Electron. Commun. Probab. **21** (2016), no. 83, 1–8.
DOI: 10.1214/16-ECP34
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Large deviations for homozygosity*

Donald A. Dawson† Shui Feng‡

Abstract

For any $m \geq 2$, the homozygosity of order m of a population is the probability that a sample of size m from the population consists of the same type individuals. Assume that the type proportions follow Kingman's Poisson-Dirichlet distribution with parameter θ. In this paper we establish the large deviation principle for the naturally scaled homozygosity as θ tends to infinity. The key step in the proof is a new representation of the homozygosity. This settles an open problem raised in [1]. The result is then generalized to the two-parameter Poisson-Dirichlet distribution.

Keywords: Dirichlet process; Poisson-Dirichlet distribution; homozygosity; large deviation.
AMS MSC 2010: 65G17; 65G60.
Submitted to ECP on August 17, 2016, final version accepted on November 29, 2016.

1 Introduction

Let $\gamma(t)$ denote the gamma subordinator with Lévy measure

$$\Lambda(dx) = x^{-1}e^{-x}dx, \quad x > 0.$$

For any $\theta > 0$, let $J_1(\theta) \geq J_2(\theta) \geq \cdots$ denote the jump sizes of $\gamma(t)$ over the interval $[0, \theta]$ in descending order. If we set $P_i(\theta) = J_i(\theta)/\gamma(\theta), i \geq 1$, then the law of

$$\mathbf{P}(\theta) = (P_1(\theta), P_2(\theta), \ldots)$$

is Kingman's Poisson-Dirichlet distribution $PD(\theta)$(cf.[10]). It is a probability on the infinite-dimensional simplex

$$\nabla_\infty = \{\mathbf{p} = (p_1, p_2, \ldots) : p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} p_i \leq 1\}.$$

For any integer $m \geq 2$, the function

$$H(\mathbf{p}; m) = \sum_{i=1}^{\infty} p_i^m$$

*Supported by the Natural Sciences and Engineering Research Council of Canada.
†School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6. E-mail: ddawson@math.carleton.ca
‡Department of Mathematics and Statistics, McMaster University, Hamilton, Canada L8S 4K1.
E-mail: shuifeng@mcmaster.ca

is loosely called the homozygosity of order m. The name is taken from population genetics where the homozygosity corresponds to $m = 2$. The function is closely associated with the Shannon entropy in communication, the Herfindahl-Hirschman index in economics, and the Gini-Simpson index in ecology. It can be used to measure the population diversity in terms of the number of different types and the evenness in the distribution among those types. The value of $H(\mathbf{p}; m)$ decreases when the number of types increases and the distribution among those types becomes more even.

In this paper we are interested in the behaviour of the random variable $H(\mathbf{P}(\theta); m)$ when θ tends to infinity. When a random sample of size m is selected from a population whose individual types have distribution $PD(\theta)$, the probability that all samples are of the same type is given by $H(\mathbf{P}(\theta); m)$. Since $H(\mathbf{P}(\theta); m) \leq P_1^{m-1}(\theta)$, it follows that $H(\mathbf{P}(\theta); m)$ converges to zero as θ approaches infinity. In [7] and [9] it is shown that $H(\mathbf{P}(\theta); m)$ goes to zero at a magnitude of $\frac{\Gamma(m)}{\theta^{m-1}}$, and

$$\sqrt{\theta}\left[\frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m) - 1\right] \Rightarrow Z_m \tag{1.1}$$

where $\Rightarrow$ denotes convergence in distribution and Z_m is a normal random variable with mean zero and variance

$$\frac{\Gamma(2m)}{\Gamma^2(m)} - m^2.$$

It is natural to investigate more refined structures associated with the limits

$$H(\mathbf{P}(\theta); m) \to 0, \ \theta \to \infty$$

and

$$\frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m) \to 1, \ \theta \to \infty.$$

In [1], a full large deviation principle is established for $H(\mathbf{P}(\theta); m)$ describing the deviations from zero. For l in $(0, 1/2)$, the quantity $\theta^l \left(\frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m) - 1\right)$ converges to zero in probability as θ tends to infinity. Large deviations associated with this limit are called the *moderate deviation principle* for $\{\frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m) : \theta > 0\}$. In [5], the moderate deviation principles are shown to hold for l in $(\frac{m-1}{2m-1}, \frac{1}{2})$. The large deviation principle corresponding to $l = 0$ remains an open problem.

