Japan J. Indust. Appl. Math., 24 (2007), 119–130 Area ⟨1⟩

Effect of the Finite Size of Population on the Distribution of Family Names

Kazunori SATO* and Atsushi OGURI

Shizuoka University
Hamamatsu, 432–8561, Japan
**E-mail: sato@sys.eng.shizuoka.ac.jp*

Received August 4, 2006
Revised October 14, 2006

We examine the effect of finite size of population on the distribution of family names. As the result we observe that the power-law behavior of size-frequency distribution in Reed–Hughes ([15]) model collapses to show the convex shape on the logarithmic graph. We can approximately calculate the average distribution of size-frequency distribution of family names obtained by the similar method for Ewens sampling formula.

Key words: family names, distribution on logarithmic graph, finite size of population, random partitions

1. Introduction

Up to the present, various data indicate the power-law of size-frequency distribution for the family names, and the several models are shown to reveal the mechanism to generate it ([12, 14, 15, 16, 17, 18]). Especially, our model in this paper is based on the model by [15], so we briefly explain it here.

Reed and Hughes distinguish two dynamics for the population dynamics within each family name and the number of family names. The former is defined by the continuous-time birth process with immigration, and the latter by discrete-time Galton-Watson branching process. Then they obtain an asymptotic size-frequency distribution by power-law, which determines the relative effect of the model parameters on the data, i.e. the birth rate for the number of family names and the population growth rate within each family name. In this asymptotic behavior the immigration rate never concerns to the distribution, so one may consider that the immigration term has no effect on the dynamics, but we have the trivial state as an equilibrium, i.e. no individuals and no family names, when the immigration rate becomes zero.

In this paper we examine that the finite size effect of population by the introduction of density dependent population growth to Reed–Hughes model, which naturally restrict the number of family names under the population dynamics. In the study of population ecology, it is common to assume that the resource such as habitat or food are restricted, and we observe the finite size of population as the result. Especially logistic model have often been used and succeeded to explain the phenomena qualitatively, partially because of its tractability to analyse and its simplicity to be modelling. Also in the dynamics of family names it seems to be natural to introduce the finiteness of population size and to be interesting to examine its effect.

When we analyze our model, we find it rather difficult to evaluate its characteristics, e.g. size-frequency distribution, quantitatively. Therefore we adopt an approximation under the assumption that the total population size remains around an average and the probability distribution for the size-frequency distribution reaches an equilibrium. In that case we can apply the technique of random partition, which have been used in the analyses of neutral allele models in the field of population genetics. Ewens sampling formula may be one of the most investigated distribution since [6], which is proved by [10]. Though the derivation of this Ewens sampling formula seems rather complicated (e.g. [13]), several authors independently devise its derivation by the urn model or its modified ([8, 9, 11]). Among these, we adopt the "finitary characterization" method proposed recently by the group of Costantini, which has been developed in the field of economics (e.g. [2, 5, 7] and the references therein), which seems to be much easier to understand. They also clarify the differences between their method and others' ([5]).

2. Model

We construct the population dynamics as the stochastic processes in continuous time for each family name with logistic type of density dependent effect, and as well as for the number of family names can change exponentially (alteration or modification as an interpretation) with immigration from the outside. As the former process, we simply assume that the birth occurs at a proportinal (or linear) to the population size and the death occurs by the term of second degree, although there are infinitely many combinations for these rates to be considered (e.g. [1]).

First of all, we consider the dynamics of population size in the case of fixed number of family names. In the present at time t there exists l family names with population size $m^{(i)} = m^{(i)}(t)$, $i = 1, 2, \ldots, l$, then the joint probability of population size $P(m^{(1)}, \ldots, m^{(l)})$ at time t will change as

$$\begin{aligned}
&\frac{dP(m^{(1)}, \ldots, m^{(l)})(t)}{dt} \\
&= \delta \sum_{i=1}^{l} (m^{(i)} - 1) P(m^{(1)}, \ldots, m^{(i)} - 1, \ldots, m^{(l)})(t) \\
&\quad + \delta \sum_{i=1}^{l} (m^{(i)} + 1) \frac{1 + \sum_{j=1}^{l} m^{(j)}}{K} P(m^{(1)}, \ldots, m^{(i)} + 1, \ldots, m^{(l)})(t) \qquad (1) \\
&\quad - \delta \left(\sum_{i=1}^{l} m^{(i)} \right) P(m^{(1)}, \ldots, m^{(l)})(t) \\
&\quad - \delta \frac{\left(\sum_{i=1}^{l} m^{(i)} \right)^2}{K} P(m^{(1)}, \ldots, m^{(l)})(t),
\end{aligned}$$

