CYCLES AND DESCENTS OF RANDOM PERMUTATIONS

BY

PERSI DIACONIS, MICHAEL MCGRATH, and JIM PITMAN

TECHNICAL REPORT NO. 387

OCTOBER 1991

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Cycles and Descents of Random Permutations

By

Persi Diaconis, Michael McGrath, and Jim Pitman

Technical Report No. 387

October 1991

Prepared Under the Auspices

Of

National Science Foundation Grant DMS89–05874

Department of Statistics
Stanford University
Stanford, California
Cycles and Descents of Random Permutations

Persi Diaconis\(^1\), Michael McGrath \(^2\), and Jim Pitman \(^3\) *

Preliminary Draft
October 1, 1991

Abstract

Formulae for the joint distribution of the cycle structure and number of descents of a random permutation are derived from simpler formulae for the distribution of the cycle structure of certain random riffle shuffles with at most \(a - 1\) descents. The results for the cycle structure of riffle shuffles assume a product form parallel to classical results for uniform random permutations, and involve the number of aperiodic circular words of \(a\) letters (or necklaces of \(a\) colors) of length \(n\).


Contents

1 Introduction. 2
2 Preliminaries concerning $a$-shuffles. 5
3 Bijections. 7
4 Derangements with one descent. 17
5 Word lengths in the Lyndon decomposition of a random word. 19
6 Distribution of the number of $i$-cycles. 20
7 Generating Functions. 23

1 Introduction.

According to a formula of Cauchy, for non-negative integers $n_j$ with $\sum jn_j = n$, the number of permutations $\pi \in S_n$ such that $\pi$ has $n_j$ cycles of length $j$ is

$$n! \prod_{j=1}^{n} [j^{n_j} n_j!]^{-1}$$

(1)

Let $D(\pi) = \# \{ i : \pi(i+1) > \pi(i) \}$, the number of descents of $\pi$. The number of $\pi \in S_n$ with a given number of descents is also known (see e.g. Foata (19??) ...): for $1 \leq k \leq n$

$$\# \{ \pi \in S_n : D(\pi) = k - 1 \} = A_{n,k}$$

where the $A_{n,k}$ are the Eulerian numbers

$$A_{n,k} = \sum_{a=1}^{k} (-1)^{k-a} \binom{n+1}{k-1} a^n.$$  

(2)

A central result of the present paper is the following formula for the number of permutations $\pi$ with prescribed cycle counts and a prescribed number of descents.
Proposition 1 For non-negative \( n_j \) with \( \sum j n_j = n \), and \( 1 \leq k \leq n \), the number of permutation \( \pi \in S_n \) such that \( \pi \) has \( n_j \) cycles of length \( j \), \( 1 \leq j \leq n \), and \( k-1 \) descents, is

\[
\sum_{a=1}^{k} (-1)^{k-a} \left( \begin{array}{c} n + 1 \\ k - a \end{array} \right) \prod_{j=1}^{n} \left( \begin{array}{c} f_{ja} \\ n_j \end{array} \right),
\]

where \( f_{ja} \) is the number of aperiodic circular words of length \( j \) for an alphabet of \( a \) letters, and

\[
\left( \begin{array}{c} f \\ m \end{array} \right) = \left( \begin{array}{c} f + m - 1 \\ m \end{array} \right)
\]

is the number of sequences of \( f \) non-negative integers with sum \( m \).

By a standard application of Mobius inversion (see e.g. Berge (1971, p.87)),

\[
f_{ja} = \frac{1}{j} \sum_{d|j} \mu(d) a^{j/d},
\]

where the sum is over all divisors \( d \) of \( j \), and \( \mu \) is the Mobius function

\[
\begin{align*}
\mu(1) &= 1 \\
\mu(d) &= (-1)^k \text{ if } d = p_1 p_2 \ldots p_k \text{ for distinct primes } p_1, \ldots, p_k \\
&= 0 \text{ else .}
\end{align*}
\]

We understand Proposition 1 as a combination of two results. The first, stated as Proposition 7 below expresses the uniform counting distribution on \( \{ \pi \in S_n : D(\pi) = k-1 \} \) as a linear combination, with weights

\[
(-1)^{k-a} \left( \begin{array}{c} n + 1 \\ k - a \end{array} \right) a^n, 1 \leq a \leq k,
\]

as appearing in (2), of certain non-uniform probability distributions \( P_{n,a} \) on \( S_n \). Here \( P_{n,a} \) is the distribution on \( S_n \) of a random \( a \)-shuffle, as defined in the next paragraph following Bayer-Diaconis (1990). The second ingredient in Proposition 1 is the simple description of the distribution of the cycle counts for a random \( a \)-shuffle stated as Proposition 3 below. Proposition 1 follows immediately from Proposition 7 and 3.
Roughly speaking, a random \(a\)-shuffle of \(n\) cards is obtained by first cutting the \(n\) cards at random into \(a\) packets of random sizes, then randomly riffling the packets together. To be more precise, let \(A = \{0, 1, \ldots, a - 1\}\), and for \(x \in A^n\) let \(x^\dagger\) be the non-decreasing rearrangement of \(x\). Let \(\pi_x \in S_n\) be the unique permutation that for each \(b \in A\) is the increasing map from

\[
\{ i : x_i^\dagger = b \} \rightarrow \{ j : x_j = b \}.
\]

Note that \(\pi_x\) has at most \(a - 1\) descents.

**Example 2** Example goes here.

The card shuffling interpretation is that \(\{ i : x_i^\dagger = b \}\) is the set of places in the deck which form the \(b\)th packet, and \(\{ j : x_j = b \}\) is the set of places occupied by cards in the \(b\)th packet after the \(a\) packets are riffled together. (Note that some packets may be empty). Then \(\pi_x(i)\) is the place after the shuffle of the card initially in place \(i\). And \(P_{n,a}\) is the distribution of an \(a\)-shuffle, that is to say the distribution of \(\pi_x\) when \(x\) is picked uniformly at random from \(A^n\).

**Proposition 3** For non-negative \(n_j\) with \(\sum jn_j = n\), and \(a = 1, 2, \ldots\), the probability that a random \(a\)-shuffle of \(n\) cards produces a permutation with \(n_j\) cycles of length \(j\), \(1 \leq j \leq n\), is

\[
a^{-n} \prod_{j=1}^{n} \frac{f_{j}^{n_j}}{n_j}
\]  

(6)

As shown in Section 3, Proposition 3 follows easily from properties of a bijection discovered by Gessel (refs ???) between \(A^n\) and a the collection of multisets of aperiodic necklaces with total length \(n\). While Gessel’s result has been part of the folklore of combinatorics for several years, it seems that no proof has yet been published. So we include an elementary proof in Section 3.

For fixed \(n\), as \(a \rightarrow \infty\), the distribution of an \(a\)-shuffle approaches the uniform distribution on \(S_n\). Consequently, for fixed \(n\) as \(a \rightarrow \infty\), the expression (6) converges to the corresponding probability for a uniformly distributed \(\pi\), found by dividing Cauchy’s expression in (1) by \(n!\). This is easily checked directly, using the relation

\[
f_{ja} \sim j^{-1} a^j \quad \text{for each } j \quad \text{as } a \rightarrow \infty
\]  

(7)
In the special case $n_n = 1$, $n_j = 0$ else, Proposition 3 shows that the probability that an $a$-shuffle produces an $n$-cycle is $f_n / a^n$, while Proposition 1 implies the probability that a uniform $\pi$ consists of a single $n$-cycle with $k - 1$ descents is

$$\frac{1}{n!} \sum_{a=1}^{k} (-1)^{k-a} \left( \begin{array}{c} n + 1 \\ k - 1 \end{array} \right) f_{n-a}. \tag{8}$$

Replacing $n!$ by $(n - 1)!$ gives the conditional probability that $\pi$ has $(k - 1)$ descents given that $\pi$ is an $n$ cycle.

