

A CENTRAL LIMIT THEOREM ASSOCIATED WITH THE TRANSFORMED TWO-PARAMETER POISSON–DIRICHLET DISTRIBUTION

FANG XU,* *Beijing Normal University and McMaster University*

Abstract

In this paper we introduce the transformed two-parameter Poisson–Dirichlet distribution $\Pi_{\theta,\alpha}^\sigma$ on the ordered infinite simplex. Furthermore, we prove the central limit theorem related to this distribution when both the mutation rate θ and the selection rate σ become large in a specified manner. As a consequence, we find that the properly scaled homozygosities have asymptotical normal behavior. In particular, there is a certain phase transition with the limit depending on the relative strength of σ and θ .

Keywords: Poisson–Dirichlet distribution; two-parameter Poisson–Dirichlet distribution; homozygosity; GEM representation; selection

2000 Mathematics Subject Classification: Primary 60F05
Secondary 92D10

1. Introduction

In 1975 Kingman *et al.* [9] introduced the Poisson–Dirichlet distribution denoted by Π_θ to describe allele frequencies in descending order with parameter $\theta > 0$. It is a probability measure on the ordered infinite simplex

$$\nabla_\infty = \left\{ (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, \sum x_i = 1 \right\}.$$

It plays an important role as the stationary distribution of the neutral infinite-alleles model. We refer the reader to [8] for more literature on this distribution. In population genetics, the parameter $\theta > 0$ reflects the mutation rate and $\theta = 4N_e u$, where u is the individual mutation rate and N_e is the effective population size. For convenience, throughout the paper, we call θ the mutation rate.

Applying a size-biased sampling procedure to the random frequencies $\mathbf{X} = (X_1, X_2, \dots)$ with the Poisson–Dirichlet distribution, we can obtain the GEM representation $(\tilde{X}_1, \tilde{X}_2, \dots)$ of the Poisson–Dirichlet distribution. The GEM distribution is defined as follows. Let V_i , $i = 1, 2, \dots$, be a sequence of independent, identically distributed (i.i.d.) random variables with Beta(1, θ) density $f(v) = \theta(1-v)^{\theta-1} \mathbf{1}_{(0,1)}(v)$. A size-biased permutation of $(X_i)_1^\infty$ is given by $\tilde{X}_1 = V_1$, $\tilde{X}_k = (1 - V_1) \cdots (1 - V_{k-1}) V_k$, $k \geq 2$, which is called the GEM representation. By ordering $(\tilde{X}_i)_1^\infty$ in descending order we can retrieve the $(X_i)_1^\infty$. In 1997 Pitman and Yor [12] generalized the GEM distribution to the two-parameter case in which the distribution of V_i is Beta(1 - α , $\theta + i\alpha$). Ranking the two-parameter GEM in descending order gives the two-parameter Poisson–Dirichlet distribution denoted by $\Pi_{\theta,\alpha}$ for $0 < \alpha < 1$ and $\theta > -\alpha$.

Received 18 February 2009; revision received 3 May 2009.

* Postal address: Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1. Email address: xuf23@math.mcmaster.ca

Besides mutation, selection is another important factor in the population evolution. In the infinite-alleles model with overdominant selection intensity $\sigma > 0$, the stationary distribution denoted by Π_θ^σ has been shown to be absolutely continuous with the Poisson–Dirichlet distribution Π_θ . The corresponding Radon–Nikodym derivative is given explicitly as a special case in [2]:

$$\frac{d\Pi_\theta^\sigma}{d\Pi_\theta}(\mathbf{x}) = \frac{\exp\{-\sigma H_2(\mathbf{x})\}}{C(\sigma, \theta)},$$

where the population homozygosity

$$H_2(\mathbf{x}) = H(x_1, x_2, \dots) = \sum_{i=1}^{\infty} x_i^2, \quad \mathbf{x} = (x_1, x_2, \dots) \in \nabla,$$

and $C(\sigma, \theta) = \int_{\nabla} \exp\{-\sigma H_2(\mathbf{x})\} \Pi_\theta(d\mathbf{x})$ is a normalizing constant. In this paper we generalize the selection density to the two-parameter case. For convenience, we name it the two-parameter Poisson–Dirichlet distribution with selection though the two-parameter Poisson–Dirichlet distribution has no direct application in population genetics. Define the two-parameter Poisson–Dirichlet distribution with selection, denoted by $\Pi_{\theta,\alpha}^\sigma$, as