where the summation of the first term in RHS takes i only for $m^{(i)} \geq 1$. The parameter δ corresponds to the intrinsic rate of natural increase for the average size of total population, which will be explained below, and here we assume that the dynamics follows by two processes: density-independent birth and the density-dependent death. The first and the third terms indicate the increase of population size by exponential growth corresponding to Galton–Watson branching processes by Reed–Hughes model. The remaining second and fourth terms represent the death events by density dependent effect, which is characterized by the multiple of the population size in the focal family name by the total populaton size.

This equation is rather complicated, so we consider the following form of the same equation instead. Let define $P_{m,n}$ as the probability that the popoulation size in focal family name is m and the total population size is $n = n(t) = \sum_{i=1}^{l} m^{(i)}(t)$. Here notice that $m \leq n$ always holds by definition. Then we can rewrite the population dynamics by this probability as

$$\begin{aligned}\frac{dP_{m,n}(t)}{dt} &= \delta(m-1)P_{m-1,n-1}(t) + \delta(n-1-m)P_{m,n-1}(t)\\ &\quad + \frac{\delta(m+1)(n+1)}{K}P_{m+1,n+1}(t) + \frac{\delta(n+1-m)(n+1)}{K}P_{m,n+1}(t)\\ &\quad - \delta n P_{m,n} - \frac{\delta n^2}{K}P_{m,n}(t),\end{aligned} \tag{2}$$

where the first and the second temrs disappear when m or n is less than 1. When we multiply n and take summation on m and n on both sides, then

$$\text{LHS} = \frac{d}{dt}\left(\sum_{n=0}^{\infty} n \sum_{m=0}^{n} P_{m,n}(t)\right) = \frac{d\widehat{n}}{dt}$$

and

$$\begin{aligned}\text{RHS} = \delta &\sum_{n=1}^{\infty}\sum_{m=1}^{n}(m-1)nP_{m-1,n-1}(t)\\ &+ \delta\sum_{n=1}^{\infty}\sum_{m=0}^{n-1} n(n-1-m)P_{m,n-1}(t)\\ &+ \frac{\delta}{K}\sum_{n=0}^{\infty}\sum_{m=0}^{n} n(m+1)(n+1)P_{m+1,n+1}(t)\\ &+ \frac{\delta}{K}\sum_{n=0}^{\infty}\sum_{m=0}^{n+1} n(n+1-m)(n+1)P_{m,n+1}(t)\\ &- \delta\sum_{n=0}^{\infty}\sum_{m=0}^{n} n^2 P_{m,n} - \frac{\delta}{K}\sum_{n=0}^{\infty}\sum_{m=0}^{n} n^3 P_{m,n}(t)\\ = \delta\widehat{n} &- \frac{\delta}{K}\widehat{n^2},\end{aligned}$$

where we use the relation $\sum_{m=0}^{n} P_{m,n} = P_n$, which is the probability that the total population size is equal to n, and introduce new notations as $\widehat{n} = \sum_{n=0}^{\infty} nP_n$ (i.e. mean or first order moment) and $\widehat{n^2} = \sum_{n=0}^{\infty} n^2 P_n$ (i.e. mean square or second order moment).

In the second place, the probability of i distinct kinds of family names q_i will change by the following exponential growth at rate λ with immigration at rate ρ in the case of no death of family names, i.e. no family names from one individual to zero (see [15])

$$\begin{cases} \dfrac{dq_i(t)}{dt} = \lambda(i-1)q_{i-1}(t) - \lambda i q_i(t) + \rho q_{i-1}(t) - \rho q_i(t), \quad (i \geq 1) \\ \dfrac{dq_0(t)}{dt} = -\rho q_0(t). \end{cases} \tag{3}$$

Namely, the average size of total population is considered to follow by the logistic type of dynamics with intrinsic rate of natural increase δ and the carrying capacity K, and the number of family names changes by the exponential growth (at rate λ) with immigration (at rate ρ). We regard it that the latter dynamics is exactly the same as Reed–Hughes model but the former is modified for Reed–Hughes model with two differences: time development by discrete-time to continuous-time and branching process (without carrying capacity) to logistic growth (with carrying capacity).