In this paper we will solve this open problem, namely, the large deviation principle for $\frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m)$ describing deviations from 1. The two-parameter generalization is also obtained. The key in the proof is a new representation of the homozygosity.

2 Large deviations

Let m is any integer that is greater than or equal to 2. The objective of this section is to establish the large deviation principle for

$$L(\mathbf{P}(\theta); m) = \frac{\theta^{m-1}}{\Gamma(m)} H(\mathbf{P}(\theta); m).$$

We begin with the case that θ takes integer values. For any $1 \leq k \leq \theta$, let $J_i^k, i = 1, 2, \ldots$ denote all the jump sizes of $\gamma(t)$ over $[k-1, k]$. Since the subordinator $\gamma(t)$ will not jump at $t = 0, 1, \ldots, \theta$ with probability one, it follows that

$$H(\mathbf{P}(\theta);m) = \frac{1}{\gamma^m(\theta)}\sum_{k=1}^{\theta}(\gamma(k)-\gamma(k-1))^m\sum_{i=1}^{\infty}\left(\frac{J_i^k}{\gamma(k)-\gamma(k-1)}\right)^m \tag{2.1}$$
$$= \frac{1}{\gamma^m(\theta)}\sum_{k=1}^{\theta}W_k^m H_k$$

where $W_1,\dots,W_\theta$ are independent copies of $\gamma(1)$, and independently, $H_1,\dots,H_\theta$ are independent copies of $H(\mathbf{P}(1);m)$. Set

$$L_0(\mathbf{P}(\theta);m) = \frac{1}{\Gamma(m)\theta}\sum_{k=1}^{\theta}W_k^m H_k.$$

Then we have

$$L(\mathbf{P}(\theta);m) = \frac{\theta^m}{\gamma^m(\theta)}L_0(\mathbf{P}(\theta);m).$$

Theorem 2.1. *A large deviation principle holds for $L(\mathbf{P}(\theta);m)$ as θ converges to infinity on space $\mathbb{R}$ with rate $\theta^{1/m}$ and good rate function*

$$I(x) = \begin{cases} [\Gamma(m)(x-1)]^{1/m}, & x \geq 1, \\ +\infty, & otherwise. \end{cases}$$

Proof: By Ewens sampling formula and direct calculation we have

$$\mathbb{E}[W_1^m H_1] = \Gamma(m).$$

Let

$$J(y) = \sup\left\{\lambda y - \log\mathbb{E}[e^{\lambda\frac{W_1^m H_1}{\Gamma(m)}}] : \lambda\in\mathbb{R}\right\}.$$

Since $\mathbb{E}[e^{\lambda\frac{W_1^m H_1}{\Gamma(m)}}] = \infty$ for $\lambda > 0$, it follows that

$$J(y) = \sup\left\{\lambda y - \log\mathbb{E}[e^{\lambda\frac{W_1^m H_1}{\Gamma(m)}}] : \lambda \leq 0\right\}.$$

By Cramér's theorem (cf. Theorem 2.2.3 in [2]), we have that for any x

$$\limsup_{\theta\to\infty}\theta^{-1}\log\mathbb{P}\{L_0(\mathbf{P}(\theta);m)\leq x\} \leq -\inf_{y\leq x}J(y). \tag{2.2}$$

Rewrite $J(y)$ as

$$\sup\left\{\log e^{\lambda y} - \log\mathbb{E}[e^{\lambda\frac{W_1^m H_1}{\Gamma(m)}}] : \lambda\leq 0\right\}$$
$$= \sup\left\{-\log\mathbb{E}\left[\exp\left\{\lambda\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\right\}\right] : \lambda\leq 0\right\}$$
$$= -\inf\left\{\log\mathbb{E}\left[\exp\left\{\lambda\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\right\}\right] : \lambda\leq 0\right\}$$
$$= -\log\inf\left\{\mathbb{E}\left[\exp\left\{\lambda\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\right\}\right] : \lambda\leq 0\right\}.$$

