Random preorders and alignments

Peter Cameron[1], Mihyun Kang[2] and Dudley Stark[1]

[1]School of Mathematical Sciences
Queen Mary, University of London
Mile End, London, E1 4NS U.K.
{p.j.cameron, dstark}@qmul.ac.uk
[2]Humboldt-Universität zu Berlin, Institut für Informatik
Unter den Linden 6, 10099 Berlin, Germany
kang@informatik.hu-berlin.de

July 21, 2006

Abstract

A preorder consists of linearly ordered equivalence classes called *blocks*, and an alignment is a sequence of cycles on n labelled elements. We investigate the block structure of a random preorder chosen uniformly at random among all preorders on n elements, and also the distribution of cycles in a random alignment chosen uniformly at random among all alignments on n elements, as $n \to \infty$.

1 Introduction

A preorder consists of linearly ordered equivalence classes called *blocks*, and an alignment is a sequence of cycles on n labelled elements. The number of preorders and alignments are closely related to the Stirling number of the second kind and of the first kind, which are well studied numbers [7, 9]. Let n and r be positive integers with $1 \leq r \leq n$. Let $p(n)$ denote the number of preorders possible on a set of n elements, say $[n] := \{1, 2, \cdots, n\}$.

Let $S(n, k)$ denote the Stirling number of the second kind. Note that the number of preorders with exactly r blocks is

$$p(n, r) = r!S(n, r) \tag{1}$$

Therefore, $p(n) = \sum_{r=0}^{n} r!S(n, r)$. Let $q(n)$ denote the number of alignment possible on $[n]$, and let $q(n, r)$ denote the number of alignment possible on $[n]$ with exactly r cycles. Let $s(n, r)$ denote the Stirling number of the first kind, which is defined by the rule that $(-1)^{n-r}s(n, r)$ counts the number of permutations of $[n]$ with exactly r cycles. Note that the number of alignments on $[n]$ with exactly r cycles is

$$q(n, r) = r!(-1)^{n-r}s(n, r) \tag{2}$$

and therefore $q(n) = \sum_{r=1}^{n} q(n, r) = \sum_{r=1}^{n} r!|s(n, r)|$. These numbers can be extended for all nonnegative integers n and r by defining $p(n, r) = 0$ unless $1 \le r \le n$, but $p(0, 0) = 1$, and $s(n, r) = 0$ unless $1 \le r \le n$, but $s(0, 0) = 1$. The identity

$$\sum_{n=0}^{\infty} \frac{S(n, r)z^n}{n!} = \frac{(e^z - 1)^r}{r!} \tag{3}$$

is proven in Proposition (5.4.1) of [7] using inclusion-exclusion. It is also proved that the exponential generating function of the Stirling number of the first kind $s(n, r)$ with r fixed has an explicit form (see e.g., Proposition (5.4.4) of [7])

$$\sum_{n=0}^{\infty} \frac{s(n, r)z^n}{n!} = \sum_{n=r}^{\infty} \frac{s(n, r)z^n}{n!} = \frac{(\ln(1 + z))^r}{r!}. \tag{4}$$

In this paper we derive a key lemma (Lemma 3.1) from (1) to look at the block structure of a random preorder. We give asymptotic estimates of the number of blocks, the size of a typical block, and the number of blocks of a particular size. We show that the maximal size of a block asymptotically takes on one of two values.

There have been previous results about random preorders. It was shown in [3] that the number of blocks is asymptotically normal. In [9], Proposition V.6, the asymptotic expectation and variance of the number of blocks of size s are derived for fixed s only, while we find the asymptotic number of blocks of size s for $s = s(n)$, thereby giving a complete description of the block structure of random preorders. The distribution of the maximal block size of a random preorder was looked at in Example 8 (Largest number of pre-images in surjections) of [12]. We go further and get a two point concentration result.

We also use (2) to derive another key lemma (Lemma 4.1) and use it to study the cycle structure of random alignments. We get a complete description of the cycle structure and size of a typcial cycle as well as an almost sure asymptotic for the size of the largest cycle. Previous results on alignments, such as the moments of the number of cycles of fixed size, can be found in [9].

Preorders and alignments are orbits of particular oligomorphic permutation groups and the structure of orbits of other oligomorphic permutation groups may be the subject of further related study. Some further remarks to that effect are made in the concluding section.

2 Preliminaries

If $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is a power series, then we use the notation $[z^n]\, f(z)$ to denote a_n.

Lemma 2.1 *Let $f(z) = (1-z)^{-\alpha}$ with $\alpha \notin \{0, -1, -2, \ldots\}$. For large n the coefficients $[z^n]\, f(z)$ has a singular expansion in descending powers of n,*

$$[z^n]\, f(z) = \binom{n+\alpha-1}{n} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + \sum_{k=1}^{\infty} \frac{e_k(k)}{n^k}\right) \tag{5}$$

where $\Gamma(\alpha) := \int_0^\infty e^{-t} t^{\alpha-1} dt$ for $\alpha \notin \{0, -1, -2, \ldots\}$, and $e_k(k)$ is a polynomial in k of degree $2k$.

Note in particular that $\Gamma(k+1) = k!$ for positive integer k.

A random variable X has the logarithmic distribution with parameter $p \in (0,1)$ if for $k = 1, 2, 3, \cdots$,

$$\mathbb{P}(X = k) = \frac{p^k}{k \ln(1-p)^{-1}}.$$

It is denoted by $\mathrm{Log}(p)$ and its mean and variance are given by

$$\mathbb{E}(X) = \frac{p}{(1-p)\ln(1-p)^{-1}}, \quad \mathrm{Var}(X) = \frac{-p(p+\ln(1-p))}{(1-p)^2 \ln^2(1-p)}.$$

We use the notation $a_n \sim b_n$ for sequences a_n, b_n to mean $\lim_{n\to\infty} a_n/b_n = 1$. We write $X_n \overset{\text{a.a.s.}}{\sim} a_n$ (X_n converges to a_n asymptotically almost surely) to mean $\lim_{n\to\infty} \mathbb{P}(|X_n/a_n - 1| > \epsilon) = 0$ for all $\epsilon > 0$.

We need Chebyshev's inequality: for a random variable X and any $\lambda > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2} \tag{6}$$

The kth falling factorial of a real number x is defined to be

$$(x)_k = x(x-1)(x-2)\cdots(x-k+1)$$

(and thus $(x)_k = 0$ for a nonnegative integer $x < k$). The kth factorial moment of a random variable X to be

$$\mathbb{E}(X)_k = \mathbb{E}X(X-1)(X-2)\cdots(X-k+1).$$

Observe that for any x

$$\sum_{r=0}^{\infty}(r)_k x^r = \frac{k!x^k}{(1-x)^{k+1}}. \tag{7}$$

$$\sum_{r=0}^{\infty}(k+r)_k\, x^r = \frac{d^k}{dx^k}\frac{x^k}{1-x} \quad = \quad \sum_{j=0}^{k}\binom{k}{j}\frac{(k-j)!(k)_j x^{k-j}}{(1-x)^{k-j+1}}, \tag{8}$$

which follows from the Leibnitz formula $\frac{d^k}{dx^k}uv = \sum_{j=0}^{k}\binom{k}{j}\frac{d^j}{dx^j}u\frac{d^{k-j}}{dx^{k-j}}v$ for functions $u(x)$ and $v(x)$.

