

# THE SAMPLING FORMULA AND LAPLACE TRANSFORM ASSOCIATED WITH THE TWO-PARAMETER POISSON–DIRICHLET DISTRIBUTION

FANG XU,\* *McMaster University*

## Abstract

In this paper we investigate the relationship between the sampling formula and Laplace transform associated with the two-parameter Poisson–Dirichlet distribution. We conclude that they are equivalent to determining the corresponding infinite-dimensional distribution. With these tools, a central limit theorem is established associated with the infinitely-many-neutral-alleles model at any fixed time. We also obtain the probability generating function of random sampling from a generalized two-parameter diffusion process. At the end of the paper a selection case is considered.

*Keywords:* Poisson–Dirichlet distribution; two-parameter Poisson–Dirichlet distribution; sampling formula; Laplace transform

2010 Mathematics Subject Classification: Primary 60C05

Secondary 60F05

## 1. Introduction

In this paper we start with the random partition structure associated with the two-parameter Poisson–Dirichlet distribution which is defined in the following way. For  $0 \leq \alpha < 1$  and  $\theta > -\alpha$ , let  $U_k$ ,  $k = 1, 2, \dots$ , be a sequence of independent random variables such that  $U_k$  has a Beta( $1 - \alpha$ ,  $\theta + k\alpha$ ) distribution. Define an infinite sequence  $\tilde{V}_1, \tilde{V}_2, \dots$  via a stick-breaking process such that

$$\tilde{V}_1 = U_1, \quad \tilde{V}_n = (1 - U_1) \cdots (1 - U_{n-1})U_n, \quad n \geq 2.$$

Rank  $(\tilde{V}_1, \tilde{V}_2, \dots)$  in decreasing order, and let  $V_1 \geq V_2 \geq \dots$ . The distributions of  $(\tilde{V}_1, \tilde{V}_2, \dots)$  and  $(V_1, V_2, \dots)$  are called the two-parameter Griffiths–Engen–McCloskey (GEM) distribution and the two-parameter Poisson–Dirichlet distribution, denoted by  $\text{GEM}(\alpha, \theta)$  and  $\text{PD}(\alpha, \theta)$ , respectively.  $\text{PD}(\alpha, \theta)$  is a family of distributions generalized from the well-known Poisson–Dirichlet distribution with  $\alpha = 0$ . The distribution was introduced by Kingman *et al.* [14] and has been associated with a number of models in different areas, especially in population genetics. We refer the reader to [5] and the references therein for more background and applications of the family of distributions. Hereinafter, we denote  $\text{PD}(0, \theta)$  simply by  $\text{PD}(\theta)$ .

The two-parameter Poisson–Dirichlet distribution first appeared in Perman *et al.* [15] in 1992. We refer the reader to Pitman and Yor [19] for a comprehensive survey of  $\text{PD}(\alpha, \theta)$  from the perspective of subordination. Recently, Handa [11] systematically studied the two-parameter family by means of random point processes. In fact, the  $\text{PD}(\alpha, \theta)$  distribution can

Received 1 April 2011; revision received 16 June 2011.

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

be regarded as the distribution of a random mass partition on the infinite-dimensional ordered simplex

$$\nabla_\infty := \left\{ \mathbf{x} = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, \sum_{i=1}^\infty x_i = 1 \right\}.$$

The partition probability functions of the exchangeable random partition associated with  $\text{PD}(\theta)$  and  $\text{PD}(\alpha, \theta)$  are obtained as the well-known Ewens' sampling formula and Pitman's sampling formula. In fact,  $\text{PD}(\theta)$  describes the equilibrium of the infinitely-many-neutral-alleles (IMNA) model with mutation rate  $\theta > 0$  (cf. [4]). That is,  $\text{PD}(\theta)$  is the stationary distribution of frequencies of different alleles that are ordered according to their family sizes. Consider a random sample of  $n$  individuals from the population. The composition of the sample can be given by a partition vector  $\mathbf{n} = (n_1, \dots, n_k)$  of integer  $n$ , which indicates  $k$  different types of allele in the sample and  $n_i$  individuals of each type. In the literature, the partition structure of  $n$  samples is often denoted by a different but equivalent vector  $\mathbf{a} = (a_1, \dots, a_n)$ , where  $a_i$  is the number of families containing  $i$  individuals. Therefore, for each  $i$ ,  $a_i \geq 0$  and  $\sum_{i=1}^n i a_i = n$ . Given  $\mathbf{x} = (x_1, x_2, \dots) \in \nabla_\infty$ , where  $x_i$  denotes the frequency of the  $i$ th largest family in the population, the conditional probability of a random sample of size  $n$  with partition structure  $A_n = \mathbf{a}$  is expressed as

$$P(A_n = \mathbf{a} \mid \mathbf{x} = (x_1, x_2, \dots)) = \frac{n!}{\prod_{i=1}^n (i!)^{a_i}} \sum_{\mathbf{n}} x_1^{n_1} x_2^{n_2} \dots,$$

where the summation is over the set  $\{\mathbf{n} = (n_1, n_2, \dots) : \# \{i : n_i = j\} = a_j, n_i \geq 0\}$ . When  $x \sim \text{PD}(\theta)$ , the probability

$$P(A_n = \mathbf{a}) = \int_{\nabla_\infty} P(A_n = \mathbf{a} \mid \mathbf{x} = (x_1, x_2, \dots)) \text{PD}(\theta)(d\mathbf{x})$$

is given by Ewens' sampling formula (ESF):

$$P(A_n = \mathbf{a}) = \frac{n!}{\theta_{(n)}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{a_j} \frac{1}{a_j!}. \tag{1.1}$$

Here  $\theta_{(n)} = \theta(\theta + 1) \cdots (\theta + n - 1)$ . In the two-parameter case, i.e.  $x \sim \text{PD}(\alpha, \theta)$ , we have the generalized formula obtained by Pitman [17], known as Pitman's sampling formula (PSF):

$$P(A_n = \mathbf{a}) = \frac{n!}{\theta_{(n)}} \prod_{\ell=0}^{k-1} (\theta + \ell\alpha) \prod_{j=1}^n \frac{(1 - \alpha)_{(j-1)}^{a_j}}{(j!)^{a_j} (a_j!)}. \tag{1.2}$$

As  $\text{PD}(\theta)$  plays an important role in population genetics, the sampling formula has wide applications, including the estimation of the mutation rate  $\theta$  and the comparison of different models, such as the neutral and nonneutral models. Here, neutral means that mutation is considered only in the evolution. Moreover, the proof of the sampling formula also inspires recursive constructions of some well-known partition structures, such as Hoppe's urn model, Blackwell–MacQueen's urn, and the Chinese restaurant process. See, e.g. [18], which also contains applications of the sampling formula in Bayesian statistics. In this paper we display an alternative derivation of the sampling formula from the Laplace transform of the  $\text{PD}(\alpha, \theta)$  distribution. Moreover, we show that the Laplace transform is also determined uniquely by

the sampling formula. This implies that they both equivalently characterize the  $PD(\alpha, \theta)$  distribution .

In Pitman and Yor [19] and Handa [11], the Laplace transform of the two-parameter Poisson–Dirichlet distribution is obtained as a probability generating functional of a random vector  $\mathbf{x}$  with the  $PD(\alpha, \theta)$  distribution. As shown in [11], the Laplace transform leads to some fundamental results for  $PD(\alpha, \theta)$ , such as the joint probability density, the moment formula, the asymptotic behavior, etc. By Campbell’s theorem from the Poisson point processes (see, e.g. [13]), the particular form of this Laplace transform can be interpreted as the connection between  $PD(\theta)$  and  $PD(\alpha, \theta)$  with some pure-jump subordinators. In [11] the key to deriving the Laplace transform was the correlation measure of the two-parameter point process, which was obtained based on the equivalence of the size-biased permutations for  $PD(\alpha, \theta)$  and  $GEM(\alpha, \theta)$ . In this paper we take an alternative approach from the sampling formula (1.2). This can be regarded as an interpretation of Kingman’s conclusion that the partition function and random mass partition  $PD(\alpha, \theta)$  are uniquely determined by each other.

The paper is organized as follows. In Section 2 we introduce the necessary definitions and notation. We gather the above results and their proofs in Section 3. In Section 4 we consider the Laplace transform for the IMNA model. The result leads us to establish a central limit theorem associated with homozygosity functions at any time  $t > 0$  as  $\theta$  goes to  $\infty$  (recall that  $\theta$  denotes the mutation rate). Moreover, the limiting behavior is the same as that found for the  $PD(\theta)$  distribution. The latter case has been studied by several authors, including Joyce *et al.* [12] and Handa [11]. Our conclusion indicates that the population distribution at a finite time is almost the same as the stationary state when mutation becomes extremely strong. We also consider a generalized two-parameter infinitely-many-alleles diffusion model. The two-parameter diffusion  $X_{\alpha, \theta}(t)$  is constructed by different means in [6] and [16]. Here we compute the probability generating function of random sampling from  $X_{\alpha, \theta}(t)$ , which can be seen as a two-parameter generalization of Theorem 4.3 of [9]. The proof of [9, Theorem 4.3] relied on the approximation of IMNA diffusion processes by the finite-alleles Wright–Fisher model. However, a finite-dimensional version of the two-parameter model is yet to be found. Thanks to the correlation measure, we find the corresponding two-parameter result, which makes it possible for us to compare the sampling formula at a finite time with the sampling formula at the stationary state.