$$\frac{d\Pi_{\theta,\alpha}^\sigma}{d\Pi_{\theta,\alpha}}(\mathbf{x}) = \frac{\exp\{-\sigma H_m(\mathbf{x})\}}{C(\sigma, \theta, \alpha)},$$

where $H_m(\mathbf{x}) = \sum_{i=1}^{\infty} x_i^m$, $\mathbf{x} = (x_1, x_2, \dots) \in \nabla$, $m = 2, 3, \dots$, denotes the m th-order homozygosity, and $C(\sigma, \theta, \alpha) = \int_{\nabla} \exp\{-\sigma H_m(\mathbf{x})\} \Pi_{\theta,\alpha}(d\mathbf{x})$.

Population genetics applications have motivated much study of the limiting behavior of the one-parameter model when the mutation rate θ goes to ∞ . In this setting, θ is the scaled population mutation rate and large θ corresponds to a large population size. Relevant results include the limit theorem of Griffiths [4], the associated large deviation estimates of Dawson and Feng [1], and the central limit theorem obtained by Joyce *et al.* [5].

The Poisson–Dirichlet distribution and its two-parameter counterpart share much structural similarities, including the GEM representation and the sampling formula (see [11]). It is natural to ask if we can establish similar limiting theorems for the two-parameter model. These results include the large deviation principle in [3] and a comprehensive study of the two-parameter Poisson–Dirichlet distribution in [7].

This paper is inspired by the central limit theorem obtained for the one-parameter model in [6]. There the authors showed that the limiting behavior of the infinite-alleles model with overdominant selection is similar to the neutral infinite-alleles model when both the mutation rate and the selection rate become large. The goal of this paper is to find the limiting distribution of the properly scaled p th population homozygosity under the two-parameter Poisson–Dirichlet distribution with selection density as θ approaches ∞ . In particular, letting the selection intensity σ be some power of the mutation rate leads to similar results, which indicates that both the large selection intensity and the large mutation behave like the large mutation alone. Also, there exists a certain phase transition which depends on the relative strength of the mutation rate and the selection intensity. We use the approach of proving the convergence of the characteristic function. The main difficulty is obtaining the uniform integrability of certain exponential functions of homozygosity. This property comes intuitively from the corresponding result in the one-parameter case. But the two-parameter GEM loses the identical property. We solve this problem via decomposing the two-parameter Poisson–Dirichlet distribution $\Pi_{\theta,\alpha}$

into a random variable with Poisson–Dirichlet distribution Π_θ and an i.i.d. sequence of random variables with distribution $\Pi_{0,\alpha}$.

We can identify from the results obtained that the Poisson–Dirichlet distribution with selection and its two-parameter counterpart have similar asymptotic properties for large θ when α is fixed. It also indicates similar mathematical structure between the Poisson–Dirichlet distribution and its two-parameter counterpart.

To state the main result, we need to introduce some notation and related results.

For $p = 2, 3, \dots$ and $\mathbf{X} = (X_1, X_2, \dots) \sim \Pi_{\theta,\alpha}$, define

$$Z_{p,\theta} = \sqrt{\theta} \left(\frac{\Gamma(1-\alpha)}{\Gamma(p-\alpha)} \theta^{p-1} H_p(\mathbf{X}) - 1 \right), \tag{1.1}$$

where $H_p(\mathbf{X})$ denotes the p th homozygosity of \mathbf{X} . Similarly, for $\mathbf{Y} = (Y_1, Y_2, \dots) \sim \Pi_{\theta,\alpha}^\sigma$, where $\sigma = c\theta^{m-1/2-\beta}$, $m = 2, 3, \dots$, $\beta \geq 0$, and c is a constant, set

$$Z_{p,\theta}^\beta = \sqrt{\theta} \left(\frac{\Gamma(1-\alpha)}{\Gamma(p-\alpha)} \theta^{p-1} H_p(\mathbf{Y}) - 1 \right). \tag{1.2}$$

The main ingredient used in the proof of our theorem is the following central limit theorem for $(Z_{p,\theta})_{p=2}^\infty$ obtained in [7]. We remark that this result can be proved by an argument similar to that given in [5].