3 Random preorders

Let R be a binary relation on a set X. We say R is *reflexive* if $(x,x) \in R$ for all $x \in X$. We say R is *transitive* if $(x,y) \in R$ and $(y,z) \in R$ implies $(x,z) \in R$. A *partial preorder* is a relation R on X which is reflexive and transitive. A relation R is said to satisfy *trichotomy* if, for any $x, y \in X$, one of the cases $(x,y) \in R$, $x = y$, or $(y,x) \in R$ holds. We say that R is a *preorder* if it is a partial preorder that satisfies trichotomy. The members of X are said to be the *elements* of the preorder.

A relation R is *antisymmetric* if, whenever $(x, y) \in R$ and $(y, x) \in R$ both hold, then $x = y$. A relation R on X is a *partial order* if it is reflexive, transitive, and antisymmetric. A relation is a *total order*, if it is a partial order which satisfies trichotomy. Given a partial preorder R on X, define a new relation S on X by the rule that $(x, y) \in S$ if and only if both (x, y) and (y, x) belong to R. Then S is an equivalence relation. Moreover, R induces a partial order $\overline{x}$ on the set of equivalence classes of S in a natural way: if $(x, y) \in R$, then $(\overline{x}, \overline{y}) \in \overline{R}$, where $\overline{x}$ is the S-equivalence class containing x and similarly for y. We will call an S-equivalence class a *block*. If R is a preorder, then the relation $\overline{R}$ on the equivalence classes of S is a total order. See Section 3.8 and question 19 of Section 3.13 in [7] for more on the above definitions and results.

Preorders are used in [14] to model the voting preferences of voters. (A different but equivalent definition of preorders is used in [14], where they are called weak orders.) We suppose that there are n candidates and m voters. Suppose that X is a finite set representing a collection of candidates. Let R_i, $i = 1, 2, \ldots, m$, be a set of weak orders on X. Then $(x, y) \in R_i$ means that the ith voter prefers candidate y to candidate x. The R_i blocks correspond to sets of candidates to which voter i is indifferent.

The assumption is made in [14] that each voter chooses his voting preference uniformly at random from all of the $p(n)$ possibilities independently of the other voters. An algorithm for generating a random preorder is given in [14] and the ideas behind the algorithm are used to derive a formula for the probability of the occurrence of "Condorcet's paradox". See [11] for a survey of assumptions on voter preferences used in the study of Condorcet's paradox.

3.1 The number of preorders

In this subsection we study the asymptotic number of preorders and the distribution of blocks in a random preorder. The following consequence of (1) and (3) will be used to derive complete information about the block structure of a random preorder.

Lemma 3.1 *For any sequence θ_r and any nonnegative integer n,*

$$\sum_{r=0}^{n} \theta_r p(n, r) = n! \, [z^n] \left(\sum_{n=0}^{\infty} \theta_n (e^z - 1)^n \right).$$

Proof We observe that

$$\begin{aligned}
\sum_{n=0}^{\infty}\left(\sum_{r=0}^{n}\theta_r p(n,r)\right)\frac{z^n}{n!} &= \sum_{r=0}^{\infty}\left(\sum_{n=r}^{\infty}\frac{p(n,r)z^n}{n!}\right)\theta_r \\
&\overset{(1)}{=} \sum_{r=0}^{\infty}\left(\sum_{n=r}^{\infty}\frac{S(n,r)z^n}{n!}\right)r!\,\theta_r \\
&\overset{(3)}{=} \sum_{r=0}^{\infty}\theta_r(e^z-1)^r.
\end{aligned}$$

The lemma follows immediately. ■

If we take $\theta_r = 1$ in Lemma 3.1, then we find that

$$p(n) = n![z^n]\left(2-e^z\right)^{-1},$$

an identity proved in [2]. The singularity of smallest modulus of $(2-e^z)^{-1}$ occurs at $z = \log 2$ with residue

$$\lim_{z\to\log 2}\left(\frac{z-\log 2}{2-e^z}\right) = \lim_{z\to\log 2}\left(\frac{1}{-e^z}\right) = -\frac{1}{2}, \tag{9}$$

by l'Hôpital's rule. So the function

$$\frac{1}{2-e^z} + \frac{1}{2(z-\log 2)}$$

is analytic in a circle with centre at the origin and the next singularities of $(2-e^z)^{-1}$ (at $\log 2 \pm 2\pi i$) on the boundary. Thus

$$p(n) \sim \frac{n!}{2}\left(\frac{1}{\log 2}\right)^{n+1}, \tag{10}$$

and indeed it follows from Theorem 10.2 of [15] that the difference between the two sides is $o((r-\epsilon)^{-n})$, where $r = |\log 2 + 2\pi i|$; that is, exponentially small. An exact expression for $(2-e^z)^{-1}$ is given in [1] in terms of its singularities and the truncation error from using only a finite number of singularities is estimated.

3.2 The number of blocks

We are interested in the size of the blocks in a random preorder. Let B_1 be the size of the first block, let B_2 be the size of the second, and let B_i be the size of the ith block. If the preorder has N blocks we define $B_i = 0$ for $i > N$. It is an identity that

$$\sum_{i=1}^{\infty} B_i = n. \tag{11}$$

We can represent a preorder on the set X by the sequence $(B_1, B_2, \ldots)$, where the B_i are disjoint and $\bigcup_i B_i = X$. A related combinatorial object to preorders is set partitions, for which the blocks are not ordered. The block structure of random set partitions has been studied in [16].

We denote the number of blocks of a random preorder on n elements by X_n. In terms of the block sizes B_i we may express X_n as $X_n = \sum_{i=1}^{\infty} I[B_i > 0]$, where $I[B_i > 0]$ is the indicator variable that the ith block has positive size. In this section we give asymptotics for X_n.

Define λ_n to be

$$\lambda_n = \frac{n}{2\log 2}. \tag{12}$$

We will show that $\mathbb{E}(X_n)_k \sim \lambda_n^k$ for each fixed $k \geq 0$. By a standard argument using Chebyshev's inequality, the asymptotics of the first two moments implies that $X_n \overset{\text{a.a.s.}}{\sim} \frac{n}{2\log 2}$. Theorem 3.1 agrees with Example 3.4 of [3], where it is shown that $(X_n - \lambda_n)/\sqrt{\lambda_n}$ converges in distribution to a standard normal.

Theorem 3.1 *The kth falling moment of the number of blocks of a random preorder equals*

$$\mathbb{E}(X_n)_k = \frac{k!n!}{p(n)}[z^n]\frac{(e^z-1)^k}{(2-e^z)^{k+1}}. \tag{13}$$

It follows that for fixed k

$$\mathbb{E}(X_n)_k \sim \lambda_n^k \tag{14}$$

and that

$$X_n \overset{\text{a.a.s.}}{\sim} \lambda_n,$$

where λ_n is defined by (12).