Finally, in Section 5 we consider a nonneutral model incorporating selection. In the classic Wright–Fisher model with selection, the stationary distribution is shown to possess a density  $e^{sH_2(\mathbf{X})}/C$  with respect to the neutral case, where  $H_2 = \sum X_i^2$  is the homozygosity function,  $s \in \mathbb{R}$  denotes the selection intensity, and  $C$  is the normalization constant. The density follows from the assumption that all the homozygotes (pairs of alleles of the same type) have equal fitness  $1 + s$  and all the heterozygotes (pairs of alleles with different types) have fitness 1. In the infinite-alleles model with selection, the stationary distribution  $PD^{s,q}(\theta)$  is absolutely continuous with respect to the Poisson–Dirichlet distribution  $PD(\theta)$ , and its Radon–Nikodym derivative is

$$\frac{e^{sH_q(\mathbf{X})}}{C}, \quad \mathbf{X} = (X_1, X_2, \dots) \in \nabla_\infty,$$

where  $H_q(\mathbf{X}) = \sum X_i^q$ ,  $q = 2, 3, \dots$ , denotes the  $q$ th order homozygosity. Given this distribution, Handa [10] considered the more general partition structure

$$P_{\theta, s, q}(\mathbf{A}_n = \mathbf{a}) = \frac{E[e^{sH_q(\mathbf{X})}\psi_{\mathbf{a}}(\mathbf{X})]}{E[e^{sH_q(\mathbf{X})}]},$$

where

$$\psi_\alpha(\mathbf{x}) = \frac{n!}{n_1! \cdots n_k! a_1! \cdots a_n!} \sum_{i_1, \dots, i_k \neq} x_{i_1}^{n_1} \cdots x_{i_k}^{n_k} \tag{1.3}$$

and  $\sum_{i_1, \dots, i_k \neq}$  indicates that the sum is taken over all  $k$ -tuples of distinct indices. The expectation is taken with respect to the  $PD(\theta)$  distribution. Because of the conclusion in Section 3, we obtain the correlation function and Laplace transform in the selection case.

### 2. Preliminaries and notation

First we introduce some definitions in the theory of point processes, which can be found in [1]. The random element  $(X_i)_{i=1}^\infty$  governed by the  $PD(\alpha, \theta)$  distribution can be viewed as a random point process  $\xi := \sum_{i=1}^\infty \delta_{X_i}$ , which is called the (two-parameter) Poisson–Dirichlet point process with parameters  $(\alpha, \theta)$ , or simply the  $PD(\alpha, \theta)$  process in [11]. For any positive integer  $n$ , the  $n$ th correlation measure of  $\xi$ , if it exists, is defined to be a  $\sigma$ -finite Borel measure  $\mu_n$  such that, for any nonnegative measurable function  $f$  on  $\mathbb{R}^n$ ,

$$E \left[ \sum_{i_1, \dots, i_n \neq} f(X_{i_1}, \dots, X_{i_n}) \right] = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) \mu_n(dx_1 \cdots dx_n),$$

where  $\sum_{i_1, \dots, i_n \neq}$  denotes that the sum is taken over all  $n$ -tuples of distinct indices. If  $\mu_n$  is absolutely continuous with respect to the  $n$ -dimensional Lebesgue measure, the density is called the  $n$ th correlation function of  $\xi$ .

It is also worth noting that the sampling formula (1.2) is equivalent to the following statement: given an arbitrary partition  $(n_1, \dots, n_k)$  of  $n$ , we have

$$E_{\alpha, \theta} \left[ \sum_{i_1, \dots, i_k \neq} X_{i_1}^{n_1} \cdots X_{i_k}^{n_k} \right] = \frac{\prod_{l=0}^{k-1} (\theta + l\alpha)}{\theta^{(n)}} \prod_{i=1}^k (1 - \alpha)^{(n_i-1)}, \tag{2.1}$$

where  $E_{\alpha, \theta}$  means that the expectation is taken with respect to the  $PD(\alpha, \theta)$  distribution. This statement is key to showing our first result for the correlation function of the  $PD(\alpha, \theta)$  process in the next section.

### 3. Sampling formula and Laplace transform

**Theorem 3.1.** *Suppose that the sampling formula (1.2), which is equivalent to (2.1), holds for a random element  $(X_i)_{i=1}^\infty \in \nabla_\infty$ . Then the  $n$ th correlation function of the point process  $\xi = \sum_{i=1}^\infty \delta_{X_i}$  for each  $k = 1, 2, \dots$  is given by*

$$q_k(x_1, \dots, x_k) = c_{k, \alpha, \theta} \prod_{i=1}^k x_i^{-(\alpha+1)} \left( 1 - \sum_{j=1}^k x_j \right)^{\theta + \alpha k - 1} \mathbf{1}_{\Delta_k}(x_1, \dots, x_k), \tag{3.1}$$

where

$$\Delta_k = \left\{ (x_1, \dots, x_k) : x_1 \geq 0, \dots, x_k \geq 0, \sum_{i=1}^k x_i \leq 1 \right\}$$

and

$$c_{k, \alpha, \theta} = \prod_{i=1}^k \frac{\Gamma(\theta + 1 + (i - 1)\alpha)}{\Gamma(1 - \alpha)\Gamma(\theta + i\alpha)}.$$

*Proof.* In fact, we need only show that, for any nonnegative measurable function  $f(x_1, \dots, x_k)$  on  $\Delta_k$ , the identity

$$E \left[ \sum_{i_1, \dots, i_k \neq} f(X_{i_1}, \dots, X_{i_k}) \right] = \int_{\Delta_k} f(x_1, \dots, x_k) q_k(x_1, \dots, x_k) dx_1 \cdots dx_k \tag{3.2}$$

holds. Set  $f(x_1, \dots, x_k) = x_1^{n_1} \cdots x_k^{n_k}$ . Then the right-hand side of (3.2) is equal to

$$\begin{aligned} & c_{k, \alpha, \theta} \int_{\Delta_k} \prod_{i=1}^k x_i^{n_i - \alpha - 1} \left( 1 - \sum_{j=1}^k x_j \right)^{\theta + \alpha k - 1} dx_1 \cdots dx_k \\ &= \prod_{i=1}^k \frac{\Gamma(\theta + 1 + (i - 1)\alpha)}{\Gamma(1 - \alpha)\Gamma(\theta + i\alpha)} \frac{\prod_{i=1}^k \Gamma(n_i - \alpha)\Gamma(\theta + k\alpha)}{\Gamma(\theta + n)} \\ &= \frac{\prod_{i=1}^k (1 - \alpha)_{(n_i - 1)} \prod_{i=0}^{k-1} (\theta + i\alpha)}{\theta_{(n)}}, \end{aligned}$$

which, upon combining with (2.1), indicates that (3.2) holds for all polynomial functions on  $\Delta_k$ . Let  $\mathcal{L}$  be the set of all the nonnegative functions satisfying (3.2). It is easy to check that  $\mathcal{L}$  is a  $\lambda$ -system, which means that  $\mathcal{L}$  satisfies the following three conditions:

1.  $1 \in \mathcal{L}$ ;
2.  $\mathcal{L}$  is closed under a finite linear combination;
3. If  $f_n \in \mathcal{L}$  and  $f_n \uparrow f$ , then  $f \in \mathcal{L}$ .

It suffices to show that  $\mathcal{L}$  contains all the nonnegative bounded continuous functions on  $\Delta_k$  by the monotone theorem.

Since  $\Delta_k$  is compact, by the Stone–Weierstrass theorem, any continuous function  $f \in C(\Delta_k)$  can be approximated by a sequence of polynomial functions uniformly. Thus, for any nonnegative  $f \in C(\Delta_k)$ , there exists a sequence of polynomials  $\{B_m\}$  such that  $\sup_{\Delta_k} |B_m(x) - f(x)| \rightarrow 0$  as  $m \rightarrow \infty$ . Without loss of generality, we can assume that the sequence  $\{B_m\}$  is also nonnegative. We consider the following two cases.

