Cycle lengths in a permutation are typically Poisson

Andrew Granville*

Départment de mathématiques et de statistique
Université de Montréal, Montréal QC H3C 3J7, Canada
andrew@dms.umontreal.ca

Submitted: May 3, 2006; Accepted: Nov 10, 2006; Published: Nov 17, 2006
Mathematics Subject Classifications: Primary 62E20; Secondary 62E17, 05A16.

Abstract

The set of cycle lengths of almost all permutations in S_n are "Poisson distributed": we show that this remains true even when we restrict the number of cycles in the permutation. The formulas we develop allow us to also show that almost all permutations with a given number of cycles have a certain "normal order" (in the spirit of the Erdős-Turán theorem). Our results were inspired by analogous questions about the size of the prime divisors of "typical" integers.

1 Introduction

Define S_n to be the set of permutations on n letters, and let $\ell(\sigma)$ be the number of cycles of $\sigma \in S_n$. It is well-known that

$$\ell(\sigma) \sim \log n \text{ for almost all } \sigma \in S_n$$

(a fact we will reprove in Section 2). More precisely we mean that for any $\delta, \epsilon > 0$ if n is sufficiently large then $(1+\delta)\log n > \ell(\sigma) > (1-\delta)\log n$ for all but at most $\epsilon\ n!$ permutations $\sigma \in S_n$.

Write $\sigma = C_1 C_2 \cdots C_\ell$ where the C_i's are cycles and $\ell = \ell(\sigma)$, and let $d_i(\sigma) = d(C_i)$ be the number of elements of C_i. We may order the cycles so that

$$1 \leq d_1(\sigma) \leq d_2(\sigma) \leq \cdots \leq d_\ell(\sigma) \leq n$$

and therefore

$$0 \leq \log d_1(\sigma) \leq \log d_2(\sigma) \leq \cdots \leq \log d_\ell(\sigma) \leq \log n.$$

Thus, for almost all $\sigma \in S_n$ we have $\sim \log n$ numbers $\log d_i(\sigma)$ in an interval $[0, \log n]$ of length $\log n$. How are these numbers distributed within the interval? Other than near the

*L'auteur est partiellement soutenu par une bourse de la CRSNG du Canada.

beginning and end of the interval we might, for want of a better idea, guess that these numbers are "randomly distributed" in some appropriate sense, given that the average gap is 1. That guess, correctly formulated, turns out to be correct. In probability theory one uses the notion of a "Poisson point process" when one wishes to show that the event times of a random variable are "randomly distributed". However, in our question we do not have random variables. Indeed the set of permutations on n letters are pre-determined, as are their cycle lengths, so we need to create an analogy of the Poisson point process for this non-random situation. A little loosely we proceed as follows:

A sequence of finite sets $S_1, S_2, \cdots$ is called "Poisson distributed" if there exist functions $m_j, K_j, L_j \to \infty$ monotonically as $j \to \infty$ such that $S_j \subseteq [0, m_j]$ and $|S_j| \sim m_j$; and for all $\lambda,\ 1/L_j \leq \lambda \leq L_j$ and integers k in the range $0 \leq k \leq K_j$ we have

$$\frac{1}{m_j} \int\limits_{\substack{0 \\ \#\{S_j \cap [t, t+\lambda]\} = k}}^{m_j} 1 \ \mathrm{d}t \sim e^{-\lambda} \frac{\lambda^k}{k!}.$$

For example if each S_m is a set of m real numbers chosen uniformly and independently in the interval $[0, m]$, then this sequence of sets is almost surely Poisson distributed.

With this definition we prove in section 4 the following result (which can also be deduced from the much stronger theorem of DeLaurentis and Pittel [4]):

Theorem 1 *As $n \to \infty$, the sets of numbers*

$$D_\sigma := \{\log d_1(\sigma), \log d_2(\sigma), \cdots \log d_\ell(\sigma)\}$$

are Poisson distributed, for almost all $\sigma \in S_n$.

The precise statement of what we prove is: *There exist functions $K(n), L(n) \to \infty$ as $n \to \infty$ such that for all $\epsilon > 0$, if n is sufficiently large (depending on ϵ) then we have*

$$(1 - \epsilon) e^{-\lambda} \frac{\lambda^k}{k!} \leq \frac{1}{\log n} \int\limits_{\substack{t=0 \\ \#\{D_\sigma \cap [t, t+\lambda]\} = k}}^{\log n} 1 \ \mathrm{d}t \leq (1 + \epsilon) e^{-\lambda} \frac{\lambda^k}{k!}.$$

for any λ in the range $1/L(n) \leq \lambda \leq L(n)$ and any non-negative integer $k \leq K(n)$, for at least $(1 - \epsilon)\ n!$ elements $\sigma \in S_n$.

Notice that if for each integer $m \geq 1$ we select a permutation $\sigma_m \in S_m$ at random (that is, each permutation is selected with probability $1/m!$) then Theorem 1 implies that the sequence of sets $D_{\sigma_m},\ m \geq 1$ is almost surely Poisson distributed.

Evidently D_σ can only be distributed as in Theorem 1 if $\ell(\sigma) \sim \log n$. So what happens if $\ell(\sigma)$ is considerably smaller or larger? In other words, if we fix $\ell, 1 \leq \ell \leq n$ then what do the sets D_σ typically look like when we consider those $\sigma \in S_n$ with $\ell(\sigma) = \ell$? In this case, the average gap between elements is $(\log n)/\ell$ so we might expect a Poisson distribution with this parameter. However, there are three obvious problems with this guess:

• If ℓ is bounded then there cannot be a non-discrete distribution function for gaps between elements of D_σ for each individual σ since there are a bounded number of elements of D_σ. We deal with this relatively easy case separately and find the following in section 3.3:

Theorem 2 *For large n and $2 \le \ell \le \frac{1}{2}\log\log n$ consider $S_{n,\ell}$ the set of $\sigma \in S_n$ with $\ell(\sigma) = \ell$. The distribution of the points*

$$\{\log d_i(\sigma)/\log n : \quad 1 \le i \le \ell - 1\}$$

on $(0,1)$ as we vary over $\sigma \in S_{n,\ell}$, is the same as the distribution of $\ell - 1$ numbers chosen independently at random with uniform distribution on $(0,1)$. More precisely, for any ϵ in the range $1/\ell > \epsilon > (e/\ell)(\ell/\log n)^{1/(\ell-1)}$, for any $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{\ell-1} \le \alpha_\ell = 1$ with $\alpha_{j+1} - \alpha_j > \epsilon$, there are $(\ell-1)!\epsilon^{\ell-1}\{1 + O(\ell/\log n)\}|S_{n,\ell}|$ elements $\sigma \in S_{n,\ell}$ with $\log d_i(\sigma)/\log n \in (\alpha_i, \alpha_i + \epsilon)$ for each $1 \le i \le \ell - 1$.