Lemma 1.1. *Let $(Z_{2,\alpha}, Z_{3,\alpha}, \dots)$ be multivariate normal with mean 0, and let the covariance of $Z_{i,\alpha}$ and $Z_{j,\alpha}$ be given by*

$$\frac{\Gamma(1-\alpha)\Gamma(i+j-\alpha)}{\Gamma(i-\alpha)\Gamma(j-\alpha)} + \alpha - ij. \tag{1.3}$$

Then, as $\theta \rightarrow \infty$, we have

$$(Z_{2,\theta}, Z_{3,\theta}, \dots) \xrightarrow{D} (Z_{2,\alpha}, Z_{3,\alpha}, \dots).$$

Here and in what follows, ‘ \xrightarrow{D} ’ denotes convergence in distribution. The next theorem is the main result of this paper.

Theorem 1.1. *Suppose that $\mathbf{Y} = (Y_1, Y_2, \dots) \sim \Pi_{\theta,\alpha}^\sigma$ and $\sigma = c\theta^{m-1/2-\beta}$, where $\beta \geq 0$, $c > 0$ is a constant, and $m = 2, 3, \dots$. Let $(Z_{2,\theta}^\beta, Z_{3,\theta}^\beta, \dots)$ be defined as in (1.2), and let $(Z_{2,\alpha}, Z_{3,\alpha}, \dots)$ be given as in Lemma 1.1. As $\theta \rightarrow \infty$,*

$$(Z_{2,\theta}^\beta, Z_{3,\theta}^\beta, \dots) \xrightarrow{D} \begin{cases} (Z_{2,\alpha}^*, Z_{3,\alpha}^*, \dots) & \text{if } \beta = 0, \\ (Z_{2,\alpha}, Z_{3,\alpha}, \dots) & \text{if } \beta > 0, \end{cases}$$

where

$$Z_{p,\alpha}^* = Z_{p,\alpha} - cb_p, \quad b_p = (1-\alpha)_{(p-1)} \text{cov}(Z_{p,\alpha}, Z_{m,\alpha}),$$

and $(1-\alpha)_{(p-1)} = \frac{\Gamma(p-\alpha)}{\Gamma(1-\alpha)} = (1-\alpha) \cdots (p-1-\alpha)$ for $p = 2, 3, \dots$

Theorem 1.1 is proved in Section 3. In Section 2 we present a uniform integrability argument for $\text{exp}\{-tZ_{p,\theta}\}$. We also use several results obtained in [6].

2. Uniform integrability

To obtain the uniform integrability of $\exp\{-tZ_{p,\theta}\}$, it suffices to show the following lemma.

Lemma 2.1. *Suppose that, for any fixed $t > 0$ and $p \geq 2$, $Z_{p,\theta}$ is defined as in (1.1). Then there exists $\theta_0 = \theta_0(t) < \infty$ such that*

$$\sup_{\theta > \theta_0} E(\exp\{-tZ_{p,\theta}\}) < \infty.$$

We first need to introduce an important result of [10] in the following lemma.

Lemma 2.2. ([10].) *Suppose that (X'_n) is given by the Poisson–Dirichlet distribution with parameter $\theta > 0$. Independent of (X'_n) , let (U_{mn}) , $m = 1, 2, \dots$, be a sequence of independent copies of (U_n) which has the two-parameter Poisson–Dirichlet distribution with $\theta = 0$ and $\alpha > 0$, i.e. $\Pi_{0,\alpha}$. Let (X_n) be defined by ranking the collection of products $\{X'_m U_{mn}, m, n \in \mathbb{N}\}$. Then (X_n) has the two-parameter Poisson–Dirichlet distribution with parameters θ and α , i.e. $\Pi_{\theta,\alpha}$.*

In virtue of the above representation, we can rewrite the homozygosity corresponding to the two-parameter Poisson–Dirichlet distribution as well as $Z_{p,\theta}$ in a form that keeps the i.i.d. property.

Suppose that W_m , $m = 1, 2, \dots$, is a sequence of i.i.d. random variables as copies of $\sum_{n=1}^\infty U_n^p$ for fixed $p \geq 2$. Therefore, $0 < W_m \leq 1$ and its moments depend on α alone.

