Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 150 (2022) 1139–1164

stochastic processes and their applications

www.elsevier.com/locate/spa

The Feller Coupling for random derangements

Poly H. da Silva[a], Arash Jamshidpey[b], Simon Tavaré[a,c,*]

[a] Columbia University, Department of Statistics, 1255 Amsterdam Avenue, New York, NY 10027, USA
[b] Columbia University, Department of Mathematics, 2990 Broadway, New York, NY 10027, USA
[c] Columbia University, Irving Institute for Cancer Dynamics, Schermerhorn Hall, Suite 601, 1190 Amsterdam Avenue, New York, NY 10027, USA

Received 3 January 2021; received in revised form 14 July 2021; accepted 6 September 2021
Available online 20 September 2021

Abstract

We study derangements of $\{1, 2, \ldots, n\}$ under the Ewens distribution with parameter θ. We give basic properties of derangements, such as the moments and marginal distributions of the cycle counts, the number of cycles, and asymptotic distributions for large n, and we construct, for any given n, a $\{0, 1\}$-valued non-homogeneous Markov chain $\vec{\eta}^{(n)}$ with the property that the counts of lengths of spacings between the 1s have the same distribution as the cycle counts of the random derangement of size n. Unlike the Feller Coupling, this chain does not couple realizations for different values of n – the chain must be rerun to get derangements of other sizes. To resolve this issue we construct another $\{0, 1\}$-valued Markov chain η whose law coincides with that of the Feller Coupling conditional on no consecutive 1s. The distribution of η, the so-called "Feller Coupling for random derangements", arises as the weak limit as $n \to \infty$ of the distributions of $\vec{\eta}^{(n)}$. Consequently, the asymptotic behavior of finite random derangements may be studied via this coupling. The rate of convergence of $\vec{\eta}^{(n)}$ to η is studied via an estimate of their total variation distance. We provide extensive comparisons of these methods, and show that the Markov chain methods generate derangements in time independent of θ for a given n and linear in the size of the derangement.

MSC: 60C05; 60J10; 65C05; 97K20; 97K60; 65C40

Keywords: Feller Coupling; Simulation; Poisson approximation; Poisson–Dirichlet and GEM distributions; Probabilistic combinatorics

* Corresponding author at: Columbia University, Department of Statistics, 1255 Amsterdam Avenue, New York, NY 10027, USA.

E-mail addresses: phd2120@columbia.edu (P.H. da Silva), aj2963@columbia.edu (A. Jamshidpey), st3193@columbia.edu (S. Tavaré).

https://doi.org/10.1016/j.spa.2021.09.003
0304-4149/

1. Introduction

The Ewens Sampling Formula [10] arose in population genetics as the joint probability distribution of the number of selectively neutral alleles $C_j(n)$ represented j times in a sample of n genes, for $j = 1, 2, \ldots, n$. For non-negative integers $c_1, c_2, \ldots, c_n$ satisfying $\sum_{j=1}^{n} jc_j = n$, we have

$$\mathbb{P}_\theta(C_1(n) = c_1, \ldots, C_n(n) = c_n) = \frac{n!}{\theta_{(n)}} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j!}, \tag{1}$$

for $\theta \in (0, \infty)$, $\theta_{(n)} := \theta(\theta+1)\ldots(\theta+n-1) = \Gamma(n+\theta)/\Gamma(\theta), n \geq 1$. We set $\theta_{(0)} = 1, \theta_{(-k)} = 0, k \in \mathbb{N}$. In its original formulation, θ is a parameter related to the rate at which novel alleles appear. In what follows we denote the law in (1) by $\mathrm{ESF}_n(\theta)$; to simplify the notation, we suppress the θ in $\mathbb{P}_\theta$ in what follows, where there is no cause for confusion.

The ESF has been studied extensively, and it arises in many different settings in probability and statistics. [13, Chapter 41] provides an overview, [6] describes numerous applications in combinatorics, and [9,20] provide many other examples. Of particular interest here is its appearance as the distribution of the cycle counts of a θ-biased permutation. Let π be a permutation of $\{1, 2, \ldots, n\}$ decomposed as a product of cycles. If π is chosen uniformly with probability $1/n!$, then Cauchy's formula establishes that the cycle counts $(C_1(n), \ldots, C_n(n))$ have the $\mathrm{ESF}_n(1)$ law [12], and if a permutation π having k cycles is chosen with probability proportional to θ^k, then the cycle counts have the $\mathrm{ESF}_n(\theta)$ law. In this case,

$$\mathbb{P}(\pi) = \frac{\theta^k}{\theta_{(n)}}, \tag{2}$$

if the permutation π has k cycles. See [6, Chapters 1 and 2.5] for more detailed discussion and history. We note that the number K_n of cycles has probability generating function given by

$$\mathbb{E}(s^{K_n}) = \frac{(\theta s)_{(n)}}{\theta_{(n)}}, 0 \leq s \leq 1. \tag{3}$$

The Feller Coupling was introduced in [5] as a way to generate the cycles in a growing permutation one at a time, and it has proved very useful in the study of the asymptotics of properties of the ESF; [7] illustrates some of these. To describe the Feller Coupling, define independent Bernoulli random variables ξ_i satisfying

$$\mathbb{P}(\xi_i = 1) = \frac{\theta}{\theta + i - 1}, \quad \mathbb{P}(\xi_i = 0) = \frac{i-1}{\theta + i - 1}, \quad i \geq 1.$$

The cycle counts for a permutation of size n are determined by the spacings between the 1s in realizations of $\xi_n, \xi_{n-1}, \ldots, \xi_1 = 1$, starting from a 1 placed at position $n+1$. If we define

$$C_j(n) = \text{number of spacings of length } j \text{ between the 1s in } 1\xi_n\xi_{n-1}\ldots\xi_2 1, \tag{4}$$

then

(i) The law of $(C_1(n), \ldots, C_n(n))$ is $\mathrm{ESF}_n(\theta)$;
(ii) $Z_j = C_j(\infty) =$ the number of spacings of length j between the 1s in $1\xi_2\xi_3\ldots$ are independent Poisson-distributed random variables having $\mathbb{E}Z_j = \theta/j$.
(iii) To generate a particular permutation, we can use auxiliary randomization; cf. [6, page 95].

Further details may be found in [5] and [6, Chapter 5].

In this paper we study the behavior of the cycle counts of derangements, permutations with no fixed points, obtained from $\mathrm{ESF}_n(\theta)$, conditional on having no cycles of length 1, via the relation

$$\mathcal{L}_\theta(\widetilde{C}_2(n), \ldots, \widetilde{C}_n(n)) = \mathcal{L}_\theta(C_2(n), \ldots, C_n(n) \mid C_1(n) = 0), \tag{5}$$

where $\widetilde{C}_j(n)$ denotes the number of cycles of length j in the derangement. After recording some basic properties of random derangements in Section 2, our focus in Section 3 is to construct a Markov chain $\overleftarrow{\eta}^{(n)} = (\overleftarrow{\eta}_n^{(n)} = 0, \ldots, \overleftarrow{\eta}_1^{(n)} = 1)$ that allows us to generate the cycle counts via the analog of (4):

$$\widetilde{C}_j(n) = \text{number of spacings of length } j \text{ between the 1s in } 1, \overleftarrow{\eta}_n^{(n)}, \ldots, \overleftarrow{\eta}_1^{(n)} = 1.$$

We show that the law of the resulting cycle counts is indeed given by (5).

In contrast to the Feller coupling, which, as its name suggests, allows generation of the cycles sequentially for each n from a single observation of $1 = \xi_1, \xi_2, \xi_3 \ldots$, the chains $\overleftarrow{\eta}^{(n)}$ have to be rerun for each n to produce derangements of different sizes. The main results of this paper, in Section 4, concern the asymptotic behavior of derangements, and provide a coupling that can be used to generate derangements sequentially from a single run of a Markov chain. We analyze the behavior of this chain in some detail. In Section 5 we discuss methods for simulating derangements, and in Section 6 provide a number of examples of properties of derangements, for some of which explicit results for finite n are known, for some of which asymptotics are known, and for some of which simulation seems to be the only recourse. Section 7 provides some numerical experiments, and the paper concludes with a brief discussion.

2. Derangements

Students of probability often meet derangements in the context of (versions of) the so-called hat-check problem [11, Chapter IV]: n diners leave their hats at a restaurant before their meal and hats are returned at random after the meal. What is the probability that no diner gets back their own hat? Label the diners 1,2,..., n and construct a permutation π by setting π_j to be the label of the diner whose hat was returned to j. The question asks us to evaluate the probability that π has no singleton cycles, and inclusion-exclusion is typically used to show that the required probability is

$$\mathbb{P}_1(C_1(n) = 0) = \frac{D_n}{n!} = \sum_{l=0}^{n} (-1)^l \frac{1}{l!}, \tag{6}$$

where D_n is the nth derangement number, the number of n-permutations with no fixed points. The cycles of a derangement describe groups of diners who share hats among themselves, with no diner getting his own. The distribution of the cycle counts $(\widetilde{C}_2(n), \ldots, \widetilde{C}_n(n))$ is given by (5) with $\theta = 1$. For arbitrary θ, the distribution (5) is determined by $\mathrm{ESF}_n(\theta)$, and (6) is replaced by

$$\lambda_n(\theta) := \mathbb{P}_\theta(C_1(n) = 0) = \frac{n!}{\theta_{(n)}} \sum_{j=0}^{n} (-1)^j \frac{\theta^j}{j!} \frac{\theta_{(n-j)}}{(n-j)!}, \; n = 2, 3, \ldots \tag{7}$$

with $\lambda_0(\theta) = 1$, $\lambda_1(\theta) = 0$; cf. [13, eqn. (41.10)] We will see later that $\lambda_n(\theta)$ can be represented as a confluent hypergeometric function. We can also see that, for any given n, the mapping

$\theta \mapsto \lambda_n(\theta)$ is decreasing. To show this, for $j \in \mathbb{N}$ let

$$\Delta_j^* = \left\{(a_j, a_{j-1}, \ldots, a_1) \in \{0,1\}^j : a_j + \sum_{i=1}^{j-1} a_i a_{i+1} = 0\right\},$$

consisting of all $(0, a_{j-1}, \ldots, a_1)$ with no consecutive 1's, and let

$$\Delta_j = \{(a_j, a_{j-1}, \ldots, a_1) \in \Delta_j^* : a_1 = 1\}.$$

Note that $\Delta_1 = \emptyset$. In particular, $\lambda_n(\theta) = \mathbb{P}((\xi_n, \ldots, \xi_1) \in \Delta_n)$. Now, let $0 < \theta_1 \le \theta_2$, and for the moment, denote by $\xi^{\theta_1} = (\xi_i^{\theta_1})_{i=1}^\infty$, $\xi^{\theta_2} = (\xi_i^{\theta_2})_{i=1}^\infty$ the Feller couplings with parameters θ_1 and θ_2, respectively. Note that $\frac{\theta_1}{\theta_1+i} \le \frac{\theta_2}{\theta_2+i}$, for $i \in \mathbb{N}$. The idea is to couple $\xi_i^{\theta_1}$ and $\xi_i^{\theta_2}$ for $i \in \mathbb{N}$ such that

$$\mathbb{P}(\xi_i^{\theta_1} = 1, \xi_i^{\theta_2} = 1) = \frac{\theta_1}{\theta_1 + i},$$
$$\mathbb{P}(\xi_i^{\theta_1} = 0, \xi_i^{\theta_2} = 1) = \frac{\theta_2}{\theta_2 + i} - \frac{\theta_1}{\theta_1 + i},$$
$$\mathbb{P}(\xi_i^{\theta_1} = 0, \xi_i^{\theta_2} = 0) = \frac{i}{\theta_2 + i}.$$

Thus, it is clear that ξ^{θ_1} and ξ^{θ_2} are Feller couplings with parameters θ_1 and θ_2 such that whenever $\xi_i^{\theta_2} = 0$, we have $\xi_i^{\theta_1} = 0$, but not vice versa, that is if $(\xi_n^{\theta_2}, \ldots, \xi_1^{\theta_2}) \in \Delta_n$, then $(\xi_n^{\theta_1}, \ldots, \xi_1^{\theta_1}) \in \Delta_n$. Hence $\lambda_n(\theta_2) \le \lambda_n(\theta_1)$, from the definition of λ_n.