(Prosumably the asymptotic distribution for the number of descents is the same conditionally on an $n$-cycle as unconditionally. Can we estimate the mean and variance?).

In the special case $a = 2$, Proposition 1 shows that the number of permutations $\pi \in S_n$ such that $\pi$ has exactly 1 descent, and $n_i$ cycles of size $i$, is

$$\prod_{i=1}^{n} \left[ f_{i2} \right] \left[ n_i \right] \tag{9}$$

except in case $n_1 = n$, $n_i = 0$ else, when there is obviously no such $\pi$.

2 Preliminaries concerning $a$-shuffles.

Let $x \in A^n$. Let $n_k = \#\{j: x_j = k\}$, so $\sum_{k=0}^{n-1} n_k = n$. Then we may understand $\pi_x$ as follows: Take a deck of cards with

$n_0$ cards labelled 0 on top

$n_1$ cards labelled 1 below the 0’s

\ldots

\ldots

$n_{a-1}$ cards labelled $a - 1$ on bottom.

The map $j \rightarrow \pi_x(j)$ gives the position (counting down from the top) of the card initially $j$ from the top for a certain riffling of the $a$ packets of sizes $n_0, \ldots, n_{a-1}$. To be precise:

Definition 4 For $x \in A^n$, $n_0, \ldots, n_{a-1}$ defined in terms of $x$ as above, the permutation $\pi_x \in S_n$ is the unique permutation $\pi$ such that for each $b = 0, \ldots, a - 1$ with $n_b > 0$

$$\pi(n_0 + \ldots + n_{b-1}) < \ldots < \pi(n_0 + \ldots + n_b),$$

\ldots
and $x_{\pi(j)} = b$ for $n_0 + \ldots + n_{b-1} < j \leq n_0 + \ldots + n_b$.

Notice that if $x^\dagger$ denotes the non-decreasing rearrangement of $x$, so
\[ x^\dagger_j = b \text{ for } n_0 + \ldots + n_{b-1} < j \leq n_0 + \ldots + n_b, \]
then
\[ x^\dagger = x \circ \pi_x \] (10)

hence
\[ x = x^\dagger \circ \pi_x^{-1}. \] (11)

**Proposition 5** The map
\[ x \rightarrow (\pi_x, x^\dagger) \]
from $A^n$ to $S_n \times A^n$ is injective. The range of this map is the set
\[ \{(\pi, y) \in S_n \times A^n : y \text{ is non-decreasing, and } Desc(\pi) \subseteq Asc(y)\} \]
where
\[ Desc(\pi) = \{i : \pi(i) > \pi(i + 1)\} \]
is the descent set of $\pi$, and
\[ Asc(y) = \{i : y(i) < y(i + 1)\} \]
is the ascent set of $y$.

**Proof.** The first sentence is immediate from (11). As for the second, notice that by definition of $\pi_x$,
\[ \text{if } x^\dagger_i = x^\dagger_{i+1} \text{ then } \pi_x(i) < \pi_x(i + 1). \]

That is to say
\[ \text{if } \pi_x(i) > \pi_x(i + 1) \text{ then } x^\dagger_i < x^\dagger_{i+1}, \text{ or} \]
\[ Descents(\pi_x) \subseteq Ascents(x^\dagger). \]
If $\pi \in S^n$ and $y \in A^n$ is non-decreasing, then the sequence $x$ defined by
\[ x_{\pi(j)} = y_j \]
is clearly such that $\pi_x = \pi$ and $x^\dagger = y$. 

If $\pi$ has $d$ descents, then the number of non-decreasing $y \in A^n$ such that $Desc(\pi) \subseteq Asc(y)$ is $\begin{bmatrix} a - d \\ n \end{bmatrix}$ by a standard argument. Thus the above proposition implies
Proposition 6 (Bayer-Diaconis.) The range of the map \( x \rightarrow \pi_x \) from \( A^n \rightarrow S_n \) is the set of all permutations in \( S^n \) with at most \( a - 1 \) descents. If \( \pi \) has \( d \) descents then \( \pi = \pi_x \) for exactly \( \binom{a - d}{n} \) distinct sequences \( x \) in \( A^n \).

Proposition 7 For \( n = 1,2,\ldots, k = 1,\ldots,n \), let \( \#_{n,k} \) be the measure on \( S_n \) defined by restriction of counting measure to the set \( \{ \pi \in S_n : \pi \text{ has } k - 1 \text{ descents} \} \). For \( a = 1,2,\ldots \) let \( M_{n,a} \) be the distribution on \( S_n \) of \( \pi_x \) when \( x \) is given counting distribution on \( \{0,1,\ldots,a - 1\}^n \). Then

\[
M_{n,a} = \sum_{k=1}^{a} \binom{n + a - k}{n} \#_{n,k} \tag{12}
\]

\[
\#_{n,k} = \sum_{a=1}^{k} (-1)^{k-a} \binom{n + 1}{k-a} M_{n,a}. \tag{13}
\]

Proof. The first formula is a restatement of the preceding proposition. The second follows easily from the first by inversion.

Note. When evaluated on the whole set \( S_n \), these identities for counting measures reduce to classical identities relating linear combinations of the Eulerian numbers

\[ A_{n,k} = \#_{n,k}(S_n) \]

to linear combinations of powers

\[ a^n = M_{n,a}(S_n). \]

In particular (13) is an extension of (2).

Let \( P_{n,a} = M_{n,a}/a^n \) denote the distribution on \( S_n \) of a random \( a \)-shuffle, \( P_n = \lim_{a \to \infty} P_{n,a} \) the uniform probability distribution on \( S_n \). Formula (13) shows that from knowledge of the \( P_{n,a} \) distribution of some function \( Y \) of \( \pi \) for all \( a = 1,\ldots,k \) we obtain the \( P_n \) distribution of \( Y(\pi) \) jointly with the number of descents \( D(\pi) \). Thus, for \( k = 1,\ldots,n \)

\[
P_n(\cdot \cap D = k - 1) = \frac{1}{n!} \sum_{a=1}^{k} (-1)^{k-a} \binom{n + 1}{k-a} a^n P_{n,a}(\cdot).
\]

As a case in point, Proposition 3 implies Proposition 1.
3 Bijectons.

Fix $a \geq 2$. Let $A = \{0, \ldots, a - 1\}$. To simplify terminology, we regard $A$ as a totally ordered alphabet of $a$ letters (rather than numbers). Letters in $A$ are used to label the packets for an $a$-shuffle. More precisely, the letters describe the places in the deck defined by the packets for an $a$-shuffle. The places of cards in the first packet are labelled 0, places of cards in the second packet are labelled 1, and so on. Note that some packets may be empty. Regard $x \in A^n$ as a word of length $n$. For a word $x$ with non-decreasing rearrangement $x^1$, the letter for the packet containing place $k$ in the deck is by definition $x^1(k)$. The shuffle $\pi_x$ induced by the word $x$ cuts the cards into packets in accordance with $x^1$, then riffles them into new places dictated by $x$. To be precise, if card $i$ is initially the $m$th card in packet $b$, then $\pi_x(i)$, the place of card $i$ after the shuffle, is the place of the $m$th letter $b$ in word $x$. Now write $\pi_x$ in cycle notation in the standard way, say

$$\pi_x = (C_{x1})(C_{x2}) \cdots$$

where for example

$$C_{x1} = (1, \pi_x(1), \pi_x^2(1), \cdots)$$

is the sequence of places in the deck defined by the orbit of card 1, $C_{x2}$ is the orbit of the first card in the deck not in $C_{x1}$, and so on.

