Random mappings with Ewens cycle structure

Hansen, Jennie Charlotte

Published in:
Ars Combinatoria

Publication date:
2013

Citation for published version (APA):
Hansen, J. C. (2013). Random mappings with Ewens cycle structure. Ars Combinatoria, 112, 307.

Random mappings with Ewens cycle structure

Jennie C. Hansen* and Jerzy Jaworski†‡

Abstract

In this paper we consider a random mapping, $\hat{T}_{n,\theta}$, of the finite set $\{1, 2, ..., n\}$ into itself for which the digraph representation $\hat{G}_{n,\theta}$ is constructed by: (1) selecting a random number, $\hat{L}_n$, of cyclic vertices, (2) constructing a uniform random forest of size n with the selected cyclic vertices as roots, and (3) forming 'cycles' of trees by applying to the selected cyclic vertices a random permutation with cycle structure given by the Ewens sampling formula with parameter θ. We investigate $\hat{k}_{n,\theta}$, the size of a 'typical' component of $\hat{G}_{n,\theta}$, and we obtain the asymptotic distribution of $\hat{k}_{n,\theta}$ conditioned on $\hat{L}_n = m(n)$. As an application of our results, we show in Section 3 that provided $\hat{L}_n$ is of order much larger than $\sqrt{n}$, then the joint distribution of the normalized order statistics of the component sizes of $\hat{G}_{n,\theta}$ converges to the Poisson-Dirichlet(θ) distribution as $n \to \infty$.

1 Introduction

Random mapping models have been extensively studied in the literature. The classical approach to constructing such models, including the well-studied uniform model, is to define a probability measure P_n on $\mathcal{M}_n$, the set of all functions from $[n] \equiv \{1, 2, ..., n\}$ into $[n]$, and to define the random mapping T_{P_n} to be a random element of $\mathcal{M}_n$ with distribution given by P_n. For the model T_{P_n}, it is natural to investigate the structural features of the random directed graph that represents T_{P_n} and much work has been done in this direction (see, for example, [9], [11], [12] and the references therein). In this paper we take a different approach, first introduced in

*Actuarial Mathematics and Statistics Department and The Maxwell Institute for Mathematical Sciences, Heriot–Watt University, Edinburgh EH14 4AS, UK; email: J.Hansen@hw.ac.uk

†Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland; email: jaworski@amu.edu.pl

‡J. Jaworski acknowledges the support by the Marie Curie Intra-European Fellowship No. 236845 (RANDOMAPP) within the 7th European Community Framework Programme and by National Science Centre - DEC-2011/01/B/ST1/03943.

a companion paper [10], to constructing a random mapping model. Our approach is based on building a random mapping model such that certain structural features of the model are specified at the outset.

In order to define the model, we begin with a few definitions. First, we note that for any $f \in \mathcal{M}_n$, we can represent f by a directed graph $G(f)$ on n labelled vertices such that a directed edge from vertex i to vertex j exists in $G(f)$ if and only if $f(i) = j$. For any $f \in \mathcal{M}_n$ and any positive integer ℓ, let $f^{(\ell)}$ denote the ℓ^{th} iterate of f, and for every $i \in [n]$, define $f^{(0)}(i) \equiv i$. We say that vertex i in $G(f)$ is a cyclic vertex if $f^{(\ell)}(i) = i$ for some $\ell \geq 1$ and we let $L(f)$ denote the number of cyclic vertices in $G(f)$. We also note that since each vertex in $G(f)$ has out-degree 1, the components of $G(f)$ consist of directed cycles with directed trees attached. This observation motivates the following construction, first introduced in [10], of a random mapping digraph $\hat{G}_n$ on n labelled vertices: (1) Select a random number, $\hat{L}_n$, of cyclic vertices; (2) Construct a random forest of size n with the selected cyclic vertices as roots; (3) Form 'cycles' of trees by applying a random permutation to the selected cyclic vertices to obtain $\hat{G}_n$. Given the random digraph $\hat{G}_n$, we let $\hat{T}_n$ denote the corresponding random element of $\mathcal{M}_n$ which satisfies $G(\hat{T}_n) = \hat{G}_n$. It is clear from the definition above, that the distribution of $\hat{G}_n$ will depend on the distribution of $\hat{L}_n$, on the distribution of the random forest constructed in Step 2, and on the distribution of the random permutation which is used to form the 'cycles' of trees in Step 3.

Questions concerning the structure of random mappings arise in many applications, such as cryptographic systems (see, for example, [5], [14]), the analysis of Pollard's algorithm (see [13], [16]), simulations of shift register sequences, and computational number theory and random number generation. These applications have motivated our investigation of $\hat{G}_n$. We note that in these applications, and more generally, it is of interest to determine how the structure of $\hat{G}_n$ depends on the number of cyclic vertices $\hat{L}_n$ and on the cycle structure of the random permutation that is used to construct the cycle of trees. In [10] we considered random mapping model $\hat{T}_n$ where both the random forest and the random permutation in the construction of $\hat{G}_n$ are uniformly distributed. For this model we obtained both the exact and the asymptotic distribution of $\hat{k}_n$, the size of a 'typical' component of $\hat{G}_n$, conditioned on the number of cyclic vertices $\hat{L}_n$. It was noted in [10] that it would be interesting to investigate the structure of $\hat{G}_n$ when the random permutation used to construct the cycles is not uniformly distributed. We consider this question in this paper and we extend our earlier results by investigating a more general family of random digraphs, $\hat{G}_{n,\theta}$, which are obtained when the probability distribution of the random permutation of the cyclical vertices is associated with the well-known Ewens sampling formula (see [3]) with parameter $\theta > 0$. Such permutations are defined as follows: Suppose that A is a finite set of size m and let S_A denote the set of all permutations $\sigma : A \to A$. For $\theta > 0$, let $\hat{\sigma}_{A,\theta}$ denote a random element of S_A with distribution given by

$$\Pr\{\hat{\sigma}_{A,\theta} = \sigma\} = \frac{\theta^{|\sigma|}}{\theta^{(m)}} \tag{1.1}$$

where $|\sigma|$ equals the number of cycles in the permutation $\sigma \in S_A$ and $\theta^{(m)} = \theta(\theta+1)(\theta+2)\cdots(\theta+m-1)$. In the case where $\theta = 1$, we obtain the uniform distribution on S_A, so the mappings considered in [10] are a special case of the models that we consider in this paper. Now suppose that for any permutation σ and $i > 0$, we define $C_i(\sigma)$ to be the number of cycles in σ of size i, then it follows from (1.1) that for any finite set A of size m and any non-negative integers $c_1, c_2, ..., c_m$ such that $\sum_{i=1}^{m} ic_i = m$,

$$\Pr\{C_i(\hat{\sigma}_{A,\theta}) = c_i, i = 1, ..., m\} = \frac{m!}{\theta^{(m)}} \prod_{i=1}^{m} \left(\frac{\theta}{i}\right)^{c_i} \frac{1}{c_i!}. \tag{1.2}$$