For any $p = 2, 3, \dots$, by the GEM representation and the exchangeability of H_p as a function of $(X_n)^\infty_1$, we have

$$H_p = \sum_{n=1}^\infty X_n^p = \sum_{m,n=1}^\infty (X'_m)^p U_{mn}^p \stackrel{D}{=} \sum_{m=1}^\infty (X'_m)^p W_m = V_1^p W_1 + (1 - V_1)^p \tilde{H}_p,$$

where $V_1 \sim \text{Beta}(1, \theta)$ and \tilde{H}_p is a random variable with the same distribution as H_p and independent of V_1 and W_1 . Here ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution.

Since

$$Z_{p,\theta} = \sqrt{\theta} \left(\frac{\Gamma(1 - \alpha)}{\Gamma(p - \alpha)} \theta^{p-1} H_p - 1 \right) = \sqrt{\theta} \left(\frac{\theta^{p-1}}{(1 - \alpha)_{(p-1)}} H_p - 1 \right),$$

we deduce that

$$\begin{aligned} (1 - \alpha)_{(p-1)} Z_{p,\theta} &= \sqrt{\theta} (\theta^{p-1} H_p - (1 - \alpha)_{(p-1)}) \\ &\stackrel{D}{=} \sqrt{\theta} (\theta^{p-1} V_1^p W_1 + \theta^{p-1} (1 - V_1)^p \tilde{H}_p - (1 - \alpha)_{(p-1)}) \\ &= \sqrt{\theta} (\theta^{p-1} V_1^p W_1 + (1 - V_1)^p (1 - \alpha)_{(p-1)} - (1 - \alpha)_{(p-1)}) \\ &\quad + (1 - V_1)^p (\theta^{p-1} \tilde{H}_p - (1 - \alpha)_{(p-1)}) \\ &= f_p(V_1, W_1) + (1 - \alpha)_{(p-1)} (1 - V_1)^p \tilde{Z}_{p,\theta}, \end{aligned}$$

where

$$f_p(V_1, W_1) = \sqrt{\theta} (\theta^{p-1} V_1^p W_1 - (1 - \alpha)_{(p-1)} (1 - (1 - V_1)^p)), \tag{2.1}$$

and

$$\tilde{Z}_{p,\theta} = \sqrt{\theta} \left(\frac{\theta^{p-1}}{(1 - \alpha)_{(p-1)}} \tilde{H}_p - 1 \right),$$

with the same distribution as $Z_{p,\theta}$, is independent of V_1 and W_1 . Since $0 < 1 - (1 - V_1)^p \leq 2^p V_1$ and $0 < W_1 \leq 1$,

$$\frac{1}{\sqrt{\theta}} (\theta^p V_1^p W_1 - (1 - \alpha)_{(p-1)} 2^p \theta V_1) \leq f_p \leq \frac{1}{\sqrt{\theta}} (\theta^p V_1^p W_1 + (1 - \alpha)_{(p-1)} 2^p \theta V_1). \tag{2.2}$$

Also,

$$|f_p(V_1, W_1)| \leq \frac{1}{\sqrt{\theta}} (\theta^p V_1^p + (1 - \alpha)_{(p-1)} 2^p \theta V_1).$$

For $k \geq 1$,

$$\begin{aligned} E(|f_p(V_1, W_1)|^k) &\leq \theta^{-k/2} E((\theta^p V_1^p + (1 - \alpha)_{(p-1)} 2^p \theta V_1)^k) \\ &\leq \theta^{-k/2} 2^k [E((\theta V_1)^{pk}) + (1 - \alpha)_{(p-1)}^k 2^{pk} E((\theta V_1)^k)] \\ &\leq \left(\frac{2}{\sqrt{\theta}}\right)^k [(pk)! + (1 - \alpha)_{(p-1)}^k 2^{pk} k!] \\ &\sim O(\theta^{-k/2}), \end{aligned}$$

where in the last step we used the fact that $0 \leq E((\theta V_1)^j) = j! \theta^j / (\theta + 1) \cdots (\theta + j) \leq j!$, $j \geq 1$.

Next we give the estimation with respect to f_p .