Let $F(\lambda) = \mathbb{E}[\exp\{\lambda(\frac{W_1^m H_1}{\Gamma(m)} - y)\}]$. Then we have

$$F'(\lambda) = \mathbb{E}\left[\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\exp\left\{\lambda\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\right\}\right]$$

and

$$F''(\lambda) = \mathbb{E}\left[\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)^2 \exp\left\{\lambda\left(\frac{W_1^m H_1}{\Gamma(m)} - y\right)\right\}\right] > 0.$$

If $y < 1$, then $F(0) = 1, F'(0) = 1 - y > 0$. Thus there exists $\lambda < 0$ such that $F(\lambda) < 1$ which implies that $J(y) > 0$ for $y < 1$. This combined with (2.2) and the fact that $J(\cdot)$ is non-increasing implies that for any $x < 1$

$$\begin{aligned}
&\limsup_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) \le x\} \\
&= \liminf_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) < x\} = -\infty
\end{aligned} \tag{2.3}$$

and thus

$$\begin{aligned}
&\lim_{\delta\to 0} \limsup_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{|L_0(\mathbf{P}(\theta); m) - x| \le \delta\} \\
&= \lim_{\delta\to 0} \liminf_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{|L_0(\mathbf{P}(\theta); m) - x| < \delta\} \\
&= -\infty.
\end{aligned} \tag{2.4}$$

Now consider the case $x > 1$. Since

$$\mathbb{P}\{W_1^m H_1 > \Gamma(m)x\} \le \mathbb{P}\{W_1 > [\Gamma(m)x]^{1/m}\} = e^{-[\Gamma(m)x]^{1/m}},$$

it follows from Theorem 3 in [11] that

$$\limsup_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) \ge x\} \le -[\Gamma(m)(x-1)]^{1/m}. \tag{2.5}$$

On the other hand, for any $\epsilon > 0$ and $0 < \delta < 1$

$$\begin{aligned}
&\mathbb{P}\{L_0(\mathbf{P}(\theta); m) > x\} \\
&\ge \mathbb{P}\left\{\frac{1}{\theta}\left(\frac{W_1^m H_1}{\Gamma(m)} - 1\right) \ge x - 1 + \epsilon\right\} \mathbb{P}\left\{\frac{1}{\theta}\sum_{k=2}^{\theta}\left(\frac{W_k^m H_k}{\Gamma(m)} - 1\right) \ge -\epsilon\right\} \\
&\ge \mathbb{P}\{H_1 \ge \delta\}\mathbb{P}\{W_1^m > \delta^{-1}\Gamma(m)[1 + \theta(x - 1 + \epsilon)]\}\mathbb{P}\left\{\frac{1}{\theta}\sum_{k=2}^{\theta}\left(\frac{W_k^m H_k}{\Gamma(m)} - 1\right) \ge -\epsilon\right\} \\
&= \mathbb{P}\{H_1 \ge \delta\} \\
&\quad \times \mathbb{P}\left\{\frac{1}{\theta}\sum_{k=2}^{\theta}\left(\frac{W_k^m H_k}{\Gamma(m)} - 1\right) \ge -\epsilon\right\} \exp\left\{-\left(\delta^{-1}\Gamma(m)[1 + \theta(x - 1 + \epsilon)]\right)^{1/m}\right\}
\end{aligned}$$

Since $\{\frac{1}{\theta}\sum_{i=2}^{\theta}(\frac{W_k^m H_k}{\Gamma(m)} - 1) \ge -\epsilon\}$ converges to one as θ tends to infinity, it follows that

$$\liminf_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) > x\} \ge -[\delta^{-1}\Gamma(m)(x - 1 + \epsilon)]^{1/m}.$$