The formula on the right side of (1.2) is known as the Ewens Sampling Formula from populations genetics and, based on (1.2), we say that the random permutation $\hat{\sigma}_{A,\theta}$ has a Ewens cycle structure. It is also interesting to note that realisations of $\hat{\sigma}_{A,\theta}$ can be generated by the following variant of the sequential Chinese Restaurant Process: Label the elements in A by $1, 2, \ldots, n$ and put the integer 1 in the first cycle. Next, integer 2 joins the first cycle, to the right of 1, with probability $\theta/(\theta + 1)$ or it starts a second cycle. In general, after $k - 1$ insertions, integer k either starts a new cycle with probability $\theta/(\theta + k - 1)$ or it is inserted to the right of a randomly chosen integer that has already been assigned to a cycle.

In order to give a more detailed description of the directed random graph $\hat{G}_{n,\theta}$ and the corresponding random mapping $\hat{T}_{n,\theta}$, we introduce some additional notation. For $n \geq 1$, let V_n denote a set of vertices labelled $1, 2, ..., n$ and let $\hat{L}_n$ denote a discrete random variable such that $1 \leq \hat{L}_n \leq n$. With this notation, we define the random digraph $\hat{G}_{n\,\theta}$ with parameter θ, using $\hat{L}_n$, as follows: *Given* $\hat{L}_n = m$, let $\mathcal{A}_m$ denote a uniform random subset of size m from the vertices V_n (i.e. all subsets of size m are equally likely). *Given* $\mathcal{A}_m = A \subseteq V_n$, let $\mathcal{F}_n(A)$ denote the uniform random rooted forest on the vertices V_n, where A is the set of roots, and the edges in the trees of $\mathcal{F}_n(A)$ are directed such that any path from a vertex to a root is directed towards the root. Finally, suppose that $\hat{\sigma}_{A,\theta}$ is a random permutation of A with Ewens cycle structure which is also *independent* of the random forest $\mathcal{F}_n(A)$. We form the directed graph $\hat{G}_{n,\theta}$ from the rooted forest $\mathcal{F}_n(A)$ by adding a directed edge from $i \in A$ to $j \in A$ if $\hat{\sigma}_{A,\theta}(i) = j$, and we let $\hat{T}_{n,\theta}$

denote the random mapping which is represented by $\hat{G}_{n,\theta}$. It follows from the construction of $\hat{G}_{n,\theta}$ and the definition of $\hat{T}_{n,\theta}$, that for any $f \in \mathcal{M}_n$ and $1 \le m \le n$

$$\Pr\{\hat{T}_{n,\theta} = f \mid \hat{L}_n = m\} = \frac{1}{\binom{n}{m} m n^{n-1-m}} \frac{\theta^{|\pi(f)|}}{\theta^{(m)}} \tag{1.3}$$

where mn^{n-1-m} equals the number of forests on n vertices with m roots and $\pi(f)$ is the permutation that is obtained by restricting the mapping f to its cyclical vertices.

Let $\mathbb{Z}_+$ be the set of non-negative integers. It is clear from the construction above that if $\hat{T}_{n,\theta}$ is a random mapping with Ewens cycle structure, and if $\phi : \mathcal{M}_n \to \mathbb{Z}_+$ is a functional, then the distribution of $\phi(\hat{T}_{n,\theta})$ is determined by the distribution of $\hat{L}_n$ and by the conditional probabilities

$$\Pr\{\phi(\hat{T}_{n,\theta}) = k \,\big|\, \hat{L}_n = m\}\,. \tag{1.4}$$

Hence the conditional probabilities in (1.4) are fundamental to any investigation of $\hat{T}_{n,\theta}$ and of the structure of the corresponding random digraph $\hat{G}_{n,\theta}$. We note that sometimes it can be very easy to compute the conditional probabilities in (1.4). For example, suppose that, for $f \in \mathcal{M}_n$, $\phi(f)$ equals the number of components in $G(f)$, then the conditional distribution of $\phi(\hat{T}_{n,\theta})$ *given* that $\hat{L}_n = m$ is the same as the distribution of the number of cycles in the random permutation $\hat{\sigma}_{A,\theta}$ where $|A| = m$. In this case, both the exact and asymptotic distributions of the number of cycles in $\hat{\sigma}_{A,\theta}$ are well-known (see, [4], [6], in the case $\theta = 1$, and [2], [7] for general $\theta > 0$). However, for most functionals it can be much more complicated to compute the conditional probabilities in (1.4).

In this paper we investigate the conditional distribution of $\hat{k}_{n,\theta} \equiv k_n(\hat{T}_{n,\theta})$ given $\hat{L}_n$, where, for $f \in \mathcal{M}_n$, $k_n(f)$ is defined to be the size of the component in $G(f)$ which contains the vertex 1. Since vertex 1 is 'arbitrary', we can say that $\hat{k}_{n,\theta}$ is the size of a 'typical' component in $\hat{G}_{n,\theta}$. We also note that by selecting a component in $\hat{G}_{n,\theta}$ which contains the vertex 1, we are selecting a component using 'size-biased' sampling. Size-biased sampling has been studied and used in the context of both random mappings and random permutations with Ewens cycle structure (see Aldous [1], Vershik and Schmidt [15] and also the discussion in Hansen and Jaworski [9]). In Section 2, we determine an exact formula for the conditional probabilities

$$\Pr\{\hat{k}_{n,\theta} = k \,\big|\, \hat{L}_n = m\}, \tag{1.5}$$

and we show that conditioned on $\hat{L}_n = m(n)$, where $\sqrt{n} = o(m(n))$, the asymptotic distribution of $\hat{k}_{n,\theta}$ is, in some sense, 'independent' of the number of cyclic vertices as $n \to \infty$. Using this result, we show that provided

$\sqrt{n} = o(\hat{L}_n)$ (in a sense which we make precise in the statement of Theorem 2) and $\hat{\sigma}_{A,\theta}$ is a random permutation with Ewens cycle structure with parameter $\theta > 0$, then the joint distribution of the normalized order statistics of the component sizes of $\hat{G}_{n,\theta}$ converges, as $n \to \infty$, to the $Poisson - Dirichlet(\theta)$ distribution with parameter θ, which we denote by $\mathcal{PD}(\theta)$, on the simplex

$$\nabla = \left\{ \{x_i\} : \sum x_i \leq 1, x_i \geq x_{i+1} \geq 0 \text{ for every } i \geq 1 \right\}.$$

The key point of this result is that we always obtain the *same* limiting distribution for the normalized order statistics of the component sizes provided *only* that, with high probability, the number of cyclic vertices, $\hat{L}_n$, is much larger than $\sqrt{n}$.