Proof In order to prove (13) it suffices to note that

$$\mathbb{E}(X_n)_k = \sum_{r=0}^{n} \frac{p(n,r)}{p(n)}(r)_k = \frac{1}{p(n)} \sum_{r=0}^{n} p(n,r)(r)_k,$$

to apply Lemma 3.1 with $\theta_r = (r)_k$, and to observe (7).

We now show (14). An analysis similar to (9) shows that

$$\lim_{z \to \log 2} \frac{(z-\log 2)^{k+1}(e^z-1)^k}{(2-e^z)^{k+1}} = \left(-\frac{1}{2}\right)^{k+1}$$

and that

$$\frac{(e^z-1)^k}{(2-e^z)^{k+1}} - \frac{(-1/2)^{k+1}}{(z-\log 2)^{k+1}} \tag{15}$$

is analytic on any disc of radius less than $|\log 2 + 2\pi i|$. Singularity analysis (Section 11 of [15]) implies that

$$\begin{aligned}
[z^n]\frac{(e^z-1)^k}{(2-e^z)^{k+1}} &\overset{(15)}{\sim} \left(\frac{1}{2\log 2}\right)^{k+1} [z^n](1-z/\log 2)^{-k-1} \\
&\overset{(5)}{\sim} \left(\frac{1}{2\log 2}\right)^{k+1} \frac{n^k}{\Gamma(k+1)(\log 2)^n}.
\end{aligned} \tag{16}$$

Therefore,

$$\begin{aligned}
\mathbb{E}(X_n)_k &\overset{(13)}{\sim} \frac{k!n!}{p(n)}[z^n]\frac{(e^z-1)^k}{(2-e^z)^{k+1}} \\
&\overset{(16)}{\sim} \frac{n!}{p(n)}\left(\frac{1}{2\log 2}\right)^{k+1} \frac{n^k}{(\log 2)^n} \\
&\overset{(10)}{\sim} \lambda_n^k.
\end{aligned}$$

The variance of X_n is asymptotically $\mathrm{Var}(X_n) = \mathbb{E}X_n(X_n-1) + \mathbb{E}X_n - (\mathbb{E}X_n)^2 = \lambda_n^2 + o(\lambda_n^2) + (\lambda_n + o(\lambda_n)) - (\lambda_n + o(\lambda_n))^2 = \lambda_n + o(\lambda_n^2)$. The conclusion that $X_n \overset{\text{a.a.s.}}{\sim} \lambda_n$ is a consequence of

$$\mathbb{P}\left(|X_n/\lambda_n - 1| > \epsilon\right) = \mathbb{P}\left(|X_n - \lambda_n| > \epsilon\lambda_n\right) \le \mathrm{Var}(X_n)/(\epsilon\lambda_n)^2 = o(1).$$

■

3.3 The size of a typical block

Because the blocks in a random preorder are linearly ordered, we may take B_1 as the size of a typical block. Given a preorder $(B_1, B_2, \ldots)$ on X, we may define a new preorder on $X \backslash B_1$ by the sequence $(B_2, B_3, \ldots)$. This operation can be reversed: given a preorder on $X \backslash B_1$, $(B_2, B_3, \ldots)$, we can insert B_1 to get the original preorder on X. The above correspondence implies

$$\mathbb{P}(B_1 = k) = \binom{n}{k} \frac{p(n-k)}{p(n)}$$

and for fixed k the asymptotic (10) gives

$$\mathbb{P}(B_1 = k) \sim \binom{n}{k} \frac{(n-k)!}{n!} (\log 2)^k = \frac{(\log 2)^k}{k!}. \tag{17}$$

It is easily checked that the distribution defined by the right hand side of (17) is the same as the distribution of the conditioned random variable $(Z|Z > 0)$, where Z is Poisson$(\log 2)$ distributed.

We will use an argument similar to the one above and the results of Section 2 to show that the distribution of fixed block sizes are asymptotically i.i.d. and distributed as $(Z|Z > 0)$.

Theorem 3.2 *Let a finite set of indices $i_1, i_2, \ldots, i_L$ and a sequence of non-negative integers $a_1, a_2, \ldots, a_L$ be given. Then*

$$\mathbb{P}(B_{i_1} = a_1, B_{i_2} = a_2, \ldots, B_{i_L} = a_L) \sim \prod_{i=1}^{L} \frac{(\log 2)^{a_i}}{a_i!}.$$

That is, the distribution of the B_{i_j} converges weakly to an i.i.d. sequence of random variables distributed as $(Z|Z > 0)$, where Z is Poisson$(\log 2)$ distributed.

Proof Given a preorder with blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_L}$ on X, we can form a new preorder by $(B_1, \ldots, B_{i_1-1}, B_{i-1+1}, \ldots, B_{i_2-1}, B_{i_2+1} \ldots)$ on $X \backslash \bigcup_{l=1}^{L} B_{i_l}$. On the other hand, a preorder $(B_1, \ldots, B_{i_1-1}, B_{i-1+1}, \ldots, B_{i_2-1}, B_{i_2+1} \ldots)$ on $X \backslash \bigcup_{l=1}^{L} B_{i_l}$ forms a valid preorder $(B_1, B_2, \ldots)$ on X by the insertion of the blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_L}$ if and only if $(B_1, \ldots, B_{i_1-1}, B_{i-1+1}, \ldots, B_{i_2-1}, B_{i_2+1} \ldots)$ is a preorder with at least $i_L - L$ nonempty blocks. Therefore, with b defined as $b = \sum_{l=1}^{L} a_l$,

$$\begin{aligned} &\mathbb{P}(B_{i_1} = a_1, B_{i_2} = a_2, \ldots, B_{i_L} = a_L) \\ = &\binom{n}{a_1, a_2, \ldots, a_L, n-b} \frac{\sum_{r=i_L-L}^{\infty} p(n-b, r)}{p(n)} \\ = &\binom{n}{a_1, a_2, \ldots, a_L, n-b} \frac{p(n-b)}{p(n)} \mathbb{P}(X_{n-b} \geq a_L - L). \end{aligned} \tag{18}$$

The probability in (18) approaches 1 because of Theorem 3.1. The other factors have asymptotics that give the theorem. ■

3.4 The number of blocks of fixed size

Define $X_n^{(s)}$ to be the number of blocks of size $s = s(n)$ in a random preorder on n elements. Define $\lambda_n^{(s)}$ to be

$$\lambda_n^{(s)} = \frac{(\log 2)^{s-1} n}{2s!}. \tag{19}$$

Theorem 3.3 *The kth falling moment of the number of s-blocks of a random preorder equals*

$$\mathbb{E}(X_n^{(s)})_k = \frac{k!n!}{p(n)(s!)^k} [z^{n-ks}] \sum_{j=0}^{k} \frac{(k)_j (e^z - 1)^{k-j}}{j!(2-e^z)^{k-j+1}}. \tag{20}$$

It follows that for fixed k and $s = o(n)$ such that $\lambda_n^{(s)} \to \infty$,

$$\mathbb{E}(X_n^{(s)})_k \sim (\lambda_n^{(s)})^k$$

and that

$$X_n^{(s)} \overset{\text{a.a.s.}}{\sim} \lambda_n^{(s)}, \tag{21}$$

where $\lambda_n^{(s)}$ is defined by (19).