2.1. *Properties of derangements*

In the following sections we collect some results for derangements obtainable directly from (5).

2.1.1. *Factorial moments of the cycle counts*

The falling factorial moments are straightforward to compute. For $x \in \mathbb{R}$ and $r \in \mathbb{Z}_+$, define $x^{[r]} = x(x-1)\ldots(x-r+1), r \ge 1$ with $x^{[0]} = 1$. For $r_2, r_3, \ldots, r_b \ge 0$ with $2r_2 + \cdots + br_b = m \le n$,

$$\lambda_n(\theta)\, \mathbb{E}(\widetilde{C}_2^{[r_2]} \cdots \widetilde{C}_b^{[r_b]}) = \sum{}' c_2^{[r_2]} \cdots c_b^{[r_b]} \frac{n!}{\theta_{(n)}} \prod_{j=2}^n \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j!},$$

where the sum in $\sum'$ is over $2c_2 + \cdots + nc_n = n, c_2 \ge r_2, \ldots, c_b \ge r_b, c_{b+1}, \ldots, c_n \ge 0$. Letting $c_j' = c_j - r_j$, for $j \le b$, and noting that $1/c_j'! = c_j^{[r_j]}/c_j!$, the r.h.s. of the last equality equals to

$$\frac{n!}{\theta_{(n)}} \prod_{j=2}^b \left(\frac{\theta}{j}\right)^{r_j} \frac{\theta_{(n-m)}}{(n-m)!} \sum{}'' \frac{(n-m)!}{\theta_{(n-m)}} \prod_{j=2}^b \left(\frac{\theta}{j}\right)^{c_j'} \frac{1}{c_j'!} \prod_{j=b+1}^n \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j!}$$
$$= \frac{n!}{\theta_{(n)}} \prod_{j=2}^b \left(\frac{\theta}{j}\right)^{r_j} \frac{\theta_{(n-m)}}{(n-m)!} \lambda_{n-m}(\theta),$$

since the last sum is just the probability that a random permutation of $(n-m)$ objects is a derangement; the sum $\sum''$ is over $c_2', \ldots, c_b', c_{b+1}, \ldots, c_n \ge 0$ satisfying $2c_2' + \cdots + bc_b' +$

$(b+1)c_{b+1}+\cdots+nc_n = n-m$. Hence

$$\mathbb{E}(\widetilde{C}_2^{[r_2]}\cdots\widetilde{C}_b^{[r_b]}) = \mathbb{1}(m \le n)\,\frac{n!}{\lambda_n(\theta)\theta_{(n)}}\,\frac{\lambda_{n-m}(\theta)\theta_{(n-m)}}{(n-m)!}\prod_{j=2}^{b}\left(\frac{\theta}{j}\right)^{r_j}. \tag{8}$$

In particular, for $j = 2,\ldots,n$,

$$\mathbb{E}\widetilde{C}_j(n) = \frac{n!}{\lambda_n(\theta)\theta_{(n)}}\,\frac{\lambda_{n-j}(\theta)\theta_{(n-j)}}{(n-j)!}\,\frac{\theta}{j}. \tag{9}$$

Note that $\mathbb{P}(\widetilde{C}_{n-1}(n) = 0) = 1$, and indeed $\mathbb{E}\widetilde{C}_{n-1}(n) = 0$.

2.1.2. Distribution of the cycle counts

To compute the distribution of the cycle counts, suppose that X is a discrete random variable taking values in $\{0,1,2,\ldots,n\}$, with distribution $p_l = \mathbb{P}(X = l)$, $0 \le l \le n$. Define

$$u_j = \mathbb{E}X^{[j]} = \sum_{l=j}^{n} l^{[j]}\,p_l, \quad j = 1,2,\ldots,n,$$

where $u_0 = 1$. Inverting this relationship gives

$$p_r = \frac{1}{r!}\sum_{l=0}^{n-r}(-1)^l\,\frac{1}{l!}\,u_{r+l} = \frac{1}{r!}\sum_{i=r}^{n}(-1)^{i-r}\,\frac{1}{(i-r)!}\,u_i.$$

Using the result in (8), choose $j \in \{2,\ldots,n\}$, $i \le \lfloor n/j \rfloor$ and set

$$u_i = \mathbb{E}\widetilde{C}_j(n)^{[i]} = \frac{n!}{\lambda_n(\theta)\theta_{(n)}}\,\frac{\lambda_{n-ji}(\theta)\theta_{(n-ji)}}{(n-ji)!}\left(\frac{\theta}{j}\right)^i.$$

Then for $0 \le r \le \lfloor n/j \rfloor$,

$$\mathbb{P}(\widetilde{C}_j(n) = r) = \left(\frac{\theta}{j}\right)^r\frac{1}{r!}\frac{n!}{\lambda_n(\theta)\theta_{(n)}} \times \sum_{i=r}^{\lfloor n/j \rfloor}(-1)^{i-r}\frac{1}{(i-r)!}\frac{\lambda_{n-ji}(\theta)\theta_{(n-ji)}}{(n-ji)!}\left(\frac{\theta}{j}\right)^{i-r}.$$

The special case $j = n, r = 1$ is used in Section 6.1.

Remark 1. Many of these results are well known in the case of random derangements, for which $\theta = 1$. For example,

$$\mathbb{P}_1(\widetilde{C}_2(n) = 0) = \frac{n!}{D_n}\sum_{i=0}^{\lfloor n/2 \rfloor}(-1)^i\,\frac{1}{i!}\left(\frac{1}{2}\right)^i\frac{D_{(n-2i)}}{(n-2i)!}.$$

The integers

$$a(n) = D_n\,\mathbb{P}_1(\widetilde{C}_2(n) = 0), \quad n = 1,2,3,\ldots$$

give the number of derangements of n objects that have all cycles of length at least 3; computing the first few values gives

$$a(2) = 0, a(3) = 2, a(4) = 6, a(5) = 24, a(6) = 160, a(7) = 1140, a(8) = 8988, \ldots.$$

It is readily checked that this is (the start of) sequence A038205 in the Online Encyclopedia of Integer Sequences [18], where other formulae are provided.

2.1.3. The number of cycles

The distribution of the number of cycles, $\widetilde{K}_n$, may be found from the fact that the number $D(n,k)$ of derangements of size n having k cycles is

$$D(n,k) = \sum_{l=0}^{k} (-1)^l \binom{n}{l} \begin{bmatrix} n-l \\ k-l \end{bmatrix},$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ is the unsigned Stirling number of the first kind. Letting $|\pi|$ denote the number of cycles in π, it follows that

$$\begin{aligned} \mathbb{P}(\widetilde{K}_n = k) &= \frac{1}{\lambda_n(\theta)} \sum_{\substack{\pi:|\pi|=k \\ \pi \text{ a derangement}}} \mathbb{P}(\pi) \\ &= \frac{1}{\lambda_n(\theta)} \sum_{\substack{\pi:|\pi|=k \\ \pi \text{ a derangement}}} \frac{\theta^k}{\theta_{(n)}} \qquad \text{(from (2))} \\ &= \frac{\theta^k D(n,k)}{\lambda_n(\theta)\theta_{(n)}}, \quad k = 1, 2, \ldots, \lfloor n/2 \rfloor. \end{aligned}$$

The probability generating function of $\widetilde{K}_n$ satisfies

$$\mathbb{E}s^{\widetilde{K}_n} = \frac{\lambda_n(\theta s)}{\lambda_n(\theta)} \mathbb{E}s^{K_n}, \tag{10}$$

which follows from the fact that

$$\begin{aligned} \mathbb{E}s^{\widetilde{K}_n} &= \sum_{k=1}^{n} \mathbb{P}_\theta(\widetilde{K}_n = k) s^k = \sum_{k=1}^{n} \frac{(\theta s)^k D(n,k)}{\lambda_n(\theta)\theta_{(n)}} \\ &= \frac{1}{\lambda_n(\theta)\theta_{(n)}} \sum_{k=1}^{n} \sum_{\ell=0}^{k} (-1)^\ell \binom{n}{\ell} \begin{bmatrix} n-\ell \\ k-\ell \end{bmatrix} (\theta s)^k \\ &= \frac{1}{\lambda_n(\theta)\theta_{(n)}} \sum_{\ell=0}^{n} (-1)^\ell \binom{n}{\ell} (\theta s)^\ell \sum_{k=\ell}^{n} \begin{bmatrix} n-\ell \\ k-\ell \end{bmatrix} (\theta s)^{k-\ell} \\ &= \frac{1}{\lambda_n(\theta)\theta_{(n)}} \sum_{\ell=0}^{n} (-1)^\ell \binom{n}{\ell} (\theta s)^\ell (\theta s)_{(n-\ell)} \\ &= \frac{1}{\lambda_n(\theta)\theta_{(n)}} (\theta s)_{(n)} \lambda_n(\theta s), \qquad \text{from (7)} \\ &= \frac{\lambda_n(\theta s)}{\lambda_n(\theta)} \mathbb{E}s^{K_n}, \qquad \text{from (3)}. \end{aligned}$$

The mean of $\widetilde{K}_n$ is given by

$$\mathbb{E}\widetilde{K}_n = \mathbb{E}(\widetilde{C}_2(n) + \cdots + \widetilde{C}_n(n)) = \frac{1}{\lambda_n(\theta)} \frac{n!}{\theta_{(n)}} \sum_{j=2}^{n} \lambda_{n-j}(\theta) \frac{\theta_{(n-j)}}{(n-j)!} \frac{\theta}{j}.$$

It follows from (10) that

$$\mathbb{E}K_n - \mathbb{E}\widetilde{K}_n = -\frac{\theta \lambda_n'(\theta)}{\lambda_n(\theta)} > 0,$$

since $\lambda_n(\theta)$ is a decreasing function of θ for fixed n, in agreement with one's intuition.

2.1.4. Properties derived from the conditioning relation

$\mathrm{ESF}_n(\theta)$ may be represented as the law of independent Poisson random variables $Z_1, Z_2, \dots, Z_n$ with

$$\mathbb{E}Z_j = x^j\theta/j \text{ for any } x > 0, \tag{11}$$

conditioned on $T_n := Z_1 + 2Z_2 + \cdots + nZ_n = n$. This is known as the Conditioning Relation, and is exploited in the context of combinatorial structures in [6]. The same relationship holds for derangements too: defining $T_{1n} = 2Z_2 + \cdots + nZ_n$, we have

$$\mathcal{L}(\widetilde{C}_2(n), \dots, \widetilde{C}_n(n)) = \mathcal{L}(Z_2, \dots, Z_n \mid T_{1n} = n). \tag{12}$$

To see this, note that for $c_2 \geq 0, \dots, c_n \geq 0$ satisfying $2c_2 + \cdots + nc_n = n$,

$$\begin{aligned}
&\mathbb{P}(Z_2 = c_2, \dots, Z_n = c_n \mid T_{1n} = n) = \\
&= \frac{\mathbb{P}(Z_2 = c_2, \dots, Z_n = c_n, T_{1n} = n)\mathbb{P}(Z_1 = 0)\,/\,\mathbb{P}(T_n = n)}{\mathbb{P}(T_{1n} = n)\mathbb{P}(Z_1 = 0)\,/\,\mathbb{P}(T_n = n)} \\
&= \frac{\mathbb{P}(Z_1 = 0, Z_2 = c_2, \dots, Z_n = c_n \mid T_n = n)}{\mathbb{P}(Z_1 = 0 \mid T_n = n)} \\
&= \frac{\mathbb{P}(C_1(n) = 0, C_2(n) = c_2, \dots, C_n(n) = c_n)}{\mathbb{P}(C_1(n) = 0)},
\end{aligned}$$

the last equality following from the conditioning relation for $(C_1(n), \dots, C_n(n))$ and $C_1(n)$.

The relationship in (12) means that asymptotic results can be read off from the general theory in [6]. For example, the $\widetilde{C}_j(n)$ are asymptotically independent Poisson random variables with mean θ/j, which follows from (8) as well. We note for later use the consequence that

$$\lambda_n(\theta) := \mathbb{P}_\theta(C_1(n) = 0) \to \mathbb{P}_\theta(Z_1 = 0) = e^{-\theta}, \quad n \to \infty. \tag{13}$$

The largest cycles, when scaled by n, have asymptotically the Poisson–Dirichlet law with parameter θ. Total variation estimates for the Poisson result also follow from [6] and [5], and methods akin to those in [7] may be used to derive central limit results, for example. We will not pursue this aspect further in this paper.