So far it is not clear the relation between the probabilities $P_{m,n}$ and q_i whose dynamics are given by equation (1) (or equation (2)) and equation (3). Indeed we should introduce the terms corresponding to the dynamics of population size with the creation of new family names and the annihilation of existent family names with only one individual by the death, and these terms combine the two kinds of dynamics dependent each other. However, the equations becomes complicated but not useful even when we explicitly add the expression of these terms, so we omit it.

In the case that the total population remains around the carrying capacity n, we assume to consider approximately the population size n as the quasi-steady state then we can evaluate qualitative results for the average size-frequency distribution or the probability distribution of the number of family names (or the average number of family names). In this paper we only concentrate on the former distribution.

Then we can approximately consider the above process at the quasi-steady state as the following dynamics with fixed population size. Let n as the fixed number of total population size. Assume that there exist k family names in the present. We choose an individual at random with equal probability, and replace it either by one of the remaining $n-1$ individuals with probability $\delta/(\delta+\rho+k\lambda)$ or by an individual from outside of the population with probability $(\rho+k\lambda)/(\delta+\rho+k\lambda)$, which means that an individual with new family name immigrates from outside or the novel family name is created from each family name by modification or mistake.

We can interpret the above dynamics as the following random partition ([5]). We fix the population size as n, and introduce g kinds of family names, which includes nonexistent ones, and this number g is large enough compared to the total population size n. Let $\mathbf{n} = (n_1, \ldots, n_g)$ as g-dimensional vector, whose i-th element n_i indicates the population size of i-th family name. In the initial condition we set $k \leq n$ distinct family names and so $\mathbf{n} = (n_1, \ldots, n_k, 0, \ldots, 0)$. We define the set $A = \{i \colon n_i > 0\}$, then the number of family names k equals $|A|$, which represents the cardinality of A. We put the conditional probability

$$P(\mathbf{n}_i^j \,|\, \mathbf{n}) = \begin{cases} \dfrac{n_i}{n} \dfrac{n_j - \delta_{ij}}{(\rho + k\lambda)/\delta + n - 1} & \text{for} \quad n_j > 0 \\ \dfrac{n_i}{n} \dfrac{1}{g-k} \dfrac{(\rho + k\lambda)/\delta}{(\rho + k\lambda)/\delta + n - 1} & \text{for} \quad n_j = 0, \end{cases}$$

where δ_{ij} indicates the Kronecker's delta, $\mathbf{n}_i^j = (n_1, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_g)$ for $i \neq j$, and $\mathbf{n}_i^i = \mathbf{n}$. This vector denotes the resulting state after the transitions that the death of an individual in i-th family name occurs and then the birth in j-th does. The rate of the event of birth differs dependent on either the present family names or the absent. Notice that this process gives Ewens sampling formula, when we replace $(\rho + k\lambda)/\delta$ with θ, because the event by this rate, which is normalized so that ratio of the rate of newborn by the absent family names to the present, corresponds to the mutation, i.e. the normalized rate that the novel family names is created ([5]). Using this conditional probability with detailed balance condition, then we can obtain the following equilibrium distribution (Appendix A):

$$P(\mathbf{n}) \propto \frac{(g-k)!\,(\lambda/\delta)^k \Gamma(k + \rho/\lambda)}{\prod_{j \in A} n_j} \tag{4}$$

where Γ indicates the gamma function, then we can rewrite it by the partition vector $\mathbf{z} = (z_1, \ldots, z_n)$, $z_i = |\{j \colon n_j = i,\, j = 1, \ldots, n\}|$ as

$$P(\mathbf{z}) \propto \left\{ \prod_{i=1}^{n} \left(\frac{\lambda}{i\delta} \right)^{z_i} \frac{1}{z_i!} \right\} \Gamma\left(\sum_{i=1}^{n} z_i + \frac{\rho}{\lambda} \right), \tag{5}$$