• Since we are modelling D_σ with a continuous distribution function, it should be very unlikely that there are repeated values in D_σ. However, in Proposition 1 below we prove that there are $\sim \ell/m\nu$ cycles of length m in σ, for almost all $\sigma \in S_{n,\ell}$ whenever $m = o(\min\{\ell/\nu, n/(\ell/\nu)\})$ where, here and henceforth,

$$\frac{e^\nu - 1}{\nu} = \frac{n}{\ell}.$$

Therefore if $\ell \le n^{1/2-\epsilon}$ then we have this "discrete spectrum" for cycle lengths up to around ℓ/ν, containing a total of $\sim (\ell/\nu)\log(\ell/\nu)$ cycles. Since $\nu \sim \log(n/\ell)$ this is $o(\ell)$ if $\ell = n^{o(1)}$, in which case these cycles are irrelevant in our statistical investigation. If ℓ is bigger, say $\ell = n^{\alpha+o(1)}$ with $\alpha < 1/2$, then there are $\sim (\alpha/(1-\alpha))\ell$ cycles in this discrete spectrum.

• We cannot have many i with $d_i(\sigma) > (n/\ell)\log(n/\ell)$: in fact, evidently no more than $\ell/\log(n/\ell) = o(\ell)$ if $\ell = o(n)$.

From these last two points we see that we should restrict our attention to cycle lengths in the interval $[\ell, (n/\ell)\log(n/\ell)]$. Notice that the average gap between the logarithm of cycle lengths in this interval is $\sim \log(n/\ell)/\ell$, provided $\ell \ll n^{1/2-\epsilon}$. Therefore we will prove in section 5 (by modifying the proof of Theorem 1):

Theorem 3 *Given ℓ and n with $\ell, n/\ell \to \infty$ and $\ell \ll n^{1/2-\epsilon}$ consider $S_{n,\ell}$ the set of $\sigma \in S_n$ with $\ell(\sigma) = \ell$. Almost all $\sigma \in S_{n,\ell}$ contain $\sim \ell/m\nu$ cycles of length m, for almost all $m = o(\ell/v)$. Moreover the elements of the set*

$$D_{\sigma,\ell} := \{(\log d_i(\sigma))/(\log(n/\ell)/\ell) : \ \log d_i(\sigma) \in D_\sigma, \text{ and } \ell \le d_i(\sigma) \le (n/\ell)\log(n/\ell)\}$$

are Poisson distributed for almost all $\sigma \in S_{n,\ell}$.

When $\ell \gg n^{1/2+\epsilon}$ almost all cycles have length $< n/\ell$; indeed almost all $\sigma \in S_{n,\ell}$ contain $\sim \ell/m\nu$ cycles of length m, for almost all $m \leq n/\ell$ (by Proposition 1 below). This cannot be modelled by any continuous distribution function.

Theorem 3 is proved by incorporating precise estimates on Stirling numbers of the first kind (as proved in section 3) into the proof of Theorem 1. In reviewing the literature we found that these estimates allowed us to generalize one of the first results of statistical group theory: Erdős and Turán [5] proved that almost all $\sigma \in S_n$ have order $\exp(\{\frac{1}{2} + o(1)\} \log^2 n)$. This follows easily from our Theorem 1: The order of σ is given by $\mathrm{lcm}[d_1(\sigma), d_2(\sigma), \ldots, d_\ell(\sigma)]$. By Theorem 1 we know that $\log(d_1(\sigma)d_2(\sigma)\ldots d_\ell(\sigma)) \sim \frac{1}{2}\log^2 n$, moreover a number theorist knows that $\log n$ "random integers" up to n, where m chosen with probability $1/m$, are unlikely to have many large common factors, and thus the result: we formalize this last step in section 6 to complete the proof. Moreover, from the estimates used to prove Theorem 3 it is not difficult to deduce the following generalization by the same type of proof:

Theorem 4 *Suppose that $k \to \infty$ and $\log(n/k^2)/\log\log n \to \infty$ as $n \to \infty$. Then almost all $\sigma \in S_{n,k}$ have order*

$$\exp\left(\left\{\frac{1}{2} + o(1)\right\} k \frac{\log n \log(n/k^2)}{\log(n/k)}\right).$$

After proving this in section 6 we also prove that if $\log(k^2/n)/\log\log n \to \infty$ as $n \to \infty$ with $k \ll n/(\log n)^C$, then almost all $\sigma \in S_{n,k}$ have order

$$\exp\left(\{1 + o(1)\}\frac{n}{k}\log(n/k)\log(k^2/n)\right).$$

These results are given more precisely in section 6.4. However an interesting range remains to be understood, where $k = \sqrt{n}(\log n)^{O(1)}$. It is evident that there is a transition between these two types of estimates (in fact the transition occurs as k runs through multiples of $\sqrt{n}\log n$), but I have been unable to obtain satisfactory results in this range.

There have been many recent developments in number theory and combinatorics examining the distributions of sets of eigenvalues and zeros, and of natural invariants of permutations (for example, the "largest increasing subsequence" of a permutation). It struck me that there are various "spectra" in multiplicative number theory that had not been properly investigated, for example the set of all prime divisors of a given integer: Hardy and Ramanujan showed that almost all integers have $\sim \log\log x$ prime factors, and it has been shown that if $p_j(n)$ is the jth smallest prime factor of an integer then $\log\log p_j(n)$ is "randomly distributed" with mean j, for a certain range of j, as we vary over all integers n. Nonetheless the literature seems to lack an investigation of all of the prime factors of n taken together, and in particular whether $\{\log\log p : p|n\}$ is "Poisson distributed" on $[0, \log\log n]$, something we prove in a companion paper to this. In fact having proved this we started to wonder whether one can prove analogous results about the distribution of $\{\log\log p : p|n\}$ for integers n with exactly k prime factors for values of k in an appropriate range. We found that we could only prove such a result in the limited range $k = (\log n)^{o(1)}$, and we wished to better understand the obstructions to extending our proof.

Arratia, Barbour and Tavaré [1] explained how certain aspects of the distribution of cycle lengths in a random permutation are analogous to the distribution of prime divisors of random integers (and see Billingsley [2] and Knuth and Trabb Prado [10]). I thought that maybe I should try to work out the analogous results for permutations, which should be substantially easier, and hopefully be able to identify the obstructions to my earlier proof in this new context. Thus Theorem 1 here is the analogy to the result I had already proved about almost all integers, and working with exactly k cycles is analogous to working with integers with exactly k prime factors. The discussion of the restriction of the domain preceding the statement of Theorem 3 is indeed precisely what I was hoping to find in this auxiliary investigation, and I have subsequently proved all that I was hoping to prove about the distribution of prime divisors of integers (see [7]).

In the course of this research I have determined several more analogies between the distribution of prime factors of integers and the distribution of cycle lengths in a permutation, something I will discuss in detail in a further paper (see [8]). It may well be that such results will allow us new insights into the structure of factorization of integers.

I believe it would be interesting to try to develop similar results to Theorem 1 for other infinite families of groups. Obviously one will obtain much the same results for finite index subgroups of S_n, but how about for other classical families?

It may well be that Theorems 1, 2 and 3 can be proved more easily in the spirit of the ideas discussed in Shepp and Lloyd [13] (and thence Arratia, Barbour and Tavaré [1]), since the distribution of cycle lengths in permutations follows a Poisson-Dirichlet distribution (and the questions above involve aspects of that distribution, conditioning on certain linear equations). However to do so, one would need to show that this distribution holds here with a high level of uniformity and I have been unable to determine whether this can be deduced from the existing literature.