Definition 8 The cycle sentence of $x$, is the sequence of words

$$s_x = (w_{x1})(w_{x2}) \cdots$$

obtained from the cycle notation of $\pi_x$ by replacing each $k$ by its packet letter $x^1(k)$, wherever $k$ appears in the cycle notation, $1 \leq k \leq n$. The cycle words of $x$, are the words in the cycle sentence of $x$.

By definition, the cycle sentence of $x$ is a sentence of $n$ letters. And the lengths of the cycle words $w_{zi}$ are identical to the lengths of corresponding cycles $C_{zi}$ of $\pi_x$.

Example 9 Let $a = 2$, $A = \{0, 1\}$, $n = 14$. Let $x = (x(i), 1 \leq i \leq n)$ be the word defined by the second row of the following table. Then $x^1(i)$ and
\( \pi_x(i) \) are as in the third and fourth rows. The \( \pi_x(i) \) are obtained by listing the places of the 0's in word \( x \), followed by the places of the 1's.

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x(i) )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x^1(i) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \pi_x(i) )</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

In cycle notation:

\[ \pi_x = (1)(2, 5, 10)(3, 6, 11)(4, 8, 13, 7, 12)(9, 14). \]

Replacing \( k \) by \( x^1(k) \) gives the cycle sentence of \( x \):

\[ s_x = (0) (0, 0, 1) (0, 0, 1) (0, 0, 1, 0, 1) (0, 1). \]

Note how the word \( (001) \) repeated twice in the cycle sentence corresponds to the fact that there is a pair of cards 2 and 3, each part of a 3 cycle, which move under \( \pi_x \) as if glued to each other. The shuffle \( \pi_x \) moves these cards around in the deck, but they never are separated by the shuffle, no matter how many times the shuffle is repeated. As a general rule, a word repeated \( m \) times in the cycle sentence corresponds to a clump of \( m \) consecutive cards left unseparated by any number of repetitions of the shuffle. This follows easily from Lemma 17 (ii) below.

Call a finite sequence of words, comprising a total of \( n \) letters, a sentence of \( n \) letters. A word \( w \) of length \( j \) is a Lyndon word iff \( w \) is lexicographically strictly smaller than each of the \( j - 1 \) cyclic shifts of \( w \). A total order on the collection of Lyndon words of any finite length, called here the repeat lexicographic order, is defined as follows: \( w < z \) iff \( w^* \) is lexicographically smaller than \( z^* \), where \( y^* \) is the infinite word obtained by indefinite repition of \( y \). The following proposition and its corollary present variations of a bijection due to Gessel. Refer to Gessel-Reutenauer. More credits ???

**Proposition 10** The map from \( x \) to \( s_x \), the cycle sentence of \( x \), defines a bijection between the set \( A^n \) of words \( x \) of length \( n \) and the set of all \( n \) letter sentences \( s \) such that

i) each word in the sentence is a Lyndon word
ii) the words in the sentence are non-decreasing in the repeat lexicographical order.

Proof of this proposition is given later in this section. We present first an immediate corollary, and the derivation of Proposition 3 from this corollary.

Clearly, a sentence $s$ of non-decreasing Lyndon words is uniquely determined by its word distribution, that is the numbers of times $n(w)$ that each Lyndon word $w$ appears in $s$. Thus immediately for Proposition 10 we obtain:

**Corollary 11** Let $L_j$ be the set of Lyndon words of length $j$ for the alphabet $A$. For $x \in A^n$ let $n_x(w)$ be the number of times the Lyndon word $w$ appears in the cycle sentence of $x$. The map $x \to (n_x(w), w \in \bigcup_j L_j)$ defines a bijection between $A^n$ and all arrays of non-negative integers

$$(n(w), w \in \bigcup_j L_j : \sum_j \sum_{w \in L_j} n(w) = n).$$

This bijection is such that the number of cycles of length $j$ in $\pi_x$ is

$$N_j(\pi_x) = \sum_{w \in L_j} n_x(w).$$

**Proof of Proposition 3.** There is an obvious one to one correspondence between Lyndon words and aperiodic circular words. So $#(L_j) = f_{ja}$ as in Proposition 3. According to Corollary 11, for $n_j$ with $\sum_j jn_j = n$, the number of $x \in A^n$ such that $N_j(\pi_x) = n_j$ equals the number of arrays of non-negative integers

$$(n(w), w \in \bigcup_j L_j : \sum_{w \in L_j} n(w) = n_j).$$

Since for each $j$ there are \( \binom{f_{ja}}{n_j} \) different ways to choose the $n(w)$,\( w \in L_j \), there are $\prod_{j=1}^n \binom{f_{ja}}{n_j}$ different ways to choose the entire array. □

We now develop some preliminaries for the proof of Proposition 10. The argument falls into two parts. The first part is to show that the cycle sentence of $x$ is composed of a non-decreasing sequence of Lyndon words. The second is to show that every such sequence comes from a unique $x$. Consider the permutations $\pi_x^m$ obtained by $m$ iterations of $\pi_x$ for a fixed $x$. So $\pi_x^m(k)$ is the place of card $k$ in the deck after a sequence of $m$ shuffles according to $\pi_x$. 

10
**Definition 12** For each $k = 1, \ldots, n$, the $x$-signature of card $k$ is the infinite word $u_{xk}$ whose letters indicate the successive packets containing card $k$ as it moves through the deck under repetitions of $\pi_x$. Formally, $u_{xk}$ is the infinite word whose $m$th letter is

$$u_{xkm} = x^1 \circ \pi_x^{m-1}(k), \ m = 1, 2, \ldots$$

To illustrate, for the 3 packet riffle $\pi_x$ induced by $x \in \{0, 1, 2\}^n$,

$$u_{xk} = 012112\cdots$$

means that under repeated riffle shuffles according to $\pi_x$, card $k$ starts in the top packet 0, after the first shuffle appears in packet 1 for the second shuffle, then in packet 2 for the third shuffle, and so on. Since $\pi_x^m$ is the identity for some $m$, the word $u_{xk}$ is obviously periodic. That is to say, we have a map

$$u_x : \{1, \ldots, n\} \rightarrow A^*$$

where $A^*$ is the set of all periodic infinite words with letters in the alphabet $A$. It is easy to see that the map $u_x$ can be recovered from the cycle sentence of $x$. It is also true, though less obvious, that the cycle sentence of $x$ can be recovered from $u_x$. This follows from Lemma 13 below. The argument eventually shows that even $x$ can be recovered from $u_x$.

To visualize $u_x$ it is helpful to identify $A^*$ as a subset of $[0, 1]$ via the usual expansion in base $a$. Give $A^*$ the lexicographical order, corresponding to the usual order on $[0, 1]$. Let $\theta$ denote the shift map on $A^*$:

$$\theta(a_1, a_2, \ldots) = (a_2, a_3, \ldots).$$

Note that $\theta$ is invertible on $A^*$, so $\theta^m : A^* \rightarrow A^*$ is defined for every integer $m$.

