Adv. Appl. Prob. **49,** 581–602 (2017)
*doi:*10.1017/apr.2017.13

LIMIT THEOREMS ASSOCIATED WITH THE PITMAN–YOR PROCESS

SHUI FENG,* *McMaster University*
FUQING GAO,** *Wuhan University*
YOUZHOU ZHOU,*** *Zhongnan University of Economics and Law*

Abstract

The Pitman–Yor process is a random discrete measure. The random weights or masses follow the two-parameter Poisson–Dirichlet distribution with parameters $0 < \alpha < 1$, $\theta > -\alpha$. The parameters α and θ correspond to the stable and gamma components, respectively. The distribution of atoms is given by a probability ν. In this paper we consider the limit theorems for the Pitman–Yor process and the two-parameter Poisson–Dirichlet distribution. These include the law of large numbers, fluctuations, and moderate or large deviation principles. The limiting procedures involve either α tending to 0 or 1. They arise naturally in genetics and physics such as the asymptotic coalescence time for explosive branching process and the approximation to the generalized random energy model for disordered systems.

Keywords: Pitman–Yor process; explosive branching; large deviation; phase transition; Poisson–Dirichlet distribution; random energy model

2010 Mathematics Subject Classification: Primary 60F10
Secondary 92D10

1. Introduction

For any $0 \leq \alpha < 1, \theta + \alpha > 0$, let $U_1(\alpha, \theta), U_2(\alpha, \theta), \dots$ be a sequence of independent random variables with $U_i(\alpha, \theta)$ having distribution $\text{beta}(1-\alpha, \theta + i\alpha)$ for $i \geq 1$. If we define

$$\begin{aligned} V_1(\alpha, \theta) &= U_1(\alpha, \theta), \\ V_n(\alpha, \theta) &= (1 - U_1(\alpha, \theta)) \cdots (1 - U_{n-1}(\alpha, \theta)) U_n(\alpha, \theta), \qquad n \geq 2, \end{aligned}$$

then the law of the decreasing order statistic

$$\mathbb{P}(\alpha, \theta) = (P_1(\alpha, \theta), P_2(\alpha, \theta), \dots)$$

of $(V_1(\alpha, \theta), V_2(\alpha, \theta), \dots)$ is the two-parameter Poisson–Dirichlet distribution $\text{PD}(\alpha, \theta)$. It is a probability on the infinite-dimensional simplex

$$\nabla_\infty = \left\{ \boldsymbol{p} = (p_1, p_2, \dots) \colon p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} p_i \leq 1 \right\}.$$

Received 3 March 2016; revision received 14 December 2016.
* Postal address: Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, L8S 4K1, Canada. Email address: shuifeng@mcmaster.ca
** Postal address: School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China.
Email address: fqgao@whu.edu.cn
*** Postal address: School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, 430073, China. Email address: youzhouzhou2014@zuel.edu.cn

Let S be a Polish space and ν a probability on S satisfying $\nu(\{x\}) = 0$ for all x in S. In this case we say that ν is diffuse. The Pitman–Yor process with parameters α, θ, and ν is the random measure

$$\Xi_{\alpha,\theta,\nu} = \sum_{i=1}^{\infty} P_i(\alpha, \theta)\delta_{\xi_i},$$

where $\xi_1, \xi_2, \ldots$ are independent and identically distributed (i.i.d.) with common distribution ν and independent of $\mathbb{P}(\alpha, \theta)$. The $\alpha = 0$ case corresponds to the Dirichlet process constructed in [20].

The distribution $\mathrm{PD}(0, \theta)$ was introduced by Kingman in [27] as the law of relative jump sizes of a gamma subordinator over the interval $[0, \theta]$. It also arises in other contexts most notably population genetics. The distribution $\mathrm{PD}(\alpha, 0)$ was also introduced in Kingman [27] through the stable subordinator. In [29] and [33], $\mathrm{PD}(\alpha, 0)$ was constructed from the ranked length of excursion intervals between the zeros of a Brownian motion ($\alpha = \frac{1}{2}$) or a recurrent Bessel process of order $2(1 - \alpha)$ for general α.

In this paper we focus on the $\theta = 0$ case. This implies that the parameter α is in $(0, 1)$. Without loss of generality, we choose the space S to be $[0, 1]$ and the probability ν to be the uniform distribution on $[0, 1]$. Our main objective is to study the asymptotic behaviour of $\mathrm{PD}(\alpha, 0)$ when α converges to 0, and the behaviour of both $\mathrm{PD}(\alpha, 0)$ and $\Xi_{\alpha,0,\nu}$ when α converges to 1. There are many scenarios where the limiting procedure of α approaching 1 or 0 arises naturally. We consider two examples below.

The first example is Derrida's random energy model (REM) introduced in [9] and [10]. This is a toy model for a disordered system such as spin glasses. For any $N \geq 1$, let $S_N = \{-1, 1\}^N$ denote the configuration space. Then the REM is a family of i.i.d. random variables $\{H_N(\sigma) : \sigma \in S_N\}$ with common normal distribution of mean zero and variance N. Here $H_N(\sigma)$ is the Hamiltonian. Given the temperature T and $\beta = T^{-1}$, the Gibbs measure is a probability on S_N given by

$$Z_N^{-1} \exp\{-\beta H_N(\sigma)\},$$

where

$$Z_N = \sum_{\sigma \in S_N} \exp\{-\beta H_N(\sigma)\}$$

is the partition function. Let the critical temperature $T_\mathrm{c} = 1/\sqrt{2 \ln 2}$ and $\alpha = T/T_\mathrm{c}$. Then, for $T < T_\mathrm{c}$, or, equivalently, $\beta > \sqrt{2 \ln 2}$, the decreasing order statistic of the Gibbs measure is known (see [37]) to converge to the Poisson–Dirichlet distribution $\mathrm{PD}(\alpha, 0)$ as N tends to ∞. Thus, α converging to 0 corresponds to temperature going to 0, while α converging to 1 corresponds to temperature rising to the critical value. To account for correlations, the generalized random energy model (GREM) involving hierarchical levels was introduced and studied in [11] and [12]. The generalization to continuum levels was carried out in [4] and the genealogy of the hierarchical systems was described by the Bolthausen–Sznitman coalescent. In deriving the infinitesimal rate of the coalescent (Proposition 4.11 in [3]), one needs to consider the limit of $\mathrm{PD}(\mathrm{e}^{-t}, 0)$ as t converges to 0 or, equivalently, $\alpha = \mathrm{e}^{-t}$ converging to 1.

The second example is concerned with the coalescence time for an explosive branching process. Consider a Galton–Watson branching process with offspring distribution $\{p_j : j \geq 0\}$ in the domain of attraction of a stable law of order $0 < \gamma < 1$ and $p_0 = 0$. Let X_n denote the coalescence time of any two individuals chosen at random at generation n. Then it was shown in [1] that

$$\mathbb{P}\{n - X_n \leq k\} \to \pi(k) \quad \text{as } n \to \infty,$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

where $\pi(k)$ can be calculated explicitly through $\mathrm{PD}(\gamma^k, 0)$. In this case, $\alpha = \gamma^k$ converging to 0 corresponds to k converging to ∞.

There have been intensive studies of the asymptotic behaviour for the Poisson–Dirichlet distribution and the Pitman–Yor process in recent years with motivations from probability theory, population genetics, and Bayesian statistics. The limiting procedures include the following three cases.

Case A. The parameter θ converges to ∞.

Case B. Both θ and α converge to 0.

Case C. The parameter α converges to 1.

Results in case A include central limit theorems in [22]–[25], large deviations in [6], [7], [14], and [28], and moderate deviations in [17]. It also includes the work in [13] and [21] where the large and moderate deviations for $\Xi_{0,\theta,\nu}$ were studied in a Bayesian context. In case B, the large and moderate deviations were obtained in [15] and [18]. Recently, the large deviations for $\mathrm{PD}(\alpha, 0)$ in case C was obtained in [19]. An extended account of asymptotic results for the Poisson–Dirichlet distribution and the Pitman–Yor process can be found in [16] and the references therein.

The main results of this paper include

(i) the fluctuation and large deviations for $\mathrm{PD}(\alpha, 0)$ in case B when each coordinate is raised to a certain power (Theorem 2 and Theorem 4);

(ii) the fluctuation theorem for $\mathrm{PD}(\alpha, 0)$ in case C (Theorem 3);

(iii) the large deviations for $\Xi_{\alpha,0,\nu}$ in case C (Theorem 8).

The existing asymptotic results for $\mathrm{PD}(\alpha, \theta)$ in the literature in case B involve only power 1. The scalings appear either as a shift or a multiplier. In (i) we studied the asymptotic behaviour of $\mathrm{PD}(\alpha, 0)$ with various powers and discovered an interesting phase transition or critical behaviour in terms of the magnitude of the power. The scaling property of the stable subordinator plays the essential role here. The result in (ii) complements the result in [18] and reveals a non-Gaussian fluctuation.

The large deviations in (iii) provide more information on the microscopic transition structure at the critical temperature for the REM. At the instant when the temperature starts to move below the critical value T_c, a portion of mass of the uniform measure ν may be lost and is replaced by an atomic portion with finite atoms. This represents the emerging of a finite number of energy valleys and the energy landscape of the system becomes a mixture of valleys and 'flat' regions. The emerging of energy valleys follows the order where the small number of energy valleys is more likely to occur than a large number of valleys. Our result will validate this picture and show that large deviations may lead to the disappearance of all 'flat' regions. In comparison with the large deviation result in [19], the result in (iii) reveals the new information that no singular components arise through large deviations.

The behaviour of $\Xi_{\alpha,0,\nu}$ in case C resembles the behaviour of $\Xi_{\alpha,\theta,\nu}$ in case A in terms of the law of large numbers. But our result in (iii) will show that the two cases are very different in terms of large deviations.

The paper is organized as follows. In Section 2 we introduce the necessary terminologies and review the subordinator representation for $\mathrm{PD}(\alpha, 0)$. Section 3 contains the law of large numbers, fluctuation, and large deviations associated with $\mathrm{PD}(\alpha, 0)$ as α converges to 0 or 1.

In Section 4 we establish the large deviation principle for $\Xi_{\alpha,0,\nu}$ under the limit of α converging to 1. Large deviations for PD(α, 0) are closely related to the large deviations for $\Xi_{\alpha,0,\nu}$. A result for one usually gives some hint for the other. But a direct derivation from one to the other is difficult due to the topologies involved and the lack of a continuous map. Thus, our proof here will be completely new in comparison to the result in [19]. We close the paper in Section 5 with some concluding remarks on the differences between case A and case C.