Acknowledgements: On hearing a delightful proof, Paul Erdős would say that we have been allowed to glimpse "The Book" in which the "supreme being" records the most elegant proofs of each theorem. I would like to thank Rod Canfield for sharing with me his delicious proof of (3.1) which I sketch there, a proof that, if not itself in "The Book", must at least appear in the pirated version! Thanks also to the referee for help in putting a few phantoms to rest.

2 Poisson and Permutations

For $\sigma \in S_n$ let $C(\sigma)$ be the set of cycles of σ of degrees

$$1 \leq d_1(\sigma) \leq d_2(\sigma) \leq \cdots \leq d_k(\sigma) \leq n,$$

where $k = k(\sigma)$, the number of cycles of σ. The expected number of cycles of length m in σ is

$$\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{\substack{C\in C(\sigma)\\ d(C)=m}} 1 = \frac{1}{n!}\sum_{C:\ d(C)=m}\ \sum_{\substack{\sigma\in S_n\\ C\in C(\sigma)}} 1 = \sum_{m=1}^{n}\frac{n\ldots(n+1-m)}{m}\cdot\frac{(n-m)!}{n!} = \frac{1}{m},$$

so that the expected length of the union of the cycles of length m in $\sigma \in S_n$, is 1. We deduce that the expected value of $k(\sigma)$ is $(1/n!)\sum_{\sigma\in S_n} k(\sigma) = \sum_{m=1}^{n}\frac{1}{m}\ := \mu_n$. Moreover

$$\frac{1}{n!}\sum_{\sigma\in S_n} k(\sigma)^2 = \sum_{j=1}^{n}\frac{1}{j} + \sum_{C_1,C_2\ \text{disjoint cycles}}\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ C_1,C_2\in C(\sigma)}} 1 = \sum_{j=1}^{n}\frac{1}{j} + \sum_{\substack{1\le i,j\\ i+j\le n}}\frac{1}{ij},$$

so that

$$\frac{1}{n!}\sum_{\sigma\in S_n}(k(\sigma)-\mu_n)^2 = \sum_{j=1}^{n}\frac{1}{j} - \sum_{k=n+1}^{2n}\sum_{\substack{1\le i,j\le n\\ i+j=k}}\frac{1}{ij} \le \mu_n,$$

where $\mu_n = \log n + \gamma + O(1/n)$. Thus $k(\sigma)$ has normal order μ_n for $\sigma \in S_n$. In fact Feller [6] elegantly showed that $k(\sigma)$ is normally distributed with mean μ_n and variance $\sim \mu_n$, a result we will reprove in a stronger form below.

Lemma 1 *For any $A > 0$ we have*

$$\sum_{r\ge m}\frac{A^r}{r!} \le \frac{1}{e^{A+m}}$$

provided $m \ge 2 + 25A/3$.

Proof. Since $m \ge 2A$,

$$\frac{A^{r+1}}{(r+1)!} \le \frac{A}{m+1}\frac{A^r}{r!} \le \frac{1}{2}\frac{A^r}{r!}$$

and so

$$\sum_{r\ge m}\frac{A^r}{r!} \le 2\frac{A^m}{m!} \le \frac{2}{3}\left(\frac{eA}{m}\right)^m$$

by Stirling's formula, in the form $m! \ge 3(m/e)^m$ for all integers $m \ge 2$; and the result follows.

Let $k_m(\sigma)$ be the number of cycles of length $\le m$ in σ. Then

$$\begin{aligned}
\frac{1}{|S_n|}\sum_{\sigma\in S_n}\binom{k_m(\sigma)}{r} &= \sum_{\substack{C_1,\ldots,C_r\in S_n\\ \ell(C_i)\le m}}\frac{1}{n!}\sum_{\substack{\sigma\in S_n\\ C_1,\ldots,C_r\in\sigma}} 1\\
&= \sum_{\substack{a_1+\cdots+a_m=r\\ a_1+2a_2+\cdots+ma_m\le n}}\frac{1}{a_1!1^{a_1}a_2!2^{a_2}\ldots a_m!m^{a_m}}\\
&\le \text{coefficient of } x^r \text{ in } \exp\left(x+\frac{x}{2}+\cdot+\frac{x}{m}\right) = \frac{\mu_m^r}{r!},
\end{aligned}$$

and equality holds if $rm \le n$. Therefore the proportion of permutations in S_n with no cycles of length $\le m$ is, by the inclusion-exclusion principle,

$$\begin{aligned}\sum_{r\ge 0}(-1)^r \frac{1}{|S_n|}\sum_{\sigma\in S_n}\binom{k_m(\sigma)}{r} &= \sum_{r\ge 0}(-1)^r\frac{\mu_m^r}{r!}+O\left(\sum_{r>n/m}\frac{\mu_m^r}{r!}\right)\\ &= e^{-\mu_m}+O(2^{-n/m}) \qquad (2.1)\end{aligned}$$

provided $n \ge 2em\mu_m$, by Lemma 1. Since $\mu_m = \log m + \gamma + O(1/m)$ the quantity in (2.1) equals

$$e^{-\gamma}/m + O(1/m^2) \qquad (2.2)$$

in this range.

The above also implies that $k_m(\sigma)$ is Poisson distributed with Poisson parameter μ_m. This holds uniformly for $m \ll n/\log n$. Since the average number of cycles of length $\gg n/\log n$ is $\log\log n + O(1)$, we deduce a rather strong version of Feller's result that $k(\sigma)$ is normally distributed with mean and variance $\sim \log n$.

The Buchstab function $\omega(u)$ is defined by $\omega(u) = 0$ for $0 < u < 1$,

$$\omega(u) = 1/u \quad \text{for } 1 \le u \le 2$$

and

$$u\omega(u) = \int_0^{u-1}\omega(t)\ \mathrm{d}t \quad \text{for all } u > 2.$$

It is known that $\omega(u) \to e^{-\gamma}$ as $u \to \infty$; in fact $\omega(u) = e^{-\gamma} + O(1/u^2)$. We prove

Theorem 5 *Define $A(n,m)$ to be the number of permutations on n letters all of whose cycles have length $\ge m$. Then*

$$\frac{A(n,m)}{n!} = \frac{\omega(n/m)}{m} + O\left(\frac{\log\log m}{m^2}\right)$$

Proof. Define $a(n,m) = mA(n,m)/n!$ and $\Delta(n/m) = a(n,m) - \omega(n/m)$. Now

$$\begin{aligned} nA(n,m) &= \sum_{\substack{\sigma\in S_n\\ C\in\sigma\Rightarrow \ell(C)\ge m}}\ \sum_{C\in\sigma}\ell(C) = \sum_{b=m}^{n} b \sum_{\substack{C\in S_n\\ \ell(C)=b}} 1 \sum_{\substack{\sigma\in S_n,\ C\in\sigma\\ C'\in\sigma\Rightarrow\ell(C')\ge m}} 1\\ &= \sum_{b=m}^{n} b\ \frac{n!}{(n-b)!}\ \frac{1}{b}\ A(n-b,m)\end{aligned}$$

and so, taking $r = n-b$,

$$a(n,m) = \frac{1}{n}\sum_{r=0}^{n-m} a(r,m). \qquad (2.3)$$

Note that $A(0,m)=1, A(n,m)=0$ if $1\le n\le m-1$ and $A(n,m)=A(n,n)=n!/n$ if $m\le n\le 2m-1$. Therefore

$$\Delta(u)=0 \quad \text{for } 0<u<2$$

(when u is of the form n/m). Now by (2.3), whenever $n\le N-m$,

$$\begin{aligned}(N+m)a(N+m,m)-(n+m)a(n+m,m) &= \sum_{r=n+1}^{N} a(r,m)\\ &= \sum_{r=n+1}^{N}\omega(r/m)+\sum_{r=n+1}^{N}\Delta(r/m).\end{aligned}$$

