REGENERATIVE RANDOM PERMUTATIONS OF INTEGERS

JIM PITMAN AND WENPIN TANG

Abstract. Motivated by recent studies of large Mallows(q) permutations, we propose a class of random permutations of $\mathbb{N}_+$ and of $\mathbb{Z}$, called *regenerative permutations*. Many previous results of the limiting Mallows(q) permutations are recovered and extended. Three special examples: blocked permutations, p-shifted permutations and p-biased permutations are studied.

Key words : Bernoulli sieve, cycle structure, indecomposable permutations, Mallows permutations, regenerative processes, renewal processes, size biasing.

AMS 2010 Mathematics Subject Classification: 05A05, 60C05, 60K05.

Contents

1. Introduction and main results

Random permutations have been extensively studied in combinatorics and probability theory. They have a variety of applications including:

- statistical theory, e.g. Fisher-Pitman permutation test [36, 80], ranked data analysis [23, 24];
- population genetics, e.g. Ewens' sampling formula [32] for the distribution of allele frequencies in a population with neutral selection;
- quantum physics, e.g. spatial random permutations [100, 12] arising from the Feynman representation of interacting Bose gas;
- computer science, e.g. data streaming algorithms [78, 53], interleaver designs for channel coding [27, 11].

Date: May 1, 2017.

Interesting mathematical problems are (i). understanding the asymptotic behavior of large random permutations, and (ii). constructing countably infinite random permutations. Over the past few decades, considerable progress has been made in these two directions:

- (i). Shepp and Lloyd [93], Vershik and Shmidt [101, 102] studied the distribution of cycles in a large uniform random permutation. The result was extended by Diaconis, McGrath and Pitman [25], Lalley [71] for a class of large non-uniform permutations. Hammersley [52] first considered the longest increasing subsequences in a large uniform random permutation. The law of large numbers was proved by Logan and Shepp [74], Kerov and Vershik [64] via representation theory, and by Aldous and Diaconis [3], Seppäläinen [91] using probabilistic arguments. The long-standing conjectured central limit theorem was solved by Baik, Deift and Johansson [8]. Recently, limit theorems for large Mallows permutations have been considered by Mueller and Starr [77], Bhatnagar and Peled [13], Basu and Bhatnagar [9], Gladkich and Peled [38].
- (ii). Pitman [83, 84] provided a sequential construction of random permutations of positive integers according to the cycle structure. This is known as the *Chinese restaurant process*, which induces partially exchangeable random partitions of positive integers. Kerov, Olshanski and Vershik [66, 65] introduced the same object under the name of *virtual permutations* via the projective limit. A description of the Chinese restaurant process in terms of records was given by Kerov [62], Kerov and Tsilevich [63]. See also Pitman [81]. Various random permutations of countably infinite sets have been devised by Gnedin and Olshanski [45, 46, 47], Gnedin [39], Gnedin and Gorin [40, 41] in a sequential way, and by Fichtner [35], Betz and Ueltschi [12], Biskup and Richthammer [14] in a Gibbsian way.

The inspiration for this article is a series of recent studies of random permutations of countably infinite sets by Gnedin and Olshanski [46, 47], Basu and Bhatnagar [9], Gladkich and Peled [38]. Typically, these models are obtained as limits in distribution, as $n \to \infty$, of some sequence of random permutations $\Pi^{[n]}$, with some given distributions Q_n on the set $\mathfrak{S}_n$ of permutations of the finite set $[n] := \{1, \dots, n\}$. The distribution of a limiting injection $\Pi : \mathbb{N}_+ \to \mathbb{N}_+$ is then defined by

$$\mathbb{P}(\Pi_i = n_i, 1 \le i \le k) := \lim_{n\to\infty} \mathbb{P}(\Pi_i^{[n]} = n_i, 1 \le i \le k), \tag{1.1}$$

for every sequence of k values $n_i \in \mathbb{N}_+ := \{1, 2, \dots\}$, provided these limits exist and sum to 1 over all choices of $(n_i, 1 \le i \le k) \in \mathbb{N}_+^k$. It is easy to see that for $Q_n = U_n$ the uniform distribution on $\mathfrak{S}_n$, the limits in (1.1) are identically equal to 0, so this program fails to produce a limiting permutation of $\mathbb{N}_+$. However, it was shown by Gnedin and Olshanski [46, Proposition A.1] that for every $0 < q < 1$ this program is successful for $Q_n = M_{n,q}$, the *Mallows*(q) *distribution on* $\mathfrak{S}_n$ [75], which assigns each permutation π of $[n]$ probability

$$\mathbb{P}(\Pi^{[n]} = \pi) = M_{n,q}(\pi) := Z_{n,q}^{-1}\, q^{\mathrm{inv}(\pi)} \qquad \text{for } \pi \in \mathfrak{S}_n, \tag{1.2}$$

where $\mathrm{inv}(\pi) := \{(i,j) : 1 \le i < j \le n, \pi(i) > \pi(j)\}$ is the *number of inversions* of π, and the normalization constant $Z_{n,q}$ is well known to be the *q-factorial function*

$$Z_{n,q} = \prod_{j=1}^{n} \sum_{i=1}^{j} q^{i-1} = (1-q)^{-n} \prod_{j=1}^{n} (1-q^j) \qquad \text{for } 0 < q < 1. \tag{1.3}$$

See Diaconis and Ram [26, Section 2.e] for algebraic properties of Mallows(q) distributions, and additional references. Note that it is possible to define the projective limit for both $Q_n = U_n$ and $Q_n = M_{n,q}$ in the space of virtual permutations.

- For $Q_n = U_n$, the consistency of the family $(U_n;\ n \ge 1)$ with respect to the projection is closely related to the *Fisher-Yates-Durstenfeld-Knuth shuffle* [69, Section 3.4.2]. The projective limit is the Chinese restaurant process with $\theta = 1$.
- For $Q_n = M_{n,q}$, the fact that $(M_{n,q};\ n \ge 1)$ are consistent relative to the projection is a consequence of the *Lehmer code* [70, Section 5.1.1]. Moreover, Gnedin and Olshanski [46, Proposition A.6] proved that the projective limit coincides with the limit in distribution (1.1).

Gnedin and Olshanski [46] gave a number of other characterizations of the limiting distribution of Π so obtained for each $0 < q < 1$, including the fact that Π is a permutation of $\mathbb{N}_+$ with probability one. They continued in [47] to show that there exists a two-sided random permutation Π^* of $\mathbb{Z}$, which is a similar limit in distribution of Mallows(q) distributions of $[n]$, shifted to act on intervals of integers $[1 - a_n, n - a_n]$, for any sequence of integers a_n with both $a_n \to \infty$ and $n - a_n \to \infty$ as $n \to \infty$. They also showed that for each $0 < q < 1$ the process Π^* is *stationary*, meaning that the *process of displacements* $(D^*_z := \Pi^*_z - z;\ z \in \mathbb{Z})$ is a stationary process:

$$(D^*_z;\ z \in \mathbb{Z}) \overset{(d)}{=} (D^*_{a+z};\ z \in \mathbb{Z}) \qquad \text{for } a \in \mathbb{Z}. \tag{1.4}$$

These results were further extended by Basu and Bhatnagar [9], Gladkich and Peled [38], who established a number of properties of the limiting Mallows(q) permutations of $\mathbb{N}_+$ and of $\mathbb{Z}$, as well as providing many finer asymptotic results regarding the behavior of various functionals of Mallows(q) permutations of $[n]$, including cycle counts, and longest increasing subsequences, in various finer limit regimes with q approaching either 0 or 1 as $n \to \infty$. The analysis of limiting Mallows(q) permutations Π of $\mathbb{N}_+$ by these authors relies on a key regenerative property of these permutations, which is generalized in this paper to provide companion results for a much larger class of random permutations of $\mathbb{N}_+$ and of $\mathbb{Z}$.

For a permutation Π of a countably infinite set I, however it may be constructed, there is the basic question:

- *is every orbit of* Π *a finite cycle?* (1.5)

If so, say Π *has only finite cycles.* Note that the random permutation Π of $\mathbb{N}_+$ constructed by the Chinese restaurant process has infinite cycles with probability one. For $I = \mathbb{N}_+$ or $\mathbb{Z}$, one way to show Π has only finite cycles, and to gain some control on the distribution of cycle lengths, is to establish the stronger property that constructed, there is the basic question:

- *every block of* Π *has finite length.* (1.6)

Here we need some vocabulary. Let $I \subseteq \mathbb{Z}$ be an interval of integers, and $\Pi : I \to I$ be a permutation of I. Call $n \in I$ a *splitting time* of Π, or say that Π *splits at* n, if Π maps $(-\infty, n]$ to itself, or equivalently, Π maps $I \cap [n+1, \infty)$ to itself. The set of splitting times of Π, called the *connectivity set* by Stanley [96], is the collection of finite right endpoints of some finite or infinite family of *components of* Π, say $\{I_j\}$. These components I_j form a partition of I, which is coarser than the partition by cycles of Π. For example, the permutation $\pi = (1)(2,4)(3) \in \mathfrak{S}_4$ induces the partition by components $[1][2,3,4]$. So Π acts on each of its components I_j as an *indecomposable permutation* of I_j, meaning that Π does not act as a permutation on any proper subinterval of I_j. For any *block* J of Π with $\#J = n$, that is a component of Π, or a union of adjacent components of Π, the *reduced block of* Π *on* J is the permutation of $[n]$ defined via conjugation of Π by the shift from J to $[n]$.

For any permutation Π of $\mathbb{N}_+$, there are two ways to express the event $\{\Pi \text{ splits at } n\}$ as an intersection of n events:

$$\{\Pi \text{ splits at } n\} = \bigcap_{i=1}^{n} \{\Pi_i \le n\} = \bigcap_{i=1}^{n} \{\Pi_i^{-1} \le n\}.$$

An alternative way of writing this event is:

$$\{\Pi \text{ splits at } n\} = \bigcap_{i=1}^{n} \left\{ \Pi_i^{-1} < \min_{j>n} \Pi_j^{-1} \right\}.$$

For if Π splits at n, then $\Pi_i^{-1} < n+1 = \min_{j>n} \Pi_j^{-1}$ for every $1 \le i \le n$. Conversely, if $\min_{j>n} \Pi_j^{-1} = m+1$ say, and $\Pi_i^{-1} < m+1$ for every $1 \le i \le n$, then the image of $[n]$ via Π^{-1} is equal to $[m]$, so $m = n$ and $\Pi_i^{-1} \le n$ for every $1 \le i \le n$. Let

$$A_{n,i} := \left\{ \min_{j>n} \Pi_j^{-1} < \Pi_i^{-1} \right\}, \tag{1.7}$$

be the complement of the i^{th} event in the above intersection. Then by the principle of inclusion-exclusion

$$\mathbb{P}(\Pi \text{ splits at } n) = 1 + \sum_{j=1}^{n} (-1)^j \Sigma_{n,j}, \tag{1.8}$$

where

$$\Sigma_{n,j} := \sum_{1 \le i_1 < \cdots < i_j \le n} \mathbb{P}\left(\bigcap_{k=1}^{j} A_{n,i_k} \right). \tag{1.9}$$

So there are the Bonferroni bounds

$$\mathbb{P}(\Pi \text{ splits at } n) \ge 1 - \Sigma_{n,1}, \quad \mathbb{P}(\Pi \text{ splits at } n) \le 1 - \Sigma_{n,1} + \Sigma_{n,2},$$

and so on. Moreover, each of the intersections of the $A_{n,i}$ is an event of the form

$$F_{B,C} := \left\{ \min_{j \in B} \Pi_j^{-1} < \min_{h \in C} \Pi_h^{-1} \right\},$$

for instance $A_{n,i} A_{n,j} = F_{B,C}$ for $F = \{n, n+1, \ldots\}$ and $C = \{i, j\}$.

An approach to the problem of whether Π has almost surely finite block lengths for a number of interesting models, including the limiting Mallows(q) model, is provided by the following structure. Let

$$\mathbb{N}_+ := \{1, 2, \ldots\} \quad \text{and} \quad \mathbb{N}_0 := \{0, 1, 2, \ldots\}.$$

Call $(T_n;\ n \ge 0)$ a *delayed renewal process* if

$$T_n := T_0 + Y_1 + \cdots + Y_n,$$

with $T_0 \in \mathbb{N}_0, Y_1, Y_2, \ldots \in \mathbb{N}_+ \cup \{\infty\}$ independent, and the Y_i identically distributed, allowing also the transient case with $\mathbb{P}(Y_1 < \infty) < 1$. When $T_0 := 0$, call $(T_n;\ n \ge 0)$ a *renewal process with zero delay.* The definition below is tailored to the general theory of regenerative processes presented by Asmussen [6, Chapter VI].

Definition 1.1. If a permutation Π of $\mathbb{N}_+$ splits at n, let Π^n be the *residual permutation* of $\mathbb{N}_+$ defined by conjugating the action of Π on $\mathbb{N}_+ \setminus [n]$ by a shift back to $\mathbb{N}_+$:

$$\Pi_i^n := \Pi_{n+i} - n \qquad \text{for } i \in \mathbb{N}_+.$$

(1). Call a random permutation of Π of $\mathbb{N}_+$ *regenerative with respect to the possibly delayed renewal process* $T_0, T_1, T_2, \ldots$ with renewal indicators $R_n := \sum_{k=0}^{\infty} 1(T_k = n)$ if every T_i is a splitting time of Π, and for each $n > 0$ such that the renewal probability $\mathbb{P}(R_n = 1)$ is strictly positive, conditionally given a renewal at n,
   (i). there is the equality in distribution
   $$(\Pi^n, R_{n+1}, R_{n+2}, \ldots) \stackrel{(d)}{=} (\Pi^0, R_1^0, R_2^0, \ldots)$$
   between the joint distribution of Π^n with the residual renewal indicators $(R_{n+1}, R_{n+2}, \ldots)$, and the joint distribution of some random permutation Π^0 of $\mathbb{N}_+$ with renewal indicators $(R_1^0, R_2^0, \ldots)$ with zero delay;
   (ii). the initial segment $(R_0, R_1, \ldots R_n)$ of the delayed renewal process is independent of $(\Pi^n, R_{n+1}, R_{n+2}, \ldots)$.
(2). Call a random permutation of Π of $\mathbb{N}_+$ *regenerative* if Π is regenerative with respect to some renewal process $T_0, T_1, T_2, \ldots$
(3). Call a random permutation of Π of $\mathbb{N}_+$ *strictly regenerative* if Π is regenerative with respect to its own splitting times.