3. A Markov chain for derangements of a given size

In the spirit of the Feller Coupling, we seek to construct a sequence of random variables $\overleftarrow{\eta}^{(n)} = (\overleftarrow{\eta}_n^{(n)}, \overleftarrow{\eta}_{n-1}^{(n)}, \dots, \overleftarrow{\eta}_1^{(n)} = 1)$ with the property that the law of the counts of spacings between the 1s in $1\,\overleftarrow{\eta}_n^{(n)}\overleftarrow{\eta}_{n-1}^{(n)} \dots \overleftarrow{\eta}_2^{(n)}\,1$ is precisely that of (5). As might be anticipated, $\overleftarrow{\eta}^{(n)}$ is no longer a sequence of independent random variables, but rather a Markov chain. We identify the structure of this chain, and provide some applications of its use.

3.1. Constructing the Markov chain

Define $\mathcal{R}_1 = \{(1)\}$ and for $j \geq 2$, $\mathcal{R}_j = \{0,1\}^{j-1} \times \{1\}$, including all $(a_1, \dots, a_j) \in \{0,1\}^j$ with $a_j = 1$. For $1 \leq i \leq n$ and $r = (r_n, \dots, r_1) \in \mathcal{R}_n$, let $N_i(r)$ be the number of i-spacings in $1, r_n, \dots, r_1$, i.e., the number of sub-patterns $10^{i-1}1$ in it, and define $\rho(a_1, \dots, a_n) = \{r \in \mathcal{R}_n : N_1(r) = a_1, \dots, N_n(r) = a_n\}$. We seek to construct a random sequence of 0s and 1s, $\overleftarrow{\eta}^{(n)} = (\overleftarrow{\eta}_n^{(n)}, \dots, \overleftarrow{\eta}_2^{(n)}, \overleftarrow{\eta}_1^{(n)} = 1)$ such that if $\widetilde{C}_i(n) := N_i(\overleftarrow{\eta}^{(n)})$, $i = 2, \dots, n$,

$$\mathbb{P}(\widetilde{C}_2(n) = c_2, \dots, \widetilde{C}_n(n) = c_n) = \mathbb{P}(C_1(n) = 0, \dots, C_n(n) = c_n | C_1(n) = 0). \tag{14}$$

Simplifying the r.h.s. of (14), we have

$$\mathbb{P}(\widetilde{C}_j(n) = c_j, 2 \le j \le n) = (\lambda_n(\theta))^{-1} \sum_{\substack{(r_n, r_{n-1}, \dots, r_1) \\ \in \rho(0, c_2, \dots, c_n)}} \mathbb{P}(\xi_n = r_n, \dots, \xi_1 = r_1).$$

Note that if $r = (r_n, r_{n-1}, \dots, r_1) \in \rho(0, c_2, \dots, c_n)$, then $r_n = r_2 = 0$. This suggests defining $\overleftarrow{\eta}_n^{(n)}, \overleftarrow{\eta}_{n-1}^{(n)}, \dots, \overleftarrow{\eta}_2^{(n)}, \overleftarrow{\eta}_1^{(n)} = 1$ with law

$$\begin{aligned} \mathbb{P}(\overleftarrow{\eta}_n^{(n)} = r_n, \dots, \overleftarrow{\eta}_1^{(n)} = r_1) &= \mathbb{P}(\xi_n = r_n, \dots, \xi_1 = r_1 \mid (\xi_n, \dots, \xi_1) \in \Delta_n) \\ &= \begin{cases} (\lambda_n(\theta))^{-1}\, \mathbb{P}(\xi_n = r_n, \dots, \xi_1 = r_1), & \text{if } r \in \Delta_n \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

By construction, $(\widetilde{C}_2(n), \dots, \widetilde{C}_n(n))$ has the law of $(C_1(n), \dots, C_n(n))$ conditioned on $C_1(n) = 0$. Since ξ_j are independent random variables, given $\overleftarrow{\eta}_i^{(n)}$, the vectors $(\overleftarrow{\eta}_n^{(n)}, \overleftarrow{\eta}_{n-1}^{(n)}, \dots, \overleftarrow{\eta}_{i+1}^{(n)})$ and $(\overleftarrow{\eta}_{i-1}^{(n)}, \dots, \overleftarrow{\eta}_1^{(n)})$ are independent and hence $\overleftarrow{\eta}_n^{(n)}, \overleftarrow{\eta}_{n-1}^{(n)}, \dots, \overleftarrow{\eta}_1^{(n)}$ is a Markov chain, starting from $\overleftarrow{\eta}_{n+1}^{(n)} = 1$.

More explicitly, for $3 \le i \le n-1$, $(r_n, r_{n-1}, \dots, r_{i+2}, x) \in \Delta^*_{n-i}$ and $y \in \{0, 1\}$, let

$$\begin{aligned} \tau_{i+1}(x, y) &:= \mathbb{P}(\overleftarrow{\eta}_i^{(n)} = y \mid \overleftarrow{\eta}_n^{(n)} = r_n, \dots, \overleftarrow{\eta}_{i+2}^{(n)} = r_{i+2}, \overleftarrow{\eta}_{i+1}^{(n)} = x) \\ &= \frac{\mathbb{P}(\overleftarrow{\eta}_n^{(n)} = r_n, \dots, \overleftarrow{\eta}_{i+2}^{(n)} = r_{i+2}, \overleftarrow{\eta}_{i+1}^{(n)} = x, \overleftarrow{\eta}_i^{(n)} = y)}{\mathbb{P}(\overleftarrow{\eta}_n^{(n)} = r_n, \dots, \overleftarrow{\eta}_{i+2}^{(n)} = r_{i+2}, \overleftarrow{\eta}_{i+1}^{(n)} = x)} \end{aligned}$$

We compute this for $x, y \in \{0, 1\}$. Starting with the case $x = y = 0$, and for $3 \le i \le n-1$, we write $\tau_{i+1}(0, 0) = A/B$, where

$$A = (\lambda_n(\theta))^{-1}\mathbb{P}(\xi_n = r_n, \dots, \xi_{i+2} = r_{i+2}, \xi_{i+1} = 0)\mathbb{P}((\xi_i, \dots, \xi_1) \in \Delta_i)$$

and

$$\begin{aligned} B &= (\lambda_n(\theta))^{-1}\mathbb{P}(\xi_n = r_n, \dots, \xi_{i+2} = r_{i+2}, \xi_{i+1} = 0) \\ &\quad \times \{\mathbb{P}((\xi_i, \dots, \xi_1) \in \Delta_i) + \mathbb{P}(\xi_i = 1)\mathbb{P}((\xi_{i-1}, \dots, \xi_1) \in \Delta_{i-1})\} \end{aligned}$$

so that

$$\tau_{i+1}(0, 0) = \frac{\lambda_i(\theta)}{\lambda_i(\theta) + \frac{\theta}{\theta+i-1}\lambda_{i-1}(\theta)}.$$

On the other hand,

$$\mathbb{P}(\overleftarrow{\eta}_i^{(n)} = 0 \mid \overleftarrow{\eta}_{i+1}^{(n)} = 0) = \mathbb{P}(\overleftarrow{\eta}_{i+1}^{(n)} = 0, \overleftarrow{\eta}_i^{(n)} = 0)/\mathbb{P}(\overleftarrow{\eta}_{i+1}^{(n)} = 0) = C/D,$$

where

$$C = (\lambda_n(\theta))^{-1}\, \mathbb{P}((\xi_n, \dots, \xi_{i+2}) \in \Delta^*_{n-i-1}, \xi_{i+1} = 0)\mathbb{P}((\xi_i, \dots, \xi_1) \in \Delta_i)$$

and

$$\begin{aligned} D &= (\lambda_n(\theta))^{-1}\, \mathbb{P}((\xi_n, \dots, \xi_{i+2}) \in \Delta^*_{n-i-1}, \xi_{i+1} = 0) \\ &\quad \times \{\mathbb{P}((\xi_i, \dots, \xi_1) \in \Delta_i) + \mathbb{P}(\xi_i = 1)\mathbb{P}((\xi_{i-1}, \dots, \xi_1) \in \Delta_{i-1})\} \end{aligned}$$

Hence

$$\mathbb{P}(\overleftarrow{\eta}_i^{(n)} = 0 \mid \overleftarrow{\eta}_{i+1}^{(n)} = 0) = \frac{\lambda_i(\theta)}{\lambda_i(\theta) + \frac{\theta}{\theta+i-1}\lambda_{i-1}(\theta)}.$$

Similarly we can deduce that

$$\tau_{i+1}(0,1) = \frac{\frac{\theta\lambda_{i-1}(\theta)}{\theta+i-1}}{\lambda_i(\theta) + \frac{\theta\lambda_{i-1}(\theta)}{\theta+i-1}} = \mathbb{P}(\overleftarrow{\eta}_i^{(n)} = 1 \mid \overleftarrow{\eta}_{i+1}^{(n)} = 0),$$

$\tau_{i+1}(1,1) = 0 = \mathbb{P}(\overleftarrow{\eta}_i^{(n)} = 1 \mid \overleftarrow{\eta}_{i+1}^{(n)} = 1)$ and $\tau_{i+1}(1,0) = 1 = \mathbb{P}(\overleftarrow{\eta}_i^{(n)} = 0 | \overleftarrow{\eta}_{i+1}^{(n)} = 1)$.

We summarize the discussion as follows.

Theorem 1. *(i) For each $n \geq 3$, the sequence of random variables $\overleftarrow{\eta}_{n+1}^{(n)} = 1, \overleftarrow{\eta}_n^{(n)}, \ldots, \overleftarrow{\eta}_2^{(n)}$, $\overleftarrow{\eta}_1^{(n)} = 1$ is a non-homogeneous Markov chain with transition matrices*

$$\overleftarrow{P}_r^{(n)} := \begin{pmatrix} \mathbb{P}(\overleftarrow{\eta}_r^{(n)} = 0 \mid \overleftarrow{\eta}_{r+1}^{(n)} = 0) & \mathbb{P}(\overleftarrow{\eta}_r^{(n)} = 1 \mid \overleftarrow{\eta}_{r+1}^{(n)} = 0) \\ \mathbb{P}(\overleftarrow{\eta}_r^{(n)} = 0 \mid \overleftarrow{\eta}_{r+1}^{(n)} = 1) & \mathbb{P}(\overleftarrow{\eta}_r^{(n)} = 1 \mid \overleftarrow{\eta}_{r+1}^{(n)} = 1) \end{pmatrix}$$

given by

$$\overleftarrow{P}_r^{(n)} = \begin{pmatrix} \dfrac{(\theta+r-1)\lambda_r(\theta)}{(\theta+r-1)\lambda_r(\theta)+\theta\lambda_{r-1}(\theta)} & \dfrac{\theta\lambda_{r-1}(\theta)}{(\theta+r-1)\lambda_r(\theta)+\theta\lambda_{r-1}(\theta)} \\ 1 & 0 \end{pmatrix}, \tag{15}$$

for $2 < r < n$,

$$\overleftarrow{P}_n^{(n)} = \overleftarrow{P}_2^{(n)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \overleftarrow{P}_1^{(n)} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

(ii) The counts $\widetilde{C}_j(n)$, $j = 2, \ldots, n$ of the j-spacings between consecutive 1s in the sequence $1\overleftarrow{\eta}_n^{(n)} \ldots \overleftarrow{\eta}_2^{(n)}1$ have joint distribution given by

$$\mathcal{L}(\widetilde{C}_2(n), \ldots, \widetilde{C}_n(n)) = \mathcal{L}(C_2(n), \ldots, C_n(n) \mid C_1(n) = 0). \tag{16}$$

4. Coupling derangements

The chain $\overleftarrow{\eta}^{(n)}$ does not generate derangements of size $n+1$ from one of size n, a property satisfied by the Feller Coupling. Rather, the chain produces a derangement for a given value of n, and needs to be re-run to generate one of size $n+1$. This constraint raises a natural question: Is there an infinite Markov chain that somehow serves as the limit of the $\overleftarrow{\eta}^{(n)}$ process and can generate approximately θ-biased random derangements of relatively large size while, on the other hand, it can use the derangements of size n to generate one of size $n+1$. The ensuing discussion explores this further and explains how the reversed Markov chain for $\overleftarrow{\eta}^{(n)}$, denoted by $\eta^{(n)}$, provides an appropriate framework to achieve this goal.