The upshot of the following lemma is that so far as anything to do with the cycle structure of $\pi_x$ is concerned, the action of the shift $\theta$ on the card signatures $u_{xk}$ is a faithful representation of the action of $\pi_x$ on the cards $k$.

**Lemma 13** For each $x$ in $A^n$ the map $u_x : \{1, \ldots, n\} \rightarrow A^*$ is such that

(i) $u_{x1} \leq u_{x2} \leq \ldots \leq u_{xn}$.

(ii) If $u_{xi} = u_{xk}$ for some $i < k$, then for every $i \leq j \leq k$ and every integer $m$,

$$\pi_x^m(j) - \pi_x^m(i) = j - i.$$
(iii) For all integers $m$

$$\theta^m \circ u_x = u_x \circ \pi_x^m$$

(iv) For each $k \in \{1, \ldots, n\}$ the period of $k$ under the action of $\pi_x$ (i.e. the length of the $\pi_x$-cycle containing $k$) is identical to the period of $u_{x \cdot k}$ under the action of $\theta$.

**Remark.** In terms of card shuffling, (ii) means that when the shuffle $\pi_x$ is iterated, cards $i$ to $k$ inclusive move like a clump of $k - i + 1$ cards glued together. No matter how often the shuffle is repeated, cards in this clump never become separated, though the whole clump will typically move around between packets in the deck.

**Proof of (i).** Let $i < k$. Follow cards $i$ and $k$ under iterates of the shuffle $\pi_x$.

If $x^1(i) < x^1(k)$, then obviously $u_{xi} < u_{xk}$, since $u_{xj1} = x^1(j)$. So suppose $x^1(i) = x^1(k)$, meaning $i$ and $k$ start in the same packet. Then $u_{x1} = u_{xk1}$.

Now the key observation is that because $\pi_x$ is increasing when restricted to each packet,

$$\text{if } \pi_x^{m-1}(i) < \pi_x^{m-1}(k) \text{ and } u_{xim} = u_{xkm} \text{ then } \pi_x^m(i) < \pi_x^m(k).$$

Consequently, if

$$u_{xim} = u_{xkm} \text{ for } m = 1, \ldots M$$

meaning that cards $i$ and $k$ are in the same packet for each of the first $M$-shuffles (though which packet may vary as the shuffles proceed), then also

$$\pi_x^M(i) < \pi_x^M(k),$$

hence

$$u_{xi,M+1} = x^1 \circ \pi_x^M(i) \leq x^1 \circ \pi_x^M(k) = u_{xk,M+1}.$$

Consequently, if $M + 1$ is the first $m$ such that which $u_{xim} \neq u_{xkm}$, then $u_{xi,M+1} < u_{xk,M+1}$. That is to say $u_{xi} < u_{xk}$ in lexicographical order. The only other possibility is that no such $M + 1$ exists. That is to say, $u_{xi} = u_{xk}$.

**Proof of (ii).** If $i < k$ and $u_{xi} = u_{xk}$, then by similar reasoning to that above, the number of cards between cards $i$ and $k$ can never decrease as the shuffles proceed. Since $\pi_x^m$ is the identity for some $m$, this number of cards must remain constant. ■

12
Proof of (iii). This just says that if $\pi_x^m$ maps $j$ to $k$, then $\theta^m$ maps $u_{xj}$ to $u_{xk}$. This follows at once from the definitions, first for $m = 1$, then for any integer $m$.

Proof of (iv). Fix $k$ and suppose that $u_{xk}$ has period $d$ under the action of $\theta$. Then the $\theta^m(u_{xk}), 0 \leq m < d$ are distinct, and $\theta^d(u_{xk}) = u_{xk}$. Due to (iii), the $\pi_x^m(k), 0 \leq m < d$ are distinct. To see that $k$ has period $d$ under $\pi_x$ it only remains to show that $\pi_x^d(k) = k$. To this end let

$$B = u_{x}^{-1}(u_{xk}),$$

the set of all $j \in \{1, \ldots, n\}$ with the same $x$-signature as $k$. By (iii) for $m = d$, $B$ is $\pi_x^d$-invariant. Now (ii) shows $\pi_x^d$ must act as the identity on $B$, hence $\pi_x^d(k) = k$.

Proof of Proposition 10, Part I: that the cycle sentence of $x$ is a non-decreasing sequence of Lyndon words.

Suppose $w$ is a word of length $j$ in the cycle sentence. Say the first place in the corresponding cycle is $k$. Then $k$ has period $j$ under $\pi_x$. By definition of $w$, $u_{xk} = w^*$, that is $w$ repeated indefinitely. And by (iii) above the signatures $u_{xi}$ of the $j - 1$ other cards $i$ in the $\pi_x$ cycle starting from $k$ are the shifts $\theta^m(w^*)$ for $m = 1, \ldots, j - 1$. By (iv) above, these signatures are all distinct. Also, by definition of the cycle sentence, $k$ is the least element in a $\pi_x$-cycle of length $j$. Consequently, by (i) $u_{xk} < u_{xi}$ for every other card $i$ in the cycle of length $j$ containing $k$. That is to say:

$$w^* < \theta^m(w^*) \text{ for } m = 1, \ldots, j - 1.$$ 

Thus $w$ is a Lyndon word. That these words are non-decreasing follows immediately from (i) and the fact that each cycle is defined to start at the least index not in any previous cycles. 

Proof of Proposition 10, part II: that every $n$-letter sentence composed of a non-decreasing sequence of Lyndon words is the cycle-sentence of $x$ for a unique $x$.

Given a non-decreasing sequence of Lyndon words, say

$$w_1 \leq w_2 \leq \ldots, \tag{16}$$

with lengths say $j_1, j_2, \ldots$, with $\sum j_i = n$, let

$$u_1 \leq u_2 \leq \ldots \leq u_n \tag{17}$$

13
be the non-decreasing sequence of \( n \) infinite words obtained by putting the
array of \( n \) infinite words

\[
(\theta^m(w_i), 0 \leq m \leq j_i - 1, i = 1, 2, \ldots)
\]  

(18)
in non-decreasing order. Since \( \theta \) acts as a permutation on \( A^* \), it makes sense
to define a word \( x \) by

\[
x = (x_1, \ldots, x_n) \text{ where } x_k \text{ is the first letter of } \theta^{-1}(u_k).
\]  

(19)

(Note that if \( u_k = \theta^m(w_i^j) \) where \( w_i \) has length \( j \) say, then \( u_k \) has period \( j \),
so the first letter of \( \theta^{-1}(u_k) \) is the \( j \)th letter of \( u_k \)).

**Claim 14** The above definition (19) of \( x \) makes

\[
u_j = u_{x_j}, \quad 1 \leq j \leq n
\]  

(20)
and

\[
x \text{ defined in (19) is the unique } x \in A^n \text{ such that (20) holds.}
\]  

(21)

The claim will proved following the next Lemma. Granted the claim the
desired conclusion follows. For it is immediate from the definition (17) that
for each \( x \in A^n \), (20) is equivalent to

\[
w_i = w_{x_i} \text{ for every } i
\]  

(22)
where \( w_{x_i} \) is the \( i \)th word in the cycle sentence of \( x \). To summarize, once the
claim is established, the preceding argument proves

**Proposition 15** Formulae (17) and (19) define the unique word \( x \) whose
cycle sentence is the non-decreasing sequence of Lyndon words (16).

