{n}{\theta+n} \tag{A.1}$$

and that

$$0 \le \sum_{j=1}^{n} p_j^2 - \frac{n\theta}{n+\theta} \le 1. \tag{A.2}$$

Especially, in Case A, it holds that $\sum_{j=1}^{n} p_j \sim \theta \log(1+n/\theta)$, *and that if* $\theta \to \infty$ *then* $\sum_{j=1}^{n} p_j^2 \sim \theta$.

(ii) *It holds that*

$$\frac{n(n-1)}{2(\theta+n)} \le \sum_{j=1}^{n} q_j \le \frac{n(n-1)}{2\theta} \tag{A.3}$$

and that

$$\frac{n(n-1)(2n-1)}{6(\theta+n)^2} \le \sum_{j=1}^{n} q_j^2 \le \frac{n(n-1)(2n-1)}{6\theta^2}. \tag{A.4}$$

Especially, in Case C, it holds that $\sum_{j=1}^{n} q_j \sim n^2/(2\theta)$ *and* $\sum_{j=1}^{n} q_j^2 \sim n^3/(3\theta^2)$.

PROOF. (i) Since $x \mapsto 1/x$ is convex, it holds for any $j = 1, 2, \dots$ that

$$\int_{\theta+j-1}^{\theta+j} \frac{dx}{x} \le \frac{1}{2}\left(\frac{1}{\theta+j-1} + \frac{1}{\theta+j}\right),$$

which is equivalent to

$$\frac{1}{2}\left(\frac{1}{\theta+j-1} - \frac{1}{\theta+j}\right) \le \frac{1}{\theta+j-1} - \int_{\theta+j-1}^{\theta+j} \frac{dx}{x}.$$

It yields that

$$\frac{1}{2}\sum_{j=1}^{n}\left(\frac{1}{\theta+j-1} - \frac{1}{\theta+j}\right) \le \sum_{j=1}^{n} \frac{1}{\theta+j-1} - \int_{\theta}^{\theta+n} \frac{dx}{x}.$$

On the other hand, it holds that

$$\sum_{j=1}^{n} \frac{1}{\theta+j-1} - \int_{\theta}^{\theta+n} \frac{dx}{x} \le \sum_{j=1}^{n}\left(\frac{1}{\theta+j-1} - \frac{1}{\theta+j}\right).$$

These inequalities lead

$$\frac{1}{2}\frac{n}{\theta(\theta+n)} \le \sum_{j=1}^{n} \frac{1}{\theta+j-1} - \log\left(\frac{\theta+n}{\theta}\right) \le \frac{n}{\theta(\theta+n)}$$

and so the result (A.1) holds. Since

$$\int_{\theta}^{\theta+n} \frac{dx}{x^2} \le \sum_{j=1}^{n} \frac{1}{(\theta+j-1)^2} \le \frac{1}{\theta^2} + \int_{\theta}^{\theta+n} \frac{dx}{x^2},$$

the result (A.2) holds.

(ii) From

$$\frac{j-1}{\theta+n} \le q_j \le \frac{j-1}{\theta}$$

and from

$$\frac{(j-1)^2}{(\theta+n)^2} \le q_j^2 \le \frac{(j-1)^2}{\theta^2},$$

the results (A.3) and (A.4) follow, respectively. This completes the proof. □

PROPOSITION A.2. *In Case A, it holds that*

$$\sum_{j=1}^{n} p_j - \theta\big(\log n - \psi(\theta)\big) = O\left(\frac{\theta^2}{n}\right). \tag{A.5}$$

PROOF. It follows from

$$\sum_{j=1}^{n} p_j = \theta\big(\psi(n+\theta) - \psi(\theta)\big) \tag{A.6}$$

that the left-hand side of (A.5) is $\theta(\psi(n+\theta) - \log n)$. Since $\psi(n+\theta) = -\gamma - 1/(n+\theta) + (n+\theta)\sum_{j=1}^{\infty} 1/\{j(n+\theta+j)\}$ and $\log n = \sum_{j=1}^{n} 1/j - \gamma + O(1/n)$ as $n \to \infty$ where γ is the Euler constant, it holds that

$$\begin{aligned}
\psi(n+\theta) - \log n &= -\sum_{j=1}^{n} \frac{1}{j} + \sum_{j=1}^{\infty}\left(\frac{1}{j} - \frac{1}{n+\theta+j}\right) + O\left(\frac{1}{n}\right) \\
&\le -\sum_{j=1}^{n} \frac{1}{j} + \sum_{j=1}^{\infty}\left(\frac{1}{j} - \frac{1}{n+\lfloor\theta\rfloor+1+j}\right) + O\left(\frac{1}{n}\right) \\
&= \sum_{j=n+1}^{n+\lfloor\theta\rfloor+1} \frac{1}{j} + O\left(\frac{1}{n}\right).
\end{aligned}$$

In Case A, the first term in the right-hand side is $O(\theta/n)$. This completes the proof.