The latter term is $\le (N-n)\max_{n/m<t\le N/m}|\Delta(t)|$; and so writing $u=N/m+1$ and $v=n/m$ with $v\le u-2$,

$$|\Delta(u)|\le \max_{v<t\le u-1}|\Delta(t)|+\frac{1}{um}\left|\sum_{r=n+1}^{N}\omega(r/m)-\int_n^N w\left(\frac{t}{m}\right)\mathrm{d}t\right|. \tag{2.4}$$

Now Maier [11]) showed that $\omega'(t)$ changes sign $O(1)$ times in any interval of length 1; and so

$$\left|\sum_{r=n+1}^{N}w(r/m)-\int_n^N w\left(\frac{t}{m}\right)\mathrm{d}t\right|\ll 1$$

since $\omega(u)=e^{-\gamma}+O(1/u^2)$. With $v=u-2$, (2.4) becomes $|\Delta(u)|\le\Delta^*(u-1)+O(1/um)$ where $\Delta^*(u):=\max_{0<t\le u}|\Delta(t)|$. Therefore, for $u\ge 2$,

$$\Delta^*(u)\le\Delta^*(u-1)+O(1/um)\ll(\log u)/m$$

by induction. This gives the theorem for $u<\log^2 m$ and (2.2) does so for $u\gg\log m$.

3 Asymptotics for quotients of neighboring Stirling numbers of the first kind

$S(n,k)$, the Stirling numbers of the first kind, are defined as the size of $S_{n,k}$, the set of $\sigma\in S_n$ with exactly k cycles. Moser and Wyman [12] proved the following estimate for $S(n,k)$ when k and $n/k\to\infty$ as $n\to\infty$: Define $T=T(n,k)$ so that

$$\sum_{i=0}^{n-1}\frac{T}{T+i}=k, \quad \text{and let} \quad \ell=k-\sum_{i=0}^{n-1}\frac{T^2}{(T+i)^2}.$$

Then

$$S(n,k) = \frac{\Gamma(n+T)}{\Gamma(T)} \frac{1}{(2\pi\ell)^{1/2}} \frac{1}{T^k} \left\{ 1 + O\left(\frac{1}{\ell}\right) \right\}. \tag{3.1}$$

Proof from "The Book" (see Canfield [3]) Let $X_0, X_1, \ldots$ be independent (binomial) random variables with $\text{Prob}(X_i = 1) = T/(T+i)$ and $\text{Prob}(X_i = 0) = i/(T+i)$, where T is chosen as above so that $\mathbb{E}(X_0 + X_1 + \cdots + X_{n-1}) = k$. By the central limit theorem we know that the random variable $X_0 + X_1 + \cdots + X_{n-1}$ satisfies a Poisson type distribution with mean k and variance

$$\sum_{i=0}^{n-1} (\mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2) = \sum_{i=0}^{n-1} \left(\frac{T}{T+i} - \left(\frac{T}{T+i}\right)^2 \right) = \ell;$$

therefore $\text{Prob}(X_0 + X_1 + \cdots + X_{n-1} = k) \approx 1/(2\pi\ell)^{1/2}$. On the other hand $\text{Prob}(X_0 + X_1 + \cdots + X_{n-1} = k)$ equals the coefficient of X^k in

$$\prod_{i=0}^{n-1} \left(\frac{TX+i}{T+i} \right) = \frac{\Gamma(T)}{\Gamma(n+T)} \sum_{k \geq 0} S(n,k) T^k X^k,$$

and the result follows, being more precise about the "$\approx$".

We need the following consequence of (3.1): If $k, m = o(n)$ and $k \to \infty$, with $1 \leq m \ll (n/k)\log(n/k)$ and $r \ll \min\{\sqrt{k}, \log(n/k)\}$ then

$$\frac{S(n-m,k-r)}{S(n,k)} = \frac{(n-m)!}{n!} \left(\frac{k}{\nu}\right)^r \left\{ 1 + O\left(\frac{r^2}{k} + \frac{m}{\frac{n}{k}\log(n/k)} + \frac{1}{\log(n/k)} + \frac{m}{n} \right) \right\} \tag{3.2}$$

where ν satisfies $e^v - 1 = v(n/k)$.

Proof. Note that $v \to \infty$ in our range as $n \to \infty$. Now

$$k = \sum_{i=0}^{n-1} \frac{T}{T+i} = 1 + T \sum_{i=1}^{n-1} \frac{1}{T+i} = T \log\left(\frac{n+T}{1+T}\right) + O(1)$$

so that for $K = k + O(1)$ we have $e^{K/T} - 1 = (n-1)/(1+T)$ from which one can deduce that $T = \{k + O(1)\}/v = n(1 + O(1/k))/(e^v - 1)$. Moreover

$$\begin{aligned} \ell = k - T^2(1/T - 1/(T+n) + O(1/T^2)) &= k - nT/(T+n) + O(1) \\ &= k \left\{ 1 + O\left(\frac{1}{v} + \frac{1}{k}\right) \right\} \end{aligned}$$

We wish to compare this with v', T' and ℓ' which come from replacing n and k by $n-m$ and $k-r$. Note that $v' = v + O(m/n + r/k) = v + o(1) \sim v$, and $\ell' = \ell(1 + O(\frac{r}{k} + \frac{1}{v}))$. Define

$$g_n(t) := \sum_{i=0}^{n-1} \frac{t}{t+i}$$

so that $g_n(T) = k$ and $g_{n-m}(T') = k - r$. Since $g'_n(t) \sim g_n(t)/t$ for all $t \sim T$, and $g_{n-m}(t) - g_n(t) \sim -mt/n$ in our range thus

$$|T' - T| \ll \frac{T}{k}\left(\frac{mT}{n} + r\right) \ll \frac{r}{v}. \tag{3.3}$$

Let τ be the integer nearest to T. Using (3.1) and results above we have

$$\begin{aligned}\frac{S(n-m,k-r)}{S(n,k)} = \frac{\Gamma(n+\tau)}{\Gamma(n+T)}\frac{\Gamma(n-m+T')}{\Gamma(n-m+\tau)} \quad &\cdot \quad \frac{\Gamma(T)}{\Gamma(T')}\frac{\Gamma(n+1)}{\Gamma(n+\tau)}\frac{\Gamma(n-m+\tau)}{\Gamma(n-m+1)}\frac{(n-m)!}{n!} \\ &\cdot \quad \left(\frac{T}{T'}\right)^{k-r}(T)^r\left\{1 + O\left(\frac{r}{k} + \frac{1}{v}\right)\right\}\end{aligned}$$

Now

$$\begin{aligned}\frac{\Gamma(n+1)}{\Gamma(n+\tau)}\frac{\Gamma(n-m+\tau)}{\Gamma(n-m+1)} &= \prod_{j=1}^{\tau-1}\left(\frac{n-m+j}{n+j}\right) = \left(1 + O\left(\frac{m}{n}\right)\right)^T \\ &= 1 + O\left(\frac{mT}{n}\right) = 1 + O\left(\frac{m}{e^v}\right).\end{aligned}$$

