Elect. Comm. in Probab. **10** (2005), 223–234

ELECTRONIC COMMUNICATIONS in PROBABILITY

SAMPLING FORMULAE FOR SYMMETRIC SELECTION

KENJI HANDA[1]
Department of Mathematics, Saga University, Saga 840-8502, Japan
email: handa@ms.saga-u.ac.jp

Submitted 2 March 2005, *accepted in final form* 12 October 2005

AMS 2000 Subject classification: 92D15, 60C05
Keywords: Ewens sampling formula, partition structure, symmetric selection, homozygosity, Dirichlet distribution, gamma process

Abstract
We study partition distributions in a population genetics model incorporating symmetric selection and mutation. They generalize Ewens distributions in the infinitely-many-neutral-alleles model, an explicit expression of which is known as the Ewens sampling formula. A sampling formula for the generalized model is obtained by means of calculus for Poisson and gamma processes.

Introduction and the main result

Random partitions of integers arise in various contexts of mathematics (see e.g. §2.1 of [1] for several combinatorial examples) and the natural sciences, like population genetics (e.g. [3]). Among various partition distributions (i.e., probability distributions on a set of integer partitions), we are concerned in this paper with partition structure, which was introduced by Kingman [9] in connection with the sampling theory in population genetics. In this theory, we observe a partition $\mathbf{a} = (a_1, \ldots, a_n)$ generated by a random sample of n genes from a population, i.e., a_i is the number of alleles in the sample which appeared exactly i times. Thus a_i are nonnegative integers such that $\sum_{i=1}^n ia_i = n$. In Kingman's papers [9], [10] on partition structures, Ewens distributions [3], which describe laws of random partitions $\mathbf{A}_n$ in the stationary infinitely-many-neutral-alleles model, play a central role. These distributions form a one-parameter family $\{P_\theta : \theta > 0\}$ of partition structures, each of which admits an explicit expression

$$P_\theta(\mathbf{A}_n = \mathbf{a}) = \frac{n!\Gamma(\theta)}{\Gamma(\theta+n)} \prod_{i=1}^n \frac{\theta^{a_i}}{i^{a_i} a_i!} \tag{1}$$

often referred to as the Ewens sampling formula [4]. It is well known (see e.g. [11]) that (1) can be described in terms of a random discrete distribution derived from the points of a

[1]RESEARCH SUPPORTED IN PART BY JSPS, GRANT-IN-AID FOR SCIENTIFIC RESEARCH (A) NO. 14204010 AND (C) NO. 14540128

Poisson point process, say $Z_1 > Z_2 > \cdots$, on $(0, \infty)$ with intensity $\theta dz/(ze^z)$. More precisely, points $\{Y_j\}$ of the normalized process given by $Y_j = Z_j/T$ with $T = \sum_j Z_j$ are interpreted as (random) ranked frequencies of alleles in the stationary infinitely-many-neutral-alleles model with mutation rate θ. Consider an infinite dimensional simplex

$$\nabla = \left\{ \mathbf{y} = \{y_j\} : y_1 \geq y_2 \geq \cdots \geq 0,\ \sum_j y_j = 1 \right\},$$

which is equipped with the topology of coordinate-wise convergence. The above $\{Y_j\}$ is nothing but the ranked jumps of a (standard) Dirichlet process with parameter θ, and its law ν_θ on ∇ is called the Poisson-Dirichlet distribution with parameter θ. The conditional probability that $\mathbf{A}_n = \mathbf{a}$ given an arbitrary sequence of allele frequencies $\{y_j\} \in \nabla$ is evaluated in general as

$$\mathrm{Prob}(\mathbf{A}_n = \mathbf{a} |\ \{y_j\}) = \frac{n!}{\prod_{i=1}^{n} i!^{a_i}} \sum_{\mathbf{m}:\mathbf{m} \in \langle \mathbf{a} \rangle} y_1^{m_1} y_2^{m_2} \cdots =: \Phi_{\mathbf{a}}(\{y_j\}), \tag{2}$$

where $\langle \mathbf{a} \rangle$ stands for the totality of sequences $\mathbf{m} = (m_1, m_2, \ldots)$ of nonnegative integers such that $n = m_1 + m_2 + \cdots$ and this equality (when ignoring the vanishing terms) defines the integer partition $\mathbf{a}$, i.e.,

$$\sharp\{\alpha : m_\alpha = i\} = a_i \quad (i = 1, \ldots, n), \tag{3}$$

where $\sharp$ stands for the cardinality. The left side of (1) is given by

$$P_\theta(\mathbf{A}_n = \mathbf{a}) = E^{\nu_\theta} \left[\Phi_{\mathbf{a}}(\{Y_j\}) \right]. \tag{4}$$

