# Structural transition in random mappings

Jennie C. Hansen

Actuarial Mathematics and Statistics Department
and The Maxwell Institute for Mathematical Sciences
Heriot–Watt University
Edinburgh, UK

J.Hansen@hw.ac.uk

Jerzy Jaworski*

Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Poznań, Poland

jaworski@amu.edu.pl

Submitted: Jul 11, 2013; Accepted: Jan 17, 2014; Published: Jan 24, 2014
Mathematics Subject Classifications: 05C05, 05C80, 60C05, 60G09

**Abstract**

In this paper we characterise the structural transition in random mappings with in-degree restrictions. Specifically, for integers $0 \leqslant r \leqslant n$, we consider a random mapping model $\hat{T}_n^r$ from $[n] = \{1, 2, \ldots, n\}$ into $[n]$ such that $\hat{G}_n^r$, the directed graph on $n$ labelled vertices which represents the mapping $\hat{T}_n^r$, has $r$ vertices that are constrained to have in-degree at most 1 and the remaining vertices have in-degree at most 2. When $r = n$, $\hat{T}_n^r$ is a uniform random permutation and when $r < n$, we can view $\hat{T}_n^r$ as a 'corrupted' permutation. We investigate structural transition in $\hat{G}_n^r$ as we vary the integer parameter $r$ relative to the total number of vertices $n$. We obtain exact and asymptotic distributions for the number of cyclic vertices, the number of components, and the size of the typical component in $\hat{G}_n^r$, and we characterise the dependence of the limiting distributions of these variables on the relationship between the parameters $n$ and $r$ as $n \to \infty$. We show that the number of cyclic vertices in $\hat{G}_n^r$ is $\Theta(\frac{n}{\sqrt{a}})$ and the number of components is $\Theta(\log(\frac{n}{\sqrt{a}}))$ where $a = n - r$. In contrast, provided only that $a = n - r \to \infty$, we show that the asymptotic distribution of the order statistics of the normalised component sizes of $\hat{G}_n^r$ is *always* the *Poisson-Dirichlet(1/2)* distribution as in the case of uniform random mappings with no in-degree restrictions.

*Supported by National Science Centre - DEC-2011/01/B/ST1/03943.

**Keywords:** restricted random mappings; exchangeable in–degrees; anti-preferential attachment; urn schemes; component structure

# 1 Introduction

The motivation for the results in this paper comes from earlier work on the component structure of random mapping models. Random mapping models have been studied since the 1950's and have applications in modelling epidemic processes, the analysis of cryptographic systems (e.g. DES) and of Pollard's algorithm, and random number generation (see, for example, [3, 9, 10, 24, 25, 26] and the references therein). The most extensively studied model is the uniform random mapping $T_n$ from $[n] = \{1, 2, \ldots, n\}$ into $[n]$, with

$$\Pr\{T_n = f\} = \frac{1}{n^n}$$

for any $f \in \mathcal{M}_n$, where $\mathcal{M}_n$ denotes the set of all mappings from $[n]$ into $[n]$. Since any mapping $f \in \mathcal{M}_n$ can be represented by a directed graph $G(f)$ on $n$ labelled vertices such that there is a directed edge from $i$ to $j$ if and only if $f(i) = j$, it is natural to consider the structure of the random digraph $G_n \equiv G(T_n)$ which represents $T_n$. Much is known about the structure of $G_n$ (see, for example, [9, 20]). In particular, Aldous [1] has shown that the joint distribution of the normalized order statistics for the component sizes in $G_n$ converges to the $Poisson-Dirichlet(1/2)$ distribution, denoted $\mathcal{PD}(1/2)$, on the simplex $\nabla = \{\{x_i\} : \sum x_i \leqslant 1, x_i \geqslant x_{i+1} \text{ for every } i \geqslant 1\}$. The component structure of other natural variants of the uniform model have also been studied. For example, a key property of $T_n$ is that the 'vertex image' variables $T_n(1), T_n(2), \ldots, T_n(n)$ are independent and uniformly distributed on $[n]$. So it is natural to consider how the structure of the random mapping digraph changes if we assume that the vertex-image variables are independent but not necessarily either uniform or identically distributed (see [2, 18]). In this case, it is known (see, for example, [5, 23, 28]) that even 'small' perturbations of the distributions of the vertex-image variables can result in a very different asymptotic component structure for the corresponding random mapping digraph.

In another direction, other authors have considered random mappings with structural constraints. This approach is based on the observation that since for any mapping $f \in \mathcal{M}_n$, each vertex in $G(f)$ has out-degree 1, the components of $G(f)$ consist of directed cycles with directed trees attached to the cycles. In the uniform case, the number of cyclic vertices in $G_n$ is $\Theta(\sqrt{n})$ and there are no constraints on the in-degrees of vertices in $G_n$. However, in applications such as the analysis of shift register data, it is natural to consider random mapping digraphs where the in-degree of each vertex is at most $m$, where $m \geqslant 2$ is a fixed integer. Such models were considered by Arney and Bender [3] and, more recently, by the authors in [15]. This later work shows that even in the case $m = 2$, the 'macroscopic' structure of the constrained random mapping digraph remains similar to the structure of $G_n$, e.g. there are still $\Theta(\sqrt{n})$ cyclic vertices in the constrained digraph and the joint distribution of the normalised order statistics of the component sizes still converges to the $\mathcal{PD}(1/2)$ distribution on $\nabla$ as the number of vertices tends to infinity.

In contrast, the asymptotic component structure of random mappings with a constrained number of cyclic vertices, but no in-degree restrictions, can be quite different from the structure of $G_n$ (see [16]). Loosely speaking, such mappings are constructed as follows: First, select a uniform random set of $\ell(n)$ vertices from a set of vertices labelled $1, 2, \ldots, n$. Next, construct a uniform random forest on the $n$ vertices which is rooted at the $\ell(n)$ selected vertices and direct the edges in the forest so that any path from a vertex to the root is directed towards the root. Lastly, complete the construction by constructing a uniform random permutation on the $\ell(n)$ selected vertices. It turns out that the asymptotic structure of random mappings with $\ell(n)$ cyclic vertices (but with no constraints on vertex in-degrees) depends on whether $\sqrt{n} = o(\ell(n))$, $\ell(n) = \Theta(\sqrt{n})$, or $\ell(n) = o(\sqrt{n})$. In particular, if $\sqrt{n} = o(\ell(n))$, then as $n \to \infty$, the joint distribution of the normalised order statistics of the component sizes converges to the $\mathcal{PD}(1)$ distribution rather than the $\mathcal{PD}(1/2)$ distribution. We note that the $\mathcal{PD}(1)$ distribution arises as the limiting distribution of the order statistics for the normalised cycle lengths in a uniform random permutation (see [29]). So the results described above indicate that when the number of cyclic vertices, $\ell(n)$, is much greater than $\sqrt{n}$, then the asymptotic cycle structure of the underlying permutation on the $\ell(n)$ cyclic vertices also determines the relative sizes of the components in the entire random mapping.

In this paper we investigate random mappings with stricter in-degree constraints than those considered by Arney and Bender in [3] and by the authors in [15]. Specifically, we consider random mapping digraphs on $n$ vertices where $r(n)$ vertices are constrained to have in-degree at most 1 and the remaining $n - r(n)$ vertices have in-degree at most 2. These mappings can be viewed as 'corrupted' permutation with $n - r(n)$ 'corrupted' vertices that may have in-degree 2. Note that for such mappings the number of vertices of in-degree 0 is equal to the number of vertices of in-degree 2 and therefore the number of vertices of in-degree 1 is always at least $2r(n) - n = n - 2(n - r(n))$. So, in some sense, the smaller $n - r(n)$ relative to $n$, the more vertices in the mapping are forced to have in-degree 1 and the 'closer' the mapping is to a one-to-one permutation. In this paper, we are interested in characterising how these in-degree constraints influence the graphical structure of the random mapping. For this model we determine (precisely) how the exact and asymptotic cycle and component structure of the digraph depends on the parameter $r(n)$. In particular, we show that as $n \to \infty$, the number of cyclic vertices in the digraph is $\Theta(\frac{n}{\sqrt{n-r(n)}})$. In light of this and the results for random mappings with $\ell(n)$ cyclic vertices described above, one might expect that when $n - r(n) = o(n)$, then the limiting distribution for the normalised order statistics of the component sizes would also converge to the $\mathcal{PD}(1)$ distribution, but this is not the case. In fact, we show that provided $n - r(n) \to \infty$, then *no matter* how slowly $n - r(n)$ grows relative to $n$, the limiting distribution of the order statistics of the normalised component sizes converges to the $\mathcal{PD}(1/2)$ distribution. In other words, in this case, the structure of the permutation on the cyclic vertices of the mapping does *not* determine the relative sizes of the components of the mapping.

The rest of this paper is organised as follows. In Section 2 we give a careful description of our model for random mappings with in-degree constraints and discuss its connection

with models for random mappings with anti-preferential attachment. In Section 3 we obtain the exact distributions of the number of cyclic vertices, the number of components, and the size of a typical component in the digraph which represents the model. In Section 4 we investigate the limiting distributions of the variables considered in Section 3 and we identify how these limiting distributions depend on the relationship between $n$ and $r$ as $n \to \infty$. We also determine the limiting distribution of the normalised order statistics of the component sizes of $\hat{G}_n^r$ as $n \to \infty$. Throughout this paper we adopt the following notational conventions. We write $(X_1, X_2, \ldots, X_k) \overset{d}{\sim} (Y_1, Y_2, \ldots, Y_k)$ when random vectors $(X_1, X_2, \ldots, X_k)$ and $(Y_1, Y_2, \ldots, Y_k)$ have the same joint distribution. Many of the summations in this paper involve products of binomial coefficients and have complicated limits of summation. For such summations, we write $\sum_y$ to denote that the sum is over all values of $y$ such that the binomial coefficients in the sum are defined, and we assume that $\binom{0}{0} = 1$. Finally, we denote the falling factorial by $(n)_k = n(n-1)(n-2)\ldots(n-k+1)$.

# 2 The Model

The model considered in this paper is a natural extension of a model for random mappings with anti-preferential attachment which was first introduced in [15]. Random mappings with anti-preferential attachment can be defined in terms of an urn model as follows: Suppose that $m$ and $n$ are positive integers and suppose that we have an urn such that for each $1 \leqslant k \leqslant n$, the urn contains $m$ balls numbered $k$. We select a sequence of $n$ balls, one at a time, uniformly at random, and without replacement, from the urn and define the random mapping $T_n^m : [n] \to [n]$ by

$$T_n^m(i) = j$$

for $1 \leqslant i, j \leqslant n$, if the ball selected on the $i^{th}$ draw is numbered $j$. This sequential construction of $T_n^m$ can be viewed as a process of anti-preferential attachment in a directed graph on $n$ labelled vertices. Starting with $n$ vertices (and no edges), we add directed edges to the graph as balls are removed from the urn according to the rule that if the ball selected on the $i^{th}$ draw is numbered $j$, then a directed edge from $i$ to $j$ is added to the graph and we set $T_n^m(i) = j$. After $n$ selections from the urn, we obtain the directed graph $G_n^m$ and the corresponding random mapping $T_n^m : [n] \to [n]$. It is the parameter $m$ that determines the strength of the anti-preferential effect in the construction of $T_n^m$. More precisely, the smaller the value of $m$, the stronger the anti-preferential effect. For values of $m$ much larger than $n$, the anti-preferential effect is negligible and $T_n^m$ is essentially equivalent to the uniform random mapping model, whereas when $m = 1$, $T_n^1$ is a uniform permutation on $[n]$.

It is clear from the construction of $T_n^m$ that the in-degree of each vertex in $G_n^m$ is at most $m$, so random mappings with anti-preferential attachment also provide a natural model for random mappings with constrained in-degrees. In particular, we can modify the construction described above to construct random mappings with even stronger constraints on the vertex in-degrees. We refer to this model as the interpolation model

because, in some sense, it is sandwiched 'between' the anti-preferential models $T_n^2$ and $T_n^1$. The model is defined as follows: Suppose that $n > 0$ and $0 \leqslant r \leqslant n$ are integers, and suppose that we have an urn which contains $n$ red balls numbered 1 to $n$ and $n$ blue balls numbered 1 to $n$. The interpolation random mapping $\hat{T}_n^r$ is constructed in two stages:

1. Select a subset of $r$ red balls, uniformly and at random, from the set of red balls and remove these balls from the urn.