The formulation of Definition 1.1 was motivated by its application to three particular models of random permutations of $\mathbb{N}_+$, introduced by the next three definitions. Each of these models is parameterized by a discrete probability distribution on $\mathbb{N}_+$, say $p = (p_1, p_2, \ldots)$. These models are close in spirit to the similarly parameterized models of *p-mappings* and *p-trees* studied in [5, 4]. See also [48, 42, 50] for closely related ideas of regeneration in random combinatorial structures.

General properties of a regenerative random permutation Π of $\mathbb{N}_+$ with zero delay can be read from the standard theory of regenerative processes [34, Chapter XIII]. Let $u_0 := 0$, and

$$u_n := \mathbb{P}(\Pi \text{ regenerates at } n), \tag{1.10}$$
$$f_n := \mathbb{P}(\Pi \text{ regenerates for the first time at } n). \tag{1.11}$$

Each of these sequences determines the other by the recursion

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_n u_0 \quad \text{for all } n > 0, \tag{1.12}$$

which may be expressed in terms of the generating functions $U(z) := \sum_{n=0}^{\infty} u_n z^n$ and $F(z) := \sum_{n=1}^{\infty} f_n z^n$ as

$$U(z) = (1 - F(z))^{-1}. \tag{1.13}$$

According to the discrete renewal theorem, either

(i). (*transient case*) $\sum_{n=1}^{\infty} u_n < \infty$, when $\mathbb{P}(Y_1 < \infty) < 1$, and Π has only finitely many regenerations with probability one, or
(ii). (*recurrent case*) $\sum_{n=1}^{\infty} u_n = \infty$, when $\mathbb{P}(Y_1 < \infty) = 1$, and with probability one Π has infinitely many regenerations, hence only finite components, and only finite cycles.

Here is a simple way of constructing recurrent regenerative random permutations of $\mathbb{N}_+$:

Definition 1.2. For a probability distribution $p := (p_1, p_2, \ldots)$ on $\mathbb{N}_+$, and Q_n for each $n \in \mathbb{N}_+$ a probability distribution on $\mathfrak{S}_n$, call a random permutation Π of $\mathbb{N}_+$ *recurrent regenerative with block length distribution p and blocks governed by* $(Q_n;\ n \geq 1)$, if Π is a concatenation of an infinite sequence $\texttt{Block}_i$, $i \geq 0$ such that

- (i). the lengths Y_i of $\texttt{Block}_i$, $i \geq 1$ are independent and identically distributed (i.i.d.) with common distribution p, and are independent of the length T_0 of $\texttt{Block}_0$ which is finite almost surely;
- (ii). conditionally given the block lengths, say $Y_i = n_i$ for $i = 1, 2, \ldots$, the reduced blocks of Π are independent random permutations of $[n_i]$ with distributions Q_{n_i}.

The main focus here is the *positive recurrent case*, with mean block length $\mu := \mathbb{E}(Y_1) < \infty$, and an aperiodic distribution of Y_1, which according to the discrete renewal theorem [34, Chapter XIII, Theorem 3] makes

$$\lim_{n \to \infty} u_n = 1/\mu > 0. \tag{1.14}$$

Then numerous asymptotic properties of the recurrent regenerative permutation Π with this distribution of block lengths can be read from standard results in renewal theory, as discussed further in Section 3. In particular, starting from any positive recurrent random permutation Π of $\mathbb{N}_+$, renewal theory gives an explicit construction of a stationary, two-sided version Π^* of Π, acting as a random permutation of $\mathbb{Z}$, along with ergodic theorems indicating the existence of limiting frequencies for various counts of cycles and components, for both the one-sided and two-sided versions. This greatly simplifies the construction of stationary versions of the limiting Mallows(q) permutations in [47, 38].

Observe that for every recurrent, strictly regenerative permutation of $\mathbb{N}_+$, the support of Q_n is necessarily contained in the set $\mathfrak{S}_n^\dagger$ of indecomposable permutations of $[n]$. As will be seen in Section 4, it is not easy to describe tractable models for distributions on $\mathfrak{S}_n^\dagger$ besides the uniform distribution, and even the uniform distribution on $\mathfrak{S}_n^\dagger$ has a nasty denominator for which there is no very simple formula. These difficulties motivate the study of other constructions of random permutations of $\mathbb{N}_+$, such as the following:

Definition 1.3. For p a probability distribution on $\mathbb{N}_+$ with $p_1 > 0$, call a random permutation Π of $\mathbb{N}_+$ a *p-shifted permutation* of $\mathbb{N}_+$, if Π has the distribution defined by the following construction from an i.i.d. sample $(X_j;\ j \geq 1)$ from p. Inductively, let

- $\Pi_1 := X_1$,
- for $i \geq 2$, let $\Pi_i := \psi(X_i)$ where ψ is the increasing bijection from $\mathbb{N}_+$ to $\mathbb{N}_+ \setminus \{\Pi_1, \Pi_2, \cdots, \Pi_{i-1}\}$.

For example, if $X_1 = 2$, $X_2 = 1$, $X_3 = 2$, $X_4 = 3$, $X_5 = 4$, $X_6 = 1 \ldots$, then the associated permutation is $(2, 1, 4, 6, 8, 3, \ldots)$.

Gnedin and Olshanski [46] introduced this construction of p-shifted permutations of $\mathbb{N}_+$ for p the geometric$(1-q)$ distribution. They proved that the limiting Mallows(q) permutations of $[n]$ is the geometric$(1-q)$-shifted permutation of $\mathbb{N}_+$. The regenerative feature of geometric$(1-q)$-shifted permutations was pointed out and exploited in [9, 38]. This regenerative feature is in fact a property of p-shifted permutations of $\mathbb{N}_+$ for any p with $p_1 > 0$. This observation allows a number of previous results for limiting Mallows(q) permutations to be extended as follows.

Proposition 1.4. *For each fixed probability distribution p on $\mathbb{N}_+$ with $p_1 > 0$, and Π a p-shifted random permutation of $\mathbb{N}_+$:*

(i). The joint distribution of the random injection $(\Pi_1, \ldots, \Pi_n) : [n] \to \mathbb{N}_+$ is given by the formula

$$\mathbb{P}(\Pi_i = \pi_i, 1 \le i \le n) = \prod_{j=1}^{n} p\left(\pi_j - \sum_{1 \le i < j} 1(\pi_i < \pi_j)\right), \tag{1.15}$$

for every fixed injection $(\pi_i, 1 \le i \le n) : [n] \to \mathbb{N}_+$, and $p(k) := p_k$.

(ii). The probability that Π maps $[n]$ to $[n]$ is

$$u_n := \mathbb{P}([n] \text{ is a block of } \Pi) = \prod_{j=1}^{n} \sum_{i=1}^{j} p_i. \tag{1.16}$$

(iii). The random permutation Π is strictly regenerative, with regeneration at every n such that $[n]$ is a block of $\mathbb{N}_+$, and the renewal sequence $(u_n;\ n \ge 1)$ as above.

(iv). The distribution of component lengths $f_n := \mathbb{P}(Y_1 = n)$ where Y_1 is the length of the first component of Π is given by the probability generating function

$$\mathbb{E}z^{Y_1} = \sum_{n=1}^{\infty} f_n z^n = 1 - \frac{1}{U(z)} \quad \text{where } U(z) := 1 + \sum_{n=1}^{\infty} u_n z^n. \tag{1.17}$$

(v). If $\mathbb{E}X_1 = m := \sum_i i p_i < \infty$, then $\mu := \mathbb{E}(Y_1) < \infty$, so Π is positive recurrent, with limiting renewal probability

$$u_\infty := \lim_{n \to \infty} u_n = \mu^{-1} = \prod_{j=1}^{\infty} (1 - \mathbb{P}(X > j)), \tag{1.18}$$

*for X with distribution p. Then Π has cycle counts with limit frequencies detailed later in (3.2), and there is a stationary version Π^* of Π acting on $\mathbb{Z}$, call it a p-*shifted random permutation of $\mathbb{Z}$.

(vi). If $m = \infty$ then Π is either null recurrent or transient, according to whether $U(1)$ is infinite or finite, and there is no stationary version of Π acting on $\mathbb{Z}$.

Even for the extensively studied limiting Mallows(q) model, Proposition 1.4 contains some new formulas and characterizations of the distribution, which are discussed in Section 5. An interesting byproduct of this proposition for a general p-shifted permutation is the following classical result of Kaluza [59]:

Corollary 1.5. [59] *Every sequence $(u_n;\ n \ge 0)$ with*

$$0 < u_n \le u_0 = 1 \quad \text{and} \quad u_n^2 \le u_{n-1} u_{n+1} \text{ for all } n \ge 1, \tag{1.19}$$

is a renewal sequence. The sequence $(u_n;\ n \ge 0)$ satisfying (1.19) is called a Kaluza sequence. The renewal process associated with a Kaluza sequence is generated by the random sequence of times n at which $[n]$ is a block of Π, for Π a p-shifted permutation of $\mathbb{N}_+$, with

$$p_1 := u_1 \quad \text{and} \quad p_n := \frac{u_n}{u_{n-1}} - \frac{u_{n-1}}{u_{n-2}} \text{ for } n \ge 2,$$

and the convention that if $p_\infty := 1 - \sum_{i=1}^{\infty} p_i > 0$, and $X_1, X_2, \ldots$ is the sequence of independent choices from this distribution on $\{1, 2, \ldots, \infty\}$ used to drive the construction of Π, then *the construction is terminated by assigning some arbitrarily distributed infinite component on $[n+1,\infty)$ following the last splitting time n such that $X_1+\cdots+X_n<\infty$, for instance by a shifting to $[n+1,\infty)$ the deterministic permutation of $\mathbb{N}_+$ with no finite components*

$$\cdots 6 \to 4 \to 2 \to 1 \to 3 \to 5 \to \cdots$$

See also [61, 55, 68, 92, 73, 37] for other derivations and interpretations of Kaluza's result, all of which now aquire some expression in terms of p-shifted permutations.

Some further instances of regenerative permutations are provided by the following close relative of the p-shifted permutation:

Definition 1.6. For p with $p_i>0$ for every i, call a random permutation Π of $\mathbb{N}_+$ a *p-biased permutation of* $\mathbb{N}_+$ if the random sequence $(p_{\Pi_1},p_{\Pi_2},\ldots)$ is what is commonly called a *sized biased random permutation* of $(p_1,p_2,\ldots)$. That is to say, $(\Pi_1,\Pi_2,\ldots)$ is the sequence of distinct values, in order of appearance, of a random sample of positive integers $(X_1,X_2,\ldots)$, which are independent and identically distributed (i.i.d.) with distribution $(p_1,p_2,\ldots)$. Inductively, let

- $\Pi_1:=X_1$, and $J_1:=1$,
- for $i\geq 2$, let $\Pi_i:=X_{J_i}$, where J_i is the least $j>J_{i-1}$ such that
$$X_j\in\mathbb{N}_+\setminus\{X_{\Pi_1},X_{\Pi_2},\cdots,X_{\Pi_{i-1}}\}.$$

See [90, 54, 28, 79, 87] for various studies of this model of size-biased permutation, with emphasis on the annealed model, where p is determined by a random discrete distribution $P:=(P_1,P_2,\ldots)$, and given $P=p$, the X_j are i.i.d. with distribution p. In particular, the joint distribution of the random injection $(\Pi_1,\ldots\Pi_n):[n]\to\mathbb{N}_+$ is

$$\mathbb{P}(\Pi_i=\pi_i,\,1\leq i\leq n)=\mathbb{E}\left(P_{\pi_1}\prod_{i=2}^{n}\frac{P_{\pi_i}}{1-\sum_{j=1}^{i-1}P_{\pi_j}}\right). \tag{1.20}$$

for every fixed injection $(\pi_i,\,1\leq i\leq n):[n]\to\mathbb{N}_+$. A tractable model of this kind, known as a *residual allocation model* (*RAM*), has the stick-breaking representation:

$$P_i:=(1-W_1)\cdots(1-W_{i-1})W_i, \tag{1.21}$$

with $0<W_i<1$ and the W_i's are independent and identically distributed. This model is of special interest for Bayesian non-parametric inference and machine learning [15, 16]. In those contexts, the distribution of P represents a prior distribution on the underlying probability model p, which may be updated in response to observations such as the values in the sample $(X_1,\ldots,X_n)$, or values of $(\Pi_1,\ldots,\Pi_n)$. A model of particular interest arises when each W_i has the beta$(1,\theta)$ density $\theta(1-w)^{\theta-1}$ at $w\in(0,1)$ for some $0<\theta<\infty$. This distribution of $(P_1,P_2,\ldots)$ is known as the GEM(θ) distribution, after Griffiths, Engen and McCloskey who discovered the remarkable properties of this model, including McCloskey's result that the GEM(θ) model is the only RAM that is is invariant under P-biased permutation, meaning that there is the equality in distribution

$$(P_{\Pi_1},P_{\Pi_2},\ldots)\overset{(d)}{=}(P_1,P_2,\ldots)\quad\text{for }\Pi\text{ a }P\text{-biased permutation of }\mathbb{N}_+. \tag{1.22}$$

The following result reveals the regeneration of sized-biased random permutations of $\mathbb{N}_+$.

Proposition 1.7. *For every residual allocation model* (1.21) *for a random discrete distribution P with i.i.d. residual factors W_i, and Π a P-biased random permutation of $\mathbb{N}_+$:*

(i). The random permutation Π is strictly regenerative, with regeneration at every n such that $[n]$ is a block of $\mathbb{N}_+$, and the renewal sequence $(u_n;\ n \geq 1)$ defined by

$$u_n := \mathbb{P}([n] \text{ is a block of } \Pi) = \int_0^\infty e^{-x} \mathbb{E} \prod_{i=1}^n \left(1 - \exp\left(-\frac{xW_i}{T_i}\right)\right) dx, \tag{1.23}$$

where $T_i := (1 - W_1) \cdots (1 - W_i)$. Then Π is positive recurrent if

$$\sum_{i=2}^\infty \mathbb{E} \exp\left(-\frac{xW_i}{T_i}\right) < \infty \quad \text{for some } x > 0. \tag{1.24}$$

(ii). If each W_i is the constant $1 - q$ for some $0 < q < 1$, so P is the geometric$(1 - q)$ distribution on $\mathbb{N}_+$, then Π is positive recurrent. Hence Π has all blocks finite and limiting frequencies of cycle counts as in (3.2), and there is a stationary version Π^* of Π acting on $\mathbb{Z}$, called a p-biased random permutation of $\mathbb{Z}$.

(iii). If the W_i are i.i.d. beta$(1, \theta)$ for some $\theta > 0$, so P has the GEM(θ) distribution, then Π is positive recurrent, with the same further implications.