Proof Let $p_s(n, k)$ be the number of preorders on n elements with exactly k blocks of size s. The kth falling moment of $X_n^{(s)}$ is

$$\mathbb{E}(X_n^{(s)})_k = \frac{1}{p(n)} \sum_{r=0}^{\infty} (r)_k p_s(n, r).$$

The quantity $\sum_{r=0}^{\infty}(r)_k p_s(n,r)$ counts the number of preorders with k labelled s-blocks, where each of the labelled s-blocks is given a unique label from the set $\{1,2,\ldots,k\}$. This number is also counted by: first, choosing k s-blocks to be the ones marked; second, forming a preorder on the $n-ks$ remaining elements (with r blocks); third, inserting the s-blocks into the preorder in the order they were chosen in one of $\binom{k+r}{k}$ ways; fourth, marking the inserted s-blocks in one of $k!$ ways. We therefore have

$$\begin{aligned}
\mathbb{E}(X_n^{(s)})_k &= \frac{1}{p(n)}\sum_{r=0}^{\infty}\binom{n}{s}\binom{n-s}{s}\cdots\binom{n-(k-1)s}{s}p(n-ks,r)\binom{k+r}{k}k! \\
&= \frac{n!}{p(n)(s!)^k(n-ks)!}\sum_{r=0}^{\infty}p(n-ks,r)(k+r)_k \\
&= \frac{n!}{p(n)(s!)^k}[z^{n-ks}]\sum_{n=0}^{\infty}(k+n)_k(e^z-1)^n
\end{aligned} \tag{22}$$

where we have made use of Lemma 3.1 at (22). We use the identity (8) in (22). After substitution of the identity and simplification (22) becomes (20).

In (20), the singularity of largest degree occurs at $z=\log 2$ when $j=0$. The asymptotics of $\mathbb{E}(X_n^{(s)})_k$ are given by

$$\begin{aligned}
\mathbb{E}(X_n^{(s)})_k &\sim \frac{n!k!}{p(n)(s!)^k}[z^{n-ks}]\frac{(e^z-1)^k}{(2-e^z)^{k+1}} \\
&\sim \frac{2k!(\log 2)^{n+1}}{(s!)^k}\left(\frac{1}{2\log 2}\right)^{k+1}\frac{(n-ks)^k}{\Gamma(k+1)(\log 2)^{n-ks}} \\
&\sim \left(\frac{(\log 2)^{s-1}n}{2s!}\right)^k,
\end{aligned}$$

where we have used (10), (16), and the assumption $s=o(n)$. The almost sure convergence result (21) is an application of Chebyshev's inequality as in the proof of Theorem 3.1. ∎

The method of the proof of Theorem 3.3 can be used to derive asymptotics of joint falling moments. For example, $\mathbb{E}((X_n^{(s_1)})_{k_1}(X_n^{(s_2)})_{k_2}) \sim (\lambda_n^{(s_1)})^{k_1}(\lambda_n^{(s_2)})^{k_2}$ for fixed s_1, s_2, k_1, k_2.

Observe that $\sum_{s=1}^{\infty} s\lambda_n^{(s)} = n$ and $\sum_{s=1}^{\infty} \lambda_n^{(s)} = \lambda_n$, showing that Theorem 20 agrees with (11) and Theorem 3.1, respectively, and indicating that Theorem 3.3 gives a good picture of the block structure of a random preorder.

3.5 Maximal block size

Let $M_n = \max_{i\geq 1} B_i$ be the maximal size of a block in a random preorder. We are able to closely estimate the maximum size of a block in a random preorder. It was stated in [12] that

$$\mathbb{P}\left(M_n \leq m\right) = \exp\left(-\lambda_n^{(m+1)}(1+o(1))\right)\left(1+O(e^{-m\epsilon})\right) \tag{23}$$

for some $\epsilon > 0$. We will show that, asymptotically, M_n is concentrated on at most two values.

Define μ_n to be

$$\mu_n = \max\left\{s : \lambda_n^{(s)} \geq 1\right\}$$

and define

$$\nu_n = \begin{cases} \mu_n & \text{if } \lambda_n(\mu_n) \geq \sqrt{\mu_n}, \\ \mu_n - 1 & \text{if } \lambda_n(\mu_n) < \sqrt{\mu_n}. \end{cases} \tag{24}$$

Theorem 3.4 *Let $M_n = \max_{i\geq 1} B_i$ be the maximal size of a block in a random preorder. Let ν_n be defined by (24). Then $\nu_n \sim \log n/\log\log n$ and*

$$\mathbb{P}\left(M_n \in \{\nu_n, \nu_n + 1\}\right) \to 1 \quad \text{as } n \to \infty. \tag{25}$$

Proof Clearly, $\lambda_n^{(s)}$ is monotone decreasing in s for $s \geq 2$. Taking the logarithm of $\lambda_n^{(s)}$ produces

$$\begin{aligned} \log \lambda_n^{(s)} &= \log n + (s-1)\log\log 2 - \log s! - \log 2 \\ &= \log n - s\log s + O(s). \end{aligned} \tag{26}$$

Plugging $s = \frac{\log n}{\log\log n}$ into (26) gives

$$\log \lambda_n\left(\frac{\log n}{\log\log n}\right) = \frac{\log n \log\log\log n}{\log\log n} + O\left(\frac{\log n}{\log\log n}\right) \to \infty,$$

from which it follows that for large enough n, $\mu_n > \log n/\log\log n$. On the other hand, if we plug $\frac{\log n}{\log\log n}\left(1 + \frac{2\log\log\log n}{\log\log n}\right)$ into the right hand side of

(26) we get

$$\begin{aligned}
&\log \lambda_n\left(\frac{\log n}{\log\log n}\left(1+\frac{2\log\log\log n}{\log\log n}\right)\right)\\
=\ & -\frac{\log n}{\log\log n}\left(1+\frac{2\log\log\log n}{\log\log n}\right)\left(\log\log n-\log\log\log n+O\left(\frac{\log\log\log n}{\log\log n}\right)\right)\\
&+\log n+O\left(\frac{\log n}{\log\log n}\right)\\
=\ & -\frac{\log n\log\log\log n}{\log\log n}+O\left(\frac{\log n(\log\log\log n)^2}{(\log\log n)^2}\right)\to-\infty,
\end{aligned}$$

so that $\mu_n < \frac{\log n}{\log\log n}\left(1+\frac{2\log\log\log n}{\log\log n}\right)$ for large enough n. We have shown that

$$\frac{\log n}{\log\log n} < \mu_n < \frac{\log n}{\log\log n}\left(1+\frac{2\log\log\log n}{\log\log n}\right)$$

for large enough n and, in particular, that $\mu_n \sim \frac{\log n}{\log\log n}$ and $\nu_n \sim \frac{\log n}{\log\log n}$.