□

REMARK A.2. As it is stated in (2.3), Yamato (2013) discussed the asymptotic normality of K_n standardized by $\theta(\log n - \psi(\theta))$, which means that $\psi(n+\theta)$ is approximated by $\log n$ from (A.6). If $\theta^3/(n^2\log(n/\theta)) \to \infty$, the bound in (A.5) is meaningless to discuss CLT. On the other hand, if $\theta^2/n \to 0$ the centering by $\theta(\log n - \psi(\theta))$ is better than centering by $\theta\log(1+n/\theta)$, which was used in Corollary 2 of Tsukuda (2017b), because $\sum_{j=1}^n p_j - \theta\log(1+n/\theta) = \Theta(n/(n+\theta))$.

PROPOSITION A.3. *In Case C, it holds that*

$$\sum_{j=1}^n p_j - \sum_{j=1}^n \frac{\theta}{j}\left(\frac{n}{\theta}\right)^j = O\left(\frac{n^2}{\theta}\right).$$

PROOF. The triangle inequality yields that

$$\begin{aligned}
&\left|\sum_{j=1}^n \left\{p_j - \frac{\theta}{j}\left(\frac{n}{\theta}\right)^j\right\}\right| \\
&\quad \le \left|\sum_{j=1}^n p_j - \theta\log\left(1+\frac{n}{\theta}\right)\right| + \left|\theta\log\left(1+\frac{n}{\theta}\right) - \log\left(1-\frac{n}{\theta}\right)^{-\theta}\right| \\
&\qquad + \left|\log\left(1-\frac{n}{\theta}\right)^{-\theta} - \sum_{j=1}^n \frac{\theta}{j}\left(\frac{n}{\theta}\right)^j\right|.
\end{aligned}$$

The first term is $O(n/(n+\theta)) = O(n/\theta)$, the second term is $\theta\log(1-n^2/\theta^2) = O(n^2/\theta)$, and from $\log(1-x)^{-1} = x + x^2/2 + \cdots$ as $x \to 0$ the third term is

$$\sum_{j=n+1}^\infty \frac{\theta}{j}\left(\frac{n}{\theta}\right)^j \le \frac{\theta}{n}\sum_{j=n+1}^\infty \left(\frac{n}{\theta}\right)^j = \frac{\theta}{n}\frac{(n/\theta)^{n+1}}{1-n/\theta} = O\left(\left(\frac{n}{\theta}\right)^n\right).$$

This completes the proof. $\square$

A.2. Technical lemmas.

LEMMA A.2. *In Case A, if* $\theta^2/n \to 0$ *then*

$$\Gamma(\theta)\frac{(\theta)_n}{n!} = n^{\theta-1}\left\{1+\frac{\theta(\theta-1)}{2n}\right\} + O\left(n^{\theta-1}\frac{\theta^4}{n^2}\right). \tag{A.7}$$

PROOF. The left-hand side of (A.7) equals $\Gamma(n+\theta)/(n\Gamma(n))$. By using the asymptotic series expansion $\Gamma(x) = \sqrt{2\pi}e^{-x}x^{x-1/2}\{1+1/(12x)\} +$

$O(e^{-x}x^{x-5/2})$ as $x \to \infty$, it holds that

$$\begin{aligned}
&\Gamma(n+\theta)\\
&\quad = \sqrt{2\pi}e^{-(n+\theta)}(n+\theta)^{n+\theta-1/2}\left\{1+\frac{1}{12(n+\theta)}\right\}\\
&\qquad + O\big(e^{-(n+\theta)}(n+\theta)^{n+\theta-5/2}\big)\\
&\quad = \sqrt{2\pi}e^{-(n+\theta)}n^{n+\theta-1/2}\left(1+\frac{\theta}{n}\right)^{n+\theta-1/2}\left(1+\frac{1}{12n}+O\left(\frac{\theta}{n^2}\right)\right)
\end{aligned}$$

and that

$$\begin{aligned}
n\Gamma(n) &= n\sqrt{2\pi}e^{-n}n^{n-1/2}\left(1+\frac{1}{12n}\right)+O\big(e^{-n}n^{n-5/2}\big)\\
&= \sqrt{2\pi}e^{-n}n^{n+1/2}\left(1+\frac{1}{12n}\right)\left(1+O\left(\frac{1}{n^2}\right)\right).
\end{aligned}$$