4.1. The finite reversed Markov chain

Having an outcome of $\overleftarrow{\eta}^{(n)}$, it is impossible to construct a θ-biased random derangement $\overleftarrow{\eta}^{(n+1)}$ with $\overleftarrow{\eta}^{(n+1)}[n, \ldots, 1] = \overleftarrow{\eta}^{(n)}[n, \ldots, 1]$. (The $[\ldots]$ notation is used to describe the indices of a vector.) In fact there is an important constraint in the Markov chain $\overleftarrow{\eta}^{(n)}$ – it always starts at $\overleftarrow{\eta}_n^{(n)} = 0$, a property that should also hold for $\overleftarrow{\eta}^{(n+1)}$. It turns out that the resulting sequence that uses a realization of the random derangement of size n to construct a derangement of size $n+1$ should be the trivial 0-sequence. A simple trick can be used to overcome this problem. Define the probability measure ν_n on $\{0, 1\}$ by $\nu_n(0) = \mathbb{P}_\theta(\overleftarrow{\eta}_n^{(n+1)} = 0)$ and slightly modify the Markov chain $\overleftarrow{\eta}^{(n+1)}$ to start at time n with the initial distribution ν_n, rather than starting at time $n+1$ with the initial distribution δ_0. Indeed, this Markov chain can be considered the same as the original one, since we always have $\overleftarrow{\eta}^{(n+1)}_{n+1} = 0$. It is clear that

$$\mathbb{P}_\theta(\overleftarrow{\eta}^{(n)}_i = 0|\overleftarrow{\eta}^{(n)}_{i+1} = 0) = \mathbb{P}_\theta(\overleftarrow{\eta}^{(m)}_i = 0|\overleftarrow{\eta}^{(m)}_{i+1} = 0),$$

for $i+1 < m \le n$, while $\nu_m \ne \nu_n$ for $m \ne n$. This will not cause a problem since in fact $\nu_n \Rightarrow \delta_0$ as n tends to infinity, which ensures the existence of a probability distribution on $\{0,1\}^{\mathbb{N}}$ as the limit of $\mathcal{L}(\overleftarrow{\eta}^{(n)})$.

To describe the limit distribution, we make use of the Markov chains $\eta^{(n)} = (\eta^{(n)}_i)_{i=1}^n$, $n \in \mathbb{N}$, which are the reversed chains for $\overleftarrow{\eta}^{(n)}$, whose laws are the same as those of $(\overleftarrow{\eta}^{(n)}_i)_{i=1}^n$, and whose transition probabilities are given via the time reversal transformations

$$\mathbb{P}(\eta^{(n)}_{i+1} = y|\eta^{(n)}_i = x) = \frac{\mathbb{P}(\overleftarrow{\eta}^{(n)}_{i+1} = y)\mathbb{P}(\overleftarrow{\eta}^{(n)}_i = x|\overleftarrow{\eta}^{(n)}_{i+1} = y)}{\mathbb{P}(\overleftarrow{\eta}^{(n)}_i = x)}.$$

We emphasize that $\eta^{(n)}_i \sim \overleftarrow{\eta}^{(n)}_i$, for $1 \le i \le n$, but we require different notation $\eta^{(n)} = (\eta^{(n)}_i)_{i=1}^n$ and $\overleftarrow{\eta}^{(n)} = (\overleftarrow{\eta}^{(n)}_{n-i+1})_{i=1}^n$ to describe two different Markov chains, which are reversals of each other. Note that the chains $\eta^{(n)}$ and $\overleftarrow{\eta}^{(n)}$ are not reversible, and hence have different laws; $(\xi_i)_{i=1}^n \nsim (\xi_{n-i+1})_{i=1}^n$. A limit distribution arises as the law of a $\{0,1\}$-valued infinite dimensional Markov chain. To identify $\eta^{(n)}$, let

$$\begin{aligned}\lambda_{i,n}(\theta) =& \mathbb{P}_\theta(\xi[i \dots n] \in \Delta^*_{n-i+1} \setminus \Delta_{n-i+1}) \\ =& \mathbb{P}_\theta(\xi_i = \xi_n = 0, \sum_{j=i+1}^{n-1} \xi_j \xi_{j+1} = 0),\end{aligned}$$

where the last term is the probability that $\xi_i = \xi_n = 0$ and there are no 11 patterns (which counts the number of 1-cycles) in $\xi[i, \dots, n]$. Similarly to the calculations given in Section 3, one can see that $\eta^{(n)}_1 = 1$, $\eta^{(n)}_n = 0$, and the transition probabilities of $\eta^{(n)}$ satisfy

$$\mathbb{P}_\theta(\eta^{(n)}_{i+1} = 0|\eta^{(n)}_i = 0) = \frac{\lambda_{i+1,n}(\theta)}{\lambda_{i+1,n}(\theta) + \frac{\theta}{i+\theta}\lambda_{i+2,n}(\theta)},$$

for $i \le n-2$. This is summarized in the following theorem.

Theorem 2. *(i) For each $n \ge 3$, the sequence of random variables $\eta^{(n)}_1 = 1, \eta^{(n)}_2, \dots, \eta^{(n)}_n = 0$ is a non-homogeneous Markov chain with transition matrices*

$$P^{(n)}_s := \begin{pmatrix} \mathbb{P}(\eta^{(n)}_{s+1} = 0 \mid \eta^{(n)}_s = 0) & \mathbb{P}(\eta^{(n)}_{s+1} = 1 \mid \eta^{(n)}_s = 0) \\ \mathbb{P}(\eta^{(n)}_{s+1} = 0 \mid \eta^{(n)}_s = 1) & \mathbb{P}(\eta^{(n)}_{s+1} = 1 \mid \eta^{(n)}_s = 1) \end{pmatrix}$$

given by

$$P^{(n)}_s = \begin{pmatrix} \dfrac{(s+\theta)\lambda_{s+1,n}(\theta)}{(s+\theta)\lambda_{s+1,n}(\theta) + \theta\lambda_{s+2,n}(\theta)} & \dfrac{\theta\lambda_{s+2,n}(\theta)}{(s+\theta)\lambda_{s+1,n}(\theta) + \theta\lambda_{s+2,n}(\theta)} \\ 1 & 0 \end{pmatrix},$$

for $s = 1, \dots, n-2$, and

$$P^{(n)}_{n-1} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

(ii) $\mathbb{P}_\theta(\eta_i^{(n)} = x_i, i = 1, \ldots, n) = \mathbb{P}_\theta(\overleftarrow{\eta}_i^{(n)} = x_i, i = 1, \ldots, n)$.

(iii) The counts $\widetilde{C}_j(n)$, $j = 2, \ldots, n$ *of the j-spacings between consecutive 1s in the sequence* $\eta_1 = 1, \ldots, \eta_n = 0, 1$ *have joint distribution given by* (16).

For $i \le n$, we define

$$\tilde{\lambda}_{n,i}(\theta) = \mathbb{P}_\theta(\xi[n, \ldots, i] \in \Delta_{n-i+1}).$$

which implies $\tilde{\lambda}_{n,n}(\theta) = \lambda_{1,n}(\theta) = 0$ and $\tilde{\lambda}_{n,1}(\theta) = \lambda_{2,n}(\theta) = \lambda_n(\theta)$, for any $n \in \mathbb{N}$.

Theorem 3. *For* $2 \le i < n$,

$$\begin{aligned}\tilde{\lambda}_{n,i}(\theta) =& \frac{\theta}{i-1+\theta}\lambda_{i+1,n}(\theta) \\ =& \frac{(n-1)!\theta_{(i-1)}}{(i-1)!\theta_{(n)}} \sum_{k=0}^{\lfloor \frac{n-i-1}{2} \rfloor} \sum_{i_0,\ldots,i_k} \frac{\theta^{k+1}}{i_0 \ldots i_k},\end{aligned} \tag{17}$$

where the inner sum is over all $i_0 = 1, i+1 \le i_1 < \cdots < i_k \le n-2$ *and* $i_{j-1} - i_j > 1$. *Moreover, for any* $i \le n$

$$\mathbb{P}_\theta(\overleftarrow{\eta}_i^{(n)} = 1) = \frac{\tilde{\lambda}_{n,i}(\theta)\lambda_{i-1}(\theta)}{\lambda_n(\theta)}.$$

Proof. The first equality is straightforward, and follows from

$$\mathbb{P}_\theta(\xi[n, \ldots, i] \in \Delta_{n-i+1}) = \mathbb{P}_\theta(\xi[i+1, \ldots, n] \in \Delta^*_{n-i} \setminus \Delta_{n-i})\mathbb{P}_\theta(\xi_i = 1).$$

To see the second equality in (17), let $A_j(n)$ denote the length of the jth cycle in the Feller coupling in the order of formation of its cycles. Note that for specific $a_1, \ldots, a_k > 1$ with $a_1 + \cdots + a_k = n - i + 1$, the probability that a θ-biased random permutation sampled from the Feller coupling process has k cycles from time n to time i s.t. $\xi_i = 1$ and $(A_1(n), \ldots, A_k(n)) = (a_1, \ldots, a_k)$, and has k' cycles from time $i-1$ to time 1 is given by

$$\frac{\theta^k\theta_{(i-1)}}{\theta_{(n)}} \cdot \frac{\theta^{k'}}{\theta_{(i-1)}}.$$

On the other hand, the number of such permutations is given by

$$\begin{aligned}&\binom{n-1}{a_1-1}(a_1-1)!\prod_{\ell=2}^{k}\binom{n-1-a_1-\cdots-a_{l-1}}{a_\ell-1}(a_\ell-1)!\begin{bmatrix} i-1 \\ k' \end{bmatrix} \\ &= \frac{(n-1)!}{(i-1)!(n-a_1)(n-a_1-a_2)\cdots(n-a_1-\cdots-a_{k-1})}\begin{bmatrix} i-1 \\ k' \end{bmatrix}\end{aligned}$$

Since

$$\sum_{k'=1}^{i-1}\begin{bmatrix} i-1 \\ k' \end{bmatrix}\theta^{k'} = \theta_{(i-1)},$$

summing over all possible k, k', and $(a_1, \ldots, a_k)$ with $a_j > 1$ for $j = 1, \ldots, k$ and $a_1 + \cdots + a_k = n - i + 1$ and letting $i_j = n - a_1 - \cdots - a_j$ gives the second equality.

For the last equality, we have

$$\begin{aligned}\mathbb{P}_\theta(\overleftarrow{\eta}_i^{(n)} = 1) =& \frac{\mathbb{P}_\theta(\xi_i = 1, \xi[n, \dots, 1] \in \Delta_n)}{\mathbb{P}_\theta(\xi[n, \dots, 1] \in \Delta_n)} \\ =& \frac{\mathbb{P}(\xi[n, \dots, i] \in \Delta_{n-i+1})\mathbb{P}_\theta(\xi[i-1, \dots, 1] \in \Delta_{i-1})}{\mathbb{P}_\theta(\xi[n, \dots, 1] \in \Delta_n)} \\ =& \frac{\tilde{\lambda}_{n,i}(\theta)\lambda_{i-1}(\theta)}{\lambda_n(\theta)},\end{aligned}$$

completing the proof. □

4.2. The limit distribution for random derangements

To study the limit distribution of the laws of $\eta^{(n)}$, we need a better understanding of the behavior of the Feller coupling conditioned on $C_1(n) = 0$, for large n. The following facts from [5] play an important role. First, we recall that $C_j(n)$ can be represented as

$$C_j(n) = \sum_{i=1}^{n-j} \xi_i(1-\xi_{i+1})\cdots(1-\xi_{i+j-1})\xi_{i+j} + \xi_{n-j+1}(1-\xi_{n-j+2})\cdots(1-\xi_n),$$

for $j \le n$, $C_j(n) = 0$ for $j > n$, and that

$$(C_1(n), C_2(n), \dots) \Rightarrow (Z_1, Z_2, \dots),$$

as $n \to \infty$, where for $j \in \mathbb{N}$,

$$Z_j = \sum_{i=1}^{\infty} \xi_i(1-\xi_{i+1})\cdots(1-\xi_{i+j-1})\xi_{i+j}$$

are independent Poisson random variables with means $\mathbb{E}Z_j = \theta/j$. In particular,

$$C_1(n) \Rightarrow Z_1 = \sum_{i=1}^{\infty} \xi_i\xi_{i+1},$$

so that

$$\lim_{n\to\infty} \lambda_n(\theta) = \lim_{n\to\infty} \mathbb{P}_\theta\left(\sum_{i=1}^{n} \xi_i\xi_{i+1} = 0\right) = \mathbb{P}_\theta\left(\sum_{i=1}^{\infty} \xi_i\xi_{i+1} = 0\right) = e^{-\theta}.$$

The preceding discussion suggests that the prime candidate for the limit distribution is the law of the Feller Coupling $\xi_1 = 1, \xi_2, \xi_3, \dots$ conditioned on $Z_1 = \sum_{i=1}^{\infty} \xi_i\xi_{i+1} = 0$.