We conclude this paper with a few remarks about the structure of $\hat{G}_{n,\theta}$ under other assumptions about the distribution of the number of cyclic vertices $\hat{L}_n$ and we suggest various directions for further investigation.

2 Results

We begin this section by recalling a result from [10] which holds for any random mapping digraph $\hat{G}_n$ which is constructed as described in Section 1.

Fact 1. *Let $\hat{l}_n$ be the length of the cycle in the connected component in $\hat{G}_n$ to which the vertex 1 belongs, $\hat{k}_n$ be the size of this component and let $\hat{L}_n$ be the total number of cyclical vertices of $\hat{G}_n$. Then, for $m = 2, 3, \ldots, n$, $k = 0, 1, \ldots, n-2$ and $j = \max\{1, m-n+k+1\}, \ldots, \min\{k+1, m-1\}$, we have*

$$\Pr\left\{\hat{k}_n = k+1,\, \hat{l}_n = j \,\middle|\, \hat{L}_n = m\right\} =$$

$$\binom{n-m}{k-j+1}\binom{m}{j}\frac{j\,(m-j)\,p_{m,j}}{nm}\left(\frac{k+1}{n}\right)^{k-j+1}\left(1-\frac{k+1}{n}\right)^{n-m-k+j-2},$$

where $p_{m,j}$ denotes the probability that in the random permutation $\hat{\sigma}_m$ of m elements used to construct $\hat{G}_n$, a given j-element set forms a cycle; and for $m = 1, \ldots, n$

$$\Pr\left\{\hat{k}_n = n,\, \hat{l}_n = m \,\middle|\, \hat{L}_n = m\right\} = p_{m,m}$$

and if $j \neq m$

$$\Pr\left\{\hat{k}_n = n,\, \hat{l}_n = j \,\middle|\, \hat{L}_n = m\right\} = 0\,. \qquad \square$$

We now consider the special case where, for $n \geq 1$ and parameter $\theta > 0$, $\hat{G}_{n,\theta}$ is a random mapping with Ewens cycle structure as constructed in Section 1. In this case, an application of Fact 1 yields the following result.

Fact 2. *Let $\hat{l}_{n,\theta}$ be the length of the cycle in the connected component in $\hat{G}_{n,\theta}$ to which the vertex 1 belongs, let $\hat{k}_{n,\theta}$ be the size of this component and let $\hat{L}_n$ be the total number of cyclical vertices of $\hat{G}_{n,\theta}$. Also, for any $x \in R$, let $(x)_j \equiv x(x-1)\cdots(x-j+1)$. Then, for $m = 2, \ldots, n$, $k = 0, 1, \ldots, n-2$ and $j = \max\{1, m-n+k+1\}, \ldots, \min\{k+1, m-1\}$, we have*

$$\Pr\{\hat{k}_{n,\theta} = k+1, \hat{l}_{n,\theta} = j \mid \hat{L}_n = m\}$$
$$= \frac{1}{n}\binom{n-m}{k-j+1} \frac{\theta(m-1)_j}{(m-1+\theta)_j} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2}$$

and for $m = 1, \ldots, n$

$$\Pr\left\{\hat{k}_{n,\theta} = n, \hat{l}_{n,\theta} = m \,\middle|\, \hat{L}_n = m\right\} = \frac{(m-1)!}{(m-1+\theta)_{m-1}}$$

and if $j \neq m$

$$\Pr\left\{\hat{k}_{n,\theta} = n, \hat{l}_{n,\theta} = j \,\middle|\, \hat{L}_n = m\right\} = 0\,.$$

Proof. Fix $\theta > 0$ and suppose that A is a set with $|A| = m$ and let $\hat{\sigma}_{A,\theta}$ be the corresponding random permutation with Ewens cycle structure. Then for $j = 1, 2, \ldots, m$

$$p_{m,j} \;=\; \frac{\theta(j-1)!}{(m-1+\theta)_j}\,. \tag{2.1}$$

where $p_{m,j}$ denotes the probability that in the random permutation $\hat{\sigma}_{A,\theta}$ a given $j-$element set forms a cycle. To see (2.1) let $B \subseteq A$ be a fixed subset of A with $|B| = j$ and let γ_B denote a fixed cyclic permutation of the elements of B. For any $\pi \in S_A$ such that $\pi|_B = \gamma_B$, we can decompose π as (γ_B, ϕ_{B^c}) where $\pi|_{B^c} = \phi_{B^c} \in S_{B^c}$. It follows from the definition of the distribution of $\hat{\sigma}_{A,\theta}$ that

$$\begin{aligned}
\Pr\{\hat{\sigma}_{A,\theta}|_B = \gamma_B\} \;&=\; \sum_{\phi_{B^c} \in S_{B^c}} \Pr\{\hat{\sigma}_{A,\theta} = (\gamma_B, \phi_{B^c})\} \\
&=\; \sum_{\phi_{B^c} \in S_{B^c}} \frac{\theta^{1+|\phi_{B^c}|}}{\theta^{(m)}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\theta}{(\theta+m-1)_j} \sum_{\phi_{B^c} \in S_{B^c}} \frac{\theta^{|\phi_{B^c}|}}{\theta^{(m-j)}} \\
&= \frac{\theta}{(\theta+m-1)_j}
\end{aligned}
$$

where the last equality follows since we are summing up a probability distribution. There are $(j-1)!$ different cyclic permutations of the set B, so it follows from the equation above that

$$
p_{m,j} = (j-1)! \Pr\{\hat{\sigma}_{A,\theta}|_B = \gamma_B\} = \frac{\theta(j-1)!}{(\theta+m-1)_j}
$$

as required. Hence the assertion of the Fact 2 follows directly from the Fact 1. □

Summing the probabilities given by Fact 2 over j leads us to the next result.