Hence, the left-hand side of (A.7) is

$$\begin{aligned}
&n^{\theta-1}e^{-\theta}\left(1+\frac{\theta}{n}\right)^n\left(1+\frac{\theta}{n}\right)^{\theta-1/2}\left(1+O\left(\frac{\theta}{n^2}\right)\right)\left(1+O\left(\frac{1}{n^2}\right)\right)\\
&\quad = n^{\theta-1}e^{-\theta}\left(1+\frac{\theta}{n}\right)^n\left(1+\frac{\theta}{n}\right)^{\theta-1/2}\left(1+O\left(\frac{\theta}{n^2}\right)\right)
\end{aligned} \tag{A.8}$$

and, from the asymptotic expansion $(1+1/x)^x = e(1-1/(2x)+O(x^{-2}))$ as $x \to \infty$ it follows that

$$\left(1+\frac{\theta}{n}\right)^n = e^{\theta}\left(1-\frac{\theta}{2n}+O\left(\frac{\theta^2}{n^2}\right)\right)^{\theta}.$$

Therefore, (A.8) is

$$\begin{aligned}
&n^{\theta-1}\left(1-\frac{\theta}{2n}+O\left(\frac{\theta^2}{n^2}\right)\right)^{\theta}\left(1+\frac{\theta}{n}\right)^{\theta}\left(1+\frac{\theta}{n}\right)^{-1/2}\left(1+O\left(\frac{\theta}{n^2}\right)\right)\\
&\quad = n^{\theta-1}\left(1+\frac{\theta}{2n}+O\left(\frac{\theta^2}{n^2}\right)\right)^{\theta}\left(1+\frac{\theta}{n}\right)^{-1/2}\left(1+O\left(\frac{\theta}{n^2}\right)\right)\\
&\quad = n^{\theta-1}\left(1+\frac{\theta^2}{2n}+O\left(\frac{\theta^4}{n^2}\right)\right)\left(1-\frac{\theta}{2n}+O\left(\frac{\theta^2}{n^2}\right)\right)\left(1+O\left(\frac{\theta}{n^2}\right)\right)\\
&\quad = n^{\theta-1}\left\{1+\frac{\theta(\theta-1)}{2n}+O\left(\frac{\theta^4}{n^2}\right)\right\}\left(1+O\left(\frac{\theta}{n^2}\right)\right)\\
&\quad = n^{\theta-1}\left\{1+\frac{\theta(\theta-1)}{2n}+O\left(\frac{\theta^4}{n^2}\right)\right\}.
\end{aligned}$$

This completes the proof. □

LEMMA A.3. *Let b be an positive integer. For $a > 1$, it holds that*

$$\sum_{j=1}^{b} \frac{a^j}{j} \leq \log b + a^b.$$

PROOF. It holds that

$$\sum_{j=1}^{b} \frac{a^j}{j} = \sum_{j=1}^{b} \int_0^a t^{j-1}\,dt = \sum_{j=1}^{b} \left(\frac{1}{j} + \int_1^a t^{j-1}\,dt \right) \leq 1 + \log b + \int_1^a \sum_{j=1}^{b} t^{j-1}\,dt.$$

As for the last term, it holds that

$$\int_1^a \sum_{j=1}^{b} t^{j-1}\,dt \leq \int_1^a bt^{b-1}\,dt = a^b - 1.$$

This completes the proof. □

LEMMA A.4. *For any $a > 0$ and any positive integer b, $(x-a)_b/(x)_b$ is increasing with respect to $x > a$.*