2. Select sequence of $n$ balls, one at a time and without replacement, from the urn and define the random mapping $\hat{T}_n^r$ by
$$\hat{T}_n^r(i) = j$$
for $1 \leqslant i, j \leqslant n$, if the ball selected on the $i^{th}$ draw is numbered $j$.

It is clear from the definition of $\hat{T}_n^r$ that if $r = 0$ (i.e. no balls are removed from the urn in the first stage), then $\hat{T}_n^0 = T_n^2$, whereas if $r = n$ (i.e. $n$ balls are removed from the urn in the first stage) then $\hat{T}_n^n = T_n^1$. If $0 < r < n$, then the model $\hat{T}_n^r$ is 'between' the models $T_n^2$ and $T_n^1$, i.e. there are $r$ vertices in the digraph $\hat{G}_n^r \equiv G(\hat{T}_n^r)$ that are constrained to have in-degree at most 1 and the other vertices have in-degree at most 2. So, the larger the value of $r$ relative to $n$, the greater the number of vertices in $\hat{G}_n^r$ with in-degree 1, and, in some sense, the closer the random mapping $\hat{T}_n^r$ is to a random permutation.

Our main goal in this paper is to identify how the component structure of the random digraph $\hat{G}_n^r$ depends on the parameter $r$ and how its structure changes as $\hat{T}_n^r$ gets 'closer' to the random permutation $T_n^1$. The main tool in this investigation is a calculus first developed in [15] for random mappings with exchangeable in-degrees. Random mappings with exchangeable in-degrees can be viewed as an analogue of the well-studied configuration model from random graph theory which was first introduced by Bollobás [6] (see also [21]). Loosely speaking, such mappings are constructed by first specifying the vertex in-degree sequence $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$, where $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ is a sequence of exchangeable, non-negative integer-valued random variables such that $\sum_{i=1}^n \hat{D}_i \equiv n$, and then selecting a mapping $T_n^{\hat{D}}$ uniformly from all mappings with the given in-degree sequence $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. This is a natural model for random mappings where no vertex or set of vertices is considered to be distinguished in some way from the other vertices (i.e. the labelling of the vertices does not matter). One of the most useful and attractive properties of random mappings with exchangeable in-degrees is that many distributions that are related to the structure of the random mapping digraph can be expressed in terms of expected values of functions of the in-degree variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$, which in turn allows us to investigate how the in-degree sequence $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ determines the structure of the digraph.

A natural class of random mappings with exchangeable in-degrees can be constructed as follows: Suppose that $D_1, D_2, \ldots, D_n$ are i.i.d. non-negative integer-valued random variables with $\Pr\{D_1 + D_2 + \ldots + D_n = n\} > 0$ and let $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ be a sequence of random variables with joint distribution is given by
$$\Pr\left\{\hat{D}_i = d_i, 1 \leqslant i \leqslant n\right\} = \Pr\left\{D_i = d_i, 1 \leqslant i \leqslant n \,\middle|\, \sum_{i=1}^n D_i = n\right\}.$$

Clearly, the variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ are exchangeable with $\sum_{i=1}^n \hat{D}_i = n$, and can be used to construct $T_n^{\hat{D}}$. It is easy to check, for example, that if $D_1, D_2, \ldots, D_n$ are i.i.d. Poisson variables, then $T_n^{\hat{D}}$ is the usual uniform random mapping $T_n$. In the case where the variables $D_1, D_2, \ldots, D_n$ have a binomial distribution $Bin(m, p)$, the random mapping $T_n^{\hat{D}}$ corresponds to the anti-preferential model $T_n^m$ described above (for more details, see [15]). In this paper we exploit the fact that the interpolation model $\hat{T}_n^r$ can also be represented as a random mapping with exchangeable in-degrees. However, it is not so easy to extract information about the structure of $\hat{G}_n^r$ because, in this case, the joint distribution of the in-degree sequence for $\hat{T}_n^r$ cannot be represented in terms of a sequence of i.i.d. random variables conditioned on the sum equalling $n$. As a consequence, the exact distribution results obtained in Section 3 for the interpolation model $\hat{T}_n^r$ are more complicated than the analogous results for the anti-preferential model $T_n^m$ and this also complicates the asymptotic analysis in Section 4.

# 3 Exact distributions for $\hat{G}_n^r$

We begin this section with some definitions and additional notation. First, for $n \geqslant 1$ and $f \in \mathcal{M}_n$, we say the vertex labelled $i$ is a cyclic vertex of $f$ if there is some $1 \leqslant k \leqslant n$ such that $f^{(k)}(i) = i$ where $f^{(k)}$ denotes the $k^{th}$ iterate of the mapping $f$. Next, for $1 \leqslant i \leqslant n$, let $d_i(f)$ denote the in-degree of vertex $i$ in the digraph $G(f)$, and let $\vec{d}(f) \equiv (d_1(f), \ldots, d_n(f))$. For any vector $\vec{d} \equiv (d_1, d_2, \ldots, d_n)$ of non-negative integers such that $\sum_{i=1}^n d_i = n$, we also define

$$\mathcal{M}_n(\vec{d}) \equiv \{f \in \mathcal{M}_n : \vec{d}(f) = \vec{d}\}.$$

Finally, for $0 \leqslant r \leqslant n$ and for $1 \leqslant i \leqslant n$, let $\hat{D}_{i,n}^r = d_i(\hat{T}_n^r)$ denote the in-degree of vertex $i$ in the random digraph $\hat{G}_n^r$ which represents $\hat{T}_n^r$. It follows from the definition of $\hat{T}_n^r$ that the variables $\hat{D}_{1,n}^r, \hat{D}_{2,n}^r, \ldots, \hat{D}_{n,n}^r$ are exchangeable and, for $1 \leqslant j \leqslant n$, $\hat{D}_{j,n}^r$ equals the number of balls labelled $j$ that are selected from the urn during Stage 2 of the construction of $\hat{T}_n^r$. It is also straightforward to verify that for any event $\{\hat{D}_{i,n}^r = d_i, \forall i \in [n]\}$ such that $\Pr\{\hat{D}_{i,n}^r = d_i, \forall i \in [n]\} > 0$, we have

$$\Pr\{\hat{T}_n^r = f \mid \hat{D}_{i,n}^r = d_i, \forall i \in [n]\} = \begin{cases} \frac{\prod_{i=1}^n d_i!}{n!} & \text{if} \quad d_i(f) = d_i, \forall i \in [n] \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In other words, *given* $(\hat{D}_{1,n}^r, \hat{D}_{2,n}^r, \ldots, \hat{D}_{n,n}^r) = (d_1, d_2, \ldots, d_n) = \vec{d}$, $\hat{T}_n^r$ is uniformly distributed over $\mathcal{M}_n(\vec{d})$. It follows from (1) that for any $f \in \mathcal{M}_n$,

$$\Pr\{\hat{T}_n^r = f\} = \frac{\prod_{i=1}^n (d_i(f))!}{n!} \Pr\left\{\hat{D}_{i,n}^r = d_i(f), \forall i \in [n]\right\}.$$

We exploit the representation of $\hat{T}_n^r$ as a random mapping with exchangeable in-degrees to prove Theorem 1, which is the main result of this section. This result gives the exact distribution of the number of cyclic vertices, $\hat{X}_n^r$, in $\hat{G}_n^r$, and from this result, it is

straightforward to determine the distributions of the number of components and of the size of a typical component in $\hat{G}_n^r$. These results are given as corollaries of Theorem 1. We also note here that the formula for the distribution of $\hat{X}_n^r$ given in Theorem 1 and the alternative representation of the distribution which is given in Proposition 3 below, may be of independent combinatorial interest.

**Theorem 1.** *Suppose that $n \geqslant 1$ and $0 \leqslant r \leqslant n-1$. Then for $1 \leqslant k \leqslant n$,*

$$\Pr\{\hat{X}_n^r = k\} = \frac{k(n-r)}{(2n-r)_{k+1}} \sum_y 2^{k-y} \binom{k-1}{y} (r)_y (n-1-r)_{k-1-y}. \tag{2}$$

*Proof.* The representation of the interpolation model $\hat{T}_n^r$ as a random mapping with exchangeable in-degrees allows us to use the calculus developed in [15] to investigate the structure of $\hat{G}_n^r$. In particular (see [15]), we have that

$$\Pr\{\hat{X}_n^r = k\} = \frac{k}{n-k} E\left( (\hat{D}_{1,n}^r - 1)\hat{D}_{1,n}^r \hat{D}_{2,n}^r \cdots \hat{D}_{k,n}^r \right) \tag{3}$$

for $1 \leqslant k \leqslant n-1$, and

$$\Pr\{\hat{X}_n^r = n\} = \Pr\{\hat{D}_{i,n}^r = 1, 1 \leqslant i \leqslant n\}. \tag{4}$$

The calculation of the right-hand side of (3) and (4) is complicated by the fact that we cannot represent the joint distribution of $(\hat{D}_{1,n}^r, \hat{D}_{2.n}^r, \ldots, \hat{D}_{n,n}^r)$ in terms of $n$ i.i.d. random variables conditioned on their sum equalling $n$. Instead, to proceed with the proof, it is easier to work with a related sequence of random variables which are defined as follows: Let $n \geqslant 1$, $0 \leqslant r \leqslant n$, and $0 \leqslant w \leqslant 2n-r$ be integers, and suppose that we have an urn with $n$ red balls and $n$ blue balls.

- Step 1: Remove $r$ red balls at random from the urn.
- Step 2: Select a random (unordered) sample of size $w$ from the urn.

Let $\mathcal{S}(r,n,w)$ denote the random sample selected in Step 2 above. Then, for $1 \leqslant j \leqslant n$, we define $D_j(r,n,w)$ to be the number of balls labelled $j$ in $\mathcal{S}(r,n,w)$. In the special case when $w = n$, we will write $\mathcal{S}(r,n) \equiv \mathcal{S}(r,n,n)$ and $D_j(r,n) \equiv D_j(r,n,n)$. We also note that in both the sampling scheme described above and in the urn scheme construction of the random mapping $\hat{T}_n^r$, the first step is to remove $r$ balls from the urn. So, conditioned on the number of red balls $r$ that are removed from the urn in the first step, the sequence $(\hat{D}_{1,n}^r, \hat{D}_{2,n}^r, \ldots, \hat{D}_{n,n}^r)$, which is obtained from an ordered sample of size $n$ without replacement from the urn, and the sequence $(D_1(r,n), \ldots, D_n(r,n))$, which is obtained from an unordered sample of size $n$ without replacement from the urn, have the same conditional distribution. So it follows from the Total Probability Theorem that

$$(\hat{D}_{1,n}^r, \ldots, \hat{D}_{n,n}^r) \overset{d}{\sim} (D_1(r,n), \ldots, D_n(r,n)). \tag{5}$$

It follows from (4) and (5), that

$$\Pr\{\hat{X}_n^r = n\} = \Pr\{D_i(r,n) = 1, 1 \leqslant i \leqslant n\} = \frac{2^{n-r}}{\binom{2n-r}{n}}$$

and this agrees with (2). Next, for $1 \leqslant k \leqslant n-1$, by (5) we have

$$E\left((\hat{D}_{1,n}^r - 1)\hat{D}_{1,n}^r \hat{D}_{2,n}^r \cdots \hat{D}_{k,n}^r\right) = E\left((D_1(r,n) - 1)D_1(r,n)D_2(r,n)\cdots D_k(r,n)\right). \quad (6)$$

We compute the right-hand side of (6) by using a conditioning argument. The events on which we condition are defined as follows. We say that a blue ball labelled $j$ is *lonely* if the red ball labelled $j$ is removed from the urn during Step 1 described above. We note that if the blue ball numbered 1 is lonely, then we must have $D_1(r,n) \leqslant 1$ and

$$(D_1(r,n) - 1)D_1(r,n)D_2(r,n)\cdots D_k(r,n) = 0.$$

Now, suppose that $1 \leqslant k \leqslant n-1$, $0 \vee (r-n+k) \leqslant y \leqslant (k-1) \wedge r$ and $0 \leqslant x \leqslant r-y$. Let $\mathcal{A}_{k,y,x}$ denote the event that

- In Step 1 above, $y$ red balls are removed from those numbered $2, 3, \ldots, k$ and $r-y$ red balls are removed from those numbered $k+1, \ldots, n$.
- In Step 2 above, all the lonely balls with labels in $\{2, \ldots, k\}$ are in $\mathcal{S}(r,n)$ and exactly $x$ of the lonely balls with labels in $\{k+1, \ldots, n\}$ are *not* in $\mathcal{S}(r,n)$.