In this paper, we discuss a more general class of partition structures of the form

$$P_{\theta,s,q}(\mathbf{A}_n = \mathbf{a}) = E^{\nu_\theta} \left[e^{sF_q(\{Y_j\})} \Phi_{\mathbf{a}}(\{Y_j\}) \right] / E^{\nu_\theta} \left[e^{sF_q(\{Y_j\})} \right], \tag{5}$$

where s is an arbitrary real number, $q \geq 1$ and $F_q(\{y_j\}) := \sum_j y_j^q$, which is a bounded measurable function on ∇. Obviously, (5) corresponds to a random element of ∇ whose law $\nu_{\theta,s,q}$ is determined by the relation

$$\frac{d\nu_{\theta,s,q}}{d\nu_\theta}(\{y_j\}) = e^{sF_q(\{y_j\})} / E^{\nu_\theta} \left[e^{sF_q(\{Y_j\})} \right], \quad \{y_j\} \in \nabla. \tag{6}$$

In the special case where $q = 2$, $F_2(\cdot)$ is known as the population homozygosity, and the distribution $\nu_{\theta,s,2}$ arises in the population genetics model incorporating symmetric selection and mutation. In this contexts, $s > 0$ means that homozygotes are selectively advantageous relative to heterozygotes (underdominant selection) while $s < 0$ implies the opposite situation (overdominant selection). Watterson [19], [20] obtained a number of asymptotic results concerning $P_{\theta,s,2}$ for small values of s, proposing a powerful statistics for the test of neutrality. Further, for the symmetric overdominance model, Grote and Speed [5] recently derived certain approximate sampling formulae with numerical discussions of interest.

The main purpose of this paper is to give an explicit expression of $P_{\theta,s,q}(\mathbf{A}_n = \mathbf{a})$ for arbitrarily fixed values of θ, s and q. It is expected that such exact formula, if available, must not only reveal mathematical structure behind the quantity studied but also be informative for applications. Note that the denominator in the right-hand side of (6) coincides formally with the numerator with $\mathbf{a} = \mathbf{0}$ ('the partition of 0'; $a_1 = a_2 = \cdots = 0$) if we define $\Phi_{\mathbf{0}} \equiv 1$. In addition to this convention, it is useful to define for any $n = 1, 2, \ldots$

$$\Delta_n = \left\{ (y_1, \ldots, y_n) : y_1 > 0, \ldots, y_n > 0, \ \sum_{i=1}^{n} y_i < 1 \right\}$$

and for any $n \in \{0, 1, 2, \ldots\} =: \mathbb{Z}_+$ and integer partition $\mathbf{a}$ of n

$$M_n(\mathbf{a}) = \frac{n!}{\prod_{i=1}^{n} (i!^{a_i} a_i!)},$$

which is a variant of multinomial coefficient. (In particular, $M_0(\mathbf{0}) = 1$.) The main result of this paper is the following.

Theorem 1 *Let $\theta > 0, s \in \mathbb{R}$ and $q \geq 1$ be arbitrary. Let $\mathbf{a}$ be a partition of $n \in \mathbb{Z}_+$. Set $k = a_1 + \cdots + a_n$ and take any k positive integers $n_1, \ldots, n_k$ such that $n = n_1 + \cdots + n_k$ defines the partition $\mathbf{a}$. Then*

$$E^{\nu_\theta}\left[e^{sF_q(\{Y_j\})} \Phi_{\mathbf{a}}(\{Y_j\}) \right] = M_n(\mathbf{a}) \theta^k \sum_{l=0}^{\infty} \frac{\theta^l}{l!} I_l(\mathbf{a}), \tag{7}$$

where $I_l(\mathbf{a}) = I_l(\mathbf{a}; \theta, s, q)$ is given by

$$I_l(\mathbf{a}) = \int_{\Delta_{k+l}} \prod_{\alpha=1}^{k} (y_\alpha^{n_\alpha} e^{s y_\alpha^q}) \prod_{\alpha=k+1}^{k+l} (e^{s y_\alpha^q} - 1) \left(1 - \sum_{\beta=1}^{k+l} y_\beta \right)^{\theta - 1} \frac{dy_1 \cdots dy_{k+l}}{y_1 \cdots y_{k+l}} \tag{8}$$

except the case of $n = 0 = l$ for which case $I_0(\mathbf{0}) = 1$. Also, it holds that

$$\left| M_n(\mathbf{a}) \theta^k I_l(\mathbf{a}) \right| \begin{cases} \leq & P_\theta(\mathbf{A}_n = \mathbf{a}) \dfrac{\max\{1, e^{s(k+l)}\} (|s|\Gamma(q))^l \Gamma(\theta + n)}{\Gamma(\theta + n + ql)}, \\ \geq & P_\theta(\mathbf{A}_n = \mathbf{a}) \dfrac{\min\{1, e^{s(k+l)}\} (|s|\Gamma(q))^l \Gamma(\theta + n)}{\Gamma(\theta + n + ql)}. \end{cases} \tag{9}$$

We shall give some remarks on immediate implications of Theorem 1 for special values of parameters.