Propositions 1.4 and 1.7 expose a close affinity between p-shifted and p-biased permutations of $\mathbb{N}_+$, at least for some choices of p, which does not seem to have been previously recognized. For instance, if p is such that p_1 is close to 1, and subsequent terms decrease rapidly to 0, then it is to be expected in either of these models that Π should be close in some sense to the identity permutation on $\mathbb{N}_+$. This intuition is confirmed by the explicit formulas described in Section 6 both for the one parameter family of geometric$(1 - q)$ distributions as $q \downarrow 0$, and for the GEM(θ) family as $\theta \downarrow 0$. This behavior is in sharp contrast to the case if Π is a uniformly distributed permutation of $[n]$, where it is well known that the expected number of fixed points of Π is 1, no matter how large n may be. See also Gladkich and Peled [38] for many finer asymptotic results for the Mallows(q) model of permutations of $[n]$, as both $n \to \infty$ and $q \downarrow 0$.

With further analysis, we derive explicit formulas for u_∞ of the GEM(θ)-biased permutations in Section 7. But there does not seem to be any simple formula for u_∞ of a P-biased permutation with P a general RAM, and the condition (1.24) for positive recurrence is not easy to check. Nevertheless, we give a simple sufficient condition for a P-biased permutation of $\mathbb{N}_+$ with P governed by a RAM to be positive recurrent.

Proposition 1.8. *Let Π be a P-biased permutation of $\mathbb{N}_+$ for P a RAM with i.i.d. residual factors $W_i \overset{(d)}{=} W$. If the distribution of $-\log(1 - W) < \infty$ is non-lattice, meaning that the distribution of $1 - W$ is not concentrated on a geometric progression, and*

$$\mathbb{E}[-\log W] < \infty \quad \text{and} \quad \mathbb{E}[-\log(1 - W)] < \infty, \tag{1.25}$$

then Π is positive recurrent regenerative permutation.

Organization of the paper: The rest of this paper is organized as follows.

- Section 2 sets the stage by recalling some basic properties of indecomposable permutations of a finite interval of integers, which are the basic building blocks of regenerative permutations.

- Section 3 indicates how the construction of a stationary random permutation of $\mathbb{Z}$ along with some limit theorems is a straightforward application of the well established theory of regenerative random processes.
- Section 4 provides an example of the regenerative permutation of $\mathbb{N}_+$, with uniform block distribution. Some explicit formulas are given.
- Section 5 sketches a proof of Proposition 1.4 for p-shifted permutations, following the template provided by [9] in the particular case of the limiting Mallows(q) models.
- Section 6 gives a proof of Proposition 1.7 for P-biased permutations. This is somewhat trickier, and the results are less explicit than in the p-shifted case.
- Section 7 provides further analysis of regenerative P-biased permutations. There Proposition 1.8 is proved. We also show that the limiting renewal probability of the GEM(1)-biased permutation is 1/3.

Acknowledgement: We thank David Aldous, Persi Diaconis, Marek Biskup and Sasha Gnedin for various pointers to the literature. Thanks to Jean-Jil Duchamps for an insightful first proof of our earlier conjecture that $u_\infty = 1/3$.

2. Indecomposable permutations

This section provides references to some basic combinatorial theory of indecomposable permutations of $[n]$ which may arise as the reduced permutations of Π on its components of finite length. For $1 \le k \le n$, let $(n,k)^\dagger$ be the number of permutations of $[n]$ with exactly k components. In particular, $(n,1)^\dagger := \#\mathfrak{S}_n^\dagger$ is the number of indecomposable permutations of $[n]$, as the sequence A003319 of OEIS. As shown by Lentin [72] and Comtet [18], the counts $((n,1)^\dagger;\ n \ge 1)$, starting from $(1,1)^\dagger = 1$, are determined by the recurrence

$$n! = \sum_{k=1}^{n} (k,1)^\dagger (n-k)!, \tag{2.1}$$

which enumerates permutations of $[n]$ according to the size k of their first component. Introducing the formal power series which is the generating function of the sequence $(n!;\ n \ge 0)$

$$G(z) := \sum_{n=0}^{\infty} n!\, z^n,$$

the recursion (2.1) gives the generating function of the sequence $((n,1)^\dagger;\ n \ge 1)$, as

$$\sum_{n=1}^{\infty} (n,1)^\dagger z^n = 1 - \frac{1}{G(z)}, \tag{2.2}$$

which implies that

$$(n,1)^\dagger = n! \left(1 - \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right).$$

Furthermore, it is derived from (2.2) that

$$\sum_{n=k}^{\infty} (n,k)^\dagger z^n = \left(1 - \frac{1}{G(z)}\right)^k \quad \text{for } 1 \le k \le n. \tag{2.3}$$

The identity (2.3) determines the triangle of numbers $(n,k)^\dagger$ for $1 \le k \le n$, as displayed for $1 \le n \le 10$ in Comtet [19, Exercise VI.14]. See also [67, 20, 21, 22, 1, 7] for various results about indecomposable permutations.

Recall that for $I \subseteq \mathbb{Z}$ an interval of integers, and $\Pi : I \to I$ a permutation of I, we say Π splits at $n \in I$, if Π maps $I \cap (-\infty, n]$ to itself. As observed by Stam [95], the splitting times of a uniform random permutation Π of a finite interval of integers $I = [a,b]$ are *regenerative* in the sense that conditionally given that Π splits at some $n \in I$ with $a \le n < b$, the restrictions of Π to $[a,n]$ and to $[n+1,b]$ are independent uniform random permutations of these two subintervals of I. However, for a uniform random permutation Π of a finite interval, the components of Π turn out not to be very interesting. In fact, for a large finite interval of integers I, most permutations of I have only one component. Assuming for simplicity that $I = [n]$, let

$$V_n := \sum_{k=1}^{n} 1(\Pi \text{ splits at } k),$$

be the number of interval components of Π, a uniformly distributed random permutation of $[n]$. Then by an obvious enumeration

$$\mathbb{E}V_n := \sum_{k=1}^{n} \mathbb{P}(\Pi \text{ splits at } k) = \sum_{k=1}^{n} \frac{k!(n-k)!}{n!} = \Sigma_n - 1,$$

where

$$\Sigma_n := \sum_{k=0}^{n} \binom{n}{k}^{-1},$$

is the sum of reciprocals of binomial coefficients. Because the binomial coefficients are first increasing and then decreasing, there are the easy estimates

$$2 \le \Sigma_n \le 2 + \frac{4}{n} \quad \text{for } n \ge 1,$$

which imply by Markov's inequality that $\mathbb{P}(V_n = 1) \to 1$ as $n \to \infty$. The sum Σ_n, as the sequence A046825 of OEIS, has been studied in a number of articles [89, 97], with some other interpretations of the sum given in [94].

The following lemma records some basic properties of the decomposition of a uniform permutation Π of $[n]$.

Lemma 2.1. *Let Π be a uniformly distributed random permutation of $[n]$. Then:*

(i). The number K_n of components of Π has distribution

$$\mathbb{P}(K_n = k) = \frac{(n,k)^\dagger}{n!} \quad \textit{for } 1 \le k \le n, \tag{2.4}$$

with the counts $(n,k)^\dagger$ determined as above.

(ii). Conditionally given $K_n = k$, the random composition of n defined by the lengths $L_{n,1}, \dots, L_{n,k}$ of these components has the exchangeable joint distribution

$$\mathbb{P}(L_{n,1} = n_1, \dots, L_{n,k} = n_k \,|\, K_n = k) = \frac{1}{(n,k)^\dagger} \prod_{i=1}^{k} (n, n_i)^\dagger, \tag{2.5}$$

for all compositions $(n_1, \dots, n_k)$ of n with k parts, meaning $n_i \ge 1$ and $\sum_{i=1}^k n_i = n$.

(*iii*). *The unconditional distribution of the length* $L_{n,1}$ *of the first component of* Π *is given by*

$$\mathbb{P}(L_{n,1} = \ell) = \frac{(\ell, 1)^{\dagger}(n-\ell)!}{n!} \quad \text{for } 1 \leq \ell \leq n, \tag{2.6}$$

while the conditional distribution of $L_{n,1}$ *given that* $K_n = k$ *is given by*

$$\mathbb{P}(L_{n,1} = \ell \,|\, K_n = k) = \frac{(\ell, 1)^{\dagger}(n-\ell, k-1)^{\dagger}}{(n,k)^{\dagger}} \quad \text{for } 1 \leq \ell \leq n, \tag{2.7}$$

with the convention that $(0,0)^{\dagger} = 1$ *but otherwise* $(n,k)^{\dagger} = 0$ *unless* $1 \leq k \leq n$.

(*iv*). *The distribution of the length* L_n^* *of a size-biased random component of* Π*, such as the length of the component of* Π *containing* U_n*, where* U_n *is independent of* Π *with uniform distribution on* $[n]$*, is given by the formula*

$$\mathbb{P}(L_n^* = \ell) = \frac{\ell\,(\ell, 1)^{\dagger}}{n \cdot n!} \sum_{k=1}^{n} k\,(n-l, k-1)^{\dagger}, \tag{2.8}$$

with the same convention.

Proof. The first three parts are just probabilistic expressions of the the preceding combinatorial discussion. Then part (*iv*) follows from the definition of the size-biased pick, using

$$\mathbb{P}(L_n^* = \ell) = \sum_{k=1}^{n} \mathbb{P}(L_n^* = \ell \,|\, K_n = k)\mathbb{P}(K_n = k).$$

Given that $K_n = k$ let the lengths of these k components listed from left to right be $L_{n,1}, \ldots, L_{n,k}$,

$$\begin{aligned}
\mathbb{P}(L_n^* = \ell \,|\, K_n = k) &= \sum_{j=1}^{k} \mathbb{P}(\text{pick } L_{n,j} \text{ and } L_{n,j} = \ell \,|\, K_n = k) \\
&= k\mathbb{P}(\text{pick } L_{n,1} \text{ and } L_{n,1} = \ell \,|\, K_n = k) \\
&= k\frac{\ell}{n}\mathbb{P}(L_{n,1} = \ell \,|\, K_n = k),
\end{aligned}$$

where the second equality is obtained by exchangeability. Now part (iv) follows by plugging in the formulas in previous parts. $\square$

Table of $n \cdot n!\,\mathbb{P}(L_n^* = \ell)$ for $1 \leq \ell \leq n \leq 7$:

n								
1	1							
2	2	2						
3	5	4	9					
4	16	10	18	52				
5	64	32	45	104	355			
6	312	128	144	260	710	2766		
7	1812	624	576	832	1775	5532	24129	
	0	1	2	3	4	5	6	ℓ

3. Regenerative and stationary permutations

This section elaborates on the structure of a regenerative permutation of $\mathbb{N}_+$, and its stationary version Π^* acting on $\mathbb{Z}$. To provide some intuitive language for discussion of a permutation Π of $I = \mathbb{N}_+$ or of $I = \mathbb{Z}$, it is convenient to regard Π as describing a motion of balls labeled by I. Initially, for each $i \in I$, ball i lies in box i. After the action of Π,

- ball i from box i is moved to box Π_i;
- box j contains the ball initially in box Π_j^{-1}.

For $i \in I$ let $D_i := \Pi_i - i$, the displacement of ball initially in box i. It follows easily from Definition 1.1 that if Π is a regenerative permutation of $\mathbb{N}_+$, then the process $(D_n;\ n \geq 1)$ is a *regenerative process with embedded delayed renewal process* $(T_k;\ k \geq 0)$. This means that if $R_n := \sum_{k=0}^{\infty} 1(T_k = n)$ is the n^{th} renewal indicator variable, then for each n such that $\mathbb{P}(R_n = 1) > 0$, conditionally given the event $\{R_n = 1\}$,

(i). there is the equality of finite dimensional joint distributions

$$((D_{n+j}, R_{n+j});\ j \geq 1) \stackrel{(d)}{=} ((D_j^0, R_j^0);\ j \geq 1),$$

where the $D_j^0 := \Pi_j^0 - j$ are the displacements of the random permutation Π^0 of $\mathbb{N}_+$, with associated renewal indicators $R_1^0, R_2^0, \ldots$ with zero delay.

(ii). the bivariate process $((D_{n+j}, R_{n+j});\ j \geq 1)$ is independent of $(R_1, \ldots, R_n)$.

This paraphrases the discrete case of the general definition of a regenerative process proposed by Asmussen [6, Chapter VI], and leads to the following Lemma.

Lemma 3.1. [6, Chapter VI, Theorem 2.1]. *Let $(D_n;\ n \geq 1)$ be a regenerative process with embedded delayed renewal process, $(T_k;\ k \geq 0)$, in the sense indicated above. Assume that the renewal process is positive recurrent with finite mean recurrence time $\mu := \mathbb{E}(Y_1) < \infty$, where $Y_1 := T_1 - T_0$, and that the distribution of Y_1 is aperiodic. Then there is the convergence in total variation of distributions of infinite sequences*

$$(D_n, D_{n+1}, \ldots) \xrightarrow{t.v.} (D_0^*, D_1^*, \ldots),$$

where $(D_z^*;\ z \in \mathbb{Z})$ is a two-sided stationary process, whose law is uniquely determined by the block formula

$$\mathbb{E}g(D_z^*, D_{z+1}^*, \ldots) = \frac{1}{\mu}\mathbb{E}\left(\sum_{k=1}^{\infty} g(D_k, D_{k+1}, \ldots) 1(Y_1 \geq k)\right), \tag{3.1}$$

for all $z \in \mathbb{Z}$ and all non-negative product measurable functions g.

The existence of a stationary limiting Mallows(q) permutation of $\mathbb{Z}$ was established by Gnedin and Olshanski [46], along with various characterizations of its distribution. Their work is difficult to follow, because they did not exploit the regenerative properties of this distribution. Gladkich and Peled [38, Section 3] provides some further information about this model, including what they call a 'stitching' construction of the two-sided model from its blocks on $(-\infty, T_0]$, (T_0, T_1) and $[T_1 + 1, \infty)$. But their construction too is difficult to follow. In fact, the structure the two-sided Mallows permutation of $\mathbb{Z}$ is typical of the general structure of stationary regenerative processes. This structure is spelled out in the following theorem, which follows easily from Lemma 3.1.