2. Preliminaries

In this section we introduce the necessary terminologies of large deviations and the subordinator representation of PD(α, 0). All results are stated in the form that is sufficient for our purposes. We refer the reader to Dembo and Zeitouni [8] for a comprehensive introduction to the general theory of large deviations.

2.1. Terminologies

Let E be a complete separable metric space equipped with metric d. Consider a family of E-valued random variables $\{Y_\lambda : \lambda > 0\}$ with corresponding distributions $\{Q_\lambda : \lambda > 0\}$. Assume that Y_λ converges in probability to a constant as λ tends to a number λ_0 in $[0, +\infty]$.

Definition 1. (i) The family of probability measures $\{Q_\lambda : \lambda > 0\}$ (or the family $\{Y_\lambda : \lambda > 0\}$) is said to satisfy a *large deviation principle* with *speed* $a(\lambda)$ and *rate function* $I(\cdot)$ if, for any closed set F and open set G in E,

(i) (upper bound)

$$\limsup_{\lambda\to\lambda_0} a^{-1}(\lambda)\log Q_\lambda\{F\} \le -\inf_{x\in F} I(x),$$

(ii) (lower bound)

$$\liminf_{\lambda\to\lambda_0} a^{-1}(\lambda)\log Q_\lambda\{G\} \ge -\inf_{x\in G} I(x),$$

and $I(\cdot)$ is lower semicontinuous. The rate function $I(\cdot)$ is *good* if, for any $c > 0$,

$$\text{the level set } \{x : I(x) \le c\} \text{ is compact.}$$

The set $\{x : I(x) < \infty\}$ is called the *effective domain* of the rate function.

(ii) The family $\{Q_\lambda : \lambda > 0\}$ is said to satisfy a *local large deviation principle* with speed $a(\lambda)$ and rate function $I(\cdot)$ if, for every x in E,

$$\lim_{\delta\to 0}\limsup_{\lambda\to\lambda_0} a^{-1}(\lambda)\log P\{d(Y_\lambda, x) \le \delta\} = \lim_{\delta\to 0}\liminf_{\lambda\to\lambda_0} a^{-1}(\lambda)\log P\{d(Y_\lambda, x) < \delta\} = -I(x).$$

(iii) The family $\{Q_\lambda : \lambda > 0\}$ is *exponentially tight* with speed $a(\lambda)$ if, for every $L > 0$, there is a compact subset K_L of E such that

$$\limsup_{\lambda\to\lambda_0} a^{-1}(\lambda)\log P\{Y_\lambda \notin K_L\} \le -L.$$

Remark 1. It is known that a local large deviation principle combined with exponential tightness implies the large deviation principle with a good rate function (see [36]).

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Definition 2. (i) The family $\{Y_\lambda : \lambda > 0\}$ is said to satisfy a *fluctuation theorem* if there exist functions $b(\lambda)$ and $c(\lambda)$, and a finite nondeterministic random variable W such that

$$\lim_{\lambda\to\lambda_0} b(\lambda) = \infty, \qquad b(\lambda)[Y_\lambda - c(\lambda)] \xrightarrow{\mathrm{D}} W, \qquad \lambda \to \lambda_0,$$

where '$\xrightarrow{\mathrm{D}}$' denotes convergence in distribution.

(ii) Assume that the family $\{Y_\lambda : \lambda > 0\}$ satisfies the fluctuation theorem above. Let $e(\lambda)$ satisfy

$$\lim_{\lambda\to\lambda_0} e(\lambda) = \infty, \qquad \lim_{\lambda\to\lambda_0} \frac{e(\lambda)}{b(\lambda)} = 0.$$

Then the family $\{Q_\lambda : \lambda > 0\}$ or, equivalently, the family $\{Y_\lambda : \lambda > 0\}$ is said to satisfy a *moderate deviation principle* with speed $a(\lambda)$ (depending on $e(\lambda)$) and (good) rate function $I(\cdot)$ if the family $\{e(\lambda)[Y_\lambda - c(\lambda)] : \lambda > 0\}$ satisfies a large deviation principle with speed $a(\lambda)$ and (good) rate function $I(\cdot)$. Thus, the moderate deviation principle for $\{Y_\lambda : \lambda > 0\}$ is the large deviation principle for $\{e(\lambda)[Y_\lambda - c(\lambda)] : \lambda > 0\}$.

2.2. Subordinator representation

For any $0 < \alpha < 1$, let ρ_t be the stable subordinator with index α and Lévy measure

$$\Lambda_\alpha(\mathrm{d}x) = \frac{\alpha}{\Gamma(1-\alpha)} x^{-(1+\alpha)}\,\mathrm{d}x, \qquad x > 0.$$

The boundary case $\alpha = 1$ corresponds to the straight line $\rho_t = t$. When α converges to 0, ρ_t becomes a killed subordinator with killing rate 1 (see [2]).

For any $t > 0$, let $J^1(\rho_t) \geq J^2(\rho_t) \geq \cdots$ denote the jump sizes of ρ_t over the interval $[0, t]$. Then the following representation holds.

Theorem 1. (Perman *et al.* [29].) *For any $t > 0$, the law of*

$$\left(\frac{J^1(\rho_t)}{\rho_t}, \frac{J^2(\rho_t)}{\rho_t}, \ldots\right) \quad \textit{is } \mathrm{PD}(\alpha, 0).$$

For any $n \geq 1$, let $Z_n = \Lambda_\alpha(J^n(\rho_1), \infty)$. Then $Z_1 < Z_2 < \cdots$ and $Z_1, Z_2 - Z_1, Z_3 - Z_2, \ldots$ are i.i.d. exponential random variables with parameter 1 (see [29]). Noting that

$$\Lambda_\alpha(x, \infty) = \frac{x^{-\alpha}}{\Gamma(1-\alpha)},$$

it follows that

$$\frac{J^n(\rho_1)}{\rho_1} = \frac{Z_n^{-1/\alpha}}{\sum_{i=1}^\infty Z_i^{-1/\alpha}}$$

and

$$\rho_1 = \Gamma(1-\alpha)^{-1/\alpha} \sum_{i=1}^\infty Z_i^{-1/\alpha}. \tag{1}$$

Thus, by Theorem 1, the law of

$$\left(\frac{Z_1^{-1/\alpha}}{\sum_{i=1}^\infty Z_i^{-1/\alpha}}, \frac{Z_2^{-1/\alpha}}{\sum_{i=1}^\infty Z_i^{-1/\alpha}}, \ldots\right) \quad \text{is } \mathrm{PD}(\alpha, 0).$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

In other words, we have the representation

$$P_n(\alpha, 0) = \frac{Z_n^{-1/\alpha}}{\sum_{i=1}^{\infty} Z_i^{-1/\alpha}}, \qquad n = 1, 2, \dots.$$

3. Limit theorems for PD(α, 0)

Take a sample of size two from a population with frequencies following the distribution PD(α, 0). Then the probability that the two samples are of the same type is given by

$$\varphi_2(\mathbb{P}(\alpha, 0)) := \sum_{i=1}^{\infty} P_i^2(\alpha, 0).$$

This function is called the homozygosity in population genetics. It is closely associated with the Shannon entropy in communication, the Herfindahl–Hirschmam index in economics, and the Gini–Simpson index in ecology. It is a measure of concentration of the population in terms of types with small values corresponding to lower concentration.

A direct application of Pitman's sampling formula (see [30] and [32]) leads to

$$\mathbb{E}_{\alpha,0}[\varphi_2(\mathbb{P}(\alpha, 0))] = 1 - \alpha.$$

This implies that $\mathbb{P}(\alpha, 0)$ converges in probability to $(1, 0, \dots)$ and $(0, 0, \dots)$ as α converges to 0 and 1, respectively. The objective of this section is to obtain more detailed information associated with these limits including fluctuation and large deviations.

3.1. Convergence and limit

Let $0 < \gamma(\alpha) \le 1$ and $\iota(\alpha) > 0$ be such that

$$\lim_{\alpha \to 0} \frac{\gamma(\alpha)}{\alpha} = c_1 \in [0, +\infty] \tag{2}$$

and

$$\lim_{\alpha \to 1} \frac{\iota(\alpha)}{\Gamma(1-\alpha)} = c_2 \in [0, \infty).$$

Theorem 2. *Let*

$$\mathbb{P}^{\gamma(\alpha)}(\alpha, 0) = (P_1^{\gamma(\alpha)}(\alpha, 0), P_2^{\gamma(\alpha)}(\alpha, 0), \dots).$$

If c_1 is finite then $\mathbb{P}^{\gamma(\alpha)}(\alpha, 0)$ converges almost surely to $(1, (Z_1/Z_2)^{c_1}, (Z_1/Z_3)^{c_1}, \dots)$ as α converges to 0. If $c_1 = \infty$ then $\mathbb{P}^{\gamma(\alpha)}(\alpha, 0)$ converges to $(1, 0, \dots)$ in probability as α converges to 0.