Fact 3. *Let $\hat{k}_{n,\theta}$ be the size of the connected component in $\hat{G}_{n,\theta}$ to which the vertex 1 belongs and let $\hat{L}_n$ be the total number of cyclical vertices of $\hat{G}_{n,\theta}$. Then, for $k = 0, \ldots, n-2;\ m = 2, \ldots, n$ we have*

$$
\Pr\{\hat{k}_{n,\theta} = k+1 \,|\, \hat{L}_n = m\} =
$$

$$
\begin{aligned}
&\sum_{j=j^*}^{\min\{k+1,m-1\}} \frac{\theta(m-1)_j}{n(m-1+\theta)_j} \binom{n-m}{k-j+1} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2} \\
&= \sum_{t=\max\{0,k+2-m\}}^{\min\{k,n-m\}} \frac{\theta(m-1)_{k-t+1}}{n(m-1+\theta)_{k-t+1}} \binom{n-m}{t} \left(\frac{k+1}{n}\right)^{t} \left(1 - \frac{k+1}{n}\right)^{n-m-t-1}
\end{aligned}
$$

where $j^* = \max\{1, m-n+k+1\}$, and for $m = 1, \ldots, n$

$$
\Pr\{\hat{k}_{n,\theta} = n \,|\, \hat{L}_n = m\} = \frac{(m-1)!}{(m-1+\theta)_{m-1}}\,.
$$

□

Before stating Theorem 1, we make a few remarks. We recall that when $\theta = 1$, $\hat{G}_{n,1}$ is constructed by using a uniform random permutation to form cycles of trees. This case was considered in detail in [10], and, in particular, the asymptotic distribution of $\hat{k}_{n,1}$ was determined under various assumptions about the number of cyclic vertices m relative to n as $n \to \infty$. In Theorems 1 and 2 below we generalise two results from [10]. We discuss other generalisations in Section 3.

Theorem 1. *Suppose that* $\sqrt{n} = o(m)$*, a is fixed,* $0 < a < \frac{1}{2}$*, and x is such that* $0 < a < x < 1 - a < 1$ *and* $xn \in \mathbb{Z}_+$*. Then, for all sufficiently large n,*

$$\Pr\left\{\frac{\hat{k}_{n,\theta}}{n} = x \,\middle|\, \hat{L}_n = m\right\} = \frac{1}{n}\theta\Big(1-x\Big)^{\theta-1} + \varepsilon(n,m,x,\theta)$$

where $\left|\varepsilon(n,m,x,\theta)\right| \le \frac{C(\theta,a)(n-m)^{1/4}}{n\sqrt{m}}$ *for some constant* $C(\theta,a)$ *that depends only on* θ *and* a*.*

Proof. First, we suppose that $n-m \to \infty$ as $n \to \infty$. We also fix $0 < a < \frac{1}{2}$ and suppose that $k+1 = xn$, where x is fixed, $0 < a < x < 1-a < 1$. Let $\Delta(j) \equiv j - mx$ and let $\delta(n,m,a) \equiv \sqrt{a\,m}(n-m)^{1/4}$. Then by Fact 3, we have

$$\begin{aligned}
&\Pr\{\hat{k}_{n,\theta} = k+1 \,|\, L_n = m\} = S_1 + S_2 + S_3 = \\
&\textstyle\sum_{j=j^*}^{\min\{k+1,m\}} \frac{\mathbf{1}_{\{|\Delta(j)|\le\delta(n,m,a)\}}}{n} \frac{\theta(m)_j}{(m-1+\theta)_j}\binom{n-m}{xn-j} x^{xn-j}(1-x)^{n-m-xn+j} \\
&-\textstyle\sum_{j=j^*}^{\min\{k+1,m\}} \frac{\mathbf{1}_{\{|\Delta(j)|\le\delta(n,m,a)\}}}{n(1-x)} \frac{\theta(m)_j\Delta(j)}{m(m-1+\theta)_j}\binom{n-m}{xn-j} x^{xn-j}(1-x)^{n-m-xn+j} \\
&+\textstyle\sum_{j=j^*}^{\min\{k+1,m\}} \frac{\mathbf{1}_{\{|\Delta(j)|>\delta(n,m,a)\}}}{n(1-x)} \frac{\theta(m)_j(m-j)}{m(m-1+\theta)_j}\binom{n-m}{xn-j} x^{xn-j}(1-x)^{n-m-xn+j}
\end{aligned}$$

where $j^* = \max\{1, m-n+k+1\}$.

We begin by bounding the sum S_3. First note that for $\theta > 0$

$$\frac{\theta(m)_j}{(m-1+\theta)_j} \le \frac{\theta(m-1)_j}{(m-1+\theta)_j}\frac{m}{m-j} \le \theta\frac{m}{m-j}. \tag{2.2}$$

Next, let X be a $Binomial(n-m,x)$ random variable and let

$$Y = (X - (n-m)x)/\sqrt{(n-m)x(1-x)}\,. \tag{2.3}$$

Then it follows from (2.2) that

$$\begin{aligned}
S_3 \;&\le\; \theta \sum_{j=j^*}^{\min\{k+1,m\}} \frac{\mathbf{1}_{\{|\Delta(j)|>\delta(n,m,a)\}}}{n(1-x)}\binom{n-m}{xn-j}x^{xn-j}(1-x)^{n-m-xn+j} \qquad (2.4)\\
&\le\; \frac{\theta}{n(1-x)}\Pr\left\{|Y| > \frac{\delta(n,m,a)}{\sqrt{(n-m)x(1-x)}}\right\} \le \frac{\theta}{n}\left(\frac{\sqrt{n-m}}{a\,m}\right)
\end{aligned}$$

where $j^* = \max\{1, m-n+k+1\}$ and the last inequality follows from Chebyshev's inequality.

