LARGE CYCLES AND A FUNCTIONAL CENTRAL LIMIT THEOREM FOR GENERALIZED WEIGHTED RANDOM PERMUTATIONS.

ASHKAN NIKEGHBALI, JULIA STORM, AND DIRK ZEINDLER

ABSTRACT. The objects of our interest are the so-called A-permutations, which are permutations whose cycle length lie in a fixed set A. They have been extensively studied with respect to the uniform or the Ewens measure. In this paper, we extend some classical results to a more general weighted probability measure which is a natural extension of the Ewens measure and which in particular allows to consider sets A_n depending on the degree n of the permutation. By means of complex analysis arguments and under reasonable conditions on generating functions we study the asymptotic behaviour of classical statistics. More precisely, we generalize results concerning large cycles of random permutations by Vershik, Shmidt and Kingman, namely the weak convergence of the size ordered cycle length to a Poisson-Dirichlet distribution. Furthermore, we apply our tools to the cycle counts and obtain a Brownian motion central limit theorem which extends results by DeLaurentis, Pittel and Hansen.

CONTENTS

1

1. Introduction

Permutations are classical objects that appear in many mathematics fields. A special class of permutations are the so-called A-permutations, where A is a non-empty subset of $\mathbb{N}$. We call an element σ of the symmetric group $\mathfrak{S}_n$ an A-permutation if σ can be written as a product of disjoint cycles whose cycle-lengths are all in A. These permutations have been extensively studied over the past thirty years, a long list of references can be found for instance in [17]. It is well-known that with respect to the uniform measure the behaviour of A-permutations is similar to those of the whole permutation group. To give a single example, in [18] it was proved that for $n \to \infty$ the cycle counts C_m (the number of cycles of length m of σ) converge in distribution for $m \in A$ to independent Poisson distributed random variables Y_m with expectation $1/m$. However, in all previous publications about A-permutations, one has only investigated its behaviour under the uniform measure and with A being independent of n. Here, we consider the following more general A_n-weighted measure.

Definition 1.1. *Let $A_n \subset \{1, \ldots, n\}$ and $\Theta = (\theta_m)_{m \geq 1}$ be given, with $\theta_m \geq 0$ for every $m \geq 1$. We define the A_n-weighted measure of $\sigma \in \mathfrak{S}_n$ as*

$$\mathbb{P}_\Theta^{(A_n)}[\sigma] := \frac{1}{h_n n!} \prod_{m=1}^{\ell(\lambda)} \theta_{\lambda_m} \mathbb{1}_{\{\lambda_m \in A_n\}} \tag{1.1}$$

with $h_n = h_n(A_n)$ a normalization constant with $h_0 := 1$, $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ the cycle-type of σ and $\ell(\lambda) = \ell$ the length of λ (see Section 2.1).

Define furthermore

$$D_n := \{1, \ldots, n\} \setminus A_n \quad \text{and} \quad d_n := \begin{cases} \max D_n & \text{if } D_n \neq \emptyset, \\ 1 & \text{otherwise}. \end{cases} \tag{1.2}$$

We investigate the behaviour of the measure $\mathbb{P}_\Theta^{(A_n)}$ for $d_n = o(n)$, that is to say the cycle lengths not contained in A_n grow slowly (the precise assumptions on d_n can be found in Theorem 3.3). This assumption is motivated by a model in [12, Section 6] about mod-Poisson convergence for an analogue of the Erdös-Kac Theorem for polynomials over finite fields.

The uniform measure or the Ewens measure on $\mathfrak{S}_n$ are special cases of the A_n-weighted measure, obtained by choosing $A_n = \mathbb{N}$ and $\theta_m \equiv 1$ or $\theta_m \equiv \theta$. Both are classical probability measures and are well-studied, see for instance [1].

For $A_n = \mathbb{N}$, one obtains the weighted measure on $\mathfrak{S}_n$, which was recently investigated in [4], [8], [14], [15] (see also the extensive background bibliography therein). Our study extends the results in [15] about the cycle counts and the total cycle number to $\mathbb{P}_\Theta^{(A_n)}$ and is based on similar argumentations as those in [15]. Furthermore, we apply our methods to objects which have so far not been considered for the weighted measure on $\mathfrak{S}_n$. More precisely, in Section 4.2 we show that the size ordered cycle lengths converge in law to a Poisson-Diriclet-distribution.

This result agrees with those by Vershik and Shmidt [16] and Kingman [11], who studied the same asymptotic behaviour with respect to the Ewens measure. Furthermore, we consider in Section 5 the number of cycles in a permutation with lengths not exceeding n^x and show that this process converges, after proper normalisation, to a standard Brownian motion. This extends the results by Delaurentis and Pittel [6] (uniform measure) and Hansen [10] (Ewens measure) to the weighted measure and the A_n-weighted measure on $\mathfrak{S}_n$. A great advantage of our argumentation is that it is much more flexible and one can obtain easily the behaviour under further restrictions, see Sections 5.2 and 5.3.

It is clear that the asymptotic behaviour of all random variables on the group $\mathfrak{S}_n$ with respect to the measure $\mathbb{P}_{\Theta}^{(A_n)}$ strongly depend on the sequence $\Theta = (\theta_m)_{m \geq 1}$ and it is thus necessary to impose appropriate assumptions on this sequence. More precisely, we will argue with generating functions, meaning that assumptions are imposed on the function

$$g(t) = g_{\Theta}(t) := \sum_{n=1}^{\infty} \frac{\theta_n}{n} t^n. \tag{1.3}$$

The link of $g_{\Theta}(t)$ and the generating series of $(h_n)_{n \geq 1}$ is the starting point of our study; for $A_n = \{1, ..., n\}$ it is given by the well-known relation

$$\sum_{n=1}^{\infty} h_n t^n = \exp(g_{\Theta}(t)).$$

For general sets A their relation will be stated in Lemma 2.4. We will choose $g_{\Theta}(t)$ in a way that allows us to apply the method of singularity analysis, see Definition 3.2. We give more details in Section 3, but a good description of the method of singularity analysis can be found for instance in the book [9] by Flajolet and Sedgewick.

The paper is organized as follows. In Section 2 some well known facts about the symmetric group $\mathfrak{S}_n$ are presented and generating functions are introduced. In particular, we will recall the *cycle index theorem*, wich will be used in computations throughout the whole paper. In Section 3 we determine the setting of our study, meaning that we properly define the assumptions on the functions $g(t)$ under consideration. With complex analysis arguments we establish our main tool, Theorem 3.3, which enables us to investigate the large-n behaviour of coefficients of relevant functions. As a direct consequence, we deduce the asymptotic behavior of the normalization constant h_n. In Section 4 we apply our methods to compute the characteristic functions of the cycle counts and of the total cycle number and we deduce a central limit theorem and a even stronger convergence result, namely mod-Poisson convergence. Further we investigate the behaviour of the large cycles and show that their asymptotic behaviour with respect to our general measure is the same as with respect to the Ewens measure. Finally, Section 5 is devoted to a functional central limit theorem giving the weak convergence of a certain functional of the cycle counts to the Brownian motion.

2. COMBINATORICS OF $\mathfrak{S}_n$ AND GENERATING FUNCTIONS

This section is devoted to some basic facts about the symmetric group $\mathfrak{S}_n$, partitions and generating functions. Further, a useful lemma which identifies averages over $\mathfrak{S}_n$ with generating functions is recalled. We give only a short overview and refer to [1] and [13] for more details.

2.1. **The symmetric group.** All probability measures and functions considered in this paper are invariant under conjugation and it is well known that the conjugation classes of $\mathfrak{S}_n$ can be parametrized with partitions of n. This can be seen as follows: Let $\sigma \in \mathfrak{S}_n$ be an arbitrary permutation and write $\sigma = \sigma_1 \cdots \sigma_\ell$ with σ_i disjoint cycles of length λ_i. Since disjoint cycles commute, we can assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell$. We call the partition $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_\ell)$ the *cycle-type* of σ and $\ell = \ell(\lambda)$ its *length*. Then two elements $\sigma, \tau \in \mathfrak{S}_n$ are conjugate if and only if σ and τ have the same cycle-type. Further details can be found for instance in [13]. For $\sigma \in \mathfrak{S}_n$ with cycle-type λ we define C_m, the number of cycles of size m, and T, the total cycle number as

$$C_m := \# \{i; \lambda_i = m\} \quad \text{and} \quad T := T_n := \sum_{m=1}^{n} C_m. \tag{2.1}$$

It will turn out that all expectations of interest have the form $\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} u(\sigma)$ for a certain class function u. Since u is constant on conjugacy classes, it is more natural to sum over all conjugacy classes. This is subject of the following lemma.

Lemma 2.1. *Let $u : \mathfrak{S}_n \to \mathbb{C}$ be a class function. For $C_m = C_m(\lambda)$ as in (2.1) and $\mathcal{C}_\lambda$ the conjugacy class corresponding to the partition λ we have*

$$|\mathcal{C}_\lambda| = \frac{|\mathfrak{S}_n|}{z_\lambda} \quad \text{with} \quad z_\lambda := \prod_{m=1}^{n} m^{C_m} C_m!$$

and

$$\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} u(\sigma) = \sum_{\lambda \vdash n} \frac{1}{z_\lambda} u(\mathcal{C}_\lambda).$$

2.2. **Generating functions.** Given a sequence $(a_n)_{n \in \mathbb{N}}$ of numbers, one can encode important information about this sequence into a formal power series called the generating series.