Remarks. (i) The formula (7) with $s = 0$ reduces to the Ewens sampling formula (1). This is seen by observing that $I_l(\mathbf{a}; \theta, 0, q) = 0$ for each $l = 1, 2, \ldots$ and that

$$\begin{aligned} I_0(\mathbf{a}; \theta, 0, q) &= \int_{\Delta_k} \prod_{\alpha=1}^{k} y_\alpha^{n_\alpha} \cdot \left(1 - \sum_{\beta=1}^{k} y_\beta \right)^{\theta - 1} \frac{dy_1 \cdots dy_k}{y_1 \cdots y_k} \\ &= \prod_{\alpha=1}^{k} \Gamma(n_\alpha) \cdot \frac{\Gamma(\theta)}{\Gamma(\theta + n)} = \prod_{i=1}^{n} (i-1)!^{a_i} \cdot \frac{\Gamma(\theta)}{\Gamma(\theta + n)}, \end{aligned} \tag{10}$$

where the second equality is known as Dirichlet's formula (e.g. [2], Appendix M12).
(ii) In case of $n = 0$ or $\mathbf{a} = \mathbf{0}$, (7) yields the following formula for the denominator in the right side of (5).

$$E^{\nu_\theta}\left[e^{sF_q(\{Y_j\})}\right] = 1 + \sum_{l=1}^{\infty} \frac{\theta^l}{l!} \int_{\Delta_l} \prod_{\alpha=1}^{l} (e^{sy_\alpha^q} - 1) \left(1 - \sum_{\beta=1}^{l} y_\beta\right)^{\theta-1} \frac{dy_1 \cdots dy_l}{y_1 \cdots y_l} \tag{11}$$

In fact, an 'implicit version' of (11) was essentially obtained by Pitman and Yor [16] (the formula (174)).
(iii) Since $F_1 \equiv 1$, the left side of (7) with $q = 1$ is equal to $e^s P_\theta(\mathbf{A}_n = \mathbf{a})$. Direct verification that the right side of (7) coincides with this value is rather involved and will be given later.

Unfortunately, our formula (7) itself seems not useful for likelihood-based statistical inference because the right side is not of product form. In general, the condition for a partition structure to be of such a form is quite restrictive as shown in Theorem 42 of [14]. On the other hand, (9) exhibits a rapid convergence of the series in (7) and hence its applicability in some numerical issues.
In the next section, we give a proof of Theorem 1 after providing some lemmas regarding technicalities. Main tools are calculus involving Poisson process and certain distinguished properties of the gamma process. The former calculus for a class of partition structures can be found in [13] (which was revised as [15]) and [6] (a condensed version of which is [7]). The latter ingredient, the use of which we call the 'Γ-trick', is now standard. (See e.g. [8], [12], [17], [18].) However, the crucial idea here is that this is exploited in an 'unusual' way: for each $s = -\sigma < 0$, it is shown that the expectation $E\left[e^{-\sigma F_q(\{Z_j\})}\Phi_{\mathbf{a}}(\{Z_j\})\right]$ with F_q and $\Phi_{\mathbf{a}}$ being naturally extended can be expressed in terms of $E^{\nu_\theta}\left[e^{-uF_q(\{Y_j\})}\Phi_{\mathbf{a}}(\{Y_j\})\right], u \in (0,\infty)$, and after a procedure of inversion we arrive at (7). Also, at the end, the verification mentioned in the Remark (iii) will be given.

1 Calculus for Poisson and gamma processes

Throughout this section let $\mathbf{a} = (a_1, \ldots, a_n)$ be an integer partition of n and set $k = a_1 + \cdots + a_n$. Suppose that k positive integers $n_1, \ldots, n_k$ satisfy $\sharp\{\alpha : n_\alpha = i\} = a_i$ $(i = 1, \ldots, n)$. Consider obvious extention of the functions $\Phi_{\mathbf{a}}$ and F_q, which were defined originally on ∇, to the functions of any sequence $\{z_j\}$ of positive numbers, i.e.,

$$\Phi_{\mathbf{a}}(\{z_j\}) = \frac{n!}{\prod_{i=1}^{n} i!^{a_i}} \sum_{\mathbf{m}:\mathbf{m}\in\langle\mathbf{a}\rangle} z_1^{m_1} z_2^{m_2} \cdots, \quad F_q(\{z_j\}) = \sum_j z_j^q.$$

Note that these functions are symmetric in $z_1, z_2, \ldots$. By suitable change of order of the sum, the following expression of $\Phi_{\mathbf{a}}$ is derived.

$$\Phi_{\mathbf{a}}(\{z_j\}) = M_n(\mathbf{a}) \sum_{j_1,\ldots,j_k:\text{distinct}} z_{j_1}^{n_1} \cdots z_{j_k}^{n_k}, \tag{12}$$

where the sum extends over k-tuples $(j_1, \ldots, j_k)$ such that $j_1, \ldots, j_k$ are mutually distinct. Our first task is calculation of the expectation of $\Phi_{\mathbf{a}}(\{Z_j\})$ for a class of Poisson point processes on $(0,\infty)$. Let $\Lambda(dz)$ be a continuous Borel measure on $(0,\infty)$ such that $\Lambda((0,\infty)) = \infty$ and

$$\int_0^\infty \min\{z,1\}\Lambda(dz) < \infty. \tag{13}$$

Assume that a realization $Z_1 > Z_2 > \cdots > 0$ of the Poisson point process with mean measure Λ is given. That is, a random discrete measure $\xi := \sum_j \delta_{Z_j}$ has Laplace transform

$$E_\Lambda\left[e^{-\langle \xi, f\rangle}\right] = \exp\left(-\int_0^\infty (1-e^{-f(z)})\Lambda(dz)\right) =: \mathcal{L}_\Lambda(f(\cdot)), \tag{14}$$

where f is an arbitrary non-negative Borel function on $(0,\infty)$ and $\langle \xi, f\rangle = \sum_j f(Z_j)$. In the above and what follows, E_Λ is used for notation of the expectation in order to indicate the process we are working on.

