LONG CYCLE OF RANDOM PERMUTATIONS WITH POLYNOMIALLY GROWING CYCLE WEIGHTS

DIRK ZEINDLER

ABSTRACT. We study random permutations of n objects with respect to multiplicative measures with polynomial growing cycle weights. We determine in this paper the asymptotic behaviour of the long cycles under this measure and also prove that the cumulative cycle numbers converge in the region of the long cycles to a Poisson process.

1. INTRODUCTION

Let $\mathfrak{S}_n$ be the symmetric group of all permutations on elements $\{1, \ldots, n\}$. For any permutation $\sigma \in \mathfrak{S}_n$, denote by $C_m = C_m(\sigma)$ the *cycle counts*, that is, the number of cycles of length $m = 1, \ldots, n$ in the cycle decomposition of σ; clearly

$$C_m \geq 0 \quad (m \geq 1), \qquad \sum_{m=1}^{n} m\, C_m = n. \tag{1.1}$$

1.1. **Random permutations.** Classical probability measures studied on $\mathfrak{S}_n$ are the uniform measure and the Ewens measure. The uniform measure is well studied and has a long history (see e.g. the first chapter of [1] for a detailed account with references). The Ewens measure originally appeared in population genetics, see [11], but has also various applications through its connection with Kingman's coalescent process, see [13]. Classical results about uniform and Ewens random permutations include convergence of joint cycle counts towards independent Poisson random variables in total variation distance [2] and a central limit theorem for cumulative cycle counts [7]. Futhermore, the longest cycles have order of magnitude n and it was established by Kingman ([14]) and by Vershik and Shimdt ([20]) that the vector of renormalized and ordered length of the cycles converges in law to a Poisson-Dirichlet distribution.

In this paper, we study random permutations with respect to the probability measure

Definition 1.1. *Let $\Theta = (\theta_k)_{k \geq 1}$ be given, with $\theta_k \geq 0$ for every $k \geq 1$. We define for $\sigma \in \mathfrak{S}_n$ the weighted measures on $\mathfrak{S}_n$ as*

$$\mathbb{P}_\Theta [\sigma] := \frac{1}{h_n n!} \prod_{k=1}^{n} \theta_k^{C_k} \tag{1.2}$$

with $h_n = h_n(\Theta)$ a normalization constant and $h_0 := 1$.

The measure $\mathbb{P}_\Theta$ has received a lot of attention in recent years and has been studied by many authors. The uniform measure and the Ewens measure are special cases of $\mathbb{P}_\Theta$ and correspond to $\theta_k \equiv 1$ and $\theta_k \equiv \vartheta$ with $\vartheta > 0$. Another well studied case are the weights $\theta_k \sim k^\alpha$, see

[10, 6]. Further studied weights are for instance $\theta_k = \log^m(k)$ and $\theta_k = \vartheta \, \mathbb{1}_{\{k \leq n^\beta\}}$, see [19] and [3]. An overview can be found in [9]. The motivation to study the measure $\mathbb{P}_\Theta$ has its origins in mathematical physics. Explicitly, it occurred in the context of the Feynman-Kac representation of the dilute Bose gas and it has been proposed in connection with the study of the Bose-Einstein condensation (see e.g. [4] and [9]).

We consider in this paper the case $\theta_k \sim k^\alpha$ with $\alpha > 0$. In this situation, a typical cycle has a length of order $n^{\frac{1}{1+\alpha}}$ and converges to a Gamma distribution after suitable normalisation. This was established by Ercolani and Ueltschi in [9] (among many further results). Dereich and Mörters extended this in [8] to a local limit theorem. Further, it was shown in [9] that the expectation of total number of cycles is $\approx n^{\frac{\alpha}{1+\alpha}}$, which was extended by Maples, Nikeghbali and the author in [17] to a central limit theorem. Furthermore, Ercolani and Ueltschi proved in [9] that the component process of the cycle counts converges in distribution to mutually *independent* Poisson random variables Y_m:

$$(C_1, C_2, \ldots, C_b) \overset{d}{\longrightarrow} (Y_1, Y_2, \ldots, Y_b), \qquad \text{as } n \to \infty, \tag{1.3}$$

where $b \in \mathbb{N}$ is fix and $(Y_k)_{k=1}^b$ are independent Poisson distributed random variables with $\mathbb{E}[Y_k] = \frac{\theta_k}{k}$. Storm and the author computed in [21] the total variation distance between the processes in (1.3) and have shown that this total variation distance tends to 0 if and only if $b = o(n^{\frac{1}{1+\alpha}})$. Further, it was shown in [6, 9] that the cumulative cycle counts

$$w_n(x) := \sum_{k \geq x n^{\frac{1}{1+\alpha}}} C_k \tag{1.4}$$

converges to a limit shape after suitable normalisation. In view of these results it is clear that one cannot expect cycles of order n as $n \to \infty$. The exact behavior of the long cycles in the case $\theta_k \sim k^\alpha$ was unknown for a long time and is the main topic of this paper.

1.2. **Main results.**
In this paper, we have two main results. In order to state these results, we have to introduce some notations. Let v_n be the solution of the equation

$$\sum_{k=1}^\infty \theta_k e^{-k v_n} = n. \tag{1.5}$$

Since $\theta_k \sim k^\alpha$ and $\alpha > 0$, we immediately see that the sum in (1.5) is monotone increasing and tending to ∞ as $v_n \searrow 0$. Thus the solution of (1.5) exists and is also uniquely determined. It follows from Lemma 2.5 that

$$v_n \sim \left(\frac{n}{\Gamma(\alpha+1)} \right)^{-\frac{1}{1+\alpha}}, \tag{1.6}$$

where we write $a_n \sim b_n$ if $\lim_{n \to \infty} a_n/b_n = 1$. Furthermore, we set

$$n^* := (v_n)^{-1} \quad \text{and} \quad \ell_n := \alpha \log(n^*) + (\alpha - 1) \log(\alpha \log(n^*)). \tag{1.7}$$

We denote by $L_k = L_k(\sigma)$ the length of the k-th longest cycle in the cycle decomposition of the permutation $\sigma \in \mathfrak{S}_n$, counted with multiplicity. The first main result of this paper is

Theorem 1.2. *Let $K \in \mathbb{N}$ be given. We have convergence in distribution of*

$$\frac{1}{n^*} \cdot (L_1 - n^* \ell_n, \ldots, L_K - n^* \ell_n) \overset{d}{\longrightarrow} \left(-\log(E_1), \ldots, -\log\left(\sum_{j=1}^K E_j \right) \right).$$

under $\mathbb{P}_\Theta$ as $n \to \infty$, where $(E_j)_{j=1}^K$ is a sequence of iid Exp(1) distributed random variables.