*Case 1:*  $\int_{\Delta_k} f(x_1, \dots, x_k) q_k(x_1, \dots, x_k) dx_1 \cdots dx_k < \infty$ . Denote the interior of  $\Delta_k$  by  $\Delta_k^\circ$ , and set  $\mathbb{Q}^k \cap \Delta_k^\circ = \{\mathbf{q}_1, \mathbf{q}_2, \dots\}$ , since it is a countable set. For any  $\varepsilon > 0$  and  $\mathbf{q}_1 = (q_{11}, \dots, q_{1k})$ , since  $f(\mathbf{q}_1) < f(\mathbf{q}_1) + \varepsilon \prod_{j=1}^k q_{1j}$  and  $B_m(\mathbf{q}_1) \rightarrow f(\mathbf{q}_1)$ , we can find a subsequence of  $B_m$ , denoted by  $B_{1m}$ , such that  $B_{1m}(\mathbf{q}_1) \leq f(\mathbf{q}_1) + \varepsilon \prod_{j=1}^k q_{1j}$ . For the same reason, we can obtain another subsequence of  $B_{1m}$ , denoted by  $B_{2m}$ , for  $\mathbf{q}_2$  such that  $B_{2m}(\mathbf{q}_2) \leq f(\mathbf{q}_2) + \varepsilon \prod_{j=1}^k q_{2j}$ . Repeating this process for each  $\mathbf{q}_i$  and using the diagonal method, we can obtain a sequence which we still call  $B_m$  such that  $B_m(\mathbf{q}_i) \leq f(\mathbf{q}_i) + \varepsilon \prod_{j=1}^k q_{ij}$  for any  $i = 1, 2, \dots$ .

By the continuity of  $B_m$  we have  $B_m(\mathbf{x}) \leq f(\mathbf{x}) + \varepsilon \prod_{j=1}^k x_j$  for all  $\mathbf{x} = (x_1, \dots, x_k) \in \Delta_k^\circ$ . Since  $\int_{\Delta_k^\circ} (f(\mathbf{x}) + \varepsilon \prod_{j=1}^k x_j) q_k(x_1, \dots, x_k) dx_1 \cdots dx_k < \infty$ , by the dominated theorem we obtain

$$\int_{\Delta_k^\circ} f(\mathbf{x}) q_k(\mathbf{x}) d\mathbf{x} = \lim_{m \rightarrow \infty} \int_{\Delta_k^\circ} B_m(\mathbf{x}) q_k(\mathbf{x}) d\mathbf{x}.$$

On the other hand, we have  $\int_{\Delta_k} f(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} = \int_{\Delta_k^\circ} f(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x}$  and  $\int_{\Delta_k} B_m(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} = \int_{\Delta_k^\circ} B_m(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x}$ , since the boundary of  $\Delta_k$  has zero Lebesgue measure. Therefore,

$$\begin{aligned} \lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} B_m(X_{i_1} \cdots X_{i_k}) \right] &= \lim_{m \rightarrow \infty} \int_{\Delta_k} B_m(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} \\ &= \lim_{m \rightarrow \infty} \int_{\Delta_k^\circ} B_m(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Delta_k^\circ} f(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Delta_k} f(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x}. \end{aligned}$$

Following Fatou’s lemma, we have

$$\mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} f(X_{i_1}, \dots, X_{i_k}) \right] \leq \lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} B_m(X_{i_1} \cdots X_{i_k}) \right]. \tag{3.3}$$

Since  $B_m(X_{i_1} \cdots X_{i_k}) \leq f(X_{i_1}, \dots, X_{i_k}) + \varepsilon \prod_{j=1}^k X_{i_j}$  almost surely,

$$\lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} B_m(X_{i_1} \cdots X_{i_k}) \right] \leq \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} f(X_{i_1}, \dots, X_{i_k}) \right] + \varepsilon.$$

Letting  $\varepsilon \downarrow 0$  and combining with (3.3), we have

$$\mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} f(X_{i_1}, \dots, X_{i_k}) \right] = \lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} B_m(X_{i_1} \cdots X_{i_k}) \right],$$

which gives the identity  $\int_{\Delta_k} f(\mathbf{x})q_k(\mathbf{x}) \, d\mathbf{x} = \mathbb{E}[\sum_{i_1, \dots, i_k \neq} f(X_{i_1}, \dots, X_{i_k})]$ .

Case 2:  $\int_{\Delta_n} f(\mathbf{x})q_n(\mathbf{x}) \, d\mathbf{x} = \infty$ . By Fatou’s lemma we have

$$\lim_{m \rightarrow \infty} \int_{\Delta_n} B_m(\mathbf{x})q_n(\mathbf{x}) \, d\mathbf{x} \geq \int_{\Delta_n} f(\mathbf{x})q_n(\mathbf{x}) \, d\mathbf{x}$$

and

$$\lim_{m \rightarrow \infty} \int_{\Delta_n} B_m(\mathbf{x})q_n(\mathbf{x}) \, d\mathbf{x} = \lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_n \neq} B_m(X_{i_1} \cdots X_{i_n}) \right] = \infty.$$

By the same method as above we can find a subsequence of  $B_m$  which we still denote by  $B_m$  such that  $B_m(\mathbf{x}) \leq f(\mathbf{x}) + [\prod_{j=1}^n x_j]$  for all  $\mathbf{x} = (x_1, \dots, x_n) \in \Delta_n^\circ$ . So

$$B_m(X_{i_1} \cdots X_{i_n}) \leq f(X_{i_1}, \dots, X_{i_n}) + 1 \quad \text{almost surely,}$$

and

$$\lim_{m \rightarrow \infty} \mathbb{E} \left[ \sum_{i_1, \dots, i_n \neq} B_m(X_{i_1} \cdots X_{i_n}) \right] \leq \mathbb{E} \left[ \sum_{i_1, \dots, i_n \neq} f(X_{i_1}, \dots, X_{i_n}) \right] + 1.$$

Thus,  $\mathbb{E}[\sum_{i_1, \dots, i_n \neq} f(X_{i_1}, \dots, X_{i_n})] = \infty$ .

Since we now know the correlation function of the point process  $\xi$ , we can derive the Laplace transform of the probability generating function as shown in Theorem 3.1 of [11].

**Theorem 3.2.** ([11].) *Suppose that  $g : (0, \infty) \rightarrow C$  is a measurable function such that*

$$\lambda_\alpha(g) := \inf \left\{ \lambda > 0 : \int_0^\infty \frac{e^{-\lambda z}}{z^{\alpha+1}} |g(z)| dz < \infty \right\} < \infty.$$

Set

$$\lambda_\alpha^*(g) = \inf \left\{ \lambda > \lambda_\alpha(g) : \frac{C_\alpha}{\lambda^\alpha} \int_0^\infty \frac{e^{-\lambda z}}{z^{\alpha+1}} |g(z)| dz \in C \setminus [1, \infty) \right\},$$

so that in particular  $\lambda_0^*(g) = \lambda_0(g)$ . The correlation measure of  $\xi = \sum_{i=1}^\infty \delta_{X_i}$  is given by (3.1). Then

$$E \left[ \prod_{i=1}^\infty (1 + |g(sX_i)|) \right] < \infty$$

for almost every  $s > 0$ , and, for  $\lambda > \lambda_\alpha^*(g)$ ,

$$\begin{aligned} & \frac{\lambda^\theta}{\Gamma(\theta + 1)} \int_0^\infty s^{\theta-1} e^{-\lambda s} E \left[ \prod_{i=1}^\infty (1 + g(sX_i)) - 1 \right] ds \\ &= \begin{cases} \frac{1}{\theta} \exp \left( \theta \int_0^\infty \frac{e^{-\lambda z}}{z} g(z) dz \right) - \frac{1}{\theta}, & \alpha = 0, \theta > 0, \\ \frac{1}{\theta} \left( 1 - \frac{C_\alpha}{\lambda^\alpha} \int_0^\infty \frac{e^{-\lambda z}}{z^{\alpha+1}} g(z) dz \right)^{-\theta/\alpha} - \frac{1}{\theta}, & 0 < \alpha < 1, \theta \neq 0, \\ -\frac{1}{\alpha} \log \left( 1 - \frac{C_\alpha}{\lambda^\alpha} \int_0^\infty \frac{e^{-\lambda z}}{z^{\alpha+1}} g(z) dz \right), & 0 < \alpha < 1, \theta = 0. \end{cases} \end{aligned}$$

As a corollary of Theorem 3.2, we have the following result for the sampling formula. It implies that the sampling formula (1.2) and the Laplace transform of the PD( $\alpha, \theta$ ) point process given in Theorem 3.2 are equivalent.