Lemma 2 *Let Λ be as above. We suppose additionally that*

$$\int_0^\infty z^{n_\alpha}\Lambda(dz) < \infty \quad (\alpha = 1,\dots,k). \tag{15}$$

Then

$$E_\Lambda\left[\Phi_{\mathbf{a}}(\{Z_j\})\right] = M_n(\mathbf{a}) \prod_{\alpha=1}^k \int_0^\infty z^{n_\alpha}\Lambda(dz). \tag{16}$$

Proof. Observe that at least formally

$$\begin{aligned}
&\sum_{j_1,\dots,j_k:\text{distinct}} Z_{j_1}^{n_1}\cdots Z_{j_k}^{n_k} \\
&= \left.\frac{\partial^k}{\partial t_1\cdots\partial t_k}\right|_{t_1=\cdots=t_k=0} \prod_j \left(1 + t_1 Z_j^{n_1} + \cdots + t_k Z_j^{n_k}\right) \\
&= \left.\frac{\partial^k}{\partial t_1\cdots\partial t_k}\right|_{t_1=\cdots=t_k=0} \exp\langle \xi, \log\left(1 + t_1 z^{n_1} + \cdots + t_k z^{n_k}\right)\rangle.
\end{aligned}$$

Here almost sure convergence of $\langle \xi, \log\left(1 + t_1 z^{n_1} + \cdots + t_k z^{n_k}\right)\rangle$ for any $t_1,\dots,t_k \geq 0$ follows from (15) by virtue of Campbell's theorem (see e.g. [11], §3.2), and therefore the above equalities hold a.s. Moreover this theorem also justifies the following calculations.

$$\begin{aligned}
&E_\Lambda\left[\sum_{j_1,\dots,j_k:\text{distinct}} Z_{j_1}^{n_1}\cdots Z_{j_k}^{n_k}\right] \\
&= \left.\frac{\partial^k}{\partial t_1\cdots\partial t_k}\right|_{t_1=\cdots=t_k=0} E_\Lambda\left[\exp\langle \xi, \log\left(1 + t_1 z^{n_1} + \cdots + t_k z^{n_k}\right)\rangle\right] \\
&= \left.\frac{\partial^k}{\partial t_1\cdots\partial t_k}\right|_{t_1=\cdots=t_k=0} \exp\left(t_1\int_0^\infty z^{n_1}\Lambda(dz) + \cdots + t_k\int_0^\infty z^{n_k}\Lambda(dz)\right).
\end{aligned}$$

Combining this with (12) shows (16). □

Next, we show that the exponential factor in the expectation in (7) can be handled by changing the measure of Poisson point process. For any nonnegative Borel function f on $(0,\infty)$, set

$$\Lambda_f(dz) = e^{-f(z)}\Lambda(dz).$$

The following lemma can be found in [6] (Proposition 1) and the proof requires only (14).

Lemma 3 *Let f and Λ_f be as above. Then for all nonnegative Borel measurable functions Φ of ξ*

$$E_\Lambda\left[e^{-\langle\xi,f\rangle}\Phi(\xi)\right] = E_{\Lambda_f}\left[\Phi(\xi)\right]\mathcal{L}_\Lambda(f(\cdot)). \tag{17}$$

So far, we prepared the auxiliaries regarding Poisson point processes on the half line. We now specify the process as in the previous section by fixing $\theta > 0$ arbitrarily and setting

$$\Lambda(dz) = \theta z^{-1}e^{-z}dz. \tag{18}$$

The associated process $\{Z_j\}$ has the distinguished property that the total sum $T := \sum_j Z_j$ and the normalized process $\{Y_j := Z_j/T\}$ are mutually independent. (See e.g. [8], [12], [17], [18].) For simplicity, we call the 'Γ-trick' use of this property. Recall also that the distribution of T on $(0,\infty)$ is given by

$$\frac{1}{\Gamma(\theta)}t^{\theta-1}e^{-t}dt. \tag{19}$$

Let $q \geq 1$ be arbitrarily. Put for each $\sigma \in \mathbb{R}$

$$J(\sigma) = E_\Lambda\left[e^{-\sigma F_q(\{Y_j\})}\Phi_{\mathbf{a}}(\{Y_j\})\right] = E^{\nu_\theta}\left[e^{-\sigma F_q(\{Y_j\})}\Phi_{\mathbf{a}}(\{Y_j\})\right],$$

which we are going to evaluate. Here is an implicit version of (7).