Proof. Set $\tilde{\boldsymbol{Z}} = (Z_1^{-1}, Z_2^{-1}, \dots)$. Then we have

$$\left(\sum_{i=1}^{\infty} Z_i^{-1/\alpha}\right)^{\alpha} = \|\tilde{\boldsymbol{Z}}\|_{1/\alpha}.$$

When α approaches to 0, $\|\tilde{\boldsymbol{Z}}\|_{1/\alpha}$ converges almost surely to $\|\tilde{\boldsymbol{Z}}\|_{\infty} = Z_1^{-1}$. This implies that

$$\mathbb{P}^{\alpha}(\alpha, 0) = \left(\frac{Z_1^{-1}}{\|\tilde{\boldsymbol{Z}}\|_{1/\alpha}}, \frac{Z_2^{-1}}{\|\tilde{\boldsymbol{Z}}\|_{1/\alpha}}, \dots\right)$$

converges almost surely to $(1, Z_1/Z_2, Z_1/Z_3, \dots)$ as α converges to 0. Write $\mathbb{P}^{\gamma(\alpha)}(\alpha, 0)$ as

$$\left(\left(\frac{Z_1^{-1}}{\|\tilde{\boldsymbol{Z}}\|_{1/\alpha}}\right)^{\gamma(\alpha)/\alpha}, \left(\frac{Z_2^{-1}}{\|\tilde{\boldsymbol{Z}}\|_{1/\alpha}}\right)^{\gamma(\alpha)/\alpha}, \dots\right).$$

Then, by continuity, it follows that $\mathbb{P}^{\gamma(\alpha)}(\alpha, 0)$ converges almost surely to

$$\left(1, \left(\frac{Z_1}{Z_2}\right)^{c_1}, \left(\frac{Z_1}{Z_3}\right)^{c_1}, \dots\right)$$

as α converges to 0. If $c_1 = \infty$ then, for any $M \geq 1$, we have $\gamma(\alpha)/\alpha > M$ for small enough α. Thus, for any $n > 1$,

$$\lim_{\alpha\to 0} P_n^{\gamma(\alpha)}(\alpha, 0) \leq \lim_{\alpha\to 0} P_n^M(\alpha, 0) = \left(\frac{Z_1}{Z_n}\right)^M.$$

Since M is arbitrary, we obtain

$$\lim_{\alpha\to 0} P_n^{\gamma(\alpha)}(\alpha, 0) = 0 \quad \text{almost surely for } n > 1.$$

Finally, for $n = 1$, we have

$$P_1(\alpha, 0) \leq P_1^{\gamma(\alpha)}(\alpha, 0) \leq 1.$$

Noting that

$$\mathbb{E}[P_1(\alpha, 0)] \leq \mathbb{E}[\varphi_2(\mathbb{P}(\alpha, 0))] = 1 - \alpha,$$

it follows that $P_1(\alpha, 0)$ converges to 1 in probability which implies that $P_1^{\gamma(\alpha)}(\alpha, 0)$ converges to 1 in probability. □

Theorem 3. *As α converges to 1, $\iota(\alpha)\mathbb{P}(\alpha, 0)$ converges to $c_2(Z_1^{-1}, Z_2^{-1}, \dots)$ in probability.*

Proof. Let $S_\alpha = \rho_1^{-\alpha}$. Then the law of S_α is the Mittag–Leffler distribution with density function

$$g_\alpha(s) = \sum_{k=0}^{\infty} \frac{(-s)^k}{k!} \Gamma(\alpha k + \alpha + 1) \frac{\sin(\alpha k \pi)}{\alpha k \pi},$$

and

$$\sum_{i=1}^{\infty} Z_i^{-1/\alpha} = \left(\frac{S_\alpha}{\Gamma(1-\alpha)}\right)^{-1/\alpha}, \qquad \mathbb{E}[S_\alpha^r] = \frac{\Gamma(r+1)}{\Gamma(\alpha r + 1)}, \qquad r > -1.$$

This implies that

$$\mathbb{E}[(S_\alpha - 1)^2] = \frac{2}{\Gamma(2\alpha + 1)} - \frac{2}{\Gamma(\alpha + 1)} + 1 \to 0, \qquad \alpha \to 1.$$

Hence, S_α converges to 1 in probability as α converges to 1.

By (1), we have

$$\begin{aligned}\iota(\alpha)\mathbb{P}(\alpha, 0) &= \frac{\iota(\alpha)}{\Gamma(1-\alpha)} \Gamma(1-\alpha)^{1-1/\alpha} \left(\left(\frac{Z_1}{S_\alpha}\right)^{-1/\alpha}, \left(\frac{Z_2}{S_\alpha}\right)^{-1/\alpha}, \dots\right) \\ &= \frac{\iota(\alpha)}{\Gamma(1-\alpha)} \Gamma(1-\alpha)^{1-1/\alpha} \exp\left\{\frac{1}{\alpha} \log S_\alpha\right\} (Z_1^{-1/\alpha}, Z_2^{-1/\alpha}, \dots).\end{aligned}$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Since S_α converges to 1 in probability and $(Z_1^{-1/\alpha}, Z_2^{-1/\alpha}, \dots)$ converges to

$$(Z_1^{-1}, Z_2^{-1}, \dots)$$

almost surely as α converges to 1, we conclude that $\iota(\alpha)\mathbb{P}(\alpha, 0)$ converges in probability to

$$c_2(Z_1^{-1}, Z_2^{-1}, \dots). \qquad \square$$

Remark 2. The random variable S_α is Pitman's α-diversity (see [32]). Consider the random partition Π_∞ of the set $\{1, 2, \dots\}$ with asymptotic frequencies following the PD$(\alpha, 0)$ distribution. For any $n \geq 2$, let Π_n denote the restriction of Π_∞ on $\{1, 2, \dots, n\}$ and K_n denote the number of blocks in the random partition Π_n. The random variable S_α is the asymptotic limit of K_n/n^α as n tends to ∞ (see Theorem 3.8 in [32]). When α converges to 1, S_α converges to 1. Thus, all blocks will become singletons.

3.2. Large deviations

In this section we consider the large deviations associated with the deterministic limits obtained in Theorem 2. In comparison with the large deviations for $\mathbb{P}(\alpha, 0)$, these results may be viewed as moderate deviations for $\log \mathbb{P}(\alpha, 0) = (\log P_1(\alpha, 0), \dots)$. We prove these results through a series of lemmas.

For any $n \geq 1$, let

$$R_n = \frac{P_{n+1}(\alpha, 0)}{P_n(\alpha, 0)}.$$

Then $\{R_n : n \geq 1\}$ is a sequence of independent beta random variables with each R_n having the beta$(n\alpha, 1)$ distribution (Proposition 8 in [34]).

Lemma 1. *Let* $\boldsymbol{R}^{\gamma(\alpha)} = (R_1^{\gamma(\alpha)}, R_2^{\gamma(\alpha)}, \dots)$. *As* α *converges to* 0, *large deviation principles hold for* $\boldsymbol{R}^{\gamma(\alpha)}$ *on the space* $[0, 1]^\infty$ *with respective speeds and rate functions* $(\alpha/\gamma(\alpha), \tilde{I}_1(\cdot))$ *and* $(\log \gamma(\alpha)/\alpha, \tilde{I}_2(\cdot))$ *depending on whether* $c_1 = 0$ *or* $c_1 = \infty$, *where*

$$\tilde{I}_1(\boldsymbol{x}) = \begin{cases} \displaystyle\sum_{n=1}^{\infty} n \log \frac{1}{x_n}, & x_n > 0 \textit{ for all } n > 1, \\ +\infty & \textit{otherwise,} \end{cases}$$

and

$$\tilde{I}_2(\boldsymbol{x}) = \#\{n \geq 1 : x_n > 0\}.$$

Proof. First note that, for any $0 \leq a < b \leq 1$, it follows that, for any $n \geq 1$,

$$\mathbb{P}\{R_n^{\gamma(\alpha)} \in (a, b)\} = \mathbb{P}\{R_n^{\gamma(\alpha)} \in [a, b]\} = b^{n(\alpha/\gamma(\alpha))} - a^{n(\alpha/\gamma(\alpha))}. \tag{3}$$

Assume that $c_1 = 0$. Then we have $\lim_{\alpha \to 0} \alpha/\gamma(\alpha) = \infty$. For any $n \geq 1$ and any x in $[0, 1]$, we have, by applying (3),

$$\begin{aligned} \lim_{\delta \to 0} \liminf_{\alpha \to 0} \frac{\gamma(\alpha)}{\alpha} \log \mathbb{P}\{|R_n^{\gamma(\alpha)} - x| < \delta\} &= \lim_{\delta \to 0} \limsup_{\alpha \to 0} \frac{\gamma(\alpha)}{\alpha} \log \mathbb{P}\{|R_n^{\gamma(\alpha)} - x| \leq \delta\} \\ &= n \log x, \end{aligned}$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

where $\log x = -\infty$ for $x = 0$. This combined with the compactness of $[0, 1]$ implies that $R_n^{\gamma(\alpha)}$ satisfies a large deviation principle on $[0, 1]$ with speed $\alpha/\gamma(\alpha)$ and rate function $n \log x$. Similarly, for $c_1 = \infty$, we have

$$\lim_{\delta\to 0}\liminf_{\alpha\to 0}\bigg(\log\frac{\gamma(\alpha)}{\alpha}\bigg)^{-1}\log\mathbb{P}\{|R_n^{\gamma(\alpha)} - x| < \delta\}$$
$$= \lim_{\delta\to 0}\limsup_{\alpha\to 0}\bigg(\log\frac{\gamma(\alpha)}{\alpha}\bigg)^{-1}\log\mathbb{P}\{|R_n^{\gamma(\alpha)} - x| \le \delta\}$$
$$= -\mathbf{1}_{(0,\infty)},$$

where $\mathbf{1}_{(0,\infty)}$ is the indicator function of set $(0, \infty)$. These combined with the independence of $R_1, R_2, \dots$ imply the large deviations for $\boldsymbol{R}^{\gamma(\alpha)}$. □

Lemma 2. *There exists $\delta \ge 1$ such that, for any $\lambda < \delta$,*

$$\mathbb{E}[\exp\{\lambda(1-\alpha)(P_1^{-1}(\alpha, 0) - 1)\}] = (1 + A_{\lambda,\alpha})^{-1} < \infty, \tag{4}$$

where

$$A_{\lambda,\alpha} = \alpha\int_0^1 (1 - \mathrm{e}^{\lambda(1-\alpha)z})z^{-(1+\alpha)}\,\mathrm{d}z.$$

Proof. Clearly, $A_{\lambda,\alpha}$ is nonnegative for $\lambda \le 0$, and converges to $-\infty$ as λ tends to $+\infty$. It is known (Equation (77) in [27]) that

$$\mathbb{E}[\exp\{\lambda(1-\alpha)(P_1^{-1}(\alpha, 0) - 1)\}] = (1 + A_{\lambda,\alpha})^{-1} < \infty \quad \text{for } \lambda \le 0.$$

For $\lambda > 0$, we have

$$A_{\lambda,\alpha} = (1-\lambda)\mathrm{e}^{\lambda(1-\alpha)} - 1 + \lambda^2(1-\alpha)\int_0^1 z^{1-\alpha}\mathrm{e}^{\lambda(1-\alpha)z}\,\mathrm{d}z$$
$$\ge (1-\lambda)\mathrm{e}^{\lambda(1-\alpha)} - 1 + \lambda^2(1-\alpha)\int_0^1 z^{1-\alpha}\mathrm{e}^{\lambda(1-\alpha)z}\,\mathrm{d}z. \tag{5}$$

If we define

$$\lambda_\alpha = \sup\{\lambda > 0\colon A_{\lambda,\alpha} + 1 > 0\},$$

then $\lambda_\alpha \ge 1$ by (5) and

$$\delta = \inf\{\lambda_\alpha : 0 < \alpha < 1\} \ge 1.$$

By Campbell's theorem, (4) holds for any $\lambda < \delta$. □

Lemma 3. *Let $\varepsilon > 0$ be arbitrarily given. If $c_1 = 0$ then*

$$\limsup_{\alpha\to 0}\frac{\gamma(\alpha)}{\alpha}\log\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha, 0) - 1| > \varepsilon\} = -\infty. \tag{6}$$