This theorem follows almost immediately from the second main result.

Theorem 1.3. *Define for $y \geq 0$*

$$P_y^{(n)} = \sum_{k \geq x_n}^{\infty} C_k \ \text{ with } \ x_n := n^*\big(\ell_n + \min\{-\log(y), \ell_n\}\big). \tag{1.8}$$

Then the stochastic process $\{P_y^{(n)}, y \geq 0\}$ converges under $\mathbb{P}_\Theta$ as $n \to \infty$ weakly in $\mathcal{D}\,[0, \infty)$ to a Poisson process with parameter 1, where $\mathcal{D}\,[0, \infty)$ denotes the space of càdlàg-functions endowed with Skorohod topology.

We prove this two theorems by computing the generating functions for the relevant quantities and then extracting their asymptotic behaviour with the saddle point method. This paper is structured as follows: In Section 2, we introduce the necessary preliminaries and then give in Section 3 the proof of Theorem 1.2 and of Theorem 1.3.

1.3. **Notation.** Apart from the notation $a_n \sim b_n$ above, we write $a_n \approx b_n$ for two real sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ if there exist constants $c_1, c_2 > 0$ such that

$$c_1 b_n \leq a_n \leq c_2 b_n \tag{1.9}$$

for large n. We also use the usual O and o notation, i.e. $f(n) = O(g(n))$ means that there exists some constant $c > 0$ so that $|f(n)| \leq c|g(n)|$ for large n, while $f(n) = o(g(n))$ means that for all $c > 0$ there exists $n_c \in \mathbb{N}$ so that the inequality $|f(n)| \leq c|g(n)|$ holds for all $n > n_c$.

2. PRELIMINARIES

We introduce in this section some notation and preliminary results.

2.1. **Generating functions.** The (ordinary) generating function of a sequence $(g_k)_{k \geq 0}$ of complex numbers is defined as the formal power series

$$g(t) := \sum_{j=0}^{\infty} g_k t^k. \tag{2.1}$$

As usual, we define the *extraction symbol* $[t^k]\, g(t) := g_k$, that is, as the coefficient of t^k in the power series expansion (2.1) of $g(t)$.

A generating function that plays an important role in this paper is

$$g_\Theta(t) := \sum_{k \geq 1} \frac{\theta_k}{k} t^k. \tag{2.2}$$

The following well-known identity is a special case of the general *Pólya's Enumeration Theorem* [18, p. 17] and is the main tool in this paper to obtain generating functions.

Lemma 2.1. *Let $(a_m)_{m \in \mathbb{N}}$ be a sequence of complex numbers. We then have as formal power series in t*

$$\sum_{n \in \mathbb{N}} \frac{t^n}{n!} \sum_{\sigma \in \mathfrak{S}_n} \prod_{j=1}^{n} a_j^{C_j} = \exp\left(\sum_{k \geq 1} \frac{a_k}{k} t^k\right). \tag{2.3}$$

If one series in (2.3) *converges absolutely, so do the other.*

We omit the proof of this lemma, but details can be found for instance in [16, p. 5]. Applying this equation to the normalisation constant h_n in Definition 1.1, we immediately get

$$\sum_{n=0}^{\infty} h_n t^n = \exp\left(\sum_{k=1}^{\infty} \frac{\theta_k}{k} t^k\right) = \exp\left(g_\Theta(t)\right), \tag{2.4}$$

with the convention $h_0 := 1$. Equation (2.3) can be reformulated as

$$\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \prod_{j=1}^{n} a_j^{C_j} = [t^n] \exp\left(\sum_{k=1}^{\infty} \frac{a_k}{k} t^k\right). \tag{2.5}$$

With this formulation, the parameters $(a_j)_{j\in\mathbb{N}}$ can depend on the system size n. In particular, we have for all $s \in \mathbb{R}$ and all integer sequences $(x_n)_{n\in\mathbb{N}}$

$$h_n \mathbb{E}\left[e^{s\sum_{k\geq x_n}^{\infty} C_k}\right] = [t^n] \exp\left((e^s - 1)\sum_{k\geq x_n}^{\infty} \frac{\theta_k}{k} t^k + g_\Theta(t)\right). \tag{2.6}$$

2.2. **Saddle point method.** The formulation in (2.6) has now the advantage that we can compute the expectation on the LHS of (2.6) by extracting series coefficients from the RHS of (2.6) with complex analysis. A way to extract these series coefficients is the saddle point method, a standard tool in asymptotic analysis. The basic idea is to rewrite a expression like (2.6) as a complex contour integral and choose the path of integration in a convenient way. The details of this procedure depend on the situation at hand and need to be done on a case by case basis. A general overview over the saddle-point method can be found in [12, page 551]. To apply the saddle point method in this paper, we introduce the following definition.