Proposition 4 *It holds that for any $\tau > 0$*

$$\begin{aligned}
&\frac{1}{\Gamma(\theta)}\int_0^\infty u^{\theta+n-1}J(u^q)e^{-\tau u}du \\
&= \frac{M_n(\mathbf{a})\theta^k}{\tau^\theta}\exp\left(-\theta\int_0^\infty \frac{1-e^{-z^q}}{z}e^{-\tau z}dz\right)\prod_{\alpha=1}^k\int_0^\infty z^{n_\alpha-1}e^{-z^q-\tau z}dz.
\end{aligned} \tag{20}$$

Proof. Let $f_q(z) = z^q$ and Λ be as in (18). Given $\sigma > 0$, consider

$$\widehat{J}(\sigma) = E_\Lambda\left[e^{-\sigma F_q(\{Z_j\})}\Phi_{\mathbf{a}}(\{Z_j\})\right] = E_\Lambda\left[e^{-\sigma\langle\xi,f_q\rangle}\Phi_{\mathbf{a}}(\{Z_j\})\right].$$

Since the symmetric function $\Phi_{\mathbf{a}}$ can be regarded also as a measurable function of ξ, Lemmas 2 and 3 imply that

$$\begin{aligned}
\widehat{J}(\sigma) &= E_{\Lambda_{\sigma f_q}}\left[\Phi_{\mathbf{a}}(\{Z_j\})\right]\mathcal{L}_\Lambda(\sigma f_q(\cdot)) \\
&= E_{\Lambda_{\sigma f_q}}\left[\Phi_{\mathbf{a}}(\{Z_j\})\right]\exp\left(-\theta\int_0^\infty \frac{1-e^{-\sigma z^q}}{z}e^{-z}dz\right) \\
&= M_n(\mathbf{a})\theta^k\exp\left(-\theta\int_0^\infty\frac{1-e^{-\sigma z^q}}{z}e^{-z}dz\right)\prod_{\alpha=1}^k\int_0^\infty z^{n_\alpha-1}e^{-\sigma z^q-z}dz.
\end{aligned} \tag{21}$$

Noting that $F_q(\{Z_j\}) = T^q F_q(\{Y_j\})$ and $\Phi_{\mathbf{a}}(\{Z_j\}) = T^n \Phi_{\mathbf{a}}(\{Y_j\})$, we can also calculate $\widehat{J}(\sigma)$ by the Γ-trick as follows.

$$\begin{aligned}
\widehat{J}(\sigma) &= E_\Lambda\left[\Phi_{\mathbf{a}}(\{Y_j\})T^n e^{-\sigma T^q F_q(\{Y_j\})}\right] \\
&= \frac{1}{\Gamma(\theta)}\int_0^\infty t^{\theta-1}e^{-t}E_\Lambda\left[\Phi_{\mathbf{a}}(\{Y_j\})t^n e^{-\sigma t^q F_q(\{Y_j\})}\right]dt \\
&= \frac{1}{\Gamma(\theta)}\int_0^\infty t^{\theta+n-1}e^{-t}J(\sigma t^q)dt \\
&= \frac{\sigma^{-(\theta+n)/q}}{\Gamma(\theta)}\int_0^\infty u^{\theta+n-1}e^{-u\sigma^{-1/q}}J(u^q)du. \qquad (22)
\end{aligned}$$

By setting $\sigma^{-1/q} =: \tau$, it follows from (21) and (22) that

$$\begin{aligned}
&\frac{1}{\Gamma(\theta)}\int_0^\infty u^{\theta+n-1}e^{-\tau u}J(u^q)du = \tau^{-(\theta+n)}\widehat{J}(\tau^{-q}) \\
&= \frac{M_n(\mathbf{a})\theta^k}{\tau^{\theta+n}}\exp\left(-\theta\int_0^\infty \frac{1-e^{-z^q\tau^{-q}}}{z}e^{-z}dz\right)\prod_{\alpha=1}^k\int_0^\infty z^{n_\alpha-1}e^{-z^q\tau^{-q}-z}dz \\
&= \frac{M_n(\mathbf{a})\theta^k}{\tau^{\theta}}\exp\left(-\theta\int_0^\infty \frac{1-e^{-z^q}}{z}e^{-\tau z}dz\right)\prod_{\alpha=1}^k\int_0^\infty z^{n_\alpha-1}e^{-z^q-\tau z}dz.
\end{aligned}$$

This completes the proof of Proposition 4 □

Proof of Theorem 1. First, we show the bound (9). Observe that for each $0 < y < 1$

$$\left|e^{sy^q}-1\right| = y^q\left|\int_0^s e^{uy^q}du\right|\left\{\begin{array}{ll}\leq & \max\{1,e^s\}y^q|s|, \\ \geq & \min\{1,e^s\}y^q|s|.\end{array}\right.$$

Hence, in view of (1), (9) is implied by (8) together with Dirichlet's formula:

$$\int_{\Delta_{k+l}}\prod_{\alpha=1}^k y_\alpha^{n_\alpha}\prod_{\alpha=k+1}^{k+l}y_\alpha^q\left(1-\sum_{\beta=1}^{k+l}y_\beta\right)^{\theta-1}\frac{dy_1\cdots dy_{k+l}}{y_1\cdots y_{k+l}} = \frac{\prod_{\alpha=1}^k\Gamma(n_\alpha)\cdot\Gamma(q)^l\Gamma(\theta)}{\Gamma(n+\theta+ql)}.$$