The following theorem is the first step to see this more explicitly. Let Δ_∞ be the set of $(r_1, r_2, \dots) \in \{0, 1\}^{\mathbb{N}}$ s.t. $r_1 = 0$ and there is no $i \in \mathbb{N}$ for which $r_i = r_{i+1} = 1$, that is

$$\Delta_\infty = \left\{(r_1, r_2, \dots) \in \{0, 1\}^{\mathbb{N}} : r_1 + \sum_{i=1}^{\infty} r_i r_{i+1} = 0\right\}.$$

Let $\lambda_{i,\infty}(\theta) = \mathbb{P}_\theta((\xi_i, \xi_{i+1}, \dots) \in \Delta_\infty)$.

Theorem 4. *For any $i \in \mathbb{N}$ and $\theta > 0$,*

$$\lim_{n\to\infty} \lambda_{i,n}(\theta) = \lambda_{i,\infty}(\theta).$$

Proof. Since $\left\{\sum_{i=n}^{\infty}\xi_i\xi_{i+1}=0\right\}\subset\left\{\sum_{i=n+1}^{\infty}\xi_i\xi_{i+1}=0\right\}$, for $n\in\mathbb{N}$ we have

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}_\theta\left(\sum_{i=n}^{\infty}\xi_i\xi_{i+1}=0\right)&=\mathbb{P}_\theta\left(\bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty}\{\xi_i\xi_{i+1}=0\}\right)\\&=1-\mathbb{P}_\theta(\xi_n=1,\xi_{n+1}=1\ i.o.)=1,\end{aligned}$$

since

$$\sum_{n=0}^{\infty}\mathbb{P}_\theta(\xi_n=1,\xi_{n+1}=1)=\sum_{i=0}^{\infty}\frac{\theta^2}{(i+\theta)(i+1+\theta)}=\theta<\infty.$$

Hence

$$\lim_{n\to\infty}\lambda_{n,\infty}(\theta)=\lim_{n\to\infty}\mathbb{P}_\theta(\xi_n=0)\mathbb{P}_\theta(\sum_{i=n+1}^{\infty}\xi_i\xi_{i+1}=0)=1.$$

Therefore

$$\begin{aligned}\lim_{n\to\infty}\lambda_{i,n}(\theta)&=\lim_{n\to\infty}\{\lambda_{i,n}(\theta)\lambda_{n+1,\infty}(\theta)+\frac{\theta}{\theta+n}\lambda_{i,n}(\theta)\lambda_{n+2,\infty}(\theta)\}\\&=\lim_{n\to\infty}\lambda_{i,\infty}(\theta)=\lambda_{i,\infty}(\theta),\end{aligned}$$

completing the proof. □

It is now more clear why the infinite Feller coupling conditional on $Z_1=0$ is the limit of the finite Markov chains $\eta^{(n)}$. To be more explicit, let $\eta=(\eta_i)_{i\geq1}$ be a random {0, 1}-valued sequence with law

$$\begin{aligned}\mathbb{P}_\theta(\eta_n=r_n;n\in\mathbb{N})&=\mathbb{P}_\theta(\xi_n=r_n;n\in\mathbb{N}\mid Z_1=0)\\&=\begin{cases}e^\theta\,\mathbb{P}_\theta(\xi_n=r_n;n\in\mathbb{N}), & \text{if }(r_1,r_2,\ldots)\in\Delta_\infty\\0, & \text{otherwise.}\end{cases}\end{aligned}\tag{18}$$

It is straightforward to see that $\eta=(\eta_1,\eta_2,\ldots)$ is indeed a Markov chain, and its transition probabilities are given by

$$\begin{aligned}\mathbb{P}_\theta(\eta_{i+1}=0|\eta_i=0)=&\frac{\mathbb{P}_\theta(\eta_{i+1}=0,\eta_i=0)}{\mathbb{P}_\theta(\eta_i=0)}\\=&\frac{\mathbb{P}_\theta(\xi[i,\ldots,1]\in\Delta_i)\lambda_{i+1,\infty}(\theta)}{\mathbb{P}_\theta(\xi[i,\ldots,1]\in\Delta_i)(\lambda_{i+1,\infty}(\theta)+\frac{\theta}{i+\theta}\lambda_{i+2,\infty}(\theta))}\\=&\frac{\lambda_{i+1,\infty}(\theta)}{\lambda_{i+1,\infty}(\theta)+\frac{\theta}{i+\theta}\lambda_{i+2,\infty}(\theta)},\end{aligned}$$

which means that, for any $i\in\mathbb{N}$

$$\mathbb{P}_\theta(\eta_{i+1}^{(n)}=0|\eta_i^{(n)}=0)\to\mathbb{P}_\theta(\eta_{i+1}=0|\eta_i=0),$$

as $n\to\infty$. The following theorem follows from the above discussion.

Theorem 5. *The sequence of random variables* $\eta_1=1,\eta_2,\eta_3,\ldots$*, defined by* (18)*, is a non-homogeneous Markov chain with transition matrices*

$$P_s:=\begin{pmatrix}\mathbb{P}(\eta_{s+1}=0\mid\eta_s=0) & \mathbb{P}(\eta_{s+1}=1\mid\eta_s=0)\\\mathbb{P}(\eta_{s+1}=0\mid\eta_s=1) & \mathbb{P}(\eta_{s+1}=1\mid\eta_s=1)\end{pmatrix}$$

given by

$$P_s = \begin{pmatrix} \frac{(s+\theta)\lambda_{s+1,\infty}(\theta)}{(s+\theta)\lambda_{s+1,\infty}(\theta)+\theta\lambda_{s+2,\infty}(\theta)} & \frac{\theta\lambda_{s+2,\infty}(\theta)}{(s+\theta)\lambda_{s+1,\infty}(\theta)+\theta\lambda_{s+2,\infty}(\theta)} \\ 1 & 0 \end{pmatrix},$$

for $s \in \mathbb{N}$.

Note that $P_s^{(n)} \to P_s$ as $n \to \infty$. To see more, let μ_n be the measure on $\{0,1\}^n$ that determines the law of $\eta^{(n)}$ and let μ be the measure on $\{0,1\}^{\mathbb{N}}$ indicating the law of η. Let $\pi_n : \{0,1\}^{\mathbb{N}} \to \{0,1\}^n$ be the projection map on the first n coordinates of an infinite $\{0,1\}$-sequence, so that $\pi_n(a_1, a_2, \ldots) = (a_1, \ldots, a_n)$ and denote the image of μ under π_n by $\mu\pi_n^{-1}$. Then

Proposition 1. $d_{TV}(\mu_n, \mu\pi_n^{-1}) \to 0$, *as* $n \to \infty$.

Proof. Since $\lambda_{n,\infty} \to 1$ and $\mathbb{P}_\theta(\xi_n = 1) \to 0$ as $n \to \infty$, we have

$$\begin{aligned}\lim_{n\to\infty} d_{TV}(\mu_n, \mu\pi_n^{-1}) &= \lim_{n\to\infty} \frac{1}{2} \sum_{r\in\{0,1\}^n} |\mathbb{P}_\theta(\eta^{(n)} = r) - \mathbb{P}_\theta(\eta[1,\ldots,n] = r)| \\ &= \lim_{n\to\infty} \frac{1}{2} \sum_{r\in\Delta_n} \mathbb{P}_\theta(\xi[n,\ldots,1] = r) \left| \frac{1}{e^{-\theta}} - \frac{1}{\lambda_n(\theta)} \right| \\ &= \lim_{n\to\infty} \frac{1}{2} |\lambda_n(\theta)e^\theta - 1| = 0. \quad \square\end{aligned}$$

After some preliminaries in the next section, we provide an estimate for the total variation distance in Section 4.4.

4.3. *Finding the values of* $\lambda_{i,\infty}$

By conditioning on whether $\xi_{i+1} = 0$ or 1, we can derive the recursion

$$\begin{aligned}\lambda_{i,\infty}(\theta) &= \mathbb{P}_\theta(\xi_i = 0)(\lambda_{i+1,\infty}(\theta) + \mathbb{P}_\theta(\xi_{i+1} = 1)\lambda_{i+2,\infty}(\theta)) \\ &= \frac{i-1}{i-1+\theta}\left(\lambda_{i+1,\infty}(\theta) + \frac{\theta}{i+\theta}\lambda_{i+2,\infty}(\theta)\right).\end{aligned} \tag{19}$$

It seems difficult to solve (19) with the initial conditions for $\lambda_{2,\infty} = e^{-\theta}$ and $\lambda_{3,\infty}$. Instead we use a Poisson representation of the Feller Coupling provided in [3,8] and due originally to Svante Janson (see [3] for historical notes on this).

Consider a Poisson point process $\mathcal{N}$ on $(0, 1)$ with intensity function $x \mapsto \theta/x$ for $x \in (0, 1)$, and denote its points by $(\tau_i)_{i\geq 1}$, where $1 > \tau_1 > \tau_2 > \cdots > 0$. Let $\kappa(x)$ be a geometric random variable with probability of success x, and mark each point τ_i by a random finite sequence $R_i = 0^{\kappa(\tau_i)-1}1$. [8] shows that the law of the infinite random $\{0,1\}$-valued sequence $1R_1R_2\cdots$ is the same as that of the Feller coupling $\xi_1 = 1, \xi_2, \xi_3, \ldots$ with parameter θ. In fact, the number of Poisson points in $(0, x)$ marked by 1 (representing cycles of size 1), denoted by $\widetilde{N}_1(x)$, is a Poisson random variable with parameter

$$\mathbb{E}\widetilde{N}_1(x) = \int_0^x \frac{\theta}{y} y\, dy = \theta x.$$

Note that, for $i \geq 2$, $\xi_i + \sum_{j=i}^{\infty} \xi_j \xi_{j+1}$ counts the number of "1 1" in $1, \xi_i, \xi_{i+1}, \ldots$.

Before stating the next result, in which the values of $\lambda_{i,\infty}$ are given, recall that the confluent hypergeometric function with parameters $a, b, z \in \mathbb{C}$ is defined by

$$M(a,b,z) = \sum_{j=0}^{\infty} \frac{a_{(j)}}{b_{(j)}} \frac{z^j}{j!};$$

cf. [1, Chapter 13]. If Re $b >$ Re $a > 0$, its integral representation is given by

$$M(a,b,z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{zu} u^{a-1} (1-u)^{b-a-1} du.$$

Theorem 6. *For $i \geq 2$,*

$$\lambda_{i,\infty}(\theta) = M(1+\theta, \theta+i-1, -\theta).$$

Proof. First note that

$$\lambda_{2,\infty}(\theta) = \lim_{n\to\infty} \lambda_{2,n}(\theta) = \lim_{n\to\infty} \lambda_n(\theta) = M(1+\theta, 1+\theta, -\theta).$$

For $i \geq 3$, the probability that $\mathcal{N}$ has no points in $(x, 1)$ is

$$\exp\left(-\int_x^1 \frac{\theta}{y} dy\right) = x^\theta.$$

Therefore the probability density that there is no point of $\mathcal{N}$ in $(x,1)$, $\tau_1 = x$ marked by $0^{i-3}1$ (i.e. $R_1 = 0^{i-3}1$), and $\widetilde{N}_1(x) = 0$ is given by

$$x^\theta \frac{\theta}{x} x(1-x)^{i-3} e^{-\theta x}.$$

Thus

$$\mathbb{P}_\theta\left(\xi_j = 0, 1 < j < i-1, \xi_{i-1} = 1, \xi_i + \sum_{k=i}^{\infty} \xi_k \xi_{k+1} = 0\right)$$
$$= \int_0^1 \theta x^\theta (1-x)^{i-3} e^{-\theta x}\, dx. \tag{20}$$

But the left hand side of (20) is

$$\left(\prod_{j=1}^{i-3} \frac{j}{\theta+j}\right) \frac{\theta}{\theta+i-2} \lambda_{i,\infty}(\theta) = \frac{(i-3)!\theta^2}{\theta_{(i-1)}} \lambda_{i,\infty}(\theta),$$

which implies that

$$\lambda_{i,\infty} = \frac{\theta_{(i-1)}}{\theta(i-3)!} \int_0^1 x^\theta (1-x)^{i-3} e^{-\theta x} dx = M(\theta+1, \theta+i-1, -\theta). \quad \square$$

Remark. Note that the recursion in (19) is precisely one of the standard recursions for the confluent hypergeometric function; see [1, 13.4.2, p. 506].