Define the index sets

$$\mathcal{N}_1 = \{n \geq 1 : \lambda_n(\mu_n) \geq \sqrt{\mu_n}\}$$

and

$$\mathcal{N}_2 = \{n \geq 1 : \lambda_n(\mu_n) < \sqrt{\mu_n}\}.$$

When $n \to \infty$ in $\mathcal{N}_1$, $\lambda_n^{(\nu_n)} \geq \sqrt{\mu_n} \to \infty$ as $n \to \infty$. The ratio $\lambda_n^{(\mu_n+1)}/\lambda_n^{(\mu_n+2)} = (\mu_n+2)/\log 2 \to \infty$ and $\lambda_n^{(\mu_n+1)} < 1$ imply $\lambda_n^{(\nu_n+2)} \to 0$. When $n \to \infty$ in $\mathcal{N}_2$, $\lambda_n^{(\mu_n)} \geq 1$ and $\lambda_n^{(\mu_n-1)}/\lambda_n^{(\mu_n)} = \mu_n/\log 2 \to \infty$ give $\lambda^{(\nu_n)} \to \infty$ and $\lambda_n^{(\nu_n)} < \sqrt{\nu_n}$ implies $\lambda^{(\nu_n+2)} \to 0$. Now (23) implies the result. ■

Asymptotic two-point concentration theorems are well known from random graph theory. See Theorem 7, page 260 of [5] for such a result regarding clique number.

4 Random alignments

Random permutations are well studied objects and fundamental results on random permutation deal with the cycle structure of a random permutation [5]. The distribution of the number of cycles of a random permutation on $[n]$ is asymptotically Normal with mean $\log n$ and variance $\log n$.

The expected length of a longest cycle converges to cn as $n \to \infty$, where $c = \int_0^\infty \exp(-x - \int_x^\infty y^{-1} \exp(-y)dy)dx = 0.624329\cdots$.

What can we say about the distribution of cycle structures if we change the weight of cycles in the cycle decomposition of a random permutation ? In particular, we are interested in the case that we weight a cycle decomposition of a permutation according to the number of ways of linearly arranging the cycles in the cycle decomposition. This corresponds exactly to an alignment, which is a sequence of cycles on n labelled elements, in other words, a collection of directed cycles arranged in a linear order.

In this section we study the asymptotic number of alignments and the distribution of cycles and the length of a longest cycle in a random alignment.

4.1 The asymptotic number of alignments

In order to investigate the structure of cycles in a random alignments we start with the following consequence of (2) and (4) which is similar to Lemma 3.1.

Lemma 4.1 *For any sequence θ_r and any nonnegative integer n*

$$\sum_{r=0}^{n} \theta_r q(n,r) = n!\,[z^n] \left(\sum_{r=0}^{\infty} \theta_r (\ln(1-z)^{-1})^r \right). \tag{27}$$

Proof Using the identities (2) and (4) gives

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} \theta_r q(n,r) \frac{z^n}{n!} &= \sum_{n=0}^{\infty} \left(\sum_{r=0}^{n} \theta_r q(n,r) \right) \frac{z^n}{n!} \\ &= \sum_{r=0}^{\infty} \theta_r (\ln(1-z)^{-1})^r. \end{aligned} \tag{28}$$

■

If we take $\theta_r = u^r$ in Equation (28), we obtain the bivariate exponential generating function for the number of alignments $q(n,r)$ on $[n]$ with exactly r cycles

$$Q(z,u) : \;=\; \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} u^r q(n,r) \frac{z^n}{n!}$$

$$
\begin{aligned}
&= \sum_{r=0}^{\infty} (u \ln(1-z)^{-1})^r \\
&= \frac{1}{1 - u \ln(1-z)^{-1}}, \qquad (29)
\end{aligned}
$$

which can also be found in Chapter IX 6. [9]. In particular taking $u = 1$ in Equation (29) we obtain the exponential generating function for the number of alignments $q(n)$ on $[n]$

$$
Q(z) := \sum_{n=0}^{\infty} q(n) \frac{z^n}{n!} = \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} q(n,r) \frac{z^n}{n!} = Q(z,1) = \frac{1}{1 - \ln(1-z)^{-1}},
$$

which yields immediately that

$$
q(n) = n! \, [z^n] \left(\frac{1}{1 - \ln(1-z)^{-1}} \right).
$$

The dominant singularity of $Q(z) = \left(1 - \ln(1-z)^{-1}\right)^{-1}$ occurs at $z = 1 - e^{-1}$ with residue

$$
\lim_{z \to 1 - e^{-1}} \left(\frac{z - (1 - e^{-1})}{1 - \ln(1-z)^{-1}} \right) = \lim_{z \to 1 - e^{-1}} \left(\frac{1}{-1/(1-z)} \right) = -e^{-1}, \qquad (30)
$$

by l'Hôpital's rule. So the function

$$
\frac{1}{1 - \ln(1-z)^{-1}} + \frac{e^{-1}}{z - (1 - e^{-1})}
$$

is analytic in a circle with centre at the origin and the next singularities of $\left(1 - \ln(1-z)^{-1}\right)^{-1}$ (at $1 - e^{-1} \pm 2\pi i$) on the boundary. Thus

$$
q(n) \sim n! \, [z^n] \frac{-e^{-1}}{z - (1 - e^{-1})} \sim \frac{n!}{e(1 - e^{-1})^{n+1}}, \qquad (31)
$$

which one can also find in Example 7 in Chapter IV of [9].

4.2 The number of cycles

We are interested in the size of the cycles in a random alignment. Let C_1 be the size of the first cycle, let C_2 be the size of the second, and let C_i be the size of the ith cycle. If the alignment has N cycles we define $C_i = 0$ for $i > N$. It holds that

$$\sum_{i=1}^{\infty} C_i = n. \tag{32}$$

We denote the number of cycles of a random alignment on n elements by Y_n. In terms of the cycle sizes C_i we may express Y_n as $Y_n = \sum_{i=1}^{\infty} I[C_i > 0]$, where $I[C_i > 0]$ is the indicator variable that the ith cycle has positive size. In this section we give asymptotics for Y_n.

Define τ_n to be

$$\tau_n = \frac{n}{e-1}. \tag{33}$$

We will show that $\mathbb{E}(Y_n)_k \sim \tau_n^k$ for each fixed $k \geq 0$.

Theorem 4.1 *The kth factorial moment of the number of cycles of a random alignment equals*

$$\mathbb{E}(Y_n)_k = \frac{n!\,k!}{q(n)}[z^n]\frac{(\ln(1-z)^{-1})^k}{(1-(\ln(1-z)^{-1}))^{k+1}}. \tag{34}$$

It follows that for fixed k

$$\mathbb{E}(Y_n)_k \sim \tau_n^k \tag{35}$$

and that

$$Y_n \overset{\text{a.a.s.}}{\sim} \tau_n,$$

where τ_n is defined by (33).