Next, observe that if $|\Delta(j)| \le \delta(n,m,a)$, then

$$\left|\frac{\Delta(j)}{m}\right| \le \frac{\sqrt{a}(n-m)^{1/4}}{\sqrt{m}}$$

and, from (2.2), for all sufficiently large m

$$\frac{\theta(m)_j}{(m-1+\theta)_j} \le \theta\frac{m}{m-mx-\delta(n,m,a)} \le \theta\frac{1}{1-x-\delta(n,m,a)/m} \le \frac{2\theta}{a}\,.$$

Using these bounds, we obtain

$$|S_2| \le \frac{2\theta}{na^2}\frac{\sqrt{a}(n-m)^{1/4}}{\sqrt{m}} \le \frac{2\theta(n-m)^{1/4}}{an\sqrt{a\,m}}. \tag{2.5}$$

Finally, suppose that $j = mx + \Delta(j)$ and $|\Delta(j)| \le \delta(n,m,a)$. Then we have, for all sufficiently large m,

$$\begin{aligned}\frac{\theta(m)_j}{(m-1+\theta)_j} &= \theta \exp\left\{-\sum_{i=0}^{j-1}\ln\left(1+\frac{\theta-1}{m-i}\right)\right\}\\ &= \theta \exp\left\{-\sum_{i=0}^{j-1}\frac{\theta-1}{m-i}+\gamma_1(\theta,m,j)\right\}\end{aligned} \tag{2.6}$$

where

$$\gamma_1(\theta,m,j) = \sum_{k=2}^{\infty}\frac{(-1)^{k+1}}{k}\sum_{i=0}^{j-1}\left(\frac{\theta-1}{m-i}\right)^k. \tag{2.7}$$

We note that since $j = mx + \Delta(j)$ and $|\Delta(j)| \le \delta(n,m,a)$, we have

$$|\gamma_1(\theta,m,j)| \le \frac{3j|\theta-1|^2}{2(m-j)^2} \le \frac{3|\theta-1|^2}{ma^2} \tag{2.8}$$

for all sufficiently large m. Furthermore,

$$-\sum_{i=0}^{j-1}\frac{\theta-1}{m-i} = (\theta-1)\ln\left(1-\frac{j}{m}\right)+\gamma_2(m,j)$$

$$= (\theta-1)\ln(1-x)+(\theta-1)\ln\left(1-\frac{\Delta(j)}{m(1-x)}\right)+\gamma_2(m,j). \tag{2.9}$$

where $|\gamma_2(m,j)| \le \frac{1}{m-j} \le \frac{2}{ma}$. It follows from calculations (2.6)-(2.9) that there is a constant $\hat{C}(\theta)$ which depends only on θ such that for j with $|\Delta(j)| \le \delta(n,m,a)$, and all sufficiently large m

$$\frac{\theta(m)_j}{n(m-1+\theta)_j} = \frac{\theta(1-x)^{\theta-1}}{n}(1+\hat{\varepsilon}(n,m,j,\theta)) \tag{2.10}$$

where $|\hat{\varepsilon}(n,m,j,\theta)| \leq \frac{\hat{C}(\theta)(n-m)^{1/4}}{a^2\sqrt{m}}$. It follows from (2.10) and the tail bounds in (2.4) that for all sufficiently large m

$$\begin{aligned} S_1 &= \frac{\theta(1-x)^{\theta-1}}{n} \sum_{j=j^*}^{\min\{k+1,m\}} \mathbf{1}_{\{|\Delta(j)|\leq\delta(n,m,a)\}}(1+\hat{\varepsilon}(n,m,j,\theta)) \times \\ &\qquad\qquad \binom{n-m}{xn-j} x^{xn-j}(1-x)^{n-m-xn+j} \\ &= \frac{\theta(1-x)^{\theta-1}}{n}(1+\tilde{\varepsilon}(n,m,x,\theta)) \end{aligned} \tag{2.11}$$

where $|\tilde{\varepsilon}(n,m,x,\theta)| \leq \frac{\hat{C}(\theta)(n-m)^{1/4}}{a^2\sqrt{m}}$. The result now follows from (2.4), (2.5), and (2.11).

Finally, we consider the case when $n-m=i=O(1)$ and, as previously, fix $0<a<\frac{1}{2}$ and $k+1=xn$, where $0<a<x<1-a<1$. Then we obtain from Fact 3

$$\Pr\{\hat{k}_{n,\theta}=k+1 \,|\, \hat{L}_n=m\} = \sum_{t=0}^{i} \frac{\theta(m-1)_{xn-t}}{n(m-1+\theta)_{xn-t}} \binom{i}{t} x^t(1-x)^{i-t-1}$$

$$= \sum_{t=0}^{i} \frac{\theta(m)_{xn-t}}{n(m-1+\theta)_{xn-t}} \frac{(m-xn+t)}{m(1-x)} \binom{i}{t} x^t(1-x)^{i-t}. \tag{2.12}$$

Since $xn=xm+xi$, we have for $0\leq t\leq i$,

$$\frac{m-xn+t}{m(1-x)} = 1+\frac{t-xi}{m(1-x)}, \tag{2.13}$$

and by (2.10), there is a constant $C(\theta)$ that depends only on θ such that

$$\frac{\theta(m)_{xn-t}}{n(m-1+\theta)_{xn-t}} = \frac{\theta(m)_{xm+xi-t}}{n(m-1+\theta)_{xm+xi-t}} = \frac{\theta(1-x)^{\theta-1}}{n}(1+\hat{\varepsilon}(n,m,i,\theta)) \tag{2.14}$$

where $|\hat{\varepsilon}(n,m,i,\theta)| \leq \frac{C(\theta)(i)^{1/4}}{a^2\sqrt{m}}$ for all sufficiently large m. The result now follows in this case from (2.12)– (2.14).

□

It follows from the above theorem that if $\hat{L}_n$ is of order greater than $\sqrt{n}$ then the conditional distribution of $\frac{k_{n,\theta}}{n}$ given $\hat{L}_n$ is asymptotically $Beta(\theta)$ distributed on $[0,1]$ and is 'independent' of the exact distribution of $\hat{L}_n$. We exploit this 'independence' of $\frac{k_{n,\theta}}{n}$ and $\hat{L}_n$ to prove our next result. To state this result, we define, for $i \geq 1$, $\hat{d}^{(i)}_{n,\theta}$ to be the size of the i^{th} largest component in $\hat{G}_{n,\theta}$. The variables $\left(\frac{\hat{d}^{(1)}_{n,\theta}}{n}, \frac{\hat{d}^{(2)}_{n,\theta}}{n}, \ldots\right)$ are the normalized order statistics of the component sizes of $\hat{G}_{n,\theta}$ and the following result characterises their asymptotic joint distribution.

Theorem 2. *Suppose that $\hat{L}_1, \hat{L}_2, \ldots$ is a sequence of discrete random variables such that for each $n \geq 1$, $1 \leq \hat{L}_n \leq n$. Also, suppose that there exists $m(n)$ such that $\sqrt{n} = o(m(n))$ and $\alpha(n) \equiv \Pr\{\hat{L}_n < m(n)\} \to 0$ as $n \to \infty$. Finally, suppose that $\theta > 0$ is fixed and for $n \geq 1$, $\hat{G}_{n,\theta}$ is a random mapping with Ewens cycle structure. Then the joint distribution of $\left(\frac{\hat{d}^{(1)}_{n,\theta}}{n}, \frac{\hat{d}^{(2)}_{n,\theta}}{n}, \ldots\right)$, converges to the $\mathcal{PD}(\theta)$ distribution on the simplex*

$$\nabla = \left\{ \{x_i\} : \sum x_i \leq 1, x_i \geq x_{i+1} \geq 0 \text{ for every } i \geq 1 \right\}$$

as $n \to \infty$.