Theorem 3.2. *Let Π be a positive recurrent regenerative random permutation of $\mathbb{N}_+$, with block length distribution p and family of block distributions Q_n on $\mathfrak{S}_n$, and $\mu := \sum_n np_n < \infty$.*

(*i*). *There exists a unique stationary regenerative random permutation Π^* of $\mathbb{Z}$, with associated stationary renewal process*

$$\{\cdots T_{-2} < T_{-1} < T_0 < T_1 < T_2 \cdots\} \subseteq \mathbb{Z},$$

with indexing defined by $T_{-1} < 0 \le T_0$, block lengths $Y_z := T_z - T_{z-1}$ for $z \in \mathbb{Z}$, and renewal indicators R_z^*, with $R_z^* = 1$ implying that Π splits at z, such that

$$\mathbb{P}(R_z^* = 1) = 1/\mu \quad \text{for } z \in \mathbb{Z},$$

and given the event $\{R_z^* = 1\}$, by letting $\Pi_i^{*,z} := \Pi_{z+i}^* - z$ for $i \in \mathbb{N}_+$,

$$(\Pi_1^{*,z}, \Pi_2^{*,z}, \ldots \mid R_z^* = 1) \stackrel{(d)}{=} (\Pi_1^n, \Pi_2^n, \ldots \mid R_n = 1) \quad \text{for } z \in \mathbb{Z},$$

for every n such that $\mathbb{P}(R_n = 1) > 0$, where $(R_n;\ n \ge 1)$ is the sequence of renewal indicators associated with the one-sided regenerative permutation Π.

(*ii*). *If Π^* is so defined, with block lengths $(Y_z;\ z \in \mathbb{Z})$, then the $(Y_z;\ z \in \mathbb{Z})$ are independent, with the $Y_z, z \neq 0$ all copies of Y_1 with distribution p, while Y_0 has the size-biased distribution*

$$\mathbb{P}(Y_0 = n) = np_n/\mu \quad \text{for } n \ge 1.$$

Conditionally given all the block lengths, the delay T_0 has uniform distribution on $\{0, 1, \ldots, Y_0 - 1\}$, and conditional on all the block lengths and on T_0, with given block lengths n_i say, the reduced permutation of Π^* on the block of n_i integers $(T_{i-1}, T_i]$ is distributed according to Q_{n_i}.

Conversely, if Π is regenerative, existence of such a stationary regenerative permutation of $\mathbb{Z}$ implies that Π is positive recurrent.

Also note that the law of the stationary regenerative random permutation Π^* is uniquely defined by the equality of joint distributions

$$(\Pi_1^*, \Pi_2^*, \ldots, T_0, T_1, T_2, \ldots \mid R_0^* = 1) \stackrel{(d)}{=} (\Pi_1^0, \Pi_2^0, \ldots, T_0, T_1, T_2, \ldots),$$

where on the left side the T_i are understood as the renewal times that are strictly positive for the stationary process Π^*, and on the right side the same notation is used for the renewal times of the regenerative random permutation Π^0 of $\mathbb{N}_+$ with zero delay, and on both sides $T_0 = 0$, the $Y_i := T_i - T_{i-1}$ for $i \ge 1$ are independent random lengths with distribution p, and conditionally given these block lengths are equal to n_i, the corresponding reduced permutations of $[n_i]$ are independent and distributed according to Q_{n_i}. So the random permutation Π^0 of $\mathbb{N}_+$ is a *Palm version* of the stationary permutation Π^* of $\mathbb{Z}$. See Thorisson [98, 99] for general background on stationary stochastic processes.

Let Π be a positive recurrent regenerative random permutation of $\mathbb{N}_+$, with block length distribution p. For $n \in \mathbb{N}_+$, let

- Cyc_n be the length of the cycle of Π containing n,
- Cmp_n be the length of the component of Π containing n,
- Blk_n be the length of the block of Π containing n.

Clearly, $1 \le \mathrm{Cyc}_n \le \mathrm{Cmp}_n \le \mathrm{Blk}_n \le \infty$, and the structure of these statistics is of obvious interest in the analysis of Π. Assuming further that p is aperiodic, it follows from Lemma 3.1 there is a limiting joint distribution of $(\mathrm{Cyc}_n, \mathrm{Cmp}_n, \mathrm{Blk}_n)$ as $n \to \infty$, and that this limiting joint distribution can in principle be computed from the block formula (3.1). In practice, however, the evaluation of this limiting joint distribution is not easy, even for the simplest regenerative models.

Suppose that a large number M of blocks of Π are formed and concatenated to make a permutation of the first N integers for $N \sim M\mu$ almost surely as $M \to \infty$. Then among these $N \sim M\mu$ integers, there are about $M\ell p_\ell$ integers contained in regenerative blocks of length ℓ. So for an integer $i = \lfloor UN \rfloor$ picked uniformly at random in $[N]$, the probability that this random integer falls in a regeneration block of length ℓ is approximately

$$\mathbb{P}(\lfloor UN \rfloor \in \text{regeneration block of length } \ell) \approx \frac{M\,\ell\,p_\ell}{M\mu} = \frac{\ell p_\ell}{\mu}.$$

This is the well known size-biased limit distribution of the length of block containing a fixed point in a renewal process. Now given that $\lfloor UN \rfloor$ falls in a regeneration block of length ℓ, the location of $\lfloor UN \rfloor$ relative to the start of this block has uniform distribution on $[\ell]$. These intuitive ideas are formalized and extended by the proposition below, which follows from Lemma 3.1, and the renewal reward theorem for ergodic averages [6, Theorem 3.1].

Proposition 3.3. *Let Π be a positive recurrent regenerative random permutation of $\mathbb{N}_+$, with block length distribution p with finite mean μ, and blocks governed by $(Q_n;\ n \ge 1)$.*

(*i*). *Let $C_{n,j}$ be the number of cycles of Π of length j that are wholly contained in $[n]$. Then the cycle counts have limit frequencies*

$$\lim_{n\to\infty} \frac{C_{n,j}}{n} = \frac{\nu_j}{\mu} \quad a.s. \quad \text{for } j \ge 1, \tag{3.2}$$

where ν_j is the expected number of cycles of length j in a generic block of Π, and $\mu = \sum_j j\nu_j$. The same conclusion holds with $C_{n,j}$ replaced by the larger number of cycles of Π of length j whose least element is contained in $[n]$.

(*ii*). *If the block length distribution p is aperiodic, then*

$$\lim_{n\to\infty} \mathbb{P}(\Pi \text{ regenerates at } n) = 1/\mu,$$

and

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{Cyc}_n = j) = \frac{j\nu_j}{\mu} \quad \text{for } j \in \mathbb{N}_+. \tag{3.3}$$

Alternatively, let L_ℓ^* be a random variable with values in $[\ell]$, which is the length of a sized-biased cycle of a random permutation of $[\ell]$ distributed as Q_ℓ. Then

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{Cyc}_n = j, \mathrm{Blk}_n = \ell) = \frac{\ell p_\ell}{\mu}\mathbb{P}(L_\ell^* = j) \quad \text{for } 1 \le j \le \ell, \tag{3.4}$$

and

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{Cyc}_n = j) = \frac{1}{\mu}\sum_{\ell=1}^{\infty} \ell p_\ell \mathbb{P}(L_\ell^* = j) \quad \text{for } j \ge 1. \tag{3.5}$$

(*iii*). *Continue to assume that p is aperiodic, there is an almost sure limiting frequency p_j° of cycles of Π of length j, relative to cycles of all lengths. These limiting frequencies are uniquely determined by*

$$p_j^\circ = \frac{\nu_j}{\sum_{j=1}^\infty \nu_j} \quad \textit{for } j \in \mathbb{N}_+, \tag{3.6}$$

or by the relations

$$p_j^\circ = \frac{\mu^\circ}{\mu} \frac{1}{j} \sum_{\ell=1}^\infty \ell p_\ell \mathbb{P}(L_\ell^* = j) \quad \textit{for } j \in \mathbb{N}_+, \tag{3.7}$$

with $\mu^\circ := \sum_{j=1}^\infty j p_j^\circ$.

(*iv*). *The statements $(i) - (iii)$ hold with cycles replaced by components, with almost sure limiting frequencies $p_j^\dagger$ of components of Π of length j.*

4. Blocked permutations

In this section we study an example of regenerative permutations where it is possible to describe the limiting cycle count frequencies explicitly. The story arises from the following observation of Shepp and Lloyd [93].

Lemma 4.1. [93] *Let N be a random variable with the geometric$(1-q)$ distribution on $\mathbb{N}_0$. That is,*

$$\mathbb{P}(N = n) = q^n(1-q) \quad \textit{for } n \geq 0.$$

Let Π be a uniform random permutation of $[N]$. Let $(N_j;\ j \geq 1)$ be the cycle counts of Π, which given $N = 0$ are identically 0, and given $N = n$ are distributed as the counts of cycles of various lengths j in a uniform random permutation of $[n]$. Then $(N_j;\ j \geq 1)$ are independent Possion random variables with means

$$\mathbb{E}N_j = \frac{q^j}{j} \quad \textit{for } j \geq 1.$$

The *Lévy-Itô representation* of N with the infinitely divisible geometric$(1-q)$ distribution as a weighted linear combination of independent Poisson variables, is realized as $N = \sum_{j=1}^\infty j N_j$. The possibility that $N = 0$ is annoying for concatenation of independent blocks. But this is avoided by simply conditioning a sequence of independent replicas of this construction on $N > 0$ for each replica. The obvious identity $N_j 1(N > 0) = N_j$ allows easy computation of

$$\mathbb{E}(N_j \,|\, N > 0) = \frac{\mathbb{E}N_j}{\mathbb{P}(N > 0)} = \frac{q^{j-1}}{j} \quad \text{for } j \geq 1. \tag{4.1}$$

Similarly, for $k = 1, 2 \ldots$

$$\mathbb{P}(N_j = k \,|\, N > 0) = \frac{1}{k!\, q} \left(\frac{q^j}{j}\right)^k \exp\left(-\frac{q^j}{j}\right) \quad \text{for } j \geq 1, \tag{4.2}$$

hence by summation

$$\mathbb{P}(N_j > 0 \,|\, N > 0) = \frac{1}{q}\left[1 - \exp\left(-\frac{q^j}{j}\right)\right] \quad \text{for } j \geq 1. \tag{4.3}$$

Proposition 4.2. *Let Π be the regenerative random permutation of $\mathbb{N}_+$, which is the concatenation of independent blocks of uniform random permutations of lengths $Y_1, Y_2, \ldots$ where each $Y_i > 0$ has the geometric$(1-q)$ distribution on $\mathbb{N}_+$. Then:*

(*i*). *The limiting cycle count frequencies ν_j/μ in (3.2) are determined by the formula $\mu := \mathbb{E}(Y_1) = (1-q)^{-1}$, and*

$$\nu_j = \frac{q^{j-1}}{j} \quad \text{for } j \in \mathbb{N}_+. \tag{4.4}$$

(*ii*). *The distribution of Π_1 is given by*

$$\mathbb{P}(\Pi_1 = k) = \frac{1-q}{q}\left(\lambda_1(q) - \sum_{h=1}^{k-1} \frac{q^h}{h}\right) \quad \text{for } k \in \mathbb{N}_+, \tag{4.5}$$

where

$$\lambda_1(q) := \sum_{h=1}^{\infty} \frac{q^h}{h} = -\log(1-q).$$

(*iii*). *The probability of the event $\{\Pi_1 = 1, \Pi_2 = 2\}$ that both 1 and 2 are fixed points of Π, is*

$$\mathbb{P}(\Pi_1 = 1, \Pi_2 = 2) = 1 - q. \tag{4.6}$$

(*iv*). *The regenerative random permutation Π is not strictly regenerative.*

Proof. (*i*). This follows readily from the formula (4.1) for the cycle counts in a generic block.
(*ii*). By conditioning on the first block length Y_1, since given $Y_1 = y$ the distribution of Π is uniform on $[y]$, there is the simple computation for $k = 1, 2, \ldots$

$$\mathbb{P}(\Pi_1 = k) = \sum_{y=k}^{\infty} q^{y-1}(1-q)\frac{1}{y},$$

which leads to (4.5). In particular, the probability that 1 is a fixed point of Π is

$$\mathbb{P}(\Pi_1 = 1) = -\frac{1-q}{q}\log(1-q).$$

(*iii*). The joint probability of the event $\{\Pi_1 = 1, \Pi_2 = 2\}$ is computed as

$$\begin{aligned}
\mathbb{P}(\Pi_1 = 1, \Pi_2 = 2) &= \mathbb{P}(Y_1 = 1, \Pi_1 = 1, \Pi_2 = 2) + \mathbb{P}(Y_1 \geq 2, \Pi_1 = 1, \Pi_2 = 2) \\
&= (1-q) \cdot \frac{1-q}{q}\lambda_1(q) + \sum_{y=2}^{\infty} q^{y-1}(1-q)\frac{1}{y(y-1)} \\
&= \frac{(1-q)^2}{q}\lambda_1(q) + \frac{1-q}{q}\lambda_2(q),
\end{aligned}$$

where

$$\lambda_2(q) := \sum_{h=2}^{\infty} \frac{q^h}{h(h-1)} = q - (1-q)\lambda_1(q). \tag{4.7}$$

But this simplifies, by cancellation of the two terms involving $\lambda_1(q)$, to the formula (6.5).
(*iv*). This follows from the fact that $\mathbb{P}(\Pi_1 = 1, \Pi_2 = 2) \neq \mathbb{P}(\Pi_1 = 1)^2$. □

More generally, the probability of the event $\{\Pi_i = i,\ 1 \le i \le k\}$ involves

$$\lambda_k(q) := \sum_{h=k}^{\infty} \frac{q^h}{h(h-1)\cdots(h-k+1)} = \frac{1}{a_k} q p_{k-2}(q) + \frac{(q-1)^{k-1}}{(k-1)!}\lambda_1(q),$$

for some $a_k \in \mathbb{Z}$ and $p_{k-2} \in \mathbb{Z}_{k-2}[q]$. The sequence $(a_k;\ k \ge 1)$ appears to be the sequence A180170 of OEIS, see [76] for related discussion.

The above example is generalized by following proposition, which is a corollary of Theorem 3.2 and Proposition 3.3.

Proposition 4.3. *Let Π be a positive recurrent random permutation of $\mathbb{N}_+$, whose block lengths Y_k are i.i.d. with distribution p, and whose reduced block permutations given their lengths are uniform on $\mathfrak{S}_n$ for each length n.*

(i). *The limiting cycle count frequencies ν_j/μ in (3.2) are determined by the formula*

$$\nu_j = j^{-1}\mathbb{P}(Y_1 \ge j) \quad \text{for } j \in \mathbb{N}_+, \tag{4.8}$$

where $\mathbb{P}(Y_1 \ge j) = \sum_{i=j}^{\infty} p_i$. So the almost sure limiting frequencies p_j° of cycles of Π of length j are given by

$$p_j^{\circ} = \frac{\sum_{i=j}^{\infty} p_i}{j\sum_{i=1}^{\infty} p_i H_i} \quad \text{for } j \in \mathbb{N}_+, \tag{4.9}$$

where $H_i := \sum_{j=1}^{i} 1/j$ is the i^{th} harmonic sum.