Definition 2.2. *Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of complex numbers. We then define the (ordinary) generating function of $(a_n)_{n \in \mathbb{N}}$ as the formal power series*

$$G(t) = G(a_n, t) = \sum_{n=0}^{\infty} a_n t^n.$$

We define $[t^n][G]$ to be the coefficient of t^n of $G(t)$, that is $[t^n][G] := a_n$.

The reason why generating functions are powerful is the possibility of recognizing them without knowing the coefficients a_n explicitly. In this case one can try to use tools from analysis to extract information about a_n, for large n, from the generating function.

The following lemma goes back to Polya and is sometimes called *cycle index theorem*. It links generating functions and averages over $\mathfrak{S}_n$.

Lemma 2.3. *Let $(a_m)_{m\in\mathbb{N}}$ be a sequence of complex numbers. Then*

$$\sum_\lambda \frac{1}{z_\lambda} \left(\prod_{m=1}^{\ell(\lambda)} a_{\lambda_m} \right) t^{|\lambda|} = \sum_\lambda \frac{1}{z_\lambda} \left(\prod_{m=1}^\infty (a_m t^m)^{C_m} \right) = \exp\left(\sum_{m=1}^\infty \frac{1}{m} a_m t^m \right)$$

(2.2)

with the same z_λ as in Lemma 2.1. If one of the sums in (2.2) is absolutely convergent then so are the others.

Proof. The proof can be found in [13] or can be directly verified using the definitions of z_λ and the exponential function. The last statement follows from the dominated convergence theorem. □

The crucial tool for our study, the relation of $g_\Theta(t)$ to the generating function of $h_n(A)$ for $A \subset \mathbb{N}$, can immediately be deduced from the previous lemma.

Lemma 2.4. *Let $A \subset \mathbb{N}$ and Θ be given as in Definition 1.1 and define $D := \mathbb{N} \backslash A$. We then have as formal power series*

$$\sum_{n=0}^\infty h_n(A) t^n = \exp\left(g_\Theta(t) - L_D(t) \right),$$

(2.3)

where $g_\Theta(t)$ is given by (1.3) and $L_D(t)$ is the formal power series

$$L_D(t) := \sum_{m \in D} \frac{\theta_m}{m} t^m.$$

Proof. We combine the definition of h_n, Lemma 2.1 and Lemma 2.3. We get

$$h_n(A) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \prod_{m=1}^{\ell(\lambda)} \left(\theta_{\lambda_m} \mathbb{1}_{\{\lambda_m \in A\}} \right) = \sum_{\lambda \vdash n} \frac{1}{z_\lambda} \prod_{m=1}^{\ell(\lambda)} \left(\theta_{\lambda_m} \mathbb{1}_{\{\lambda_m \in A\}} \right).$$

It follows that

$$\sum_{n=1}^\infty h_n(A) t^n = \exp\left(\sum_{m=1}^\infty \frac{\theta_m}{m} \mathbb{1}_{\{m \in A\}} t^m \right) = \exp\left(g_\Theta(t) - L_D(t) \right).$$

□

3. Singularity analysis for increasing cycle lengths

The main goal of this section is to provide by means of complex analysis arguments a tool, Theorem 3.3, that allows us to compute the large-n behaviour of $h_n(A_n)$ and of other related quantities.

For this purpose, as mentioned in the introduction, we have to impose assumptions on the sequence $\Theta = (\theta_m)_{m \in \mathbb{N}}$. In view of Lemma 2.4, it is natural to impose them on the function

$$g(t) = g_\Theta(t) := \sum_{m=1}^{\infty} \frac{\theta_m}{m} t^m. \tag{3.1}$$

We shall apply the method of singularity analysis to the function $g(t)$. A detailed description of singularity analysis can be found for instance in [9, Section VI]. First we need a preliminary definition.

Definition 3.1. *Let* $0 < r < R$ *and* $0 < \phi < \frac{\pi}{2}$ *be given. We then define*

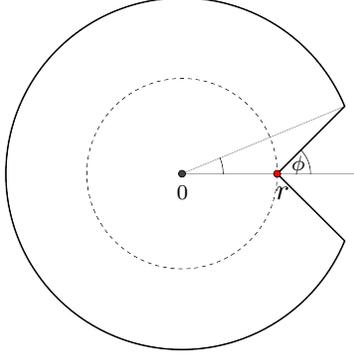$$\Delta_0 = \Delta_0(r, R, \phi) = \left\{ z \in \mathbb{C}; |z| < R, z \neq r, |\arg(z - r)| > \phi \right\}. \tag{3.2}$$

FIGURE 1. Illustration of Δ_0

Now we can introduce the family of functions we are interested in.

Definition 3.2. *Let* $r, \vartheta > 0$ *and* $K \in \mathbb{R}$ *be given. We write* $\mathcal{F}(r, \vartheta, K)$ *for the set of all functions* $g(t)$ *satisfying*

(1) $g(t)$ *is holomorphic in* $\Delta_0(r, R, \phi)$ *for some* $R > r$ *and* $0 < \phi < \frac{\pi}{2}$ *and*
(2)

$$g(t) = \vartheta \log\left(\frac{1}{1 - t/r} \right) + K + O\left(t - r \right) \ as \ t \to r. \tag{3.3}$$

Notice that $\theta_m \equiv \vartheta$ leads to $g_\Theta(t) = -\vartheta \log(1 - t) \in \mathcal{F}(1, \vartheta, 0)$ and thus the Ewens measure is covered by the family $\mathcal{F}(r, \vartheta, K)$. Also functions of the form $-\vartheta \log(1 - t) + f(t)$ with $f(t)$ holomorphic for $|t| < 1 + \epsilon$ are contained in $\mathcal{F}(1, \vartheta, f(1))$. In particular, the case $\theta_k \neq \vartheta$ for only finitely many k is covered by the family $\mathcal{F}(1, \vartheta, \cdot)$.

Remark. From from [9, Theorem VI.4] we deduce the following observation, which we will use frequently: if $g(t)$ is defined as in (3.1) and $g(t) \in \mathcal{F}(r, \vartheta, K)$, then there exists some ϵ_m such that

$$\theta_m r^m = \vartheta + \epsilon_m \quad \text{with } \epsilon_m \to 0 \text{ and } \sum_{m=1}^{\infty} \frac{|\epsilon_m|}{m} < \infty. \tag{3.4}$$

We are now ready to state the main theorem of this section.

Theorem 3.3. *Let $g(t)$ in $\mathcal{F}(r, \vartheta, K)$ and $(D_n^{(j)})_{n \in \mathbb{N}, 1 \leq j \leq k}$ with $D_n^{(j)} \subset \{1, \ldots, n\}$ be given. We define*

$$d_n^{(j)} := \begin{cases} \max D_n^{(j)} & \text{if } D_n^{(j)} \neq \emptyset, \\ 1 & \text{otherwise}, \end{cases} \quad \text{and} \quad \bar{d}_n := \max\{d_n^{(j)}\}. \tag{3.5}$$

Let further

$$G_n(t, w, v_1, ..., v_k) := \exp\left(wg(t) + \sum_{j=1}^{k} v_j L_{D_n^{(j)}}(t)\right) \tag{3.6}$$

with $w, v_1, \ldots, v_k \in \mathbb{C}$ and $L_{D_n^{(j)}}(t)$ as in Lemma 2.4. Suppose that for each $C \in \mathbb{R}$

$$C \log n - \frac{n}{d_n} \to -\infty \tag{3.7}$$

holds as $n \to \infty$. Then we have for any fixed $b \in \mathbb{N}$

$$[t^{n-b}] [G_n(t, w, v_1, ..., v_k)] \tag{3.8}$$

$$= \frac{e^{Kw} n^{w\vartheta - 1}}{r^{n-b}} \exp\left(\sum_{j=1}^{k} v_j L_{D_n^{(j)}}(r)\right) \left(\frac{1}{\Gamma(w\vartheta)} + O\left(\frac{\bar{d}_n}{n}\right)\right)$$

uniformly for bounded $|w|, |v_1|, ..., |v_k| \leq \hat{r}$ for some $\hat{r} > 0$.

Remark. We have introduced in Theorem 3.3 more than one set D_n since this will allow us to compute easily the finite dimensional distributions of the process $B_n(x)$ in Section 5.

Remark. Apparently assumption (3.7) is not fulfilled if $D_n^{(j)} = \{1, \ldots, n\}$ for some j. However, in this case

$$[t^n] \left[\exp\left(v L_{D_n^{(j)}}(t)\right)\right] = [t^n] [\exp(vg(t))]$$

holds and we can thus handle $D_n^{(j)} = \{1, \ldots, n\}$ with Theorem 3.3 by replacing $L_{D_n^{(j)}}(t)$ with $g(t)$ in (3.6).

Remark. We will mostly use (3.8) with $b = 0$, except in Subsection 4.2 where we study the behaviour of the large cycles.

Before proving Theorem 4.2 we deduce the large-n behaviour of h_n.