**Example 16** Suppose the sentence of non-decreasing Lyndon words is the
sentence

\[
(0)(001)(001)(00101)(01),
\]
obtained in Example 9 as the cycle sentence of \( x = 01110010100000 \). The array of words in (18) is

\[
(0^*), (001^*), (010^*), (100^*) \\
(001^*), (010^*), (100^*) \\
(00101^*), (01010^*), (10100^*), (01001^*), (10010^*) \\
(01^*), (10^*)
\]

Write these words \( w^* \) in increasing lexicographic order to get \( u_1, u_2, \ldots \):

\[
0^*, 001^*, 001^*, 00101^*, 010^*, 010^*, 01001^*, 01010^*, 01^*, 100^*, 100^*, 10010^*, 10100^*, 10^*.
\]

Read off the last letters of these words before they repeat to recover

\[
x = 01110010100000.
\]

To motivate the next lemma, notice that is immediate from the definition that the sequence \( (u_k) \) in (17) is such that the counting distribution of \( (u_k) \), meaning the measure \( \mu \) on \( A^* \) defined by

\[
\mu(B) = \#\{k : 1 \leq k \leq n, u_k \in B\},
\]

is \( \theta \)-invariant:

\[
\mu(B) = \mu(\theta^m B), \text{ for all integers } m.
\]

**Lemma 17** Given a sequence \( u \) in \( A^* \), with

\[
u_1 \leq u_2 \leq \ldots \leq u_n,
\]

such that the counting distribution of \( u \) is \( \theta \)-invariant, there is a unique permutation \( \pi \in S_n \) with the following properties 1) and 2):

1) \( \theta^m \circ u = u \circ \pi^m \) for all integers \( m \)

2) for each \( y \in A^* \) the permutation \( \pi \) is increasing when restricted to the set \( u^{-1}(y) \).

This \( \pi \) is given by the formula

\[
\pi(i) = \#\{j : \theta(u_j) < \theta(u_i)\} + \#\{j : \theta(u_j) = \theta(u_i) \text{ and } j \leq i\}. \quad (24)
\]

15
**Proof.** Elementary and left to the reader.

**Note.** Combining (23) with 1) and 2) above makes

\[ j < k \text{ and } u \circ \pi(j) \leq u \circ \pi(k) \Rightarrow \pi(j) < \pi(k) \quad (25) \]

which is used below.

**Lemma 18** For u as in (17), let \( x \) and \( \pi \) constructed from u as in (19) and (24). Then \( \pi = \pi_x \).

**Proof.** Let \( \xi_1 : A^* \to A \) be the first letter map. So definition (19) makes

\[ x = \xi_1 \circ \theta^{-1} \circ u. \quad (26) \]

Since \( \theta^{-1} \circ u = u \circ \pi^{-1} \) by Lemma 17 1), (26) gives also

\[ x = \xi_1 \circ u \circ \pi^{-1}. \quad (27) \]

Because \( \pi^{-1} \) is a permutation of \( \{1, \ldots, n\} \), the counting distribution of \( x \) is identical to the counting distribution of \( \xi_1 \circ u \). But \( \xi_1 \circ u \) is the composition of two non-decreasing maps, hence non-decreasing. That is to say, the non-decreasing rearrangement of \( x \) is

\[ x^\updownarrow = \xi_1 \circ u, \quad (28) \]

and (27) becomes

\[ x = x^\updownarrow \circ \pi^{-1}. \quad (29) \]

Consequently, if \( x^\updownarrow(j) = b \) say, then \( x_{\pi(j)} = b \). Also, if \( j < k \) are such that \( x^\updownarrow(j) = x^\updownarrow(k) = b \), then \( \xi_1 \circ u_j = \xi_1 \circ u_k = b \). Since \( \theta \) is an increasing map when restricted to sequences with a given first letter,

\[ u \circ \pi(j) = \theta \circ u_j \leq \theta \circ u_k = u \circ \pi(k), \]

so \( \pi(j) < \pi(k) \) by (25). That is to say, \( \pi = \pi_x \).

**Proof of Claim 14.** Let \( \xi_1 \) be the first letter map as above. Then for \( m = 1, 2, \ldots \)

\[
\begin{align*}
    u_{km} &= x^\updownarrow \circ \pi^{m-1}(k) \\
    &= \xi_1 \circ u \circ \pi^{m-1}(k) \text{ by (28)} \\
    &= \xi_1 \circ \theta^{m-1} \circ u_k \text{ by 1) of Lemma 17} \\
    &= u_{km}.
\end{align*}
\]
This is (20). Now suppose $x \in A^n$ is such that (20) holds. That $x$ is unique is easily seen by a variation of the previous argument. First of all, the $(u_{xk}, 1 \leq k \leq n)$ determine $x^!(k) = u_{xk1}$ for $1 \leq k \leq n$. Also, by definition of $u_x$ and $\pi_x$, it is immediate that $u = u_x$ and $\pi = \pi_x$ satisfy the hypotheses of Lemma 17. So $u_x$ determines both $\pi_x$ and $x^!$, hence $x = x^! \circ \pi_x^{-1}$. ■

To conclude this section we mention two variations of Proposition 10 which are implicit in the above argument:

**Proposition 19** The map $x \to u_x$ sets up a bijection between $x \in A^n$ and the set of all non-decreasing functions $u : \{1, \cdots, n\} \to A^*$ such that the counting distribution of $u$ on $A^*$ is shift-invariant.

**Proposition 20** For a $x \in A^n$, let $\mu_x$ be the counting distribution of $u_x$ on $A^*$: that is for $B \subset A^*$,

$$\mu_x(B) = \#\{k : u_{xk} \in B\}.$$

Then the map $x \to \mu_x$ sets up a bijection between $x \in A^n$ and shift-invariant counting measures on $A^*$ with total mass $n$.

**Remark.** For a Lyndon word $w$, let $O_w$ be the $\theta$-orbit containing $w^*$, so as $w$ ranges over all Lyndon words, the $O_w$ partition $A^*$ into disjoint $\theta$-orbits. Then $n_{xw}$, the number of times $w$ appears in the cycle sentence of $x$, is easily seen to be

$$n_{xw} = \mu_x(O_w)/\#(O_w),$$

where $\#(O_w)$, the number of points in $O_w$, is simply the length of $w$. Thus the composition of the bijections of Corollary 11 and Proposition 20 is the obvious correspondence between shift-invariant counting measures $\mu$ and arrays $(n_w)$ as in (14):

$$\mu(\cdot) = \sum_w n_w \#(O_w \cap \cdot).$$

## 4 Derangements with one descent.

This Section presents some consequences of the analysis of Section 3 for derangements with one descent.