Lemma 2.3. For $p \geq 2$, let $f_p(V_1, W_1)$ be defined as in (2.1), where $V_1 \sim \text{Beta}(1, \theta)$, and W_1 has the same distribution as the p th homozygosity corresponding to allele frequencies with distribution $\Pi_{0,\alpha}$. Then, for $\theta > 1$ and each $j \geq 1$, there is a positive function $g_{p,j}(t, W_1)$ increasing in t such that, for all $t > 0$,

$$\left| \sum_{k=j}^{\infty} \frac{(-t f_p)^k}{k!} \right| \leq t^j g_{p,j}(t, W_1) |f_p|^j.$$

Proof. Note that

$$\left| \sum_{k=j}^{\infty} \frac{(-t f_p)^k}{k!} \right| = |t f_p|^j \left| \sum_{k=j}^{\infty} \frac{(-t f_p)^{k-j}}{k!} \right| = t^j |f_p|^j g_j(t f_p),$$

where

$$0 < g_j(x) = \sum_{k=j}^{\infty} \frac{(-x)^{k-j}}{k!} = \begin{cases} \frac{e^{-x} - \sum_{k=0}^{j-1} (-x)^k / k!}{(-x)^j}, & x \neq 0, \\ \frac{1}{j!}, & x = 0. \end{cases}$$

Define $f(x) = W_1 x^p - (1 - \alpha)_{(p-1)} 2^p x$. Hence, f has a lower bound on $[0, \infty)$. Let $x^* = ((1 - \alpha)_{(p-1)} 2^p / W_1 p)^{1/(p-1)}$ be the point in $[0, \infty)$ at which $f(x)$ achieves its minimum, that is,

$$f(x^*) = W_1^{-1/(p-1)} \left(\frac{(1 - \alpha)_{(p-1)} 2^p}{p} \right)^{p/(p-1)} (1 - p) < 0.$$

By (2.2),

$$f_p \geq \frac{f(\theta V_1)}{\sqrt{\theta}} \geq \frac{f(x^*)}{\sqrt{\theta}} > f(x^*) \quad \text{for } \theta > 1.$$

Since $g_j(x)$ is a decreasing function of $x \in (-\infty, \infty)$, $g_j(t f_p) \leq g_j(t f(x^*)) =: g_{p,j}(t, W_1)$ for all $t > 0$.

The next lemma shows that the moments of $Z_{p,\theta}$ are uniformly bounded for large θ .

Lemma 2.4. *Let $Z_{p,\theta}$ be defined as in (1.1). Then, for all integers $r \geq 1$ and $p \geq 2$,*

$$\sup_{\theta > 1} E(|Z_{p,\theta}|^r) < \infty.$$

Proof. Suppose that $(X'_1, X'_2, \dots) \sim \Pi_\theta$. Set $H'_p = \sum_n (X'_n)^p$, and set

$$Z'_{p,\theta} = \sqrt{\theta} \left(\frac{\theta^{p-1}}{(p-1)!} H'_p - 1 \right).$$

By the sampling formulae of the Poisson–Dirichlet distribution and the two-parameter Poisson–Dirichlet distribution, we know that $E((Z'_{p,\theta})^{2r})$ and $E((Z_{p,\theta})^{2r})$ have the same highest finite order of θ . Thus,

$$\lim_{\theta \rightarrow \infty} \frac{E((Z_{p,\theta})^{2r})}{E((Z'_{p,\theta})^{2r})} = k < \infty.$$

Using the result obtained in Lemma 5 of [6], we know that

$$\sup_{\theta > 1} E(|Z'_{p,\theta}|^r) < \infty.$$

Since $E(|Z_{p,\theta}|^r) \leq \sqrt{E((Z_{p,\theta})^{2r})}$, it follows that $\sup_{\theta > 1} E(|Z_{p,\theta}|^r) < \infty$.

Now we are ready to give the proof of Lemma 2.1.

Proof of Lemma 2.1. For each integer r , define

$$\tilde{\theta}_r = (1 - \alpha)_{(p-1)} \tilde{Z}_{p,\theta} \quad \text{and} \quad \mu_{r,\theta} = (1 - \alpha)_{(p-1)}^r E(Z_{p,\theta}^r).$$

Since $\sup_{\theta > 1} E(|Z_{p,\theta}|^r) < \infty$ for each r , we only need to show that, for $t > 0$, there exist finite values $l = l(t)$ and $\theta_0 = \theta_0(l, t)$ such that

$$\sup_{\theta > \theta_0} \left| \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \right| < \infty.$$