Straightforward counting yields

$$\Pr\{\mathcal{A}_{k,y,x}\} = \frac{\binom{k-1}{y}\binom{n-k}{r-y}}{\binom{n}{r}} \frac{\binom{r-y}{x}\binom{2n-2r}{n-r+x}}{\binom{2n-r}{n}} = \frac{\binom{k-1}{y}\binom{n-k}{r-y}}{\binom{n}{r}} \frac{\binom{r-y}{x}(n)_{r-x}(n-r)_x}{(2n-r)_r}. \quad (7)$$

Now consider a triple of sets $(V_1, V_2, V_3)$ such that
(i) $V_1 \subseteq \{2, \ldots, k\}$ and $|V_1| = y$, and
(ii) $V_2, V_3 \subseteq \{k+1, \ldots, n\}$ such that $V_2 \cap V_3 = \emptyset$, $|V_2 \cup V_3| = r-y$, $|V_3| = x$.
Given the triple $(V_1, V_2, V_3)$, we define $\mathcal{E}(V_1, V_2, V_3)$ to be the event that the set of lonely blue balls corresponds to the set $V_1 \cup V_2 \cup V_3$ and the set of lonely balls in $\mathcal{S}(r,n)$ corresponds to $V_1 \cup V_2$ (and the lonely blue balls which correspond to the set $V_3$ are *not* in $\mathcal{S}(r,n)$). It is clear from the definition of $\mathcal{A}_{k,y,x}$ that we can partition $\mathcal{A}_{k,y,x}$ by events $\mathcal{E}(V_1, V_2, V_3)$ where $(V_1, V_2, V_3)$ satisfy conditions (i) and (ii) above.

Now suppose that $(V_1, V_2, V_3)$ satisfy conditions (i) and (ii) above, then for every $j \in V_1$ we must have $D_j(r,n) = 1$. It follows that

$$\begin{aligned} &E\left((D_1(r,n) - 1)D_1(r,n)D_2(r,n)\cdots D_k(r,n) \mid \mathcal{E}(V_1, V_2, V_3)\right) \\ &= E\left((D_1(r,n) - 1)D_1(r,n)D_{i_1}(r,n)\cdots D_{i_{k-1-y}}(r,n) \mid \mathcal{E}(V_1, V_2, V_3)\right) \end{aligned}$$

where $\{1, i_1, i_2, \ldots, i_{k-1-y}\} = \{1, 2, \ldots, k\} \setminus V_1$. It is straightforward, by re-labelling the indices in $\{1, 2, \ldots, n\} \setminus (V_1 \cup V_2 \cup V_3)$, to verify that the conditional joint distribution

of $(D_1(r,n), D_{i_1}(r,n), \ldots, D_{i_{k-1-y}}(r,n))$ *given* the event $\mathcal{E}(V_1,V_2,V_3))$ is equal to the joint distribution of $(D_1(0,t,t+x), D_2(0,t,t+x), \ldots, D_{k-y}(0,t,t+x))$ where $t = n-r$. Thus

$$\begin{aligned} &E\left((D_1(r,n)-1)D_1(r,n)D_2(r,n)\cdots D_k(r,n) \mid \mathcal{E}(V_1,V_2,V_3)\right) \\ &= E((D_1(0,t,t+x)-1)D_1(0,t,t+x)D_2(0,t,t+x)\cdots D_{k-y}(0,t,t+x)). \end{aligned} \tag{8}$$

We note that both sides of (8) are equal to 0 when $x > t = n-r$. To evaluate the right side of (8) in the non-trivial cases, we prove:

**Lemma 2.** *For $1 \leqslant \ell \leqslant t$ and $0 \leqslant x \leqslant t$*

$$E((D_1(0,t,t+x)-1)D_1(0,t,t+x)D_2(0,t,t+x)\cdots D_\ell(0,t,t+x)) = \frac{2^\ell \binom{2t-\ell-1}{t+x-\ell-1}}{\binom{2t}{t+x}} \tag{9}$$

*provided the binomial coefficients in (9) are defined. Otherwise, the expected value in (9) is 0.*

*Proof.* We begin by noting that if $\ell = t$ and $x = 0$ then we must have

$$(D_1(0,t,t)-1)D_1(0,t,t)D_2(0,t,t)\cdots D_\ell(0,t,t) = 0$$

since either $D_1(0,t,t) \leqslant 1$ and the product is 0 or $D_1(0,t,t) = 2$ and there is some $1 < j \leqslant t$ such that $D_j(0,t,t) = 0$. So, in this case the expected value in (9) is 0 and the result holds.

Now suppose that $1 \leqslant \ell < t$ or $0 < x \leqslant t$. For $\ell+1 \leqslant s \leqslant \min(t+x, 2\ell)$, define $\Delta(\ell,s) = \{(d_1,d_2,\ldots,d_\ell) : \sum d_i = s, 1 \leqslant d_i \leqslant 2, d_1 = 2\}$, then

$$\begin{aligned} &E\left((D_1(0,t,t+x)-1)D_1(0,t,t+x)D_2(0,t,t+x)\cdots D_\ell(0,t,t+x)\right) \\ &= \sum_{s=\ell+1}^{\min(t+x,2\ell)} \sum_{\vec{d}\in\Delta(\ell,s)} (d_1-1)d_1d_2\cdots d_\ell \times \frac{\binom{2}{d_1}\cdots\binom{2}{d_\ell}\binom{2t-2\ell}{t+x-s}}{\binom{2t}{t+x}} \\ &= 2^\ell \sum_{s=\ell+1}^{\min(t+x,2\ell)} \sum_{\vec{d}\in\Delta(\ell,s)} \frac{\binom{2t-2\ell}{t+x-s}}{\binom{2t}{t+x}} \\ &= 2^\ell \sum_{s=\ell+1}^{\min(t+x,2\ell)} \frac{\binom{\ell-1}{s-\ell-1}\binom{2t-2\ell}{t+x-s}}{\binom{2t}{t+x}} = 2^\ell \frac{\binom{2t-\ell-1}{t+x-\ell-1}}{\binom{2t}{t+x}} \end{aligned}$$

as required. □

We now complete the proof of Theorem 1. It follows from (8) and Lemma 2 that

$$\begin{aligned} E\left((D_1(r,n)-1)D_1(r,n)D_2(r,n)\cdots D_k(r,n) \mid \mathcal{E}(V_1,V_2,V_3)\right) &= \frac{2^{k-y}\binom{2n-2r-k+y-1}{n-r+x-k+y-1}}{\binom{2n-2r}{n-r+x}} \\ &= \frac{2^{k-y}(n-r+x)_{k-y+1}}{(2n-2r)_{k-y+1}} \end{aligned}$$

for *any* event $\mathcal{E}(V_1, V_2, V_3) \subseteq \mathcal{A}_{k,y,x}$ where $V_1, V_2, V_3$ satisfy conditions (i) and (ii) above. Since such events form a partition of $\mathcal{A}_{k,y,x}$, it follows that

$$E\big((D_1(r,n)-1)D_1(r,n)D_2(r,n)\cdots D_k(r,n) \;\big|\; \mathcal{A}_{k,y,x}\big) = \frac{2^{k-y}(n-r+x)_{k-y+1}}{(2n-2r)_{k-y+1}}. \quad (10)$$

So, from (3), (6), (7), and (10), we obtain for $1 \leqslant k < n$

$$\begin{aligned}
&\Pr\{\hat{X}_n^r = k\}\\
&= \frac{k}{n-k}\sum_{y=0\vee(r-n+k)}^{(k-1)\wedge r}\sum_{x=0}^{r-y} E\big((D_1(r,n)-1)\prod_{i=1}^{k} D_i(r,n) \;\big|\; \mathcal{A}_{k,y,x}\big)\Pr\{\mathcal{A}_{k,y,x}\}\\
&= \frac{k}{n-k}\sum_{y}\sum_{x=0}^{r-y} 2^{k-y}\frac{(n-r+x)_{k-y+1}}{(2n-2r)_{k-y+1}}\frac{\binom{k-1}{y}\binom{n-k}{r-y}}{\binom{n}{r}}\frac{\binom{r-y}{x}(n)_{r-x}(n-r)_x}{(2n-r)_r}\\
&= \frac{k}{n-k}\sum_{y} 2^{k-y}\frac{\binom{k-1}{y}\binom{n-k}{r-y}(n)_{k+1}}{\binom{n}{r}(2n-r)_{r+k-y+1}}\sum_{x=0}^{r-y}\binom{r-y}{x}(n-r)_x(n-k-1)_{r-y-x}\\
&= \frac{k}{n-k}\sum_{y} 2^{k-y}\frac{\binom{k-1}{y}\binom{n-k}{r-y}(n)_{k+1}(r-y)!}{\binom{n}{r}(2n-r)_{r+k-y+1}}\binom{2n-r-k-1}{r-y}\\
&= \frac{k(n-r)}{(2n-r)_{k+1}}\sum_{y} 2^{k-y}\binom{k-1}{y}(r)_y(n-1-r)_{k-1-y}\,.
\end{aligned}$$

This completes the proof of Theorem 1. □

The following alternative formula for the distribution of $\hat{X}_n^r$ will also be useful.

**Proposition 3.** *For $0 < r < n$ and $1 \leqslant k \leqslant n$,*

$$\Pr\{\hat{X}_n^r = k\} = \frac{2k(n-r)}{(2n-r)_{k+1}}\sum_{t}\binom{k-1}{t}(n-r-1)_t(n-t-1)_{k-1-t}\,. \quad (11)$$

*Proof.* The proof is combinatorial and is based on the following urn scheme: We start with an urn which contains $r$ white balls, say, $w_1, w_2, \ldots, w_r$ and $2n-2r$ black balls which are matched into $n-r$ pairs: $b_1, b_1^*, b_2, b_2^*, \ldots, b_{n-r}, b_{n-r}^*$ (i.e. there are $2n-r$ balls in the urn). Let $\mathcal{U}_k$ denote the event that in a successive sampling without replacements (i.e. we remove balls one by one) we obtain for the first time on the $k+1^{st}$ draw a black ball which matches a black ball which has already been chosen from the urn. Let $\mathcal{W}_y$ denote the event that in the first $k+1$ draws from the urn we obtain exactly $y$ white balls and let $\mathcal{B}_j$ be the event that we obtain a pair of matched black balls on the $j^{th}$ and $k+1^{st}$ draws. Then for $k = 1, 2, \ldots, n$

$$\begin{aligned}
\Pr\{\mathcal{U}_k\} &= \sum_{j=1}^{k}\sum_{y}\Pr\{\mathcal{U}_k \cap \mathcal{W}_y \cap \mathcal{B}_j\}\\
&= \sum_{j=1}^{k}\sum_{y}\binom{k-1}{y}\frac{(r)_y(n-r)2(n-r-1)_{k-1-y}2^{k-1-y}}{(2n-r)_{k+1}}\,.
\end{aligned}$$

It follows from (2) that

$$\Pr\{\mathcal{U}_k\} = \Pr\{\hat{X}_n^r = k\}\,, \tag{12}$$

and this also confirms that the formula obtained in Theorem 1 describes a proper probability distribution for the number of cyclic vertices.