Corollary 3.4. *Let $g_\Theta(t)$ in $\mathcal{F}(r, \vartheta, K)$ be given and let $(A_n)_{n \in \mathbb{N}}$ be the defining sets of the measures $\mathbb{P}_\Theta^{(A_n)}[.]$ (see Definition 1.1). We set $D_n := \{1, \ldots, n\} \setminus A_n$ and define d_n as in (1.2). If the sequence d_n fulfils the assumption (3.7), then*

$$h_n(A_n) = \exp\left(-L_{D_n}(r)\right) \frac{n^{\vartheta-1} e^K}{r^n \Gamma(\vartheta)} \left(1 + O\left(\frac{d_n}{n}\right)\right).$$

In particular, (3.7) is fulfilled if

$$d_n \sim \log n \quad or \quad d_n \sim n^\alpha \text{ for } 0 < \alpha < 1.$$

Proof. We know from Lemma 2.4 that

$$h_n(A) = [t^n] \left[\exp\left(g_\Theta(t) - L_D(t)\right)\right]$$

holds for arbitrary sets $A \subset \mathbb{N}$. Thus the first part of the corollary follows immediately from Theorem 3.3. The second part is obvious. $\qquad\square$

Proof of Theorem 3.3. For simplicity, we assume $k = 1, b = 0$ and write $D_n := D_n^{(1)}, d_n := \bar{d}_n = d_n^{(1)}$ and $v := v_1$. The proof of the general case is completely similar. We apply Cauchy's integral formula to $G_n(t, w, v)$. This gives

$$[t^n] [G_n(t, w, v)] = \frac{1}{2\pi i} \int_\gamma \exp\left(wg(t) + vL_{D_n}(t)\right) \frac{dt}{t^{n+1}} \tag{3.9}$$

for some curve γ. We follow the idea in [9, Section VI.3] and choose the curve γ as in Figure 2(a). The main difference with [9] is that we let the radius of the large

(a) $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$ (b) $\gamma' = \gamma_1' \cup \gamma_2' \cup \gamma_3'$ (c) $\gamma'' = \gamma_1'' \cup \gamma_2'' \cup \gamma_3''$
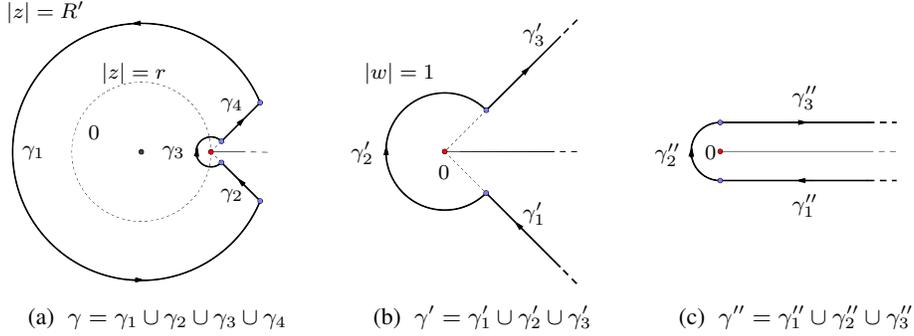
FIGURE 2. The curves used in the proof of Theorem 3.3.

circle slowly tend to r while it is fixed in [9]. More precisely, by assumption $g(t)$

is holomorphic in $\Delta_0(r, R, \phi)$ (see (3.2)) and continuous on $\overline{\Delta_0(r, R, \phi)} \setminus \{r\}$. We then define the radius of the large circle as

$$R(n) := \min\{r(1 + 1/d_n), R\}$$

and define further γ as

$$\gamma_1(\varphi) := R(n)e^{i\varphi} \qquad\qquad \text{for } \varphi \in [-\pi + \alpha_n, \pi - \alpha_n]$$

$$\gamma_2(\varphi) := r\left(1 - \frac{1}{n}e^{-i\varphi}\right) \qquad\qquad \text{for } \varphi \in [-\pi + \phi, \pi - \phi]$$

$$\gamma_3(x) := r(1 + xe^{i\phi}) \qquad\qquad \text{for } x \in [1/n, \hat{r}_n]$$

$$\gamma_4(x) := r(1 + (\hat{r}_n - x)e^{-i\phi}) \qquad\qquad \text{for } x \in [0, \hat{r}_n - 1/n]$$

where α_n and $\hat{r}_n$ are chosen such that the curve γ is closed, i.e. $r + \hat{r}_n e^{i\phi} = R(n)e^{i\alpha_n}$.

We first compute the integral over γ_1. If $\sup_n d_n = C < \infty$, we clearly have $R(n) \geq \widetilde{R} > r$ for some $\widetilde{R}$ independent of n. Thus all points of the curve γ_1 have at least a distance $|\widetilde{R} - r| > 0$ from r. Therefore $g(t)$ is uniformly bounded on γ_1. Furthermore $L_{D_n}(t)$ involves only θ_m with $m \leq C$ and is thus also uniformly bounded. We get

$$\left| \frac{1}{2\pi i} \int_{\gamma_1} \exp\left(wg(t) + vL_{D_n}(t)\right) \frac{dt}{t^{n+1}} \right| = O(\widetilde{R}^{-n}) = O\left(\frac{n^{w\vartheta-2}}{r^n} e^{vL_{D_n}(r)}\right).$$

If $\sup_n d_n = \infty$ we have to be more careful. In this case

$$\left| \frac{1}{2\pi i} \int_{\gamma_1} \exp\left(wg(t) + vL_{D_n}(t)\right) \frac{dt}{t^{n+1}} \right|$$

$$\leq \frac{1}{2\pi(R(n))^n} \int_{-\pi+\alpha_n}^{\pi-\alpha_n} \left| \exp\left(wg(R(n)e^{i\varphi}) + vL_{D_n}(R(n)e^{i\varphi})\right) \right| d\varphi$$

$$\leq \frac{1}{(R(n))^n} \exp(\max_{-\pi+\alpha_n \leq \varphi \leq \pi-\alpha_n} \{\mathrm{Re}\left[wg(R(n)e^{i\varphi}) + vL_{D_n}(R(n)e^{i\varphi})\right]\}).$$

Using that $g(t) \in \mathcal{F}(r, \vartheta, K)$ is continuous on $\overline{\Delta_0(r, R, \phi)} \setminus \{r\}$ and the expansion of $g(t)$ around r, one immediately obtains

$$\mathrm{Re}(g(t)) \leq \vartheta \log\left| \frac{1}{1 - t/r} \right| + O(1) \text{ and } \mathrm{Im}(g(t)) = O(1) \quad \text{for all } t \in \Delta_0 \setminus \{r\}.$$

This yields

$$\mathrm{Re}[wg(R(n)e^{i\varphi})] \leq |\mathrm{Re}(w)|\vartheta \log\left| \frac{1}{1 - R(n)/r} \right| + O(1) = |\mathrm{Re}(w)|\vartheta \log d_n + O(1).$$

Furthermore, we get

$$
\left| L_{D_n}\left((1+1/d_n)\,re^{i\varphi}\right) \right| \leq \sum_{m \in D_n} \frac{\theta_m}{m} r^m \left(1+1/d_n\right)^m
$$

$$
\leq \sum_{m \in D_n} \frac{\theta_m}{m} r^m \left(1 + O\left(m/d_n\right)\right)
$$

$$
\leq L_{D_n}(r) + O(1) \leq \log d_n + O(1) \qquad (3.10)
$$

since $m \leq d_n$ and $\theta_m r^m \sim \vartheta$, see (3.4). We also have

$$
\left(R(n)\right)^{-n} \leq r^{-n}(1+1/d_n)^{-n} = r^{-n}\exp(-n\log(1+1/d_n)) \leq r^{-n}\exp\left(-\frac{n}{2d_n}\right).
$$

Combining the above computations, we obtain

$$
\left| \frac{1}{2\pi i} \int_{\gamma_1} e^{wg(t)+vL_{D_n}(t)} \frac{dt}{t^{n+1}} \right| = O\left(r^{-n}\exp\left(-\frac{n}{2d_n} + (|\mathrm{Re}(w)|+|v|)\vartheta \log d_n\right)\right).
$$

It remains to prove that this is $O(n^{w\vartheta-2}\exp\left(-vL_{D_n}(r)\right))$. This holds if

$$
-\frac{n}{2d_n} + (|\mathrm{Re}(w)|\vartheta + |v|)\log d_n \leq \mathrm{Re}((w\vartheta-2)\log n - vL_{D_n}(r)) + O(1)
$$

but this follows immediately from assumption (3.7) since

$$
L_{D_n}(r) \leq \log d_n \leq \log n.
$$

The computations of the integrals over γ_2, γ_3 and γ_4 are completely similar to the computations in the proof of Theorem VI.3 in [9] and we thus give only a short overview. A simple calculation gives

$$
L_{D_n}\left(\gamma_2(\varphi)\right) = L_{D_n}(r) + O\left(d_n/n\right) \quad \text{and} \quad L_{D_n}\left(\gamma_3(x)\right) = L_{D_n}(r) + O\left(d_n x\right).
$$

$$(3.11)$$

This observations together with the computations in [9] then yields

$$
\frac{1}{2\pi i} \int_{\gamma_2 \cup \gamma_3 \cup \gamma_4} G_n(t,w,v) \frac{dt}{t^{n+1}} = \frac{1}{2\pi i} \frac{n^{w\vartheta-1}}{r^n} \int_{\gamma'} z^{-w\vartheta} e^z dt \,\, (1 + O\left(d_n/n\right))
$$

with γ' as in Figure 2(b). The variable substitution $x = -z$ and a simple contour argument then gives

$$
\frac{1}{2\pi i} \frac{n^{w\vartheta-1}}{r^n} \int_{\gamma'} z^{-w\vartheta} e^z dt = \frac{1}{2\pi i} \int_{\gamma''} (-x)^{-w\vartheta} e^{-x} dt = \frac{1}{\Gamma(w\vartheta)}
$$

with γ'' as in Figure 2(c). We have used in the second equality that that the integral is a well know expression for the inverse of Γ-function. Further details can be found for instance in [9, Section B.3]. $\qquad\square$

To investigate the bahaviour of the large cycles and a functional central limit theorem we have to consider in Section 4.2 and 5.1 expressions of the form

$$
[t^n]\left[f(t)\cdot \exp\left(g(t)+vL_{D_n}(t)\right)\right] \qquad (3.12)
$$

where the funciton $f(t)$ is either a polynomial depending on n or it is independent of n and behaves like a derivative of the logarithm near r. By suitable modifications of Theorem 3.3 we obtain in this case the following asymptotics.