If $c_1 = \infty$ and

$$\lim_{\alpha\to 0}\gamma(\alpha) = 0, \tag{7}$$

then

$$\limsup_{\alpha\to 0}\frac{1}{\log\gamma(\alpha)/\alpha}\log\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha, 0) - 1| > \varepsilon\} = -\infty. \tag{8}$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Proof. Since the limit involves only small α, we may assume that $0 < \alpha < \frac{1}{2}$ and $0 < \varepsilon < \frac{1}{2}$. Let δ be as in Lemma 2 and set $\delta_1 = \delta/4$. By direct calculation, we obtain

$$\begin{aligned}\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha,0)-1|>\varepsilon\} &= \mathbb{P}\{P_1^{-1}(\alpha,0)-1 \geq (1-\varepsilon)^{-1/\gamma(\alpha)}-1\}\\ &\leq \mathbb{E}[\exp(\delta_1(P_1^{-1}(\alpha,0)-1))]\exp(-\delta_1[(1-\varepsilon)^{-1/\gamma(\alpha)}-1])\\ &\leq (1+A_{\delta_1,\alpha})^{-1}\exp(-\delta_1[(1-\varepsilon)^{-1/\gamma(\alpha)}-1]). \end{aligned} \tag{9}$$

From (5), it follows that

$$\lim_{\alpha\to 0}(1+A_{\delta_1,\alpha}) = 1. \tag{10}$$

If $c_1 = 0$ then

$$\begin{aligned}&\limsup_{\alpha\to 0}\frac{(1-\varepsilon)^{-1/\gamma(\alpha)}-1}{\alpha/\gamma(\alpha)}\\ &\quad = \limsup_{\alpha\to 0}\frac{(1-\varepsilon)^{-1/\gamma(\alpha)}}{\alpha/\gamma(\alpha)}\\ &\quad = \limsup_{\alpha\to 0}\exp\bigg\{\frac{1}{\gamma(\alpha)}\bigg[\log\frac{1}{(1-\varepsilon)}+\gamma(\alpha)\log\gamma(\alpha)-\frac{\gamma(\alpha)}{\alpha}\alpha\log\alpha\bigg]\bigg\}\\ &\quad = \infty. \end{aligned} \tag{11}$$

Next assume that $c_1 = \infty$ and (7) holds. For any $0 < \varepsilon < \frac{1}{2}$, $(1-\varepsilon)^{1/\gamma(\alpha)}$ converges to 0 as α tends to 0. Hence, for any $k \geq 1$, we can find $\alpha_k > 0$ such that, for all $0 < \alpha < \alpha_k$,

$$\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha,0)-1|>\varepsilon\} \leq \mathbb{P}\bigg\{P_1(\alpha,0) < \frac{1}{k}\bigg\}.$$

By the large deviation principle for $P_1(\alpha, 0)$ (see Lemma 2.3 in [15]), we obtain

$$\begin{aligned}\limsup_{\alpha\to 0}\frac{1}{\log 1/\alpha}\log\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha,0)-1|>\varepsilon\} &\leq \limsup_{\alpha\to 0}\frac{1}{\log 1/\alpha}\log\mathbb{P}\bigg\{P_1(\alpha,0)<\frac{1}{k}\bigg\}\\ &\leq -(k-1).\end{aligned}$$

Noting that $\gamma(\alpha) < 1$ and k is arbitrary, it follows that

$$\begin{aligned}&\limsup_{\alpha\to 0}\frac{1}{\log\gamma(\alpha)/\alpha}\log\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha,0)-1|>\varepsilon\}\\ &\quad\leq \limsup_{\alpha\to 0}\frac{1}{\log 1/\alpha}\log\mathbb{P}\{|P_1^{\gamma(\alpha)}(\alpha,0)-1|>\varepsilon\}\\ &\quad\leq \lim_{k\to\infty}\limsup_{\alpha\to 0}\frac{1}{\log 1/\alpha}\log\mathbb{P}\bigg\{P_1(\alpha,0)\leq\frac{1}{k}\bigg\}\\ &\quad= -\infty. \end{aligned} \tag{12}$$

Combining (9)–(12), we obtain (6) and (8). □

Theorem 4. *Let $\gamma(\alpha)$ satisfy (2), and set*

$$\nabla = \{\boldsymbol{x} = (x_1, x_2, \ldots)\colon 1 \geq x_1 \geq x_2 \geq \cdots \geq 0\}.$$

Then the followings hold as α converges to 0.

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

(i) *If* $c_1 = 0$ *then the family* $\{\mathbb{P}^{\gamma(\alpha)}(\alpha, 0) \colon 0 < \alpha < 1\}$ *satisfies a large deviation principle on space* ∇ *with speed* $\alpha/\gamma(\alpha)$ *and rate function*

$$I_1(\boldsymbol{x}) = \begin{cases} \displaystyle\sum_{n=1}^{\infty} n \log \frac{x_n}{x_{n+1}}, & x_1 = 1,\ x_n > 0 \text{ for all } n > 1, \\ +\infty & \text{otherwise}. \end{cases}$$

(ii) *If* $c_1 = \infty$ *and (7) holds then the family* $\{\mathbb{P}^{\gamma(\alpha)}(\alpha, 0) \colon 0 < \alpha < 1\}$ *satisfies a large deviation principle on space* ∇ *with speed* $\log\gamma(\alpha)/\alpha$ *and rate function*

$$I_2(\boldsymbol{x}) = \begin{cases} n - 1, & x_1 = 1,\ x_n > 0,\ x_k = 0,\ k > n, \\ +\infty & \text{otherwise}. \end{cases}$$

Proof. Writing $\mathbb{P}^{\gamma(\alpha)}$ in terms of $\boldsymbol{R}^{\gamma(\alpha)}$, we have

$$\mathbb{P}^{\gamma(\alpha)} = P_1^{\gamma(\alpha)}(\alpha, 0)(1, R_1^{\gamma(\alpha)}, R_1^{\gamma(\alpha)} R_2^{\gamma(\alpha)}, \dots).$$

By Lemma 3, $P_1^{\gamma(\alpha)}(\alpha, 0)$ is exponentially equivalent to 1. Hence, by Lemma 2.1 in [17], $(1, R_1^{\gamma(\alpha)}, R_1^{\gamma(\alpha)} R_2^{\gamma(\alpha)}, \dots)$ and $\mathbb{P}^{\gamma(\alpha)}$ have the same large deviation principle. Define

$$\psi : [0, 1]^\infty \to \nabla, \qquad (x_1, x_2, \dots) \to (1, x_1, x_1 x_2, \dots).$$

Then ψ is clearly continuous and $(1, R_1^{\gamma(\alpha)}, R_1^{\gamma(\alpha)} R_2^{\gamma(\alpha)}, \dots) = \psi(\boldsymbol{R}^{\gamma(\alpha)})$. Noting that

$$I_i(\boldsymbol{x}) = \inf\{\tilde{I}_i(\boldsymbol{y}) \colon \psi(\boldsymbol{y}) = \boldsymbol{x}\}, \qquad i = 1, 2,$$

the theorem follows from Lemma 1 and the contraction principle. □

4. Asymptotic behaviour of $\Xi_{\alpha,0,\nu}$

Recall that the REM has configuration space $S_N = \{-1, 1\}^N$ and the Hamiltonian given by a family of i.i.d. normal random variables with mean 0 and variance N,

$$\{H_N(\sigma) \mid \sigma \in S_N\}.$$

The Gibbs measure $G_N(\sigma)$ at temperature T is given by

$$Z_N^{-1} \exp\{-\beta H_N(\sigma)\},$$

where $\beta = 1/T$ and $Z_N = \sum_{\sigma \in S_N} \exp\{-\beta H_N(\sigma)\}$. By making the change of variable

$$r_N(\sigma) = 1 - \sum_{i=1}^{N} (1 - \sigma_i) 2^{-i-1},$$

we can regard [0, 1] as the new configuration space. The corresponding Gibbs measure has the form

$$\mu_N^T(\mathrm{d}x) = \sum_{\sigma \in S_N} \delta_{r_N}(\mathrm{d}x) G_N(\sigma).$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

As $N \to \infty$, the limiting Gibbs measure $\mu^T = \lim_{N\to\infty} \mu_N^T$ exhibits phase transition at the critical temperature $T_c = (\sqrt{2\log 2})^{-1}$. More specifically, by Theorems 9.3.1 and 9.3.4 in [5], we have

$$\mu^T = \begin{cases} \nu & \text{if } T \geq T_c, \\ \Xi_{\alpha,0,\nu} & \text{if } T < T_c. \end{cases}$$

Thus, a phase transition occurs when the temperature crosses the critical value between high temperature and low temperature regimes. The low temperature regime has a rich structure. The transition from the low temperature regime to the critical temperature regime corresponds to α tending to 1 from below. The goal of this section is to understand the microscopic behaviour of this transition through the establishment of a large deviation principle for $\Xi_{\alpha,0,\nu}$.

4.1. Estimates for stable subordinator

Recall that ρ_t is the stable subordinator with index $0 < \alpha < 1$. For $t = 1$, the following holds.

Lemma 4. (See [26] and [35].) *The distribution function of $\rho_1^{\alpha/(1-\alpha)}$ has two integral representations:*

$$F(x) = \mathbb{P}\{\rho_1^{\alpha/(1-\alpha)} \leq x\} = \frac{1}{\pi}\int_0^\pi \mathrm{e}^{-A(u)/x}\,\mathrm{d}u, \tag{13}$$

where $A(u)$ is the Zolotarev function defined as

$$A(u) = \left\{\frac{\sin^\alpha(\alpha u)\sin^{1-\alpha}((1-\alpha)u)}{\sin u}\right\}^{1/(1-\alpha)}.$$

The distribution function of ρ_1 is, thus, $F(x^{\alpha/(1-\alpha)})$. The density function of ρ_1 has the following representation:

$$\phi_\alpha(t) = \frac{1}{\pi}\int_0^\infty \mathrm{e}^{-tu}\mathrm{e}^{-u^\alpha\cos\pi\alpha}\sin(u^\alpha\sin\pi\alpha)\,\mathrm{d}u. \tag{14}$$

Applying these representations, we obtain the following estimations.