Now let $\mathcal{B}_t^*$ be the event that in the first $k+1$ successive draws from the urn we select exactly $t+1$ black balls from those marked by $*$. It is again straightforward to verify that for $k = 1, 2, \ldots, n$

$$\begin{aligned}
\Pr\{\mathcal{U}_k\} &= \sum_{j=1}^{k} \sum_{t} \Pr\{\mathcal{U}_k \cap \mathcal{B}_j \cap \mathcal{B}_t^*\} \\
&= \sum_{j=1}^{k} \sum_{t} 2(n-r) \binom{k-1}{t} \frac{(n-r-1)_t (n-t-1)_{k-1-t}}{(2n-r)_{k+1}} \\
&= \frac{2(n-r)k}{(2n-r)_{k+1}} \sum_{t} \binom{k-1}{t} (n-r-1)_t (n-t-1)_{k-1-t}\,.
\end{aligned} \tag{13}$$

Equations (12) and (13) establish (11). We note that these equations also give us a general version of the Karl Goldberg identity as stated in Gould [11] (see (3.21)). □

The next three results are stated as corollaries of Theorem 1 because they all depend on the distribution of $\hat{X}_n^r$, the number of cyclic vertices. We begin by noting that the distribution of $\hat{N}_n^r$, the number of connected components in $\hat{G}_n^r$, is easily determined from the distribution of $\hat{X}_n^r$. This is because each component of $\hat{G}_n^r$ consists of a cycle with trees attached and the restriction of $\hat{T}_n^r$ to the cyclic vertices of $\hat{G}_n^r$ is a uniformly distributed random permutation on the cyclic vertices of $\hat{G}_n^r$. So we obtain:

**Corollary 4.** *Let $\sigma(k)$ is a uniform permutation on $k$ element set and let $N_{\sigma(k)}$ denote the number of cycles in $\sigma(k)$. Then for $0 < r < n$ and $1 \leqslant \ell \leqslant n$*

$$\Pr\{\hat{N}_n^r = \ell\} = \sum_{k=\ell}^{n} \Pr\{N_{\sigma(k)} = \ell\} \Pr\{\hat{X}_n^r = k\} = \sum_{k=\ell}^{n} \frac{|s(k,\ell)|}{k!} \Pr\{\hat{X}_n^r = k\}$$

*where $s(\cdot\,,\cdot)$ are the Stirling numbers of the first kind.*

*Proof.* The proof is based on conditioning on $\hat{X}_n^r$, the number of cyclic vertices in $\hat{G}_n^r$, and uses the well-known fact that there are $|s(k,l)|$ permutations of $k$-element set with exactly $l$ cycles, i.e.,

$$\Pr\{N_{\sigma(k)} = \ell\} = \frac{|s(k,l)|}{k!}.$$

See [15] for further details. □

Next, for $0 < r < n$, let $\mathcal{B}_n^r$ denote the event that $\hat{G}_n^r$ is connected. Then since $\Pr\{\mathcal{B}_n^r\} = \Pr\{\hat{N}_n^r = 1\}$, we obtain from Corollary 4:

**Corollary 5.** *For $0 < r < n$,*

$$\Pr\{\mathcal{B}_n^r\} = \sum_{k=1}^{n} \Pr\{N_{\sigma(k)} = 1\} \Pr\{\hat{X}_n^r = k\} = \sum_{k=1}^{n} \frac{1}{k} \Pr\{\hat{X}_n^r = k\}. \tag{14}$$

Finally, we consider the distribution of the size of a 'typical' component of $\hat{G}_n^r$. For $n > 1$ and $f \in \mathcal{M}_n$, let $\mathcal{C}_1(f)$ denote the set of vertices in the connected component in $G(f)$ which contains the vertex 1 and let $C_1(f) = |\mathcal{C}_1(f)|$ denote the size of the connected component in $G(f)$ that contains vertex 1. Then for $1 \leqslant r < n$, we define $\mathcal{C}_1^r(n) \equiv \mathcal{C}_1(\hat{T}_n^r)$ and $C_1^r(n) \equiv C_1(\hat{T}_n^r)$. The distribution of $C_1^r(n)$ is given by the following result:

**Corollary 6.** *Suppose that $0 < r < n$ and $1 \leqslant k \leqslant n$, then*

$$\Pr\{C_1^r(n) = k\} = \frac{k}{n} \sum_t \Pr\{\mathcal{B}_k^t\} \frac{\binom{2k-t}{k}\binom{2n-2k-r+t}{n-k}\binom{k}{t}\binom{n-k}{r-t}}{\binom{2n-r}{n}\binom{n}{r}}. \tag{15}$$

*Proof.* For $1 \leqslant k \leqslant n$, let $Z_n^r(k)$ denote the number of balls removed from the red balls labelled 1 to $k$ in the first step of the construction of $\hat{T}_n^r$. Then we have

$$\begin{aligned} \Pr\{C_1^r(n) = k\} &= \binom{n-1}{k-1} \Pr\{\mathcal{C}_1^r(n) = [k]\} \\ &= \binom{n-1}{k-1} \sum_t \Pr\left\{\mathcal{C}_1^r(n) = [k] \;\middle|\; Z_n^r(k) = t\right\} \frac{\binom{k}{t}\binom{n-k}{r-t}}{\binom{n}{r}} \end{aligned} \tag{16}$$

where the sum above is over all values of $t$ for which the binomial coefficients in the sum are defined. Now it follows from the two-step construction of $\hat{T}_n^r$ that

$$\Pr\left\{\mathcal{C}_1^r(n) = [k] \;\middle|\; Z_n^r(k) = t\right\} = \Pr\{\mathcal{B}_k^t\} \frac{(2k-t)_k (2n-2k-r+t)_{n-k}}{(2n-r)_n}.$$

Substituting the above formula into (16), we obtain (15). □

# 4 Asymptotic structure of $\hat{G}_n^r$

In this section we investigate the limiting distributions of the variables considered in Section 3 and identify how these limiting distributions depend on the relationship between $n$ and $r$ as $n \to \infty$. The results are stated as local limit theorems with error bounds. Keeping track of the errors in the asymptotic calculations is a little tedious, but the bounds are needed in the proofs of results later in this section. We begin by obtaining the asymptotic distribution of $\hat{X}_n^r$, the number of cyclic vertices. There are two distinct cases which correspond to the following regimes: (i) $a = n - r \to \infty$ as $n \to \infty$, and (ii) $a = n - r$ is fixed as $n \to \infty$.

**Theorem 7.** *(i) Suppose that $a = n-r$ and that $a \to \infty$ as $n \to \infty$. Then for $k = \lfloor \frac{n+a}{\sqrt{a}} x \rfloor$ where $a^{-1/32} < x < a^{1/32}$,*

$$\Pr\{\hat{X}_n^r = k\} = \frac{\sqrt{a}}{n+a} 2x \exp\left(-x^2\right)\left(1 + \epsilon(k,a,n)\right) + \delta(k,a,n)$$

*where $\epsilon(k,a,n)| \leqslant 20a^{-1/10}$ and $|\delta(k,a,n)| < 2a^{-3/5}$.*
*(ii) Suppose that $0 < x < 1$ (and $x$ is fixed). If $r = n - a$ where $a \in Z^+$ is fixed and $k = \lfloor xn \rfloor$, then*

$$\Pr\{\hat{X}_n^r = k\} \sim \frac{1}{n} 2ax\left(1 - x^2\right)^{a-1}.$$

*Proof. Part (i)* Let $T$ denote a hypergeometric random variable as defined in [19] with parameters $\underline{N} = n - 2 + a$, $\underline{n} = k - 1$ and $\underline{p} = \frac{a-1}{n-2+a}$. Recall that $E(T) = \frac{(k-1)(a-1)}{n+a-2}$ and $Var(T) = \frac{(k-1)(a-1)(n-1)(n+a-k-1)}{(n+a-2)^2(n+a-3)}$. By Proposition 3 we have

$$\begin{aligned}\Pr\{\hat{X}_n^r = k\} &= \frac{2ka}{(n+a)_2} \sum_t \binom{k-1}{t} \frac{(a-1)_t (n-1-t)_{k-1-t}}{(n-2+a)_{k-1}} \\ &= \frac{2ka}{(n+a)_2} \sum_t \frac{(n-1-t)_{k-1-t}}{(n-1)_{k-1-t}} \frac{\binom{a-1}{t}\binom{n-1}{k-1-t}}{\binom{n-2+a}{k-1}}. \end{aligned} \tag{17}$$

Let $\gamma(x,a) = x^{2/3}a^{1/3}$, then it follows from Chebyshev's inequality that

$$\begin{aligned}&\frac{2ka}{(n+a)_2} \sum_{\substack{t \ s.t. \\ |t-E(T)|>\gamma(x,a)}} \frac{(n-1-t)_{k-1-t}}{(n-1)_{k-1-t}} \frac{\binom{a-1}{t}\binom{n-1}{k-1-t}}{\binom{n-2+a}{k-1}} \\ &\leqslant \frac{2ka}{(n+a)_2} \Pr\left\{|T - E(T)| > \gamma(x,a)\right\} \\ &\leqslant \frac{2x^{2/3}a^{1/3}}{n} \leqslant 2a^{-3/5}. \end{aligned} \tag{18}$$

Next, suppose that $|t - E(T)| \leqslant \gamma(x,a)$, then routine calculations yield

$$\begin{aligned}\frac{(n-1-t)_{k-1-t}}{(n-1)_{k-1-t}} &= \exp\left(\sum_{j=1}^{k-1-t} \log\left(1 - \frac{t}{n-j}\right)\right) \\ &= \exp\left(\frac{-t(k-1-t)}{n-1} + \epsilon_1(t,k,a,n)\right) \\ &= \exp\left(-x^2 + \epsilon_2(t,k,a,n)\right)\end{aligned}$$

where $|\epsilon_1(t,k,a,n)| < 15/a^{3/8}$ and $|\epsilon_2(t,k,a,n)| < 10/a^{1/10}$ for sufficiently large $a$. It

follows that

$$\begin{aligned}
&\frac{2ka}{(n+a)_2}\sum_{\substack{t\ s.t.\\ |t-E(T)|\leqslant\gamma(x,a)}}\frac{(n-1-t)_{k-1-t}}{(n-1)_{k-1-t}}\frac{\binom{a-1}{t}\binom{n-1}{k-1-t}}{\binom{n-2+a}{k-1}}\\
&=\frac{2ka\exp(-x^2)}{(n+a)_2}\sum_{\substack{t\ s.t.\\ |t-E(T)|\leqslant\gamma(x,a)}}\exp(\epsilon_2(t,k,a,n))\frac{\binom{a-1}{t}\binom{n-1}{k-1-t}}{\binom{n-2+a}{k-1}}\\
&=\frac{2x\exp(-x^2)\sqrt{a}}{n+a}(1+\epsilon(k,a,n)) \qquad (19)
\end{aligned}$$

where $|\epsilon(k,a,n)|<20a^{-1/10}$ for $a$ sufficiently large. The result now follows from (18) and (19).

*Part (ii)* If $k=\lfloor xn\rfloor$ and $a=n-r$ is fixed as $n$ tends to infinity, we can re-write (2) to obtain

$$\begin{aligned}
\Pr\{\hat{X}_n^r=k\} &= \frac{2ka}{(n+a)_{k+1}}\sum_{j=0}^{a-1}(k-1)_j(n-a)_{k-1-j}\binom{a-1}{j}2^j\\
&= \frac{2ka}{(n+a)_2}\sum_{j=0}^{a-1}\binom{a-1}{j}2^j\frac{(n+a-k-1)_{2a-2-j}(k-1)_j}{(n+a-2)_{2a-2}}\\
&\sim \frac{1}{n}2ax\sum_{j=0}^{a-1}\binom{a-1}{j}(2x)^j(1-x)^{2a-2-j}\\
&= \frac{1}{n}2ax(1-x)^{a-1}(1+x)^{a-1}\\
&= \frac{1}{n}2ax(1-x^2)^{a-1}.
\end{aligned}$$

□

A central limit theorem for $\hat{N}_n^r$, the number of components in $\hat{G}_n^r$, follows immediately by standard arguments (see [27]) from Theorem 7 and Corollary 4. Specifically, since the number of components in $\hat{G}_n^r$ equals the number of cycles in the uniform random permutation that is obtained by restricting the mapping $\hat{T}_n^r$ to its cyclic vertices, we can condition on the number of cyclic vertices in $\hat{G}_n^r$ and appeal to the central limit theorem for the number of cycles in a uniform random permutation to obtain:

**Corollary 8.** *Suppose that $n-r=a>0$ as $n\to\infty$, then*

$$\frac{\hat{N}_n^r-\log(n/\sqrt{a})}{\sqrt{\log(n/\sqrt{a})}}\xrightarrow{d}N(0,1)$$

*as $n\to\infty$.*

Next, we consider $\Pr\{\mathcal{B}_n^r\}$, the probability that $\hat{G}_n^r$ is connected. We note that (14) and (17) give us the following crude upper bound:

$$\Pr\{\mathcal{B}_n^r\} \leqslant \frac{2na}{(n+a)_2} \tag{20}$$

where $a = n - r$. The following proposition gives more precise information about the asymptotic behaviour of the probability of connectedness under the two regimes considered in Theorem 7.