where we use the relation $k = \sum_{i=1}^{n} z_i$ and $n = \sum_{i=1}^{n} i z_i$, and the fact that the number of distinct family names with the same $\mathbf{z}$ is $g!/((g-k)!\,z_1! \cdots z_n!)$. The proportion constant is the reciprocal number of the summation of the RHS in equation (5) over all possible partitions, which we use to depict the figure in the next section. In order to check whether this equilibrium distribution obeys power-law or not, however, we need not this proportion constant for this probability, because it only affects the value of an intercept on the vertical axis but not the form of the graph.

3. Results

Here we show the time evolution by the original dynamics combined both equation (1) (or equation (2)) with the increase by new born family names and

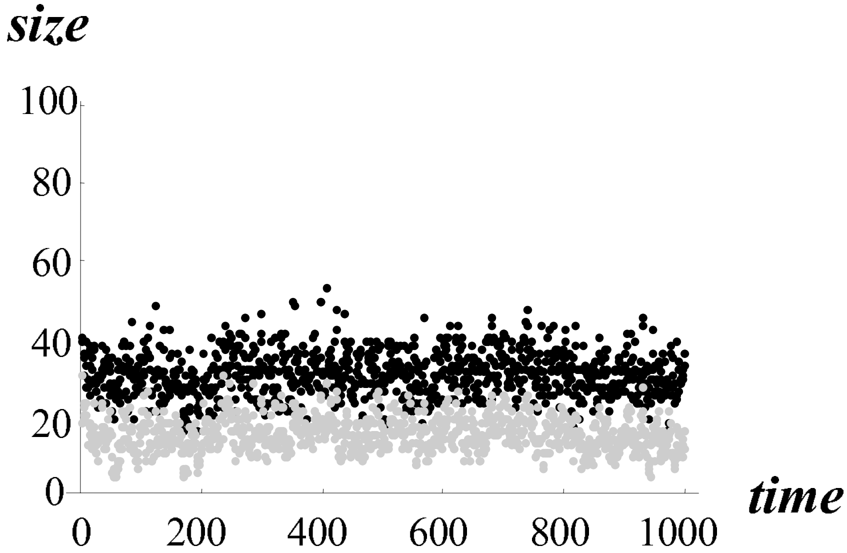

Fig. 1. Time evolution of the number of total population size (black circles) and family names (gray circles) by Monte Carlo simulation corresponding to the combined dynamics of equation (1) (or equation (2)) with the increase by new born family names and equation (3) with the decrease caused by the annihilation of family names of population size equal to unity.

equation (3) with the decrease caused by the annihilation of family names of population size equal to unity (Fig. 1). This figure suggests that the dynamics reaches an equilibrium distribution quickly which results in the quick convergence to the total population size and the number of family names within some range. Here notice that it is rather difficult to enumerate all partition vectors even for small value n, but in turn we have the large fluctuation in time when the mean total population size becomes small. For example, the number of the kinds of partition vectors can be calculated as 5604 for $n = 30$, which is obtained e.g. by Mathematica®. We depict Fig. 1 with the parameters $\delta = 1.0$, $\lambda = 0.75$, $\rho = 0.1$, in which the ratio $\lambda/\delta = 0.75$ corresponds to the data for Japanese family names (see [14]), starting from 100 individuals whose family names are different each other as an initial condition. The remaining parameter K is adjusted so that the average total population size becomes nealy equal to 30, which results in $K = 23$, because we cannot succeed to evaluate both the mean total population size and the mean number of family names on the steady state distribution, then we can compare the result by equation (5) setting with the same parameters for δ, λ, ρ and $n = 30$.