**Corollary 3.1.** *Suppose that the Laplace transform given in Theorem 3.2 holds for sufficiently large  $\lambda > 0$ . Then the sampling formula corresponding to  $\mathbf{X} = (X_1, X_2, \dots) \in \nabla_\infty$  coincides with Pitman’s sampling formula (1.2), which is equal to (2.1).*

*Proof.* Using the uniqueness of the Laplace transformation, or Lemma 3.1 of [11], and taking its inverse, we have

$$\begin{aligned} F(s) &= E \left[ \prod_{i=1}^\infty (g(sX_i) + 1) \right] - 1 \\ &= \sum_{n=1}^\infty \frac{c_{n,\alpha,\theta}}{n!} \int_{\Delta_n} \prod_{i=1}^n \frac{g(sx_i)}{x_i^{\alpha+1}} \left( 1 - \sum_{j=1}^n x_j \right)^{\theta+\alpha n-1} dx_1 \cdots dx_n \end{aligned}$$

for the function  $g(z)$  which satisfies the condition that there exists some  $\lambda > 0$  such that  $\int_0^\infty (e^{-\lambda z}/z^{\alpha+1})|g(z)| dz < \infty$ . Set  $s = 1$  and  $g(z) = t_1 z^{n_1} + \dots + t_k z^{n_k}$ , where  $n_1, \dots, n_k \geq 1$ . It is obvious that  $\int_0^\infty (e^{-\lambda z}/z^{\alpha+1})|g(z)| dz < \infty$  for all  $\lambda > 0$ . Furthermore,

$$\frac{\partial^k}{\partial t_1 \cdots \partial t_k} \Big|_{t_1=\dots=t_k=0} \prod_{i=1}^\infty (1 + g(X_i)) = \sum_{i_1, \dots, i_k \neq} X_{i_1}^{n_1} \cdots X_{i_k}^{n_k}.$$

Therefore,

$$\begin{aligned}
 & \mathbb{E} \left[ \sum_{i_1, \dots, i_k \neq} X_{i_1}^{n_1} \cdots X_{i_k}^{n_k} \right] \\
 &= \frac{\partial^k}{\partial t_1 \cdots \partial t_k} \Big|_{t_1 = \dots = t_k = 0} \mathbb{E} \left[ \prod_{i=1}^{\infty} (1 + g(X_i)) \right] \\
 &= \frac{\partial^k}{\partial t_1 \cdots \partial t_k} \Big|_{t_1 = \dots = t_k = 0} \\
 &\quad \times \left[ 1 + \sum_{n=1}^{\infty} \frac{c_{n, \alpha, \theta}}{n!} \int_{\Delta_n} \prod_{i=1}^n \frac{g(sx_i)}{x_i^{\alpha+1}} \left( 1 - \sum_{j=1}^n x_j \right)^{\theta + \alpha n - 1} dx_1 \cdots dx_n \right] \\
 &= \sum_{n=1}^{\infty} \frac{c_{n, \alpha, \theta}}{n!} \int_{\Delta_n} \frac{\partial^k}{\partial t_1 \cdots \partial t_k} \Big|_{t_1 = \dots = t_k = 0} \prod_{i=1}^n \frac{g(sx_i)}{x_i^{\alpha+1}} \left( 1 - \sum_{j=1}^n x_j \right)^{\theta + \alpha n - 1} dx_1 \cdots dx_n \\
 &= \frac{c_{k, \alpha, \theta}}{k!} \int_{\Delta_k} k! \prod_{i=1}^k x_i^{n_i - \alpha - 1} \left( 1 - \sum_{j=1}^k x_j \right)^{\theta + \alpha n - 1} dx_1 \cdots dx_n \\
 &= c_{k, \alpha, \theta} \frac{\Gamma(n_1 - \alpha) \cdots \Gamma(n_k - \alpha) \Gamma(\theta + k\alpha)}{\Gamma(\theta + n)} \\
 &= \frac{\prod_{i=1}^k (1 - \alpha)_{(n_i - 1)}}{\theta_{(n)}} \prod_{l=0}^{k-1} (\theta + l\alpha),
 \end{aligned}$$

which is exactly (2.1).

#### 4. The IMNA diffusion process

As mentioned in the introduction, the Poisson–Dirichlet distribution PD(θ) is the stationary distribution of the IMNA process X(t) with generator

$$L = \frac{1}{2} \sum_{i, j=1}^{\infty} x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} - \frac{\theta}{2} \sum_{i=1}^{\infty} x_i \frac{\partial}{\partial x_i},$$

whose domain is  $\mathcal{D}(L) = \text{span}\{1, \varphi_2, \varphi_3, \dots\} \subset C(\nabla_{\infty})$ , where  $\varphi_n(x) = \sum_{i=1}^{\infty} x_i^n$  is defined on  $\nabla_{\infty}$  and extends continuously to its closure  $\bar{\nabla}_{\infty}$ . In [2] it was shown that the transition probability P(t, x, A) of X(t) starting from  $x = (x_1, x_2, \dots) \in \nabla_{\infty}$  is absolutely continuous with respect to the stationary distribution PD(θ). Moreover, the transition density function q(t, x, y) is given as

$$\begin{aligned}
 q(t, x, y) &= 1 + \sum_{m=2}^{\infty} \frac{2m - 1 + \theta}{m!} e^{-\lambda_m t} \\
 &\quad \times \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n + \theta)_{(m-1)} p_n(x, y), \quad x, y \in \nabla_{\infty}, \quad (4.1)
 \end{aligned}$$



where  $\lambda_m = \frac{1}{2}m(m - 1 + \theta)$  and  $p_n$  is defined as

$$p_n(x, y) = \sum_{\{\mathbf{n}: |\mathbf{n}|=n\}} \frac{\psi_{\mathbf{n}}(x)\psi_{\mathbf{n}}(y)}{E[\psi_{\mathbf{n}}(z)]},$$

where  $\psi_{\mathbf{n}}$  is as defined in (1.3) and the expectation  $E[\psi_{\mathbf{n}}(z)]$  is taken with respect to the PD( $\theta$ ) distribution, which can be given by Ewens’ sampling formula (1.1). The type of expansion in (4.1) was first derived by Griffiths [8]. A very similar transition function was also obtained by Ethier and Griffiths [3] for the labeled version of the IMNA model, which is a measure-valued process called the Fleming–Viot process with parent-independent mutation. Here, we consider the Laplace transform of  $X(t) = (X_1(t), X_2(t), \dots)$  with initial value  $x$ .

**Theorem 4.1.** *Suppose that  $g(z)$  is a measurable function such that*

$$\lambda(g) := \inf \left\{ \lambda > 0: \int_0^\infty \frac{e^{-\lambda z}}{z} |g(z)| dz < \infty \right\} < \infty.$$

*Then, for fixed time  $t$ , the Laplace transform of the IMNA process  $X(t)$  starting from  $x \in \nabla_\infty$  is given by*

$$\begin{aligned} & \int_0^\infty e^{-\lambda \tau} \tau^{\theta-1} E_x \left[ \prod_{i=1}^\infty (1 + g(\tau X_i(t))) \right] d\tau \\ &= \frac{\Gamma(\theta)}{\lambda^\theta} \exp \left( \theta \int_0^\infty \frac{e^{-\lambda z}}{z} g(z) dz \right) \\ &+ \sum_{m=2}^\infty \frac{2m - 1 + \theta}{m!} e^{-\lambda m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n + \theta)_{(m-1)} \\ &\times \sum_{\substack{\mathbf{n}=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \psi_{\mathbf{n}}(x) \left[ \frac{\Gamma(\theta)}{\lambda^\theta} + \text{ESF}(\mathbf{n})^{-1} \sum_{\ell=1}^\infty \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \binom{\ell}{r} \theta^{k+\ell-r} H_{k, \ell, B} \right], \end{aligned}$$

where

$$\begin{aligned} H_{k, \ell, B} &= \int_{(0, \infty)^{k+\ell-r}} y_1^{n_1} \cdots y_k^{n_k} \left( \prod_{j \in B} g(y_j) \right) g(y_{k+1}) \cdots g(y_{k+\ell-r}) \left( \prod_{i=1}^{k+\ell-r} y_i^{-1} e^{-\lambda y_i} \right) \\ &\times dy_1 \cdots dy_{k+\ell-r} \int_0^\infty e^{-\lambda \tau} \left( \tau + \sum_{j=1}^{k+\ell-r} y_j \right)^{-n} \tau^{\theta-1} d\tau \end{aligned}$$

and  $\text{ESF}(\mathbf{n})$  is given by Ewens’ sampling formula as

$$\text{ESF}(\mathbf{n}) = \frac{\theta^k}{\theta_{(n)}} (n_1 - 1)! \cdots (n_k - 1)!.$$