**Proposition 9.** *(i) Suppose that $r = n - a$ and that $a \to \infty$ as $n \to \infty$, then*

$$\Pr\{\mathcal{B}_n^r\} = \frac{\sqrt{a\pi}}{(n+a)}\,(1 + \delta(a,n))$$

*where $|\delta(a,n) \leqslant 6a^{-1/32}$ for all $a$ sufficiently large.*
*(ii) If $r = n - a$, where $a > 0$ is a fixed integer, then*

$$\lim_{n\to\infty}(n+a)\Pr\{\mathcal{B}_n^r\} = \frac{2^{2a}}{\binom{2a}{a}}.$$

*Proof. Part (i)* We compute the right-hand side of (14) by dividing the sum into three parts. First we note that it follows from (17) that

$$\begin{aligned}\sum_{k<(n+a)a^{-17/32}} k^{-1}\Pr\{\hat{X}_n^r = k\} &= \sum_{k<(n+a)a^{-17/32}} \frac{2a}{(n+a)_2}\sum_t \frac{(n-1-t)_{k-1-t}}{(n-1)_{k-1-t}}\frac{\binom{a-1}{t}\binom{n-1}{k-1-t}}{\binom{n-2+a}{k-1}} \\ &\leqslant (n+a)a^{-17/32}\frac{2a}{(n+a)_2} = \frac{2a^{15/32}}{(n+a-1)}. \end{aligned} \tag{21}$$

We also have

$$\sum_{k>(n+a)a^{-15/32}} k^{-1}\Pr\{\hat{X}_n^r = k\} \leqslant \frac{a^{15/32}}{(n+a)}\Pr\left\{\hat{X}_n^r > \frac{a^{15/32}}{n+a}\right\} \leqslant \frac{a^{15/32}}{(n+a)}. \tag{22}$$

Finally, for $(n+a)a^{-17/32} \leqslant k \leqslant (n+a)a^{-15/32}$, we can write $k = x(n+a)/\sqrt{a}$ where $a^{-1/32} \leqslant x \leqslant a^{1/32}$, and we obtain from the proof of Theorem 7 part (i)

$$k^{-1}\Pr\{\hat{X}_n^r = k\} = \frac{2a}{(n+a)^2}\big(\exp(-x^2)(1+\epsilon(k,a,n)) + \delta'(k,a,n)\big)$$

where $|\epsilon(k,a,n)| < 20a^{-1/10}$ and $|\delta'(k,a,n)| \leqslant x^{-1/3}a^{-1/6} \leqslant a^{-1/10}$ for all $a$ sufficiently

large. It follows that

$$\begin{aligned}
&\sum_{\frac{(n+a)}{a^{17/32}}\leqslant k\leqslant\frac{(n+a)}{a^{15/32}}} k^{-1}\Pr\{\hat{X}_n^r=k\}\\
&=\frac{\sqrt{a\pi}}{(n+a)}\sum_{\frac{(n+a)}{a^{17/32}}\leqslant k\leqslant\frac{(n+a)}{a^{15/32}}}\frac{2\exp(-\frac{ak^2}{(n+a)^2})}{\sqrt{\pi}}(1+\epsilon(k,a,n))\frac{\sqrt{a}}{(n+a)}\\
&\quad+\frac{2\sqrt{a}}{(n+a)}\sum_{\frac{(n+a)}{a^{17/32}}\leqslant k\leqslant\frac{(n+a)}{a^{15/32}}}\delta'(k,a,n)\frac{\sqrt{a}}{(n+a)}\\
&=\frac{\sqrt{a\pi}}{(n+a)}(1+\hat{\delta}(a,n)) \qquad (23)
\end{aligned}$$

where $\hat{\delta}(a,n)<3a^{-1/32}$. The result now follows from (21), (22), and (23).
*Part (ii)* For each $n>0$, we define the function $f_n(x)$ on the interval $[0,1]$ by

$$f_n(x)=\begin{cases}0 & \text{if } 0\leqslant x<\frac{1}{n}\\ \frac{(n+a)n\Pr\{\hat{X}_n^r=\lfloor xn\rfloor\}}{\lfloor xn\rfloor} & \text{if } \frac{1}{n}\leqslant x\leqslant 1\end{cases}$$

It follows from the definition of $f_n$ and from Theorem 7(ii) and (14) that

$$\int_0^1 f_n(x)dx=(n+a)\Pr\{\mathcal{B}_n^r\}$$

and for any $0<x<1$

$$\lim_{n\to\infty}f_n(x)=2a(1-x^2)^{a-1}.$$

Next, suppose that $\frac{1}{n}\leqslant x\leqslant 1$ and that $\lfloor xn\rfloor=k$, then it follows from Theorem 1 that

$$\begin{aligned}
f_n(x) &= \frac{n(n+a)\Pr\{\hat{X}_n^r=k\}}{k}\\
&= \frac{n(k-1)!a}{(n+a-1)_k}\sum_y\binom{n-a}{y}\binom{a-1}{k-1-y}2^{k-y}\\
&= \frac{n\binom{n-1}{k-1}(k-1)!a}{(n+a-1)_k}\sum_y\frac{\binom{n-a}{y}\binom{a-1}{k-1-y}}{\binom{n-1}{k-1}}2^{k-y}\\
&\leqslant \frac{(n)_k}{(n+a-1)_k}a2^a\leqslant a2^a.
\end{aligned}$$

It follows by dominated convergence that

$$\lim_{n\to\infty}(n+a)\Pr\{\mathcal{B}_n^r\}=\lim_{n\to\infty}\int_0^1 f_n(x)dx=\int_0^1 2a(1-x^2)^{a-1}dx=\frac{2^{2a}}{\binom{2a}{a}}.$$

The last equality follows by integration by parts. □

The next result gives the asymptotic distribution of $\hat{C}_1^r(n)$, the size of a 'typical' component of $\hat{G}_n^r$. As in Theorem 7, the form of the asymptotic distribution will depend on whether $(i)$ $a = n - r \to \infty$ or $(ii)$ $a = n - r$ is constant as $n \to \infty$. We note that in case $(i)$, particular care must be taken to keep track of the error terms in all stages of the calculation because no assumption is made about how fast $a = n - r$ grows relative to $n$.

**Theorem 10.** *Suppose that $0 < u < v < 1$ are fixed.*
*(i) If $a = n - r \to \infty$ as $n \to \infty$, then for sufficiently large $n$ and $k \geqslant 1$ such that $k = \lfloor xn \rfloor$ for some $u < x < v$,*

$$\Pr\{C_1^r(n) = k\} = \frac{1}{2n\sqrt{1 - k/n}} \left(1 + \xi(r, k, n)\right)$$

*where $|\xi(r, k, n)| \leqslant c(u, v)a^{-1/32}$ for all large $a$ and $c(u, v)$ is a constant that depends only on $u$ and $v$.*
*(ii) Suppose that $a = n - r$ is constant as $n \to \infty$ and suppose that $0 < x < 1$ is fixed, then*

$$\Pr\{C_1^r(n) = \lfloor xn \rfloor\} \sim \frac{1}{n} \sum_{b=0}^{a} \binom{2b}{b}^{-1} \binom{a}{b}^2 (2x)^{2b}(1 - x)^{2a-2b}$$

*as $n \to \infty$ .*

*Proof. Part (i)* Throughout this proof we adopt the convention that for any integer $i > 0$, $c_i(u, v)$ denotes a constant that depends only on $u$ and $v$. Now suppose that $n$ is large and that $k = \lfloor xn \rfloor$ for some $u < x < v$. In addition, suppose that $r \leqslant n^{1/4}$. Then for $0 \leqslant t \leqslant r$, Proposition 9(i) yields

$$\Pr\{\mathcal{B}_k^t\} = \frac{\sqrt{(k - t)\pi}}{2k - t}(1 + \delta(k - t, k)) \tag{24}$$

where $\delta(k - t, k) \leqslant c_1(u, v)n^{-1/32} \leqslant c_1(u, v)a^{-1/32}$, and Stirling's formula yields

$$\frac{\binom{2k-t}{k}\binom{2n-2k-r+t}{n-k}}{\binom{2n-r}{n}} = \frac{1}{\sqrt{\pi}} \sqrt{\frac{n}{(k - t)(n - k)}} \left(1 + \gamma(t, r, k, n)\right) \tag{25}$$

where $|\gamma(t, r, k, n)| \leqslant c_2(u, v)r^2/n \leqslant c_2(u, v)/\sqrt{a}$. Substituting (24) and (25) into (15) and summing over $t$, we obtain

$$\Pr\{C_1^r(n) = k\} = \frac{1}{2n\sqrt{1 - k/n}} \left(1 + \xi(r, k, n)\right)$$

where $|\xi(r, k, n)| \leqslant c_3(u, v)a^{-1/32}$.

Now suppose that $r > n^{1/4}$. It is convenient to write $a = \alpha n$, $r = (1 - \alpha)n$ and we note that

$$\alpha(1 - \alpha)n \geqslant \frac{a^{1/4}}{2}\,, \tag{26}$$

whenever $n^{1/4} < r < n$. Next, we re-write the right side of (15) in terms of $a = \alpha n$ to obtain

$$\Pr\{C_1^r(n) = k\} = \frac{k}{n} \sum_{\ell} \Pr\{\mathcal{B}_k^{k-\ell}\} \frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}}$$

where the sum is over all $\ell$ such that the binomial coefficients in the sum are defined. Now recall that $n > 1$ and $k = \lfloor xn \rfloor$ for some $u < x < v$, and let $m = \lfloor \alpha x n \rfloor$, and for $0 \leqslant \ell \leqslant a$, let $\Delta(\ell) = \ell - m$. Then for any $\ell$ such that $|\Delta(\ell)| < (\alpha x n(1-x)(1-\alpha))^{7/12} \equiv \rho(\alpha, x, n)$ we have

$$\begin{aligned} \frac{\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n}{\alpha n}} &= \frac{\binom{xn}{m}\binom{n-xn}{\alpha n-m}}{\binom{n}{\alpha n}} \times \frac{\binom{xn}{m+\Delta(\ell)}}{\binom{xn}{m}} \times \frac{\binom{n-xn}{\alpha n-m-\Delta(\ell)}}{\binom{n-xn}{\alpha n-m}} \\ &= \frac{1}{\sqrt{2\pi\alpha(1-\alpha)x(1-x)n}} \exp\left(\frac{-\Delta^2(\ell)}{2\alpha(1-\alpha)x(1-x)n}\right)(1+\epsilon(\alpha,\ell,k,n)) \end{aligned} \tag{27}$$

where $|\epsilon(\alpha, \ell, k, n)| \leqslant c_4(u, v)(\alpha(1-\alpha)n)^{-1/4} \leqslant 2c_4(u, v)a^{-1/16}$. The last equality in (27) is obtained by a careful application of Stirling's formula to evaluate each term in the product on the right side of (27). Next, for any $\ell$ such that $|\Delta(\ell)| < \rho(\alpha, x, n)$, define $m(\ell) = \lfloor \alpha n \frac{k+\ell}{n+a} \rfloor$ and $\tilde{\Delta}(\ell) = \ell - m(\ell)$. Then similar calculations yield

$$\begin{aligned} \frac{\binom{k+\ell}{\ell}\binom{n+\alpha n-k-\ell}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}} &= \frac{\binom{k+\ell}{m(\ell)}\binom{n+\alpha n-k-\ell}{\alpha n-m(\ell)}}{\binom{n+\alpha n}{\alpha n}} \times \frac{\binom{k+\ell}{m(\ell)+\tilde{\Delta}(\ell)}}{\binom{k+\ell}{m(\ell)}} \times \frac{\binom{n+\alpha n-k-\ell}{\alpha n-m(\ell)-\tilde{\Delta}(\ell)}}{\binom{n+\alpha n-k-\ell}{\alpha n-m(\ell)}} \\ &= \sqrt{\frac{1+\alpha}{2\pi\alpha x(1-x)n}} \exp\left(\frac{-\Delta^2(\ell)}{2\alpha x(1-x)(1+\alpha)n}\right)(1+\tilde{\epsilon}(\alpha,\ell,k,n)) \end{aligned} \tag{28}$$

where $|\tilde{\epsilon}(\alpha, \ell, k, n)| \leqslant c_5(u, v)(\alpha(1-\alpha)n)^{-1/4} \leqslant 2c_5(u, v)a^{-1/16}$. To obtain the last equality in (28), we have also used the fact that

$$\tilde{\Delta}(\ell) = \frac{\Delta(\ell)}{1+\alpha} + \phi(\alpha, \ell, k, n)$$

where $|\phi(\alpha, \ell, k, n)| < 2$. Finally, for any $\ell$ such that $|\Delta(\ell)| < \rho(\alpha, x, n)$, Proposition 9(i) yields

$$\Pr\{\mathcal{B}_k^{k-\ell}\} = \frac{\sqrt{(m+\Delta(\ell))\pi}}{k+\ell}(1+\delta(\ell,k)) = \frac{\sqrt{(\alpha x n)\pi}}{(1+\alpha)xn}(1+\hat{\delta}(\ell,k)) \tag{29}$$

where $|\hat{\delta}(\ell, k)| \leqslant c_6(u, v)(\alpha n)^{-1/32} = c_6(u, v)a^{-1/32}$. It follows now from (27), (28), and