(ii). *If p is aperiodic, the limit distribution of displacements $D_n := \Pi_n - n$ as $n \to \infty$ is the common distribution of the displacement $D_z^* := \Pi_z^* - z$ for every $z \in \mathbb{Z}$, which is symmetric about 0, according to the formula*

$$\lim_{n\to\infty} \mathbb{P}(D_n = d) = \mathbb{P}(D_z^* = d) = \frac{1}{\mu}\mathbb{E}\left(\frac{(Y_1 - |d|)_+}{Y_1}\right) \quad \text{for } d \in \mathbb{Z}, \tag{4.10}$$

which implies

$$\lim_{n\to\infty} \mathbb{P}(\Pi_n > n) = \mathbb{P}(D_z^* > 0) = \frac{1}{2}\left(1 - \frac{1}{\mu}\right), \tag{4.11}$$

and the same holds for $<$ instead of $>$.

(iii). *Continuing to assume that p is aperiodic, there is also the convergence of absolute moments of all orders $r > 0$*

$$\lim_{n\to\infty} \mathbb{E}|D_n|^r = \mathbb{E}|D_z^*|^r = \frac{2}{\mu}\mathbb{E}\delta_r(Y), \tag{4.12}$$

where

$$\delta_r(n) := \sigma_r(n) - n^{-1}\sigma_{r+1}(n) \quad \text{with } \sigma_r(n) := \sum_{k=1}^{n} k^r,$$

the sum of r^{th} powers of the first n positive integers. In particular, for $r \ge 1$, $\delta_r(n)$ is a polynomial in n of degree $r+1$, for instance

$$\delta_1(n) = \frac{1}{6}(n^2 - 1), \quad \delta_2(n) = \frac{1}{12}n(n^2 - 1),$$

implying that the limit distribution of displacements has a finite absolute moment of order r if and only if $\mathbb{E}Y_1^{r+1} < \infty$.

Proof. (i). Recall the well known fact that for a uniform random permutation of $[n]$, for $1 \le j \le n$ the expected number of cycles of length j is $\mathbb{E}C_{n,j} = 1/j$. This follows from the easier fact that the length of a size-biased pick from the cycles of a uniform permutation of $[n]$ is uniformly distributed on $[n]$, and the probability $1/n$ that the size-biased pick has length j can be computed by conditioning on the cycle counts as $1/n = \mathbb{E}[jC_{n,j}/n]$. Appealing to the uniform distribution of blocks given their lengths, given Y_1 the expected number of j-cycles in the block of length Y_1 is $(1/j)1(Y_1 \ge j)$, and the conclusion follows. The limiting frequencies (4.9) are computed by injecting the formula (4.8) for cycle counts into (3.6).

(ii). This follows from Lemma 3.1, with the expression for the limit distribution of $D_z^* := \Pi_z^* - z$ given by

$$\mathbb{P}(D_z^* = d) = \frac{1}{\mu} \sum_{k=1}^{\infty} \mathbb{P}(\Pi_k = k + d, Y_1 \ge k). \tag{4.13}$$

By construction of Π, given $Y_1 = y$ for some $y \ge k$, the image of Π_k is a uniform random pick from $[y]$, so

$$\mathbb{P}(\Pi_k = k + d, Y_1 \ge k, Y_1 = y) = 1(1 \le k + d \le y) y^{-1} p_y \quad \text{for } y \ge k.$$

Sum this expression over y, then switch the order of summations over k and y, to see that for each fixed $y \ge 1$ the coefficient of $\mu^{-1} y^{-1} p_y$ in (4.13) is

$$\sum_{k=1}^{\infty} 1(1 \le k + d \le y) = (y - |d|)_+,$$

since if $d \ge 0$ the sum over k is effectively from 1 to $y - d$, and while if $d < 0$ it is from $1 + |d|$ to y, and in either case the number of non-zero terms is $y - |d|$ if $|d| < y$, and 0 otherwise. This gives the expression for the limit on the right side of (4.10), from which follow the remaining assertions.

(iii). This follows from the formula (4.10), a known result of convergence of moments in the limit theorem for regenerative stochastic processes [6, Chapter VI, Problem 1.4], and Bernoulli's formula for $\sigma_r(n)$ as a polynomial in n of degree $r + 1$, see e.g. Beardon [10]. □

Note, however, that the companion results for components of Π seem to be complicated. For instance, there is in general no simple expression for the expected number of components of Π of a fixed length. The limiting frequencies $p_j^\dagger$ of components of Π of length j are obtained by plugging (2.8) into (3.7), which are determined implicitly by the relations

$$p_j^\dagger = \frac{\mu^\dagger}{\mu} (j,1)^\dagger \sum_{\ell=1}^{\infty} \frac{p_\ell}{\ell!} \sum_{k=1}^{\ell} k \, (l - j, k - 1)^\dagger \quad \text{for } j \in \mathbb{N}_+. \tag{4.14}$$

5. p-shifted permutations

In this section we study the p-shifted permutations introduced in Definition 1.3. It is essential that p be fixed and not random to make p-shifted permutations regenerative. The point is that if p is replaced by a random P, the observation of $\Pi_1, \ldots, \Pi_n$ given a split at n allows some inference to be made about the P_i, $1 \le i \le n$. But according to the definition of the P-shifted permutation, these same values of P_i are used to create the remaining permutation of $\mathbb{N}_+ \setminus [n]$. Consequently, the independence condition required for regeneration at n will fail for any non-degenerate random P. Now we give a proof of Proposition 1.4.

Proof of Proposition 1.4. (i). This is clear from the definition of p-shifted permutations.
(ii). Observe that

$$u_n = \sum_{\pi \in \mathfrak{S}_n} \prod_{j=1}^{n} p\left(\pi_j - \sum_{1 \le i < j} 1(\pi_i < \pi_j)\right) = \sum_{\pi \in \mathfrak{S}_n} \prod_{j=1}^{n} p\left(j - \sum_{1 \le i < \pi_j^{-1}} 1(\pi_i < j)\right),$$

and the conclusion follows from the well known bijection $\mathfrak{S}_n \to [1] \times [2] \ldots \times [n]$ defined by

$$\pi \mapsto \left(j - \sum_{1 \le i < \pi_j^{-1}} 1(\pi_i < j);\ 1 \le j \le n\right).$$

(iii)-(iv). The strict regeneration is clear from the definition of p-shifted permutations, and the generating function (1.17) follows easily from the the general theory of regenerative processes [34, Chapter XIII].
(v)-(vi). The particular case of these results for p the geometric$(1-q)$ distribution was given by Basu and Bhatnagar [9, Lemmas 4.1 and 4.2]. Their argument generalizes as follows. The key observation is that for $X_1, X_2, \ldots$ the i.i.d. sample from p which drives the construction of the p-shifted permutation Π, the sequence M_n defined by $M_0 := 0$ and

$$M_n := \max(M_{n-1}, X_n) - 1,$$

has the interpretation that

$$M_n = \#\left\{i : 1 \le i \le \max_{1 \le j \le n} \Pi_j\right\} - n,$$

which can be understood as the current number of gaps in the range of $\Pi_j, 1 \le j \le n$. The event $\{\Pi \text{ regenerates at } n\}$ is then identical to the event $\{M_n = 0\}$. It is easily checked that $(M_n;\ n \ge 0)$ is a Markov chain with state space $\mathbb{N}_0$, and the unique invariant measure $(\mu_i;\ i \in \mathbb{N}_0)$ for the Markov chain $(M_n;\ n \ge 0)$ is given by

$$\mu_0 = 0 \quad \text{and} \quad \mu_i = \frac{\mathbb{P}(X_1 > i)}{\prod_{j=1}^{i}[1 - \mathbb{P}(X_1 > j)]} \quad \text{for } i \ge 1. \tag{5.1}$$

Moreover, it follows by standard analysis that this sequence μ_j is summable if and only if the mean m of X_1 is finite. The conclusion follows from the well known theory of Markov chains [30, Chapter 6], and Theorem 3.2. □

See also Alappattu and Pitman [2, Section 3] for a similar argument used to derive the stationary distribution of the lengths of the loop-erasure in a loop-erased random walk. For the p-shifted permutation, the first splitting probabilities $f_n := \mathbb{P}(\Pi \text{ first splits at } n)$ are given by the explicit formulas

$$
\begin{aligned}
f_1 =& p_1, \\
f_2 =& p_1 p_2, \\
f_3 =& p_1 p_2^2 + p_1^2 p_3 + p_1 p_2 p_3, \\
f_4 =& p_1 p_2^3 + 2p_1^2 p_2 p_3 + 2p_1 p_2^2 p_3 + p_1^2 p_3^2 + p_1 p_2 p_3^2 \\
&+ p_1^3 p_4 + 2p_1^2 p_2 p_4 + p_1 p_2^2 p_4 + p_1^2 p_3 p_4 + p_1 p_2 p_3 p_4.
\end{aligned}
$$

It is easily seen that for each n, $f_n(p_1, p_2, \ldots)$ is a polynomial of degree n in variables $p_1, \ldots, p_n$. The polynomial so defined makes sense even for variables p_i not subject to the constraints of a probability distribution. The polynomial can be understood as an enumerator polynomial for the vector of counts

$$R_{n,j} := 1\left(\pi_j - \sum_{i=1}^{j} 1(\pi_i < \pi_j)\right) \quad \text{for } 1 \leq j \leq n.$$

In the polynomial for f_n, the choice of $\pi_1, \ldots, \pi_n$ is restricted to the set $\mathfrak{S}_n^\dagger$ of indecomposable permutations of $[n]$, and the coefficient of $p_1^{r_1} \ldots p_n^{r_n}$ is for each choice of non-negative integers $r_1, \ldots, r_n$ with $\sum_{i=1}^n r_i = n$ is the number of indecomposable permutations of $[n]$ such that $\sum_{j=1}^n 1(R_{n,j} = i) = r_i$ for each $1 \leq i \leq n$. In particular, the sum of all the integer coefficients of these monomials is

$$f_n(1, 1, \ldots) = (n, 1)^\dagger,$$

which is the number of indecomposable permutations of $[n]$ discussed in Section 2.

In general, a recurrent strictly regenerative random permutation of $\mathbb{N}_+$ is specified by two sequences:

(*i*). an arbitrary sequence $(f_n;\ n \geq 1)$ of first splitting probabilities, subject only to the constraints $f_n \geq 0$ and $\sum_{n=1}^\infty f_n = 1$;

(*ii*). an arbitrary sequence $(Q_n;\ n \geq 1)$ of probability distributions on $\mathfrak{S}_n^\dagger$, the set of indecomposable permutations of $[n]$.

The construction is provided by the following rather obvious proposition.

Proposition 5.1. *Given the two ingredients (i) and (ii), there is a unique distribution of a recurrent strictly regenerative permutation Π of $\mathbb{N}_+$ such that*

$$\mathbb{P}(\Pi \textit{ first splits at } n) = f_n,$$

and

$$(\Pi_1, \ldots, \Pi_n \,|\, \Pi \textit{ first splits at } n) \textit{ is distributed as } Q_n.$$

This distribution is obtained by specifying that Π has splitting times $0 = T_0 < T_1 < \ldots$ where the $T_i - T_{i-1}$ are i.i.d. with distribution $(f_n;\ n \geq 1)$, and conditionally given $T_0, T_1, \ldots$ with $T_i - T_{i-1} = n_i$ for $i \geq 1$, the permutation Π is constructed by picking independent random permutations of $[n_i]$ according to Q_{n_i}, and then shifting these permutations to define the restriction of Π to $[T_{i-1}+1, T_i]$ for every $i \geq 1$.

The analog of the above proposition for a transient regenerative permutation Π of $\mathbb{N}_+$ is even nastier. For f_n with $F(1) := \sum_n f_n \in (0,1)$ the construction can proceed just as above up to time T_N for T_n the sum of n independent copies of T_1 with distribution $\mathbb{P}(T_1 = n) = f_n/F(1)$, and N has geometric$(1-F(1))$ distribution on $\mathbb{N}_0$, independent of the sequence $(T_n;\ n \geq 1)$. Then T_N is the time of the last finite split point of Π, and given $T_N = n$ and the restriction of Π to $[n]$ so created, the restriction of Π to $[n, \infty)$, shifted back to be a permutation of $\mathbb{N}_+$, can be constructed according to any fixed probability distribution on the set of all permutations of $\mathbb{N}_+$ with no splitting times.

Properties of the limiting Mallows(q) permutations of $\mathbb{N}_+$ and of $\mathbb{Z}$ are obtained by specializing Proposition 1.4 with p the geometric$(1-q)$ distribution on $\mathbb{N}_+$. Many results of [46, 38] acquire simpler proofs by this approach. The following corollary also exposes a number of properties of the limiting Mallows(q) models which were not mentioned in previous works.

Corollary 5.2. *For each $0 < q < 1$, with $\mathbb{P}_q$ governing Π as a geometric$(1-q)$-shifted permutation of $\mathbb{N}_+$, the conclusions of Proposition 1.4 apply with the following reductions:*

(i). *The formula* (1.15) *reduces to*

$$\mathbb{P}_q(\Pi_i = \pi_i, 1 \leq i \leq n) = (1-q)^n \, q^{\mathrm{inv}(\pi)+\delta(n,\pi)} \tag{5.2}$$

where $\mathrm{inv}(\pi)$ *is the number of inversions of* π, *and* $\delta(n,\pi) := \sum_{i=1}^n \pi_i - \frac{1}{2}n(n+1)$. *In particular,* (5.2) *holds with the further simplification* $\delta(n,p) = 0$ *if and only if* π *is a permutation of* $[n]$.

(ii). *The $\mathbb{P}_q$ distribution is the unique probability distribution on permutations of $\mathbb{N}_+$ such that for every n, the conditional distribution of $(\Pi_1, \ldots, \Pi_n)$ given that $[n]$ is a block of Π is the Mallows(q) distribution $M_{n,q}$ on permutations of $[n]$, as in* (1.2).

(iii). *The probability that Π maps $[n]$ to $[n]$ is*

$$u_{n,q} := \mathbb{P}_q([n] \textit{ is a block of } \Pi) = (1-q)^n Z_{n,q}, \tag{5.3}$$

where $Z_{n,q}$ is defined by (1.3).