Corollary 3.5. *Let the assumptions of Theorem 3.3 be fulfilled with $k = 1$ and write $D_n := D_n^{(1)}, d_n := \bar{d}_n = d_n^{(1)}$. If $f(t)$ is a holomorphic function in Δ_0 and there exists a constant $\beta \geq 0$ such that*

$$f(t) = (1 - t/r)^{-\beta}(1 + O(t - r)), \quad t \to r \text{ and } t \in \Delta_0, \tag{3.13}$$

then

$$[t^n] \left[f(t) \cdot \exp\big(g(t) + vL_{D_n}(t)\big) \right]$$
$$= \frac{e^K n^{\vartheta + \beta - 1}}{r^n} \exp\big(vL_{D_n}(r)\big) \left(\frac{1}{\Gamma(\vartheta + \beta)} + O\Big(\frac{d_n}{n}\Big) \right). \tag{3.14}$$

Proof. Since we have $g(t) \in \mathcal{F}(r, \vartheta, K)$, we get with (3.3) and (3.13)

$$\log f(t) + g(t) = -(\vartheta + \beta)\log(1 - t/r) + K + O(t - r) \text{ as } t \to r.$$

We thus see that $\log f(t) + g(t) \in \mathcal{F}\big(r, \vartheta + \beta, K\big)$ and the corollary follows immediately from Theorem 3.3 with $g(t)$ replaced by $\log f(t) + g(t)$. □

Corollary 3.6. *Let the assumptions of Theorem 3.3 be fulfilled with $k = 1$ and write $D_n := D_n^{(1)}, d_n := \bar{d}_n = d_n^{(1)}$. Let further $P_n(t)$ be a sequence of polynomials with*

$$P_n(t) = \sum_k p_{k,n} t^k, \quad \text{where } p_{n,k} \geq 0,$$

such that $P_n\big(r(1 + 1/d_n)\big) = P_n(r)(1 + o(1))$. We then have for each $v \in \mathbb{R}$

$$[t^n] \left[P_n(t) \cdot \exp\big(g(t) + vL_{D_n}(t)\big) \right]$$
$$= P_n(r) \frac{n^{\vartheta - 1}}{r^n} \exp\big(vL_{D_n}(r)\big) \left(\frac{1}{\Gamma(w\vartheta)} + o(1) \right) \tag{3.15}$$

Proof. Since the computations for this proof are very similar to those of the proof of Theorem 3.3, we only illustrate the estimate over γ_1 with $\gamma_1(\varphi) = r(1 + 1/d_n)e^{i\varphi}$. We argue as in the first part of the proof of Theorem 3.3 and use that $P_n(t)$ has positive coefficients. We obtain

$$\left| \frac{1}{2\pi i} \int_{\gamma_1} P_n(t) \exp\big(g(t) + vL_{D_n}(t)\big) \frac{dt}{t^n} \right|$$
$$\leq \frac{P_n\big(r(1 + 1/d_n)\big)}{2\pi \big(R(n)\big)^n} \int_{-\pi + \alpha_n}^{\pi - \alpha_n} \left| \exp\big(wg(R(n)e^{i\varphi}) + vL_{D_n}(R(n)e^{i\varphi})\big) \right| d\varphi$$

The latter integral is now the same as in the proof of Theorem 3.3. Using the estimate in the proof of Theorem 3.3 and the assumption on $P_n(t)$ then completes the proof. □

4. CYCLE COUNTS, TOTAL NUMBER OF CYCLES AND LARGE CYCLES

4.1. **The cycle counts and the total number of cycles.** We consider here the cycle counts C_m and the total number of cycles T, defined in (2.1). First, we compute their generating functions and then deduce with Theorem 3.3 their asymptotic behaviour. As mentioned in the introduction, the required computations are quite similar to those in [15]. Therefore, we give here only a short overview and refer to [15] for more details.

Lemma 4.1. *Let $A \subset \mathbb{N}$, $D := \mathbb{N} \setminus A$ be given. We then have for $w \in \mathbb{C}$ as formal power series*

$$\sum_{n=1}^{\infty} h_n(A)\mathbb{E}_{\Theta}^{(A)} \left[\exp(wT)\right] t^n = \exp\left(e^w g_{\Theta}(t) - e^w L_D(t)\right).$$

Let further $M = \{m_1, \ldots, m_d\} \subset A$ be given. We then have for $w_1, \ldots, w_d \in \mathbb{C}$ as formal power series

$$\sum_{n=1}^{\infty} h_n(A)\mathbb{E}_{\Theta}^{(A)} \left[\exp\left(\sum_{j=1}^{d} w_{m_j} C_{m_j}\right)\right] t^n$$

$$= \exp\left(\sum_{j=1}^{d} \frac{\theta_{m_j}}{m_j}(e^{w_{m_j}} - 1)t^{m_j}\right) \exp\left(g_{\Theta}(t) - L_D(t)\right).$$

We omit the proof since it is a simple application of Lemma 2.3 and the computations are similar to the proof of Lemma 2.4.

It follows with Lemma 4.1 that

$$h_n(A)\mathbb{E}_{\Theta}^{(A)} \left[\exp(wT)\right] = [t^n] \left[\exp\left(e^w g_{\Theta}(t) - e^w L_D(t)\right)\right] \tag{4.1}$$

with $A \subset \mathbb{N}$ arbitrary. We can thus replace A in (4.1) by any A_n depending on n. Note that this is not possible in Lemma 4.1. Now combine Theorem 3.3 and Lemma 4.1 to obtain the asymptotic behaviour of the cycle counts.

Theorem 4.2. *Suppose that $g_{\Theta}(t)$ is in $\mathcal{F}(r, \vartheta, K)$. Let further $M = \{m_1, \ldots, m_d\}$ and $(A_n)_{n \in \mathbb{N}}$ with $A_n \subset \{1, ..., n\}$ be given and let d_n be defined as in (1.2). Suppose that*

(1) *d_n fulfils the assumption (3.7) and*
(2) *there exists $n_0 \in \mathbb{N}$ such that $M \subset A_n$ for all $n \geq n_0$.*

Then

$$\mathbb{E}_{\Theta}^{(A_n)} \left[\exp\left(\sum_{j=1}^{d} w_{m_j} C_{m_j}\right)\right] = \exp\left(\sum_{j=1}^{d} \frac{\theta_{m_j}}{m_j}(e^{w_{m_j}} - 1)r^{m_j}\right) + O\left(\frac{d_n}{n}\right)$$

$$\tag{4.2}$$

uniformly in $w_{m_1}, \ldots, w_{m_d}$ *for bounded* $\mathrm{Re}(w_{m_1}), \cdots, \mathrm{Re}(w_{m_d})$. *In particular, the random variables* $C_{m_j}, m_j \in M$ *converge in law to independent Poisson distributed random variables* Y_{m_j} *with* $\mathbb{E}\left[Y_{m_j}\right] = \theta_{m_j} r^{m_j}/m_j$.

Proof. Equation (4.2) follows immediately from Lemma 4.1 and Theorem 3.3. The error is uniform for bounded $\mathrm{Re}(w_{m_1}), \cdots, \mathrm{Re}(w_{m_d})$ since all $C_{m_j} \in \mathbb{N}$ and thus the function on the left-hand side of (4.2) is periodic. $\qquad\square$

The asymptotic behaviour of the total cycle number T is computed analogously.

Theorem 4.3. *Let* $g_\Theta(t)$, $(A_n)_{n\in\mathbb{N}}$ *and* d_n *be defined as in Theorem 4.2. Then*

$$\mathbb{E}_\Theta^{(A_n)}\left[\exp(isT)\right] = n^{\vartheta(e^{is}-1)} e^{(K-L_{D_n}(r))(e^{is}-1)} \left(\frac{\Gamma(\vartheta)}{\Gamma(e^{is}\vartheta)} + O\left(\frac{d_n}{n}\right)\right) \quad (4.3)$$

uniformly in s *for bounded* $\mathrm{Re}(is)$.

We will not give the proof here since it is quite similar to that of the previous theorem. However, an analogue result, Theorem 5.1, is proved in Section 5.

Given the characteristic function of the total cycle number, one can show the following central limit theorem, in analogy to Theorem 4.2 in [15].