It is obvious that

$$\begin{aligned} [(1 - \alpha)_{(p-1)} Z_{p,\theta}]^r &\stackrel{D}{=} (f_p(V_1, W_1) + (1 - \alpha)_{(p-1)}(1 - V_1)^p \tilde{Z}_{p,\theta})^r \\ &= \sum_{k=0}^r \binom{r}{k} f_p^k(V_1, W_1) [(1 - \alpha)_{(p-1)} \tilde{Z}_{p,\theta}]^{r-k} (1 - V_1)^{p(r-k)} \\ &= (1 - V_1)^{pr} [(1 - \alpha)_{(p-1)} \tilde{Z}_{p,\theta}]^r \\ &\quad + \sum_{k=1}^r \binom{r}{k} f_p^k(V_1, W_1) (1 - V_1)^{p(r-k)} [(1 - \alpha)_{(p-1)} \tilde{Z}_{p,\theta}]^{r-k}. \end{aligned}$$

Taking the expectation of both sides and using independence, we have

$$\mu_{r,\theta} = E((1 - V_1)^{pr}) \mu_{r,\theta} + \sum_{k=1}^r \binom{r}{k} E(f_p^k(V_1, W_1) (1 - V_1)^{p(r-k)}) \mu_{r-k,\theta}.$$

Thus,

$$\mu_{r,\theta} = \frac{pr + \theta}{pr} \sum_{k=1}^r \binom{r}{k} \mathbb{E}(f_p^k(V_1, W_1)(1 - V_1)^{p(r-k)})\mu_{r-k,\theta}.$$

Define two sequences of random variables:

$$M_k := \frac{f_p^k}{k!}, \quad N_l := \frac{(\tilde{Z}_\theta(1 - V_1)^p)^l}{l!}.$$

Thus,

$$\begin{aligned} \left| \sum_{r=l}^\infty \frac{(-t)^r}{r!} \mu_{r,\theta} \right| &= \left| \sum_{r=l}^\infty \frac{pr + \theta}{pr} \frac{(-t)^r}{r!} \sum_{k=1}^r \binom{r}{k} \mu_{r-k,\theta} \mathbb{E}(f_p^k(V_1, W_1)(1 - V_1)^{p(r-k)}) \right| \\ &= \left| \mathbb{E} \left(\sum_{r=l}^\infty \sum_{k=1}^r \left(1 + \frac{\theta}{pr} \right) M_k N_{r-k} (-t)^r \right) \right| \\ &\leq \left| \mathbb{E} \left(\sum_{r=l}^\infty \sum_{k=1}^r M_k N_{r-k} (-t)^r \right) \right| + \left| \mathbb{E} \left(\sum_{r=l}^\infty \sum_{k=1}^r \frac{\theta}{pr} M_k N_{r-k} (-t)^r \right) \right| \\ &=: |E(A_1)| + |E(A_2)|. \end{aligned}$$

Using the inequality in Lemma 6 of [6] and interchanging the order of summation, we obtain

$$|A_1| \leq \left| \sum_{j=1}^\infty (-t)^j M_j \sum_{q=0}^\infty (-t)^q N_q \right| + t K_{1l}(t),$$

where

$$K_{1l}(t) := \sum_{j=1}^{l-1} |M_j| t^{j-1} \sum_{i=0}^{l-j-1} |N_i| t^i$$

is an increasing function of t . By Lemma 4 we obtain

$$\left| \sum_{j=1}^\infty (-t)^j M_j \right| = \left| \sum_{j=1}^\infty \frac{(-t f_p)^j}{j!} \right| \leq t g_{p,1}(t, W_1) |f_p|$$

and

$$\begin{aligned} \left| \sum_{q=0}^\infty (-t)^q N_q \right| &= \exp\{-t \tilde{Z}_\theta(1 - V_1)^p\} \\ &\leq \max\{1, \exp\{-t \tilde{Z}_\theta\}\} \\ &\leq 1 + \exp\{-t \tilde{Z}_\theta\}. \end{aligned}$$

Hence,

$$|A_1| \leq t g_{p,1}(t, W_1) |f_p| (1 + \exp\{-t \tilde{Z}_\theta\}) + t K_{1l}(t).$$

It follows that

$$\begin{aligned} |E(A_1)| &\leq E(|A_1|) \\ &\leq t E(g_{p,1}(t, W_1) |f_p|) + t E(f_{p,1}(t, W_1) |f_p|) E(\exp\{-t \tilde{Z}_\theta\}) + t E(K_{1l}(t)). \end{aligned}$$