PROOF. The proof is by induction on b. When $b = 1$, $(x-a)/x = 1 - a/x$ is increasing. Let x_1 and x_2 satisfy $a < x_1 < x_2$. If the conclusion of the lemma is true for b, then the conclusion is also true for $b+1$ because

$$\begin{aligned}
\frac{(x_2-a)_{b+1}}{(x_2)_{b+1}} &= \frac{x_2+b-a}{x_2+b} \frac{(x_2-a)_b}{(x_2)_b} \\
&> \frac{x_2+b-a}{x_2+b} \frac{(x_1-a)_b}{(x_1)_b} \\
&> \frac{x_1+b-a}{x_1+b} \frac{(x_1-a)_b}{(x_1)_b} \\
&= \frac{(x_1-a)_{b+1}}{(x_1)_{b+1}}.
\end{aligned}$$

This completes the proof. □

LEMMA A.5. *Let $(N_t)_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$ satisfying $N_0 = 0$. Define the nondecreasing function $(n,u) \mapsto s_n(u)$ with respect to $0 \leq u \leq 1$ and with respect to $n = 1, 2, \ldots$ which satisfies* $\inf_{u\in(\tau,1)} s_n(u) > 0$ *for all* $0 < \tau < 1$,

$$\lim_{n\to\infty} \left(\frac{\sup_{u\in(0,1)} |s_n(u) - uf(n)|}{f(n)} \right) = 0 \tag{A.9}$$

with $n \mapsto f(n)$ an increasing function of n satisfying $\lim_{n\to\infty} f(n) = \infty$, and

$$\lim_{n\to\infty}\left\{\int_0^1 \frac{du}{(s_n(u))^\delta}\right\} = 0$$

for some $\delta > 0$. Then the random process

$$\left(\frac{N_{s_n(u)} - \lambda s_n(u)}{\sqrt{\lambda s_n(u)}}\right)_{0<u<1}$$

converges weakly to $(B(u)/\sqrt{u})_{0<u<1}$ in $L^2(0,1)$ as $n \to \infty$.

REMARK A.3. Lemma A.5 is a slight generalization of Lemma 2.1 of Tsukuda (2018). The only difference is the condition (A.9), where corresponding condition (2.1) of Tsukuda (2018) is the case that $f(n) = K \log n$ with a positive constant K. To show Lemma A.5, the equation

$$\lim_{n\to\infty}\left(\frac{s_n(u) \wedge s_n(v)}{\sqrt{s_n(u)s_n(v)}}\right) = \lim_{n\to\infty}\left(\frac{K((u\log n) \wedge (v \log n))}{\sqrt{K^2 \log n^u \log n^v}}\right) = \frac{u \wedge v}{\sqrt{uv}}$$

in the proof of Lemma 2.1 of Tsukuda (2018) should be replaced by

$$\lim_{n\to\infty}\left(\frac{s_n(u) \wedge s_n(v)}{\sqrt{s_n(u)s_n(v)}}\right) = \lim_{n\to\infty}\left\{\frac{(uf(n)) \wedge (vf(n))}{\sqrt{uf(n)vf(n)}}\right\} = \frac{u \wedge v}{\sqrt{uv}},$$

and the other part has no need to change.

Acknowledgment. The author would like to express his heartfelt gratitude to Professor Shuhei Mano for a lot of constructive comments, and to the Editor and the referees for helpful comments and suggestions.

REFERENCES

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.* **2** 519–535. MR1177897

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2000). Limits of logarithmic combinatorial structures. *Ann. Probab.* **28** 1620–1644. MR1813836

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2016). Exploiting the Feller coupling for the Ewens sampling formula [comment on MR3458585]. *Statist. Sci.* **31** 27–29. MR3458588

ARRATIA, R. and DESALVO, S. (2016). Probabilistic divide-and-conquer: A new exact simulation method, with integer partitions as an example. *Combin. Probab. Comput.* **25** 324–351. MR3482658

ARRATIA, R., STARK, D. and TAVARÉ, S. (1995). Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies. *Ann. Probab.* **23** 1347–1388. MR1349176

ARRATIA, R. and TAVARÉ, S. (1992a). Limit theorems for combinatorial structures via discrete process approximations. *Random Structures Algorithms* **3** 321–345. MR1164844

ARRATIA, R. and TAVARÉ, S. (1992b). The cycle structure of random permutations. *Ann. Probab.* **20** 1567–1591. MR1175278

ARRATIA, R., BARBOUR, A. D., EWENS, W. J. and TAVARÉ, S. (2018). Dual diffusions, killed diffusions, and the age distribution problem in population genetics. *Theor. Popul. Biol.* **122** 5–11.