Letting ϵ go to zero followed by δ going to one, we obtain

$$\liminf_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) > x\} \ge -[\Gamma(m)(x-1)]^{1/m}$$

which combined with (2.5) implies that

$$\begin{aligned}
&\lim_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) \ge x\} \\
&\quad = \lim_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) > x\} \\
&\quad = -[\Gamma(m)(x-1)]^{1/m}.
\end{aligned} \tag{2.6}$$

Since $I(x)$ is strictly increasing in x for $x > 1$ and for any $\delta > 0$ such that $x - \delta > 1$

$$\mathbb{P}\{L_0(\mathbf{P}(\theta); m) \geq x - \delta\} = \mathbb{P}\{|L_0(\mathbf{P}(\theta); m) - x| \leq \delta\} + \mathbb{P}\{L_0(\mathbf{P}(\theta); m) \geq x + \delta\},$$

it follows from (2.6) that

$$\begin{aligned} &\lim_{\delta\to 0}\lim_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{|L_0(\mathbf{P}(\theta); m) - x| < \delta\} \\ &\quad = \lim_{\delta\to 0}\lim_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{|L_0(\mathbf{P}(\theta); m) - x| \leq \delta\} \\ &\quad = -[\Gamma(m)(x-1)]^{1/m}. \end{aligned} \tag{2.7}$$

Putting together (2.3) and (2.6) we obtain that for any $M > 0$ and $c = 1 + \frac{M^m}{\Gamma(m)}$

$$\limsup_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\{L_0(\mathbf{P}(\theta); m) \notin [-c, c]\} \leq -M, \tag{2.8}$$

which combined with (2.4), (2.7) and Theorem (P) in [13] implies the large deviation principle for $L_0(\mathbf{P}(\theta); m)$ with speed $\theta^{1/m}$ and good rate function $I(\cdot)$.

By direct calculation

$$\lim_{\theta\to\infty} \theta^{-1/m} \log \mathbb{P}\left\{\left|\left(\frac{\gamma(\theta)}{\theta}\right)^m - 1\right| > \delta\right\} = -\infty \tag{2.9}$$

By Lemma 2.1 in [5], the large deviation principle for $L(\mathbf{P}(\theta); m)$ is the same as $L_0(\mathbf{P}(\theta); m)$.

Finally for general $\theta \geq 1$, let $[\theta]$ denote the integer part of θ. By direct calculation we have that

$$\left(\frac{\gamma([\theta])}{\gamma(\theta)}\right)^m L(\mathbf{P}([\theta]); m) \leq L(\mathbf{P}(\theta); m) \leq \left(\frac{\gamma([\theta]+1)}{\gamma(\theta)}\right)^m L(\mathbf{P}([\theta]+1); m). \tag{2.10}$$

For any $0 < \delta < 1$,

$$\begin{aligned} \mathbb{P}\left\{\left|\left(\frac{\gamma([\theta])}{\gamma(\theta)}\right)^m - 1\right| > \delta\right\} &\leq \mathbb{P}\left\{\left|\frac{\gamma([\theta])}{\gamma(\theta)} - 1\right| > m^{-1}\delta\right\} \\ &= \mathbb{P}\left\{\frac{\gamma([\theta])}{\gamma(\theta)} < 1 - \frac{\delta}{m}\right\} \\ &= \frac{\Gamma(\theta)}{\Gamma([\theta])\Gamma(\theta-[\theta])} \int_0^{1-\frac{\delta}{m}} x^{[\theta]-1}(1-x)^{\theta-[\theta]-1} d\,x \\ &\leq \frac{\Gamma(\theta)}{\Gamma([\theta])\Gamma(\theta-[\theta])} \frac{m}{\delta} \frac{(1-\delta/m)^{[\theta]}}{[\theta]} \end{aligned}$$

where the second equality follows from the fact that $\gamma([\theta])/\gamma(\theta)$ follows a $Beta([\theta], \theta - [\theta])$ distribution. This implies that for any $0 < r < 1$

$$\lim_{\theta\to\infty} \frac{1}{\theta^r} \log \mathbb{P}\left\{\left|\left(\frac{\gamma([\theta])}{\gamma(\theta)}\right)^m - 1\right| > \delta\right\} = -\infty. \tag{2.11}$$