Corollary 4.4. *Under the same assumptions as in Theorem 4.2, we have*

$$\frac{T - \vartheta \log n}{\sqrt{\vartheta \log n}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution and* $\mathcal{N}(0,1)$ *a standard normal random variable.*

We will state and prove in Section 5 a similar result, see Corollary 5.4. We thus omit the proof here. In fact, still in analogy to [15], it follows immediately from equation (4.3) that T converges in a stronger sense, namely it is mod-Poisson convergent.

Corollary 4.5. *Under the same assumptions as in Theorem 4.2, the sequence* $(T_n)_{n\in\mathbb{N}}$ *converges in the strong mod-Poisson sense with parameter* $K + \vartheta \log n - L_{D_n}(r)$ *and limiting function* $\Gamma(\vartheta)/\Gamma(\vartheta e^{is})$.

As in [15] one can now approximate T_n by a Poisson random variable with mean $K + \vartheta \log n - L_{D_n}(t)$ or compute large deviations estimates. But as all computations are completely similar and we thus omit them. However, in Section 5 we state and prove an analogue result, see Corollary 5.2.

4.2. **Behaviour of large cycles.** The goal of this section is to study the asymptotic behaviour of the large cycles. The main result, Theorem 4.6, yields the same asymptotic behaviour as in the Ewens case, see for instance Vershik and Shmidt [16] and Kingman [11].

Let $\ell^{(1)}(\sigma)$ be the length of the longest cycle of $\sigma \in \mathfrak{S}_n$, $\ell^{(2)}(\sigma)$ the length of the second longest cycle and so on. If σ has cycle type $\lambda = (\lambda_1, \lambda_2, \cdots)$ this means that $\ell^{(j)} = \lambda_j$ for $j \in \mathbb{N}$.

Theorem 4.6. *Let $(A_n)_{n \in \mathbb{N}}$ be the defining sets of the measures $\mathbb{P}_\Theta^{(A_n)}$ and define $D_n := \{1, \ldots, n\} \setminus A_n$ and d_n as in (1.2). Suppose that $g_\Theta(t) \in \mathcal{F}(r, \vartheta, K)$, d_n fulfills assumption 3.7 and that for all $b \geq 0$*

$$\left(\frac{\partial}{\partial t}\right)^{b+1} g_\Theta(t) = \frac{\vartheta b!}{r^{b+1}(1 - t/r)^{b+1}}(1 + O(t - r)) \tag{4.4}$$

as $t \to r$. We then have, as $n \to \infty$,

$$\left(\frac{\ell^{(1)}}{n}, \frac{\ell^{(2)}}{n}, \ldots\right) \xrightarrow{d} \mathcal{PD}(\vartheta)$$

where $\mathcal{PD}(\vartheta)$ denotes the Poisson-Dirichlet distribution with parameter ϑ (see [5]).

Proof. Let $\ell_1 = \ell_1(\sigma)$ be the length of the cycle containing 1, $\ell_2 = \ell_2(\sigma)$ containing the least element not contained in the cycle containing 1 and so on. We prove that for each fixed $m \in \mathbb{N}$, as $n \to \infty$,

$$\left(\frac{\ell_1}{n}, \frac{\ell_2}{n - \ell_1}, \ldots, \frac{\ell_m}{n - \sum_{j=1}^{m-1} \ell_j}\right) \xrightarrow{d} (B_1, \ldots, B_m) \tag{4.5}$$

holds, where $B_1, \ldots, B_m$ are independent beta random variables with parameter $(1, \vartheta)$. This result immediately implies the theorem, see for instance [16].

We start with the case $m = 1$. We first compute the distribution of ℓ_1. If $k \in A_n$ is given, then there are $(n-1) \cdots (n-k+1)$ possible cycles of length k containing the element 1, and the choice of such a cycle does not influence the cycle lengths of the remaining cycles. Using the definition of $h_n(A_n)$ and a small computation then gives

$$\mathbb{P}_\Theta^{(A_n)}[\ell_1 = k] = \frac{\theta_k}{n} \frac{h_{n-k}(A_n)}{h_n(A_n)} \mathbb{1}_{\{k \in A_n\}}.$$

We use the Pochhammer symbol $(k)_b = k(k-1) \cdots (k-b+1)$ and get for $b \geq 1$

$$\mathbb{E}_\Theta^{(A_n)}[(\ell_1 - 1)_b] = \frac{1}{n} \sum_{k=b+1}^{n} (k-1)_b \cdot \theta_k \mathbb{1}_{\{k \in A_n\}} \frac{h_{n-k}(A_n)}{h_n(A_n)}.$$

On the other hand we have

$$\left(\frac{\partial}{\partial t}\right)^{b+1} g_\Theta(t) = \sum_{k=b+1}^{\infty} (k-1)_b \cdot \theta_k t^{k-b-1}.$$

This together with the definition of $L_{D_n}(t)$ and Lemma 2.4 gives

$$h_n(A_n)\mathbb{E}_\Theta^{(A_n)}[(\ell_1 - 1)_b] = [t^{n-b-1}]\left[\frac{e^{g_\Theta(t) - L_{D_n}(t)}}{n}\left(\frac{\partial}{\partial t}\right)^{b+1}\left(g_\Theta(t) - L_{D_n}(t)\right)\right].$$

We can now use Corollary 3.5 and 3.6 to compute the asymptotic behaviour of this expression. If follows with Corollary 3.5 and assumption (4.4) that

$$[t^{n-b-1}]\left[\exp\left(g_\Theta(t)-L_{D_n}(t)\right)\left(\frac{\partial}{\partial t}\right)^{b+1}g_\Theta(t)\right]$$

$$= \vartheta b!\frac{n^{\vartheta+b}e^K}{r^n\Gamma(\vartheta+b+1)}\exp\left(-L_{D_n}(r)\right)\left(1+O(d_n/n)\right). \qquad (4.6)$$

We show as next that the remaining part can be neglected with respect to (4.6). We get with (3.4)

$$L_{D_n}(r) = \sum_{k=1}^{d_n}\frac{\theta_k}{k}r^k = O\left(\sum_{k=1}^{d_n}\frac{1}{k}\right) = O(\log d_n)$$

and

$$\left(\frac{\partial}{\partial t}\right)^{b+1}L_{D_n}(r) = \sum_{k=1}^{d_n}(k-1)_b\theta_k r^{k-b-1} = O\left(\sum_{k=1}^{d_n}\binom{k-1}{b}\right) = O\left(\binom{d_n}{b+1}\right)$$

$$= O\left((d_n)^{b+1}\right).$$

Using Corollary 3.6 together with this computations gives

$$[t^{n-b-1}]\left[e^{g_\Theta(t)-L_{D_n}(t)}\left(\frac{\partial}{\partial t}\right)^{b+1}L_{D_n}(t)\right] = O\left(\frac{n^{\vartheta-1}(d_n)^{b+1}\exp\left(-L_{D_n}(r)\right)}{r^n}\right)$$

Comparing this to (4.6), we see that we can neglect it since $d_n = o(n)$. Thus the leading term of $h_n(A_n)\mathbb{E}_\Theta^{(A_n)}[(\ell_1-1)_b]$ comes from (4.6) and combined with the asymptotic behaviour of h_n, see Corollary 3.4, we obtain

$$\mathbb{E}_\Theta^{(A_n)}[(\ell_1-1)_b] = \vartheta n^b\frac{b!\,\Gamma(\vartheta)}{\Gamma(\vartheta+b+1)}\left(1+O\left(\frac{d_n}{n}\right)\right).$$

It follows that

$$\mathbb{E}_\Theta^{(A_n)}\left[\left(\frac{\ell_1}{n}\right)^b\right] = \frac{b!\,\Gamma(\vartheta+1)}{\Gamma(\vartheta+b+1)} = \mathbb{E}\left[B_1^b\right]$$

with B_1 a beta random variable with parameter ϑ. This completes the proof in the case $m=1$.

Equation (4.5) now can be proved for arbitrary m by induction over m. The argumentation is (almost) the same as in the proof of Proposition 5.2 in [3]. One only has to check that

$$\mathbb{P}_{\Theta,n}^{(A_n)}\left[\frac{\ell_{m+1}}{n-\sum_{j=1}^m\ell j}\leq a_{m+1}\Big|\ell=a_1,\ldots,\ell_n=a_m\right]$$

$$= \mathbb{P}_{\Theta,n-\sum_{j=1}^m a_j}^{(A_n)}\left[\frac{\ell_{m+1}}{n-\sum_{j=1}^m a_j}\leq a_{m+1}\right].$$

$\square$

5. A FUNCTIONAL CENTRAL LIMIT THEOREM

The object of this section is to prove that the number of cycles with length not exceeding n^x converges, after normalisation, weakly to the standard Brownian motion with respect to the Skorohod topology. (Details about the Skorohod topology and weak convergence of processes can be found for instance in [5]). Formally, this means we consider the functional

$$B_n(x) := \sum_{m=1}^{\lfloor n^x \rfloor} C_m. \tag{5.1}$$

It was first shown by DeLaurentis and Pittel [6], with respect to the uniform measure on $\mathfrak{S}_n$, that the process

$$\widetilde{B}_n(x) := \frac{B_n(x) - x \log n}{\sqrt{\log n}} \tag{5.2}$$

converges weakly to the standard Brownian motion for $0 \le x \le 1$. A corresponding result for the Ewens measure ($\theta_j = \vartheta$ for all $j \ge 1$) was shown by Hansen [10] and Donelly, Kurtz and Tavarè [7]. For this, $\log n$ in (5.2) needs to be replaced by $\vartheta \log n$. By an appropriate rescaling, we will show in this section the validity of an analogue result for our more general measure $\mathbb{P}_\Theta^{(A_n)}$ with the usual assumptions on the parameters θ_j.