Theorem 5. *For any given $1 > \delta > 0$, we have*

$$\lim_{\alpha\to1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 < 1-\delta\}} = \lim_{\alpha\to1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 \leq 1-\delta\}} = \log\frac{1}{1-\delta} \tag{15}$$

and

$$\lim_{\alpha\to1}\frac{1}{\log 1/(1-\alpha)}\log\mathbb{P}\{\rho_1 > 1+\delta\} = \lim_{\alpha\to1}\frac{1}{\log 1/(1-\alpha)}\log\mathbb{P}\{\rho_1 \geq 1+\delta\} = -1. \tag{16}$$

Proof. For any $u \in (0,\pi)$, $v \in (0,1)$, we have

$$\frac{\mathrm{d}[v\cot(vu) - \cot u]}{\mathrm{d}v} = \frac{1}{2\sin^2(vu)}(\sin(vu) - 2vu) \leq \frac{1}{2\sin^2(vu)}(\sin(vu) - vu) \leq 0,$$

which implies that

$$\frac{\mathrm{d}\log\sin(vu)/\sin u}{\mathrm{d}u} = v\cot(vu) - \cot u \geq 0.$$

Hence,

$$A(u) = \exp\left\{\alpha\log\frac{\sin(\alpha u)}{\sin u} + (1-\alpha)\log\frac{\sin((1-\alpha)u)}{\sin u}\right\}$$

is nondecreasing in u. Furthermore, it follows from direct calculation that

$$A(0+) = \lim_{u\to 0} A(u) = (1-\alpha)\alpha^{\alpha/(1-\alpha)}, \qquad \lim_{u\to\pi} A(u) = \infty.$$

Therefore, applying representation (13), we have, for any $\epsilon > 0$,

$$\begin{aligned}\frac{\pi-\epsilon}{\pi}\exp\bigg\{-\frac{A(\pi-\epsilon)}{(1-\delta)^{\alpha/(1-\alpha)}}\bigg\} &\le \frac{1}{\pi}\int_0^{\pi-\epsilon}\exp\bigg(-\frac{A(u)}{(1-\delta)^{\alpha/(1-\alpha)}}\bigg)\,\mathrm{d}u \\ &= \mathbb{P}\{\rho_1 \le 1-\delta\} \\ &\le \exp\bigg\{-\frac{A(0+)}{(1-\delta)^{\alpha/(1-\alpha)}}\bigg\}.\end{aligned}$$

This implies that

$$\frac{1}{\mathbb{P}\{\rho_1 \le 1-\delta\}} \ge \exp\bigg\{\exp\bigg\{\log A(0+) + \frac{\alpha}{1-\alpha}\log\frac{1}{1-\delta}\bigg\}\bigg\}$$

and

$$\frac{1}{\mathbb{P}\{\rho_1 \le 1-\delta\}} \le \frac{\pi}{\pi-\varepsilon}\exp\bigg\{\exp\bigg\{\log A(\pi-\varepsilon) + \frac{\alpha}{1-\alpha}\log\frac{1}{1-\delta}\bigg\}\bigg\}.$$

Taking the double logarithm and letting α converge to 1, we have

$$\begin{aligned}\log\frac{1}{1-\delta} &\le \liminf_{\alpha\to 1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 \le 1-\delta\}} \\ &= \liminf_{\alpha\to 1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 < 1-\delta\}} \\ &\le \limsup_{\alpha\to 1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 < 1-\delta\}} \\ &= \limsup_{\alpha\to 1}(1-\alpha)\log\log\frac{1}{\mathbb{P}\{\rho_1 \le 1-\delta\}} \\ &\le \log\bigg(\frac{1}{1-\delta}\bigg),\end{aligned}$$

which is (15).

To prove (16), we apply (14) to obtain

$$\begin{aligned}\mathbb{P}\{\rho_1 > 1+\delta\} &= \mathbb{P}\{\rho_1 \ge 1+\delta\} \\ &= \frac{1}{\pi}\int_{1+\delta}^{\infty}\int_0^{\infty} u^{-1}\mathrm{e}^{-(1+\delta)u}\mathrm{e}^{-u^{\alpha}\cos\pi\alpha}\sin(u^{\alpha}\sin\pi\alpha)\,\mathrm{d}u\,\mathrm{d}t \\ &= \frac{\sin\pi\alpha}{\pi}\int_0^{\infty} u^{-(1-\alpha)}\mathrm{e}^{-\delta u}\bigg[\mathrm{e}^{-u-u^{\alpha}\cos\pi\alpha}\frac{\sin(u^{\alpha}\sin\pi\alpha)}{u^{\alpha}\sin\pi\alpha}\bigg]\,\mathrm{d}u.\end{aligned}$$

Noting that $\sin(u^{\alpha}\sin\pi\alpha)/u^{\alpha}\sin\pi\alpha$ is bounded and

$$\lim_{\alpha\to 1}\frac{\sin\pi\alpha}{\pi(1-\alpha)} = 1,$$

it follows that (16) holds. □

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Theorem 6. *The family* $\{\rho_1 : 0 < \alpha < 1\}$ *satisfies a large deviation principle on* $(0, \infty)$ *as* α *tends to* 1 *with speed* $-\log(1-\alpha)$ *and rate function (not good in this case)*

$$J(x) = \begin{cases} 1, & x > 1, \\ 0, & x = 1, \\ +\infty & \text{otherwise}. \end{cases}$$

Proof. Let A be a closed set in $(0, \infty)$. If A contains 1 then $\inf_{x\in A} J(x) = 0$ and the upper estimate holds. If A does not contain 1 then we can find $0 < a < 1 < b$ such that A is either a subset of $(0, a]$, a subset of $[b, \infty)$, or a subset of $(0, a] \cup [b, \infty)$. For each case, we can apply Theorem 5 to obtain the upper estimate.

The proof for the lower estimates is as follows. Let B be any open set. If B intersects with $[0, 1)$ then the lower estimates are trivial. If B does not intersect with $[0, 1)$ then B cannot contain 1. Hence, we can find $1 < a < b < \infty$ such that $(a, b) \subset B$ and

$$\mathbb{P}\{\rho_1 \in B\} \ge \mathbb{P}\{\rho_1 \in (a, b)\} \ge \frac{b-a}{\pi} \int_0^\infty u^{-1} \mathrm{e}^{-bu} \mathrm{e}^{-u^\alpha \cos \pi\alpha} \sin(u^\alpha \sin \pi\alpha)\, \mathrm{d}u,$$

which implies that

$$\liminf_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{\rho_1 \in B\} \ge -1 = -\inf_{x\in B} J(x). \qquad \square$$

For any $n \ge 1$, let $\tau_1, \dots, \tau_{n+1}$ be independent copies of ρ_1. Set

$$\sigma_i = \frac{\tau_i}{\tau_1}, \qquad i = 2, \dots, n+1. \tag{17}$$

Set

$$\tilde{\sigma}_n = \min\{\sigma_i : 2 \le i \le n+1\}$$

and let r_n denote the frequency of $\tilde{\sigma}_n$ among $\{\sigma_i\}_{i=2,\dots,n+1}$. Define

$$J_n(u_1, \dots, u_n) = \begin{cases} n + 1 - r_n, & \tilde{\sigma}_n < 1, \\ n - r_n, & \tilde{\sigma}_n = 1, \\ n, & \tilde{\sigma}_n > 1. \end{cases}$$

Clearly, $J_n(\cdot)$ is a rate function on $(0, \infty)^n$.

Theorem 7. *The family* $\{(\sigma_2, \dots, \sigma_{n+1}) : 0 < \alpha < 1\}$ *satisfies a large deviation principle on* $(0, \infty)^n$ *with speed* $-\log(1-\alpha)$ *and rate function* $J_n(\cdot)$ *as* α *tends to* 1.

Proof. Note that the map

$$\Phi : (0, \infty)^{n+1} \to (0, \infty)^n, \qquad (x_1, \dots, x_{n+1}) \to \left(\frac{x_2}{x_1}, \dots, \frac{x_{n+1}}{x_1}\right)$$

is clearly continuous. It follows from the contraction principle that large deviation upper and lower estimates hold for the family $\{(\sigma_2, \dots, \sigma_{n+1}) : 0 < \alpha < 1\}$ with the bounds given by the function

$$\tilde{J}_n(u_1, \dots, u_n) = \inf\left\{\sum_{i=1}^{n+1} J(x_i) : x_{j+1} = u_j x_1,\ j = 1, \dots, n\right\}.$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Since $J(x) = \infty$ for x in $(0, 1)$, it follows that

$$\begin{aligned}\tilde{J}_n(u_1, \dots, u_n) &= \inf\{\sum_{i=1}^{n+1} J(x_i) \colon x_1 \geq 1,\ x_{j+1} = u_j x_1 \geq 1,\ j = 1, \dots, n\} \\ &= J_n(u_1, \dots, u_n)\end{aligned}$$

and the theorem follows. □

Remark 3. The contraction principle used in Theorem 7 does not lead to a large deviation principle in general due to the fact that the starting rate function is not good. But here and later on, direct calculations show that the upper and lower bounds are all given by rate functions.

4.2. Large deviations for $\Xi_{\alpha,0,\nu}$

Let $M_1([0, 1])$ denote the space of probabilities on $[0, 1]$ equipped with the weak topology. For any μ in $M_1([0, 1])$, define

$$\mathcal{I}(\mu) = \begin{cases} 0, & \mu = \nu, \\ n, & \mu = \sum_{i=1}^n p_i \delta_{x_i} + (1 - \sum_{i=1}^n p_i)\nu, \\ \infty, & \text{otherwise.} \end{cases}$$

The main result of this subsection is the following theorem.

Theorem 8. *The family* $\{\Xi_{\alpha,0,\nu} \colon 0 < \alpha < 1\}$ *satisfies a large deviation principle on* $M_1([0, 1])$ *with speed* $-\log(1 - \alpha)$ *and good rate function* $\mathcal{I}(\cdot)$ *as* α *tends to* 1.

We prove this theorem through a series of lemmas.