Definition 2.2. *Let* $\left(g_n(t)\right)_{n\in\mathbb{N}}$ *with* $g_n(t) = \sum_{k=0}^{\infty} g_{k,n} t^k$ *be given with radius of convergence* $\rho > 0$ *and* $g_{k,n} \geq 0$. *We say that* $\left(g_n(t)\right)_{n\in\mathbb{N}}$ *is* log-n-admissible *if there exist functions* $a_n, b_n :$ $[0, \rho) \to \mathbb{R}^+$, $R_n : [0, \rho) \times (-\pi/2, \pi/2) \to \mathbb{R}^+$ *and sequences* $(\delta_n)_{n\in\mathbb{N}}$ *and* $(r_n)_{n\in\mathbb{N}}$ *s.t.*

Saddle-point: *For each* n, *we have* $r_n \in [0, \rho)$ *and*

$$a_n(r_n) = n + o\left(\sqrt{b_n(r_n)}\right). \tag{2.7}$$

Divergence: $b_n(r_n) \to \infty$ *and* $\delta_n \to 0$ *as* $n \to \infty$.
Approximation: *For all* $|\varphi| \leq \delta_n$ *we have the expansion*

$$g_n(r_n e^{i\varphi}) = g_n(r_n) + i\varphi a_n(r_n) - \frac{\varphi^2}{2} b_n(r_n) + R_n(r_n, \varphi) \tag{2.8}$$

where $R_n(r_n, \varphi) = o(\varphi^3 \delta_n^{-3})$.
Width of convergence: *We have* $\delta_n^2 b_n(r_n) - \log b_n(r_n) \to +\infty$ *as* $n \to +\infty$.
Monotonicity: *For all* $|\varphi| > \delta_n$, *we have*

$$\text{Re}\left(g_n(r_n e^{i\varphi})\right) \leq \text{Re}\left(g(r_n e^{\pm i\delta_n})\right). \tag{2.9}$$

The approximation condition allows us to compute the functions a_n and b_n exactly. We have

$$a_n(r) = r g_n'(r), \tag{2.10}$$
$$b_n(r) = r g_n'(r) + r^2 g_n''(r). \tag{2.11}$$

Clearly a_n and b_n are strictly increasing real analytic functions in $[0, \rho)$. The error in the approximation can similarly be bounded, so that

$$R_n(r, \varphi) = \varphi^3 O \left(a_n(r) + b_n(r) + r^3 g_n'''(r) \right). \tag{2.12}$$

In view of (2.6), we will use in Section 3

$$g_n(t) = \sum_{k \geq x_n}^{\infty} \frac{\theta_k}{k} t^k + g_\Theta(t) \tag{2.13}$$

with x_n as in Theorem 1.3 and show that this sequence is log-n-admissible with $r_n = e^{-v_n}$.

Further, we need

Theorem 2.3. *Let* $\left(g_n(t) \right)_{n \in \mathbb{N}}$ *be* log-n-*admissible with associated functions* a_n, b_n *and constants* r_n. *Call*

$$G_n := [t^n] \exp \left(g_n(t) \right).$$

Then G_n *has the asymptotic expansion*

$$G_n = \frac{1}{\sqrt{2\pi}} (r_n)^{-n} b_n(r_n)^{-1/2} \exp \left(g_n(r_n) \right) (1 + o(1)). \tag{2.14}$$

The proof of this theorem can be found in [6]. Further

Corollary 2.4. *Let* $\left(g_n(t) \right)_{n \in \mathbb{N}}$ *be* log-n-*admissible with associated functions* a_n, b_n *and constants* r_n. *Let further* $\left(f_n(t) \right)_{n \in \mathbb{N}}$ *with* $f_n(t) = \sum_{k=0}^{\infty} f_{k,n} t^k$ *be given with radius of convergence* $\geq \rho$ *and* $f_{k,n} \geq 0$. *Then there exists* n_0 *only depending on* $\left(g_n(t) \right)_{n \in \mathbb{N}}$ *such that*

$$[t^n] f_n(t) \exp \left(g_n(t) \right) \leq 2 f_n(r_n) [t^n] \exp \left(g_n(t) \right) \text{ for all } n \geq n_0. \tag{2.15}$$

Proof. Since $f_{k,n} \geq 0$ and $g_{k,n} \geq 0$, it follows immediately that $[t^n] f_n(t) \exp \left(g_n(t) \right) \geq 0$. We get with Cauchy's intergral formula and the curve $\gamma(\varphi) = r_n e^{i\varphi}$ that

$$[t^n] f_n(t) \exp \left(g_n(t) \right) = \frac{1}{2\pi i (r_n)^n} \int_{-\pi}^{\pi} f_n(r_n e^{i\varphi}) \exp \left(g_n(r_n e^{i\varphi}) - in\varphi \right) d\varphi$$

$$\leq \frac{1}{2\pi (r_n)^n} \int_{-\pi}^{\pi} \left| f_n(r_n e^{i\varphi}) \exp \left(g_n(r_n e^{i\varphi}) - in\varphi \right) \right| d\varphi$$

$$\leq \frac{f_n(r_n)}{2\pi (r_n)^n} \int_{-\pi}^{\pi} \exp \left(\mathrm{Re}(g_n(r_n e^{i\varphi})) \right) d\varphi. \tag{2.16}$$

We now can compute the last integral as in the proof of Theorem 2.3. These computations are (almost) identical and we thus omit them. We then get

$$\int_{-\pi}^{\pi} \exp \left(\mathrm{Re}(g_n(r_n e^{i\varphi})) \right) d\varphi = \sqrt{2\pi} b_n(r_n)^{-1/2} \exp \left(g_n(r_n) \right) (1 + o(1)). \tag{2.17}$$

Note that the $(1 + o(1))$ depends on $\left(g_n(t) \right)_{n \in \mathbb{N}}$, but not on $\left(f_n(t) \right)_{n \in \mathbb{N}}$. Combining (2.16) and (2.17) with Theorem 2.3 then completes the proof. $\qquad \square$

2.3. **Approximation of Sums.** We require for our argumentation the asymptotic behaviour of the generating function $g_\Theta(t)$ as t tends to the radius of convergence, which is 1 in our case.