Similarly we can prove that

$$\lim_{\theta\to\infty} \frac{1}{\theta^r} \log \mathbb{P}\left\{\left|\left(\frac{\gamma([\theta]+1)}{\gamma(\theta)}\right)^m - 1\right| > \delta\right\} = -\infty. \tag{2.12}$$

Applying Lemma 2.1 in [5] again, we conclude that the large deviations at scales of θ^r of $\left(\frac{\gamma([\theta])}{\gamma(\theta)}\right)^m L(\mathbf{P}([\theta]); m)$ and $\left(\frac{\gamma([\theta]+1)}{\gamma(\theta)}\right)^m L(\mathbf{P}([\theta]+1); m)$ are the same as the corresponding large deviations of $L(\mathbf{P}([\theta]); m)$ and $L(\mathbf{P}([\theta]+1); m)$. □

Remark 2.2. Considering Kingman's coalescent and the time T_n when n ancestors are found. Large deviation estimates were obtained in [3] for the scaled probability of two randomly selected individuals at time zero having the same ancestor at time T_n. In our notation this probability has the form

$$\left(\frac{n}{\gamma(n)}\right)^2 \frac{1}{n}\sum_{k=1}^{n} W_k^2.$$

This is the same as $L_0(\mathbf{P}(n);2)$ except H_k is replaced by 1. Our result shows that the corresponding work in [3] can be generalized to any $m \geq 2$.

Remark 2.3. (Connections to the result in [1]). The large deviation principle for $H(\mathbf{P}(\theta);m)$ obtained in [1] has speed θ and rate function

$$S(y) = \begin{cases} -\log(1-y^{1/m}), & y \in [0,1] \\ \infty, & otherwise \end{cases}$$

Since $H(\mathbf{P}(\theta);m)$ and $H(\mathbf{P}(\theta);m) - \frac{\Gamma(m)}{\theta^{m-1}}$ are exponentially equivalent, the same large deviation principle holds for $H(\mathbf{P}(\theta);m) - \frac{\Gamma(m)}{\theta^{m-1}}$. Since

$$L(\mathbf{P}(\theta);m) = \frac{\theta^{m-1}}{\Gamma(m)}\left[H(\mathbf{P}(\theta);m) - \frac{\Gamma(m)}{\theta^{m-1}}\right] + 1,$$

one has that for $L(\mathbf{P}(\theta);m) = x \geq 1$ and $y = \frac{\Gamma(m)}{\theta^{m-1}}(x-1)$

$$\begin{aligned} \exp\{-\theta S(y)\} &= \exp\left\{-\theta^{1/m}\theta^{(m-1)/m}\log\frac{1}{1-(\frac{\Gamma(m)}{\theta^{m-1}}(x-1))^{1/m}}\right\} \\ &\approx \exp\{-\theta^{1/m} I(x)\}. \end{aligned}$$

3 Two-parameter generalization

For any $0 < \alpha < 1$ and $\theta > 0$, let $\rho(t)$ denote the subordinator with Lévy measure

$$\alpha C_\alpha x^{-(1+\alpha)} e^{-x} dx, \quad x > 0$$

where $C_\alpha > 0$. Set

$$\tau(\alpha,\theta) = \frac{\gamma(\theta/\alpha)}{C_\alpha \Gamma(1-\alpha)}, \ \sigma_{\alpha,\theta} = \rho(\tau(\alpha,\theta))$$

and let

$$\mathbf{P}(\alpha,\theta) = (P_1(\alpha,\theta), P_2(\alpha,\theta), \ldots)$$

denote the descending order statistics of the normalized jump sizes of $\rho(t)$ over the random interval $[0,\tau(\alpha,\theta)]$. By Proposition 21 in [12], the law of $\mathbf{P}(\alpha,\theta)$ is the two-parameter Poisson-Dirichlet distribution, $\sigma_{\alpha,\theta}$ is a gamma random variable with parameters $(\theta,1)$ and is independent of $\mathbf{P}(\alpha,\theta)$. The case $\alpha = 0, \theta > 0$ can be recovered by choosing C_α such that $\lim_{\alpha\to 0} \alpha C_\alpha = 1$.