(iv). *The $\mathbb{P}_q$ distribution of Π is strictly regenerative, with regeneration at every n such that $[n]$ is a block of $\mathbb{N}_+$, and renewal sequence $(u_{n,q};\ n \geq 1)$ as above.*

(v). *The $\mathbb{P}_q$ distribution of component lengths $f_{n,q} = \mathbb{P}_q(Y_1 = n)$, where Y_1 is the length of the first component of Π, is given by the probability generating function*

$$\sum_{n=1}^{\infty} f_{n,q} z^n = 1 - \frac{1}{U_q(z)} \quad \textit{where } U_q(z) = 1 + \sum_{n=1}^{\infty} u_{q,n} z^n, \tag{5.4}$$

as well as by the formula

$$f_{n,q} = (1-q)^n \, Z^{\dagger}_{n,q}, \tag{5.5}$$

where $Z^{\dagger}_{n,q} := \sum_{\pi \in \mathfrak{S}_n} 1(\pi \in \mathfrak{S}^{\dagger}_n) q^{\mathrm{inv}(\pi)}$ *is the restricted partition function of the Mallows(q) distribution $M_{n,q}$ on the set $\mathfrak{S}^{\dagger}_n$ of indecomposable permutations of $[n]$.*

(vi). *Under $\mathbb{P}_q$, conditionally given the component lengths, say $Y_i = n_i$ for $i = 1, 2, \ldots$, the reduced components of Π are independent random permutations of $[n_i]$ with conditional Mallows(q) distributions $M^{\dagger}_{n_i,q}$ defined by*

$$M^{\dagger}_{n_i,q}(\pi) := \frac{1}{Z^{\dagger}_{n_i,q}} 1(\pi \in \mathfrak{S}^{\dagger}_{n_i}) q^{\mathrm{inv}(\pi)} \quad \textit{for } \pi \in \mathfrak{S}_n. \tag{5.6}$$

6. p-biased permutations

This section provides a detailed study of p-biased permutations introduced in Definition 1.6. For a P-biased permutation Π of $\mathbb{N}_+$ with $P = (P_1, P_2, \ldots)$ a random discrete distribution, the joint distribution of $(\Pi_1, \ldots \Pi_n)$ is computed by the formula (1.20). In particular, the distribution of Π_1 is given by the vector of means $(\mathbb{E}(P_1), \mathbb{E}(P_2), \ldots)$. So if P is the GEM(θ) distribution, then Π_1 has the geometric$(\theta/(1+\theta))$ distribution on $\mathbb{N}_+$. The index Π_1 of a single size-biased pick from $(P_1, P_2, \ldots)$, and especially the random size P_{Π_1} of this pick from $(P_1, P_2, \ldots)$ plays an important role in the theory of random discrete distributions and associated random partitions of positive integers [84]. Features of the joint distribution of $(P_{\Pi_1}, \ldots, P_{\Pi_n})$ also play an important role in this setting [83], but we are unaware of any previous study of $(\Pi_1, \Pi_2, \ldots)$ regarded as a random permutation of $\mathbb{N}_+$.

We start with the following construction of size-biased permutations from Perman, Pitman and Yor [79]. See also Gordon [51] where this construction is indicated in the abstract, and Pitman and Tran [85] for further references to size-biased permutations.

Lemma 6.1. [79, Lemma 4.4] *Let $(L_i;\ 1 \le i \le n)$ be a possibly random sequence such that $\sum_{i=1}^n L_i = 1$, and $(\varepsilon_i;\ 1 \le i \le n)$ be i.i.d. standard exponential variables, independent of the L_i's. Define*

$$Y_i := \frac{\varepsilon_i}{L_i} \quad \text{for } 1 \le i \le n.$$

Let $Y_{(1)} < \cdots < Y_{(n)}$ be the order statistics of the Y_i's, and $L_1^*, \cdots, L_n^*$ be the corresponding L values. Then $(L_i^*;\ 1 \le i \le n)$ is a size-biased permutation of $(L_i;\ 1 \le i \le n)$.

By applying the formula (1.8) and Lemma 6.1, we evaluate the splitting probabilities for P-biased permutation with P a random discrete distribution.

Proposition 6.2. *Let Π be a P-biased permutation of $\mathbb{N}_+$ of a random discrete distribution $P = (P_1, P_2, \ldots)$, and $T_n := 1 - \sum_{i=1}^n P_i$. Then the probability that Π maps $[n]$ to $[n]$ is given by (1.8) with*

$$\Sigma_{n,j} = \sum_{1 \le i_1 < \cdots < i_j \le n} \mathbb{E}\left(\frac{T_n}{T_n + P_{i_1} + \cdots + P_{i_j}}\right). \tag{6.1}$$

Proof. Recall the definition of $A_{n,i}$ from (1.7). By Lemma 6.1, for $1 \le i_1 < \cdots < i_j \le n$,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{k=1}^{j} A_{n,i_k}\right) &= \mathbb{P}\left(\min_{1 \le k \le j}\left(\frac{\varepsilon_{i_k}}{P_{i_k}}\right) > \frac{\varepsilon}{T_n}\right) \\ &= \mathbb{E}\exp\left(-\frac{\varepsilon(P_{i_1} + \cdots + P_{i_j})}{T_n}\right) \\ &= \mathbb{E}\left(\frac{T_n}{T_n + P_{i_1} + \cdots + P_{i_j}}\right),\end{aligned}$$

which leads to the desired result. □

In terms of the occupancy scheme by throwing balls independently into an infinite array of boxes indexed by $\mathbb{N}_+$ with random frequencies $P = (P_1, P_2, \ldots)$, the quantity $\Sigma_{n,j}$ has the following interpretation. Let C_n be the count of empty boxes when the first box in $\{n+1, n+2, \cdots\}$ is filled. Then

$$\Sigma_{n,j} = \mathbb{E}\binom{C_n}{j}. \tag{6.2}$$

Further analysis of C_n and $\Sigma_{n,j}$ for the GEM(θ) model will be presented in the forthcoming article [29]. Contrary to p-shifted permutations, we consider P-biased permutations where P is determined by a RAM (1.21). In the latter case, the only model with P fixed is the geometric($1-q$)-biased permutation for $0 < q < 1$. Now we give a proof of Proposition 1.7.

Proof of Proposition 1.7. (i). The strict regeneration follows easily from the stick breaking property of RAM models. By Lemma 6.1, the renewal probabilities u_n are given by

$$u_n = \mathbb{P}\left(\max_{1 \le i \le n} \frac{\varepsilon_i}{P_i} < \frac{\varepsilon}{T_n}\right),$$

where $P_i = W_i \prod_{j=1}^{i-1}(1-W_j)$, $T_n = \prod_{j=1}^{n}(1-W_j)$, and the ε_i's and ε are independent standard exponential variables. Note that for each $x > 0$,

$$\mathbb{P}\left(\max_{1 \le i \le n} \frac{\varepsilon_i}{P_i} < \frac{x}{T_n}\right) = \mathbb{P}\left(\bigcap_{i=1}^{n}\left\{\varepsilon_i < \frac{xP_i}{T_n}\right\}\right) = \mathbb{E}\prod_{i=1}^{n}\left(1 - e^{-\frac{xP_i}{T_n}}\right),$$

which, by conditioning on $\varepsilon = x$, leads to

$$u_n = \int_0^\infty e^{-x}\mathbb{E}\prod_{i=1}^{n}\left(1 - e^{-xP_i/T_n}\right)dx. \tag{6.3}$$

Since $(W_1, \ldots, W_n) \overset{(d)}{=} (W_n, \ldots, W_1)$ for every $n \ge 1$, the formula (6.3) simplifies to (1.23). So to prove $u_\infty > 0$, it suffices to prove (1.24).
(ii). This is the deterministic case where $P_i = q^{i-1}(1-q)$ and $T_n = q^n$. So the formula (1.23) specializes to

$$u_n = \int_0^\infty e^{-x}\prod_{i=1}^{n}\left(1 - e^{-x(1-q)/q^i}\right)dx.$$

It follows by standard analysis that $u_\infty := \lim_{n\to\infty} u_n > 0$ if and only if

$$\sum_{i=1}^{\infty} e^{-x(1-q)/q^i} < \infty.$$

But this is obvious for $0 < q < 1$, which implies that Π is positive recurrent.
(iii). This case corresponds to $P_i = W_i \prod_{j=1}^{i-1}(1-W_j)$ and $T_n = \prod_{j=1}^{n}(1-W_j)$, where W_i are i.i.d. beta$(1,\theta)$ variables. Note that for each i, W_i is independent of T_{i-1}. By conditioning on T_{i-1}, we get

$$\begin{aligned}
\sum_{i=2}^{\infty}\mathbb{E}\exp\left(-\frac{xW_i}{T_i}\right) &= \int_0^1 \mathbb{E}\exp\left(-\frac{xw}{T_{i-1}(1-w)}\right)\cdot\theta(1-w)^{\theta-1}dw \\
&= \int_0^1\int_0^1 \exp\left(-\frac{xw}{t(1-w)}\right)\cdot\frac{\theta}{u}\cdot\theta(1-w)^{\theta-1}dtdw,
\end{aligned}$$

where the second equality follows from Ignatov's description [56] of GEM(θ) variables as a Poisson point process on $(0,1)$ with intensity $\theta(1-u)^{-1}du$. Note that

$$\int_0^1 \int_0^1 \exp\left(-\frac{xw}{t(1-w)}\right) u^{-1}(1-w)^{\theta-1} dt dw = \int_0^1 E_1\left(\frac{xw}{1-w}\right)(1-w)^{\theta-1}dw,$$

where $E_1(x) := \int_x^\infty u^{-1}e^{-u}du$ with $E_1(x) \sim -\log(x)$ as $x \to 0^+$. It follows by elementary estimates that the above integral is finite, which leads to the desired result. □

Let Π the P-biased permutation of $\mathbb{N}_+$ for P the geometric$(1-q)$ distribution, with the renewal sequence $(u_{n,q};\ n \ge 1)$. Let $C_{n,1,q}$ be the number of fixed points of Π contained in $[n]$, and $\nu_{1,q}$ be the expected number of fixed points of Π in a generic component. According to Proposition 3.3,

$$\lim_{n\to\infty} \frac{C_{n,1,q}}{n} = \nu_{1,q} u_{\infty,q} \quad a.s.,$$

where $u_{\infty,q} := \lim_{n\to\infty} u_{n,q}$. Note that with probability $1-q$, a generic component has only one element. This implies that $\nu_{1,q} \ge 1-q$. It follows from Proposition 6.2 that $\lim_{q\downarrow 0} u_{\infty,q} = 1$. As a result,

$$\lim_{n\to\infty} \frac{C_{n,1,q}}{n} = \alpha(q) \quad a.s. \quad \text{with } \lim_{q\downarrow 0} \alpha(q) = 1. \tag{6.4}$$

Similarly, by letting Π be the P-biased permutation of $\mathbb{N}_+$ for P the GEM(θ) distribution, and $C_{n,1,\theta}$ be the number of fixed points of Π contained in $[n]$,

$$\lim_{n\to\infty} \frac{C_{n,1,\theta}}{n} = \beta(\theta) \quad a.s. \quad \text{with } \lim_{\theta\downarrow 0} \beta(\theta) = 1. \tag{6.5}$$

7. Regenerative P-biased permutations

This section provides further analysis of P-biased permutations of $\mathbb{N}_+$, especially for P the GEM(1) distribution, with the W_i's i.i.d. uniform on $(0,1)$. While the formulas provided by (1.8), (1.23), or by summing the r.h.s. of (1.20) over all permutations $\pi \in \mathfrak{S}_k$ for the renewal probabilities u_k and their limit u_∞ are quite explicit, it is not easy to evaluate these integrals and their limit directly. For instance, even in the simplest case where the W_i's are uniform on $(0,1)$, explicit evaluation of u_k for $k \ge 2$ involves the values of $\zeta(j)$ of the Riemann zeta function at $j = 2, \dots, k$, as indicated later in Proposition 7.2.

We start with an exact simulation of the P-biased permutation for any $P = (P_1, P_2, \dots)$ with $P_i > 0$ for all $i \ge 1$ involving the following construction of a process $(\mathcal{W}_k;\ k \ge 1)$ with state space the set of finite unions of open subintervals of $(0,1)$, from P and a collection of i.i.d. uniform variables $U_1, U_2, \dots$ on $(0,1)$ independent of P.

- Construct $F_j = P_1 + \cdots + P_j$ until the least j such that $F_j > U_1$. Then set

$$\mathcal{W}_1 = \left(\bigcup_{i=1}^{j-1}(F_{i-1}, F_i)\right) \cup (F_j, 1),$$

  with convention $F_0 := 0$.

- Assume that $\mathcal{W}_{k-1}$ has been constructed for some $k \geq 2$ as a finite union of open intervals with the rightmost interval $(F_j, 1)$ for some $j \geq 1$. If U_k lands in one of the intervals of $\mathcal{W}_{k-1}$ that is not the rightmost interval, then remove that interval from $\mathcal{W}_{k-1}$ to create $\mathcal{W}_k$. If U_k hits the rightmost interval $(F_j, 1)$, then construct $F_\ell = F_j + P_{j+1} + \cdots + P_\ell$ for $\ell > j$ until the least ℓ such that $F_\ell > U_k$. Set
$$\mathcal{W}_k = (\mathcal{W}_{k-1} \cap (0, F_j)) \cup \left(\bigcup_{i=j+1}^{\ell-1} (F_{i-1}, F_i) \right) \cup (F_\ell, 1).$$

It is not hard to see that a P-biased permutation Π of $\mathbb{N}_+$ can be recovered from the process $(\mathcal{W}_k;\ k \geq 1)$ driven by P, with Π_k a function of $\mathcal{W}_1, \ldots, \mathcal{W}_k$. In particular, the length Y_1 of the first component of Π is
$$Y_1 = \min\{k \geq 1 : \mathcal{W}_k \text{ is composed of a single interval } (F_\ell, 1) \text{ for some } l\}.$$
In the sequel, the notation $\stackrel{\theta}{=}$ or $\stackrel{\theta}{\approx}$ indicates exact or approximate evaluations for the GEM(θ) model; that is the residual factors W_i are i.i.d. beta$(1, \theta)$ distributed. By a simulation of the process $(\mathcal{W}_k;\ k \geq 1)$ for GEM(1), we get some surprising results:
$$\mathbb{E}Y_1 \stackrel{1}{\approx} 3 \quad \text{and} \quad Var(Y_1) \stackrel{1}{\approx} 11, \tag{7.1}$$
which suggests that
$$u_\infty = 1/\mathbb{E}Y_1 \stackrel{1}{=} 1/3. \tag{7.2}$$

These simulation results (7.1) are explained by the following lemma, which provides an alternative approach to the evaluation of u_k derived from a RAM. This lemma is suggested by work of Gnedin and coauthors on the Bernoulli sieve [49, 44, 43], and following work on extremes and gaps in sampling from a RAM by Pitman and Yakubovich [86, 82].