An immediate consequence of (9) is that the right side of (7) defines a real analytic function of s. On the other hand, since F_q is bounded on ∇, the left side of (7) is also real analytic in s. So, it is sufficient to prove (7) for $0 > s =: -\sigma$ only. Define

$$I(\sigma) = M_n(\mathbf{a})\theta^k\sum_{l=0}^\infty\frac{\theta^l}{l!}I_l(\mathbf{a};\theta,-\sigma,q), \qquad (23)$$

where I_l $(l = 0, 1, \ldots)$ are given by (8). By virtue of Proposition 4 and the uniqueness of Laplace transform, we only have to verify the equality (20) with $J(\cdot)$ being replaced by $I(\cdot)$. For each $l = 1, 2, \ldots$, let

$$C_l = \{(z_1, \ldots, z_l) : z_1 > 0, \ldots, z_l > 0\}$$

and define a σ-finite measure m_l on C_l by

$$m_l(dz_1 \cdots dz_l) = \frac{dz_1 \cdots dz_l}{z_1 \cdots z_l},$$

which is invariant under arbitrary multiplications of components. For any $u > 0$, set

$$\Delta_l(u) = \left\{ (z_1, \ldots, z_l) : z_1 > 0, \ldots, z_l > 0, \ \sum_{\alpha=1}^{l} z_\alpha < u \right\}.$$

With the above notation, we have by Fubini's theorem

$$\begin{aligned}
&\int_0^\infty u^{\theta+n-1} I_l(\mathbf{a}; \theta, -u^q, q) e^{-\tau u} du \\
&= \int_0^\infty e^{-\tau u} du \int_{\Delta_{k+l}} \prod_{\alpha=1}^{k} ((uy_\alpha)^{n_\alpha} e^{-(uy_\alpha)^q}) \\
&\quad \times \prod_{\alpha=k+1}^{k+l} (e^{-(uy_\alpha)^q} - 1) \left(u - \sum_{\beta=1}^{k+l} (uy_\beta) \right)^{\theta-1} m_{k+l}(dy_1 \cdots dy_{k+l}) \\
&= \int_0^\infty e^{-\tau u} du \int_{\Delta_{k+l}(u)} \prod_{\alpha=1}^{k} (z_\alpha^{n_\alpha} e^{-z_\alpha^q}) \\
&\quad \times \prod_{\alpha=k+1}^{k+l} (e^{-z_\alpha^q} - 1) \left(u - \sum_{\beta=1}^{k+l} z_\beta \right)^{\theta-1} m_{k+l}(dz_1 \cdots dz_{k+l}) \\
&= \int_{C_{k+l}} m_{k+l}(dz_1 \cdots dz_{k+l}) \prod_{\alpha=1}^{k} (z_\alpha^{n_\alpha} e^{-z_\alpha^q}) \prod_{\alpha=k+1}^{k+l} (e^{-z_\alpha^q} - 1) \\
&\quad \times \int_{\sum_{\alpha=1}^{k+l} z_\alpha}^\infty e^{-\tau u} \left(u - \sum_{\beta=1}^{k+l} z_\beta \right)^{\theta-1} du.
\end{aligned}$$

Since this last (one-dimensional) integral is

$$\exp\left(-\tau \sum_{\beta=1}^{k+l} z_\beta \right) \frac{\Gamma(\theta)}{\tau^\theta},$$

we obtain for each $l = 0, 1, \ldots$

$$\begin{aligned}
&\int_0^\infty u^{\theta+n-1} I_l(\mathbf{a}; \theta, -u^q, q) e^{-\tau u} du \\
&= \prod_{\alpha=1}^{k} \int_0^\infty z^{n_\alpha - 1} e^{-z^q - \tau z} dz \cdot \left(\int_0^\infty \frac{e^{-z^q} - 1}{z} e^{-\tau z} dz \right)^l \frac{\Gamma(\theta)}{\tau^\theta}.
\end{aligned} \tag{24}$$

Similarly

$$\begin{aligned}
&\int_0^\infty u^{\theta+n-1}\left|I_l(\mathbf{a};\theta,-u^q,q)\right|e^{-\tau u}du \\
&= \prod_{\alpha=1}^{k}\int_0^\infty z^{n_\alpha-1}e^{-z^q-\tau z}dz\cdot\left|\int_0^\infty \frac{e^{-z^q}-1}{z}e^{-\tau z}dz\right|^l\frac{\Gamma(\theta)}{\tau^\theta}.
\end{aligned}$$

This implies that it is possible to integrate (23) with σ replaced by u^q term by term, and therefore

$$\begin{aligned}
&\int_0^\infty u^{\theta+n-1}I(u^q)e^{-\tau u}du \\
&= M_n(\mathbf{a})\theta^k\prod_{\alpha=1}^{k}\int_0^\infty z^{n_\alpha-1}e^{-z^q-\tau z}dz\cdot\exp\left(-\theta\int_0^\infty\frac{1-e^{-z^q}}{z}e^{-\tau z}dz\right)\frac{\Gamma(\theta)}{\tau^\theta}.
\end{aligned}$$

Comparing this with (20), we completes the proof of Theorem 1. □

At the end of this section, we give a direct proof of the fact claimed in the Remark (iii). That is,