4.4. Estimating the total variation distance

Theorem 7. *As $n \to \infty$,*

$$d_{TV}(\mu_n, \mu\pi_n^{-1}) = \frac{\theta}{n} - \frac{\theta(3\theta-1)}{n^2} + O(n^{-3}).$$

To prove this, we need the following lemmas.

Lemma 1. *For $n \geq 2$,*

$$d_{TV}(\mu_n, \mu\pi_n^{-1}) = \frac{1}{2}\left|e^{\theta}\lambda_n(\theta)\lambda_{n,\infty}(\theta)\left(1+\frac{\theta}{n-1}\right)-1\right| + \frac{\theta e^{\theta}\lambda_{n-1}(\theta)\lambda_{n+1,\infty}(\theta)}{2(n-1+\theta)}.$$

Proof.

$$\begin{aligned}
d_{TV}(\mu_n,\mu\pi_n^{-1}) &= \frac{1}{2}\sum_{r\in\{0,1\}^n} |\mathbb{P}(\eta[1,\ldots,n]=r) - \mathbb{P}(\eta^{(n)}=r)| \\
&= \frac{1}{2}\sum_{r\in\Delta_n}\left|e^{\theta}\mathbb{P}(\xi[n,\ldots,1]=r, \sum_{j=n+1}^{\infty}\xi_j\xi_{j+1}=0) - \frac{\mathbb{P}(\xi[n,\ldots,1]=r)}{\lambda_n(\theta)}\right| \\
&\quad + \frac{1}{2}\sum_{r'\in\Delta_{n-1}} e^{\theta}\mathbb{P}(\xi[n-1,\ldots,1]=r', \xi_n=1, \xi_{n+1}+\sum_{j=n+1}^{\infty}\xi_j\xi_{j+1}=0) \\
&= \frac{1}{2}\left|\frac{e^{\theta}\lambda_{n,\infty}(\theta)}{\mathbb{P}(\xi_n=0)} - \frac{1}{\lambda_n(\theta)}\right|\sum_{r\in\Delta_n}\mathbb{P}(\xi[n,\ldots,1]=r) \\
&\quad + \frac{1}{2}e^{\theta}\lambda_{n+1,\infty}(\theta)\mathbb{P}(\xi_n=1)\sum_{r'\in\Delta_{n-1}}\mathbb{P}(\xi[n-1,\ldots,1]=r') \\
&= \frac{1}{2}\left|\frac{e^{\theta}\lambda_n(\theta)\lambda_{n,\infty}(\theta)(n-1+\theta)}{n-1} - 1\right| + \frac{\theta e^{\theta}\lambda_{n-1}(\theta)\lambda_{n+1,\infty}(\theta)}{2(n-1+\theta)},
\end{aligned}$$

as required. □

The following lemma estimates $\lambda_n(\theta)$ and $\lambda_{n,\infty}(\theta)$.

Lemma 2. *As $n \to \infty$,*

$$\begin{aligned}
\lambda_n(\theta) &= e^{-\theta}\left(1+\frac{\theta(\theta-1)}{n}+\frac{\theta(\theta-1)(\theta-2-\sqrt{2})(\theta-2+\sqrt{2})}{2n^2}\right)+O(n^{-3}). \\
\lambda_{n,\infty}(\theta) &= 1-\frac{\theta(\theta+1)}{n}+\frac{\theta(\theta+1)(\theta+2+\sqrt{6})(\theta+2-\sqrt{6})}{2n^2}+O(n^{-3}).
\end{aligned} \tag{21}$$

Proof. Noting that $a^{[r]} = (-1)^r(-a)_{(r)}$, we can write

$$\begin{aligned}
\lambda_n(\theta) &= \sum_{j=0}^{n}(-1)^j\frac{\theta^j}{j!}\frac{n^{[j]}}{(n+\theta-1)^{[j]}} \\
&= \sum_{j=0}^{n}\frac{(-\theta)^j}{j!}\frac{(-n)_{(j)}}{(1-n-\theta)_{(j)}}
\end{aligned}$$

$$
\begin{aligned}
&= M(-n, -n-\theta+1, -\theta) \\
&= e^{-\theta}\ M(1-\theta, -n-\theta+1, \theta),
\end{aligned}
\tag{22}
$$

the last equality derived from Kummer's transformation. Making use of Kummer's series expansion for the last line of (22), we can represent $\lambda_n(\theta)$ by the infinite sum

$$
\lambda_n(\theta) = e^{-\theta} \sum_{j=0}^{\infty} \frac{\theta^j}{j!} \frac{(1-\theta)_{(j)}}{(-n-\theta+1)_{(j)}} = e^{-\theta} \sum_{j=0}^{\infty} \binom{\theta-1}{j} \frac{\theta^j}{(n+\theta-1)^{[j]}},
$$

for $\theta \notin \mathbb{N}$. For $\theta \in \mathbb{N}$, since $(\theta-1)_{(\theta)} = 0$, we notice that all but the first θ terms vanish in the above representation. Let $v_j(n,\theta) = (n+\theta-1)^{[j]} - n^j$, for $j \in \mathbb{Z}_+$. Then

$$
\frac{1}{(n+\theta-1)^{[j]}} = \frac{1}{n^j\left(1+\frac{v_j(n,\theta)}{n^j}\right)} = \frac{1}{n^j}\sum_{i=0}^{\infty}(-1)^i\left(\frac{v_j(n,\theta)}{n^j}\right)^i,
$$

which implies

$$
\lambda_n(\theta) = e^{-\theta} \sum_{j=0}^{\infty} \binom{\theta-1}{j} \frac{\theta^j}{n^j} \sum_{i=0}^{\infty}(-1)^i\left(\frac{v_j(n,\theta)}{n^j}\right)^i. \tag{23}
$$

Similarly, letting $\tilde{v}_j(n,\theta) = (n+\theta-1)_{(j)} - n^j$, for $j \in \mathbb{Z}_+$, we have

$$
\frac{1}{(n+\theta-1)_{(j)}} = \frac{1}{n^j}\sum_{i=0}^{\infty}(-1)^i\left(\frac{\tilde{v}_j(n,\theta)}{n^j}\right)^i,
$$

and therefore

$$
\begin{aligned}
\lambda_{n,\infty}(\theta) &= M(\theta+1, n+\theta-1, -\theta) \\
&= \sum_{j=0}^{\infty} \frac{(-\theta)^j(\theta+1)_{(j)}}{j!(n+\theta-1)_{(j)}} \\
&= \sum_{j=0}^{\infty} \frac{(-\theta)^j(\theta+1)_{(j)}}{n^j j!} \sum_{i=0}^{\infty}(-1)^i\left(\frac{\tilde{v}_j(n,\theta)}{n^j}\right)^i.
\end{aligned}
\tag{24}
$$

To find the contribution of n^{-k} in $\lambda_n(\theta)$ and $\lambda_{n,\infty}(\theta)$, it suffices to find its contribution in the first $k+1$ terms in the right of (23) and (24), so that evaluating $v_0(n,\theta) = \tilde{v}_0(n,\theta) = 0$, $v_1(n,\theta) = \tilde{v}_1(n,\theta) = \theta - 1$, for $k = 1, 2$, we obtain

$$
\begin{aligned}
\lambda_n(\theta) &= e^{-\theta} + \frac{e^{-\theta}}{n}\binom{\theta-1}{1}\theta\left(1 - \frac{\theta-1}{n} + O(n^{-2})\right) \\
&\quad + \frac{e^{-\theta}}{n^2}\binom{\theta-1}{2}\theta^2\left(1 - O(n^{-1})\right) + O(n^{-3}) \\
&= e^{-\theta} + \frac{\theta(\theta-1)e^{-\theta}}{n} + \frac{e^{-\theta}\theta(\theta-1)\left(\frac{\theta^2}{2} - 2\theta + 1\right)}{n^2} + O(n^{-3}),
\end{aligned}
$$

and similarly,

$$\begin{aligned}\lambda_{n,\infty}(\theta) &= 1 + \frac{-\theta(\theta+1)}{n}\left(1 - \frac{\theta-1}{n} + O(n^{-2})\right)\\ &\quad + \frac{(-\theta)^2(\theta+1)(\theta+2)}{2!n^2}\left(1 - O(n^{-1})\right) + O(n^{-3})\\ &= 1 - \frac{\theta(\theta+1)}{n} + \frac{\theta(\theta+1)\left(\frac{\theta^2}{2}+2\theta-1\right)}{n^2} + O(n^{-3}). \quad \square\end{aligned}$$

Remark 2. For $\theta = 1$, the second and third terms in the right hand side of (21), for $\lambda_n(\theta)$, vanish. In fact, we can see much more in this case, using the power series of the exponential function. More precisely,

$$\lambda_n(1) = \sum_{j=0}^{n} \frac{(-1)^j}{j!} = e^{-1} - \sum_{j=n+1}^{\infty} \frac{(-1)^j}{j!} = e^{-1} + O\left(\frac{1}{(n+1)!}\right).$$

We are now ready to prove Theorem 7.

Proof of Theorem 7. It follows from Lemma 2 that, as $n \to \infty$,

$$\begin{aligned}\left(1+\frac{\theta}{n-1}\right)e^{\theta}\lambda_n(\theta)\lambda_{n,\infty}(\theta) &= \left(1 + \frac{\theta}{n} + \frac{\theta}{n^2} + O(n^{-3})\right)\\ &\quad \times \left(1 + \frac{\theta(\theta-1)}{n} + \frac{\theta(\theta-1)(\theta^2/2-2\theta+1)}{n^2} + O(n^{-3})\right)\\ &\quad \times \left(1 - \frac{\theta(\theta+1)}{n} + \frac{\theta(\theta+1)(\theta^2/2+2\theta-1)}{n^2} + O(n^{-3})\right)\\ &= 1 - \frac{\theta}{n} + \frac{\theta(3\theta-1)}{n^2} + O(n^{-3}),\end{aligned}$$

$$\begin{aligned}\frac{\theta e^{\theta}\lambda_{n-1}(\theta)\lambda_{n+1,\infty}(\theta)}{n-1+\theta} &= \theta\left(1 + \frac{\theta(\theta-1)}{n} + O(n^{-2})\right)\\ &\quad \times \left(1 - \frac{\theta(\theta+1)}{n} + O(n^{-2})\right)\left(\frac{1}{n} + \frac{1-\theta}{n^2} + O(n^{-3})\right)\\ &= \frac{\theta}{n} - \frac{\theta(3\theta-1)}{n^2} + O(n^{-3}).\end{aligned}$$

Therefore, from Lemma 1, as $n \to \infty$,

$$\begin{aligned}d_{TV}(\mu_n, \mu\pi_n^{-1}) &= \frac{1}{2}\left(\frac{\theta}{n} - \frac{\theta(3\theta-1)}{n^2}\right) + \frac{1}{2}\left(\frac{\theta}{n} - \frac{\theta(3\theta-1)}{n^2}\right) + O(n^{-3})\\ &= \frac{\theta}{n} - \frac{\theta(3\theta-1)}{n^2} + O(n^{-3}),\end{aligned}$$

as was to be shown. $\square$

In the following sections, we exploit the limiting process to produce coupled simulations of derangements.