*Proof.* For an arbitrary fixed time  $t > 0$ ,

$$\begin{aligned} E_x \left[ \prod_{i=1}^{\infty} (1 + g(\tau X_i(t))) \right] &= E \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) q(t, x, Y) \right] \\ &= E \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \right] \\ &\quad + \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} \\ &\quad \times E \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) p_n(x, Y) \right], \end{aligned}$$

where  $Y = (Y_1, Y_2, \dots)$  is distributed as  $PD(\theta)$ . Since

$$\prod_{i=1}^{\infty} (1 + g(\tau Y_i)) p_n(x, Y) = \sum_{\substack{n=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \phi_n(Y) \psi_n(x) \text{ESF}(\mathbf{n})^{-1},$$

where

$$\phi_n(Y) = \sum_{i_1, \dots, i_k \neq} Y_{i_1}^{n_1} \dots Y_{i_k}^{n_k},$$

we have

$$\begin{aligned} \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \phi_n(Y) &= \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \sum_{i_1, \dots, i_k \neq} Y_{i_1}^{n_1} \dots Y_{i_k}^{n_k} \\ &= \sum_{i_1, \dots, i_k \neq} Y_{i_1}^{n_1} \dots Y_{i_k}^{n_k} + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{\substack{j_1, \dots, j_{\ell} \neq \\ i_1, \dots, i_k \neq}} g(\tau Y_{j_1}) \dots g(\tau Y_{j_{\ell}}) Y_{i_1}^{n_1} \dots Y_{i_k}^{n_k} \\ &= \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \\ &\quad \times \sum_{i_1, \dots, i_{k+\ell-r} \neq} Y_{i_1}^{n_1} \dots Y_{i_k}^{n_k} \left( \prod_{j \in B} g(\tau Y_{j_j}) \right) g(\tau Y_{i_{k+1}}) \dots g(\tau Y_{i_{k+\ell-r}}). \end{aligned}$$

Using the  $k$ th correlation measure  $\mu_k$  obtained in last section, we have

$$\begin{aligned} E \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \phi_n(Y) \right] &= \text{ESF}(\mathbf{n}) + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \\ &\quad \times \int_{\Delta_{k+\ell-r}} y_1^{n_1} \dots y_k^{n_k} \left( \prod_{j \in B} g(\tau y_j) \right) g(\tau y_{k+1}) \dots g(\tau y_{k+\ell-r}) \\ &\quad \times \mu_{k+\ell-r}(dy_1 \dots dy_{k+\ell-r}) \end{aligned}$$

$$\begin{aligned}
 &= \text{ESF}(\mathbf{n}) + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \theta^{k+\ell-r} \\
 &\quad \times \int_{\Delta_{k+\ell-r}} y_1^{n_1} \cdots y_k^{n_k} \left( \prod_{j \in B} g(\tau y_j) \right) g(\tau y_{k+1}) \cdots g(\tau y_{k+\ell-r}) \\
 &\quad \times \left( 1 - \sum_{i=1}^{k+\ell-r} y_i \right)^{\theta-1} \frac{dy_1 \cdots dy_{k+\ell-r}}{y_1 \cdots y_{k+\ell-r}}.
 \end{aligned}$$

Define

$$\begin{aligned}
 H_{k,\ell,B} &:= \int_0^{\infty} e^{-\lambda\tau} \tau^{\theta-1} d\tau \int_{\Delta_{k+\ell-r}} y_1^{n_1} \cdots y_k^{n_k} \left( \prod_{j \in B} g(\tau y_j) \right) g(\tau y_{k+1}) \cdots g(\tau y_{k+\ell-r}) \\
 &\quad \times \left( 1 - \sum_{i=1}^{k+\ell-r} y_i \right)^{\theta-1} \frac{dy_1 \cdots dy_{k+\ell-r}}{y_1 \cdots y_{k+\ell-r}}.
 \end{aligned}$$

Using Fubini’s theorem, we have

$$|H_{k,\ell,B}| \leq \Gamma(\theta) \left( \int_0^{\infty} \frac{e^{-\lambda z} |g(z)|}{z} dz \right)^{\ell}.$$

Consequently, the series

$$\sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \theta^{k+\ell-r} H_{k,\ell,B}$$

is absolutely convergent and upper bounded by  $\Gamma(\theta)e^{\theta M} \theta^k (1 + M)^k$ , where

$$M = \int_0^{\infty} e^{-\lambda x} |g(x)| x^{-1} dx < \infty.$$

In addition, we obtained previously

$$\int_0^{\infty} e^{-\lambda\tau} \tau^{\theta-1} \mathbb{E} \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \right] d\tau = \frac{\Gamma(\theta)}{\lambda^{\theta}} \exp \left( \theta \int_0^{\infty} \frac{e^{-\lambda z}}{z} g(z) dz \right).$$

Therefore,

$$\begin{aligned}
 &\int_0^{\infty} e^{-\lambda\tau} \tau^{\theta-1} \mathbb{E}_x \left[ \prod_{i=1}^{\infty} (1 + g(\tau X_i(t))) \right] d\tau \\
 &= \int_0^{\infty} e^{-\lambda\tau} \tau^{\theta-1} \mathbb{E} \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) \right] d\tau \\
 &\quad + \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} \\
 &\quad \times \int_0^{\infty} e^{-\lambda\tau} \tau^{\theta-1} \mathbb{E} \left[ \prod_{i=1}^{\infty} (1 + g(\tau Y_i)) p_n(x, Y) \right] d\tau
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\Gamma(\theta)}{\lambda^\theta} \exp\left(\theta \int_0^\infty \frac{e^{-\lambda z}}{z} g(z) dz\right) \\
 &+ \sum_{m=2}^\infty \frac{2m-1+\theta}{m!} e^{-\lambda m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} \\
 &\times \sum_{\substack{n=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \psi_n(x) \left[ \frac{\Gamma(\theta)}{\lambda^\theta} + \text{ESF}(n)^{-1} \sum_{\ell=1}^\infty \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \theta^{k+\ell-r} H_{k, \ell, B} \right],
 \end{aligned}$$

where

$$\begin{aligned}
 H_{k, \ell, B} &= \int_{(0, \infty)^{k+\ell-r}} y_1^{n_1} \cdots y_k^{n_k} \left( \prod_{j \in B} g(y_j) \right) g(y_{k+1}) \cdots g(y_{k+\ell-r}) \frac{dy_1 \cdots dy_{k+\ell-r}}{y_1 \cdots y_{k+\ell-r}} \\
 &\times \int_{\sum_{j=1}^{k+\ell-r} y_j = \tau}^\infty e^{-\lambda \tau} \tau^{-n} \left( \tau - \sum_{j=1}^{k+\ell-r} y_j \right)^{\theta-1} d\tau \\
 &= \int_{(0, \infty)^{k+\ell-r}} y_1^{n_1} \cdots y_k^{n_k} \left( \prod_{j \in B} g(y_j) \right) g(y_{k+1}) \cdots g(y_{k+\ell-r}) \\
 &\times \prod_{i=1}^{k+\ell-r} (y_i^{-1} e^{-\lambda y_i}) dy_1 \cdots dy_{k+\ell-r} \int_0^\infty e^{-\lambda \tau} \left( \tau + \sum_{j=1}^{k+\ell-r} y_j \right)^{-n} \tau^{\theta-1} d\tau.
 \end{aligned}$$

The upper bound of

$$\sum_{\ell=1}^\infty \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \theta^{k+\ell-r} H_{k, \ell, B}$$

guarantees the absolute convergence of the above series.

### 4.1. Central limit theorem for the IMNA process

The Laplace transform obtained in Theorem 4.1 enables us to derive the central limit theorem (CLT) associated with the distribution of the IMNA process  $X(t)$ .

**Theorem 4.2.** For any time  $t > 0$ , set

$$W_p(t) = \sqrt{\theta} \left( \frac{\theta^{p-1}}{(p-1)!} H_p(t) - 1 \right),$$

where  $H_p(t) = \sum_{i=1}^\infty X_i(t)^p$  denotes the  $p$ th homozygosity at time  $t$  for  $p = 2, 3, \dots$ . Then, as  $\theta \rightarrow \infty$ ,  $W_p(t)$  converges in law to a normal random variable with mean 0 and variance  $\sigma_p^2 = \Gamma(2p)/\Gamma(p)^2 - p^2$ , which is strictly positive.

*Proof.* To prove the CLT for  $W_p(t)$ , it suffices to show that the characteristic function of  $W_p(t)$ , which is  $\psi_p(x) = E[\exp(ix W_p(t))]$ , converges to  $\exp(-\sigma_p^2 x^2/2)$  as  $\theta \rightarrow \infty$ .