Using this parameter set, Fig. 2 gives the mean equilibrium distribution for size-frequency to examine the quantitative accuracy of equation (5) compared to Monte Carlo simulation. At a glance, our approximation scheme by random partition with

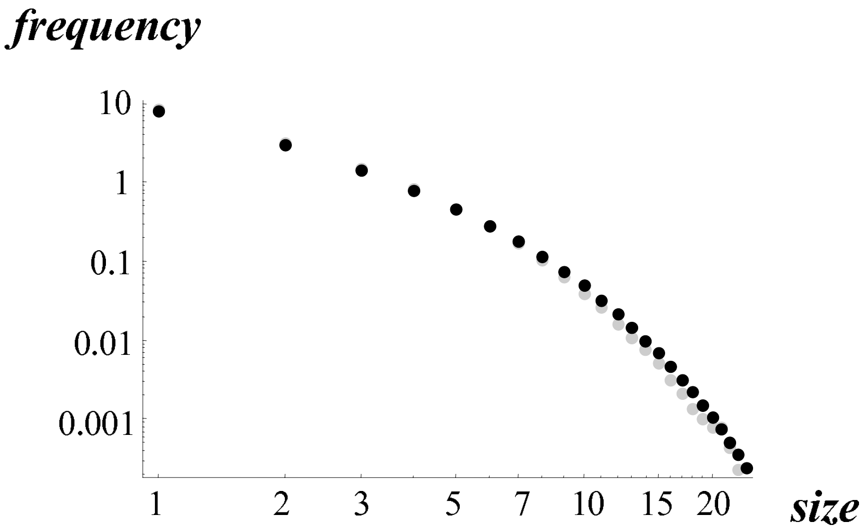

Fig. 2. Size-frequency distribution at equilibrium. Black circles: equation (5), Gray circles: Monte Carlo simulation.

fixed number of total population size succeeds to obtain the satisfactory accuracy for the average size-frequency distribution.

We calculate the average distribution as follows. In the approximation using the partition vector equation (5), it becomes

$$\sum_{\mathbf{z}} P(\mathbf{z})\mathbf{z} = \left(\sum_{\mathbf{z}} P(\mathbf{z})z_1, \ldots, \sum_{\mathbf{z}} P(\mathbf{z})z_{30}\right),$$

where the summation is taken over 5604 kinds of partition vector $\mathbf{z}$. On the other hand in Monte Carlo simulation, we consider the partition vector

$$\mathbf{z}^{(i)} = \left(z_1^{(i)}, \ldots, z_{23}^{(i)}\right), \quad i = 1, \ldots, 1000,$$

in which the dimension of 23 and the number 1000 correspond to the value of K and the number of realizations, respectively. We show an average of these 1000 vectors in the figure:

$$\frac{1}{1000}\sum_{i=1}^{1000} \mathbf{z}^{(i)} = \left(\frac{\sum_{i=1}^{1000} z_1^{(i)}}{1000}, \ldots, \frac{\sum_{i=1}^{1000} z_{23}^{(i)}}{1000}\right).$$

We can see by the graphs that the size-frequency distribution becomes a convex function instead of straight line, which may be attributed to the effect of finite size population.

4. Discussion

In this paper we approximate the dynamics of Reed–Hughes model with finite size population under the assumption that the total population size reaches an equilibrium. When we compare this approximation to Monte Carlo simulation, it seems that this approximation gives quantitaively accurate distribution for this model but we should continue to do exact analyses.

In the real population dynamics in the world, we have not yet had the evidence to support the logistic growth of population, but several people consider that the population in the world have already reached into the region of carrying capacity ([4]). Besides that, it is theoretically meaningful to consider the case of finite size of population caused by logistic-type density dependent effect as the starting point.

As the result we obtained the equilibrium distribution similar convex graph as the exponential growth population model instead of power-law. When [3] succeeded to construct the scale-free network with power-law distribution for degree distribution, they pointed out two important features of network construction; the preference attachement and the network growth. Our model only has the former character but not the latter which may result in discrepancy with the power-law distribution.

We have not yet determined both the constant parameter for the eqilibrium distribution and the average distribution exactly, and we leave them as the future problems.

One of the interesting and important factors for the dynamics of family names may be the introduction of spatial structure of populations. For example, we can consider the metapopulation for Japanese populations divided by 47 prefectures. Although Tokyo or Osaka has large populations with large degrees, Shizuoka belongs to intermediate size of population. Directed graph (or network) can be a possible model to capture the real dynamics. Besides that we will try to find the general mechanism for power-law by the analyses in the case of foreign countries.