Call $\pi \in S_n$ a **derigif** if $\pi$ is a derangement with exactly one descent.
A subset \( A \) of \( \{1, \ldots, n\} \) is \( \pi \)-invariant if \( \pi \) maps \( A \) to \( A \). If \( A \) has \( m \) elements, the action of \( \pi \) on \( A \) identifies a unique permutation \( \sigma \in S_m \) via the increasing correspondence between \( \{1, \ldots, m\} \) and \( A \). Say \( \pi \) acts like \( \sigma \) on \( A \). For \( m \geq 2 \) let

\[
D_m = \{ \text{deriffler} \} \subset S_m
\]

\[
C_m = \{ \text{deriffler with a single cycle of size } m \} \subset S_m.
\]

**Note.** \( \#(D_m) = 2^{m-2}, \#(C_m) = f_{2,m} \).

**Proposition 21** For \( \alpha \in D_m, \beta \in D_n \), there is a unique \( \pi \in D_{m+n} \), denoted \( \pi = \alpha + \beta \), which admits disjoint invariant sets of sizes \( m \) and \( n \) on which \( \pi \) acts like \( \alpha \) and \( \beta \) respectively. If \( d(\alpha) \) is the unique place \( d: \alpha_d > \alpha_{d+1} \), then \( d(\alpha + \beta) = d(\alpha) + d(\beta) \). The operation \( + \) on \( \bigcup_{m=2}^{\infty} D_m \) is commutative and associative. If \( C_1, C_2, \ldots \) are the \( \pi \) invariant subsets defined by cycles of \( \pi \in D_n \), and \( \pi \) acts like \( \gamma_j \) on \( C_j \), then

\[
\pi = \sum_j \gamma_j.
\]

Apart from the order of the terms, this is the unique representation of \( \pi \) as a sum of cyclic derifflers.

**Proof.** By Proposition 6, for each \( \beta \in D_n \), there is a unique sequence \( x \in \{0,1\}^n \) such that \( \beta = \pi_x \). Corollary 11 then identifies \( x \), hence \( \beta \), with an array of non-negative integers:

\[
\beta \leftrightarrow (n(w), w \in \bigcup_j L_j : n(w) = 0 \text{ for } w \in L_1, \text{ and } \sum_j \sum_{w \in L_j} n(w) = n). \quad (30)
\]

Now \( \alpha \in D_m \) corresponds to a similar array of coefficients, say \( (n'(w)) \). From the definition of the cycle words, it is clear that a permutation \( \pi \) in \( S_n \) acts like \( \alpha \) and \( \beta \) on disjoint \( \pi \)-invariant subsets iff

\[
\pi \leftrightarrow (n(w) + n'(w)). \quad (31)
\]

But by Corollary 11 again, there is a unique such \( \pi \in D_{n+m} \).

**Note.** To see disjoint invariant \( F \) and \( G \) sets of \( \{1, \ldots, m+n\} \) on which \( \pi \) acts like \( \alpha \) and \( \beta \) respectively, let \( F \) be the union over \( w \) of the union of \( n(w) \)
disjoint \( \pi \) cycles whose associated Lyndon word is \( w \), \( G = F^c \). The number of different ways to choose \( F \) is

\[
\prod_w \left( \frac{n(w) + n'(w)}{n(w)} \right).
\]

But \( \pi \) acts on \( F \) like \( \alpha \), no matter which of these choices is made. The set \( F \) is unique if and only if \( n(w) \wedge n'(w) = 0 \) for all \( w \). That is to say \( \alpha \) and \( \beta \) do not contain any cyclic component in common.

**Proof that** \( d(\beta) = d(\alpha) + d(\beta) \).

Simply note that for \( \alpha \in D^m \), \( d(\alpha) \) is the number of zeros in the sequence \( x \in \{0,1\}^m \) such that \( \pi_x = \alpha \). And the sequences \( x \) and \( y \) corresponding to \( \alpha \) and \( \beta \) are disjoint subsequences of the sequence \( z \) corresponding to \( \alpha + \beta \).

**Note.** The signature \( u_x \) of the unique \( x \in \{0,1\}^n \) associated with \( \pi \in D_n \) is given by

\[
u_{x_{kj}} = \begin{cases} 1 & \text{if } \pi^k(j) > d(\pi) \\ 0 & \text{else} \end{cases}
\]

5 Word lengths in the Lyndon decomposition of a random word.

The joint distribution appearing in Proposition 3 turns out to be identical to the joint decomposition of wordlengths in the Lyndon decomposition of a random word of length \( n \) from an alphabet of \( a \) letters. This would appear to provide the most elementary proof that the formula (6) does define a joint probability distribution over \( n \)-tuples of non-negative integers \( (n_j) \) with \( \sum_j jn_j = n \).

To formulate this precisely, fix the alphabet \( A \) with \( a \) letters. Let \( L_j \) be the set of Lyndon words of length \( j \). The set of all Lyndon words is \( \cup_j L_j \). Instead of the repeat lexicographic order imposed on \( \cup_j L_j \) in Section 3, now give \( \cup_j L_j \) the ordinary lexicographic order for words of finite length, in which \( v < w \) for any word \( w = vx \) obtained by concatenating \( v \) and another word \( x \). (e.g. 01 \( \& \) 01001 in this ordinary order, but 01 \( \& \) 01001 in the repeat order).

According to the fundamental result of Lyndon (see e.g. Lothaire (1983), Theorem 5.15), when \( \cup_j L_j \) is given the ordinary lexicographical order, every word \( x \in A^n \) can be written uniquely as the concatenation of a non-increasing sequence of Lyndon words. Call this the **Lyndon decomposition** of \( x \). Let
$M_x(w)$ denote the number of times the Lyndon word $w$ appears in the Lyndon decomposition of $x$. An immediate consequence of the Lyndon decomposition is that the map

$$x 	o (M_x(w), w \in \cup_j L_j)$$  \hspace{1cm} (32)

induces a bijection between words $x \in A^n$ and arrays of non-negative integers

$$(n(w), w \in \cup_j L_j : \sum_j \sum_{w \in L_j} n(w) = n).$$  \hspace{1cm} (33)

This should be compared with the different bijection between the same sets displayed in 11. It follows immediately, by the argument below Corollary 11, that, formula (6) gives the probability the Lyndon decomposition of a word picked at random from $A^n$ contains $n_j$ Lyndon words of length $j$, $1 \leq j \leq n$. That is to say, we have the following:

**Proposition 22** Let $M_j = M_{x_j} = \sum_j \sum_{w \in L_j} M_x(w)$ be the number of Lyndon words of length $j$ in the Lyndon decomposition of a word $x$ picked uniformly at random from $A^n$. Then the joint distribution of the counts $(M_j, 1 \leq j \leq n)$ is identical to the joint distribution of the cycle counts $(N_j, 1 \leq j \leq n)$ derived from an $a$-shuffle of $n$ cards, as described in Proposition 3.

As a consequence of this proposition, every result presented in the following sections concerning the distribution of $N_j$ and asymptotic distribution of $(N_j, 1 \leq j \leq n)$ as $n \to \infty$, applies verbatim to $M_j$ and the asymptotic distribution of $(M_j, 1 \leq j \leq n)$ as $n \to \infty$.

**Remark.** The composition of the two different bijections between $A^n$ and arrays (33) defines permutation of $A^n$ which acts on each word by rearranging its letters. Thus we have another way of inducing a permutation in $S_n$ from a word in $A^n$, besides the $a$-shuffle. This may be rather artificial, but perhaps worth studying further.