Proof In order to prove (34) it suffices to note that

$$\mathbb{E}(Y_n)_k = \sum_{r=0}^{n} \frac{q(n,r)}{q(n)}(r)_k = \frac{1}{q(n)}\sum_{r=0}^{n} q(n,r)(r)_k,$$

and to apply Lemma 4.1 with $\theta_r = (r)_k$. Using (7) and (27) we get

$$\begin{aligned}\frac{1}{q(n)}\sum_{r=0}^{n} q(n,r)(r)_k \quad &\overset{(27)}{=} \quad \frac{n!}{q(n)}\,[z^n]\sum_{r=0}^{\infty}(\ln(1-z)^{-1})^r(r)_k \\ &\overset{(7)}{=} \quad \frac{n!\,k!}{q(n)}\,[z^n]\frac{(\ln(1-z)^{-1})^k}{(1-(\ln(1-z)^{-1}))^{k+1}}.\end{aligned}$$

We now proceed to show (35). An analysis similar to (30) shows that

$$\lim_{z\to 1-e^{-1}} \frac{(z-(1-e^{-1}))^{k+1}(\ln(1-z)^{-1})^k}{(1-\ln(1-z)^{-1})^{k+1}} = \left(-e^{-1}\right)^{k+1}$$

and that

$$\frac{(\ln(1-z)^{-1})^k}{(1-(\ln(1-z)^{-1}))^{k+1}} - \frac{(-e^{-1})^{k+1}}{(z-(1-e^{-1}))^{k+1}} \tag{36}$$

is analytic on any disc of radius less than $|1-e^{-1}+2\pi i|$. Singularity analysis (Section 11 of [15]) implies that

$$\begin{aligned}
[z^n]\frac{(\ln(1-z)^{-1})^k}{(1-(\ln(1-z)^{-1}))^{k+1}} &\overset{(36)}{\sim} \left(-\frac{1}{e}\right)^{k+1}[z^n](z-(1-e^{-1}))^{-k-1} \\
&\sim \left(\frac{1}{e(1-e^{-1})}\right)^{k+1}[z^n](1-z/(1-e^{-1}))^{-(k+1)} \\
&\overset{(5)}{\sim} \left(\frac{1}{e-1}\right)^{k+1}\left(\frac{1}{1-e^{-1}}\right)^n \frac{n^k}{k!}. \qquad (37)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}(Y_n)_k &\overset{(34)}{=} \frac{n!k!}{q(n)}[z^n]\frac{(\ln(1-z)^{-1})^k}{(1-(\ln(1-z)^{-1}))^{k+1}} \\
&\overset{(37)}{\sim} \frac{n!}{q(n)}\left(\frac{1}{e-1}\right)^{k+1}\frac{n^k}{(1-e^{-1})^n} \\
&\overset{(31)}{\sim} \left(\frac{n}{e-1}\right)^k = \tau_n^k.
\end{aligned}$$

The proof that $Y_n \overset{\text{a.a.s.}}{\sim} \tau_n$ proceeds as in the last paragraph of the proof of Theorem 3.3. ■

One would expect from Theorem 4.1 that $(Y_n-\tau_n)/\sqrt{\tau_n}$ converges weakly to the standard normal distribution. Indeed one can find such a result in in Chapter IX 6 in [9].

4.3 The size of a typical cycle

Because the cycles in a random alignment are linearly ordered, we may take C_1 as the size of a typical cycle. Given an alignment $(C_1, C_2, \ldots)$ on Y, we may define a new alignment on $Y \backslash C_1$ by the sequence $(C_2, C_3, \ldots)$. This operation can be reversed: given an alignment on $Y \backslash C_1$, $(C_2, C_3, \ldots)$, we can insert C_1 to get the original alignment on Y. The above correspondence implies

$$\mathbb{P}(C_1 = k) = \binom{n}{k}(k-1)!\frac{q(n-k)}{q(n)}$$

and for fixed k (indeed for any k with $n-k \to \infty$) the asymptotic (31) gives

$$\mathbb{P}(C_1 = k) \sim \binom{n}{k}(k-1)!\frac{(n-k)!e(1-e^{-1})^{n+1}}{e(1-e^{-1})^{n-k+1}n!} = \frac{(1-e^{-1})^k}{k}. \tag{38}$$

It is easily checked that the distribution defined by the right hand side of (38) is the same as the distribution of the conditioned random variable has the $\mathrm{Log}(1-e^{-1})$ distribution (defined in Section 2).

We will use an argument similar to the one above and the results of Section 2 to show that the distribution of fixed cycle sizes are asymptotically i.i.d. $\mathrm{Log}(1-e^{-1})$ distributed.

Theorem 4.2 *Let a finite set of indices $i_1, i_2, \ldots, i_L$ and a sequence of non-negative integers $a_1, a_2, \ldots, a_L$ be given. Then*

$$\mathbb{P}(C_{i_1} = a_1, C_{i_2} = a_2, \ldots, C_{i_L} = a_L) \sim \prod_{i=1}^{L} \frac{(1-e^{-1})^{a_i}}{a_i}.$$

That is, the distribution of the C_{i_j} converges weakly to an i.i.d. sequence of random variables with $\mathrm{Log}(1-e^{-1})$ distribution.

Proof Given an alignment with cycles $C_{i_1}, C_{i_2}, \ldots, C_{i_L}$ on Y, we can form a new alignment by $(C_1, \ldots, C_{i_1-1}, C_{i_1+1}, \ldots, C_{i_2-1}, C_{i_2+1} \ldots)$ on $Y \backslash \bigcup_{l=1}^{L} C_{i_l}$. On the other hand, an alignment $(C_1, \ldots, C_{i_1-1}, C_{i_1+1}, \ldots, C_{i_2-1}, C_{i_2+1} \ldots)$ on $Y \backslash \bigcup_{l=1}^{L} C_{i_l}$ forms a valid alignment $(C_1, C_2, \ldots)$ on X by the insertion of the cycles $C_{i_1}, C_{i_2}, \ldots, C_{i_L}$ if and only if $(C_1, \ldots, C_{i_1-1}, C_{i_1+1}, \ldots, C_{i_2-1}, C_{i_2+1} \ldots)$ is an alignment with at least $i_L - L$ nonempty cycles. Therefore, with b defined as $b = \sum_{l=1}^{L} a_l$,

$$\begin{aligned} &\mathbb{P}(C_{i_1} = a_1, C_{i_2} = a_2, \ldots, C_{i_L} = a_L) \\ = &\binom{n}{a_1, a_2, \ldots, a_L, n-b}(a_1-1)!(a_2-1)!\cdots(a_L-1)!\frac{\sum_{r=i_L-L}^{\infty} q(n-b,r)}{q(n)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{n!}{a_1 a_2 \cdots a_L (n-b)!} \frac{q(n-b)}{q(n)} \, \mathbb{P}(Y_{n-b} \geq a_L - L) \\
&\overset{(31)}{\sim} \frac{(1-e^{-1})^b}{a_1 a_2 \cdots a_L} \, \mathbb{P}(Y_{n-b} \geq a_L - L) \\
&= \prod_{i=1}^{L} \frac{(1-e^{-1})^{a_i}}{a_i} \, \mathbb{P}(Y_{n-b} \geq a_L - L). \qquad (39)
\end{aligned}
$$