BARBOUR, A. D. (1992). Refined approximations for the Ewens sampling formula. *Random Structures Algorithms* **3** 267–276. MR1164840

BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480. MR0755837

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. The Clarendon Press, New York. MR1163825

CRANE, H. (2016). The ubiquitous Ewens sampling formula. *Statist. Sci.* **31** 1–19. MR3458585

DELAURENTIS, J. M. and PITTEL, B. G. (1985). Random permutations and Brownian motion. *Pacific J. Math.* **119** 287–301. MR0803120

DESALVO, S. (2018). Probabilistic divide-and-conquer: Deterministic second half. *Adv. in Appl. Math.* **92** 17–50. MR3699112

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3** 87–112; erratum, ibid. 3 (1972), 240; erratum, ibid. 3 (1972), 376. MR0325177

FAVARO, S. and JAMES, L. F. (2016). Relatives of the Ewens sampling formula in Bayesian nonparametrics [comment on MR3458585]. *Statist. Sci.* **31** 30–33. MR3458589

FENG, S. (2007). Large deviations associated with Poisson–Dirichlet distribution and Ewens sampling formula. *Ann. Appl. Probab.* **17** 1570–1595. MR2358634

FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics: Models and Asymptotic Behaviors*. Springer, Heidelberg. MR2663265

FENG, S. (2016). Diffusion processes and the Ewens sampling formula [comment on MR3458585]. *Statist. Sci.* **31** 20–22. MR3458586

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FLAJOLET, P. and SORIA, M. (1990). Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combin. Theory Ser. A* **53** 165–182. MR1041444

GONCHAROV, V. L. (1944). Some facts from combinatorics. *Izv. Akad. Nauk SSSR, Ser. Mat.* **8** 3–48.

HANSEN, J. C. (1990). A functional central limit theorem for the Ewens sampling formula. *J. Appl. Probab.* **27** 28–43. MR1039182

JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York. MR1429617

KNUTH, D. E. and WILF, H. S. (1989). A short proof of Darboux's lemma. *Appl. Math. Lett.* **2** 139–140. MR1003845

MANO, S. (2017). Extreme sizes in Gibbs-type exchangeable random partitions. *Ann. Inst. Statist. Math.* **69** 1–37. MR3590710

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249

SHEPP, L. A. and LLOYD, S. P. (1966). Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.* **121** 340–357. MR0195117

TEH, Y. W. (2016). Bayesian nonparametric modeling and the ubiquitous Ewens sampling formula [comment on MR3458585]. *Statist. Sci.* **31** 34–36. MR3458590

TSUKUDA, K. (2017a). A change detection procedure for an ergodic diffusion process. *Ann. Inst. Statist. Math.* **69** 833–864. MR3671640

TSUKUDA, K. (2017b). Estimating the large mutation parameter of the Ewens sampling formula. *J. Appl. Probab.* **54** 42–54. Correction: to appear in *J. Appl. Probab.* **55**, no. 3. MR3632604

TSUKUDA, K. (2018). Functional central limit theorems in $L^2(0, 1)$ for logarithmic combinatorial assemblies. *Bernoulli* **24** 1033–1052. MR3706786

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

VARRON, D. (2014). Donsker and Glivenko–Cantelli theorems for a class of processes generalizing the empirical process. *Electron*. *J*. *Stat*. **8** 2296–2320. MR3275745

WATTERSON, G. A. (1974a). Models for the logarithmic species abundance distributions. *Theor*. *Popul*. *Biol*. **6** 217–250. MR0368829

WATTERSON, G. A. (1974b). The sampling theory of selectively neutral alleles. *Adv. in Appl*. *Probab*. **6** 463–488. MR0351504

YAMATO, H. (2013). Edgeworth expansions for the number of distinct components associated with the Ewens sampling formula. *J*. *Japan Statist*. *Soc*. **43** 17–28. MR3154716

YANNAROS, N. (1991). Poisson approximation for random sums of Bernoulli random variables. *Statist*. *Probab*. *Lett*. **11** 161–165. MR1092977

GRADUATE SCHOOL OF ARTS AND SCIENCES
THE UNIVERSITY OF TOKYO
3-8-1 KOMABA, MEGURO-KU
TOKYO 153-8902
JAPAN
E-MAIL: ctsukuda@g.ecc.u-tokyo.ac.jp