Finally we want to comment on the relation between the size-frequency distribution and the order-size distribution because we are concerned with the possibility of power-law behavior for various points of view by the same data. As shown in [14], these two relations are sometimes considered separately whether the power-law distributions hold or not. However, we can estimate the latter distribution when we assume the former distribution in the following. Indeed, as [15] shows, suppose that the size(m)-frequency(f) distribution obeys the nearly truncated power-law distribution with maximum size m_{max} and the exponent $\gamma = 2$

$$\begin{aligned} f(m) &= \left(1 + \frac{1}{m_{\max}}\right) \frac{\Gamma(m)}{\Gamma(m+2)} \\ &= \left(1 + \frac{1}{m_{\max}}\right) \frac{1}{m(m+1)} \end{aligned} \tag{6}$$

then the order(x)-size(m) distribution has the following distribution

$$m(x) = \frac{n+1-x}{x-1+n/m_{\max}}, \quad (x \geq 1), \tag{7}$$

which can be obtained as the solution of the following equation on m

$$\begin{aligned} x &= 1 + n\left(1 - \frac{\sum_{i=1}^{m} \Gamma(i)/\Gamma(i+2)}{\sum_{i=1}^{m_{\max}} \Gamma(i)/\Gamma(i+2)}\right) \\ &= 1 + \frac{n(m_{\max}-m)}{m_{\max}(m+1)}. \end{aligned}$$

Here the term $1-\left(\sum_{i=1}^{m} \Gamma(i)/\Gamma(i+2)\right)/\left(\sum_{i=1}^{m_{\max}\Gamma(i)/\Gamma(i+2)}\right)$ represents the fraction of individuals who belong to the family names with population size larger than m, and so the multiplication by total population size n equals to the population size of them. As an continuum approximation for x, which should be integer by definition in fact, it is natural to assume that the order of the family name whose population size is m becomes this value added to unity. We show these relations in Fig. 3, and it is clear that the latter distribution does not follow the power-law when the former does but instead we can observe almost straight lines with different slopes for both sides of the graph (power-law in each part) as indicated by [14]. In the similar way, we can consider the case exactly following the truncated power-law distribution with continuum approximation for x (Appendix B).

Appendix A. Probability function of n at steady state

We derive the probability function of $\mathbf{n}$ at steady state following the method by [5]. Suppose that $P(\mathbf{n}) \propto f(k)/g(\mathbf{n})$, where $k = k(\mathbf{n})$ is the cardinality of $A = \{i\colon n_i > 0\}$. For $i \neq j$ we can derive $P(\mathbf{n})$ at steady state by the detailed balance equation:

$$P(\mathbf{n}_i^j)P(\mathbf{n} \mid \mathbf{n}_i^j) = P(\mathbf{n})P(\mathbf{n}_i^j \mid \mathbf{n}).$$

(i) The case for $n_j > 0$.
We have $k(\mathbf{n}_i^j) = k(\mathbf{n})$, $P(\mathbf{n}_i^j \mid \mathbf{n}) = Bn_in_j$, $P(\mathbf{n} \mid \mathbf{n}_i^j) = B(n_i-1)(n_j+1)$, where B is the constant parameter equal to $\{n((\rho+k\lambda)/\delta+n-1)\}^{-1}$. Then

$$\frac{P(\mathbf{n}_i^j)}{P(\mathbf{n})} = \frac{P(\mathbf{n}_i^j \mid \mathbf{n})}{P(\mathbf{n} \mid \mathbf{n}_i^j)} = \frac{n_i}{n_i-1}\frac{n_j}{n_j+1},$$

which is satisfied by $g(\mathbf{n}) = \prod_{j \in A} n_j$.

(ii) The case that $n_j = 0$.
(ii-1) The case that $n_i > 1$.
We have $k(\mathbf{n}_i^j) = k(\mathbf{n}) + 1$, $P(\mathbf{n}_i^j \mid \mathbf{n}) = B(1/(g-k))n_i((\rho+k\lambda)/\delta)$, $P(\mathbf{n} \mid \mathbf{n}_i^j) = B(n_i - 1)$. Then

$$\frac{P(\mathbf{n}_i^j)}{P(\mathbf{n})} = \frac{P(\mathbf{n}_i^j \mid \mathbf{n})}{P(\mathbf{n} \mid \mathbf{n}_i^j)} = \frac{n_i}{n_i-1}\frac{\rho/\lambda+k}{(g-k)\delta/\lambda},$$