Recall from Theorem 4.1 that, for any fixed time  $t > 0$ , we have

$$\begin{aligned}
 &\frac{1}{\Gamma(\theta)} \int_0^\infty e^{-\tau} \tau^{\theta-1} E_x \left[ \prod_{i=1}^\infty (1 + g(\tau X_i(t))) \right] d\tau \\
 &= \exp\left(\theta \int_0^\infty \frac{e^{-z}}{z} g(z) dz\right)
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} \sum_{\substack{n=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \psi_n(x) \\
 & \times \left[ 1 + \frac{1}{\Gamma(\theta)} \text{ESF}(\mathbf{n})^{-1} \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \binom{\ell}{r} \theta^{k+\ell-r} H_{k, \ell, B} \right], \tag{4.2}
 \end{aligned}$$

where  $H_{k, \ell, B}$  and  $\text{ESF}(\mathbf{n})$  are as defined in Theorem 4.1. Set  $g(z) = \exp(c_1 z^p + d_1 z) - 1$ , where  $c_1 = ix/\Gamma(p)\sqrt{\theta}$  and  $d_1 = -ix/\sqrt{\theta}$  for  $x \in \mathbb{R}$ . Then  $E_x[\prod_{i=1}^{\infty} (1 + g(\tau X_i(t)))] = E_x[\exp(c_1 \tau^p H_p(t) + d_1 \tau)]$  and the first term of (4.2) becomes

$$\begin{aligned}
 I & = \exp\left(\theta \int_0^{\infty} \frac{e^{-z}}{z} \left(\exp\left(\frac{ix}{\sqrt{\theta}} \frac{z^p}{\Gamma(p)} - \frac{ix}{\sqrt{\theta}} z\right) - 1\right) dz\right) \\
 & = \exp\left(\theta \int_0^{\infty} \frac{e^{-z}}{z} \left[\left(\frac{ix}{\sqrt{\theta}} \frac{z^p}{\Gamma(p)} - \frac{ix}{\sqrt{\theta}} z\right) - \frac{x^2}{2\theta} \left(\frac{z^p}{\Gamma(p)} - z\right)^2\right] dz + o(1)\right) \\
 & = \exp\left(-\frac{x^2}{2} (\sigma_p^2 + (p-1)^2) + o(1)\right) \quad \text{as } \theta \rightarrow \infty.
 \end{aligned}$$

Next, we show that the second term of (4.2), i.e. the series, goes to 0 as  $\theta \rightarrow \infty$ . Since  $|H_{k, \ell, B}| \leq \Gamma(\theta) \left(\int_0^{\infty} (e^{-z} |g(z)|/z) dz\right)^\ell$ , we have

$$|H_{k, \ell, B}| \leq \Gamma(\theta) \left(\int_0^{\infty} \frac{e^{-z}}{z} \left|\frac{x}{\sqrt{\theta}} \left(\frac{z^p}{\Gamma(p)} - z\right)\right| dz\right)^\ell \leq \Gamma(\theta) \left(\frac{2x}{\sqrt{\theta}}\right)^\ell.$$

For all  $\theta \geq 1$ , we have

$$\begin{aligned}
 \left| \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \frac{\theta^{k+\ell-r}}{\Gamma(\theta)} H_{k, \ell, B} \right| & \leq \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{\ell!}{(\ell-r)!} \theta^{k+\ell-r} \left(\frac{2x}{\sqrt{\theta}}\right)^\ell \\
 & = \theta^k \sum_{\ell=1}^{\infty} \sum_{r=0}^{\ell \wedge k} \sum_{\substack{B \subset \{1, \dots, k\} \\ |B|=r}} \frac{1}{(\ell-r)!} \theta^{\ell-r} \left(\frac{2x}{\sqrt{\theta}}\right)^\ell \\
 & = \theta^k \sum_{r=0}^k \binom{k}{r} \sum_{\ell=r \vee 1}^{\infty} \frac{1}{(\ell-r)!} \theta^{\ell-r} \left(\frac{2x}{\sqrt{\theta}}\right)^\ell \\
 & \leq \theta^k e^{2x\sqrt{\theta}} \sum_{r=0}^k \binom{k}{r} \left(\frac{2x}{\sqrt{\theta}}\right)^r \\
 & = \theta^k e^{2x\sqrt{\theta}} \left(1 + \frac{2x}{\sqrt{\theta}}\right)^k \\
 & \leq \theta^n e^{2x\sqrt{\theta}} (1 + 2x)^n.
 \end{aligned}$$

Using the facts that

$$\sum_{\substack{n=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \psi_n(x) = 1 \quad \text{and} \quad \text{ESF}(\mathbf{n}) = (n_1-1)! \cdots (n_k-1)! \frac{\theta^k}{\theta(n)},$$

we have

$$\sum_{\substack{n=(n_1, \dots, n_k) \\ \sum_{i=1}^k n_i = n}} \psi_n(x) \text{ESF}(\mathbf{n})^{-1} \leq \theta_{(n)}.$$

Thus, the second term of (4.2) is estimated as

$$\begin{aligned} |II| &\leq \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} (n+\theta)_{(m-1)} \\ &\quad + \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} (n+\theta)_{(m-1)} \theta_{(n)} \theta^n e^{2x\sqrt{\theta}}. \end{aligned}$$

If we can show that the above two series are uniformly convergent then they converge to 0 as  $\theta \rightarrow \infty$ , since every term in the series goes to 0.

First, we can see that

$$\begin{aligned} \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} (n+\theta)_{(m-1)} &\leq \sum_{m=2}^{\infty} \frac{(2m-1+\theta)(m+\theta)_{(m-1)}}{m!} e^{-\lambda_m t} 2^m \\ &= \sum_{m=2}^{\infty} \frac{(m+\theta)_{(m)}}{m!} e^{-\lambda_m t} 2^m. \end{aligned}$$

Since

$$(m+\theta)_{(m)} e^{-\lambda_m t} \leq (\theta+2m-1)^m e^{-m(m-1+\theta)t/2} \leq \left( \frac{2(\theta+2m-1)}{t(m-1+\theta)} e^{-1} \right)^m \leq \left( \frac{4}{te} \right)^m,$$

we have

$$\sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} (n+\theta)_{(m-1)} \leq \sum_{m=2}^{\infty} \left( \frac{4}{te} \right)^m \frac{1}{m!} \leq \exp\left( \frac{8}{te} \right) < \infty,$$

i.e. the series converges uniformly. Similarly, we have the estimation of the second term:

$$\begin{aligned} &\sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} (n+\theta)_{(m-1)} \theta_{(n)} \theta^n e^{2x\sqrt{\theta}} \\ &\leq \sum_{m=2}^{\infty} \frac{2m-1+\theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m \binom{m}{n} \theta_{(n+m-1)} \theta^n e^{2x\sqrt{\theta}} \\ &\leq \sum_{m=2}^{\infty} \frac{\theta_{(2m)}}{m!} e^{-\lambda_m t} (\theta+1)^m e^{2x\sqrt{\theta}} \end{aligned}$$

and

$$\theta_{(2m)} (\theta+1)^m e^{-m(m-1+\theta)t/2+2x\sqrt{\theta}} \leq ((\theta+2m-1)(\theta+1) e^{-(m-1+\theta)t/2+2x\sqrt{\theta}/m})^m \leq C^m.$$

The second series in *II* is uniformly convergent. Therefore, the right-hand side of the Laplace transform converges to  $\exp(-x^2(\sigma_p^2 + (p-1)^2)/2)$ .

On the other hand, the left-hand side becomes

$$\begin{aligned} & \frac{1}{\Gamma(\theta)} \int_0^\infty e^{-\tau} \tau^{\theta-1} E_x[\exp(c_1 \tau^p H_p(t) + d_1 \tau)] d\tau \\ &= \left( \int_0^\infty e^{-\theta\tau} \tau^{\theta-1} E_x[\exp(c_1 \theta^p \tau^p H_p(t) + d_1 \theta \tau)] d\tau \right) \left( \int_0^\infty \tau^{\theta-1} e^{-\theta\tau} d\tau \right)^{-1} \\ &= \left( \int_0^\infty e^{-\theta h(\tau)} \tau^{-1} \psi_p(x\tau^p) \exp(ix\sqrt{\theta}(\tau^p - \tau)) d\tau \right) \left( \int_0^\infty \tau^{\theta-1} e^{-\theta h(\tau)} d\tau \right)^{-1}, \end{aligned}$$

where  $h(\tau) = \tau - \log \tau - 1$ . So we need only to prove that

$$\frac{\int_0^\infty e^{-\theta h(\tau)} \tau^{-1} \psi_p(x\tau^p) \exp(ix\sqrt{\theta}(\tau^p - \tau)) d\tau}{\int_0^\infty \tau^{\theta-1} e^{-\theta\tau} d\tau} - \psi_p(x) e^{-(p-1)^2 x^2 / 2} \rightarrow 0.$$

The rest of the proof follows that of Theorem 6.2 of [11]. All that remains is to show that  $\sup_{\theta > 1} E[|W_p(t)|] < \infty$ . Since  $q(t, x, y) \rightarrow 1$  as  $\theta \rightarrow \infty$ , we have  $\sup_{\theta > 1} q(t, x, y) \leq K$  for all  $\theta \geq 1$ . Since

$$(E_x[|W_p(t)|])^2 \leq E_x[|W_p(t)|^2] = E[W_p^2 q(t, x, Y)^2] \leq K^2 (\text{var}(W_p) + (E[W_p])^2),$$

where  $W_p$  is defined as  $W_p(t)$  with the replacement of  $(X_i(t))_{i=1}^\infty$  by  $Y = (Y_i)_{i=1}^\infty$  with the PD( $\theta$ ) distribution, then  $\text{var}(W_p) + (E[W_p])^2 \rightarrow \sigma_p^2$  as  $\theta \rightarrow \infty$ , and  $\sup_{\theta > 1} E[|W_p(t)|] < \infty$ .