Proposition 5 *For any $\theta > 0$ and $s \in \mathbb{R}$, let $I_l(\mathbf{a};\theta,s,1)$ be given by the right side of (8) with $q = 1$. Then*

$$M_n(\mathbf{a})\theta^k\sum_{l=0}^\infty\frac{\theta^l}{l!}I_l(\mathbf{a};\theta,s,1) = e^sP_\theta(\mathbf{A}_n=\mathbf{a}). \tag{25}$$

Proof. For notational simplicity, put $y_{l+1} = 1-(y_1+\cdots+y_l)$ for $(y_1,\ldots,y_l)\in\Delta_l$. First, we assume that $s > 0$. This assumption makes us possible to exchange sums appearing in the subsequent calculations. Expansions

$$e^{sy_\alpha} = \sum_{m_\alpha=0}^\infty\frac{s^{m_\alpha}y_\alpha^{m_\alpha}}{m_\alpha!}\quad(\alpha=1,\ldots,k)$$

and

$$e^{sy_{k+\beta}}-1 = \sum_{p_\beta=1}^\infty\frac{s^{p_\beta}y_{k+\beta}^{p_\beta}}{p_\beta!}\quad(\beta=1,\ldots,l)$$

reduce (8) with $q = 1$ to

$$\begin{aligned}
&I_l(\mathbf{a};\theta,s,1) \\
&= \sum_{m=0}^\infty s^m\ {\sum}^*\frac{1}{\prod_{\alpha=1}^{k}m_\alpha!\prod_{\beta=1}^{l}p_\beta!} \\
&\quad\times\int_{\Delta_{k+l}}\prod_{\alpha=1}^{k}y_\alpha^{n_\alpha+m_\alpha}\prod_{\beta=1}^{l}y_{k+\beta}^{p_\beta}(y_{k+l+1})^{\theta-1}\frac{dy_1\cdots dy_{k+l}}{y_1\cdots y_{k+l}},
\end{aligned}$$

where the sum $\sum^*$ is taken over k-tuples $(m_1, \ldots, m_k)$ of nonnegative integers and l-tuples $(p_1, \ldots, p_l)$ of positive integers such that $m_1 + \cdots + m_k + p_1 + \cdots + p_l = m$. By using Dirichlet's formula, this can be rewritten into

$$\sum_{m=0}^{\infty} s^m \sum{}^{*} \frac{1}{\prod_{\alpha=1}^{k} m_\alpha! \prod_{\beta=1}^{l} p_\beta!} \cdot \frac{\prod_{\alpha=1}^{k} \Gamma(n_\alpha + m_\alpha) \prod_{\beta=1}^{l} \Gamma(p_\beta) \cdot \Gamma(\theta)}{\Gamma(n+m+\theta)},$$

and therefore

$$\begin{aligned} &\sum_{l=0}^{\infty} \frac{\theta^l}{l!} I_l(\mathbf{a}; \theta, s, 1) \\ &= \sum_{m=0}^{\infty} s^m \sum{}^{**} \frac{\prod_{\alpha=1}^{k} \Gamma(n_\alpha + m_\alpha) \cdot \Gamma(\theta)}{\prod_{\alpha=1}^{k} m_\alpha! \cdot \Gamma(n+m+\theta)} \sum{}^{***} \frac{\theta^l \prod_{\beta=1}^{l} \Gamma(p_\beta)}{l! \prod_{\beta=1}^{l} p_\beta!}, \end{aligned} \tag{26}$$

where $\sum^{**}$ indicates the sum taken over k-tuples $(m_1, \ldots, m_k)$ of nonnegative integers such that $0 \le m - (m_1 + \cdots + m_k) =: m_{k+1}$ and where $\sum^{***}$ is the sum taken over finite sequences $(p_1, \ldots, p_l)$ (with l varying) of positive integers summing up to m_{k+1}. By setting $b_i = \sharp\{\beta : p_\beta = i\}$ $(i = 1, \ldots, m_{k+1})$, this last sum is nothing but the following sum taken over all integer partitions $\mathbf{b} = (b_1, \ldots, b_{m_{k+1}})$ of m_{k+1}:

$$\sum{}^{***} \frac{\theta^l \prod_{\beta=1}^{l} \Gamma(p_\beta)}{l! \prod_{\beta=1}^{l} p_\beta!} = \frac{1}{m_{k+1}!} \sum_{\mathbf{b}} M_{m_{k+1}}(\mathbf{b}) \prod_{i=1}^{m_{k+1}} \left(\theta^{b_i} \Gamma(i)^{b_i}\right) = \frac{\Gamma(m_{k+1} + \theta)}{m_{k+1}! \, \Gamma(\theta)}.$$

Here the last equality follows from

$$M_{m_{k+1}}(\mathbf{b}) \prod_{i=1}^{m_{k+1}} \left(\theta^{b_i} \Gamma(i)^{b_i}\right) = P_\theta(\mathbf{A}_{m_{k+1}} = \mathbf{b}) \frac{\Gamma(m_{k+1} + \theta)}{\Gamma(\theta)}.$$