5. Simulating derangements

While we have a good understanding of the asymptotics of the distribution of cycle counts, for small values of n simulation may be a useful approach to answer more detailed questions where explicit results are hard to find. Simulating derangements for the uniform case ($\theta = 1$) is a classical problem, and there have been many suggested methods, including [2,14] and [16] which use a modification of the Fisher–Yates algorithm for random permutations and a rejection step, and improved in [17]. [15] exploits two different techniques, one based on random restricted transpositions and one on sequential importance sampling. We are not aware of explicit methods for the case of arbitrary θ, but the Markov chain approach provides an efficient way to do this.

In this section we discuss methods for simulating derangements for particular values of n and θ, and a coupling approach that can be used for a collection of n-values. Here we focus on the relative speed of the methods. In the following section, we compare the methods for computing properties of derangements via a collection of numerical experiments.

5.1. Rejection methods

There are at least two such methods. For example, we can use the Feller Coupling to simulate $(C_1(n), C_2(n), \ldots, C_n(n))$ from $\mathrm{ESF}_n(\theta)$ and set

$$(\widetilde{C}_2(n), \ldots, \widetilde{C}_n(n)) = (C_2(n), \ldots, C_n(n)) \text{ if } C_1(n) = 0,$$

producing an observation from (5). The acceptance probability is just $\lambda_n(\theta)$, which is $\approx e^{-\theta}$, so this strategy is slow if θ is large. Indicative results are shown in Table 1.

The Conditioning Relation (12) provides another approach: the naive implementation takes $x = 1$ in (11), and simulates independent Poisson random variables $Z_2, \ldots, Z_n$ with $\mathbb{E}Z_j = \theta/j$ and accepts $(Z_2, \ldots, Z_n)$ as an observation of the counts $(\widetilde{C}_2(n), \ldots, \widetilde{C}_n(n))$ if $T_{1n} = 2Z_2 + \cdots + nZ_n = n$. The acceptance probability is $\mathbb{P}(T_{1n} = n)$; for large n, [6, Theorem 4.13] shows that $n\mathbb{P}(T_{1n} = n) \sim e^{-\gamma\theta}/\Gamma(\theta)$, where γ is Euler's constant. We can do much better by adapting the argument in [4, Section 5] by choosing $x = x(n)$ more carefully: choose c as the solution of the equation $\theta(1 - e^{-c}) = c$, and set $x = e^{-c/n}$. We then have

$$n\mathbb{P}(T_{1n} = n) \sim e^{-\gamma\theta} e^{u(c)}/\Gamma(\theta), \quad n \to \infty, \tag{25}$$

where $u(c) = -c + \theta \int_0^1 v^{-1}(1 - e^{-cv})dv$. The quantity $e^{u(c)}$ is the asymptotic factor by which the acceptance rate increases compared to the naive rate when $x = 1, c = 0$. For example, when $\theta = 5$, this is 379.6, indicating a dramatic speed up over the naive version. Indicative results are shown in Table 2.

Thus one of these methods is slow for large n, the other for large θ. In contrast, the Markov chain approaches provide methods that are acceptable for any values of n and θ.

5.2. Simulating derangements via the Markov chain $\overleftarrow{\eta}^{(n)}$

It is straightforward to use the transition mechanism from Theorem 1 to generate a derangement from the spacings between the 1s in the Markovian sequence $\overleftarrow{\eta}^{(n)} = (1, \overleftarrow{\eta}_n, \overleftarrow{\eta}_{n-1}, \ldots \overleftarrow{\eta}_2, 1)$. Indicative results are shown in Table 3. As anticipated, the run time of the Markov chain method is essentially constant as a function of θ for a fixed value of n, a property obviously not shared by the rejection methods. Comparing timings of these methods (which were implemented in

Table 1
Rejection method. Derangements of size n, estimates based on 10,000 accepted runs.

n	$\theta = 0.5$ Time (s)	Accept rate	Theory (7)	$\theta = 1.0$ Time (s)	Accept rate	Theory (7)	$\theta = 5.0$ Time (s)	Accept rate	Theory (7)
10	0.38	0.590	0.591	0.62	0.372	0.368	9.58	0.024	0.023
50	1.40	0.607	0.604	2.17	0.372	0.368	86.19	0.010	0.010
250	6.89	0.600	0.606	10.71	0.367	0.368	532.1	0.007	0.007

Table 2
Conditioning relation method. Derangements of size n, estimates based on 10,000 accepted runs. Values of c: -1.256 ($\theta = 0.5$), 0 ($\theta = 1$), 4.965 ($\theta = 5$).

n	$\theta = 0.5$ Time (s)	Accept rate	Theory (25)	$\theta = 1.0$ Time (s)	Accept rate	Theory (25)	$\theta = 5.0$ Time (s)	Accept rate	Theory (25)
10	2.46	0.059	0.061	2.59	0.055	0.056	4.41	0.032	0.088
50	62.84	0.012	0.012	68.04	0.011	0.011	51.23	0.015	0.018
250	1730	0.002	0.002	1884	0.002	0.002	1223	0.003	0.004

R) depends of course on the details of the code and the computer they are run on (in this case, a 3 GHz iMac Pro with 10 cores, 128 Gb RAM, running OS 10.15.7 on a single core) so they should only be viewed as relative. It is interesting to note that the acceptance rate of the Conditioning Relation method is not monotone in θ, because of the nature of the conditioning event.

5.3. Coupling derangements via the infinite Markov chain η

One application of the infinite chain η is the coupled generation of derangements with asymptotically the correct distribution, as guaranteed by Proposition 1 and Theorem 7. For a single value of n, we generate the η chain for $n-1$ steps, and then set $\eta_n = 0, \eta_{n+1} = 1$ (this is the analog of the artificial boundary at $n+1$ for the Feller Coupling). The ordered cycle lengths of the derangement are then read of as the spacings between the 1s in the sequence $1, 0, \eta_{n-1}, \dots, \eta_1 = 1$. For coupled simulations of derangements of length $n_1 < n_2 < \dots < n_k$, generate an η chain of length at least n_k, and for each j, use the subsequence $1, 0, \eta_{n_j-1}, \dots, \eta_1 = 1$ as above to construct the cycle counts for the derangement of size n_j. The efficacy of this strategy is discussed in the next section.

5.4. Relative timing information

Here we report the time behavior of the algorithms, beginning with the rejection and conditioning relation methods illustrated in Tables 1 and 2.

Implementation of the chain $\overleftarrow{\eta}^{(n)}$ in Theorem 1 requires computation of the quantities $\lambda_n(\theta)$ in (7). The explicit formula there is subject to computational errors for large values of n and θ and might require high precision arithmetic (as available in the R package `Rmpfr` for example). An alternative, numerically stable approach is via recursion. A conditioning argument shows that for $n \geq 3$,

$$\lambda_n(\theta) = \frac{n-1}{\theta+n-1}\left(\frac{\theta}{\theta+n-2}\lambda_{n-2}(\theta) + \lambda_{n-1}(\theta)\right), \tag{26}$$

Table 3
Markov chain method using chains in Theorems 1 and 5. Derangements of size n, estimates based on 10,000 accepted runs. The runs for the infinite chain are separate for each n.

	Run time (s)					
	$\theta = 0.5$		$\theta = 1.0$		$\theta = 5.0$	
n	Finite MC	Infinite MC	Finite MC	Infinite MC	Finite MC	Infinite MC
10	0.273	0.231	0.213	0.233	0.210	0.224
50	0.918	0.972	0.947	0.919	0.875	0.879
250	4.358	4.445	4.369	4.558	4.249	4.262

Table 4
Finite and approximate infinite Markov chain methods. The total run time for derangements of sizes 50, 100, 150, 200, 250 and 300. Estimates based on 10,000 accepted runs. The finite Markov chain method was run separately for each n and the total time for that is the sum of the run times for different n. The runs were coupled for the approximate infinite Markov chain method, i.e., the simulation is for 299 steps and "01" is added to steps 49, 99, 149, 199, 249, 299.

	Total run time (s)		
	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
Finite MC	19.190	18.787	17.955
Infinite MC	6.263	6.124	6.332

with initial conditions $\lambda_1(\theta) = 0$, $\lambda_2(\theta) = 1/(\theta + 1)$. The results below used this method.

Implementation of the infinite chain method in Theorem 5 requires evaluation of the quantities $\lambda_{i,\infty}$ in Theorem 6. The recursion (19) can be unstable due to the alternating signs, and instead we employed the `kummerM` function from the R package `fAsianOptions`.

For an example of coupled simulation from the infinite chain, we computed the run time for derangements of size $n = 50, 100, 150, 200, 250$ and 300, first by summing the times for each derangement size separately, and comparing this to the time for a coupled run based on 300 steps. For a coupled run of length 300, we computed the derangement of size $m \leq 300$ by using $\eta_1, \ldots, \eta_{m-1}$, appending 0 1 to the end and computing the cycle sizes by calculating the spacings between the 1s. The comparison is given in Table 4.

We conclude this section by reporting the behavior of the exact simulation via $\overleftarrow{\eta}^{(n)}$ and the infinite chain η for simulating very large derangements. Implementing the finite chain using (7), (26) or even the `kummerM` function, although in principle exact, proved difficult due to the time taken to compute $\lambda_n(\theta)$. On the other hand the Feller Coupling via the η chain was feasible via the integral representation for $\lambda_{n,\infty}(\theta)$. Table 5 supports the view that the Feller Coupling is the only feasible method for generating the cycle counts of very large derangements.

6. Probabilistic examples

In this section we study some further properties of the cycle structure of derangements. In some of the examples explicit formulae are available for a given value of n; such examples provide a way to test the adequacy of the approximate coupling method of simulation. For some quantities explicit results are not known; for these, simulation is the only option. In some cases, asymptotic behavior is known as well. Numerical results for the following examples are given in Section 7.

Table 5
Markov chain method using chain in Theorem 5. Derangements of size n, estimates based on 10,000 accepted runs. The runs are separate for each n.

	Run time (s)		
n	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
10,000	212.7	207.1	215.5
20,000	426.6	410.4	390.2
30,000	597.8	608.3	616.2
40,000	801.1	809.8	807.8

6.1. The probability of a single cycle

Since

$$\mathbb{P}(\widetilde{C}_n(n) = 1) = \mathbb{P}(C_n(n) = 1 \mid C_1(n) = 0) = \mathbb{P}(C_n(n) = 1)/\mathbb{P}(C_1(n) = 0),$$

we obtain

$$\mathbb{P}(\widetilde{C}_n(n) = 1) = \frac{n!}{\theta_{(n)}} \frac{\theta}{n} \frac{1}{\lambda_n(\theta)}, \tag{27}$$

which also follows from (9) because $\mathbb{P}(\widetilde{C}_n(n) = 1) = \mathbb{E}\widetilde{C}_n(n)$.

The asymptotics of (27) follow readily, using the fact that $n^{-\alpha}\Gamma(n+\alpha)/\Gamma(n) \to 1$ as $n \to \infty$ to obtain

$$\mathbb{P}(\widetilde{C}_n(n) = 1) \sim \Gamma(\theta+1)\left(\frac{e}{n}\right)^{\theta}, \quad n \to \infty. \tag{28}$$

6.2. The probability that all cycle lengths are distinct

This is a variant of the problem discussed in [4], for which there is no easy analytical answer. The difference between the number of cycles and the number of distinct cycle lengths is

$$\widetilde{D}_n = \sum_{j=2}^{n} (\widetilde{C}_j(n) - 1)_+,$$

where $(x)_+ = \max(0, x)$. We want $\mathbb{P}(\widetilde{D}_n = 0)$, which can be estimated by simulation.