The probability in (39) approaches 1 because of Theorem 4.1. ■

4.4 The number of cycles of fixed size

Define $Y_n^{(s)}$ to be the number of cycles of size $s = s(n)$ in a random alignment on n elements. Define $\tau_n^{(s)}$ to be

$$
\tau_n^{(s)} = \frac{(1-e^{-1})^{s-1} n}{es}. \qquad (40)
$$

Theorem 4.3 *The kth factorial moment of the number of s-cycles of a random alignment equals*

$$
\mathbb{E}(Y_n^{(s)})_k = \frac{n!k!}{q(n)s^k} [z^{n-ks}] \sum_{j=0}^{k} \frac{(k)_j (\ln(1-z)^{-1})^{k-j}}{j!(1-\ln(1-z)^{-1})^{k-j+1}}. \qquad (41)
$$

It follows that for fixed k and $s = o(n)$ such that $\tau_n^{(s)} \to \infty$,

$$
\mathbb{E}(Y_n^{(s)})_k \sim (\tau_n^{(s)})^k
$$

and that

$$
Y_n^{(s)} \overset{\text{a.a.s.}}{\sim} \tau_n^{(s)}, \qquad (42)
$$

where $\tau_n^{(s)}$ is defined by (40).

Proof Let $q_s(n,r)$ be the number of alignments on n elements with exactly r cycles of size s. The kth factorial moment of $Y_n^{(s)}$ is

$$
\mathbb{E}(Y_n^{(s)})_k = \frac{1}{q(n)} \sum_{r=0}^{\infty} (r)_k q_s(n,r).
$$

The quantity $\sum_{r=0}^{\infty}(r)_k q_s(n,r) = \sum_{r=k}^{\infty}(r)_k q_s(n,r)$ counts the number of alignments with k labelled s-cycles, where each of the labelled s-cycles is given a unique label from the set $\{1,2,\ldots,k\}$. This number is also counted by: first, choosing k s-cycles to be the ones marked; second, forming an alignment on the $n-ks$ remaining elements (with r cycles); third, inserting the s-cycles into the alignment in the order they were chosen in one of $\binom{k+r}{k}$ ways; fourth, marking the inserted s-cycles in one of $k!$ ways. We therefore have

$$\begin{aligned}
\mathbb{E}(Y_n^{(s)})_k &= \frac{1}{q(n)}\sum_{r=0}^{n-ks}\binom{n}{s}\binom{n-s}{s}\cdots\binom{n-(k-1)s}{s}((s-1)!)^k \cdot \\
&\quad q(n-ks,r)\cdot\binom{k+r}{k}\cdot k! \\
&= \frac{n!((s-1)!)^k}{q(n)(s!)^k(n-ks)!}\sum_{r=0}^{n-ks} q(n-ks,r)(k+r)_k \\
&\overset{(27)}{=} \frac{n!}{q(n)s^k}[z^{n-ks}]\sum_{r=0}^{\infty}(k+r)_k(\ln(1-z)^{-1})^r
\end{aligned} \tag{43}$$

where we have made use of Lemma 4.1 with $\theta_r = (k+r)_k$ at (43). We use the identity (8) in (43) and obtain

$$\begin{aligned}
\mathbb{E}(Y_n^{(s)})_k &= \frac{n!}{q(n)s^k}[z^{n-ks}]\sum_{j=0}^{k}\binom{k}{j}\frac{(k-j)!(k)_j(\ln(1-z)^{-1})^{k-j}}{(1-\ln(1-z)^{-1})^{k-j+1}} \\
&= \frac{n!k!}{q(n)s^k}[z^{n-ks}]\sum_{j=0}^{k}\frac{(k)_j(\ln(1-z)^{-1})^{k-j}}{j!(1-\ln(1-z)^{-1})^{k-j+1}}.
\end{aligned}$$

In (41), the singularity of largest degree occurs at $z = 1-e^{-1}$ when $j=0$. The asymptotics of $\mathbb{E}(Y_n^{(s)})_k$ are given by

$$\begin{aligned}
\mathbb{E}(Y_n^{(s)})_k &\sim \frac{n!k!}{q(n)s^k}[z^{n-ks}]\frac{(\ln(1-z)^{-1})^k}{(1-\ln(1-z)^{-1})^{k+1}} \\
&\overset{(37)}{\sim} \frac{n!k!}{q(n)s^k}\left(\frac{1}{e-1}\right)^{k+1}\left(\frac{1}{1-e^{-1}}\right)^{n-ks}\frac{(n-ks)^k}{k!} \\
&\overset{(31)}{\sim} \left(\frac{(1-e^{-1})^{s-1}n}{es}\right)^k.
\end{aligned}$$

where we have used the assumption $s = o(n)$. The almost sure convergence result (42) is an application of Chebyshev's inequality as in the proof of Theorem 4.1. ■

Observe that $\sum_{s=1}^{\infty} s\tau_n^{(s)} = n$ and $\sum_{s=1}^{\infty} \tau_n^{(s)} = \tau_n$, showing that Theorem 41 agrees with (32) and Theorem 4.1, respectively, and indicating that Theorem 4.3 gives a good picture of the cycle structure of a random alignment.

4.5 Maximal cycle size

In this subsection we get an estimate on the maximum size of a cycle in a random alignment.

Theorem 4.4 *Let $N_n = \max_{i\geq 1} C_i$ be the maximal size of a cycle in a random alignment. Then for any constant $K > 1/\ln((1-e^{-1})^{-1})$*

$$\mathbb{P}\left(\frac{\ln n}{\ln(1-e^{-1})^{-1}} - K\ln\ln n \leq N_n \leq \frac{\ln n}{\ln(1-e^{-1})^{-1}}\right) \to 1.$$

Proof Clearly, $\tau_n^{(s)}$ is monotone decreasing in s for $s \geq 2$. Define

$$s_K = \frac{\ln n}{\ln(1-e^{-1})^{-1}} - K\ln\ln n.$$

We have

$$\tau_n^{(s_K)} = \frac{n(1-e^{-1})^{\frac{\ln n}{\ln(1-e^{-1})^{-1}} - K\ln\ln n - 1}}{e\left(\frac{\ln n}{\ln(1-e^{-1})^{-1}} - K\ln\ln n\right)} = \Theta\left(\frac{(1-e^{-1})^{-K\ln\ln n}}{\ln n}\right) \to \infty$$

and therefore because of Theorem 4.3 it follows that $\mathbb{P}(N_n \geq s_K) \to 1$. Furthermore, with s_0 defined to be

$$s_0 = \frac{\ln n}{\ln(1-e^{-1})^{-1}},$$

we have

$$\mathbb{P}(N_n > s_0) \quad \leq \quad \sum_{s>s_0} \mathbb{E}(Y_n^{(s)})$$

$$
\begin{aligned}
&\overset{(41)}{=} \sum_{s \geq s_K} \frac{n!}{q(n)s} [z^{n-s}] \frac{1}{(1-\ln(1-z)^{-1})^2} \\
&= O\left(\sum_{s > s_0} \frac{n!(n-s)}{q(n)s(1-e^{-1})^{n-s}} \right) \qquad (44) \\
&\overset{(31)}{=} O\left(n \sum_{s > s_0} \frac{(1-e^{-1})^s}{s} \right) \\
&= O\left(n \frac{(1-e^{-1})^{s_0}}{s_0} \right) \\
&= o(1),
\end{aligned}
$$

where we have used the $O(\cdot)$ version of singularity analysis to give the upper bound indicated by (37) at (44). ■

5 Concluding remarks

Preorders and alignments are examples of structures which are linearly ordered sets of structures of a simpler type (sets and cycles, respectively). The problems studied here could be generalized simply by taking linearly ordered sets of other types of structures. However, there is a context in which these objects arise naturally, and which suggests further problems to study.