### 6 Distribution of the number of $i$-cycles.

The following proposition gives a formula for the distribution of the number of $i$-cycles in an $a$-shuffle. Combined with the inversion formula of proposition 7, this yields the joint distribution of the number of $i$ cycles and the number of descents for a uniformly distributed random permutation.
Proposition 23 Let $P_{n,a}$ denote the distribution of an $a$-shuffle, $N_i(\pi)$ the number of cycles of length $i$ in permutation $\pi$. Then for $m = 0, 1, 2, \ldots$

$$P_{n,a}(N_i = m) = \binom{f_{ia}}{m} a^{-im} P_{n-\text{im},a,i}$$

where

$$p_{n,a,i} = P_{n,a}(N_i = 0) = \sum_{k=0}^{\lfloor n/i \rfloor} \binom{f_{ia}}{k} (1)^k a^{-ik}$$

Note.

$$p_{0,a,i} = 1$$
$$p_{n,a,i} = (1 - a^{-i})^{f_{ia}} \text{ for } n \geq if_{ia}.$$ 

So for fixed $a$, as $n \to \infty$, the $P_{n,a}$ distribution of $N_i$ converges to that of $N_{ooai}$ with the negative binomial $(f_{ia}, a^{-i})$ distribution:

$$P(N_{ooai} = m) = \binom{f_{ia}}{m} a^{-im} (1 - a^{-i})^{f_{ia}}.$$ 

(36)

For fixed $i$, as $a \to \infty$,

$$f_{ia} \sim i^{-1} a^i.$$ 

(37)

So

$$EN_{ooai} = f_{ia} a^{-i} \to i^{-1} \text{ as } a \to \infty,$$

as is to be expected from the classical Poisson asymptotics as $n \to \infty$ for the cycle counts of a uniform random permutation of $n$ elements. Also from (37), for fixed $i$, as $a \to \infty$, in (35) we see

$$\binom{f_{ia}}{k} \sim \frac{i^{-k} a^{ik}}{k!}$$

$$p_{n,a} \to p_{n,ooi} = \sum_{k=0}^{\lfloor n/i \rfloor} (-1)^k \frac{i^{-k}}{k!}$$

$$P(N_{ooi} = m) \to \frac{i^{-m}}{m!} p_{n-\text{im},oo,i}, \quad 0 \leq m \leq \lfloor n/i \rfloor.$$
These limits define the distribution of $N_i$ for a uniformly distributed random permutation. See e.g. Wilf (1990, formula (4.2.14)) for the generating function in this case, which can be seen to be consistent with the above expression. Presumably a simple formula for the g.f. for finite $a$ can be obtained using the methods of section 7.

Following is a proof of the above proposition in the case $i = 1$. The general case is only slightly more complicated. Change of notation: instead of $N_1 = \text{number of fixed points},$ let

$$F_{na} = \#\text{ fixed points of an a-shuffle} = \sum_{k=0}^{a-1} F_{nak}$$

where $F_{nak}$ is the number of fixed points in the $k$th packet. In terms of the random counting measure $\mu = \mu_x$ on $A^*$ induced by the random $x \in A^n$, as in Proposition 20,

$$F_{nak} = \mu(k^*)$$

where $k^* = (k, k, k, \ldots) \in A^*$. Since $\mu$ is uniform on all shift-invariant counting measures on $A^*$ with total mass $n$, the $F_{nak}, k = 0, 1, \ldots, a - 1$ are exchangeable, with

$$P(F_{nak} = m_k, k = 0, 1, \ldots, a - 1 | F_{na} = m) = \left[ \frac{a}{m} \right]^{-1}$$

for every choice of $m_k \geq 0$ with $\sum_k m_k = m$. Consequently, for $j = 1, 2, \ldots, n$

$$P(F_{nak} \geq j) = P(F_{na0} \geq j) = P(\text{first } j \text{ digits of } x \text{ are } 0) = a^{-j}, \quad (39)$$

$$E(F_{nak}) = \sum_{j=1}^{n} a^{-j}, \quad (40)$$

$$E(F_{na}) = \sum_{k=0}^{a-1} E(F_{nak}) = a \sum_{j=1}^{n} a^{-j} = \frac{a^n - 1}{a^n - a^{n-1}}. \quad (41)$$

Also

$$P(F_{na} = m) = a^{-n} \left[ \frac{a}{m} \right] \text{[number ways to put mass } n-m \text{ to orbits not fixed points]}$$

22
hence
\[ P(F_{na} = m) = \binom{a}{m} a^{-m} P(F_{n-m,a} = 0). \] (42)

This is (34) for the case \(i = 1\). So as in the classical (uniform) case, it suffices to calculate the derangement probabilities
\[
\begin{align*}
  p_{na} &= P(\text{an } a\text{-shuffle of } n \text{ cards has no fixed point}) \\
     &= P(F_{na} = 0).
\end{align*}
\]

To do this, let \(d_{na} = a^n p_{na}\), the number of sequences \(x \in A^n\) such that the associated permutation \(\pi_x\) is a derangement. The identity
\[
1 = \sum_{m=0}^{n} P(F_{na} = m)
\]
combined with (42) gives
\[
a^n = \sum_{m=0}^{n} \binom{a}{m} d_{n-m,a} \quad n = 0, 1, \ldots \] (43)

Let
\[
\begin{align*}
  D_a(z) &= \sum_{n=0}^{\infty} d_{na} z^n \\
  G_a(z) &= \sum_{n=0}^{\infty} \binom{a}{n} z^n = (1 - z)^{-a} \\
  H_a(z) &= \sum_{n=0}^{\infty} a^n z^n = (1 - az)^{-1}
\end{align*}
\]

Then (43) amounts to
\[
H_a(z) = G_a(z) D_a(z),
\]
whence
\[
D_a(z) = (1 - z)^a (1 - az)^{-1}.
\]

Collecting coefficients gives
\[
d_{n,a} = a^n \sum_{k=0}^{n} \binom{a}{k} (-1)^k a^{-k},
\]

23
whence the formula (35) for $p_{n,a} = p_{n,a,1}$.

**Remark.** The above argument shows that for fixed $a$, as $n \to \infty$, the joint distribution of

$$(F_{nak}, k = 0, \ldots, a - 1)$$

converges to that of

$$(F_{oak}, k = 0, \ldots, a - 1),$$

where the $F_{oak}$ are independent geometric random variables with parameter $1/a$.

The above proof of Proposition 23 for case $i = 1$ extends straightforwardly to general $i = 1, 2, \ldots$. Formula (34) is derived exactly as before, and (35) follows from (34) by a generating function argument. In this case $N_i$ is the sum of $f_{ia}$ exchangeable terms, each corresponding the number of $i$ cycles whose signatures induce the $f_{ia}$ possible circular words. For fixed $a$ as $n \to \infty$, the joint distribution of these $f_{ia}$ variables approaches that of $f_{ia}$ independent geometric ($a^{-i}$) variables. The analysis of Section 7 combined with the present discussion shows further that as $i$ varies there is joint convergence to independent arrays of such independent geometric variables. (Formulate as a proposition ???) See also Section 7 for formulae for the factorial moments of the $N_i$ for an $a$-shuffle.

## 7 Generating Functions.

The formula of Proposition 3 for the distribution of the cycle counts under an $a$-shuffle can be recast in the language of generating functions. This allows organized computations for moments and a probabilistic interpretation in terms of randomizing $n$. To be definite we develop results assuming that $N_i = N_i(\pi)$ is the number of cycles of $\pi$ of length $i$ for $\pi$ derived as a random $a$-shuffle of $n$ cards. But as a consequence of Proposition 22, the results apply just as well to $N_i = N_i(x)$ defined as the number of words of length $i$ in the Lyndon decomposition of a uniformly distributed random word $x$ of length $n$, for an alphabet of $a$ letters.