(29) that

$$\begin{aligned}
&\frac{k}{n}\sum_{\substack{\ell\ s.t.\\ |\Delta(\ell)|<\rho(\alpha,x,n)}} \Pr\{\mathcal{B}_k^{k-\ell}\}\frac{\binom{k+\ell}{\ell}\binom{n-k+a-\ell}{a-\ell}\binom{k}{\ell}\binom{n-k}{a-\ell}}{\binom{n+a}{a}\binom{n}{a}}\\
&= \frac{1}{2n\sqrt{1-x}}\frac{1}{\sqrt{\pi\alpha(1-\alpha^2)x(1-x)n}}\times\\
&\qquad \sum_{\substack{\ell\ s.t.\\ |\Delta(\ell)|<\rho(\alpha,x,n)}} \exp\left(\frac{-\Delta^2(\ell)}{\alpha(1-\alpha^2)x(1-x)n}\right)(1+\hat{\epsilon}(\alpha,\ell,k,n))\\
&= \frac{1}{2n\sqrt{1-k/n}}(1+\gamma(\alpha,k,n)) \qquad (30)
\end{aligned}$$

where $|\hat{\epsilon}(\alpha,\ell,k,n)| \leqslant c_7(u,v)a^{-1/32}$ and $|\gamma(\alpha,k,n)| \leqslant c_8(u,v)a^{-1/32}$. Lastly, it follows from (20) that for sufficiently large $a$

$$\begin{aligned}
&\sum_{\substack{\ell\ s.t.\\ |\Delta(\ell)|\geqslant\rho(\alpha,x,n)}} \frac{k\,\Pr\{\mathcal{B}_k^{k-\ell}\}}{n}\frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}}\\
&\leqslant \sum_{\substack{\ell\ s.t.\\ |\Delta(\ell)|\geqslant\rho(\alpha,x,n)}} \frac{2k^2\ell}{n(k+\ell)_2}\frac{\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n}{\alpha n}} \leqslant \sum_{\substack{\ell\ s.t.\\ |\Delta(\ell)|\geqslant\rho(\alpha,x,n)}} \frac{2\ell}{n}\frac{\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n}{\alpha n}}\\
&\leqslant \frac{(\alpha n)^2}{n}\frac{\exp(-(\alpha(1-\alpha)x(1-x)n)^{1/6}/2)}{\sqrt{\alpha(1-\alpha)x(1-x)n}}. \qquad (31)
\end{aligned}$$

The last inequality above follows from (27) and the unimodality of the hypergeometric distribution. Finally, it follows from the lower bound (26) that

$$\frac{(\alpha n)^2}{n}\frac{\exp(-(\alpha(1-\alpha)x(1-x)n)^{1/6}/2)}{\sqrt{\alpha(1-\alpha)x(1-x)n}} \leqslant \frac{c_9(u,v)a^{15/8}}{2n\sqrt{1-k/n}}\exp(-c_{10}(u,v)a^{1/24}). \qquad (32)$$

Part (i), in the case $n^{1/4} < r < n$, now follows from (30) – (32).
*Part (ii)* Now suppose that $0 < x < 1$ and $a = n - r$ is fixed as $n \to \infty$. In the calculation below, let $k'$ denote $\lfloor xn \rfloor$. Then for all sufficiently large $n$, (15) and Proposition 9(ii) yield

$$\begin{aligned}
\Pr\{C_1^r(n) = k'\} &= \frac{k'}{n}\sum_{k'-a\leqslant t\leqslant k'} \Pr\{\mathcal{B}_{k'}^t\}\frac{\binom{2k'-t}{k'}\binom{2n-2k'-r+t}{n-k'}\binom{k'}{t}\binom{n-k'}{r-t}}{\binom{2n-r}{n}\binom{n}{r}}\\
&= \frac{k'}{n}\sum_{b=0}^{a}\Pr\{\mathcal{B}_{k'}^{k'-b}\}\frac{\binom{k'+b}{b}\binom{n-k'-a-b}{a-b}\binom{k'}{b}\binom{n-k'}{a-b}}{\binom{n+a}{a}\binom{n}{a}}\\
&\sim \frac{1}{n}\sum_{b=0}^{a} 2^{2b}\binom{2b}{b}^{-1}\binom{a}{b}^2 (x)^{2b}(1-x)^{2a-2b}.
\end{aligned}$$

We note that

$$\begin{aligned}\int_0^1 \sum_{b=0}^{a} 2^{2b}\binom{2b}{b}^{-1}\binom{a}{b}^2 (x)^{2b}(1-x)^{2a-2b}dx &= \sum_{b=0}^{a} \frac{2^b(a)_b}{(2a+1)(2a-1)\cdots(2a-2b+1)} \\ &= \sum_{b=0}^{a} 2^{2b}\frac{(a)_b(a)_b}{(2a+1)(2a)_{2b}} \\ &= \frac{1}{(2a+1)\binom{2a}{a}}\sum_{b=0}^{a} 2^{2b}\binom{2a-2b}{a-b} \\ &= \frac{2^{2a}}{(2a+1)\binom{2a}{a}}\sum_{j=0}^{a} 2^{-2j}\binom{2j}{j} = 1,\end{aligned}$$

where the last identity (see e.g. (1.109) in [11]) can be easily proved by induction. □

It follows from Theorem 10 that if $a \to \infty$ as $n \to \infty$, then $\frac{C_1^r(n)}{n}$ converges in distribution to a $Beta(1, 1/2)$ random variable with density given by $f(u) = \frac{1}{2}(1-u)^{-1/2}$ on the interval $(0, 1)$. On the other hand, if $a = n - r$ is fixed as $n \to \infty$, then $\frac{C_1^r(n)}{n}$ converges in distribution to a non-degenerate random variable with density given by $f_a(u) = \sum_{b=0}^{a}\binom{2b}{b}^{-1}\binom{a}{b}^2(2u)^{2b}(1-u)^{2a-2b}$ on the interval $(0, 1)$. Theorem 10 is also key to the next result which identifies the limiting distribution of the order statistics of the normalised component sizes of $\hat{G}_n^r$. To state the result, we define, for $0 \leqslant r < n$ and $i \geqslant 1$, $\hat{Y}_i^r(n)$ to be the size of the $i^{th}$ largest connected component in $\hat{G}_n^r$, where $\hat{Y}_i^r(n) = 0$ if the number of components in $\hat{G}_n^r$ is less than $i$, and we let $Q_i^r(n) = \hat{Y}_i^r(n)/n$ denote the $i^{th}$ normalised order statistic of the component sizes of $\hat{G}_n^r$. With this notation, we can state:

**Theorem 11.** *Suppose that $a = n - r \to \infty$ as $n \to \infty$, then the joint distribution of $(Q_1^r(n), Q_2^r(n), Q_3^r(n), \ldots)$, the normalised order statistics for the component sizes of $\hat{G}_n^r$, converges in distribution to the $\mathcal{PD}(1/2)$ distribution on the simplex*

$$\nabla = \{(x_1, x_2, \ldots) : \sum x_i \leqslant 1, x_i \geqslant x_{i+1} \geqslant 0\}$$

*as $n \to \infty$.*

*Sketch of the proof.* The proof of this result depends on a well-known connection between size-biased sampling of components in a random combinatorial structure and the *Poisson-Dirichlet($\theta$)* distribution, denoted $\mathcal{PD}(\theta)$, on $\nabla$. To describe how we use this connection to prove the result above, we must introduce some additional notation. First, recall that $\mathcal{C}_1^r(n) = \mathcal{C}_1(\hat{T}_n^r)$ denotes the component in $\hat{G}_n^r$ that contains the vertex labelled 1 and that $C_1^r(n) = |\mathcal{C}_1(\hat{T}_n^r)|$. If $\mathcal{C}_1^r(n) \neq \hat{G}_n^r$, let $\mathcal{C}_2^r(n)$ denote the component in $\hat{G}_n^r \setminus \mathcal{C}_1^r(n)$ which contains the vertex with smallest label; otherwise, set $\mathcal{C}_2^r(n) = \emptyset$. For $i > 2$, we define $\mathcal{C}_i^r(n)$ iteratively: If $\hat{G}_n^r \setminus (\mathcal{C}_1^r(n) \cup \ldots \cup \mathcal{C}_{i-1}^r(n)) \neq \emptyset$, let $\mathcal{C}_i^r(n)$ denote the component in $\hat{G}_n^r \setminus (\mathcal{C}_1^r(n) \cup \ldots \cup \mathcal{C}_{i-1}^r(n))$ which contains the vertex with smallest label; otherwise, set

$\mathcal{C}_i^r(n) = \emptyset$. So we obtain the sequence of components $\mathcal{C}_1^r(n), \mathcal{C}_2^r(n), \ldots$ by removing, at the $i^{th}$ step, a 'typical' component of the remaining graph $\hat{G}_n^r \setminus (\mathcal{C}_1^r(n) \cup \ldots \cup \mathcal{C}_{i-1}^r(n))$ and this selection process is size-biased. For $1 \leqslant r < n$ and $i > 0$, we define $C_i^r(n) = |\mathcal{C}_1^r(n)|$, and we define a sequence of normalised component sizes $(U_1^r(n), U_2^r(n), \ \ldots)$ as follows: Let $U_1^r(n) = \frac{C_1^r(n)}{n}$, and for $i > 1$, let $U_i^r(n) = 0$ if $n - C_1^r(n) - \cdots - C_{i-1}^r(n) = 0$; otherwise let

$$U_i^r(n) = \frac{C_i^r(n)}{n - C_1^r(n) - \cdots - C_{i-1}^r(n)}.$$

So, for $i \geqslant 1$, $U_i^r(n)$ equals the relative size of the size-biased component $\mathcal{C}_i^r(n)$ in the digraph $\hat{G}_n^r \setminus (\mathcal{C}_1^r(n) \cup \ldots \cup \mathcal{C}_{i-1}^r(n))$.