Lemma 5. *For any* $n \geq 1$*, let* $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = 1$ *and* $B_1, \dots, B_{n+1}$ *be a measurable partition of* $[0, 1]$ *such that* $\nu(B_i) = t_i - t_{i-1}$*. Then*

$$\begin{aligned}&(\Xi_{\alpha,0,\nu}(B_1), \dots, \Xi_{\alpha,0,\nu}(B_{n+1})) \\ &\quad \overset{\mathrm{D}}{=} \rho_1^{-1}(\rho_{t_1}, \rho_{t_2} - \rho_{t_1}, \dots, \rho_{t_k} - \rho_{t_{k-1}}, \rho_1 - \rho_{t_k}) \\ &\quad \overset{\mathrm{D}}{=} \bigg(t_1^{1/\alpha} + \sum_{k=2}^{n+1} (t_k - t_{k-1})^{1/\alpha} \sigma_k\bigg)^{-1} (t_1^{1/\alpha}, (t_2 - t_1)^{1/\alpha} \sigma_2, \dots, (1 - t_n)^{1/\alpha} \sigma_{n+1}),\end{aligned}$$

where $\sigma_2, \dots, \sigma_{n+1}$ *are defined in (17), and '*$\overset{\mathrm{D}}{=}$*' denotes equality in distribution.*

Proof. The first equality is from [31] and the second equality follows from the independent increments of the stable subordinator and $\rho_t \overset{\mathrm{D}}{=} t^{1/\alpha} \rho_1$. □

Lemma 6. *Let*

$$\triangle_{n+1} := \bigg\{(y_1, \dots, y_{n+1}) \colon y_i \geq 0, \sum_{k=1}^{n+1} y_k = 1\bigg\}.$$

Then the family $\{(\Xi_{\alpha,0,\nu}(B_1), \dots, \Xi_{\alpha,0,\nu}(B_{n+1})) \colon 0 < \alpha < 1\}$ *satisfies a large deviation principle on* $\triangle_{n+1}$ *with speed* $-\log(1-\alpha)$ *and good rate function* $\mathcal{I}_n(\cdot)$ *as* α *tends to* 1*, where*

$$\mathcal{I}_n(y_1, \dots, y_{n+1}) = (n + 1) - \gamma(y_1, \dots, y_{n+1})$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

with

$$\gamma(y_1,\dots,y_{n+1}) = \#\bigg\{1\le i\le n+1\colon \frac{y_i}{t_i-t_{i-1}} = \min\bigg\{\frac{y_k}{t_k-t_{k-1}}\colon 1\le k\le n+1\bigg\}\bigg\}.$$

Proof. First note that the map

$$H:[0,1]^n\times(0,\infty)^n\to[0,1],$$
$$(v_1,\dots,v_{n+1};\,u_1,\dots,u_n)\to\bigg(v_1+\sum_{k=2}^{n+1}v_ku_{k-1}\bigg)^{-1}(v_1,v_2u_1,\dots,v_{n+1}u_n)$$

is continuous and $(\Xi_{\alpha,0,\nu}(B_1),\dots,\Xi_{\alpha,0,\nu}(B_{n+1}))$ has the same distribution as

$$H(t_1^{1/\alpha},\dots,(1-t_n)^{1/\alpha};\,\sigma_2,\dots,\sigma_{n+1}).$$

Noting that $(t_1^{1/\alpha},\dots,(1-t_n)^{1/\alpha})$ satisfies a full large deviation principle with effective domain $(t_1,\dots,(1-t_n))$. It follows from Theorem 7, the independence between $(t_1^{1/\alpha},\dots,(1-t_n)^{1/\alpha})$ and $(\sigma_2,\dots,\sigma_{n+1})$, and the contraction principle that large deviation estimates hold for $(\Xi_{\alpha,0,\nu}(B_1),\dots,\Xi_{\alpha,0,\nu}(B_{n+1}))$ with upper and lower bounds given by the function

$$\begin{aligned}\tilde{\mathcal{I}}_n(y_1,\dots,y_{n+1}) &= \inf\bigg\{J_n(u_1,\dots,u_n)\colon u_i\in(0,\infty),\ u_i=\frac{t_1}{y_1}\frac{y_{i+1}}{t_{i+1}-t_i},\ i=1,\dots,n\bigg\}\\ &= \begin{cases} n+1-\tilde r_n, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}<y_1/t_1,\\ n-\tilde r_n, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}=y_1/t_1,\\ n, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}>y_1/t_1,\end{cases}\end{aligned}$$

where $\tilde r_n$ is the frequency of $\min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}$ among $y_2/(t_2-t_1),\dots,y_{n+1}/(1-t_n)$. On the other hand,

$$\gamma(y_1,\dots,y_{n+1}) = \begin{cases}\tilde r_n, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1}\}<y_1/t_1,\\ \tilde r_n+1, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}=y_1/t_1,\\ 1, & \min_{2\le i\le n+1}\{y_i/(t_i-t_{i-1})\}>y_1/t_1.\end{cases}$$

Hence, we obtain $\tilde{\mathcal{I}}_n(\cdot)=\mathcal{I}_n(\cdot)$. It remains to show that $\mathcal{I}_n(\cdot)$ is a good rate function. Since $\triangle_{n+1}$ is compact, it suffices to verify the lower semicontinuity of the $\mathcal{I}_n(\cdot)$. For any point $(y_1,\dots,y_{n+1})$ in $\triangle_{n+1}$, let $\gamma(y_1,\dots,y_{n+1})=m$. If the neighbourhood of $(y_1,\dots,y_{n+1})$ is small enough then the frequency of the minimum in each point inside the neighbourhood is at least m. Hence, $\mathcal{I}(\cdot)$ is lower semicontinuous. □

Lemma 7. *We have*

$$\mathcal{I}(\mu)=\sup\{\mathcal{I}_n(\mu([0,t_1]),\mu((t_1,t_2]),\dots,\mu((t_n,1])\colon 0=t_0<t_1<\cdots<t_n<t_{n+1}=1,\ n=1,2,\dots\}. \tag{18}$$

The supremum can be taken over all continuity points $t_1,\dots,t_n$ of μ.

Proof. We divide the proof into several cases. Let μ be any probability in $M_1([0,1])$. By Lebesgue's decomposition theorem, we can write

$$\mu=\lambda_1\mu_{\mathrm{a}}+\lambda_2\mu_{\mathrm{s}}+\lambda_3\mu_{\mathrm{ac}},$$

where μ_a is atomic, μ_{s} is diffuse and singular with respect to ν, μ_{ac} is absolutely continuous with respect to ν, and

$$\lambda_1 + \lambda_2 + \lambda_3 = 1, \qquad \lambda_i \geq 0, \quad i = 1, 2, 3.$$

Set

$$F_{\mathrm{s}}(x) = \mu_{\mathrm{s}}([0, x]), \quad f(x) = \frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\nu}(x).$$

Case 1. The probability μ *has a countable number of atoms.* Since the total mass of μ_{a} is equal to 1, there exists a countable infinite number of atoms all with different values of masses. Let the masses of these atoms be ranked in descending order and the corresponding atoms are $x_1, x_2, \dots$. Clearly, $\mu_{\mathrm{s}}(\{x_i\}) = \mu_{\mathrm{ac}}(\{x_i\}) = 0$ for all $i \geq 1$. For any $m \geq 2$, by the continuity of probabilities, we can choose small positive numbers $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ such that $x_i \pm \varepsilon_i$, $1 \leq i \leq m$, are the continuity points of μ, $(x_i - \varepsilon_i, x_i + \varepsilon_i] \subset [0, 1]$, $1 \leq i \leq m$, are disjoint, and

$$\mu((x_1 - \varepsilon_1, x_1 + \varepsilon_1]) > \mu((x_2 - \varepsilon_2, x_2 + \varepsilon_2]) > \cdots > \mu((x_m - \varepsilon_m, x_m + \varepsilon_m]).$$

The partition based on the points $\{x_i \pm \varepsilon_i,\ i = 1, 2, \dots, m\}$ clearly yields a lower bound $m - 1$ for $\mathcal{I}(\cdot)$. Since m is arbitrary, the supremum taken over continuity points of μ gives the value of ∞ which is the same as $\mathcal{I}(\cdot)$.

Case 2. The probability μ *has at most a finite number of atoms and* $\nu(\{f(x) \neq 1\}) > 0$. Let $A = \{x \in [0, 1] \colon f(x) < 1\}$, $B = \{x \in [0, 1] \colon f(x) > 1\}$, and $C = \{x \in [0, 1] \colon f(x) = 1\}$. Then, we have

$$\mu_{\mathrm{ac}}(A) < \nu(A), \qquad \mu_{\mathrm{ac}}(B) > \nu(B), \qquad \mu_{\mathrm{ac}}(C) = \nu(C),$$

and

$$\nu(A) - \mu_{\mathrm{ac}}(A) = \mu_{\mathrm{ac}}(B) - \nu(B).$$

The fact that $\nu\{C\} < 1$ thus implies that $\nu(A) > 0$ and $\nu(B) > 0$. For any $m \geq 1$, we can find $0 < s_1 < \cdots < s_m < 1$ and $0 < t_1 < \cdots < t_m < 1$ such that

$$\{s_i\}_{1 \leq i \leq m} \subset A, \qquad \{t_i\}_{1 \leq i \leq m} \subset B,$$

$\{s_i, t_i\}_{i \geq 1}$ does not contain atoms of μ when $\lambda_2 > 0$, $F_{\mathrm{s}}'(x) = 0$ for $x = s_i$ or, t_i, $i \geq 1$.

For any $i, j \geq 1$, we then have

$$\begin{aligned}
\lim_{\varepsilon \to 0} \frac{\mu((s_i - \varepsilon, s_i + \varepsilon])}{2\varepsilon} &= \lambda_3 \lim_{\varepsilon \to 0} \frac{\mu_{\mathrm{ac}}((s_i - \varepsilon, s_i + \varepsilon])}{2\varepsilon} \\
&= \lambda_3 f(s_i) \\
&< \lambda_3 f(t_j) \\
&= \lambda_3 \lim_{\varepsilon \to 0} \frac{\mu_{\mathrm{ac}}((t_j - \varepsilon, t_j + \varepsilon])}{2\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\mu((t_j - \varepsilon, t_j + \varepsilon])}{2\varepsilon}.
\end{aligned}$$

This makes it possible to choose $\varepsilon_i > 0$ such that $s_i \pm \varepsilon_i$, $t_j \pm \varepsilon_j$ are all continuity points of μ and

$$\frac{\mu((s_i - \varepsilon_i, s_i + \varepsilon_i])}{\nu(s_i - \varepsilon_i, s_i + \varepsilon_i]} < \frac{\mu((t_j - \varepsilon_j, t_j + \varepsilon_j])}{\nu(t_j - \varepsilon_j, t_j + \varepsilon_j]}.$$

This provides a lower bound of m for $\mathcal{I}(\mu)$. Since m is arbitrary, we have established (18) in this case.