5.1. **Without restriction.** Throughout this subsection we assume no restrictions on the cycle lengths, that is $A_n = \{1, ..., n\}$ in Definition 1.1, and write $\mathbb{E}_\Theta [.]$ instead of $\mathbb{E}_\Theta^{(A_n)} [.]$, $\mathbb{P}_\Theta$ instead of $\mathbb{P}_\Theta^{(A_n)}$ and h_n instead of $h_n(A_n)$. First, we compute the characteristic function of the process given by (5.1).

Theorem 5.1. *Suppose that $g_\Theta(t)$ is in $\mathcal{F}(r, \vartheta, K)$ and let $B_n(.)$ be defined as in (5.1). Then, for any fixed $x \in [0, 1)$, we have*

$$\mathbb{E}_\Theta \left[\exp(is B_n(x)) \right] = \exp \left((e^{is} - 1) L_{D_x}(r) \right) \left(1 + O(n^{x-1}) \right), \tag{5.3}$$

whith $D_x = \{1, ..., \lfloor n^x \rfloor\}$.

Remark. Note that $B_n(1) = T$, and thus Theorem 4.3 states a similar behaviour as in (5.3) for $x = 1$, except that the 1 on the right-hand side in (5.3) is replaced by the quotient $\Gamma(\vartheta)/\Gamma(e^{is}\vartheta)$.

Proof. Consider $B_b := \sum_{k=1}^b C_k$. By Lemma 2.1 we get

$$h_n \mathbb{E}_\Theta \left[\exp(is B_b) \right] = \sum_{\lambda \vdash n} \frac{1}{z_\lambda} \exp \left(\sum_{k=1}^b C_k \right) \prod_{k=1}^{\ell(\lambda)} \theta_{\lambda_k}.$$

Now apply Lemma 2.3 to obtain

$$\sum_{n=0}^{\infty} h_n \mathbb{E}_{\Theta}\left[\exp(isB_b)\right] t^n = \sum_{\lambda} \frac{1}{z_{\lambda}} \left(\prod_{k=1}^{b} \left(\theta_k e^{is}\right)^{C_k} \right) \left(\prod_{k=b+1}^{\infty} \theta_k^{C_k} \right) t^{|\lambda|}$$

$$= \exp\left(\sum_{k=1}^{b} \frac{\theta_k e^{is}}{k} t^k + \sum_{k=b+1}^{\infty} \frac{\theta_k}{k} t^k \right)$$

$$= \exp\left((e^{is} - 1)L_{D_b}(t) + g_{\Theta}(t) \right),$$

where $D_b = \{1, ..., b\}$. Then set $b = \lfloor n^x \rfloor$ and Theorem 3.3 gives the result.   □

Corollary 5.2. *Let $g_{\Theta}(t)$ and $B_n(.)$ be as in Theorem 5.1. Then, for any fixed $x \in [0,1)$, the sequence $(B_n(x))_{n\in\mathbb{N}}$ is strongly mod-Poisson convergent with limiting function 1 and parameter $L_{D_x}(r)$.*

This corollary follows immediately from (5.3). Note again that a similar result for $x = 1$ can be found in Corollary 4.5. For the definition and details of mod-convergence we refer to [2].

We obtain from (3.4) that

$$L_{D_x}(r) = \sum_{m=1}^{\lfloor n^x \rfloor} \frac{\vartheta}{m} + \sum_{m=1}^{\lfloor n^x \rfloor} \frac{\epsilon_m}{m} = x\vartheta \log n + c + o(1) \qquad (5.4)$$

as $n \to \infty$ with some $c \in \mathbb{R}$. This shows that the mod-Poisson convergence in Corollary 5.3 does also hold with parameter $x\vartheta \log n + c$. Given this, we can estimate the distance of $B_n(x)$ and a Poisson random variable with mean $x\vartheta \log n + c$, analogously to Lemma 4.6 in [15]. This is done in terms of the *point metric d_{loc}* and the *Kolmogorov distance d_K*.

Corollary 5.3. *Let $g_{\Theta}(t)$ and $B_n(.)$ be as in Theorem 5.1 and let P_{ϑ} be a Poisson distributed random variable with mean $x\vartheta \log n + c$. Then, for any fixed $x \in [0,1)$,*

$$d_{loc}\left(B_n(x), P_{\vartheta}\right) = O\left(\frac{n^{x-1}}{\log n} \right) \quad and \quad d_K\left(B_n(x), P_{\vartheta}\right) = O\left(\frac{n^{x-1}}{\log n} \right).$$

Proof. These estimates can be established with Proposition 3.1 and Corollary 3.2 in [2] with $\chi(s) = \exp((e^{is}-1)(x\vartheta \log n + \gamma + c))$, $\psi_{\nu}(s) = 1$ and $\psi_{\mu}(s) = 1$.   □

Another consequence of Theorem 5.1 is the following central limit result.

Corollary 5.4. *Let $g_{\Theta}(t)$ and $B_n(.)$ be as in Theorem 5.1. Then, for any fixed $x \in [0,1]$,*

$$\widetilde{B}_n(x) := \frac{B_n(x) - x\vartheta \log n}{\sqrt{\vartheta \log n}} \xrightarrow{d} \mathcal{N}(0, x),$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\mathcal{N}(0, x)$ a centred Gaussian random variable with variance x.

Proof. The case $x = 1$ follows immediately from Corollary 4.4 and we can thus assume $x < 1$. We know from (5.4) that $L_{D_x}(r) = x\vartheta \log n + O(1)$, as $n \to \infty$. We obtain with (5.3)

$$\mathbb{E}_\Theta \left[\exp \left(\frac{is}{\sqrt{\vartheta \log n}} B_n(x) \right) \right] = \exp \left(\left(e^{\frac{is}{\sqrt{\vartheta \log n}}} - 1 \right) x\vartheta \log n \right) (1 + o(1))$$

$$= \exp \left(isx\sqrt{\vartheta \log n} - \frac{s^2 x}{2} \right) (1 + o(1))$$

and the result follows. $\qquad\square$

We now turn to the main result of this section. As already mentioned in the beginning of this section, it was shown that the process $\widetilde{B}_n(.)$ given in (5.2), considered with respect to the uniform measure (and with respect to the Ewens measure, when $\widetilde{B}_n(.)$ is properly rescaled) converges weakly to the standard Brownian motion. In our setting, the analogue statement is the following.

Theorem 5.5. *Suppose that $g_\Theta(t)$ is in $\mathcal{F}(r, \vartheta, K)$ and define*

$$\widetilde{B}_n(x) := \frac{B_n(x) - x\vartheta \log n}{\sqrt{\vartheta \log n}}.$$

Then, as $n \to \infty$ and for $0 \leq x \leq 1$, $\widetilde{B}_n$ converges weakly to the standard Brownian motion $\mathcal{W}$ on $[0, 1]$.

Proof. We will proof this statement following the arguments of Hansen [10]. We first define a process

$$B_n^*(x) := \left(B_n(x) - L_{D_x}(r) \right) / \sqrt{\vartheta \log n} \tag{5.5}$$

with D_x as in Theorem 5.1. It follows that

$$|\widetilde{B}_n(x) - B_n^*(x)| = |L_{D_x}(r) - x\vartheta \log n| / \sqrt{\vartheta \log n} = o(1), \quad \text{as } n \to \infty$$

with $o(1)$ uniform in $x \in [0, 1]$. Therefore, the distance between $\widetilde{B}_n(x)$ and $B_n^*(x)$ is asymptotically vanishing with respect to the Skorohod topology on the space of right-continuous functions with left limits. It is thus sufficient to prove $B_n^* \overset{d}{\longrightarrow} \mathcal{W}$. We will proceed in two steps: first, we will show that the process $B_n^*(x)$ converges to $\mathcal{W}(.)$ in terms of finite-dimensional distributions and then its tightness.

Convergence of the finite dimensional distributions. We have to show that for any $k \in \mathbb{N}$ and $0 \leq x_1 < x_2 < ... < x_k \leq 1$ the random vector $\{B_n^*(x_j)\}_{j=1}^k$ converges in distribution to the vector $\{\mathcal{N}(0, x_j)\}_{j=1}^k$ with independent increments. We know from Corollary 5.4 that $B_n^*(x_j) \overset{d}{\longrightarrow} \mathcal{N}(0, x_j)$ for all $x_j \in [0, 1]$. It remains to show that the increments are independent. We define the sets $D_n^{(j)} := \left\{ \lfloor n^{x_{j-1}} \rfloor + 1, \ldots, \lfloor n^{x_j} \rfloor \right\}$ with $x_0 := 0$ and a straight forward application of Lemma 2.3 gives

$$h_n \mathbb{E}_\Theta \left[e^{\sum_{j=1}^k \left(is_j(B_n(x_j) - B_n(x_{j-1})) \right)} \right] = [t^n] \left[e^{g_\Theta(t) + \sum_{j=1}^k (e^{is_j} - 1) L_{D_n^{(j)}}(t)} \right]. \tag{5.6}$$

We can thus apply Theorem 3.3. The remaining computations are the same as in the proof of Corollary 5.4. Only the case $x_k = 1$ needs further explanation. This is because we have in this case $D_n^{(k)} = \{n^{\lfloor x_{k-1} \rfloor} + 1, \ldots, n\}$ and thus assumption (3.7) is not satisfied. However, we then have

$$L_{D_n^{(k)}}(t) = g_\Theta(t) - L_{D_c}(t) + t^{n+1} f_n(t)$$

with $D_c := \{1, \ldots, \lfloor n^{x_{k-1}} \rfloor\}$ and $f_n(t)$ a holomorphic function around the origin. Inserting this into (5.6), one can see that the term $t^{n+1} f_n(t)$ can be neglected. Hence, Theorem 3.3 does also apply for $x_k = 1$.