Lemma 2.5. *Let $(v_n)_{n\in\mathbb{N}}$ a sequence of positive numbers with $v_n \downarrow 0$ as $n \to +\infty$. We have for all $\delta \in \mathbb{R} \setminus \{-1, -2, -3, \dots\}$*

$$\sum_{k=1}^{\infty} k^\delta \exp(-kv_n) = \Gamma(\delta+1)v_n^{-\delta-1} + \zeta(-\delta) + O(v_n). \qquad (2.18)$$

$\zeta(\cdot)$ *indicates the Riemann Zeta function.*

This lemma can be proven with Euler Maclaurin summation formula or with the Mellin transformation. These computations are straightforward and the details of the proof with the Mellin transformation can be found for instance in [12, Chapter VI.8]. We thus omit the proof of this lemma.

We require also the behaviour of partial sum $\sum_{k\geq x_n}^{\infty} k^\delta \exp(-kv_n)$ as $x_n \to \infty$ and as $v_n \to 0$. We have

Proposition 2.6. *Let $\delta \in \mathbb{R}$ be given. Let further $(v_n)_{n\in\mathbb{N}}$ and $(x_n)_{n\in\mathbb{N}}$ be sequences with $v_n > 0$, $v_n \to 0$ and $x_n v_n \to \infty$. We then have*

$$\sum_{k=x_n}^{\infty} k^\delta e^{-kv_n} = \int_{x_n}^{\infty} x^\delta e^{-xv_n} dx + x_n^\delta e^{-x_n v_n} \left(\sum_{k=0}^{N} Q_k(1/x_n, v_n) + O\left(v_n^{-N-1}\right) \right) \qquad (2.19)$$

where $Q_k(\cdot, \cdot)$ are a homogeneous polynomials of degree k with $Q_0 = 1$. Furthermore

$$\int_{x_n}^{\infty} x^\delta \exp(-xv_n)dx = \frac{x_n^\delta e^{-x_n v_n}}{v_n} \left(\sum_{j=0}^{N} \frac{(\delta)_j}{(x_n v_n)^j} + O(x_n v_n)^{-N-1} \right), \qquad (2.20)$$

where $(\delta)_0 = 1$ and $(\delta)_j = \delta(\delta-1)\cdots(\delta-j+1)$ for $j \geq 1$.

Proof. We first proof (2.20). We get with N times partial integration

$$\int_{x_n}^{\infty} x^\delta \exp(-xv_n)dx = \sum_{j=0}^{N} (\delta)_j x_n^{\delta-j} v_n^{-j-1} e^{-x_n v_n} + (\delta)_{N+1} v_n^{-N-2} \int_{x_n}^{\infty} x^{\delta-N-1} e^{-xv_n} dx$$

$$= \frac{x_n^\delta e^{-x_n v_n}}{v_n} \sum_{j=0}^{N} \frac{(\delta)_j}{(x_n v_n)^j} + O\left(v_n^{-N-2} \int_{x_n}^{\infty} x^{\delta-N-1} e^{-xv_n} dx \right). \qquad (2.21)$$

We now can assume that $N > \delta$ and thus

$$\int_{x_n}^{\infty} x^{\delta-N-1} e^{-xv_n} dx \leq x_n^{\delta-N-1} \int_{x_n}^{\infty} e^{-xv_n} dx = \frac{x_n^{\delta-N-1}}{v_n} e^{-x_n v_n}. \qquad (2.22)$$

Inserting this into (2.21) completes the proof of (2.20). For the proof of (2.19), we use the Euler-Maclaurin summation formula and obtain

$$\sum_{k=x_n}^{\infty} k^\delta e^{-kv_n} = \int_{x_n}^{\infty} x^\delta e^{-xv_n} dx + \sum_{k=0}^{N+1} \frac{B_{k+1}}{(k+1)!} f^{(k)}(x_n) + O\left(\int_{x_n}^{\infty} |f^{(N+2)}(x)|dx \right), \quad (2.23)$$

where B_k is the kth Bernoulli number and $f(x) = x^\delta e^{-xv_n}$ and $f^{(k)}$ the kth derivative of f. A straight forward computation shows that

$$f^{(k)}(x) = P_k(1/x, v_n)f(x), \qquad (2.24)$$

where $P_k(\cdot, \cdot)$ is a homogeneous polynomial of degree k. Thus we have

$$\sum_{k=x_n}^{\infty} k^{\delta} e^{-kv_n} = \int_{x_n}^{\infty} x^{\delta} e^{-xv_n} dx + x_n^{\delta} e^{-x_n v_n} \left(\sum_{k=0}^{N+1} Q_k(1/x_n, v_n) \right) + O \left(\int_{x_n}^{\infty} |f^{(N+2)}(x)| dx \right),$$

with $Q_k(\cdot, \cdot) = \frac{B_{k+1}}{(k+1)!} P_k(\cdot, \cdot)$. Since $x_n v_n \to \infty$, we have that $1/x_n = o(v_n)$ and thus

$$\int_{x_n}^{\infty} |f^{(N+2)}(x)| dx = \int_{x_n}^{\infty} |P_{N+2}(1/x, v_n)| f(x) dx = O \left(v_n^{N+2} \int_{x_n}^{\infty} f(x) dx \right). \qquad (2.25)$$

Using (2.21) and that $1/x_n = o(v_n)$ completes the proof. $\qquad \square$

3. Proof of the main results

We give in this section the proofs of the Theorems 1.2 and 1.3. We begin with the proof of Theorem 1.3. We proceed in two steps. In Section 3.1, we show that the finite dimensional distributions of the process $P_y^{(n)}$ converges to the finite dimensional distributions of a Poisson process. In Section 3.2, we show that the sequence $(P^{(n)})_{n \in \mathbb{N}}$ is tight. This then completes the proof of Theorem 1.3. Finally, we use in Section 3.3 the Theorem 1.3 to prove Theorem 1.2.