Lemma 7.1. *Let $X_1, X_2, \ldots$ be a sample from the RAM (1.21) with i.i.d. stick-breaking factors $W_i \stackrel{(d)}{=} W$ for some distribution of W on $(0,1)$. For positive integers n and $k = 0, 1, \ldots$ let*
$$Q_n^*(k) := \sum_{i=1}^n 1(X_i > k) \tag{7.3}$$
represent the number of the first n balls which land outside the first k boxes. For $m = 1, 2, \ldots$ let $n(k,m) := \min\{n : Q_n^*(k) = m\}$ be the first time n that there are m balls outside the first k boxes. Then:

- *For each k and m there is the equality of joint distributions*
$$\left(Q^*_{n(k,m)}(k-j), 0 \leq j \leq k \right) \stackrel{(d)}{=} (\widehat{Q}_j, 0 \leq j \leq k \,|\, \widehat{Q}_0 = m) \tag{7.4}$$
where $(\widehat{Q}_0, \widehat{Q}_1, \ldots)$ with $1 \leq \widehat{Q}_0 \leq \widehat{Q}_1 \cdots$ is a Markov chain with state space $\mathbb{N}_+$ and stationary transition probability function
$$\widehat{q}(m,n) := \binom{n-1}{m-1} \mathbb{E}W^{n-m}(1-W)^m \quad \text{for } m \leq n. \tag{7.5}$$
So $\widehat{q}(m, \bullet)$ is the mixture of Pascal $(m, 1-W)$ distributions, and the distribution of the $\widehat{Q}$ increment from state m is mixed negative binomial $(m, 1-W)$.

- *For each $k \geq 1$ the renewal probability u_k for the P-biased permutation of $\mathbb{N}_+$ for P a RAM is the probability that the Markov chain $\widehat{Q}$ started in state 1 is strictly increasing for its first k steps:*
$$u_k = \mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots < \widehat{Q}_k \,|\, \widehat{Q}_0 = 1). \tag{7.6}$$
- *The sequence u_k is strictly decreasing, with limit $u_\infty \geq 0$ which is the probability that the Markov chain $\widehat{Q}$ started in state 1 is strictly increasing forever:*
$$u_\infty = \mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots \,|\, \widehat{Q}_0 = 1). \tag{7.7}$$

Proof. For $0 < v < 1$ and $U_1, U_2, \ldots$ a sequence of i.i.d. uniform $[0,1]$ variables, let
$$N_n(v,1) := \sum_{i=1}^{n} 1(v < U_i < 1)$$
be the number of the first n values that fall in $(v,1)$, and let
$$g(v,m) := \min\{n \geq 1 : N_n(v,1) = m\}$$
be the random time when $N_n(v,1)$ first reaches m. So $g(v,m)$ has the Pascal$(m, 1-v)$ distribution of the sum of m independent random variables with geometric $(1-v)$ distribution on $\mathbb{N}_+$. Then there is the well known identity in distribution of Pascal counting processes [33, 17]
$$\left(N_{g(v,m)}(u,1), 0 \leq u \leq v\right) \overset{(d)}{=} \left(Y_m\left(\log\left(\frac{1-v}{1-u}\right)\right), 0 \leq u \leq v\right), \tag{7.8}$$
where $(Y_m(t), t \geq 0)$ is a *standard Yule process*; that is the pure birth process on positive integers with birth rate k in state k, with initial state $Y_m(0) = m$. Let the sample $X_1, X_2, \ldots$ from the RAM be constructed as $X_i = j$ iff $U_i \in (F_{j-1}, F_j]$ where $F_j := 1 - \prod_{i=1}^{j}(1 - W_i)$ for a sequence of stick-breaking factors $(W_i;\ i \geq 1)$ independent of the uniform sample points $(U_i;\ i \geq 1)$. Then by construction $Q^*_{n(k,m)}(i) = N_{g(F_k,m)}(F_i, 1)$ for each $0 \leq i \leq k$. The identity in distribution (7.8) yields
$$\left(Q^*_{n(k,m)}(i), 0 \leq i \leq k\right) \overset{(d)}{=} \left(Y_m\left(\log\left(\frac{1-F_k}{1-F_i}\right)\right), 0 \leq i \leq k\right), \tag{7.9}$$
first conditionally on $F_1, \ldots, F_k$, then also unconditionally, where on the right side it is assumed that the Yule process Y_m is independent of $F_1, \ldots, F_k$. By a reversal of indexing, and the equality in distribution $(W_k, \ldots, W_1) \overset{(d)}{=} (W_1, \ldots, W_k)$, this gives
$$\left(Q^*_{n(k,m)}(k-j), 0 \leq j \leq k\right) \overset{(d)}{=} \left(Y_m(\tau_j), 0 \leq j \leq k\right), \tag{7.10}$$
where $\tau_j := \sum_{i=1}^{j} - \log(1 - W_i)$, and the W_i are independent of the Yule process Y_m. It is easily shown that the process on the right side of (7.10) is a Markov chain with stationary transition function $\widehat{q}$ as in (7.5). This gives the first part of the lemma, and the remaining parts follow easily. □

For P the GEM(θ) distribution, the transition probability function $\widehat{q}$ of the $\widehat{Q}$ chain simplifies to
$$\widehat{q}(m,n) \overset{\theta}{=} \frac{(m)_{n-m}(\theta)_m}{(1+\theta)_n} \overset{1}{=} \frac{m}{n(n+1)} \quad \text{for } m \leq n, \tag{7.11}$$

where

$$(x)_j := x(x+1)\cdots(x+j-1) = \frac{\Gamma(x+j)}{\Gamma(x)}.$$

The Markov chain $\widehat{Q}$ with the transition probability function (7.11) for $\theta = 1$ was first encountered by Erdös, Rényi and Szüsz in their study [31] of *Engel's series* derived from U with uniform $(0,1)$ distribution, that is

$$U = \frac{1}{q_1} + \frac{1}{q_1 q_2} + \cdots + \frac{1}{q_1 q_2 \cdots q_n} + \cdots,$$

for a sequence of random positive integers $q_j \geq 2$. They showed that

$$(q_{k+1} - 1, k \geq 0) \overset{(d)}{=} (\widehat{Q}_k, k \geq 0),$$

for $\widehat{Q}$ with transition matrix $\widehat{q}$ as in (7.11) for $\theta = 1$, and initial distribution

$$\mathbb{P}(\widehat{Q}_0 = m) = \frac{1}{m(m+1)} \quad \text{for } m \geq 1. \tag{7.12}$$

Rényi [88, Theorem 1] showed that for this Markov chain derived from Engel's series, the occupation times

$$G_j := \sum_{k=0}^{\infty} 1(\widehat{Q}_k = j) \quad \text{for } j \geq 1, \tag{7.13}$$

are independent random variables with geometric$(j/(j+1))$ distributions on $\mathbb{N}_0$. Rényi deduced that with probability one the chain $\widehat{Q}$ is eventually strictly increasing, and [88, (4.5)] that for the initial distribution (7.12) of $\widehat{Q}_0$

$$\mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots) \overset{1}{=} \prod_{j=1}^{\infty} \mathbb{P}(G_j \leq 1) \overset{1}{=} \prod_{j=1}^{\infty} \frac{j(j+2)}{(j+1)^2} = \frac{1}{2}, \tag{7.14}$$

by telescopic cancellation of the infinite product. A slight variation of Rényi's calculation gives for each possible initial state m of the chain

$$\mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots \mid \widehat{Q}_0 = m) \overset{1}{=} \frac{m}{m+1} \prod_{j=m+1}^{\infty} \frac{j(j+2)}{(j+1)^2} = \frac{m}{m+2}. \tag{7.15}$$

The instance $m = 1$ of this formula, combined with (7.7), proves the formula (7.2) for u_∞ for the GEM(1) model. A straightforward variation of these calculations gives the corresponding result for P the GEM(θ) distribution:

$$u_\infty \overset{\theta}{=} \frac{1}{1+\theta} \prod_{j=2}^{\infty} \frac{j(j+2\theta)}{(j+\theta)^2} = \frac{\Gamma(\theta+2)\Gamma(\theta+1)}{\Gamma(2\theta+2)}. \tag{7.16}$$

A key ingredient in this evaluation is the fact that in the GEM(θ) model the random occupation times G_j of $\widehat{Q}$ are independent geometric variables, see [86, 82]. For a more general RAM, the G_j's may not be independent, and they may not be exactly geometric, only conditionally so given $G_j \geq 1$. The Yule representation (7.10) of $\widehat{Q}$ given $\widehat{Q}(0) = 1$ as $\widehat{Q}(j) = Y_1(\tau_j)$ combined with Kendall's representation [60, Theorem 1] of $Y_1(t) = 1 + N(\varepsilon(e^t - 1))$ for N a rate 1 Poisson process and ε standard exponential independent of N, only reduces the expression (7.7) for u_∞ back to the limit form as $n \to \infty$ of the previous expression (1.23).

So u_∞ for a RAM is always an integral over x of the expected value of an infinite product of random variables. See also [58, 57, 86] for treatment of closely related problems. Now we give a proof of Proposition 1.8.

Proof of Proposition 1.8. The result of [43, Theorem 3.3] shows that under the assumptions of the proposition, if L_n is the number of empty boxes to the left of the rightmost box when n balls are thrown, then

$$L_n \xrightarrow{(d)} L_\infty := \sum_{j=1}^{\infty} (G_j - 1)_+ \quad \text{as } n \to \infty,$$

where the right side is defined by the occupation counts (7.13) of the Markov chain $\widehat{Q}$ for the special entrance law

$$\mathbb{P}(\widehat{Q}_0 = m) = \frac{\mathbb{E}W^m}{m\,\mathbb{E}\left[-\log(1-W)\right]} \quad \text{for } m \geq 1, \tag{7.17}$$

which is the limit distribution of Z_n, the number of balls in the rightmost occupied box, as $n \to \infty$, and that also

$$\mathbb{E}L_n \to \mathbb{E}L_\infty = \frac{\mathbb{E}\left[-\log W\right]}{\mathbb{E}\left[-\log(1-W)\right]},$$

which is finite by assumption. It follows that $\mathbb{P}(L_\infty < \infty) = 1$, hence also that $\mathbb{P}_m(L_\infty < \infty) = 1$ for every m, where $\mathbb{P}_m(\bullet) := \mathbb{P}(\bullet \,|\, \widehat{Q}_0 = m)$. Let $R := \max\{j : G_j > 1\}$ be the index of the last repeated value of the Markov chain. From $\mathbb{P}_1(L_\infty < \infty) = 1$ it follows that $\mathbb{P}_1(R < \infty) = 1$, hence that $\mathbb{P}_1(R = r) > 0$ for some positive integer r. But for $r = 2, 3, \ldots$, a last exit decomposition gives

$$\mathbb{P}_1(R = r) = \left(\sum_{k=1}^{\infty} \mathbb{P}_1(\widehat{Q}_{k-1} = \widehat{Q}_k = r) \right) \mathbb{P}_r(L_\infty = 0),$$

where both factors on the right side must be strictly positive to make $\mathbb{P}_1(R = r) > 0$. Combined with a similar argument if $\mathbb{P}_1(R = 1) > 0$, this implies $\mathbb{P}_r(L_\infty = 0) > 0$ for some $r \geq 1$, hence also

$$u_\infty = \mathbb{P}_1(L_\infty = 0) \geq \mathbb{P}_1(G_i \leq 1 \text{ for } 0 \leq i < r, G_r = 1)\mathbb{P}_r(L_\infty = 0) > 0,$$

which is the desired conclusion. $\square$

To conclude, we present explicit formulas for u_k of a GEM(1)-biased permutation of $\mathbb{N}_+$. The proof is deferred to the forthcoming article [29].

Proposition 7.2. [29] *Let Π be a GEM(1)-biased permutation of $\mathbb{N}_+$, with the renewal sequence $(u_k;\ k \geq 0)$. Then $(u_k;\ k \geq 0)$ is characterized by any one of the following equivalent conditions:*

(i). The sequence $(u_k;\ k \geq 0)$ is defined recursively by

$$2u_k + 3u_{k-1} + u_{k-2} \stackrel{1}{=} 2\zeta(k) \quad \text{with } u_0 = 1,\ u_1 = 1/2, \tag{7.18}$$

where $\zeta(k) := \sum_{n=1}^{\infty} 1/n^k$ is the Riemann zeta function.

(ii). For all $k \geq 0$,

$$u_k \stackrel{1}{=} (-1)^{k-1}\left(2 - \frac{3}{2^k}\right) + \sum_{j=2}^{k} (-1)^{k-j}\left(2 - \frac{1}{2^{k-j}}\right)\zeta(j). \tag{7.19}$$

(iii). For all $k \geq 0$,

$$u_k \stackrel{1}{=} \sum_{j=1}^{\infty} \frac{2}{j^k(j+1)(j+2)}. \tag{7.20}$$

(iv). The generating function of $(u_k;\ k \geq 0)$ is

$$U(z) := \sum_{k=0}^{\infty} u_k z^k \stackrel{1}{=} \frac{2}{(1+z)(2+z)}\left[1 + \left(2 - \gamma - \Psi(1-z)\right)z\right], \tag{7.21}$$

where $\gamma := \lim_{n\to\infty}(\sum_{k=1}^{n} 1/k - \ln n) \approx 0.577$ is the Euler constant*, and $\Psi(z) := \Gamma'(z)/\Gamma(z)$ with $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}dt$, is the* digamma function.

The distribution of Y_1, that is $f_k := \mathbb{P}(Y_1 = k)$ for all $k \geq 1$, is determined by $(u_k;\ k \geq 0)$ or $U(z)$ via the relations (1.12)-(1.13). It is easy to see that the generating function $F(z)$ of $(f_k;\ k \geq 1)$ is real analytic on $(0, z_0)$ with $z_0 \approx 1.29$. This implies that all moments of Y_1 are finite. By expanding $F(z)$ into power series at $z = 1$, we get:

$$F(z) \stackrel{1}{=} 1 + 3(z-1) + \frac{17}{2}(z-1)^2 + \frac{1}{2}(47 + \pi^2)(z-1)^3 + \cdots, \tag{7.22}$$

which agrees with the simulation (7.1), since $\mathbb{E}Y_1 = F'(1) \stackrel{1}{=} 3$ and $Var(Y_1) = 2F''(1) + F'(1) - F'(1)^2 \stackrel{1}{=} 11$.