The two-parameter homozygosity of order m is defined as

$$H(\mathbf{P}(\alpha,\theta);m) = \sum_{i=1}^{\infty} P_i^m(\alpha,\theta).$$

Set

$$L(\mathbf{P}(\alpha,\theta);m) = \frac{\theta^{m-1}\Gamma(1-\alpha)}{\Gamma(m-\alpha)} H(\mathbf{P}(\alpha,\theta);m).$$

It is known ([8], [4]) that

$$\sqrt{\theta}[L(\mathbf{P}(\alpha,\theta);m)-1] \Rightarrow Z_m^\alpha, \quad m\to\infty \tag{3.1}$$

where Z_m^α is a normal random variable with mean zero and variance

$$\frac{\Gamma(1-\alpha)\Gamma(2m-\alpha)}{\Gamma^2(m-\alpha)}+\alpha-m^2.$$

As in the one-parameter case, the moderate deviation principles hold for the two-parameter Poisson-Dirichlet distribution ([6]), i.e., for any l in $(\frac{m-1}{2m-1},\frac{1}{2})$ large deviation principles hold for

$$\theta^l[L(\mathbf{P}(\alpha,\theta);m)-1].$$

Our next result establishes the large deviation principle for $L(\mathbf{P}(\alpha,\theta);m)$.

Theorem 3.1. *A large deviation principle holds for $L(\mathbf{P}(\alpha,\theta);m)$ as θ converges to infinity on space $\mathbb{R}$ with rate $\theta^{1/m}$ and good rate function*

$$I_\alpha(x)=\begin{cases}\left[\frac{\Gamma(m-\alpha)(x-1)}{\Gamma(1-\alpha)}\right]^{1/m}, & x\geq 1,\\ +\infty, & otherwise.\end{cases}$$

Proof: It suffices to consider the case that θ is an integer. Let $\xi_1,\ldots,\xi_\theta$ be i.i.d. copies of $(C_\alpha\Gamma(1-\alpha))\gamma(1/\alpha)$ and set

$$W_k^\alpha=\rho(\sum_{i=1}^k\xi_i)-\rho(\sum_{i=1}^{k-1}\xi_i), k=1,\ldots,\theta.$$

It is clear from the definition that $W_1^\alpha,\ldots,W_\theta^\alpha$ are independent. For any $\lambda>0$, the Laplace transform of W_k^α has the form

$$\begin{aligned}\mathbb{E}[e^{-\lambda W_k^\alpha}] &= \mathbb{E}[e^{-\lambda W_1^\alpha}]\\ &= \mathbb{E}[\exp\{-\gamma(1/\alpha)[(\lambda+1)^\alpha-1]\}]\\ &= (1+\lambda)^{-1}.\end{aligned}$$

Hence $W_1^\alpha,\ldots,W_\theta^\alpha$ are i.i.d. exponential with parameter one. In other words, $W_1^\alpha,\ldots,W_\theta^\alpha$ and $W_1,\ldots,W_\theta$ in the previous section are equal in distribution.

For each $1\leq k\leq\theta$, let $\{J_j^k(\alpha,\theta): j\geq 1\}$ denote all the jump sizes of $\rho(\cdot)$ in the interval $[\sum_{i=1}^{k-1}\xi_i,\sum_{i=1}^k\xi_i]$ and set

$$H_{\alpha,k}=\sum_j\left(\frac{J_j^k(\alpha,\theta)}{W_k^\alpha}\right)^m.$$

By Proposition 21 in [12], $H_{\alpha,k}$ is independent of W_k^α. It is not difficult to see that $H_{\alpha,1},\ldots,H_{\alpha,\theta}$ are i.i.d. with the same distribution as $H(\mathbf{P}(\alpha,1);m)$.