3.1. Finite dimensional distributions.
We first have to show for all $0 \leq y_1 < y_2 < \cdots < y_K$ that we have

$$\left(P_{y_1}^{(n)}, P_{y_2}^{(n)} - P_{y_1}^{(n)}, P_{y_3}^{(n)} - P_{y_2}^{(n)}, \ldots, P_{y_K}^{(n)} - P_{y_{K-1}}^{(n)} \right) \xrightarrow{d} (Y_1, \ldots, Y_K), \qquad (3.1)$$

where $(Y_k)_{k=1}^{K}$ is a sequence of independent, Poisson distributed random variables such that $\mathbb{E}[Y_k] = y_k - y_{k-1}$ for all $1 \leq k \leq K$.

We begin with the proof of (3.1) for the case $K = 1$. Thus we have to determine the asymptotic behaviour of $P_y^{(n)}$ for $y \geq 0$ fix. We do this by determining the asymptotic behaviour of the moment generating function of $P_y^{(n)}$. Recall,

$$P_y^{(n)} = \sum_{k \geq x_n}^{\infty} C_k \quad \text{with} \quad x_n := n^* \left(\ell_n + \min\{-\log(y), \ell_n\} \right),$$

where

$$n^* = (v_n)^{-1} \quad \text{and} \quad \ell_n = \alpha \log(n^*) + (\alpha - 1) \log \left(\alpha \log(n^*) \right) \qquad (3.2)$$

and v_n is the solution of the equation $\sum_{k=1}^{\infty} \theta_k e^{-kv_n} = n$. We now have to distinguish the two cases $y > 0$ and $y = 0$. We begin with the case $y > 0$. In this case we have for n large enough

$$x_n = n^* \left(\ell_n - \log(y) \right). \qquad (3.3)$$

Using (2.6), we obtain for $s \in \mathbb{R}$.

$$h_n \mathbb{E} \left[e^{s P_y^{(n)}} \right] = [t^n] \exp \left((e^s - 1) \sum_{k \geq x_n} \frac{\theta_k}{k} t^k + g_{\Theta}(t) \right) =: [t^n] \exp \left(g_{n,s}(t) \right) \qquad (3.4)$$

with h_n as in Definition 1.1. Note that the dependence on y of the RHS of (3.4) is hidden in x_n. We now have

Lemma 3.1. *Let $g_{n,s}(t)$ be as in (3.4). Then sequence $(g_{n,s})_{n\in\mathbb{N}}$ is $\log$-n-admissible for all $s \in \mathbb{R}$. Further, the*

$$b_n(r_n) \sim \Gamma(\alpha + 2)(n^*)^{\alpha+2}, r_n = e^{-v_n} \text{ and } \delta_n = (v_n)^\xi, \tag{3.5}$$

where ξ is any real number such that $\frac{\alpha+3}{3} < \xi < \frac{\alpha+2}{2}$.

This lemma shows that the sum depending on s and y on the right hand side of (3.4) has only a minor influence on the saddle point as r_n and δ_n can be chosen independent of y and s.

Proof. We begin with the case $\theta_k = k^\alpha$. We use (2.10) and get with Lemma 2.5, Proposition 2.6 and the definition of v_n that

$$a_n(r_n) = (e^s - 1)\sum_{k\geq x_n}^{\infty} k^\alpha e^{-kv_n} + \sum_{k=1}^{\infty} k^\alpha e^{-kv_n} = n + (e^s - 1)\frac{x_n^\alpha e^{-x_n v_n}}{v_n}(1 + o(1))$$

$$= n + (e^s - 1)(n^*)^\alpha(\ell_n - \log(y))^\alpha \cdot yn^* e^{-\ell_n}(1 + o(1))$$

$$= n + \alpha y(e^s - 1)n^* \log(n^*)(1 + o(1)).$$

Similarly, we get

$$b_n(r_n) = \Gamma(\alpha + 2)(n^*)^{\alpha+2}(1 + o(1)),$$

$$R_n(r, \varphi) = \varphi^3 O\left((n^*)^{\alpha+3}\right).$$

It is straight forward to see that replacing $\theta_k = k^\alpha$ by $\theta_k \sim k^\alpha$ in the above computations has only an influence to the $o(1)$ terms. For the $\log$-n-admissibility, we have now to check five conditions. For these, we need also that $v_n \sim n^{-\frac{1}{1+\alpha}}(\Gamma(\alpha+1))^{\frac{1}{1+\alpha}}$.

Divergence: We clearly have $\delta_n \to 0$ and $b_n(r_n) \to \infty$. Thus this condition is fulfilled.

Saddle-point: We require $a_n(r_n) = n + o\left(\sqrt{b_n(r_n)}\right)$. This condition is clearly fulfilled.

Approximation: We need that $R_n(r, \varphi) = o(\varphi^3 \delta_n^{-3})$. By the definition of δ_n and ξ, we have $\delta_n^{-3} = (n^*)^{3\xi}$ and $3\xi > \alpha + 3$. Thus $(n^*)^{\alpha+3} = o(\delta_n^{-3})$. Thus this condition is fulfilled.

Width of convergence: We have $\delta_n^2 b_n(r_n) \sim \Gamma(\alpha+2)(n^*)^{\alpha+2-2\xi}$. Since $\alpha + 2 - 2\xi > 0$, this condition is also fulfilled.

Monotonicity: The computations for this point are a little bit more involved, but are almost the same as in [6, Page 25] and we thus omit it.