Since $E(|f_p(V_1, W_1)|^k) \sim O(\theta^{-k/2})$ and $E(g_{p,1}(t, W_1))$ depends only on α , there exists $\theta_0 = \theta_0(t) < \infty$ such that $t E(g_{p,1}(t, W_1)|_{p|}) < \frac{1}{4}$ for all $\theta > \theta_0(t)$ by Hölder’s inequality. In addition, $K_{1l}(t)$ is a sum of a finite number of random variables, and note that the mean of each random variable is bounded for all θ . Thus, $\sup_{\theta > 1} E(K_{1l}(t)) < \infty$. Set $\alpha_1(t) = \sup_{\theta > 1} E(K_{1l}(t)) + \frac{1}{4}$. Then, for $\theta > \theta_0(t)$,

$$|E(A_1)| \leq \frac{1}{4} E(\exp\{-t\tilde{Z}_\theta\}) + \alpha_1(t) = \frac{1}{4} \sum_{r=0}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} + \alpha_1(t).$$

We also have the same inequality for A_2 :

$$|E(A_2)| \leq \frac{1}{4} \sum_{r=0}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} + \alpha_2(t).$$

Since A_2 involves the factor θ/pr , it needs a more delicate estimation, but the argument is similar to that of A_1 . The reader can find a similar detailed proof in Lemma 9 of [6], which we will not repeat here.

By the inequalities for A_1 and A_2 above we have

$$\left| \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \right| \leq \frac{1}{2} \sum_{r=0}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} + \alpha_1(t) + \alpha_2(t).$$

Thus,

$$\left| \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \right| - \frac{1}{2} \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \leq \frac{1}{2} \sum_{r=0}^{l-1} \frac{t^r}{r!} |\mu_{r,\theta}| + \alpha_1(t) + \alpha_2(t).$$

It follows that

$$\frac{1}{2} \left| \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \right| \leq \frac{1}{2} \sum_{r=0}^{l-1} \frac{t^r}{r!} |\mu_{r,\theta}| + \alpha_1(t) + \alpha_2(t).$$

Therefore, $|\sum_{r=l}^{\infty} (-t)^r \mu_{r,\theta}/r!|$ is bounded by the sum of a finite number of terms, which are uniformly bounded for θ . It can be concluded that

$$\sup_{\theta > \theta_0} \left| \sum_{r=l}^{\infty} \frac{(-t)^r}{r!} \mu_{r,\theta} \right| < \infty.$$

3. Proof of Theorem 1.1

For $\sigma = c\theta^{m-1/2-\beta}$, $\beta \geq 0$ and $m = 2, 3, \dots$, we know that the characteristic function of $Z_{p,\theta}^\beta$ can be calculated as follows:

$$\begin{aligned} \psi_p(x) &= E(\exp\{ix Z_{p,\theta}^\beta\}) \\ &= \frac{E(\exp\{ix Z_{p,\theta}\} \exp\{-\sigma H_m\})}{E(\exp\{-\sigma H_m\})} \\ &= \frac{E(\exp\{ix Z_{p,\theta} - c(1 - \alpha)_{(m-1)}\theta^{-\beta} Z_{m,\theta}\})}{E(\exp\{-c(1 - \alpha)_{(m-1)}\theta^{-\beta} Z_{m,\theta}\})}. \end{aligned}$$

By Lemma 1.1, $(Z_{2,\theta}, Z_{3,\theta}, \dots) \xrightarrow{D} (Z_{2,\alpha}, Z_{3,\alpha}, \dots)$ as θ goes to ∞ , where $(Z_{2,\alpha}, Z_{3,\alpha}, \dots)$ is multivariate normal with mean 0 and the covariance matrix of $Z_{i,\alpha}$ and $Z_{j,\alpha}$ is given by (1.3).