We note that the above theorem gives us a continuum of random mapping models based on the value of the parameter θ for which the joint distribution of the order statistics of the normalised component sizes converges to the $\mathcal{PD}(\theta)$ distribution. To describe the main steps in the proof of Theorem 2, we introduce some notation and state a sufficient condition for convergence to the $\mathcal{PD}(\theta)$ distribution on ∇.

First, given $\hat{G}_{n,\theta}$, let $\hat{\mathcal{K}}^{(1)}_{n,\theta}$ denote the component in $\hat{G}_{n,\theta}$ which contains vertex labelled 1. If $\hat{\mathcal{K}}^{(1)}_{n,\theta} \neq \hat{G}_{n,\theta}$, then let $\hat{\mathcal{K}}^{(2)}_{n,\theta}$ denote the component in $\hat{G}_{n,\theta} \setminus \hat{\mathcal{K}}^{(1)}_{n,\theta}$ which contains the smallest vertex; otherwise, set $\hat{\mathcal{K}}^{(2)}_{n,\theta} = \emptyset$. For $i > 2$, we define $\hat{\mathcal{K}}^{(i)}_{n,\theta}$ iteratively: If $\hat{G}_{n,\theta} \setminus (\hat{\mathcal{K}}^{(1)}_{n,\theta} \cup \ldots \cup \hat{\mathcal{K}}^{(i-1)}_{n,\theta}) \neq \emptyset$, then let $\hat{\mathcal{K}}^{(i)}_{n,\theta}$ denote the component in $\hat{G}_{n,\theta} \setminus (\hat{\mathcal{K}}^{(1)}_{n,\theta} \cup \ldots \cup \hat{\mathcal{K}}^{(i-1)}_{n,\theta})$ which contains the smallest vertex; otherwise, set $\hat{\mathcal{K}}^{(i)}_{n,\theta} = \emptyset$. For $i \geq 1$, let $\hat{k}^{(i)}_{n,\theta} = |\hat{\mathcal{K}}^{(i)}_{n,\theta}|$ and define the sequence $(\hat{z}^{(1)}_{n,\theta}, \hat{z}^{(2)}_{n,\theta}, \ldots)$ by

$$\hat{z}^{(1)}_{n,\theta} = \frac{\hat{k}^{(1)}_{n,\theta}}{n},\ \hat{z}^{(2)}_{n,\theta} = \frac{\hat{k}^{(2)}_{n,\theta}}{n - \hat{k}^{(1)}_{n,\theta}}, \ldots,\ \hat{z}^{(i)}_{n,\theta} = \frac{\hat{k}^{(i)}_{n,\theta}}{n - \hat{k}^{(1)}_{n,\theta} - \hat{k}^{(2)}_{n,\theta} - \ldots - \hat{k}^{(i-1)}_{n,\theta}}, \ldots$$

where $\hat{z}^{(i)}_{n,\theta} = 0$ if $n - \hat{k}^{(1)}_{n,\theta} - \hat{k}^{(2)}_{n,\theta} - \ldots - \hat{k}^{(i-1)}_{n,\theta} = 0$.

Now it is well-known (see, for example, Hansen [8] and references therein) that to show that the joint distribution of the normalized order statistics, $\left(\frac{\hat{d}^{(1)}_{n,\theta}}{n}, \frac{\hat{d}^{(2)}_{n,\theta}}{n}, \ldots\right)$, converges to the $\mathcal{PD}(\theta)$ distribution on ∇, it is sufficient to show that for each $t \geq 1$ and $0 < a_i < b_i < 1$, $i = 1, 2, \ldots, t$, we have

$$\lim_{n\to\infty} \Pr\left\{ a_i < \hat{z}^{(i)}_{n,\theta} \leq b_i \,:\, 1 \leq i \leq t \right\} = \prod_{i=1}^{t} \int_{a_1}^{b_i} \theta(1-x)^{\theta-1} dx \,. \tag{2.15}$$

The proof of (2.15) is by induction on t and we give a sketch of this proof below.
Sketch of proof of Theorem 2. First, suppose that $t = 1$ and let $\gamma_1 = \min(a_1, 1 - b_1)$. Then we have

$$\Pr\big\{a_1 \leq \frac{\hat{k}^{(1)}_{n,\theta}}{n} \leq b_1\big\}$$

$$= \sum_{m \geq m(n)} \Pr\left\{a_1 \leq \frac{\hat{k}^{(1)}_{n,\theta}}{n} \leq b_1 \,\Big|\, \hat{L}_n = m\right\} \Pr\{\hat{L}_n = m\big\} + \alpha(a_1, b_1, n) \tag{2.16}$$

where $0 \leq \alpha(a_1, b_1, n) \leq \alpha(n)$. Now it follows from Theorem 1 that

$$\sum_{m \geq m(n)} \Pr\{a_1 \leq \frac{\hat{k}^{(1)}_{n,\theta}}{n} \leq b_1 \,\big|\, \hat{L}_n = m\} \Pr\{\hat{L}_n = m\}$$

$$= \sum_{m \geq m(n)} \left(\int_{a_1}^{b_1} \theta(1-x)^{\theta-1} dx + \varepsilon(a_1, b_1, n, m) \right) \Pr\{\hat{L}_n = m\}$$

$$= \left(\int_{a_1}^{b_1} \theta(1-x)^{\theta-1} dx \right) \Pr\{\hat{L}_m \geq m(n)\} + \sum_{m \geq m(n)} \varepsilon(a_1, b_1, n, m) \Pr\{\hat{L}_n = m\} \tag{2.17}$$

where

$$|\varepsilon(a_1, b_1, n, m)| \leq \frac{C(\theta, \gamma_1)(n - m(n))^{1/4}}{\sqrt{\lambda_1 m(n)}} \,.$$

for all sufficiently large n and $m \geq m(n)$, and some constant $C(\theta, \gamma_1)$ which depends only on θ and γ_1. Since $\Pr\{\hat{L}_n \geq m(n)\} \to 1$ and $\frac{(n-m(n))^{1/4}}{\sqrt{\lambda_1 m(n)}} \to 0$ as $n \to \infty$, the result for $t = 1$ follows from (2.16) and (2.17).