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 3.1 shows that we can apply Theorem 2.3 to (3.4) and thus compute the asymptotic behaviour of $\mathbb{E}\left[e^{sP_y^{(n)}}\right]$. Furthermore, using $s = 0$ in (3.4), we see that $g_{n,0}(t) = g_\Theta(t)$. Thus we can use Lemma 3.1 to compute the asymptotic behaviour of h_n. Note that r_n and the leading term of $b_n(r_n)$ do not depend on s. This implies together with Theorem 2.3 that

$$\mathbb{E}\left[e^{sP_y^{(n)}}\right] = \exp\left((e^s - 1)\sum_{k\geq x_n}\frac{\theta_k}{k}(r_n)^k\right)(1 + o(1)). \tag{3.6}$$

Using again Proposition 2.6

$$\sum_{k\geq x_n} k^{\alpha-1}e^{-x_n v_n} \sim \frac{x_n^{\alpha-1}}{v_n}e^{-x_n v_n} \sim (n^*)^\alpha(\alpha\log(n^*))^{\alpha-1} \cdot \frac{y(n^*)^{-\alpha}}{(\alpha\log(n^*))^{\alpha-1}} = y. \tag{3.7}$$

We have by assumption $\theta_k \sim k^\alpha$. Thus there exists for all $\epsilon > 0$ a $k_0 = k_0(\epsilon)$ such that $(1 - \epsilon)k^\alpha \leq \theta_k \leq (1 + \epsilon)k^\alpha$ for all $k \geq k_0$. By definition, we have $x_n \to \infty$ and therefore we immediately get also

$$\sum_{k \geq x_n} \frac{\theta_k}{k} e^{-x_n v_n} \sim y. \tag{3.8}$$

Since $s \in \mathbb{R}$ is fix, this implies that $\mathbb{E}\left[e^{sP_y^{(n)}}\right] \to \exp\left((e^s - 1)y\right)$. This completes the proof for the case $y > 0$. For $y = 0$, we have $x_n = 2n^* \ell_n$ and get as in (3.7)

$$\sum_{k \geq x_n} k^{\alpha - 1} e^{-x_n v_n} \sim \frac{(n^*)^\alpha (2\alpha \log(n^*))^{\alpha - 1} \cdot (n^*)^{-2\alpha}}{(\alpha \log(n^*))^{2(\alpha - 1)}} = O\left((n^*)^{-\alpha}(\log(n^*))^{1-\alpha}\right). \tag{3.9}$$

The remaining computations are the same as for $y > 0$ and thus we get $P_0^{(n)} \xrightarrow{d} 0$. This completes the proof of (3.1) for the case $K = 1$. For the general case, define

$$x_{n,j} := n^*(\ell_n + \min\{-\log(y_j), \ell_n\}).$$

We then have for $s_1, \ldots, s_K \in \mathbb{R}$

$$h_n \mathbb{E}\left[e^{\sum_{j=1}^K s_j \left(P_{y_j}^{(n)} - P_{y_{j-1}}^{(n)}\right)}\right] = [t^n] \exp\left(\sum_{j=1}^L (e^{s_j} - 1) \sum_{x_{n,j} \leq k < x_{n,j-1}} \frac{\theta_k}{k} t^k + g_\Theta(t)\right). \tag{3.10}$$

It is now straight forward to see that we can use the exactly same argumentation as for $K = 1$. The only difference is that the notation is more cumbersome. This completes the proof of (3.1).

3.2. **Tightness.** To complete the proof of Theorem 1.3, we have to show that the process $\{P_y^{(n)}, y \geq 0\}$ is tight. By [5, Theorem 13.5 and (13.14)], it is sufficient to show for each $M > 0$ that

$$\mathbb{E}_\Theta\left[\left(P_y^{(n)} - P_{y_1}^{(n)}\right)^2 \left(P_{y_2}^{(n)} - P_y^{(n)}\right)^2\right] = O\left((y_2 - y_1)^2\right) \tag{3.11}$$

uniformly in y, y_1, y_2 with $0 \leq y_1 \leq y \leq y_2 \leq M$. We define

$$x_n^{(j)} := n^*(\ell_n + \min\{-\log(y_1), \ell_n\}) \text{ for } j = 1, 2.$$

Using the definition of $P_y^{(n)}$, we see that

$$\left(P_y^{(n)} - P_{y_1}^{(n)}\right)^2 \left(P_{y_2}^{(n)} - P_y^{(n)}\right)^2 = 0$$

if $x_n^{(2)} - x_n^{(1)} < 1$ since the involved sums are empty. In this case (3.11) is trivially fulfilled. We thus can assume that $x_n^{(2)} - x_n^{(1)} \geq 1$, which is equivalent to $y_2 \geq y_1 e^{v_n}$.

Using (3.10), we immediately get for $s_1, s_2 \in \mathbb{R}$

$$\mathbb{E}_\Theta\left[e^{s_1\left(P_y^{(n)} - P_{y_1}^{(n)}\right) + s_2\left(P_{y_2}^{(n)} - P_y^{(n)}\right)}\right]$$

$$= \frac{1}{h_n} [z^n] \exp\left((e^{s_1} - 1) F_{n,y_1,y}(t) + (e^{s_2} - 1) F_{n,y,y_2}(t)\right) \exp\left(g_\Theta(t)\right)$$

with

$$F_{n,u,v}(t) := \sum_{x_{n,u} \leq k < x_{n,v}} \frac{\theta_k}{k} t^k \text{ for } 0 \leq u \leq v \leq M \text{ and} \tag{3.12}$$

where $x_{n,w} := n^*(\ell_n + \min\{-\log(w), \ell_n\})$. We now have

$$\mathbb{E}_\Theta\left[\left(P_y^{(n)} - P_{y_1}^{(n)}\right)^2\left(P_{y_2}^{(n)} - P_y^{(n)}\right)^2\right] = \frac{\partial^2}{\partial s_2^2}\frac{\partial^2}{\partial s_1^2}\mathbb{E}_\Theta\left[e^{s_1\left(P_t - P_{t_1}\right) + s_2\left(P_{t_2} - P_t\right)}\right]\Bigg|_{s_1=s_2=0}.$$

Calculating the derivatives and entering $s_1 = s_2 = 0$ gives

$$\mathbb{E}_\Theta\left[\left(P_y^{(n)} - P_{y_1}^{(n)}\right)^2\left(P_{y_2}^{(n)} - P_y^{(n)}\right)^2\right] = \frac{1}{h_n}\left[t^n\right]f_n(t)\exp\left(g_\Theta(t)\right)$$

with

$$f_n(t) := F_{n,y_1,y}(t)\left(1 + F_{n,y_1,y}(t)\right)F_{n,y,y_2}(t)\left(1 + F_{n,y,y_2}(t)\right).$$