Also if t is large and $|\delta| \ll 1$ then

$$\log\Gamma(t+\delta) - \log\Gamma(t) = \delta\log t + O\left(\frac{\delta}{t}\right)$$

so that

$$\begin{aligned}\log\left(\frac{\Gamma(n+\tau)}{\Gamma(n+T)}\cdot\frac{\Gamma(n-m+T')}{\Gamma(n-m+\tau)}\cdot\frac{\Gamma(T)}{\Gamma(T')}\right) &= (\tau - T)\log(n+T) \\ &+ (T' - \tau)\log(n-m+T') \\ &+ (T - T')\log T + O\left(\frac{1}{n} + \frac{|T-T'|}{T}\right) \\ &= (T' - T)\log(n/T) + O\left(\frac{m+T}{n} + \frac{r}{k}\right)\end{aligned}$$

by (3.3), and

$$\begin{aligned}\log((T/T')^{k-r}) &= -(k-r)\log\left(1 + \frac{T'-T}{T}\right) = (T-T')\frac{(k-r)}{T} + O\left(k\frac{(T'-T)^2}{T^2}\right) \\ &= \frac{k(T-T')}{T} + O\left(\frac{r^2}{k}\right).\end{aligned}$$

by (3.3). Also

$$
\begin{aligned}
(T'-T)\log(n/T)+\frac{k(T-T')}{T} &= (T'-T)\left(\left(\log(e^v-1)-\frac{k}{T}\right)+\log(k/vT)\right)\\
&\ll \frac{r}{v}\left(\frac{1}{e^v}+\frac{v}{k}\right)\ll \frac{1}{e^v}+\frac{r}{k}
\end{aligned}
$$

since $r\ll v$. Combining these estimates together gives (3.2).

3.2. A consequence

Proposition 1 *If $k\to\infty, k=o(n)$ and $m\ll\min\{k/v, nv/k\}$ where v is the solution to $e^v-1=v(n/k)$ then*

$$\frac{1}{|S_{n,k}|}\sum_{\sigma\in S_{n,k}}\left|\sum_{\substack{C\in\sigma\\|C|=m}}1-\frac{k}{mv}\right|^2\ll\left(\frac{k}{mv}\right)^2\left\{\frac{1}{v}+\frac{m}{k/v}+\frac{m}{n/(k/v)}\right\}.$$

Proof: The mean value for the number of cycles of length m in $\sigma\in S_{n,k}$ is given by

$$
\begin{aligned}
\frac{1}{|S_{n,k}|}\sum_{\sigma\in S_{n,k}}\sum_{\substack{C\in\sigma\\|C|=m}}1 &= \sum_{\substack{|C|=m\\C\in S_n}}\frac{1}{|S_{n,k}|}\sum_{\substack{\sigma\in S_{n,k}\\C\in\sigma}}1\\
&= \frac{1}{m}\frac{S(n-m,k-1)/(n-m)!}{S(n,k)/n!}\\
&= \frac{k}{mv}\left\{1+O\left(\frac{1}{k}+\frac{1}{v}+\frac{m}{(n/k)\log(n/k)}+\frac{m}{n}\right)\right\}
\end{aligned}
$$

by (3.2), provided $m\ll(n/k)\log(n/k), k\to\infty$ and $k,m=o(n)$. The mean square is, in the same range.

$$
\begin{aligned}
\frac{1}{|S_{n,k}|}\sum_{\sigma\in S_{n,k}}\left(\sum_{\substack{C\in\sigma\\|C|=m}}1\right)^2 &= \frac{1}{S(n,k)}\sum_{\substack{C_1,C_2\in S_n\\|C_1|=|C_2|=m}}\sum_{\substack{\sigma\in S_{n,k}\\C_1,C_2\in\sigma}}1\\
&= \frac{1}{m}\frac{S(n-m,k-1)/(n-m)!}{S(n,k)/n!}\\
&\qquad+\frac{1}{m^2}\frac{S(n-2m,k-2)/(n-2m)!}{S(n,k)/n!}\\
&= \left(\frac{k}{mv}\right)^2\left\{1+O\left(\frac{mv}{k}+\frac{1}{v}+\frac{m}{(n/k)\log(n/k)}\right)\right\}
\end{aligned}
$$

since if $(e^{v'}-1)/v'=(n-m)/(k-1)$ then $v'=v+O(\frac{1}{k}+\frac{m}{n})$. The result follows.

3.3. Upper bounds

In this section we obtain the following upper bound on $S(n-M,k-R)$ using (3.2): In the range

$$R \ll M \leq nR/2k, \quad k=o(n), \quad 1 \leq R \leq k-1$$

we have

$$\frac{S(n-M,k-R)}{S(n,k)} \ll \frac{(n-M)!}{n!}\left(\frac{k}{v}\right)^R \exp\left(O\left(\frac{R}{\log(\frac{n}{k})}+\frac{R}{k/\log k}\right)-\frac{R^2}{2k}\right). \tag{3.4}$$

Jordan [9] showed that

$$S(n,k) \sim \frac{(n-1)!}{(k-1)!}(\log n+\gamma)^{k-1}$$

when $k=o(\log n)$ (which we reprove in (3.5) below). Therefore, for all $1 \leq R \leq k-1$, we have

$$\frac{S(n-M,k-R)}{S(n,k)} \sim \frac{(n-M)!}{n!}\prod_{i=1}^{R}\left(\frac{k-i}{\log n+\gamma}\right)\frac{n}{n-M},$$

which implies (3.4) in our range with $k=o(\log n)$.

When $k-R \to \infty$ we let $n_i = n-[iM/R]$ for $0 \leq i \leq R$, so that $n_0 = n$ and $n_R = n-M$. Let v_i be the solution to $e^{v_i}-1 = v_i(n_i/(k-i))$ We have

$$\frac{n_i}{k-i} \geq \frac{n-iM/R}{k-i} \geq \frac{n-in/k}{k-i} = \frac{n}{k}$$

and thus $v_i \geq v_0 = v$ for all i. Therefore

$$\begin{aligned}
\frac{S(n-M,k-R)}{S(n,k)} &= \prod_{i=1}^{R}\frac{S(n_i,k-i)}{S(n_{i-1},k-(i-1))} \\
&\leqq \frac{(n-M)!}{n!}\prod_{i=1}^{R}\left(\frac{k-i}{v_i}\right)\cdot \\
&\qquad \cdot \exp\left(O\left(\log\left(\frac{k}{k-R}\right)+\frac{M}{\frac{n}{k}\log\frac{n}{k}}+\frac{R}{\log(\frac{n}{k})}\right)\right)
\end{aligned}$$

since each $v_i \geq v$, which implies (3.4).

We can deduce (3.4) in the rest of our range, that is when $k-R \ll 1$, by writing $S(n-M,k-R)/S(n,k) = (S(n-M,k-R)/S(n_1,k_1))(S(n_1,k_1)/S(n,k))$ with $k_1 = [\sqrt{\log n}]$ and $n_1 = [n-M(k-k_1)/(k-R)]$ and using the result in the two ranges above to bound these two terms.