Now the convergence principle for the $\mathcal{PD}(\theta)$ distribution (see, for example [13] and the references therein) says that to show that the joint distribution of $(Q_1^r(n), Q_2^r(n), \ldots)$ converges to the $\mathcal{PD}(1/2)$ distribution on $\nabla$, it is enough to show that the joint distribution of $(U_1^r(n), U_2^r(n), \ \ldots)$ converges to the joint distribution of $(U_1, U_2, \ldots)$ where $U_1, U_2, \ldots$ are i.i.d. $Beta(1, 1/2)$ random variables with density given by $f(u) = \frac{1}{2}(1-u)^{-1/2}$ on the interval $(0, 1)$. Thus it is enough to prove:

**Proposition 12.** *Let $a = n - r$ and suppose that $a \to \infty$ as $n \to \infty$, then for any integer $j \geqslant 1$ and constants $0 < u_i < v_i < 1$, where $1 \leqslant i \leqslant j$,*

$$\lim_{n\to\infty} \Pr\{u_i < U_i^r(n) < v_i, 1 \leqslant i \leqslant j\} = \prod_{i=1}^{j} \int_{u_i}^{v_i} \frac{1}{2\sqrt{1-x}} dx.$$

*Proof.* Throughout the proof we adopt the convention that for any integer $i > 0$ and any vectors $\vec{w}$ and $\vec{z}$, $c_i(\vec{w}, \vec{z})$ is a constant which depends only on $\vec{w}$ and $\vec{z}$. Now for any $j \geqslant 1$, and for any $\vec{u} = (u_1, \ldots, u_j)$ and $\vec{v} = (v_1, \ldots, v_j)$ such that $0 < u_i < v_i < 1$ for $1 \leqslant i \leqslant j$, we define, for $0 < r < n$, the event

$$\mathcal{A}_n^r(j, \vec{u}, \vec{v}) \equiv \left\{ u_i < \frac{C_i^r(n)}{n - C_1^r(n) - \cdots - C_{i-1}^r(n)} < v_i, 1 \leqslant i \leqslant j \right\}.$$

We show by induction on $j$, that for $\vec{u} = (u_1, u_2, \ldots, u_j)$ and $\vec{v} = (v_1, v_2, \ldots, v_j)$ such that $0 < u_i < v_i < 1$ for $1 \leqslant i \leqslant j$, and $0 < r < n$,

$$\Pr\{\mathcal{A}_n^r(j, \vec{u}, \vec{v})\} = \left( \prod_{i=1}^{j} \int_{u_i}^{v_i} \frac{1}{2\sqrt{1-x}} dx \right) \big(1 + \eta(\vec{u}, \vec{v}, r, n)\big) \tag{33}$$

where $|\eta(\vec{u}, \vec{v}, r, n)| \leqslant c_1(\vec{u}, \vec{v}) a^{-\zeta(j)}$ for some constant $\zeta(j) > 0$ and all sufficiently large $a$. First, it is clear from Theorem 10, that the result holds for $j = 1$. Next, suppose that $j \geqslant 2$ and the result holds for $j - 1$, and suppose that $0 < u_i < v_i < 1$, for $1 \leqslant i \leqslant j$. Then we have

$$\Pr\{\mathcal{A}_n^r(j, \vec{u}, \vec{v})\} = \sum_{k > u_1 n}^{v_1 n} \Pr\left\{ C_1^r(n) = k, \ u_i < \frac{C_i^r(n)}{n - k - \cdots - C_{i-1}^r(n)} < v_i, 2 \leqslant i \leqslant j \right\}. \tag{34}$$

Now for any $u_1 n < k < v_1 n$, we have

$$\begin{aligned}
&\Pr\left\{C_1^r(n) = k,\ u_i < \frac{C_i^r(n)}{n-k-\cdots-C_{i-1}^r(n)} < v_i, 2 \leqslant i \leqslant j\right\} \\
&= \binom{n-1}{k-1} \Pr\left\{\mathcal{C}_1^r(n) = [k],\ u_i < \frac{C_i^r(n)}{n-k-\cdots-C_{i-1}^r(n)} < v_i, 2 \leqslant i \leqslant j\right\}
\end{aligned}$$

by the invariance of the distribution of $\hat{G}_n^r$ under the re-labelling of its vertices. Next, for $u_1 n < k < v_1 n$, we define

$$\begin{aligned}
\mathcal{F}_n(k, \hat{u}, \hat{v}) &= \left\{f \in \mathcal{M}_n : \mathcal{C}_1(f) = [k],\ u_i < \frac{C_i(f)}{n-k-\cdots-C_{i-1}(f)} < v_i,\ 2 \leqslant i \leqslant j\right\}; \\
\tilde{\mathcal{M}}_k &= \left\{g \in \mathcal{M}_k : \mathcal{C}_1(g) = [k]\right\}; \\
\mathcal{H}_{n-k}(\hat{u}, \hat{v}) &= \left\{h \in \mathcal{M}_{n-k} : u_i < \frac{C_{i-1}(h)}{n-k-\cdots-C_{i-2}(h)} < v_i, 2 \leqslant i \leqslant j\right\}
\end{aligned}$$

where $\hat{u} = (u_2, \ldots, u_j)$ and $\hat{v} = (v_2, \ldots, v_j)$. We note that any $f \in \mathcal{F}_n(k, \hat{u}, \hat{v})$ can be identified with a pair of functions $g \in \tilde{M}_k$ and $h \in \mathcal{H}_{n-k}(\hat{u}, \hat{v})$, such that for $1 \leqslant \ell \leqslant k$, $f(\ell) = g(\ell)$ and for $k+1 \leqslant \ell \leqslant n$, $f(\ell) = h(\ell - k) + k$. This is a 1-to-1 correspondence and we denote this correspondence by $f \equiv (g, h)$. Using this notation, we have

$$\begin{aligned}
&\binom{n-1}{k-1} \Pr\left\{\mathcal{C}_1^r(n) = [k], u_i < \frac{C_i^r(n)}{n-k-\cdots-C_{i-1}^r(n)} < v_i, 2 \leqslant i \leqslant j\right\} \\
&= \binom{n-1}{k-1} \sum_{f \in \mathcal{F}_n(k, \hat{u}, \hat{v})} \Pr\{\hat{T}_n^r = f\} \\
&= \binom{n-1}{k-1} \sum_t \sum_{f \in \mathcal{F}_n(k, \hat{u}, \hat{v})} \Pr\{\hat{T}_n^r = f \mid Z_n^r(k) = t\} \frac{\binom{k}{t}\binom{n-k}{r-t}}{\binom{n}{r}} \qquad (35)
\end{aligned}$$

where the last sum above is over all $t$ for which the binomial coefficients are defined and $Z_n^r(k)$ denotes the number of balls removed from the red balls labelled 1 to $k$ in the first step of the construction of $\hat{T}_n^r$. Next we note that for any $f \in \mathcal{F}(k, \hat{u}, \hat{v})$

$$\begin{aligned}
\Pr\{\hat{T}_n^r = f \mid Z_n^r(k) = t\} &= \Pr\{\hat{T}_k^t = g\} \Pr\{\hat{T}_{n-k}^{r-t} = h\} \frac{(2k-t)_k (2(n-k)-r+t)_{n-k}}{(2n-r)_n} \\
&= \binom{n}{k}^{-1} \Pr\{\hat{T}_k^t = g\} \Pr\{\hat{T}_{n-k}^{r-t} = h\} \frac{\binom{2k-t}{k}\binom{2(n-k)-r+t}{n-k}}{\binom{2n-r}{n}} \qquad (36)
\end{aligned}$$

where $f \equiv (g, h)$ for some $g \in \tilde{\mathcal{M}}_k$ and $h \in \mathcal{H}_{n-k}(\hat{u}, \hat{v})$. We also note that

$$\sum_{g \in \tilde{\mathcal{M}}_k} \Pr\{\hat{T}_k^t = g\} = \Pr\{\mathcal{B}_k^t\} \qquad (37)$$

and

$$\sum_{h\in\mathcal{H}_{n-k}(\hat{u},\hat{v})}\Pr\{\hat{T}_{n-k}^{r-t}=h\}=\Pr\{\mathcal{A}_{n-k}^{r-t}(j-1,\hat{u},\hat{v})\}. \tag{38}$$

Finally by the induction hypothesis, we have

$$\Pr\{\mathcal{A}_{n-k}^{r-t}(j-1,\hat{u},\hat{v})\}=\left(\prod_{i=2}^{j}\int_{u_i}^{v_i}\frac{1}{2\sqrt{1-x}}dx\right)\big(1+\eta(\hat{u},\hat{v},r-t,n-k)\big) \tag{39}$$

for $0\leqslant t\leqslant r$, where $|\eta(\hat{u},\hat{v},r-t,n-k)|\leqslant c_2(\hat{u},\hat{v})(n-k-r+t)^{-\zeta(j-1)}$.

Now suppose that $r<n^{1/4}$. Then for $0\leqslant t\leqslant r$, we have the uniform bound $c_2(\hat{u},\hat{v})(n-k-r+t)^{-\zeta(j-1)}\leqslant c_3(\vec{u},\vec{v})a^{-\zeta(j-1)}$, and it follows from (15) and (34) – (39) that

$$\begin{aligned}\Pr\{\mathcal{A}_n^r(j,\vec{u},\vec{v})\} &= \sum_{k>u_1n}^{v_1n}\frac{k}{n}\sum_t\Pr\{\mathcal{B}_k^t\}\Pr\{\mathcal{A}_{n-k}^{r-t}(j-1,\hat{u},\hat{v})\}\frac{\binom{2k-t}{k}\binom{2(n-k)-r+t}{n-k}}{\binom{2n-r}{n}}\frac{\binom{k}{t}\binom{n-k}{r-t}}{\binom{n}{r}}\\ &= \Pr\{u_1n<C_1^r(n)<v_1n\}\left(\prod_{i=2}^{j}\int_{u_i}^{v_i}\frac{dx}{2\sqrt{1-x}}\right)(1+\xi(\vec{u},\vec{v},r,n)) \end{aligned}\tag{40}$$

where $|\xi(\vec{u},\vec{v},r,n)|\leqslant c_4(\vec{u},\vec{v})a^{-\zeta(j-1)}$. Equation (33) now follows from (40) and Theorem 10.

Next, suppose that $r\geqslant n^{1/4}$, then as is the proof of Theorem 10, we write $a=\alpha n$ and $r=(1-\alpha)n$ and we re-write the sum on the right-hand side of the first equality in (40) in terms of $\alpha$ to obtain:

$$\begin{aligned}&\Pr\{\mathcal{A}_n^r(j,\vec{u},\vec{v})\}\\ &=\sum_{k>u_1n}^{v_1n}\frac{k}{n}\sum_{\ell}\Pr\{\mathcal{B}_k^{k-\ell}\}\Pr\{\mathcal{A}_{n-k}^{r-k+\ell}(j-1,\hat{u},\hat{v})\}\frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}}\\ &=\Pr\{u_1n<C_1^r(n)<v_1n\}\left(\prod_{i=2}^{j}\int_{u_i}^{v_i}\frac{dx}{2\sqrt{1-x}}\right)\\ &\quad+\sum_{k>u_1n}^{v_1n}\frac{k}{n}\sum_{\ell}\Pr\{\mathcal{B}_k^{k-\ell}\}\Phi(\ell,k,\alpha,n,\hat{u},\hat{v})\frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}}\end{aligned}\tag{41}$$

where the sums above are over those values of $\ell$ for which the binomial coefficients are defined and where

$$\Phi(\ell,k,\alpha,n,\hat{u},\hat{v})\equiv\Pr\{\mathcal{A}_{n-k}^{r-k+\ell}(j-1,\hat{u},\hat{v})\}-\left(\prod_{i=2}^{j}\int_{u_i}^{v_i}\frac{dx}{2\sqrt{1-x}}\right).$$