For a function $X(\pi)$ of permutations $\pi$, let

$$E_{n,a}X = \sum_{\pi \in S_n} P_{n,a}(\pi)X(\pi),$$

24
the expectation of $X(\pi)$ when $\pi$ is a random $a$-shuffle of $n$ cards. From Proposition 3 we obtain at once the following:

**Proposition 24** Let

$$g_{na} = g_{na}(x_1, \ldots, x_n) = E_{n,a} \prod_{i=1}^{n} x_i^{N_i},$$  \hspace{1cm} (44)

the joint probability generating function for the cycle counts $(N_i, 1 \leq i \leq n)$ derived from a random $a$-shuffle of $n$ cards. Then

$$g_{na} = a^{-n} \sum_{\sum^n_{i=1} n_i = n} \prod_{i=1}^{n} \left( \begin{array}{c} f_{ia} \newline n_i \end{array} \right) x_i^{n_i},$$  \hspace{1cm} (45)

where the sum is over all sequences of non-negative integers $(n_1, \ldots, n_n)$ with $\sum in_i = n$, and

$$\sum_{n=0}^{\infty} z^n a^n g_{na} = \prod_{i=1}^{\infty} [1 - z^i x_i]^{-f_{ia}}.$$  \hspace{1cm} (46)

**Proof.** Formula (45) is immediate from Theorem. To see (46), note that for $f = 1, 2, \ldots$

$$(1 - z)^{-f} = \sum_{m=0}^{\infty} \left( \begin{array}{c} f \newline m \end{array} \right) z^m.$$  \hspace{1cm} (47)

Apply this identity on the right side of (46), multiply out, and use (45) to recognise $g_{na}$ as the coefficient of $t^n$. ■

The identity (46) is an extension of the cyclotomic identity

$$(1 - az)^{-1} = \prod_{i=1}^{\infty} (1 - z^i)^{-f_{ia}}.$$  \hspace{1cm} (48)

This is the special case of (46) when $x + i \equiv 1$. The above argument shows this identity amounts to the fact that for each $n$ the right side of formula (6) defines a probability distribution over over all sequences of non-negative integers $(n_1, \ldots, n_n)$ with $\sum in_i = n$. The simplest way to see this appears to be via Proposition 22. See Metropolis and Rota (1983, 1984) for discussion and alternative proofs of (48). A probabilistic interpretation of (46) can be given as follows:
Proposition 25 Fix $t$ with $0 < t < 1$. Let $N$ be picked at random from \{0, 1, 2, \ldots\} according to the geometric distribution

$$P(N = n) = (1 - t)^n.$$ 

Given $N$, let $\pi$ be a random $a$-shuffle of $N$ cards, and let $N_i$ be the number of cycles of $\pi$ of length $i$. Then the random variables $(N_i, 1 \leq i < \infty)$ are independent, and $N_i$ has negative binomial distribution with parameters $f_{ia}$ and $(t/a)^i$.

Proof. Using (48) in (46) gives

$$\sum_{n=0}^{\infty} (1 - t)^n g_n = \prod_{i=1}^{\infty} \left[ \frac{1 - (t/a)^i}{1 - (t/a)^i x_i} \right]^{f_{ia}}. \tag{49}$$

This is valid in the ring of formal power series. If all but finitely many $x_i$ are set equal to 1, it is valid for $0 \leq t < 1$ and the remaining $x_i$ in $[0, 1]$, for all $a = 1, 2, 3, \ldots$. This shows that any finite collection of $N_i$ have the stated distribution which is enough to prove the result.

Remark. Heuristically, setting $t = 1$ corresponds to the limit as $n \to \infty$. Thus for fixed $a$ and large $n$ we expect $N_i$ to be asymptotically independent with negative binomial $(a^{-1}, f_{ia})$ distributions. The one-dimensional version of this result was obtained already in Section 6. The finite-dimensional version is proved below by the method of moments. Similar heuristics hold for "reasonable" functional of the cycle structure vector $N_i$. ???. This is parallel to the results of Shepp and Lloyd (1966) for uniformly distributed permutations.

As a consequence of (49), we now derive moments of arbitrary order for $N_i(\pi)$.

Proposition 26 Let $(n)_k = n(n-1) \ldots (n-k+1)$. The $k$th factorial moment of the number of $i$-cycles created by an $a$-shuffle of $n$ cards is

$$E_{n,a} (N_i)_k = (f_{ia})_k \sum_{m=k}^{[n/i]} \binom{m-1}{m-k} a^{-im}.$$ 

Proof. Set $x_j = 1$ for $j \neq i$ in (49) to obtain

$$\sum_{n=0}^{\infty} (1 - t)^n E_{n,a} x^{N_i} = \left[ \frac{1 - (t/a)^i}{1 - (t/a)^i x} \right]^{f_{ia}}.$$ 

26
Differentiate \( k \) times with respect to \( x \) and then set \( x = 1 \) to get

\[
\sum_{n=0}^{\infty} (1-t)^n E_{n,a}(N_i)_k = (f_{ia})_k (t/a)^{ki} \left[ 1 - (t/a)^i \right]^k.
\]

The result follows by comparing coefficients of \( t^n \) on both sides. \( \blacksquare \)

**Remark.** A negative binomial variable with parameters \( f \) and \( p \) has \( k \)th falling factorial moments \((f)_k \left( \frac{p}{1-p} \right)^k\). Letting \( n \) tend to infinity with \( a \) fixed in Proposition 26 confirms the result of Section 6 that \( N_i \) has a limiting negative binomial distribution with parameters \( f_{ia} \) and \( a^{-i} \). The same argument works for the joint distribution of any finite number of \( N_i \). We record this formally as follows:

**Proposition 27** For fixed \( a \), as \( n \) tends to infinity, the limiting joint distribution of the numbers \( N_i \) of \( i \)-cycles for a random \( a \)-shuffle is the distribution of independent negative binomial variables with parameters \((f_{ia}, a^{-i})\).

The analysis of riffle shuffles given by Bayer and Diaconis (1991) showed that order \( \frac{3}{2} \log_2 n \) riffle shuffles are necessary and sufficient to close to the uniform distribution on \( S_{sub n} \) in total variation distance. As a final consequence of the analysis presented here we show that features that only depend on cycle structure have the correct distribution after \( c(n) \) shuffles, where \( c(n) \) tends to infinity arbitrarily slowly with \( n \).

**Proposition 28** Let \( a(n) \) tend to infinity with \( n \). As \( n \to \infty \) the \( P_{n,a(n)} \) joint distribution of the numbers \( N_i \) of \( i \)-cycles converges to the distribution of independent Poisson variables with parameters \( 1/i \).

**Proof.** For notational simplicity, the argument is given just for the one-dimensional distribution of \( N_1 \). From Proposition 26, the \( k \)th falling factorial moment of \( N_1 \) is

\[
E_{n,a(n)}(N_1)_k = \frac{(a)_k}{a^k} \sum_{j=0}^{n-k} \left( \frac{k+j-1}{\binom{j}{a^j}} \right).
\]

The sum is bounded below by 1 and above by \((1-\frac{1}{a})^{-k}\) fixed \( k \). Thus, for any fixed \( k \), \( E_{n,a(n)}(N_1)_k \) tends to 1 as \( n \) tends to infinity. This is the \( k \)th falling
factorial moment of a Poisson variable with parameter 1. The argument for
the joint moments of $N_1, \ldots, N_i$ is essentially the same for any number of
counts $i$. Further details are omitted.

**Remark.** The method of moments proof given above does not give a bound
for finite $n$ and $a(n)$. In the case of the convergence of $N_i$ to Poisson $(1/i)$
under the uniform distribution, sharp rates have been given by Barbour and
Stein (1991). The closed form expression for the distribution of $N_i$ given in
Section 6 may be used to get explicit bounds, but a more refined probabilistic
representation is needed to get rates for the joint distribution of $N_1, \ldots, N_i$. 

28
REFERENCES