3.4. Stirling numbers $S(n,k)$ with k small, and Theorem 2.

$S(n,k)$ is the coefficient x^k in

$$\begin{aligned} x(x+1)\dots(x+n-1) &= (n-1)!x\prod_{j=1}^{n-1}(1+x/j) \\ &= (n-1)!x\exp\left(\sum_{j=1}^{n-1}\left(\frac{x}{j}+O\left(\frac{x^2}{j^2}\right)\right)\right) \\ &= (n-1)!x\exp\left(x\mu_{n-1}+O(x^2)\right), \end{aligned}$$

so that

$$S(n,k)=(n-1)!\frac{\mu_{n-1}^{k-1}}{(k-1)!}\exp(O(k^2/\log^2 n)), \tag{3.5}$$

an asymptotic estimate for $k=o(\log n)$. Note that, in particular, this gives

$$\frac{S(n-2j,k-2)/(n-2j)!}{S(n,k)/(n)!}=\frac{n}{n-2j}\frac{(k-1)(k-2)}{\log^2 n}\left\{1+O\left(\frac{k}{\log n}\right)\right\} \tag{3.6}$$

for $1\le j\le n/3$ and $k=o(\log n)$.

Suppose that $1\le j_1<j_2<\dots<j_{k-1}$ are intergers with

$$j_1+j_2+\dots+j_{k-2}+2j_{k-1}<n.$$

The number of $\sigma\in S_n$ with cycle lengths $j_1,\dots,j_{k-1},j_k(=n-j_1-j_2-\dots-j_{k-1})$ equals

$$n!/j_1j_2\dots j_k \tag{3.7}$$

This accounts for all $\sigma\in S_{n,k}$ except those in which there is a repeated cycle length. The number with a repeated cycle length is

$$\begin{aligned} &\le \sum_{j=1}^{[n/2]}\sum_{\substack{C_1,C_2\in S_n\\ |C_1|,|C_2|=j}}\#\{\sigma\in S_{n,k}:C_1,C_2\in\sigma\} \\ &\le \sum_{j=1}^{[n/2]}\frac{n!}{(n-2j)!}\cdot\frac{1}{2j^2}S(n-2j,k-2) \\ &\le \sum_{j=1}^{[n/3]}\left(\frac{n}{n-2j}\right)\frac{(k-1)(k-2)}{\log^2 n}\frac{S(n,k)}{2j^2}\left\{1+O\left(\frac{k}{\log n}\right)\right\}+\sum_{j=[n/3]+1}^{[n/2]}\frac{n!}{2j^2} \end{aligned}$$

using (3.6) and the trivial bound $S(n-2j,k-2)\le(n-2j)!$,

$$\ll S(n,k)\frac{k^2}{\log^2 n}+(n-1)!\leqq\frac{k^2}{\log^2 n}S(n,k)=o(S(n,k)). \tag{3.8}$$

for $3 \le k = o(\log n)$, using (3.5).

So select any $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{k-1} < 1$ with $\alpha_{i+1} - \alpha_i > \epsilon$ for each i (where $\alpha_0 = 0$ and $\alpha_k = 1$). Therefore the proportion of $\sigma \in S_{n,k}$ for which $\log d_i(\sigma)/\log n \in (\alpha_i, \alpha_i + \epsilon]$ for $1 \le i \le k-1$ equals, by (3.7) and (3.8),

$$\frac{n!}{S(n,k)} \sum_{\substack{n^{\alpha_1} < j_1 \le n^{\alpha_1+\epsilon} \\ n^{\alpha_2} < j_2 \le n^{\alpha_2+\epsilon} \\ n^{\alpha_{k-1}} < j_{k-1} \le n^{\alpha_{k-1}+\epsilon}}} \frac{1}{j_1 j_2 \ldots j_{k-1}(n - j_1 - \cdots - j_k)} + O\left(\frac{k^2}{\log^2 n}\right).$$

Now $j_1 + \cdots + j_{k-1} \le kn^{\alpha_{k-1}+\varepsilon}$ and so $(n - j_1 - \cdots - j_k) = n(1 + O(n^{-\delta}))$ for, say, $\delta = (1 - \varepsilon - \alpha_{k-1})/2$. Therefore, by (3.5), we get

$$(k-1)!\varepsilon^{k-1}\{1 + O(k/\log n)\} + O(k^2/\log^2 n)$$

which implies Theorem 2.

4 Poisson distribution of cycles

Taking M small and N large, say $1 \le M \le \sqrt{n} \le N \le n$, define

$$\mu_{r,L}(\sigma) = \frac{1}{\log(N/M)} \int_{\substack{t=\log M \\ \#\{D_\sigma \cap [t,t+L]\}=r}}^{\log N} 1 \ dt.$$

Let $m := [10 \log\log N/(\log\log\log N)^2]$ and assume $r \le m/10$. We will prove

$$\frac{1}{n!} \sum_{\sigma \in S_n} \left| \mu_{r,L}(\sigma) - \frac{e^{-L}L^r}{r!} \right| \le e^{-L} \frac{L^r}{r!} \frac{1}{2^m} \tag{4.1}$$

in the range

$$M \le N^{1-\epsilon}, \ L > \frac{1}{M} \log\log N \text{ and } r, L \le \frac{\log\log N}{(\log\log\log N)^2}. \tag{4.2}$$

Taking $M = (\log\log N)^2$ say, gives us Theorem 1.

In order to prove (4.1) we write

$$\sum_{r \ge R} \binom{R}{r} \mu_{r,L}(\sigma) = ((A+B)_{R,L}(\sigma))/\log(N/M)$$

where $(A+B)_{R,L}(\sigma) = A_{R,L}(\sigma) + B_{R,L}(\sigma)$ with

$$A_{R,L}(\sigma) = \sum_{\substack{C_1, C_2, \ldots, C_R \text{ disjoint cycles in } \sigma \\ \text{with } d(C_1) \le d(C_2) \le \cdots \le d(C_R) \le d(C_1)e^L \\ M \le d(C_1) \le N \text{ and } d(C_R) \le Ne^L}} \left\{ L - \log\left(\frac{d(C_R)}{d(C_1)}\right) \right\}$$

and

$$B_{R,L}(\sigma)=\sum_{\substack{C_1,C_2,\dots,C_R \text{ disjoint cycles in } \sigma\\ \text{with } N<d(C_1)\le d(C_2)\le\cdots\le d(C_R)\le Ne^L}}\left\{L-\log\left(\frac{d(C_R)}{N}\right)\right\}.$$

Now

$$\mu_{r,L}(\sigma)-\frac{e^{-L}L^r}{r!}=\sum_{R\ge r}(-1)^{R-r}\binom{R}{r}\left\{\frac{(A+B)_{R,L}(\sigma)}{\log N/M}-\frac{L^R}{R!}\right\}$$

so that

$$\frac{1}{n!}\sum_{\sigma\in S_n}\left|\mu_{r,L}(\sigma)-\frac{e^{-L}L^r}{r!}\right|\le\sum_{R\ge r}\binom{R}{r}\frac{1}{n!}\sum_{\sigma\in S_n}\left|\frac{(A+B)_{R,L}(\sigma)}{\log N/M}-\frac{L^R}{R!}\right|. \tag{4.3}$$

We will show that the overall contributions of the $B_{R,L}(\sigma)$ to the right side of (4.3) is negligible, as are the contributions of the terms $A_{R,L}(\sigma)$ with $R>m$. To bound the contributions of the remaining $A_{R,L}(\sigma)$ terms, namely those with $r\le R\le m$ we will use the Cauchy-Schwarz inequality and the bound

$$\frac{1}{n!}\sum_{\sigma\in S_n}\left|\frac{A_{R,L}(\sigma)}{\log N/M}-\frac{L^R}{R!}\right|^2\ll\frac{L^{2R}}{R!^2}\frac{1}{(\log N)^{1-o(1)}}, \tag{4.4}$$

which holds in this range, and thus obtain (4.1).