The two-parameter homozygosity can now be written as

$$H(\mathbf{P}(\alpha,\theta);m)=\left(\frac{1}{\sigma_{\alpha,\theta}}\right)^m\sum_{k=1}^\theta(W_k^\alpha)^m H_{\alpha,k}. \tag{3.2}$$

which has the same structure as (2.1) with $H_{\alpha,k}$ in place of H_k. Set

$$L_\alpha(\mathbf{P}(\alpha,\theta);m)=\frac{\Gamma(1-\alpha)}{\Gamma(m-\alpha)}\frac{1}{\theta}\sum_{k=1}^\theta(W_k^\alpha)^m H_{\alpha,k}$$

and write

$$L(\mathbf{P}(\alpha,\theta);m) = \left(\frac{\theta}{\sigma_{\alpha,\theta}}\right)^m L_\alpha(\mathbf{P}(\alpha,\theta);m),$$

the conclusion now follows from similar arguments used in the proof of Theorem 2.1. □

Remark 3.2. The subordinator representation for the two-parameter Poisson-Dirichlet distribution is a special case of subordination of a subordinator. The representations (2.1) and (3.2) can be generalized to these models. But the independency between the total jump size and the normalized individual jump sizes may no longer hold. It is not clear whether our result can be generalized to these situations.

Remark 3.3. For $0 < \alpha < 1, x > 1$, we have $I_\alpha(x) < I(x)$. Thus $L(\mathbf{P}(\alpha,\theta);m)$ is more spread out from 1 than $L(\mathbf{P}(\theta);m)$ and α can then be used to describe the diversity of the population in terms of large deviations.

Acknowledgments. We wish to thank the referee for the careful review of the paper and many helpful suggestions.

References

[1] D.A. Dawson and S. Feng (2006). Asymptotic behavior of Poisson-Dirichlet distribution for large mutation rate. *Ann. Appl. Probab.* Vol. 16, No.2, 562–582. MR-2244425

[2] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Second edition. Applications of Mathematics, **38**. Springer-Verlag, New York, 1998. MR-1619036

[3] A. Depperschmidt, P. Pfaffelhuber, and A. Scheuringer (2015). Some large deviations in Kingman's coalescent. *Electron. Commun. Probab.* **20**, 1–14. MR-3304413

[4] S. Feng. *The Poisson-Dirichlet Distribution and Related Topics*. Probability and its Applications (New York). Springer, Heidelberg, 2010. MR-2663265

[5] S. Feng and F.Q. Gao (2008). Moderate deviations for Poisson–Dirichlet distribution. *Ann. Appl. Probab.* **18**, No. 5, 1794–1824. MR-2462549

[6] S. Feng and F.Q. Gao (2010). Asymptotic results for the two-parameter Poisson–Dirichlet distribution. *Stoch. Proc. Appl.* **120**, 1159–1177. MR-2639742

[7] R.C. Griffiths (1979). On the distribution of allele frequencies in a diffusion model. *Theor. Pop. Biol.* **15**, 140–158. MR-0528914

[8] K. Handa (2009). The two-parameter Poisson-Dirichlet point process. *Bernoulli* **15**, 1082–1116. MR-2597584

[9] P. Joyce, S.M. Krone, and T.G. Kurtz (2002). Gaussian limits associated with the Poisson–Dirichlet distribution and the Ewens sampling formula. *Ann. Appl. Probab.* **12**, No. 1, 101–124. MR-1951997

[10] J.C.F. Kingman (1975). Random discrete distributions. *J. Roy. Statist. Soc. B,* **37**, 1–22. MR-0368264

[11] A.V. Nagaev (1969). Integral limit theorems taking large deviations into account when Cramér's condition does not hold, I. *Theory of Probability and Its Applications*,14(1):51–64. MR-0247651

[12] J. Pitman and M. Yor (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* Vol. 25, No. 2, 855–900. MR-1434129

[13] A.A. Puhalskii. On functional principle of large deviations. In V. Sazonov and T. Shervashidze, editors. *New Trends in Probability and Statistics.* pages 198–218. VSP Moks'las, Moskva, 1991. MR-1200917