Next, we sketch the induction step by considering the proof of the case $t = 2$ given that the result holds for $t = 1$. First, observe that we can decompose any $f \in \mathcal{M}_n$ as $f = (h_1, h_2)$ where h_1 is the restriction of f to the vertices of $\mathcal{K}^{(1)}(f)$, the component of $G(f)$ that contains vertex 1, and h_2 is the restriction of f to the vertices of $G(f) \setminus \mathcal{K}^{(1)}(f)$. Similarly, $\hat{T}_{n,\theta}$ can be decomposed into a pair of random mappings $\hat{T}^{(1)}_{n,\theta}$ and $\hat{T}^{(1^c)}_{n,\theta}$ where $\hat{T}^{(1)}_{n,\theta}$ is the restriction of $\hat{T}_{n,\theta}$ to the vertices of $\hat{K}^{(1)}_{n,\theta}$ and $\hat{T}^{(1^c)}_{n,\theta}$ is the restriction of $\hat{T}^{(1^c)}_{n,\theta}$ to $\hat{G}_{n,\theta}\backslash\hat{\mathcal{K}}^{(1)}_{n,\theta}$. It follows by arguments like the one given for (1.3) and by other straightforward calculations, that for $1 \leq m \leq n$

$$\Pr\{\hat{T}^{(1^c)}_{n,\theta} = h_2 \mid \hat{T}^{(1)}_{n,\theta} = h_1,\ \hat{L}_n = m\} = \frac{\Pr\{\hat{T}^{(1)}_{n,\theta} = h_1, \hat{T}^{(1^c)}_{n,\theta} = h_2 \mid \hat{L}_n = m\}}{\Pr\{\hat{T}^{(1)}_{n,\theta} = h_1 \mid \hat{L}_n = m\}}$$

$$= \frac{1}{\binom{n-k}{m-l}(m-l)(n-k)^{n-k-m+l-1}} \frac{\theta^{|\pi(h_2)|}}{\theta^{(m-l)}} = \Pr\{\hat{T}_{n-k,\theta} = \tilde{h} \mid \hat{L}_{n-k} = m-l\} \tag{2.18}$$

where k is the number of vertices in the connected digraph $G(h_1)$ which represents h_1, l is the number of cyclical vertices in $G(h_1)$, and $\tilde{h} \in \mathcal{M}_{n-k}$ is the mapping that is obtained by re-numbering the vertices in $G(h_2)$ by $1, 2, ..., n-k$.

As in the case $t = 1$ above, it is enough to consider

$$\Pr\left\{a_1 \leq \frac{\hat{k}^{(1)}_{n,\theta}}{n} \leq b_1, a_2 \leq \frac{\hat{k}^{(2)}_{n,\theta}}{n - \hat{k}^{(1)}_{n,\theta}} \leq b_2 \,\middle|\, \hat{L}_n = m\right\} \tag{2.19}$$

for $m \geq m(n)$. Now for $i \geq 1$, let $\hat{\ell}^{(i)}_{n,\theta}$ denote the number of cyclic vertices in the component $\hat{\mathcal{K}}^{(i)}_{n,\theta}$ (where $\hat{\ell}^{(i)}_{n,\theta} = 0$ if $\hat{\mathcal{K}}^{(i)}_{n,\theta} = \emptyset$). Then it follows from (2.18) that if $\hat{k}^{(1)}_{n,\theta} = xn$ for some $a_1 < x < b_1$ then

$$\Pr\left\{a_2 \leq \frac{\hat{k}^{(2)}_{n,\theta}}{n - xn} \leq b_2 \,\middle|\, \hat{\ell}^{(1)}_{n,\theta} = l, \hat{k}^{(1)}_{n,\theta} = xn, \hat{L}_n = m\right\}$$

$$= \Pr\left\{a_2 \leq \frac{\hat{k}^{(1)}_{n-xn,\theta}}{n - xn} \leq b_2 \,\middle|\, \hat{L}_{n-xn} = m - l\right\}. \tag{2.20}$$

Next, we claim that for $m \geq m(n)$ and all sufficiently large n,

$$\Pr\left\{\hat{\ell}^{(1)}_{n,\theta} \geq \frac{(1+b_1)m}{2} \,\middle|\, \hat{k}^{(1)}_{n,\theta} = xn,\ \hat{L}_n = m\right\} \leq C(\theta, a_1, b_1)\frac{n}{(m(n))^2} \tag{2.21}$$

where $C(\theta, a_1, b_1)$ is a constant that depends on only θ, a_1, and b_1. To prove (2.21), first note that if $m > \frac{2b_1 n}{1+b_1}$ then the left-hand side of (2.21) is 0, so we may assume that $m(n) \leq m \leq \frac{2b_1 n}{1+b_1}$. In this case, by using Fact 2 and the bound (2.2), we obtain, as in the proof of Theorem 1,

$$\Pr\left\{\hat{k}^{(1)}_{n,\theta} = xn,\ \hat{\ell}^{(1)}_{n,\theta} \geq \frac{(1+b_1)m}{2} \,\middle|\, \hat{L}_n = m\right\}$$

$$\leq \frac{\theta}{n(1-x)} \sum_{j \geq \frac{(1+b_1)m}{2}}^{\min(xn, m-1)} \binom{n-m}{xn-j} x^{xn-j}(1-x)^{n-m-xn-j}$$

$$\leq \frac{\theta}{n(1-x)} \Pr\left\{ |Y| > \left(\frac{1-b_1}{2}\right) \frac{m}{\sqrt{(n-m)x(1-x)}} \right\}$$

$$\leq \frac{\theta}{n(1-x)} \frac{n}{(1-b_1)^2(m(n))^2} \tag{2.22}$$

where Y is defined as in (2.3) and the last inequality follows from Chebyshev's inequality. The bound (2.21) now follows from Theorem 1 and (2.22). Furthermore, given $\hat{k}^{(1)}_{n,\theta} = xn$ and $\hat{L}_n = m \geq m(n)$, we obtain from (2.21) that $m - \hat{\ell}^{(1)}_{n,\theta} \geq \frac{(1-b_1)}{2} m(n)$ with (uniformly) high probability. Using this and standard arguments based on Theorem 1 and (2.20), it is straightforward now to obtain

$$\left| \Pr\left\{ a_2 \leq \frac{\hat{k}^{(2)}_{n,\theta}}{n - \hat{k}^{(1)}_{n,\theta}} \leq b_2 \,\middle|\, a_1 \leq \frac{\hat{k}^{(1)}_{n,\theta}}{n} \leq b_1, \hat{L}_n = m \right\} - \int_{a_2}^{b_2} \theta(1-x)^{\theta-1} dx \right|$$

$$\leq \hat{\varepsilon}(a_1, b_1, a_2, b_2, n, m(n)) \tag{2.23}$$

for $m \geq m(n)$, where $\hat{\varepsilon}(a_1, b_1, a_2, b_2, n, m(n)) \to 0$ as $n \to \infty$. The result for $t = 2$ now follows from (2.23), and the result for $t = 1$. The general induction argument is similar to the argument sketched above, but more cumbersome to write down.