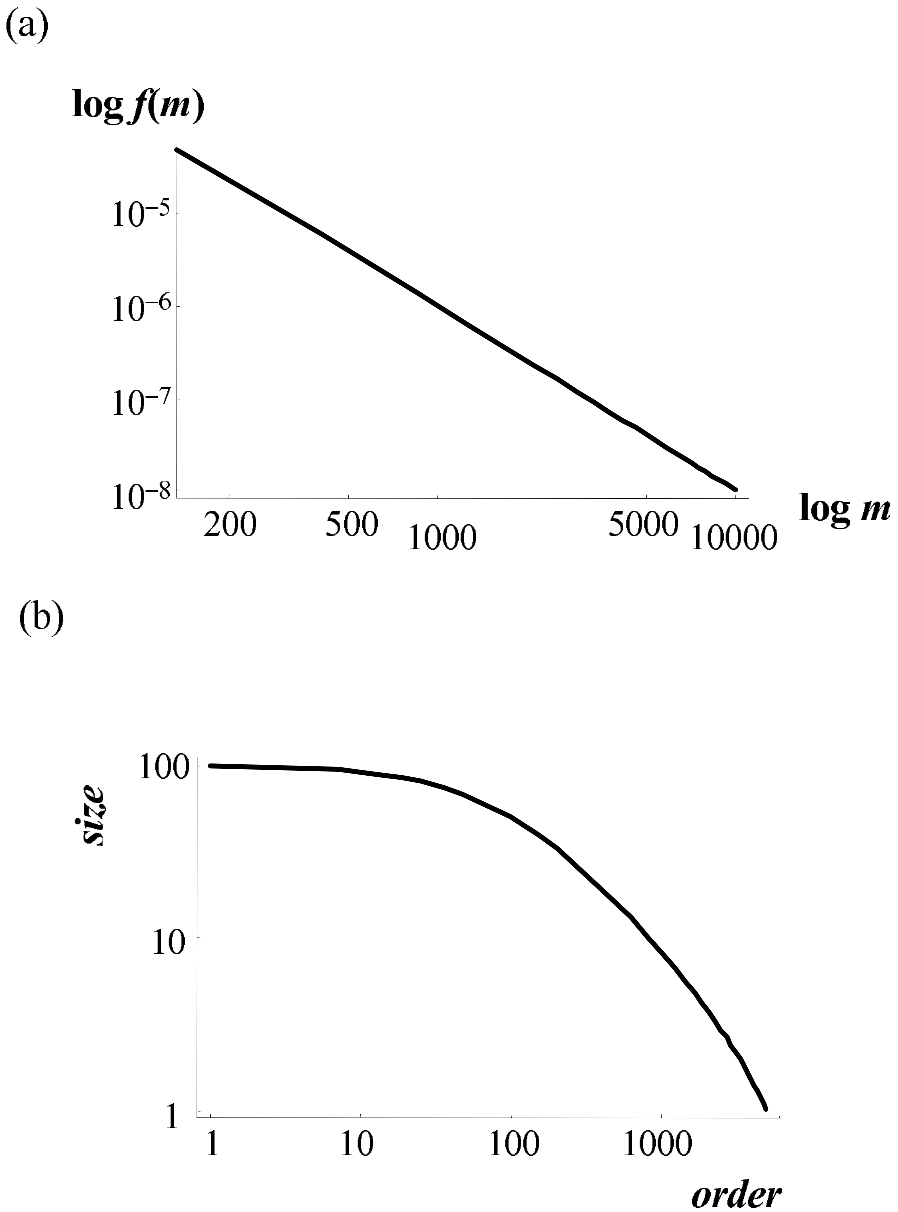

Fig. 3. Relation between size-frequency distribution and order-size distribution. (a) assumed size-frequency distribution by equation (6). (b) resulted order-size distribution by equation (7).

which is satisfied by $f(k) = (g-k)!\,(\lambda/\delta)^k \Gamma(\rho/\lambda + k)$.

(ii-2) The case that $n_i = 1$.