By the definition, we have $f_n(t) = \sum_{k=0} f_{n,k}t^k$ with all $f_{n,k} \geq 0$. Furthermore $g_\Theta(t)$ is log-n-admissible. This follows immediately from Lemma 3.1 using $s = 0$. Thus we get with Corollary 2.4 that three exists a n_0 only dependent on $g_\Theta(t)$ such that

$$\mathbb{E}_\Theta\left[\left(P_y^{(n)} - P_{y_1}^{(n)}\right)^2\left(P_{y_2}^{(n)} - P_y^{(n)}\right)^2\right] \leq 2f_n(r_n) \text{ for all } n \geq n_0.$$

We therefore have to estimate $f_n(r_n)$. We have

$$f_n(r_n) \leq \left(F_{n,y_1,y_2}(r_n)\left(1 + F_{n,y_1,y_2}(r_n)\right)\right)^2.$$

Since $0 \leq y_1 < y_2 \leq M$, it is sufficient to show that

$$F_{n,y_1,y_2}(r_n) = \sum_{x_{n,y_2} \leq k < x_{n,y_1}} k^{\alpha-1}e^{-kv_n} \leq 5(y_2 - y_1). \tag{3.13}$$

Using that the function $x^{\alpha-1}e^{-xv_n}$ is monotone decreasing for $x \geq (\alpha-1)n^*$ and the variable substitution $x = n^*(\ell_n - \log(u))$ gives

$$\sum_{x_{n,y_2} \leq k < x_{n,y_1}} k^{\alpha-1}e^{-kv_n} \leq \int_{x_{n,y_2}-1}^{x_{n,y_1}} x^{\alpha-1}e^{-xv_n}dx$$

$$= \int_{\max\{y_1, e^{-\ell_n}\}}^{e^{v_n}\max\{y_2, e^{-\ell_n}\}} (n^*)^{\alpha-1}\left(\ell_n - \log(u)\right)^{\alpha-1}e^{-\ell_n}\,du$$

$$\leq \frac{1}{(\alpha\log n^*)^{\alpha-1}}\int_{\max\{y_1, e^{-\ell_n}\}}^{e^{v_n}\max\{y_2, e^{-\ell_n}\}}\left(\ell_n - \log(u)\right)^{\alpha-1}du. \tag{3.14}$$

We have to distinguish the cases $0 < \alpha \leq 1$ and $\alpha > 1$. For $0 < \alpha \leq 1$ we use that $0 \leq y_1 < y_2 \leq M$ and that

$$\frac{\left(\ell_n - \log(M)\right)^{\alpha-1}}{(\alpha\log n^*)^{\alpha-1}} \longrightarrow 1 \text{ as } n \to \infty. \tag{3.15}$$

Thus we get for n large

$$\sum_{x_{n,y_2} \leq k < x_{n,y_1}} k^{\alpha-1}e^{-kv_n} \leq \int_{\max\{y_1, e^{-\ell_n}\}}^{e^{v_n}\max\{y_2, e^{-\ell_n}\}}\frac{\left(\ell_n - \log(M)\right)^{\alpha-1}}{(\alpha\log n^*)^{\alpha-1}}\,du \leq 2\int_{\max\{y_1, e^{-\ell_n}\}}^{e^{v_n}\max\{y_2, e^{-\ell_n}\}} 1\,du$$

$$\leq 2\left(y_2 e^{v_n} - y_1\right) \leq 2\left(y_2 e^{v_n} - y_2\right) + 2(y_2 - y_1)$$

$$\leq 2e^{v_n}(y_2 - y_1) + 2(y_2 - y_1) \leq 5(y_2 - y_1).$$

We used on the last line the assumption in (3.11) and that $v_n \to 0$. This completes the proof for $0 < \alpha \leq 1$. For $\alpha > 1$, we use that $-\log(u) \leq \ell_n$ in (3.14) and get

$$\sum_{x_{n,y_2} \leq k < x_{n,y_1}} k^{\alpha-1} e^{-kv_n} \leq \frac{1}{(\log n^*)^{\alpha-1}} \int_{\max\{y_1, e^{-\ell_n}\}}^{e^{v_n} \max\{y_2, e^{-\ell_n}\}} \left(2\ell_n\right)^{\alpha-1} du$$

The remaining computations are the same as for $0 < \alpha \leq 1$ and thus this completes the proof.

3.3. Proof of Theorem 1.2.

To prove Theorem 1.2, observe that we have for $j \in \mathbb{N}$ and $m \in \mathbb{N}$

$$L_j = \max \left\{ m; \sum_{k=m}^{n} C_k \geq j \right\} \tag{3.16}$$

where L_j is the length of the j'th longest cycle. We now define

$$\widetilde{L}_j := \max \left\{ m; \sum_{k=m}^{2n^* \ell_n} C_k \geq j \right\} \tag{3.17}$$

We thus immediately get for $y > 0$ with the definition of $P_j^{(n)}$ in (1.8) that

$$\left\{ \frac{\widetilde{L}_j - n^* \ell_n}{n^*} < -\log(y) \right\} = \left\{ \widetilde{L}_j < n^* \left(\ell_n - \log(y) \right) \right\} = \{ P_y^{(n)} < j \}. \tag{3.18}$$

By Theorem 1.3, we get that

$$\mathbb{P} \left[\frac{\widetilde{L}_j - n^* \ell_n}{n^*} < -\log(y) \right] \to e^{-y} \sum_{k=0}^{j-1} \frac{y^k}{k!}. \tag{3.19}$$