□

3 Final Remarks

The results in this paper have been obtained under the assumption that the permutation used to construct $\hat{G}_{n,\theta}$ has a Ewens cycle structure. Theorem 2 tells us that provided $\sqrt{n} = o(\hat{L}_n)$ with high probability, then $\hat{G}_{n,\theta}$ 'inherits', in some sense, its component structure from the Ewens cycle structure of the permutation on its cyclical vertices. Specifically, it is well-known (see [3]) that if $\theta > 0$ is fixed and if $\hat{\sigma}_{m,\theta}$ is a random permutation on m vertices with Ewens cycle structure, then the joint distribution of the normalised order statistics of the cycle lengths of $\hat{\sigma}_{m,\theta}$ converges as $m \to \infty$ to the Poisson-Dirichlet distribution, $\mathcal{PD}(\theta)$. So Theorem 2 establishes that, provided $\sqrt{n} = o(\hat{L}_n)$ with high probability, the normalised order statistics for the component sizes of $\hat{G}_{n,\theta}$ *also* converges, as $n \to \infty$, to the $\mathcal{PD}(\theta)$ distribution. This result follows from Theorem 1 which depends on the local approximation given by (2.10). This suggests, in turn, that if, for a family of random permutations, there is a density $f(x)$ on the interval $(0,1)$ such that $\binom{m}{j} jp_{mj} \approx f(\frac{j}{m})$ as $m \to \infty$, then an analogue of Theorem 1 should hold for the size of a 'typical' component of the random digraph $\hat{G}_n$ which is constructed using the family of random permutations.

Finally, we mention regimes treated in [10] but not presented in this paper. First, recall that if $\theta = 1$, then $\hat{G}_{n,1}$ is obtained by using a uniform random permutation to construct the cycles of trees. In this case, it is known (see [10]) that if $\hat{L}_n = m(n) = o(\sqrt{n})$ then the size of the typical component in $\hat{G}_{n,1}$ is $n - o(n)$. So, provided $\hat{L}_n = o(\sqrt{n})$ with high probability, then $\hat{G}_{n,1}$ has one 'giant' component, independent of the distribution of $\hat{L}_n$, and in this case, it is of more interest to investigate the distribution of $n - \hat{k}_{n,1}$, the number of vertices *not* in the component that contains the vertex 1 (see [10] for further details). In the case where $\hat{L}_n = m(n) = \alpha\sqrt{n}$ for some fixed $\alpha > 0$, the asymptotic distribution of $\frac{\hat{k}_{n,1}}{n}$ is parameterized by α. Thus, if $\hat{L}_n = O(\sqrt{n})$ with high probability, then the asymptotic distribution of $\frac{\hat{k}_{n,1}}{n}$ depends, in an interesting and complicated way, on the distribution of $\hat{L}_n$ as well as on the distribution of the permutation used to form the cycles of trees. For example, if $\hat{L}_n$ equals the number of cyclical vertices in the usual uniform random mapping, then the asymptotic distribution of $\frac{\hat{k}_{n,1}}{n}$ is asymptotically $Beta(1/2)$ whereas the asymptotic distribution of the normalised length of a typical cycle in the underlying uniform random permutation is $Beta(1)$. We note that when the permutation used to construct $\hat{G}_{n,\theta}$ has Ewens cycle structure with $\theta \neq 1$, the results are qualitatively the same as those described above, but the resulting distributions and formulas, especially in the case where $\hat{L}_n = m(n) = \alpha\sqrt{n}$, are more complicated and cumbersome. As suggested in [10], it would be interesting to determine other distributions for $\hat{L}_n$ (which may depend on the parameter θ) such that $\hat{L}_n = O(\sqrt{n})$ with high probability and such that the asymptotic distribution of $\frac{\hat{k}_{n,\theta}}{n}$ is $Beta(\rho(\theta))$ for some specified function $\rho(\theta)$ of θ. We have not been successful in finding such distributions, so this remains as open question.

References

[1] D. Aldous, *Exchangeability and related topics*, Lecture Notes in Math, **1117**, Springer-Verlag, New York (1985).

[2] R. Arratia, D. Stark and S. Tavaré, Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies, *Ann. of Probab.*, **23**, (1995) 1347–1388.

[3] R. Arratia, A. D. Barbour and S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, European Mathematical Society, Zurich (2003).

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

[5] J. M. DeLaurentis, Components and cycles of a random function, *Lecture Notes in Computer Science*, **293**, Advances in Cryptology - CRYPTO'87 Proceedings, (1988) 231–242.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol I, 3rd edition, John Wiley and Sons, New York (1970).

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

[8] J. C. Hansen, Order statistics for decomposable combinatorial structures, *Random Structures and Algorithms*, **5**, (1994) 517–533.

[9] J. C. Hansen and J. Jaworski, A cutting process for random mappings, *Random Structures and Algorithms*, **30**, (2007) 287–306.

[10] J. C. Hansen and J. Jaworski, Random mappings with a given number of cyclical points, *Ars Combinatoria*, **94**, (2010), 341–359.

[11] V. F. Kolchin, *Random Mappings*, Optimization Software Inc., New York (1986).

[12] L. Mutafchiev, On some stochastic problems of discrete mathematics, in: *Mathematics and Education in Mathematics*, Bulg. Akad. Nauk, Sofia, (1984) 57–80.

[13] J. M. Pollard, A Monte Carlo method for factorization, *BIT*, **15(3)**, (1975) 331–334.

[14] J. J. Quisquater and J. P. Delescaille, How easy is collision search? Application to DES, *Lecture Notes in Computer Science*, **434**, Advances in Cryptology - CRYPTO'89 Proceedings, (1990) 429–434.

[15] A. M. Vershik and A. A. Schmidt, Limit measures arising in the asymptotic theory of symmetric groups. I, *Theor. Probab. Appl.*, **22**, (1977) 70–85.

[16] P. C. van Oorschot and M. J. Wiener, Parallel collision search with cryptanalytic applications, *Journal of Cryptology*, **12(1)**, (1999) 1–28.