Now for any $k=xn$ where $u_1<x<v_1$ and for any $0\leqslant\ell\leqslant a=\alpha n$, we define $\Delta(\ell,\alpha,k,n)=\ell-\lfloor\alpha xn\rfloor$. Then by the induction hypothesis, we have for $0\leqslant\ell\leqslant\alpha n$ such that $|\Delta(\ell,\alpha,x,n)|<(\alpha xn(1-x)(1-\alpha))^{7/12}\equiv\rho(\alpha,x,n)$

$$|\Phi(\ell,k,\alpha,n,\hat{u},\hat{v})|\leqslant c_5(\hat{u},\hat{v})(n-r-\ell)^{-\zeta(j-1)}\leqslant c_6(\vec{u},\vec{v})a^{-\zeta(j-1)}, \tag{42}$$

for all sufficiently large $a$, and in all cases $|\Phi(\ell,k,\alpha,n,\hat{u},\hat{v})| \leqslant 2$. It follows from (42) and the bounds (31) and (32) obtained in the proof of Theorem 10 that

$$\begin{aligned}
&\left| \sum_{k>u_1 n}^{v_1 n} \frac{k}{n} \sum_{\ell} \Pr\{\mathcal{B}_k^{k-\ell}\} \Phi(\ell,k,\alpha,n,\hat{u},\hat{v}) \frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}} \right| \\
&\leqslant c_6(\vec{u},\vec{v}) a^{-\zeta(j-1)} \Pr\{u_1 n \leqslant C_1^r(n) \leqslant v_1 n\} \\
&\quad +2 \sum_{k>u_1 n}^{v_1 n} \frac{k}{n} \sum_{\substack{\ell \text{ s.t.} \\ \Delta(\ell,x\alpha,n)|>\rho(\alpha,x,n)}} \Pr\{\mathcal{B}_k^{k-\ell}\} \frac{\binom{k+\ell}{\ell}\binom{n-k+\alpha n-\ell}{\alpha n-\ell}\binom{k}{\ell}\binom{n-k}{\alpha n-\ell}}{\binom{n+\alpha n}{\alpha n}\binom{n}{\alpha n}} \\
&\leqslant c_6(\vec{u},\vec{v}) a^{-\zeta(j-1)} \Pr\{u_1 n \leqslant C_1^r(n) \leqslant v_1 n\} \\
&\quad +2c_7(u_1,v_1) a^{-1/32} \sum_{k>u_1 n}^{v_1 n} \frac{1}{2n\sqrt{1-k/n}} \qquad (43)
\end{aligned}$$

for all sufficiently large $a$. It follows from (41), (43) and the induction hypothesis that (33) holds for all sufficiently large $a$. This completes proof of (33) and the proposition now follows for $j$ by taking the limit of (33) as $a = n - r$ tends to infinity. □

Given Proposition 12, Theorem 11 now follows, by the convergence principle described above.

# 5 Final Remarks

In this paper we have investigated graphical structure of random mappings under stronger in-degree constraints (when $r > 0$) than those considered by Arney and Bender in [3] or by the authors in [15] and have determined the distributions for the number of cyclic vertices, the number of components, and the size of a typical component when $r$ vertices are constrained to have in-degree at most 1 and the remaining $n-r$ vertices are constrained to have in-degree at most 2. We have also determined the asymptotic distributions for these variables under the regimes: $(i)$ $a = n - r \to \infty$ and $(ii)$ $a = n - r$ constant as $n \to \infty$, and we have shown that in regime $(i)$ the limiting distribution for the order statistics of the normalised component sizes of $\hat{G}_n^r$ is always $\mathcal{PD}(1/2)$. The persistence of the $\mathcal{PD}(1/2)$ distribution as a limiting distribution is surprising because, provided $a = n - r \to \infty$, it does not depend on the number of cyclic vertices in $\hat{G}_n^r$ or on the structure of the underlying uniform permutation of the cyclic vertices of $\hat{G}_n^r$. In particular, when $a = o(n)$, the number of cyclic vertices is of order $\frac{n}{\sqrt{a}}$, but the limiting distribution of the normalised order statistics of the component sizes is still the $\mathcal{PD}(1/2)$ distribution rather than the $\mathcal{PD}(1)$ distribution which was obtained as the limiting distribution in [16] when the order of the cyclic vertices in the random mapping model is greater than $\sqrt{n}$. This persistence of the $\mathcal{PD}(1/2)$ distribution no matter how slowly $a$ grows, suggests that there is a delicate and interesting interplay between the cycle structure of $\hat{G}_n^r$ and the structure of the forest obtained by deleting the cyclic edges in $\hat{G}_n^r$ which warrants further investigation.

It is also instructive to compare the asymptotic structure $\hat{G}_n^r$ to that of logarithmic combinatorial structures with parameter $\theta$. Examples of such structures include random permutations with Ewens cycle structure, prime factorisation of integers ([8]), factorisation of the characteristic polynomial for a random matrix over a finite field ([17]), and uniform mapping patterns ([22]) (for other examples, see [4, 14]). In a logarithmic combinatorial structure with parameter $\theta$, the distribution of the order statistics for the normalised 'component' sizes converges to the $\mathcal{PD}(\theta)$ distribution as the size of the structure $n \to \infty$. In addition, the expected number of 'components' of size $k$ is asymptotic to $\frac{\theta}{k}$ as $k \to \infty$ and the total number of components is of order $\theta \log n$ as $n \to \infty$. In contrast, it follows from the Corollary 8, that the number of components, $\hat{N}_n^r$, in $\hat{G}_n^r$ is of order $\log(\frac{n}{\sqrt{a}})$. So when $a = o(n)$ and $a \to \infty$, the asymptotic structure of $\hat{G}_n^r$ is qualitatively different from that of a logarithmic combinatorial structure with parameter $\theta = 1/2$ since $\frac{1}{2}\log n = o(\hat{N}_n^r)$. We note that we cannot 'see' this difference if we only look at the large components of $\hat{G}_n^r$ because the limiting distribution of the order statistics of the normalised component sizes of $\hat{G}_n^r$ is the same as that of a logarithmic combinatorial structure with parameter $\theta = 1/2$. It would be interesting to determine the distribution of the number of components of size $k$ in $\hat{G}_n^r$ for $k = o(n)$, and to use such results to extend the central limit theorem for $\hat{N}_n^r$ obtained in this paper to a functional central limit theorem for the component sizes in $\hat{G}_n^r$, analogous to the functional central limit theorems that have been obtained for uniform permutations (see [7]), uniform random mappings (see [12]) and other logarithmic combinatorial structures.

Finally, we mention an alternative, but related, model, $\tilde{T}_n^\beta$, for random mappings with in-degree constraints. The construction of $\tilde{T}_n^\beta$ is similar in spirit to the configuration model from random graph theory and is defined as follows: Suppose that $0 \leqslant \beta \leqslant 1$ and let $D_1^\beta, D_2^\beta, \ldots$ be a sequence of i.i.d. random variables such that for $i \geqslant 1$,

$$\Pr\{D_i^\beta = 0\} = \Pr\{D_i^\beta = 2\} = \frac{1-\beta}{(2-\beta)^2} \quad \text{and} \quad \Pr\{D_i^\beta = 1\} = \frac{1+(1-\beta)^2}{(2-\beta)^2}$$

and let $D(\beta, n) = (D_{1,n}^\beta, D_{2,n}^\beta, \ldots, D_{n,n}^\beta)$ be a collection of exchangeable random variables with joint distribution given by

$$\Pr\left\{D_{i,n}^\beta = d_i, 1 \leqslant i \leqslant n\right\} = \Pr\left\{D_i^\beta = d_i, 1 \leqslant i \leqslant n \,\middle|\, \sum_{i=1}^n D_i^\beta = n\right\}.$$

Then we define $\tilde{T}_n^\beta$ so that, given the event $\{D(\beta, n) = (d_1, d_2, \ldots, d_n)\}$, $\tilde{T}_n^\beta$ is uniformly distributed over $\mathcal{M}_n(d_1, d_2, \ldots, d_n)$. It is clear from the definition of $\tilde{T}_n^\beta$ that the vertices in $\tilde{G}_n^\beta \equiv G(\tilde{T}_n^\beta)$ have in-degree at most 2. Furthermore, the larger the value of $\beta$, the greater the expected number of vertices with in-degree 1, and when $\beta = 1$, $\tilde{T}_n^1$ is a uniform random permutation. Since the variables $(D_{1,n}^\beta, D_{2,n}^\beta, \ldots, D_{n,n}^\beta)$ are exchangeable, one can use the calculus for random mappings with exchangeable in-degrees to investigate the structure of $\tilde{G}_n^\beta$, but in this case the distributions that are obtained are more complicated and cumbersome than those presented here for $\hat{G}_n^r$. Nevertheless, some preliminary calculations indicate that for $0 < \beta(n) < 1$ such that $\beta(n)n \to \infty$ and $(1-\beta(n))n \to \infty$

as $n \to \infty$, the models $\hat{G}_n^r$ and $\tilde{G}_n^{\beta(n)}$ are asymptotically related. In particular, it should be possible to translate the results obtained in this paper into results for $\tilde{G}_n^{\beta}$.


## Acknowledgements

Jerzy Jaworski acknowledges also the support by the Marie Curie Intra-European Fellowship No. 236845 (RANDOMAPP) within the 7th European Community Framework Programme.


# References


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

[2] D. Aldous, G. Miermont and J. Pitman. Brownian Bridge Asymptotics for Random $p-$mappings. *Electonic J. Probab.*, 9:37–56, 2004.

[3] J. Arney and E. A. Bender. Random mappings with constraints on coalescence and number of origins. *Pacific J. Math.*, 103:269–294, 1982.

[4] R. Arratia, A. D. Barbour, and S. Tavaré. *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS Monographs in Mathematics, European Mathematical Society, 2003.

[5] S. Berg and L. Mutafchiev. Random mappings with an attracting center: Lagrangian distributions and a regression function. *J. Appl. Probab.*, 27:622–636, 1990.

[6] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Europ. J. Combinatorics*, 1:311–316, 1980.

[7] J. M. DeLaurentis and B. Pittel. Random permutations and Brownian Motion. *Pacific J. Math.*, 119:287–301, 1985.

[8] P. Donnelly and G. Grimmett. On the asymptotic distribution of large prime factors. *Jour. of the LMS*, 47:395–404, 1993.

[9] P. Flajolet and A. M. Odlyzko,. Random mapping statistics. In *Advances in Cryptology - Eurocrypt Proceedings, volume 434 of Lecture Notes in Comput. Sci.*, pages 329–354. Springer, 1990.

[10] I. B. Gertsbakh. Epidemic processes on a random graph: some preliminary results. *J. Appl. Probab.*, 14:427–438, 1977.

[11] H. Gould. *Combinatorial Identities*, Revised Edition, Morgantown, W. Va., 1972.

[12] J. C. Hansen. A functional central limit theorem for random mappings. *Ann. of Probab.*, 17:317–332, 1989.

[13] J. C. Hansen,. Order statistics for decomposable combinatorial structures. *Random Struct. Algorithms*, 5:517–533, 1994.

[14] J. C. Hansen and J. Jaworski. A cutting process for random mappings. *Random Struct. Algorithms*, 30:287–306, 2007.

[15] J. C. Hansen and J. Jaworski. Random mappings with exchangeable in-degrees. *Random Struct. Algorithms*, 33:105–126, 2008.

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

[17] J. C. Hansen and E. Schmutz. How random is the characteristic polynomial of a random matrix? *Math. Proc. Cam. Phil. Soc.*, 114:507–515, 1993.

[18] J. Jaworski. Random mappings with independent choices of the images. In *Random Graphs*, volume 1, pages 89–101. Wiley, New York, 1990.

[19] N. L. Johnson, A. K. Kemp, S. Kotz. *Univariate Discrete Distributions*. 3rd edition, Wiley, New York, 2005.

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

[21] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms*, 6:161–179, 1995.

[22] L. Mutafchiev. Large components and cycles in a random mapping pattern. In *Random Graphs'87* (M. Karoński, J. Jaworski, A. Ruciński, eds), pages 189–202. Wiley, New York, 1990.

[23] L. Mutafchiev. On random mappings with a single attracting centre. *J. Appl. Probab.*, 24:258–264, 1987.

[24] B. Pittel. On distributions related to transitive closures of random finite mappings. *Ann. of Probab.*, 11:428–441, 1983.

[25] J. J. Quisquater and J. P. Delescaille. How easy is collision search? Application to DES. In *Advances in Cryptology - Eurocrypt Proceedings, volume 434 of Lecture Notes in Comput. Sci.*, pages 429–434. Springer, 1990.

[26] P. C. van Oorschot and M. J. Wiener. Parallel collision search with applications to hash functions and discrete logarithms. In *Proc. of the 2nd ACM Conference on Computer and Communications Security*, pages 210–218, 1994.

[27] V. E. Stepanov. Limit distributions for certain characteristics of random mappings. *Theory Probab. Appl.*, 14:612–622, 1969.

[28] V. E. Stepanov. Random mappings with a single attracting center. *Theory Probab. Appl.*, 16:155–162, 1971.

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