We have $k(\mathbf{n}_i^j) = k(\mathbf{n})$, $P(\mathbf{n}_i^j \mid \mathbf{n}) = P(\mathbf{n} \mid \mathbf{n}_i^j) = B(1/(g-k))((\rho + k\lambda)/\delta)$, and

$$\frac{P(\mathbf{n}_i^j)}{P(\mathbf{n})} = \frac{P(\mathbf{n}_i^j \mid \mathbf{n})}{P(\mathbf{n} \mid \mathbf{n}_i^j)} = 1,$$

which is satisfied by the above definition of $g(\mathbf{n})$ and $f(k)$.

Appendix B. Relation between size-frequency distribution and order-size distribution for truncated power-law distribution

We assume size(m)-frequency(f) distribution exactly obeys the truncated power-law distribution with maximum size $m_{\max}$ and the exponent γ:

$$f(m) = \frac{m^{-\gamma}}{\sum_{l=1}^{m_{\max}} l^{-\gamma}},$$

namely,

$$\log f(m) = -\gamma \log m - \log \sum_{l=1}^{m_{\max}} l^{-\gamma}.$$

By the similar continuum approximation on order x, we can obtain the relation between order (x) and size (m) as follows:

$$x = 1 + n\left(1 - \frac{\sum_{i=1}^{m} i^{-\gamma}}{\sum_{i=1}^{m_{\max}} i^{-\gamma}}\right).$$

However, it is difficult to obtain the explicit solution on the size m, and then we can evaluate the relation by numerical calculation, which gives the similar graph as Fig. 3.

Acknowledgements. We appreciate Dr. Hiromi Seno, whose comment helps us to lead to the approximate calculation using the method for Ewens sampling formula. Also we thank two anonymous reviewers to give us constructive and helpful comments to revise the manuscript.

References

[1] L.J.S. Allen, An Introduction to Stochastic Processes with Applications to Biology. Pearson Prentice Hall, New Jersey, 2003.

[2] M. Aoki, Modeling Aggregate Behavior and Fluctuations in Economics: Stochastic Views of Interacting Agents. Cambridge University Press, Cambridge, 2002.

[3] A.-L. Barabási and R. Albert, Emergence of scaling in random networks. Science, **286**, (1999) 509–512.

[4] J.E. Cohen, How Many People Can the Earth Support? W.W. Norton & Company, Inc., New York, 1995.

[5] D. Costantini, U. Garibaldi and P. Viarengo, A finitary characterization of Ewens sampling formula. `http://cinef.dibe.unige.it/WEHIA2003/Finitary_ESF.pdf`

[6] W.J. Ewens, The sampling theory of selectively neutral alleles. Theor. Popul. Biol., **3**, (1972) 87–112.

[7] U. Garibaldi, D. Costantini, S. Donadio and P. Viarengo, Herding and clustering in economics: The Yule-Zipf-Simon model. Comp. Econ., **27**, (2006) 115–134.

[8] F.M. Hoppe, Pólya-like urns and the Ewens' sampling formula. J. Math. Biol., **20**, (1984) 91–94.

[9] F.M. Hoppe, The sampling theory of neutral alleles and an urn model in population genetics. J. Math. Biol., **25**, (1987) 123–159.

[10] S. Karlin and J.L. McGregor, Addendum to a paper of W. Ewens. Theor. Popul. Biol., **3**, (1972) 113–116.

[11] F.P. Kelly, Reversibility and Stochastic Networks. John Wiley & Sons, Chichester, 1979.

[12] B.J. Kim and S.M. Park, Distribution of Korean family names. Physica A, **347**, (2005) 683–694.

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

[14] S. Miyazima, Y. Lee, T. Nagamine and H. Miyajima, Power-law distribution of family names in Japanese societies. Physica A, **278**, (2000) 282–288.

[15] W.J. Reed and B.D. Hughes, On the distribution of family names. Physica A, **319**, (2003) 579–590.

[16] Y. Sato and H. Seno, Mathematical Ecology on Succession and Extinction of Family Names. Kyoto University Press, Kyoto, 2003 (in Japanese).

[17] H.A. Simon, Models of Man. Wiley, New York, 1957.

[18] D.H. Zanette and S.C. Manrubia, Vertical transmission of culture and the distribution of family names. Physica A, **295**, (2001) 1–8.