Tightness. It remains to prove that process $B_n^*(.)$ is tight. We use the moment condition given in [5, Theorem 15.6]. More precisely, we show that for any $n \geq 0$ and $0 \leq x_1 < x < x_2 \leq 1$,

$$E_\Theta^{B_n^*} := \mathbb{E}_\Theta \left[\left(B_n^*(x) - B_n^*(x_1) \right)^2 \left(B_n^*(x_2) - B_n^*(x) \right)^2 \right] = O\left((x_2 - x_1)^2 \right). \tag{5.7}$$

We start with the identity

$$h_n E_\Theta^{B_n^*} = \frac{[t^n]}{(\vartheta \log n)^2} \left[L_{D_1}(t) L_{D_2}(t) \, e^{g_\Theta(t)} \right], \tag{5.8}$$

where $D_1 = \{\lfloor n^{x_1} \rfloor + 1, \ldots, \lfloor n^x \rfloor\}$, $D_2 = \{\lfloor n^x \rfloor + 1, \ldots, \lfloor n^{x_2} \rfloor\}$. Before we prove (5.8), we complete the proof of the tightness. It follows with Corollary 3.6 that

$$[t^n] \left[L_{D_1}(t) L_{D_2}(t) \, e^{g_\Theta(t)} \right] = O\left(L_{D_1}(r) L_{D_2}(r) \frac{n^{\vartheta - 1}}{r^n} \right) \tag{5.9}$$

and therefore

$$E_\Theta^{B_n^*} = O\left(\frac{L_{D_1}(r) L_{D_2}(r)}{(\log n)^2} \right). \tag{5.10}$$

Then, from equation (3.4) follows

$$L_{D_1}(r) = O\left((x - x_1) \log n \right) \quad \text{and} \quad L_{D_2}(r) = O\left((x_2 - x) \log n \right). \tag{5.11}$$

Finally, this yields

$$E_\Theta^{B_n^*} = O\left((x - x_1)(x_2 - x) \right) = O\left((x_2 - x_1)^2 \right).$$

which completes the proof of (5.7) and proves the tightness.

It remains to prove (5.8). One can try to proceed with Lemma 2.3, but the computations are rather technical. We prefer to follow the idea of Hansen [10]. Therefore, we consider for $0 < t < r$ a product space

$$\Omega_t := \{(k_1, k_2, \ldots) | k_i \text{ is a non-negative integer}\}$$

and a measure $\mathbb{P}_\Theta^t$ on Ω_t given such that the $m-$th coordinate of Ω_t is Poisson distributed with parameter $\theta_m t^m / m$. Then, analogously to [10, Lemma 2.1], we have

$$\mathbb{P}_\Theta^t[v = n] = t^n h_n \exp(-g_\Theta(t)), \tag{5.12}$$

where $v : \Omega_t \to \mathbb{N}$ is defined by $v(k_1, k_2, ...) = \sum_{m=1}^{\infty} mk_m$. We omit the prove of (5.12) since it is line by line the same as the proof of [10, Lemma 2.1], one simply has to replace θ by θ_k and $(1-t)^{-\theta}$ by $\exp(g_\Theta(t))$.

We obtain the following identity between $\mathbb{P}_\Theta^t$ on Ω_t and $\mathbb{P}_\Theta$ on $\mathfrak{S}_n$

$$\mathbb{P}_\Theta^t[(k_1, k_2, ...)|v = n] = \mathbb{P}_\Theta\left[C_1 = k_1, \ldots, C_n = k_n\right]. \qquad (5.13)$$

This gives in analogy to (2) in [10]

$$e^{g_\Theta(t)}\, \mathbb{E}_\Theta^t[\Psi] = \sum_{n=1}^{\infty} h_n \mathbb{E}_\Theta[\Psi_n]t^n + \Psi(0), \qquad (5.14)$$

where Ψ is a function on the space Ω_t and $\Psi_n : \mathfrak{S}_n \to \mathbb{C}$ is defined as $\Psi_n := \Psi(C_1, C_2, \dots)$. Again, we omit the proofs of (5.13) and (5.14) since they are identical to those in [10].

The identity (5.14) is true for all $0 < t < r$ and thus remains valid as formal power series. Hence, we get with (5.5)

$$h_n E_\Theta^{B_n^*} = [t^n]\left[e^{g_\Theta(t)}\, \mathbb{E}_\Theta^t\left[\left(B^*(x) - B^*(x_1)\right)^2\left(B^*(x_2) - B^*(x)\right)^2\right]\right]$$

and

$$B^*(x)(k_1, k_2, \dots) := \frac{1}{\sqrt{\vartheta \log n}} \sum_{m=1}^{\lfloor n^x \rfloor}\left(k_m - \frac{\theta_m t^m}{m}\right).$$

A small calculation using that the fact that the coordinates on Ω_t are independent Poisson distributed completes the proof of (5.8).

$\square$

5.2. **Restricted measure.** In the last subsection we only considered the weak convergence of the process $\widetilde{B}_n(.)$ without restriction of the probability measure, meaning under the condition $A_n = \{1, \ldots, n\}$. Verifying the proof in Subsection 5.1 carefully, one notices that our argumentation is based on the equations (5.6) and (5.8), but they require only minor modifications in case $A_n \neq \{1, \ldots, n\}$. Thus, one can apply the proof of Subsection 5.1 for many possible restrictions A_n (as long as the assumptions of Theorem 3.3 are satisfied). Since the argumentation for all the interesting cases are similar we restrict the investigation to $A_n = \{\lceil n^a \rceil, \ldots, n\}$ with $0 \leq a < 1$. In this case, the characteristic function of $B_n(x)$ for $0 \leq x < 1$ behaves like

$$\mathbb{E}_\Theta^{(A_n)}\left[\exp(isB_n(x))\right] = \exp\left((e^{is} - 1)L_{M_n}(r)\right)\left(1 + O\left(\frac{\max\{n^x, n^a\}}{n}\right)\right),$$

where $M_n = A_n \cap \{1, ..., \lfloor n^x \rfloor\} = \{\lceil n^a \rceil, \ldots, \lfloor n^x \rfloor\}$. We get with (3.4)

$$L_{M_n}(r) = \sum_{m=1}^{\lfloor n^x \rfloor} \frac{\theta_m}{m} r^m = (x - a)\vartheta \log n\, \mathbb{1}_{\{x \geq a\}} + O(1).$$

One can now use the same argumentation as in Section 5.1 to show that

$$\frac{B_n(x) - \max\{x - a, 0\}\,\vartheta \log n}{\sqrt{\vartheta \log n}} \xrightarrow{d} \mathcal{W}_a(x), \qquad (5.15)$$

where $\mathcal{W}_a(x)$ is the continuous process on $[0, 1]$ with

$$\mathcal{W}_a(x) \stackrel{d}{=} \begin{cases} \mathcal{N}(0, x - a) & \text{if } x \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

In other terms, for $A_n = \{\lceil n^a \rceil, \ldots, n\}$, the process defined on the left-hand side of (5.15) converges weakly to a Brownian motion started at $x = a$.

5.3. **Restriction to even and odd cycles.** This subsection is devoted to the asymptotic behaviour of the processes

$$B_n^{(ev)}(x) := \sum_{\substack{1 \leq m \leq n^x \\ m \text{ even}}} C_m \quad \text{and} \quad B_n^{(odd)}(x) := \sum_{\substack{1 \leq m \leq n^x \\ m \text{ odd}}} C_m. \qquad (5.16)$$

For simplicity we assume that we have no restrictions to the cycle lengths, that is $A_n = \{1, \ldots, n\}$ in Definition 1.1. As for the process $B_n(.)$ in Subsection 5.1, we will find an appropriate rescaling for $B_n^{(ev)}(.)$ and $B_n^{(odd)}(.)$ in order to prove joint convergence to the Brownian motion, see Theorem 5.6.