The right context is probably the notion of a *species* [13, 4]. To quote from the preface of [4],

> ... a species of structures is a rule, F, associating with each finite set U a set $F[U]$ which is "independent of the nature" of the elements of U. The members of the set $F[U]$, called F-*structures*, are interpreted as combinatorial structures on the set U given by the rule F. The fact that the rule is independent of the nature of the elements of U is expressed by an invariance under relabelling.

Examples of species include $\mathcal{S}$ (sets), $\mathcal{L}$ (linear orders), $\mathcal{C}$ (cycles), $\mathcal{T}$ (trees), $\mathcal{G}$ (graphs). So, for example, $\mathcal{G}[U]$ is the set of all graphs on the vertex set U, while $\mathcal{S}[U] = \{U\}$.

The notion of *substitution* on species is defined as follows: if F and G are species such that $G[\emptyset] = \emptyset$, then $(F \circ G)[U]$ is the set of structures consisting of a partition π of U, a G-structure on each part of π, and an F-structure on the set of parts. Thus, preorders and alignments are respectively the species $\mathcal{L} \circ \mathcal{S}$ and $\mathcal{L} \circ \mathcal{C}$ respectively.

An important source of species is the theory of infinite permutation groups [6]. A permutation group G on an infinite set Ω is said to be *oligomorphic* if it has only finitely many orbits on Ω^n for all positive integers n. Given an oligomorphic permutation group G on Ω, the structures on U in the associated species are essentially the orbits of G on $|U|$-tuples of elements of Ω. A notable special case is that where G is the automorphism group of a *homogeneous* relational structure M (one in which every isomorphism between finite substructures is induced by an automorphism). In this case, the associated species can be thought of as the *age* of M [10], the class of finite structures embeddable in M. (In fact, this is not really special, since any permutation group is associated with a canonical relational structure which is homogeneous, but this structure in general will involve infinitely many relations.)

Note that all our above examples of species except trees come from oligomorphic groups in this way. (For example, $\mathcal{L}$ is associated with the group of order-preserving permutations of the rational numbers, and $\mathcal{G}$ with the automorphism group of the "random graph".) Moreover, substitution of species corresponds to the *wreath product* of permutation groups. See [6] for more details.

Our general philosophy is that species arising from oligomorphic permutation groups should be better-behaved combinatorially than arbitrary species. This behavior could show itself in several ways: restrictions on the possible growth rate of the counting function (including "smooth growth") is an obvious one. Another, which has not been investigated, concerns cases where there is a natural notion of "connected components" in the species, in which case one can ask the questions we have considered for preorders and alignments in this paper.

Now the obvious case in which such a notion of connected components exists is a substitution species $F \circ G$, where the connected components of a $(F \circ G)$-species are the G-structures on the parts of the partition π.

If the number of G-structures grows too rapidly, then almost all $(F \circ G)$-structures will be connected, and so will be simply G-structures. If the number of F-structures grows too rapidly, then in almost all $(F \circ G)$-structures the partition π will have all its parts singletons; if there is just one G-structure on one element, then these structures are essentially just F-structures. So interesting problems arise when there is not too much imbalance between the numbers of F- and G-structures. We propose that the cases $\mathcal{S}$, $\mathcal{L}$, $\mathcal{C}$, and their close relations $\mathcal{B}$ (betweenness in a linear order) and $\mathcal{D}$ (separation in a circular order), are good candidates.

As we have observed, the cases $\mathcal{S} \circ \mathcal{S}$ (partitions) and $\mathcal{S} \circ \mathcal{C}$ (permutations) have well-developed theories, and we have considered the cases $\mathcal{L} \circ \mathcal{S}$ (preorders) and $\mathcal{L} \circ \mathcal{C}$ (alignments) in this paper. Several cases remain to be considered!

In addition, we mention the problem of deriving general results about the sizes of the components of $F \circ G$ in terms of the counting functions for F and G, and also related problems associated with other species or group constructions such as the product action of the direct product [8].

References

[1] Bailey, R. W. (1997) The number of weak orderings on a finite set. *Soc. Choice Welfare* **15** 559 – 562.

[2] Barthelemy, J. P. (1980) An asymptotic equivalent for the number of total preorders on a finite set. *Discrete Math.* **29** 311 – 313.

[3] Bender, E. A. (1973) Central and local limit theorems applied to asymptotic enumeration. *J. Comb. Theory, Ser. A* **15** 91 – 111.

[4] Bergeron, F., Labelle, G. and Leroux, P. (1998) *Combinatorial Species and Tree-Like Structures*, Encyclopedia of Mathematics and its Applications **67**, Cambridge University Press, Cambridge.

[5] Bollobás, B. (1985) *Random Graphs.* Academic Press.

[6] Cameron, P. (1990) *Oligomorphic Permutation Groups*, London Mathematical Society Lecture Notes **152**, Cambridge University Press, Cambridge.

[7] Cameron, P. (1994) *Combinatorics: Topics, Techniques, Algorithms.* Cambridge University Press.

[8] Cameron, P., Prellberg, T. and Stark, D. (2006) Asymptotic enumeration of incidence matrices, *Journal of Physics (Conference Series)* **42** 59–70.

[9] Flajolet, P. and Sedgewick, R., *Analytic combinatorics*, 0th Edition (October 1, 2005), available online at
`http://algo.inria.fr/flajolet/Publications/book051001.pdf`.

[10] Fraïssé, R. (1986) *Theory of Relations*, North-Holland, Amsterdam.

[11] Gehrlein, W. V. (2002) Condorcet's Paradox and the Likelihood of its Occurrence: Different Perspectives on Balanced Preferences. *Theory and Decision* **52** 171 – 199.

[12] Gourdon, X. (1998) Largest component in random combinatorial structures. *Discrete Math.* **180** 185 – 209.

[13] Joyal, A. (1981) Une théorie combinatoire des séries formelles, *Advances in Mathematics* **42**, 1–82.

[14] Maassen, H. and Bezembinder, T. (2002) Generating random weak orders and the probability of a Condorcet winner. *Soc. Choice Welfare* **19** 517 – 532.

[15] Odlyzko, A. M. (1995) Asymptotic enumeration methods. *Handbook of combinatorics, Vol. 1, 2*, 1063 – 1229, Elsevier, Amsterdam.

[16] Pittel, B. (1997) Random set partitions: asymptotics of subset counts. *J. Combin. Theory, Ser. A* **79** 326 – 359.