In [4, Eq. (10)] it is shown that for a permutation having the $\mathrm{ESF}_n(\theta)$ distribution, the asymptotic probability that it has no repeated cycle lengths is $e^{-\gamma\theta}/\Gamma(\theta+1)$. A modification of that argument shows that for derangements,

$$\widetilde{D}_n \Rightarrow \widetilde{D} = \sum_{j\geq 2} (Z_j - 1)_+,$$

where the Z_j are the familiar independent Poisson random variables with $\mathbb{E}Z_j = \theta/j$, so that

$$\begin{aligned}\mathbb{P}(\widetilde{D}_n = 0) \to \mathbb{P}(Z_j \leq 1, j \geq 2) &= \prod_{j\geq 2} e^{-\theta/j}(1+\theta/j) \\ &= \frac{1}{e^{-\theta}(1+\theta)} \frac{e^{-\gamma\theta}}{\Gamma(\theta+1)} = \frac{e^{-\theta(\gamma-1)}}{\Gamma(\theta+2)}.\end{aligned} \tag{29}$$

6.3. The ordered cycle lengths

The $\overleftarrow{\eta}^{(n)}$ process generates the lengths of cycles in an n-derangement in order, starting from the artificial boundary at $\overleftarrow{\eta}_{n+1} = 1$. Denoting the length of the first cycle by $A_1(n)$, we have

$$\begin{aligned}\mathbb{P}(A_1(n) > l) &= \mathbb{P}(\overleftarrow{\eta}_{n+1} = 1, \overleftarrow{\eta}_n = 0, \ldots, \overleftarrow{\eta}_{n-l+1} = 0)\\ &= \prod_{r=n-l+1}^{n-1} \frac{(\theta + r - 1)\lambda_r(\theta)}{(\theta + r - 1)\lambda_r(\theta) + \theta\lambda_{r-1}(\theta)}.\end{aligned}$$

When n is large, we have for $x \in (0, 1)$,

$$\begin{aligned}\log \mathbb{P}(A_1(n) > \lfloor nx \rfloor) &= -\sum_{r=n-\lfloor nx \rfloor+1}^{n-1} \log\left(1 + \frac{\theta}{\theta + r - 1}\frac{\lambda_{r-1}(\theta)}{\lambda_r(\theta)}\right)\\ &\sim -\theta \sum_{r=\lfloor n(1-x)\rfloor}^{n-1} \frac{1}{\theta + r - 1}\frac{\lambda_{r-1}(\theta)}{\lambda_r(\theta)}\\ &\sim -\theta \int_{1-x}^{1} u^{-1} du = \theta \log(1 - x),\end{aligned}$$

using (13). It follows that $n^{-1}A_1(n)$ has asymptotically a Beta distribution with density $\theta(1 - x)^{\theta-1}, 0 < x < 1$. The joint law of the ordered spacings may be used in a similar way to show directly that $n^{-1}(A_1(n), A_2(n), \ldots)$ has asymptotically the GEM distribution with parameter θ; see [6, Chapter 5.4].

7. Numerical experiments

In this section we assess the behavior of the exact simulation method in Theorem 1, and the coupled estimates from the approximate method in Theorem 5.[1]

7.1. Results from exact simulation using $\overleftarrow{\eta}^{(n)}$

In this section we use simulation using the finite chain $\overleftarrow{\eta}^{(n)}$ to estimate quantities discussed in Section 6. See Tables 6–8.

Table 6
Probability that a derangement has a single cycle. Estimates are based on 100,000 runs.

	$\theta = 0.5$			$\theta = 1.0$			$\theta = 5.0$		
n	Sim	Exact (27)	Asymp (28)	Sim	Exact (27)	Asymp (28)	Sim	Exact (27)	Asymp (28)
10	0.476	0.480	0.462	0.270	0.272	0.272	0.021	0.021	0.178
50	0.211	0.208	0.207	0.054	0.054	0.054	5×10^{-5}	3.29×10^{-5}	5.70×10^{-5}
250	0.092	0.093	0.092	0.011	0.011	0.011	0.0	1.62×10^{-8}	1.82×10^{-8}

[1] R code for the examples may be obtained from the authors.

Table 7
$\mathbb{P}(\widetilde{D}_n = 0)$, the probability that a derangement has distinct cycle lengths. Estimates are based on 100,000 runs. The last row comes from (29).

n	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
10	0.885	0.774	0.357
50	0.920	0.776	0.091
250	0.927	0.765	0.028
∞	0.929	0.763	0.012

Table 8
Estimates of the probability o_n that the largest cycle length is the first, the mean length $\mathbb{E}A_1(n)$ of the first cycle, and the mean length $\mathbb{E}L_1(n)$ of the longest cycle. Estimates are based on 100,000 runs.

	$\theta = 0.5$			$\theta = 1.0$			$\theta = 5.0$		
n	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$
10	0.847	7.64	8.16	0.766	6.45	7.17	0.604	3.91	4.79
50	0.775	34.33	38.43	0.652	26.51	32.15	0.356	10.78	16.99
250	0.761	167.8	190.2	0.630	126.70	157.00	0.311	44.27	76.58

Table 9
Probability that a derangement has a single cycle. Estimates are based on 100,000 runs. Compare to results in Table 6.

	$\theta = 0.5$		$\theta = 1.0$		$\theta = 5.0$	
n	Sim	Exact (27)	Sim	Exact (27)	Sim	Exact (27)
10	0.479	0.480	0.277	0.272	0.024	0.021
50	0.206	0.208	0.056	0.054	2×10^{-5}	3.29×10^{-5}
250	0.091	0.093	0.011	0.011	0.0	1.62×10^{-8}

7.2. Results from approximate simulation using η

In all of the Tables 9–11 we use the approximate infinite Markov chain method where runs are coupled. The values are very close to the corresponding ones for the finite Markov chain method illustrated in Section 7.1. The coupling generates the η chain for 250 steps, and uses the run to generate derangements of length 10, 50, 250 by appending 0 1 to the strings terminating at steps 9, 49, 249 respectively. Given the shorter run time of the coupled method, and the suggestion from the simulations in this section that the results from the approximate coupling method are sufficiently accurate, we recommend using this approach when many runs of different lengths are required.

8. Discussion

The seminal paper of Shepp and Lloyd [19] – listed by Larry as one of his “top 10” – studied the behavior of random permutations (the case $\theta = 1$ in our notation), focusing primarily on the rth largest and smallest cycle lengths. Their model is the first appearance of a conditioning relation, which gives

$$\mathrm{ESF}_n(1) = \mathcal{L}(Z_1, Z_2, \ldots \mid \sum_{i\geq 1} i Z_i = n)$$

Table 10
$\mathbb{P}(\tilde{D}_n = 0)$, the probability that a derangement has distinct cycle lengths. Estimates are based on 100,000 runs. The last row comes from (29). Compare to results in Table 7.

n	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
10	0.883	0.779	0.377
50	0.921	0.776	0.092
250	0.927	0.766	0.028
∞	0.929	0.763	0.012

Table 11
Estimates of the probability o_n that the largest cycle length is the first, the mean length $\mathbb{E}A_1(n)$ of the first cycle, and the mean length $\mathbb{E}L_1(n)$ of the longest cycle. Estimates are based on 100,000 runs. Compare to results of Table 8.

	$\theta = 0.5$			$\theta = 1.0$			$\theta = 5.0$		
n	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$	o_n	$\mathbb{E}A_1(n)$	$\mathbb{E}L_1(n)$
10	0.851	7.66	8.16	0.773	6.49	7.18	0.653	4.09	4.85
50	0.774	34.29	38.43	0.651	26.55	32.17	0.356	10.79	16.97
250	0.761	167.8	189.9	0.629	126.27	156.80	0.311	44.34	76.71

where the Z_i are independent Poisson random variables with $\mathbb{E}Z_i = z^i/i$, for $z \in (0, 1)$. The random variable $\sum_{i\geq 1} iZ_i$ is geometric with mean $z/(1-z)$. Their paper exploits a related Poisson process construction to uncover the asymptotics of the largest and smallest cycles.

Our work also focuses on permutations, albeit derangements. Our motivation was understanding how the Feller Coupling might be adapted to simulate derangements under the Ewens Sampling Formula with arbitrary parameter θ. Section 3 provides a $\{0, 1\}$-valued non-homogeneous Markov chain $\overleftarrow{\eta}^{(n)} = \overleftarrow{\eta}_n, \overleftarrow{\eta}_{n-1}, \ldots, \overleftarrow{\eta}_1 = 1$ for which the spacings between the 1s in $1\overleftarrow{\eta}_n\overleftarrow{\eta}_{n-1}\cdots\overleftarrow{\eta}_1$ produce the ordered cycle sizes of a θ-biased derangement of length n. For the uniform case $\theta = 1$, the method described in [17] may also be described as a Markov chain (although it was not in that paper), and its transition matrix reduces to that in (15) when $\theta = 1$; it is interesting to note that its construction differs dramatically from ours.

The chains $\overleftarrow{\eta}^{(n)}$ do not generate derangements of size $n + 1$ from one of size n, a property satisfied by the Feller Coupling (see the discussion after (5)). Rather, the chain produces a derangement for a given value of n, and needs to be re-run to generate one of size $n + 1$. In order to rectify this, we provide in Section 4 a construction of a single Markov process $\eta = (\eta_1, \eta_2, \ldots)$ which provides a way (see Sections 5 and 6) to generate coupled derangements from a single run of the chain. Proposition 1 and Theorem 7 indicate that these derangements have asymptotically the correct distribution.

While we have focused here on the behavior of counts of cycle lengths, the chains $\overleftarrow{\eta}^{(n)}$ and η may be used to construct the ordered permutation itself by a simple auxiliary randomization [6, Chapter 5]. We discuss various methods for generating derangements, and compare them by evaluating a number of functionals of the cycle counts. The coupled derangements produced by η are shown to behave well, even for small derangement sizes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

PHdS and ST were supported in part by US National Science Foundation grant DMS2030562.

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover Press, 1965.
[2] S.G. Akl, A new algorithm for generating derangements, BIT 20 (1980) 2–7.
[3] R. Arratia, On the amount of dependence in the prime factorization of a uniform random integer, Bolyai Society Mathematical Studies 10 (2002) 29–91.
[4] R. Arratia, A. Barbour, W. Ewens, S. Tavaré, Simulating the component counts of combinatorial structures, Theor. Popul. Biol. 122 (2018) 5–11.
[5] R. Arratia, A. Barbour, S. Tavaré, Poisson process approximations for the Ewens sampling formula, Ann. Appl. Probab. 2 (1992) 519–535.
[6] R. Arratia, A. Barbour, S. Tavaré, Logarithmic Combinatorial Structures: A Probabilistic Approach, in: EMS Monographs in Mathematics, European Mathematical Society, 2003.
[7] R. Arratia, S. Tavaré, Limit theorems for combinatorial structures via discrete process approximations, Rand. Struct. Algorithms 3 (1992) 321–345.
[8] R. Arratia, S. Tavaré, Random Partitions, Permutations, and Primes, in: Lecture notes, Mathematics Department, University of Southern California, 1996, unpublished.
[9] H. Crane, The ubiquitous Ewens sampling formula, Statist. Sci. 31 (2016) 1–19.
[10] W. Ewens, The sampling theory of selectively neutral alleles, Theor. Popul. Biol. 3 (1972) 87–112.
[11] W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, third ed., Wiley, New York, 1968.
[12] V. Goncharov, Some facts from combinatorics, Izv. Akad. Nauk. SSSR, Ser. Mat. 8 (1944) 3–48, See also: On the field of combinatory analysis. Translations of the American Mathematical Society 19, 1-46.
[13] N. Johnson, S. Kotz, N. Balakrishnan, Discrete Multivariate Distributions, Wiley, New York, 1997.
[14] C. Martinez, A. Panholzer, H. Prodinger, Generating random derangements, in: ANALCO '08: Proceedings of the Meeting on Analytic Algorithmics and Combinatorics, SIAM, 2008, pp. 234–240.
[15] J. Mendonça, Efficient generation of random derangements with the expected distribution of cycle lengths, Comput. Appl. Math. 39 (2020) 244.
[16] D. Merlini, R. Sprugnoli, M. Verri, An analysis of a simple algorithm for random derangements, in: Proceedings of the 10th Italian Conference on Theoretical Computer Science, World Scientific, 2007, pp. 139–150.
[17] K. Mikawa, K. Tanaka, Linear-time encoding of uniform random derangements encoded in cycle notation, Discrete Appl. Math. 217 (2017) 722–728.
[18] OEIS Foundation Inc, The on-Line Encyclopedia of Integer Sequences, 2020, http://oeis.org.
[19] L. Shepp, S. Lloyd, Ordered cycle lengths in a random permutation, Trans. Amer. Math. Soc. 121 (1966) 340–357.
[20] S. Tavaré, The magical Ewens sampling formula, Bull. Lond. Math. Soc. 00 (2021) 000.