First, we need to compute the characteristic function. For $0 \leq x_1, x_2 \leq 1$ we have

$$h_n \mathbb{E}_\Theta \left[\exp\left(is_1 B_n^{(ev)}(x_1) + is_2 B_n^{(odd)}(x_2) \right) \right] \qquad (5.17)$$

$$= [t^n] \left[\exp\left(g_\Theta(t) + (e^{is_1} - 1)L_{D_n^{(ev)}}(t) + (e^{is_2} - 1)L_{D_n^{(odd)}}(t) \right) \right]$$

with $D_n^{(ev)} = \{m \leq n^{x_1} | m \text{ even}\}$ and $D_n^{(odd)} = \{m \leq n^{x_1} | m \text{ odd}\}$. This is proven by our usual argumentation. We now can apply Theorem 3.3 for $0 \leq x_1, x_2 < 1$ and get

$$\mathbb{E}_\Theta \left[\exp\left(is_1 B_n^{(ev)}(x_1) + is_2 B_n^{(odd)}(x_2) \right) \right]$$

$$= \exp\left((e^{is_1} - 1)L_{D_n^{(ev)}}(r) \right) \exp\left((e^{is_2} - 1)L_{D_n^{(odd)}}(r) \right) \left(1 + O\left(\frac{1}{n} \right) \right).$$

With (3.4) follows that

$$L_{D_n^{(ev)}}(r) = x_1 \frac{\vartheta}{2} \log n + O(1) \quad \text{and} \quad L_{D_n^{(odd)}}(r) = x_2 \frac{\vartheta}{2} \log n + O(1) \qquad (5.18)$$

and therefore we define the rescaled processes

$$\widetilde{B}_n^{(ev)}(x) := \frac{B_n^{(ev)}(x) - x\frac{\vartheta}{2} \log n}{\sqrt{\frac{\vartheta}{2} \log n}} \quad \text{and} \quad \widetilde{B}_n^{(odd)}(x) := \frac{B_n^{(odd)}(x) - x\frac{\vartheta}{2} \log n}{\sqrt{\frac{\vartheta}{2} \log n}}.$$

$$(5.19)$$

Our aim is to prove the following theorem.

Theorem 5.6. *The processes $\widetilde{B}_n^{(ev)}(x)$ and $\widetilde{B}_n^{(odd)}(x)$ converge, as $n \to \infty$, to two independent standard Brownian motions for $0 \leq x \leq 1$.*

Proof. Using the same argumentation as in the proof of Corollary 5.4, we see that for $0 \leq x_1, x_2 < 1$

$$\left(\widetilde{B}_n^{(odd)}(x_1), \widetilde{B}_n^{(odd)}(x_2) \right) \xrightarrow{d} (\mathcal{N}_1, \mathcal{N}_2) , \tag{5.20}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are independent centered Gaussian random variables with variance x_1, x_2, respectively. More interesting and difficult is the behaviour for $x_1 = 1$ and/or $x_2 = 1$. We have for $x_1 = 1$

$$L_{D_n^{(ev)}}(t) = \sum_{\substack{1 \leq m \leq n \\ m \text{ even}}} \frac{\theta_m}{m} t^m = \frac{1}{2} g_\Theta(t) + \frac{1}{2} g_\Theta(-t) - t^{n+1} f_n(t)$$

where $f_n(t)$ is a holomorphic function around 0. We thus get

$$h_n \mathbb{E}_\Theta \left[\exp \left(is_1 B_n^{(ev)}(1) \right) \right] = [t^n] \left[\exp \left(g_\Theta(t) + \frac{e^{is_1} - 1}{2} (g_\Theta(t) + g_\Theta(-t)) \right) \right].$$
$$\tag{5.21}$$

By an analogue argumentation for $x_1 = x_2 = 1$ we obtain

$$h_n \mathbb{E}_\Theta \left[\exp \left(is_1 B_n^{(ev)}(1) + is_2 B_n^{(odd)}(1) \right) \right]$$
$$= [t^n] \left[\exp \left(g_\Theta(t) + \frac{e^{is_1} - 1}{2} (g_\Theta(t) + g_\Theta(-t)) + \frac{e^{is_2} - 1}{2} (g_\Theta(t) - g_\Theta(-t)) \right) \right]$$
$$= [t^n] \left[\exp \left(\frac{e^{is_1} + e^{is_2}}{2} g_\Theta(t) + \frac{e^{is_1} - e^{is_2}}{2} g_\Theta(-t) \right) \right]. \tag{5.22}$$

We cannot apply Theorem 3.3 for (5.21) and (5.22) since the functions in this equations have singularities at the points r and $-r$. However, a modification of Theorem 3.3 applies to this situation, one simply has to replace the curve γ in Figure 2(a) by the curve in Figure 3.
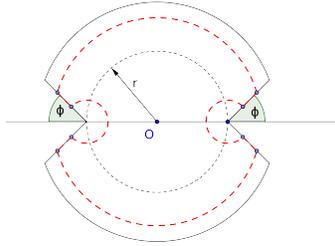

FIGURE 3. The curve γ for two singularities

This gives Theorem 5.7, see below, which we apply for (5.22) if $|s_1|, |s_2| \leq \pi/4$ and obtain

$$\mathbb{E}_\Theta \left[\exp \left(is_1 B_n^{(ev)}(1) + is_2 B_n^{(odd)}(1) \right) \right] = n^{\left(\frac{e^{is_1} + e^{is_2}}{2} - 1 \right)\vartheta} \left(F(s_1, s_2) + \frac{\bar{d}_n}{n} \right)$$

with $F(s_1, s_2)$ a holomorphic function in a neighbourhood of the origin with $F(0, 0) = 1$. Then, the same computation as in the proof of Corollary 5.4 shows that

$$\left(\widetilde{B}_n^{(odd)}(1), \widetilde{B}_n^{(odd)}(1) \right) \xrightarrow{d} (\mathcal{N}_1, \mathcal{N}_2) \tag{5.23}$$

with $\mathcal{N}_1, \mathcal{N}_2$ two independent standard Gaussian random variables.

Finally, it remains to prove that the increments of $\widetilde{B}_n^{(ev)}(.)$ and $\widetilde{B}_n^{(odd)}(.)$ are independent and the tightness of both processes. This argumentations are completely similar to those in Section 5.1, see (5.6) and (5.8), and we thus omit them. $\qquad\square$

Theorem 5.7. *Let $g(t)$ in $\mathcal{F}(r, \vartheta, K)$ be given and $D_n^{(j)}, d_n^{(j)}$ and $\bar{d}_n$ be as in Theorem 3.3. We define*

$$G_n(t, w_1, w_2, v_1, ..., v_k) := \exp \left(w_1 g(t) + w_2 g(-t) + \sum_{j=1}^{k} v_j L_{D_n^{(j)}}(t) \right)$$

with $w_1, w_2, v_1, \ldots, v_k \in \mathbb{C}$ and $L_{D_n^{(j)}}(t)$ as in Lemma 2.4. Suppose further that $\bar{d}_n$ fulfils the assumption (3.7) We then have for each $b \in \mathbb{N}$ fix

$$[t^{n-b}] [G_n(t, w_1, w_2, v_1, ..., v_k)] \tag{5.24}$$

$$= \frac{e^{Kw_1} n^{w_1\vartheta - 1} e^{w_2 g(-r)}}{r^{n-b}} \exp \left(\sum_{j=1}^{k} v_j L_{D_n^{(j)}}(r) \right) \left(\frac{1}{\Gamma(w_1\vartheta)} + O\left(\frac{\bar{d}_n}{n} \right) \right)$$

$$+ O\left(n^{\max\{\text{Re}(w_2), 0\} - 1} \right)$$

uniformly for bounded $|w_1|, |w_2|, |v_1|, ..., |v_k|$ and $\text{Re}(w_1) \geq 0$.

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] A. Barbour, E. Kowalski, and A. Nikeghbali. Mod-discrete expansions. Preprint, 2009.

[3] V. Betz and D. Ueltschi. Spatial random permutations and poisson-dirichlet law of cycle lengths. *Electron. J. Probab.*, 16:no. 41, 1173–1192, 2011.

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

[7] P. Donnelly, T. G. Kurtz, and S. Tavaré. On the functional central limit theorem for the Ewens sampling formula. *Ann. Appl. Probab.*, 1(4):539–545, 1991.

[8] N. Ercolani and D. Ueltschi. Cycle structure of random permutations with cycle weights. Preprint, 2011.

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

[10] J. C. Hansen. A functional central limit theorem for the Ewens sampling formula. *J. Appl. Probab.*, 27(1):28–43, 1990.

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

[12] E. Kowalski and A. Nikeghbali. Mod-Poisson convergence in probability and number theory. *Int. Math. Res. Not. IMRN*, 2010(18):3549–3587, 2010.

[13] 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.

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

[15] A. Nikeghbali and D. Zeindler. The generalized weighted probability measure on the symmetric group and the asymptotic behaviour of the cycles. To appear in Annales de L'Institut Poincaré, 2011.

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

[17] A. L. Yakymiv. A limit theorem for the total number of cycles of a random A-permutation. *Teor. Veroyatn. Primen.*, 52(1):69–83, 2007.

[18] A. L. Yakymiv. Random A-permutations: convergence to a Poisson process. *Mat. Zametki*, 81(6):939–947, 2007.

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERSTRASSE 190, 8057-ZÜRICH, SWITZERLAND

E-mail address: ashkan.nikeghbali@math.uzh.ch

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERSTRASSE 190, 8057-ZÜRICH, SWITZERLAND

E-mail address: julia.storm@math.uzh.ch

DEPARTMENT FÜR MATHEMATIK, SONDERFORSCHUNSGEBREICH 701, UNIVERSITÄT BIELEFELD, BIELEFELD, 33501, DEUTSCHLAND

E-mail address: zeindler@math.uni-bielfeld.de