(Compare with (1).) Accordingly, the sum $\sum^{**}$ in (26) is equal to

$$\begin{aligned} &\sum \frac{1}{\prod_{\alpha=1}^{k+1} m_\alpha!} \cdot \frac{\prod_{\alpha=1}^{k} \Gamma(n_\alpha + m_\alpha) \cdot \Gamma(m_{k+1} + \theta)}{\Gamma(n+m+\theta)} \\ &= \frac{1}{m!} \int_{\Delta_k} \sum \frac{m!}{\prod_{\alpha=1}^{k+1} m_\alpha!} \cdot \prod_{\alpha=1}^{k} y_\alpha^{n_\alpha + m_\alpha} \cdot (y_{k+1})^{m_{k+1} + \theta - 1} \frac{dy_1 \cdots dy_k}{y_1 \cdots y_k}, \end{aligned}$$

where the both sums extend over $k+1$-tuples $(m_1, \ldots, m_{k+1})$ adding to m. So the multinomial theorem and Dirichlet's formula together reduce this to

$$\frac{1}{m!} \int_{\Delta_k} \prod_{\alpha=1}^{k} y_\alpha^{n_\alpha} (y_1 + \cdots + y_k + y_{k+1})^m (y_{k+1})^{\theta-1} \frac{dy_1 \cdots dy_k}{y_1 \cdots y_k} = \frac{\prod_{\alpha=1}^{k} \Gamma(n_\alpha) \cdot \Gamma(\theta)}{m!\, \Gamma(n+\theta)}.$$

Consequently (26) becomes

$$\sum_{l=0}^{\infty} \frac{\theta^l}{l!} I_l(\mathbf{a}; \theta, s, 1) = \sum_{m=0}^{\infty} \frac{s^m}{m!} \cdot \frac{\prod_{\alpha=1}^{k} \Gamma(n_\alpha) \cdot \Gamma(\theta)}{\Gamma(\theta+n)} = e^s \frac{\Gamma(\theta)}{\Gamma(\theta+n)} \prod_{i=1}^{n} \Gamma(i)^{a_i}.$$

In view of (1), this proves (25) for all $s > 0$. All the calculations seen in the above hold true for $s < 0$ because all the series appeared are absolutely convergent. The proof of Proposition 5 is now complete. □

Acknowledgment. The author thanks Professors Iizuka and Tachida for their interest in the subject of this paper. He is indebted to Lancelot James for comments on an earlier version of the manuscript and for bringing the work of Lo and Weng [12] to his attention.

References

[1] R. Arratia, A. D. Barbour, S. Tavaré. *Logarithmic Combinatorial Structures: a Probabilistic Approach.* European Mathematical Society, Zurich, 2003.

[2] P. Billingsley. *Convergence of Probability Measures.* 2nd edition. John Wiley & Sons, Inc., New York, 1999.

[3] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology* **3** (1972), 87–112; erratum, ibid. **3** (1972), 240; erratum, ibid. **3** (1972), 376.

[4] W. J. Ewens, S. Tavaré. Multivariate Ewens distribution. in: N. Johnson, S. Kotz, N. Balakrishnan (Eds.), *Discrete Multivariate Distributions.* John Wiley & Sons, Inc., New York, 1997, pp. 232-246.

[5] M. N. Grote, T. P. Speed. Approximate Ewens formulae for symmetric overdominance selection. *Ann. Appl. Probab.* **12** (2002), 637–663.

[6] L. F. James. Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. preprint, 2002. available at http://front.math.ucdavis.edu/math.PR/0205093

[7] L. F. James. Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statistics* **33** (2005), 1771–1799.

[8] J. F. C. Kingman. Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** (1975), 1–22.

[9] J. F. C. Kingman. Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A* **361** (1978), 1–20.

[10] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)* **18** (1978), 374–380.

[11] J. F. C. Kingman. *Poisson Processes.* Oxford University Press, New York, 1993.

[12] A. Y. Lo, C.-S. Weng. On a class of Bayesian nonparametric estimates, II, Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** (1989), 227–245.

[13] J. Pitman. Poisson-Kingman partitions. preprint, 1995.

[14] J. Pitman. Combinatorial stochastic processes. Technical Report No. 621, Dept. Statistics., U. C. Berkeley, 2002; Lecture notes for St. Flour course, July 2002. available at http://www.stat.berkeley.edu/users/pitman

[15] J. Pitman. Poisson-Kingman partitions. in: D. R. Goldstein (Ed.), *Science and Statistics: A Festschrift for Terry Speed.* Institute of Mathematical Statistics Hayward, California, 2003, pp. 1-34.

[16] J. Pitman, M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** (1997), 855–900.

[17] N. Tsilevich, A. Vershik, M. Yor. Distinguished properties of the gamma process, and related topics. Prépublication du Laboratoire de Probabilités et Modèles Aléatoires. No. 575, 2000. available at http://xxx.lanl.gov/ps/math.PR/0005287

[18] N. Tsilevich, A. Vershik, M. Yor. An infinite-dimensional analogue of the Lebesgue measure and distinguished properties of the gamma process. *Journ. Funct. Anal.* **185** (2001), 274–296.

[19] G. A. Watterson. Heterosis or neutrality ? *Genetics* **85** (1977), 789–814.

[20] G. A. Watterson. The homozygosity test of neutrality. *Genetics* **88** (1978), 405–417.