**4.2. Two-parameter extension of IMNA process**

In [6] and [16] the authors used different methods to construct a two-parameter version of the IMNA model. Specifically, the two-parameter infinite-alleles diffusion is an infinite-dimensional symmetric diffusion process taking values in  $\nabla_\infty$  with generator

$$L_{\alpha,\theta} = \frac{1}{2} \left\{ \sum_{i,j=1}^\infty x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} - \sum_{i=1}^\infty (\theta x_i + \alpha) \frac{\partial}{\partial x_i} \right\}, \tag{4.3}$$

defined on the same domain of  $L$ . The process is shown to be reversible with respect to the two-parameter Poisson–Dirichlet distribution PD( $\alpha, \theta$ ). Moreover, in a recent paper [7] the transition function was proved to be of the same form as (4.1) in the IMNA model. We need only to replace Ewens’ sampling formula in  $P_n(x, y)$  with Pitman’s sampling formula (1.2). For the two-parameter Poisson–Dirichlet distribution, the CLT is established in [11], which corresponds to the one-parameter case in [12]. Intuitively, the CLT should also be true for the two-parameter process, as in the previous section. However, in this case it becomes more complicated to use the Laplace transform to do the same estimation as in the proof of Theorem 4.2.

On the other hand, the transient sampling distribution for the two-parameter diffusion model can be found by using the correlation measure.

**Theorem 4.3.** *Consider the two-parameter diffusion process  $X(t)$  with generator  $L_{\alpha,\theta}$  (4.3) and initial value  $\mathbf{x} = (x_1, x_2, \dots)$ . The probability generating function  $E[\prod_{i=1}^n u_i^{a_i} \mid X(0) = \mathbf{x}]$  of a sample  $\mathbf{a} = (a_1, a_2, \dots, a_r)$  of size  $r$  at time  $t$  is the coefficient of  $\varphi^r$  in*

$$G_0 + \sum_{m=2}^\infty \frac{2m - 1 + \theta}{m!} e^{-\lambda_m t} \sum_{n=0}^m (-1)^{m-n} \binom{m}{n} (n + \theta)_{(m-1)} G_n,$$

where

$$G_0 = G_1 = \frac{r!}{\theta_{(r)}} \left( 1 - \alpha \sum_{l=1}^r \frac{(1 - \alpha)_{(l-1)} u_l \varphi^l}{l!} \right)^{-\theta/\alpha},$$

and, for  $m \geq 2$ ,

$$G_m = r! \frac{\theta_n}{\theta_{(n+r)}} \sum_{\mathbf{n}=(n_1, \dots, n_k)} \psi_{\mathbf{n}}(\mathbf{x}) \left( 1 - \alpha \sum_{l=1}^r \frac{(1 - \alpha)_{(l-1)} u_l \varphi^l}{l!} \right)^{(-k+\theta/\alpha)} \\ \times \prod_{i=1}^k \left( 1 + \sum_{l=1}^r \frac{(n_i - \alpha)_{(l)} \varphi^l u_l}{l!} \right).$$

*Proof.* Since the probability generating function is  $r!$  times the coefficient of  $\varphi^r$  in

$$\mathbb{E} \left[ \prod_{i=1}^{\infty} \left( 1 + \sum_{l=1}^r \frac{(\varphi X_i(t))^l u_l}{l!} \right) \middle| X(0) = \mathbf{x} \right],$$

owing to the expression of the transition density  $q(t, x, y)$ , we need only to compute

$$\mathbb{E} \left[ \prod_{i=1}^{\infty} \left( 1 + \sum_{l=1}^r \frac{(\varphi X_i)^l u_l}{l!} \right) \right] \tag{4.4}$$

and

$$\text{PSF}(\mathbf{n})^{-1} \mathbb{E} \left[ \psi_{\mathbf{n}}(X) \prod_{i=1}^{\infty} \left( 1 + \sum_{l=1}^r \frac{(\varphi X_i)^l u_l}{l!} \right) \right], \tag{4.5}$$

where  $\mathbf{n} = (n_1, \dots, n_k)$ ,

$$\text{PSF}(\mathbf{n}) = \frac{\theta(\theta + \alpha) \cdots (\theta + (k - 1)\alpha)}{\theta_{(n)}} (1 - \alpha)_{(n_1-1)} \cdots (1 - \alpha)_{(n_k-1)}$$

and the expectation is taken with respect to the PD( $\alpha, \theta$ ) distribution. For simplicity, we only compute the term in (4.4). The term in (4.5) can be computed using a similar argument to that used in the proof of Theorem 4.1. Setting  $g(z) = \sum_{l=1}^n \varphi^l z^l u_l / l!$  and using the correlation function obtained in Theorem 3.1, we have

$$\mathbb{E} \left[ \prod_{i=1}^{\infty} \left( 1 + \sum_{l=1}^r \frac{(\varphi X_i(t))^l u_l}{l!} \right) \right] \\ = 1 + \sum_{n=1}^{\infty} \frac{c_{n,\alpha,\theta}}{n!} \int_{\Delta_n} \prod_{i=1}^n \frac{g(v_i)}{v_i^{1+\alpha}} \left( 1 - \sum_{j=1}^n v_j \right)^{\theta+\alpha n-1} dv_1 \cdots dv_n \\ = 1 + \sum_{n=1}^{\infty} \frac{c_{n,\alpha,\theta}}{n!} \sum_{l_1, \dots, l_n=1}^r \frac{\varphi^{l_1+\dots+l_n} u_{l_1} \cdots u_{l_n}}{l_1! \cdots l_n!} \\ \times \frac{\Gamma(l_1 - \alpha) \cdots \Gamma(l_n - \alpha) \Gamma(\theta + \alpha n)}{\Gamma(l_1 + \cdots + l_n + \theta)} \\ = 1 + \sum_{n=1}^{\infty} \frac{\Gamma(\theta) \alpha^n}{n!} \left( \frac{\theta}{\alpha} \right)_{(n)} \sum_{l_1, \dots, l_n=1}^r \frac{\varphi^{l_1+\dots+l_n} u_{l_1} \cdots u_{l_n} (1 - \alpha)_{(l_1-1)} \cdots (1 - \alpha)_{(l_n-1)}}{l_1! \cdots l_n! \Gamma(l_1 + \cdots + l_n + \theta)}.$$



Therefore, the coefficient of  $\varphi^r$  in the above corresponds to the coefficient of  $\varphi^r$  in the Taylor expansion of  $G_0/\theta_{(r)}$ . It is also worth noting that, by letting  $\alpha \rightarrow 0$ , the limit of  $G_0$  is  $r! \theta_{(r)}^{-1} \exp(\theta \sum_{l=1}^r u_l \varphi^l / l)$ , which coincides with the result obtained by Griffiths [9]. This is also true for  $G_m$ ,  $m \geq 2$ .

### 5. Symmetric selection

The Poisson–Dirichlet distribution with selection,  $PD^{s,q}(\theta)$ , is defined as an absolutely continuous distribution with respect to the Poisson–Dirichlet distribution  $PD(\theta)$  with density

$$\frac{PD^{s,q}(\theta)(d\mathbf{x})}{PD(\theta)(d\mathbf{x})} = \frac{e^{sH_q(\mathbf{x})}}{C}, \quad \mathbf{x} = (x_1, x_2, \dots) \in \nabla_\infty,$$

where  $s$  is an arbitrary real number,  $C$  is the normalization constant, and  $H_q(\mathbf{x}) = \sum x_i^q$ ,  $q = 2, 3, \dots$ . The sampling formula with selection was obtained in [10] as

$$P_{\theta,s,q}(n_1, \dots, n_k) = \frac{E_{PD(\theta)}[e^{sH_q(X)} \psi_n(X)]}{E_{PD(\theta)}[e^{sH_q(X)}]} = C^{-1} \theta^k \sum_{l=0}^\infty \frac{\theta^l}{l!} I_l(\mathbf{a}),$$

where

$$I_l(\mathbf{a}) = \int_{\Delta_{l+k}} \prod_{\alpha=1}^k (x_\alpha^{n_\alpha} e^{s x_\alpha^q}) \prod_{\alpha=k+1}^{k+l} (e^{s x_\alpha^q} - 1) \left(1 - \sum_{\beta=1}^{k+l} x_\beta\right)^{\theta-1} \frac{dx_1 \cdots dx_{k+l}}{x_1 \cdots x_{k+l}}$$

and

$$C = E[e^{sH_q(X)}] = 1 + \sum_{l=1}^\infty \frac{\theta^l}{l!} \int_{\Delta_l} \prod_{\alpha=1}^l (e^{s x_\alpha^q} - 1) \left(1 - \sum_{\beta=1}^l x_\beta\right)^{\theta-1} \frac{dx_1 \cdots dx_l}{x_1 \cdots x_l}.$$

Since the correlation measure and sampling formula are determined by each other, we can obtain the correlation measure in the selection case. The proof is similar to that of Theorem 3.1. It suffices to replace  $f$  by the polynomial functions in (3.2) and the correlation measure is found as follows. Define

$$J_l(t) = \int_{\Delta_l} \prod_{j=1}^l (e^{s t x_j^q} - 1) \left(1 - \sum_{\beta=1}^l x_\beta\right)^{\theta-1} \frac{dx_1 \cdots dx_l}{x_1 \cdots x_l} \quad \text{and} \quad F_s(t) = \sum_{l=0}^\infty \frac{\theta^l}{l!} J_l(t).$$

The correlation measure is given by

$$\begin{aligned} \mu_k(dx_1, \dots, dx_k) &= C^{-1} \theta^k \prod_{j=1}^k (e^{s x_j^q} x_j^{-1}) \left(1 - \sum_{j=1}^k x_j\right)^{\theta-1} \\ &\quad \times F_s \left(1 - \sum_{j=1}^k x_j\right) \mathbf{1}_{\Delta_k}(x_1, \dots, x_k) dx_1 \cdots dx_k. \end{aligned}$$

We obtain the Laplace transform with selection in the next theorem.