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Case 3. The probability μ has at most a finite number of atoms, $\lambda_2 > 0$ and $\nu(\{f(x) \neq 1\}) = 0$. It is clear that we have $\mu_{\text{ac}} = \nu$ in this case. For any $m \geq 1$, the singularity guarantees the existence of $0 < s_1 < \cdots < s_m < 1$ and $0 < t_1 < \cdots < t_m < 1$ such that the derivative of $F_s(x)$ is 0 for $x = t_i$, while the derivative at s_i is either ∞ or does not exist. Additionally, we can choose s_i and t_i so that none of them are atoms of μ_{a}. Let ε be small enough so that all intervals $(s_i - \varepsilon, s_i + \varepsilon]$ and $(t_i - \varepsilon, t_i + \varepsilon]$, $i = 1, \ldots, m$, are disjoint. Let $\mathcal{J}$ denote the partition of $[0, 1]$ using $\{t_i \pm \varepsilon, s_i \pm \varepsilon : i = 1, \ldots, m\}$. We can then find a refined partition, using a subsequence if necessary, $\tilde{\mathcal{J}}$ of $\mathcal{J}$, and positive numbers ε_0 and δ_0 such that $s_i \pm \varepsilon_0, t_i \pm \varepsilon_0$ are continuity points of μ and the value of $(2\varepsilon_0)^{-1}\mu_s$ on each interval containing one of the t_i is less than δ_0 while its value on each interval containing one of the s_i is greater than δ_0. In other words, we can have, for any $1 \leq i, j \leq m$,

$$\frac{\mu((s_i - \varepsilon_0, s_i + \varepsilon_0])}{\nu((s_i - \varepsilon_0, s_i + \varepsilon_0])} \neq \frac{\mu((t_j - \varepsilon_0, t_j + \varepsilon_0])}{\nu((t_j - \varepsilon_0, t_j + \varepsilon_0])}.$$

This implies that

$$\sup\{\mathcal{I}_n(\mu([0, t_1]), \mu((t_1, t_2]), \ldots, \mu((t_n, 1]) : 0 = t_0 < \cdots < t_{n+1} = 1,\ n \geq 1\} \geq m.$$

The arbitrary selection of m leads to (18) in this case.

Case 4. The probability μ has at most a finite number of atoms, $\lambda_2 = 0$ and $\nu(\{f(x) \neq 1\}) = 0$. In this case we have $\mu = \lambda_1\mu_{\text{a}} + \lambda_3\nu$. If $\lambda_1 = 0$ then $\mu = \nu$ and $\mathcal{I}(\mu)$ is clearly 0. Assume that $\lambda_1 > 0$ and the number of atoms is r. Let $F(x) = \mu([0, x])$. Since r is finite, any partition $\mathcal{J}$ of $[0, 1]$ will have at most r disjoint intervals covering these atoms. The maximum

$$\sup\{\mathcal{I}_n(\mu([0, t_1]), \mu((t_1, t_2]), \ldots, \mu((t_n, 1]) : 0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = 1, n \geq 1\}$$

is achieved at any partition with exactly r disjoint intervals covering the r atoms. □

Proof of Theorem 8. Let $C([0, 1])$ be the space of all continuous function on $[0, 1]$ equipped with the supremum norm, and $\{g_j(x) : j = 1, 2, \ldots\}$ be a countable dense subset of $C([0, 1])$. The set $\{g_j(x) : j = 1, 2, \ldots\}$ is clearly convergence determining on $M_1([0, 1])$. Let $|g_j| = \sup_{x\in[0,1]}|g_j(x)|$ and $\{h_j(x) = g_j(x)/|g_j| \vee 1 : j = 1, \ldots\}$ is also convergence determining.

For any μ, υ in $M_1([0, 1])$, define

$$d(\mu, \upsilon) = \sum_{j=1}^{\infty} \frac{1}{2^j}|\langle\mu, h_j\rangle - \langle\upsilon, h_j\rangle|.$$

Then d is a metric generating the weak topology on $M_1([0,])$.

For any $\delta > 0$, $\mu \in M_1([0, 1])$, let

$$B(\mu, \delta) = \{\upsilon \in M_1([0, 1]) : d(\upsilon, \mu) < \delta\}, \qquad \bar{B}(\mu, \delta) = \{\upsilon \in M_1([0, 1]) : d(\upsilon, \mu) \leq \delta\}.$$

Since $M_1([0, 1])$ is compact, the family of the laws of $\Xi_{\alpha,0,\nu}$ is exponentially tight. By Theorem (P) in [36], to prove the theorem it suffices to verify that

$$\begin{aligned}\lim_{\delta\to 0}\liminf_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{B(\mu, \delta)\} &= \lim_{\delta\to 0}\limsup_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{\bar{B}(\mu, \delta)\} \\ &= -\mathcal{I}(\mu).\end{aligned}$$

Let m be large enough so that

$$\Big\{\upsilon \in M_1([0, 1]) : |\langle\mu, h_j\rangle - \langle\upsilon, h_j\rangle| < \tfrac{1}{2}\delta : j = 1, \ldots, m\Big\} \subset B(\nu, \delta). \qquad (19)$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Consider $0 = t_0 < t_1 < \cdots < t_{n+1} = 1$ with $A_i = (t_{i-1}, t_i]$, $i = 1, \dots, n+1$, such that

$$\sup\{|h_j(x) - h_j(y)| \colon x, y \in A_i,\ i = 1, \dots, n;\ j = 1, \dots, m\} < \tfrac{1}{8}\delta.$$

Choose $0 < \delta_1 < \delta/4n$, and define

$$V_{t_1,\dots,t_n}(\mu, \delta_1) = \{(y_1, \dots, y_n) \in \triangle_n \colon |y_i - \mu(A_i)| < \delta_1,\ i = 1, \dots, n\}.$$

For any υ in $M_1([0, 1])$, let

$$\Psi(\upsilon) = (\upsilon(A_1), \dots, \upsilon(A_{n+1})).$$

If $\Psi(\upsilon)$ belongs to $V_{t_1,\dots,t_n}(\mu, \delta_1)$ then, for $j = 1, \dots, m$,

$$|\langle \upsilon, h_j\rangle - \langle \mu, h_j\rangle| = \left|\sum_{i=1}^{n+1} \int_{A_i} h_j(x)(\upsilon(\mathrm{d}x) - \mu(\mathrm{d}x))\right| < \frac{\delta}{4} + n\delta_1 < \frac{\delta}{2},$$

which implies that

$$\Psi^{-1}(V_{t_1,\dots,t_n}(\mu, \delta_1)) \subset \left\{\upsilon \in M_1([0, 1]) \colon |\langle \upsilon, h_j\rangle - \langle \mu, h_j\rangle| < \tfrac{1}{2}\delta \colon j = 1, \dots, m\right\}.$$

This combined with (19) implies that

$$\Psi^{-1}(V_{t_1,\dots,t_n}(\mu, \delta_1)) \subset B(\mu, \delta).$$

Since $V_{t_1,\dots,t_n}(\mu, \delta_1)$ is open in $\triangle_n$, it follows from Lemma 6 that

$$\begin{aligned}
&\lim_{\delta\to 0} \liminf_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{B(\mu, \delta)\} \\
&\quad\geq \lim_{\delta\to 0} \liminf_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{\Psi^{-1}(V_{t_1,\dots,t_n}(\mu, \delta_1))\} \\
&\quad= \lim_{\delta\to 0} \liminf_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{(\Xi_{\alpha,0,\nu}(A_1), \dots, \Xi_{\alpha,0,\nu}(A_{n+1})) \in V_{t_1,\dots,t_n}(\mu, \delta_1)\} \\
&\quad\geq -\mathcal{I}_{n+1}(\mu(A_1), \dots, \mu(A_{n+1})) \\
&\quad\geq -\mathcal{I}(\mu).
\end{aligned} \tag{20}$$

Next we assume that $t_1, \dots, t_n$ are continuity points of μ. We denote the collection of all partitions from these points by $\mathcal{J}_\mu$. This implies that $\Psi(\upsilon)$ is continuous at μ. Hence, for any $\delta_2 > 0$, we can choose $\delta > 0$ small enough such that

$$\bar{B}(\mu, \delta) \subset \Psi^{-1}(V_{t_1,\dots,t_k}(\mu, \delta_2)).$$

Let

$$\bar{V}_{t_1,\dots,t_k}(\mu, \delta_2) = \{(y_1, \dots, y_n) \in \triangle_n \colon |y_i - \mu(A_i)| \leq \delta_2,\ i = 1, \dots, n\}.$$

Then we have

$$\begin{aligned}
&\lim_{\delta\to 0} \limsup_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{\bar{B}(\mu, \delta)\} \\
&\quad\leq \limsup_{\alpha\to 1} \frac{1}{-\log(1-\alpha)} \log \mathbb{P}\{(\Xi_{\alpha,0,\nu}(A_1), \dots, \Xi_{\alpha,0,\nu}(A_{n+1})) \in \bar{V}_{t_1,\dots,t_n}(\mu, \delta_2)\}.
\end{aligned} \tag{21}$$

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Letting δ_2 go to 0 and applying Lemma 6 again, we obtain

$$\lim_{\delta\to 0}\limsup_{\theta\to\infty}\frac{1}{\theta}\log P\{\bar{B}(\mu,\delta)\}\le -\mathcal{I}_{n+1}(\mu(A_1),\dots,\mu(A_{n+1})).$$

Finally, taking supremum over $\mathcal{J}_\mu$ and applying Lemma 7, we obtain

$$\lim_{\delta\to 0}\limsup_{\alpha\to 1}\frac{1}{-\log(1-\alpha)}\log\mathbb{P}\{\bar{B}(\mu,\delta)\}\le -\mathcal{I}(\mu),$$

which combined with (20) leads to the theorem. □

5. Concluding remarks

A comparison between α converging to 1 and θ converging to ∞ reveals fundamental differences. Under these limiting procedures, we have both $\mathbb{P}(\alpha,0)$ and $\mathbb{P}(0,\theta)$ converging to $(0,0,\dots)$. This can be seen from the distributions of $\varphi_2(\mathbb{P}(\alpha,0))$ and $\varphi_2(\mathbb{P}(0,\theta))$.