References

[1] Hüseyin Acan and Boris Pittel. On the connected components of a random permutation graph with a given number of edges. *J. Combin. Theory Ser. A*, 120(8):1947–1975, 2013.

[2] Jomy Alappattu and Jim Pitman. Coloured loop-erased random walk on the complete graph. *Combin. Probab. Comput.*, 17(6):727–740, 2008.

[3] David Aldous and Persi Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probab. Theory Related Fields*, 103(2):199–213, 1995.

[4] David Aldous, Grégory Miermont, and Jim Pitman. Weak convergence of random p-mappings and the exploration process of inhomogeneous continuum random trees. *Probab. Theory Related Fields*, 133(1):1–17, 2005.

[5] David Aldous and Jim Pitman. Invariance principles for non-uniform random mappings and trees. In *Asymptotic combinatorics with application to mathematical physics (St. Petersburg, 2001)*, volume 77 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 113–147. Kluwer Acad. Publ., Dordrecht, 2002.

[6] Søren Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[7] Roland Bacher and Christophe Reutenauer. Number of right ideals and a q-analogue of indecomposable permutations. *Canad. J. Math.*, 68(3):481–503, 2016.

[8] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[9] Riddhipratim Basu and Nayantara Bhatnagar. Limit theorems for longest monotone subsequences in random Mallows permuations. *arXiv:1601.02003*, 2016.

[10] A. F. Beardon. Sums of powers of integers. *Amer. Math. Monthly*, 103(3):201–213, 1996.

[11] Sergio Benedetto and Guido Montorsi. Unveiling turbo codes: some results on parallel concatenated coding schemes. *IEEE Transactions on Information theory*, 42(2):409–428, 1996.

[12] Volker Betz and Daniel Ueltschi. Spatial random permutations and infinite cycles. *Comm. Math. Phys.*, 285(2):469–501, 2009.

[13] Nayantara Bhatnagar and Ron Peled. Lengths of monotone subsequences in a Mallows permutation. *Probab. Theory Related Fields*, 161(3-4):719–780, 2015.

[14] Marek Biskup and Thomas Richthammer. Gibbs measures on permutations over one-dimensional discrete point sets. *Ann. Appl. Probab.*, 25(2):898–929, 2015.

[15] Tamara Broderick, Michael I. Jordan, and Jim Pitman. Beta processes, stick-breaking and power laws. *Bayesian Anal.*, 7(2):439–475, 2012.

[16] Tamara Broderick, Jim Pitman, and Michael I. Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Anal.*, 8(4):801–836, 2013.

[17] F. T. Bruss and L. C. G. Rogers. Pascal processes and their characterization. *Stochastic Process. Appl.*, 37(2):331–338, 1991.

[18] Louis Comtet. Sur les coefficients de l'inverse de la série formelle $\sum n! t^n$. *C. R. Acad. Sci. Paris Sér. A-B*, 275:A569–A572, 1972.

[19] Louis Comtet. *Advanced combinatorics*. D. Reidel Publishing Co., Dordrecht, enlarged edition, 1974. The art of finite and infinite expansions.

[20] Robert Cori. Hypermaps and indecomposable permutations. *European J. Combin.*, 30(2):540–541, 2009.

[21] Robert Cori. Indecomposable permutations, hypermaps and labeled Dyck paths. *J. Combin. Theory Ser. A*, 116(8):1326–1343, 2009.

[22] Robert Cori, Claire Mathieu, and John Michael Robson. On the number of indecomposable permutations with a given number of cycles. *Electron. J. Combin.*, 19(1):Paper 49, 14, 2012.

[23] Douglas E. Critchlow. *Metric methods for analyzing partially ranked data*, volume 34 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 1985.

[24] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[25] Persi Diaconis, Michael McGrath, and Jim Pitman. Riffle shuffles, cycles, and descents. *Combinatorica*, 15(1):11–29, 1995.

[26] Persi Diaconis and Arun Ram. Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. *Michigan Math. J.*, 48:157–190, 2000.

[27] Dariush Divsalar and Fabrizio Pollara. Turbo codes for PCS applications. In *IEEE International Conference on Communications*, volume 1, pages 54–59. 1995.

[28] Peter Donnelly. The heaps process, libraries, and size-biased permutations. *J. Appl. Probab.*, 28(2):321–335, 1991.

[29] Jean-Jil Duchamps, Jim Pitman, and Wenpin Tang. Renewal sequences derived from the Riemann zeta function. In preparation.

[30] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[31] P. Erdös, A. Rényi, and P. Szüsz. On Engel's and Sylvester's series. *Ann. Univ. Sci. Budapest. Eötvös. Sect. Math.*, 1:7–32, 1958.

[32] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3:87–112, 1972.

[33] Paul D. Feigin. On the characterization of point processes with the order statistic property. *J. Appl. Probab.*, 16(2):297–304, 1979.

[34] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[35] Karl-Heinz Fichtner. Random permutations of countable sets. *Probab. Theory Related Fields*, 89(1):35–60, 1991.

[36] R. A. Fisher. *The Design of Experiments.* Oliver and Boyd, Edinburgh, 1935.

[37] Bert Fristedt. Intersections and limits of regenerative sets. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 121–151. Springer, New York, 1996.

[38] Alexey Gladkich and Ron Peled. On the cycle structure of Mallows permutations. *arXiv:1601.06991*, 2016.

[39] Alexander Gnedin. Coherent random permutations with biased record statistics. *Discrete Math.*, 311(1):80–91, 2011.

[40] Alexander Gnedin and Vadim Gorin. Record-dependent measures on the symmetric groups. *Random Structures Algorithms*, 46(4):688–706, 2015.

[41] Alexander Gnedin and Vadim Gorin. Spherically symmetric random permutations. *arXiv:1611.01860*, 2016.

[42] Alexander Gnedin, Chris Haulk, and Jim Pitman. Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In *Probability and mathematical genetics*, volume 378 of *London Math. Soc. Lecture Note Ser.*, pages 264–298. Cambridge Univ. Press, Cambridge, 2010.

[43] Alexander Gnedin, Alex Iksanov, and Uwe Roesler. Small parts in the Bernoulli sieve. In *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 235–242. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[44] Alexander Gnedin, Alexander Iksanov, and Alexander Marynych. The Bernoulli sieve: an overview. In *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, Discrete Math. Theor. Comput. Sci. Proc., AM, pages 329–341. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010.

[45] Alexander Gnedin and Grigori Olshanski. Coherent permutations with descent statistic and the boundary problem for the graph of zigzag diagrams. *Int. Math. Res. Not.*, pages Art. ID 51968, 39, 2006.

[46] Alexander Gnedin and Grigori Olshanski. q-exchangeability via quasi-invariance. *Ann. Probab.*, 38(6):2103–2135, 2010.

[47] Alexander Gnedin and Grigori Olshanski. The two-sided infinite extension of the Mallows model for random permutations. *Adv. in Appl. Math.*, 48(5):615–639, 2012.

[48] Alexander Gnedin and Jim Pitman. Regenerative composition structures. *Ann. Probab.*, 33(2):445–479, 2005.

[49] Alexander V. Gnedin. The Bernoulli sieve. *Bernoulli*, 10(1):79–96, 2004.

[50] Alexander V. Gnedin. Regeneration in random combinatorial structures. *Probab. Surv.*, 7:105–156, 2010.

[51] Louis Gordon. Successive sampling in large finite populations. *Ann. Statist.*, 11(2):702–706, 1983.

[52] J. M. Hammersley. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I: Theory of statistics*, pages 345–394, 1972.

[53] Ahmed Helmi, Jérémie Lumbroso, Conrado Martínez, and Alfredo Viola. Data streams as random permutations: the distinct element problem. In *23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms*, pages 323–338. 2012.

[54] Fred M. Hoppe. Size-biased filtering of Poisson-Dirichlet samples with an application to partition structures in genetics. *J. Appl. Probab.*, 23(4):1008–1012, 1986.

[55] Roger A. Horn. On moment sequences and renewal sequences. *J. Math. Anal. Appl.*, 31:130–135, 1970.

[56] Z. Ignatov. Point processes generated by order statistics and their applications. In *Point processes and queuing problems (Colloq., Keszthely, 1978)*, volume 24 of *Colloq. Math. Soc. János Bolyai*, pages 109–116. North-Holland, Amsterdam-New York, 1981.

[57] Alexander Iksanov. On the number of empty boxes in the Bernoulli sieve II. *Stochastic Process. Appl.*, 122(7):2701–2729, 2012.

[58] Alexander Iksanov. On the number of empty boxes in the Bernoulli sieve I. *Stochastics*, 85(6):946–959, 2013.

[59] Th. Kaluza. Über die Koeffizienten reziproker Potenzreihen. *Math. Z.*, 28(1):161–170, 1928.

[60] David G. Kendall. Branching processes since 1873. *J. London Math. Soc.*, 41:385–406. (1 plate), 1966.

[61] David G. Kendall. Renewal sequences and their arithmetic. In *Symposium on Probability Methods in Analysis (Loutraki, 1966)*, pages 147–175. Springer, Berlin, 1967.

[62] S. V. Kerov. Subordinators and the actions of permutations with quasi-invariant measure. *Journal of Mathematical Sciences*, 87(6):4094–4117, 1997.

[63] S. V. Kerov and N. V. Tsilevich. Stick breaking process generated by virtual permutations with Ewens distribution. *Journal of Mathematical Sciences*, 87(6):4082–4093, 1997.

[64] S. V. Kerov and A. M. Vershik. Asymptotics of the plancherel measure of the symmetric group and the limiting form of young tableaux. In *Soviet Math. Dokl*, volume 18, pages 527–531, 1977.

[65] Sergei Kerov, Grigori Olshanski, and Anatoly Vershik. Harmonic analysis on the infinite symmetric group. *Invent. Math.*, 158(3):551–642, 2004.

[66] Serguei Kerov, Grigori Olshanski, and Anatoli Vershik. Harmonic analysis on the infinite symmetric group. A deformation of the regular representation. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(8):773–778, 1993.
[67] Andrew King. Generating indecomposable permutations. *Discrete Math.*, 306(5):508–518, 2006.
[68] J. F. C. Kingman. *Regenerative phenomena.* John Wiley & Sons Ltd., London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.
[69] Donald E. Knuth. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms.* Addison-Wesley Longman Publishing Co., Inc. Third Edition, 1997.
[70] Donald E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and Searching.* Addison Wesley Longman Publishing Co., Inc. Second Edition, 1998.
[71] Steven P. Lalley. Cycle structure of riffle shuffles. *Ann. Probab.*, 24(1):49–73, 1996.
[72] André Lentin. *Équations dans les monoïdes libres.* Mouton; Gauthier-Villars, Paris, 1972. Mathématiques et Sciences de l'Homme, No. 16.
[73] Thomas M. Liggett. Total positivity and renewal theory. In *Probability, statistics, and mathematics*, pages 141–162. Academic Press, Boston, MA, 1989.
[74] B. F. Logan and L. A. Shepp. A variational problem for random Young tableaux. *Advances in Math.*, 26(2):206–222, 1977.
[75] C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44:114–130, 1957.
[76] Luis A. Medina, Victor H. Moll, and Eric S. Rowland. Iterated primitives of logarithmic powers. *Int. J. Number Theory*, 7(3):623–634, 2011.
[77] Carl Mueller and Shannon Starr. The length of the longest increasing subsequence of a random Mallows permutation. *J. Theoret. Probab.*, 26(2):514–540, 2013.
[78] S. Muthukrishnan. Data streams: algorithms and applications. *Found. Trends Theor. Comput. Sci.*, 1(2):117–236, 2005.
[79] Mihael Perman, Jim Pitman, and Marc Yor. Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields*, 92(1):21–39, 1992.
[80] E. J. G. Pitman. Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):119–130, 1937.
[81] Jim Pitman. The bar raising model for records and partially exchangeable partitions. In preparation.
[82] Jim Pitman. Extremes and gaps in sampling from a residual allocation model. In preparation.
[83] Jim Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, 102(2):145–158, 1995.
[84] Jim Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006.
[85] Jim Pitman and Ngoc M. Tran. Size-biased permutation of a finite sequence with independent and identically distributed terms. *Bernoulli*, 21(4):2484–2512, 2015.
[86] Jim Pitman and Yuri Yakubovich. Extremes and gaps in sampling from a GEM random discrete distribution. *arXiv:1701.06294*, 2017.
[87] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.
[88] A. Rényi. A new approach to the theory of Engel's series. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 5:25–32, 1962.
[89] Andrew M. Rockett. Sums of the inverses of binomial coefficients. *Fibonacci Quart.*, 19(5):433–437, 1981.
[90] Stanley Sawyer and Daniel Hartl. A sampling theory for local selection. *Journal of Genetics*, 64(1):21–29, 1985.
[91] Timo Seppäläinen. A microscopic model for the Burgers equation and longest increasing subsequences. *Electron. J. Probab.*, 1:no. 5, approx. 51 pp. 1996.
[92] D. N. Shanbhag. On renewal sequences. *Bull. London Math. Soc.*, 9(1):79–80, 1977.
[93] L. A. Shepp and S. P. Lloyd. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.*, 121:340–357, 1966.
[94] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences, A003149. Available at https://www.oeis.org/A003149.
[95] A. J. Stam. Regeneration points in random permutations. *Fibonacci Quart.*, 23(1):49–56, 1985.

[96] Richard P. Stanley. The descent set and connectivity set of a permutation. *J. Integer Seq.*, 8(3):Article 05.3.8, 9, 2005.
[97] B. Sury. Sum of the reciprocals of the binomial coefficients. *European J. Combin.*, 14(4):351–353, 1993.
[98] Hermann Thorisson. On time- and cycle-stationarity. *Stochastic Process. Appl.*, 55(2):183–209, 1995.
[99] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.
[100] Daniel Ueltschi. The model of interlacing spatial permutations and its relation to the Bose gas. In *Mathematical results in quantum mechanics*, pages 255–272. World Sci. Publ., Hackensack, NJ, 2008.
[101] A. M. Vershik and A. A. Shmidt. Limit measures arising in the asymptotic theory of symmetric groups, I. *Theory of Probability & Its Applications*, 22(1):70–85, 1977.
[102] A. M. Vershik and A. A. Shmidt. Limit measures arising in the asymptotic theory of symmetric groups, II. *Theory of Probability & Its Applications*, 23(1):36–49, 1978.

Statistics department, University of California, Berkeley. Email:
E-mail address: `pitman@stat.berkeley.edu`

Statistics department, University of California, Berkeley. Email:
E-mail address: `wenpintang@stat.berkeley.edu`