For $\beta = 0$, combining the uniform integrability of $\exp\{-tZ_{p,\theta}\}$ with $Z_{p,\theta} \xrightarrow{D} Z_{p,\alpha}$ for any $p = 2, 3, \dots$ we have

$$E(\exp\{-c(1 - \alpha)_{(m-1)}Z_{m,\theta}\}) \rightarrow E(\exp\{-c(1 - \alpha)_{(m-1)}Z_{m,\alpha}\}) \quad \text{as } \theta \rightarrow \infty.$$

Therefore,

$$E(\exp\{ixZ_{p,\theta} - c(1 - \alpha)_{(m-1)}Z_{m,\theta}\}) \rightarrow E(\exp\{ixZ_{p,\alpha} - c(1 - \alpha)_{(m-1)}Z_{m,\alpha}\}).$$

It follows that, when θ goes to ∞ ,

$$\psi_p(x) \rightarrow \exp\left\{-\frac{x^2}{2} \text{var } Z_{p,\alpha} - icx(1 - \alpha)_{(m-1)} \text{cov}(Z_{p,\alpha}, Z_{m,\alpha})\right\}.$$

Define $b_p = (1 - \alpha)_{(m-1)} \text{cov}(Z_{p,\alpha}, Z_{m,\alpha})$, and let $Z_{p,\alpha}^* = Z_{p,\alpha} - cb_p$. It follows that

$$Z_{p,\theta}^\beta \xrightarrow{D} Z_{p,\alpha}^* \quad \text{for } p = 2, 3, \dots \text{ as } \theta \rightarrow \infty.$$

Using the same method to calculate the characteristic function for a finite linear combination of $Z_{p,\theta}$, we obtain

$$(Z_{2,\theta}^\beta, Z_{3,\theta}^\beta, \dots) \xrightarrow{D} (Z_{2,\alpha}^*, Z_{3,\alpha}^*, \dots) \quad \text{as } \theta \rightarrow \infty.$$

For $\beta > 0$ and $\theta > 1$,

$$\exp\{-c(1 - \alpha)_{(m-1)}\theta^{-\beta}Z_{m,\theta}\} \leq \max\{\exp\{-c(1 - \alpha)_{(m-1)}Z_{m,\theta}\}, 1\}.$$

Using the dominated convergence theorem, we have

$$E(\exp\{-c(1 - \alpha)_{(p-1)}\theta^{-\beta}Z_{q,\theta}\}) \rightarrow 1 \quad \text{as } \theta \rightarrow \infty.$$

Therefore, $(Z_{2,\theta}^\beta, Z_{3,\theta}^\beta, \dots) \xrightarrow{D} (Z_{2,\alpha}, Z_{3,\alpha}, \dots)$ as $\theta \rightarrow \infty$.

Acknowledgement

The author would like to thank the anonymous referee for his/her valuable suggestions and comments.

References

- [1] DAWSON, D. A. AND FENG, S. (2006). Asymptotic behavior of the Poisson–Dirichlet distribution for large mutation rate. *Ann. Appl. Prob.* **16**, 562–582.
- [2] ETHIER, S. N. AND KURTZ, T. G. (1994). Convergence to Fleming–Viot processes in the weak atomic topology. *Stoch. Process. Appl.* **54**, 1–27.
- [3] FENG, S. (2007). Large deviations for Dirichlet processes and Poisson–Dirichlet distribution with two parameters. *Electron. J. Prob.* **12**, 787–807.
- [4] GRIFFITHS, R. C. (1979). On the distribution of allele frequencies in a diffusion model. *Theoret. Pop. Biol.* **15**, 140–158.
- [5] JOYCE, P., KRONE, S. M. AND KURTZ, T. G. (2002). Gaussian limits associated with the Poisson–Dirichlet distribution and the Ewens sampling formula. *Ann. Appl. Prob.* **12**, 101–124.

- [6] JOYCE, P., KRONE, S. M. AND KURTZ, T. G. (2003). When can one detect overdominant selection in the infinite-alleles model? *Ann. Appl. Prob.* **13**, 181–212.
- [7] KENJI, H. (2009). The two-parameter Poisson–Dirichlet point process. To appear in *Bernoulli*.
- [8] KINGMAN, J. F. C. (1977). The population structure associated with the Ewens sampling formula. *Theoret. Pop. Biol.* **11**, 274–283.
- [9] KINGMAN, J. F. C. *et al.* (1975). Random discrete distributions. *J. R. Statist. Soc. B* **37**, 1–22.
- [10] PITMAN, J. (1992). The two-parameter generalization of Ewens’ random partition structure. Tech. Rep. 345, Department of Statistics, University of California, Berkeley.
- [11] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Prob. Theory Relat. Fields* **102**, 145–158.
- [12] PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25**, 855–900.