It was shown in [22] and [25] that

$$\sqrt{\tfrac{1}{2}\theta}[\theta\varphi_2(\mathbb{P}(0,\theta))-1]\xrightarrow{\text{D}} Z\quad\text{as }\theta\to\infty,$$

where Z is the standard normal random variable. By the Ewens sampling formula, we have

$$\mathbb{E}[\varphi_2(\mathbb{P}(0,\theta))]=\frac{1}{\theta+1},\qquad \mathbb{E}[\varphi_2^2(\mathbb{P}(0,\theta))]=\frac{3!+\theta}{(\theta+1)(\theta+2)(\theta+3)},$$

and

$$\mathbb{E}[\varphi_2^3(\mathbb{P}(0,\theta))]=\frac{1}{(\theta+1)_{(5)}}(5!+3\cdot 3!\,\theta+\theta^2).$$

The skewness of $\varphi_2(\mathbb{P}(0,\theta))$ is given by

$$\frac{\mathbb{E}[\varphi_2^3(\mathbb{P}(0,\theta))]-3\mathbb{E}[\varphi_2(\mathbb{P}(0,\theta))]\mathbb{E}[\varphi_2^2(\mathbb{P}(0,\theta))]+2(\mathbb{E}[\varphi_2(\mathbb{P}(0,\theta))])^3}{(\mathbb{E}[\varphi_2^2(\mathbb{P}(0,\theta))]-(\mathbb{E}[\varphi_2(\mathbb{P}(0,\theta))])^2)^{3/2}}$$
$$=\frac{O(\theta^{-5})}{O(\theta^{-4.5})}\to 0\quad\text{as }\theta\to\infty,$$

which is consistent with the Gaussian limit.

On the other hand, for $\varphi_2(\mathbb{P}(\alpha,0))$, we have

$$\mathbb{E}[\varphi_2(\mathbb{P}(\alpha,0))]=1-\alpha,\qquad \mathbb{E}[\varphi_2^2(\mathbb{P}(\alpha,0))]=\tfrac{1}{6}[(1-\alpha)(2-\alpha)(3-\alpha)+\alpha(1-\alpha)^2],$$

and

$$\operatorname{var}(\varphi_2(\mathbb{P}(\alpha,0)))=\tfrac{1}{3}\alpha(1-\alpha),$$
$$\mathbb{E}[\varphi_2^3(\mathbb{P}(\alpha,0))]=\frac{1}{5!}[(1-\alpha)_{(5)}+3\alpha(1-\alpha)^2(2-\alpha)(3-\alpha)+\alpha^2(1-\alpha)^3].$$

This means that the skewness of $\varphi_2(\mathbb{P}(\alpha,0))$ is of order $O(1-\alpha)/O((1-\alpha)^{3/2})$ which goes to ∞ as α converges to 1. Thus, the distribution of $\varphi_2(\mathbb{P}(\alpha,0))$ is skewed strongly to the right and a Gaussian limit is unlikely. The cause for this is the fact that as α increases the frequencies become more even, and the tail becomes heavier.

https://doi.org/10.1017/apr.2017.13 Published online by Cambridge University Press

Another difference is reflected from the large deviation behaviour of the Pitman sampling formula. For any $n \geq 1$, a partition $\boldsymbol{\eta}$ of n with length l, the conditional Pitman sampling formula given $\mathbb{P}(\alpha, \theta) = \boldsymbol{p}$ is

$$F_{\boldsymbol{\eta}}(\boldsymbol{p}) = C(n, \boldsymbol{\eta}) \sum_{\text{distinct } i_1, \dots, i_l} p_{i_1}^{\eta_1} \dots p_{i_l}^{\eta_l},$$

where

$$C(n, \boldsymbol{\eta}) = \frac{n!}{\prod_{k=1}^{l} \eta_k! \prod_{j=1}^{n} a_j(\boldsymbol{\eta})}.$$

Assuming $\eta_i \geq 2$ for all i then $F_{\boldsymbol{\eta}}(\boldsymbol{p})$ is continuous on ∇_∞. By the contraction principle, large deviation principles hold for the image laws of $\mathrm{PD}(0, \theta)$ and $\mathrm{PD}(\alpha, 0)$ under $F_{\boldsymbol{\eta}}(\boldsymbol{p})$ with respect to speed θ and $-\log(1-\alpha)$.

Integrating $F_{\boldsymbol{\eta}}(\boldsymbol{p})$ with respect to $\mathrm{PD}(\alpha, \theta)$ leads to the unconditional Pitman sampling formula. The large deviation speed was shown in [14] to be $\log \theta$ under $\mathrm{PD}(0, \theta)$. In [19], the large deviation speed under $\mathrm{PD}(\alpha, 0)$ was shown to be $-\log(1-\alpha)$. In other words, under $\mathrm{PD}(0, \theta)$ the conditional and unconditional Pitman sampling formulae have different large deviation speeds due to averaging and finite sample size, while under $\mathrm{PD}(\alpha, 0)$ the corresponding speeds are the same.

In a Bayesian context, the limiting procedure $\theta \to \infty$ corresponds to the sample size tending to ∞ and the large deviations for $\Xi_{0,\theta,\nu}$ become the large deviations for the posteriors of Dirichlet prior to when the sample size converges to ∞ (see [21]). It is not clear what the natural Bayesian counterpart is for the limiting procedure $\alpha \to 1$.

Acknowledgements

We wish to thank the two anonymous referees for their insightful comments and suggestions which substantially improved the paper. S. Feng is supported by the Natural Sciences and Engineering Research Council of Canada, F. Gao is supported by the National Natural Science Foundation of China (grant number 11571262) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (grant number 20130141110076), and Y. Zhou is supported by the Fundamental Research Grant of Zhongnan University of Economics and Law (grant number 21141611208).

References

[1] Athreya, K. B. (2012). Coalescence in the recent past in a rapidly growing population. *Stoch. Process. Appl.* **122,** 3757–3766.
[2] Bertoin, J. (1996). *Lévy Processes* (Camb. Tracts Math. **121**). Cambridge University Press.
[3] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes* (Camb. Stud. Adv. Math. **102**). Cambridge University Press.
[4] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197,** 247–276.
[5] Bovier, A. (2006). *Statistical Mechanics of Disordered Systems* (Camb. Ser. Statist. Prob. Math. **18**). Cambridge University Press.
[6] Dawson, D. A. and Feng, S. (2001). Large deviations for the Fleming–Viot process with neutral mutation and selection. II. *Stoch. Process. Appl.* **92,** 131–162.
[7] Dawson, D. A. and Feng, S. (2006). Asymptotic behavior of Poisson–Dirichlet distribution for large mutation rate. *Ann. Appl. Prob.* **16,** 562–582.
[8] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications* (Appl. Math. **38**), 2nd edn. Springer, New York.
[9] Derrida, B. (1980). Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.* **45,** 79–82.

[10] Derrida, B. (1981). Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B* **24,** 2613–2626.
[11] Derrida, B. (1985). A generalization of the random energy model which includes correlations between the energies. *J. Phys. Lett.* **46,** 401–407.
[12] Derrida, B. and Gardner, E. (1986). Solution of the generalized random energy model. *J. Phys. C* **19,** 2253–2274.
[13] Eichelsbacher, P. and Ganesh, A. (2002). Moderate deviations for Bayes posteriors. *Scand. J. Statist.* **29,** 153–167.
[14] Feng, S. (2007). Large deviations for Dirichlet processes and Poisson–Dirichlet distribution with two parameters. *Electron. J. Prob.* **12,** 787–807.
[15] Feng, S. (2009). Poisson–Dirichlet distribution with small mutation rate. *Stoch. Process. Appl.* **119,** 2082–2094.
[16] Feng, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics*. Springer, Heidelberg.
[17] Feng, S. and Gao, F. (2008). Moderate deviations for Poisson–Dirichlet distribution. *Ann. Appl. Prob.* **18,** 1794–1824.
[18] Feng, S. and Gao, F. (2010). Asymptotic results for the two-parameter Poisson–Dirichlet distribution. *Stoch. Process. Appl.*, **120,** 1159–1177.
[19] Feng, S. and Zhou, Y. (2015). Asymptotic behaviour of Poisson–Dirichlet distribution and random energy model. In *XI Symp. on Probability and Stochastic Processes* (Progress in Probability **69**), eds R. H. Mena *et al.*, Birkhäuser, Cham, pp. 141–155.
[20] Ferguson, T. S. (1973). A Baysian analysis of some nonparametric problems. *Ann. Statist.* **1,** 209–230.
[21] Ganesh, A. J. and O'Connell, N. (2000). A large-deviation principle for Dirichlet posteriors. *Bernoulli* **6,** 1021–1034.
[22] Griffiths, R. C. (1979). On the distribution of allele frequencies in a diffusion model. *Theoret. Pop. Biol.* **15,** 140–158.
[23] Handa, K. (2009). The two-parameter Poisson–Dirichlet point process. *Bernoulli* **15,** 1082–1116.
[24] James, L. F. (2008). Large sample asymptotics for the two-parameter Poisson–Dirichlet process. In *Pushing the Limits of Contemporary Statistics* (Inst. Math. Statist. Collect. **3**), eds B. Clarke and S. Ghosal, Institute of Mathematical Statistics, Beachwood, OH, pp. 187–199.
[25] 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.
[26] Kanter, M. (1975). Stable densities under change of scale and total variation inequalities. *Ann. Prob.* **3,** 697–707.
[27] Kingman, J. C. F. *et al.* (1975). Random discrete distribution. *J. R. Statist. Soc. B* **37,** 1–22.
[28] Lynch, J. and Sethuraman, J. (1987). Large deviations for processes with independent increments. *Ann. Prob.* **15,** 610–627.
[29] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Prob. Theory Relat. Fields* **92,** 21–39.
[30] Pitman, J. (1992). The two-parameter generalization of Ewens' random partition structure. Tech. Rep. 345, University of California, Berkeley.
[31] Pitman, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory*, eds T. S. Ferguson, L. S. Shapley and J. B. MacQueen, Institute of Mathematical Statistics, Hayward, CA, pp. 245–267.
[32] Pitman, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**), Springer, Berlin.
[33] Pitman, J. and Yor, M. (1992). Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc.* **65,** 326–356.
[34] Pitman, J. and Yor, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25,** 855–900.
[35] Pollard, H. (1946). The representation of $\mathrm{e}^{-x^{\lambda}}$ as a Laplace integral. *Bull. Amer. Math. Soc.* **52,** 908–910.
[36] Puhalskii, A. A. (1991). On functional principle of large deviations. In *New Trends in Probability and Statistics*, Vol. 1, eds V. Sazonov and T. Shervashidze, VSP, Utrecht, pp. 198–218.
[37] Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians* (Res. Math. Relat. Areas **3**; Ser. Modern Surveys Math. **46**). Springer, Berlin.