To begin with we determine the mean values of $A_{R,L}(\sigma)$ when $R\le m$ (though the method works in a somewhat wider range):

$$\begin{aligned}\frac{1}{n!}\sum_{\sigma\in S_n}A_{R,L}(\sigma) &= \sum_{\substack{C_1,\dots,C_R \text{ disjoint cycles in } S_n\\ \text{with } d(C_1)\le\cdots\le d(C_R)\le d(C_1)e^L\\ \text{and } M\le d(C_1)\le N}}\left\{L-\log\left(\frac{d(C_R)}{d(C_1)}\right)\right\}\frac{\left(n-\sum_{i=1}^R d(C_i)\right)!}{n!}\\ &= \sum_{\substack{M\le u\le N\\ u\le v\le ue^L}}\left(L-\log\left(\frac{v}{u}\right)\right)\sum_{\substack{C_1,\dots,C_R \text{ disjoint cycles in } S_n\\ \text{with } u=d(C_1)\le d(C_2)\le\cdots\le d(C_R)=v}}\frac{\left(n-\sum_{i=1}^R d(C_i)\right)!}{n!}\end{aligned} \tag{4.5}$$

In the final sum suppose there are g_j cycles of size j so that each g_j is a non-negative integer with $g_u,g_v\ge 1$, and $g_u+g_{u+1}+\cdots+g_v=R$. The number of choices of such cycles is

$$\frac{n!}{\left(n-\sum_{i=1}^k d(C_i)\right)!}\cdot\frac{1}{u^{g_u}g_u!}\cdot\frac{1}{(u+1)^{g_{u+1}}g_{u+1}!}\cdots\frac{1}{v^{g_v}g_v!}.$$

Thus the final sum in (4.5) is

$$\frac{1}{R!}\left\{\left(\sum_{u\le j\le v}\frac{1}{j}\right)^R-\left(\sum_{u+1\le j\le v}\frac{1}{j}\right)^R-\left(\sum_{u\le j\le v-1}\frac{1}{j}\right)^R+\left(\sum_{u+1\le j\le v-1}\frac{1}{j}\right)^R\right\} \tag{4.6}$$

by the inclusion-exclusion principle, since we must have $g_u \geq 1$ and $g_v \geq 1$. It will be convenient to denote this formula by "$(4.6)_{R,u,v}$" for future use. The sum in $(4.6)_{R,u,v}$ is

$$\frac{1}{(R-2)!}\frac{\Sigma^{R-2}}{uv}\left(1+O\left(\frac{R}{\Sigma u}\right)\right) \text{ where } \Sigma := \sum_{u\leq j\leq v}\frac{1}{j} = \log(v/u)+O(1/u)$$

provided $R \leq 10u\Sigma$. If $R > 10u\Sigma$ then this is $\leq \Sigma^R/R! \ll (e/10u)^R \leq 1/(e^{10u\Sigma}u^R) \ll 1/((v/u)^{10u}u^R) \ll 1/(vu^{R-1})$.

Substituting this in above (but invalidly taking the first estimate when $R > 10u\Sigma$) gives a main term of

$$\begin{aligned}
&\frac{1}{(R-2)!}\sum_{M\leq u\leq N}\frac{1}{u}\sum_{u\leq v\leq ue^L}\left(L-\log\left(\frac{v}{u}\right)\right)\frac{1}{v}\left(\log\left(\frac{v}{u}\right)\right)^{R-2}\left(1+O\left(\frac{R}{u\log(v/u)}\right)\right)\\
=\;&\frac{1}{(R-2)!}\sum_{M\leq u\leq N}\frac{1}{u}\left(\int_{t=0}^{L}(L-t)t^{R-2}\left(1+O\left(\frac{R}{ut}\right)\right)\mathrm{d}t+O\left(\frac{L^{R-1}}{u}\right)\right)\\
=\;&\frac{L^R}{R!}\left(\log(N/M)+O\left(\frac{1}{M}\right)\right)+O\left(\frac{L^{R-1}}{M(R-2)!}\right)
\end{aligned}$$

taking $v = ue^t$. The error term from when $R > 10u\Sigma$ is

$$\ll \sum_{M\leq u\leq N}\sum_{u\leq v\leq ue^{R/u}}\frac{L}{vu^{R-1}} \ll LR\sum_{M\leq u\leq N}\frac{1}{u^R} \ll \frac{L}{M^{R-1}}.$$

Thus we get

$$\frac{1}{n!}\sum_{\sigma\in S_n}A_{R,L}(\sigma) = \frac{L^R}{R!}\left\{\log(N/M)+O\left(\frac{1}{M}+R\left(\frac{R}{eLM}+\left(\frac{R}{eLM}\right)^{R-1}\right)\right)\right\} \tag{4.7}$$

By the same methods the mean value of $B_{R,L}(\sigma)$ for $\sigma \in S_n$ is (neglecting the terms with $R > 10v\log(v/N)$)

$$\begin{aligned}
&\sum_{N<v\leq Ne^L}\left(L-\log\left(\frac{v}{N}\right)\right)\frac{1}{R!}\left(\left(\sum_{N<j\leq v}\frac{1}{j}\right)^R-\left(\sum_{N<j<v}\frac{1}{j}\right)^R\right)\\
=\;&\sum_{N<v\leq Ne^L}\left(L-\log\left(\frac{v}{N}\right)\right)\frac{1}{v}\,\frac{(\log(\frac{v}{N})+O(\frac{1}{N}))^{R-1}}{(R-1)!}\left(1+O\left(\frac{R}{v\log(v/N)}\right)\right)\\
=\;&\frac{L^{R+1}}{(R+1)!}\left(1+O\left(\frac{R}{LN}\right)\right),
\end{aligned}$$

and for the terms when $R > 10v\log(v/N)$, which belong to the interval with $N < v < N+R$, the error term is $\leq (RL/R!)(R/N)^R$, which is negligible. As $M = o(N)$ we deduce that $B_{R,L}(\sigma) = o(L^R\log(N/M)/R!)$ for almost all $\sigma \in S_n$.