**Theorem 5.1.** Suppose that  $g(z)$  is a measurable function such that

$$\lambda(g) := \inf \left\{ \lambda > 0: \int_0^\infty \frac{e^{-\lambda z}}{z} |g(z)| dz < \infty \right\} < \infty.$$

The random element  $(X_i)_{i=1}^\infty$  has the  $PD^{s,q}(\theta)$  distribution. Then

$$\int_0^\infty e^{-\lambda t} t^{\theta-1} E \left[ \prod_{i=1}^\infty (1 + g(tX_i)) \right] dt = \frac{\Gamma(\theta)}{\lambda^\theta} + C^{-1} \sum_{k=1}^\infty \frac{\theta^k}{k!} M_k,$$

where

$$\begin{aligned} M_k &= \int_{(0,\infty)^k} \prod_{j=1}^k (g(x_j)x_j^{-1} e^{-\lambda x_j}) dx_1 \cdots dx_k \\ &\times \int_0^\infty e^{-\lambda t} t^{\theta-1} \prod_{j=1}^k \exp\left(\frac{sx_j^q}{(t + \sum_{j=1}^k x_j)^q}\right) F_s\left(\frac{t}{t + \sum_{j=1}^k x_j}\right) dt. \end{aligned}$$

*Proof.* Since we know the correlation measure, it is obvious that

$$\begin{aligned} E \left[ \prod_{i=1}^\infty (1 + g(tX_i)) \right] &= 1 + C^{-1} \sum_{k=1}^\infty \frac{\theta^k}{k!} \int_{\Delta_k} \prod_{j=1}^k (e^{sx_j^q} g(tx_j)x_j^{-1}) \left(1 - \sum_{j=1}^k x_j\right)^{\theta-1} \\ &\times F_s\left(1 - \sum_{j=1}^k x_j\right) dx_1 \cdots dx_k. \end{aligned}$$

Thus,

$$\begin{aligned} &\int_0^\infty e^{-\lambda t} t^{\theta-1} E \left[ \prod_{i=1}^\infty (1 + g(tX_i)) \right] dt \\ &= \frac{\Gamma(\theta)}{\lambda^\theta} + C^{-1} \sum_{k=1}^\infty \frac{\theta^k}{k!} \int_0^\infty e^{-\lambda t} t^{\theta-1} dt \\ &\quad \times \int_{\Delta_k} \prod_{j=1}^k (e^{sx_j^q} g(tx_j)x_j^{-1}) \left(1 - \sum_{j=1}^k x_j\right)^{\theta-1} F_s\left(1 - \sum_{j=1}^k x_j\right) dx_1 \cdots dx_k \\ &= \frac{\Gamma(\theta)}{\lambda^\theta} + C^{-1} \sum_{k=1}^\infty \frac{\theta^k}{k!} \int_{(0,\infty)^k} \prod_{j=1}^k (g(x_j)x_j^{-1} e^{-\lambda x_j}) dx_1 \cdots dx_k \\ &\quad \times \int_0^\infty e^{-\lambda t} t^{\theta-1} \prod_{j=1}^k \exp\left(\frac{sx_j^q}{(t + \sum_{j=1}^k x_j)^q}\right) F_s\left(\frac{t}{t + \sum_{j=1}^k x_j}\right) dt. \end{aligned}$$

Since, for  $0 < x < 1$ ,

$$|e^{sx^q} - 1| = x^q \left| \int_0^s e^{ux^q} du \right| \leq (1 \vee e^s)x^q |s|,$$

we have

$$\begin{aligned}
 |J_l(t)| &\leq \int_{\Delta_l} (1 \vee e^{s^l})(t^q |s|)^l \prod_{i=1}^l x_i^q \left(1 - \sum_{i=1}^l x_i\right)^{\theta-1} \frac{dx_1 \cdots dx_l}{x_1 \cdots x_l} \\
 &= (1 \vee e^{s^l})(t^q |s|)^l \frac{\Gamma(q)^l \Gamma(\theta)}{\Gamma(ql + \theta)}, \\
 |F_s(t)| &\leq \sum_{l=0}^{\infty} \frac{\theta^l}{l!} |J_l(t)| \\
 &\leq \sum_{l=0}^{\infty} \frac{\theta^l (1 \vee e^{s^l})(t^q |s|)^l \Gamma(q)^l \Gamma(\theta)}{l! \Gamma(ql + \theta)} \\
 &\leq \exp(K(1 \vee e^{s^l})t^q |s| \Gamma(q)\theta^{1-q}),
 \end{aligned}$$

where  $K$  is a constant. Thus,

$$\prod_{j=1}^k \exp\left(\frac{sy_j^q}{(t + \sum_{j=1}^k y_j)^q}\right) \left| F_s\left(\frac{t}{t + \sum_{j=1}^k y_j}\right) \right| \leq m(\theta, q, s),$$

and the second term in the above, i.e. the series, can be controlled by

$$m(\theta, q, s) \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left( \int_0^{\infty} g(z)z^{-1}e^{-\lambda z} dz \right)^k \frac{\Gamma(\theta)}{\lambda^\theta}.$$

Since  $\int_0^{\infty} e^{-\lambda x} |g(x)|x^{-1} dx < \infty$ , the above series is absolutely convergent, which completes the proof.

**Remark.** The above Laplace transform can also lead to the CLT which is parallel to Theorem 4.2. However, the CLT associated with a more generalized selection under the two-parameter case is considered in [20].

### Acknowledgement

The author would like to thank Professor Shui Feng for introducing her to the subject and his help while preparing this paper.

### References

- [1] DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I, 2nd edn. Springer, New York.
- [2] ETHIER, S. N. (1992). Eigenstructure of the infinitely-many-neutral-alleles diffusion model. *J. Appl. Prob.* **29**, 487–498.
- [3] ETHIER, S. N. AND GRIFFITHS, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Prob.* **21**, 1571–1590.
- [4] ETHIER, S. N. AND KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Prob.* **13**, 429–452.
- [5] FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics*. Springer, Heidelberg.
- [6] FENG, S. AND SUN, W. (2010). Some diffusion processes associated with two parameter Poisson–Dirichlet distribution and Dirichlet process. *Prob. Theory Relat. Fields* **148**, 501–525.
- [7] FENG, S., SUN, W., WANG, F. Y. AND XU, F. (2011). Functional inequalities for the two-parameter extension of the infinitely-many-neutral-alleles diffusion. *J. Funct. Anal.* **260**, 399–413.

- [8] GRIFFITHS, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. Appl. Prob.* **11**, 310–325.
- [9] GRIFFITHS, R. C. (1979). Exact sampling distributions from the infinite neutral alleles model. *Adv. Appl. Prob.* **11**, 326–354.
- [10] HANDA, K. (2005). Sampling formulae for symmetric selection. *Electron. Commun. Probab.* **10**, 223–234.
- [11] HANDA, K. (2009). The two-parameter Poisson–Dirichlet point process. *Bernoulli* **15**, 1082–1116.
- [12] 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.
- [13] KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press.
- [14] KINGMAN, J. F. C. *et al.* (1975). Random discrete distributions. *J. R. Statist. Soc. B* **37**, 1–22.
- [15] PERMAN, M., PITMAN, J. AND YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Prob. Theory Relat. Fields* **92**, 21–39.
- [16] PETROV, L. A. (2009). A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funct. Anal. Appl.* **43**, 279–296.
- [17] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Prob. Theory Relat. Fields* **102**, 145–158.
- [18] PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory* (IMS Lecture Notes Monogr. **30**), Institute Mathematical Statistics, Hayward, CA, pp. 245–267.
- [19] PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25**, 855–900.
- [20] XU, F. (2009). A central limit theorem associated with the transformed two-parameter Poisson–Dirichlet distribution. *J. Appl. Prob.* **46**, 392–401.