Recall, the cumulative distribution function of an $\mathrm{Exp}(1)$ distributed random variable E_1 is $\mathbb{P}\left[E_1 \leq u\right] = 1 - e^{-u}$. We thus get with $j = 1$ and $y = e^{-x}$ that

$$\mathbb{P}\left[\frac{\widetilde{L}_1 - n^* \ell_n}{n^*} < x \right] \to e^{-e^{-x}} = \mathbb{P}\left[E_1 \geq e^{-x}\right] = \mathbb{P}\left[-\log(E_1) \leq x\right]. \tag{3.20}$$

This establishes the asymptotic behaviour of $\widetilde{L}_1$. We could determine the asymptotic behaviour of the vector $(\widetilde{L}_1, \ldots, \widetilde{L}_K)$ in similar way as for $\widetilde{L}_1$. However, instead to do this directly, it is easier to us the distributions of the jump times of the Poisson process (see,e.g. [15, p.5]). Denote by $y_1^{(n)} < y_2^{(n)} < \ldots$ the jump times of the process $P_y^{(n)}$. These jumps corresponds to the $\widetilde{L}_j$ by the identity $\widetilde{L}_j = n^* \left(\ell_n - \log(y_j^{(n)}) \right)$. Further, since the process $P_y^{(n)}$ converges to a Poisson process, we know that

$$\left(y_1^{(n)}, y_2^{(n)}, \ldots, y_K^{(n)} \right) \xrightarrow{d} \left(E_1, E_1 + E_2, \ldots, \sum_{j=1}^{K} E_j \right), \tag{3.21}$$

where $(E_j)_{j=1}^K$ is a sequence of iid $\mathrm{Exp}(1)$ distributed random variables. This then implies that

$$\frac{1}{n^*} \cdot \left(\widetilde{L}_1 - n^* \ell_n, \ldots, \widetilde{L}_K - n^* \ell_n \right) \xrightarrow{d} \left(-\log(E_1), \ldots, -\log\left(\sum_{j=1}^{K} E_j \right) \right).$$

To complete the proof of Theorem 1.2, we consider the event $B_n := \{\sum_{k>2n^*\ell_n} C_k \geq 1\}$. We now get with the Markov inequality

$$\mathbb{P}_\Theta\left[B_n\right] = \mathbb{P}_\Theta\left[\sum_{k>2n^*\ell_n} C_k \geq 1\right] \leq \mathbb{E}_\Theta\left[\sum_{k>2n^*\ell_n} C_k\right]. \qquad (3.22)$$

Using Corollary 2.4, we get with a similar computation as in Section 3.2 that

$$\mathbb{P}_\Theta\left[B_n\right] \leq 2 \sum_{k>2n^*\ell_n} \frac{\theta_k}{k} e^{-kv_n} = O\left(\frac{(2n^*\ell_n)^{\alpha-1} e^{-v_n 2n^*\ell_n}}{v_n}\right) = O\left((n^*)^{-\alpha} \log^{1-\alpha} n\right). \quad (3.23)$$

Thus $\mathbb{P}_\Theta\left[B_n\right] \to 0$. Further $(\widetilde{L}_1, \ldots, \widetilde{L}_K)$ and $(L_1, \ldots, L_K)$ agree on the complement of B_n. This completes the proof of Theorem 1.2.

REFERENCES

[1] R. Arratia, A. Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.

[2] R. Arratia and S. Tavaré. The cycle structure of random permutations. *Ann. Probab.*, 20(3):1567–1591, 1992.

[3] V. Betz, H. Schäfer, and D. Zeindler. Random permutations without macroscopic cycles. Dec. 2017.

[4] V. Betz, D. Ueltschi, and Y. Velenik. Random permutations with cycle weights. *Ann. Appl. Probab.*, 21(1):312–331, 2011.

[5] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[6] A. Cipriani and D. Zeindler. The limit shape of random permutations with polynomially growing cycle weights. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(2):971–999, 2015.

[7] J. M. DeLaurentis and B. G. Pittel. Random permutations and Brownian motion. *Pacific J. Math.*, 119(2):287–301, 1985.

[8] S. Dereich and P. Mörters. Cycle length distributions in random permutations with diverging cycle weights. *Random Structures Algorithms*, 46(4):635–650, 2015.

[9] N. M. Ercolani and D. Ueltschi. Cycle structure of random permutations with cycle weights. *Random Structures Algorithms*, 44(1):109–133, 2014.

[10] M. M. Erlihson and B. L. Granovsky. Limit shapes of Gibbs distributions on the set of integer partitions: the expansive case. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(5):915–945, 2008.

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

[12] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, New York, NY, USA, 2009.

[13] F. M. Hoppe. The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.*, 25(2):123–159, 1987.

[14] J. F. C. Kingman. The population structure associated with the Ewens sampling formula. *Theoret. Population Biology*, 11(2):274–283, 1977.

[15] T. M. Liggett. *Continuous time Markov processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. An introduction.

[16] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[17] K. Maples, A. Nikeghbali, and D. Zeindler. The number of cycles in a random permutation. *Electron. Commun. Probab.*, 17:no. 20, 1–13, 2012.

[18] G. Pólya. Kombinatorische anzahlbestimmungen für gruppen, graphen, und chemische verbindungen. *Acta Mathematica*, 68:145–254, 1937.

[19] N. Robles and D. Zeindler. Random permutations with logarithmic cycle weights. *arXiv e-prints*, page arXiv:1806.04700, Jun 2018.

[20] A. Shimdt and A. M. Vershik. Limit measures arising in the asymptotic theory of symmetric groups. *Theory Probab. Appl.*, 22, No.1:70–85, 1977.

[21] J. Storm and D. Zeindler. Total variation distance and the Erdös-Turán law for random permutations with polynomially growing cycle weights. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1614–1640, 2016.

DEPARTMENT OF MATHEMATICS AND STATISTICS, LANCASTER UNIVERSITY, FYLDE COLLEGE, BAILRIGG, LANCASTER LA1 4YF, UNITED KINGDOM

Email address: d.zeindler@lancaster.ac.uk