To compute the second moment of $A_{R,L}(\sigma)$ we proceed analogously to (4.5) to obtain

$$\frac{1}{n!}\sum_{\sigma\in S_n} A_{R,L}(\sigma)^2 = \sideset{}{'}\sum_{\substack{C_1,\dots,C_R \text{ disjoint}\\ C_1',\dots,C_R' \text{ disjoint}}} \left(L-\log\left(\frac{d(C_R)}{d(C_1)}\right)\right)\left(L-\log\left(\frac{d(C_R')}{d(C_1')}\right)\right)\frac{(n-D)!}{n!} \tag{4.8}$$

where $\sum'$ denotes that $d(C_R)\le d(C_1)e^L$, $d(C_R')\le d(C_1')e^L$, each C_i is either disjoint from each C_j' or equal, and $D=\sum_{C\in B} d(C)$ where $B=\{C_1,C_2,\dots,C_R,C_1',\dots,C_R'\}$. In the sum we fix $d(C_1)=u, d(C_R)=v, d(C_1')=u', d(C_R')=v'$ for now. Suppose there are g_j cycles of size j in B for each j. We sum over all possible sets B with these parameters, so that this subsum of the right side of (4.8) becomes

$$\left(L-\log\left(\frac{v}{u}\right)\right)\left(L-\log\left(\frac{v'}{u'}\right)\right)\prod_j \frac{1}{g_j!j^{g_j}}$$

Therefore the non-zero terms of

$$\left|\frac{1}{n!}\sum_{\sigma\in S_n} A_{R,L}(\sigma)^2-\left(\frac{1}{n!}\sum_{\sigma\in S_n} A_{R,L}(\sigma)\right)^2\right| \tag{4.9}$$

come from

- Those B with $2R$ distinct elements, with $d(C_i)=d(C_j')$ for some i and j.
- Those B with less than $2R$ distinct elements (so that $C_i=C_j'$ for some i and j).

Either way $u\le d(C_i)=d(C_j')\le u'e^L$ so that $v\le u'e^{2L}$, and similarly $v'\le ue^{2L}$. We now determine the contribution of such terms:

If B is a given set of $(R+i)$ disjoint cycles in S_n, with $0\le i\le R-1$, then $C_1,\dots,C_R\subseteq B$ can be selected in $\binom{R+i}{R}$ ways. Therefore $B\setminus\{C_1,\dots,C_R\}\subseteq\{C_1',\dots,C_R'\}$; that is, i elements of this set are predetermined, and thus the final $R-i$ elements may be chosen from $\{C_1,\dots,C_R\}$, so in $\binom{R}{R-i}$ ways. Taking $u=\min_{C\in B} d(C)$, so that $v=\max_{C\in B} d(C)\le ue^{2L}$, the total contribution of such sets to (4.9) is

$$\begin{aligned}
&\le L\sum_{M\le u\le N}\sum_{u<v<ue^{2L}}\binom{R+i}{R}\binom{R}{i}\left(L-\log\left(\frac{v}{u}\right)\right)\ (4.6)_{R+i,u,v}\\
&= L\frac{(2L)^{R+i}}{i!^2(R-i)!}\left\{\log(N/M)+O\left(\frac{1}{M}+R\left(\frac{R}{eLM}+\left(\frac{R}{eLM}\right)^{R+i-1}\right)\right)\right\}\\
&\ll (2R)^R\left(\frac{L^R}{R!}\right)^2\left\{\log(N/M)+O\left(\frac{1}{M}+R\left(\frac{R}{eLM}+\left(\frac{R}{eLM}\right)^{2R-2}\right)\right)\right\}
\end{aligned}$$

proceeding as in the proof of (4.7).

We now bound the contribution of sets B, with $2R$ distinct elements and $d(C_i)=d(C_j')$ for some i and j, to (4.9). Note that if $u'\ge u$ then $u'\le d(C_j')=d(C_i)\le ue^L$. In the first line below we have $u\le j\le ue^L$ in the first sum and $u'\le j\le u'e^L$ in the second sum;

which means that $u \le j \le ue^{2L}$ in either case, which is the range on j that we will use in subsequent lines.

$$
\begin{aligned}
&\leqq \quad 2L^2 \sum_{M\le u\le N} \sum_{u\le u'\le ue^{2L}} \sum_{\sum_j h_j = R} \prod_j \frac{1}{h_j! j^{h_j}} \sum_{\substack{h_i, h'_i \ge 1 \text{ for some } i \\ \sum_j h'_j = R}} \prod_j \frac{1}{h'_j! j^{h'_j}} \\
&\leqq \quad 2L^2 \sum_{1\le u\le N} \sum_{\sum g_j = 2R} \prod_j \frac{1}{j^{g_j} g_j!} \sum_{\substack{h_j + h'_j = g_j \text{ for each } j \\ h_i, h'_i \ge 1 \text{ for some } i}} \prod_i \binom{g_j}{h_j} \\
&\le \quad 2L^2 \sum_{M\le u\le N} \sum_{\substack{\sum g_j = 2R \\ \text{some } g_i \ge 2}} \prod_j \frac{2^{g_j}}{j^{g_j} g_j!} \\
&\le \quad 2^{2R+1} L^2 \sum_{M\le u\le N} \sum_{u\le i\le ue^{2L}} \frac{1}{i^2 2!} \frac{1}{(2R-2)!} \left(\sum_{u\le j\le ue^{2L}} \frac{1}{j} \right)^{2R-2} \\
&\ll \quad 2^{2R} L^2 \sum_{M\le u\le N} \frac{1}{u} \frac{1}{(2R-2)!} \left(2L + O(1/u)\right)^{2R-2} \\
&\ll \quad \left(\frac{L^R}{R!} \right)^2 \log(N/M) \cdot 2^{2R} R^{5/2}
\end{aligned}
$$

since $R \le m \ll ML$ and $L/R \le L \ll \log N$.

Combining the above we have proved (4.4) in our range. We deduce that the contribution of the $A_{R,L}(\sigma)$ terms in (4.3) with $r \le R \le m$ can be bounded as follows:

$$
\begin{aligned}
& \frac{1}{n!} \sum_{\sigma \in S_n} \left(\sum_{r\le R\le m} \binom{R}{r} \left| \frac{A_{R,L}(\sigma)}{\log N/M} - \frac{L^R}{R!} \right| \right) \\
\leqq \quad & \left(\frac{1}{n!} \sum_{\sigma\in S_n} \sum_{r\le R\le m} 1 \right)^{1/2} \left(\frac{1}{n!} \sum_{\sigma\in S_n} \sum_{r\le R\le m} \binom{R}{r}^2 \left| \frac{A_{R,L}(\sigma)}{\log N/M} - \frac{L^R}{R!} \right|^2 \right)^{1/2} \\
\le \quad & \left(m \frac{L^{2r}}{r!^2} \sum_{r\le R\le m} \frac{L^{2(R-r)}}{(R-r)!^2} \frac{1}{(\log N/M)^{1-o(1)}} \right)^{1/2} \\
= \quad & \left(\frac{me^{O(L)}}{(\log N/M)^{1-o(1)}} \right)^{1/2} = \frac{1}{(\log N/M)^{1/2-o(1)}}
\end{aligned}
$$

by Cauchy-Schwarz and then (4.4). Note also that

$$
\frac{1}{n!} \sum_{\sigma\in S_n} \sum_{r\le R\le m} \binom{R}{r} \frac{B_{R,L}(\sigma)}{\log N/M} \ll \frac{1}{\log N/M} \sum_{r\le R\le m} \binom{R}{r} \frac{L^{R+1}}{(R+1)!} \le \frac{1}{(\log N/M)^{1-o(1)}}.
$$

To get an upper bound in (4.5) (even adding in the $B_{R,L}(\sigma)$ terms) for terms with

$R \geq m$ we modify the above argument to obtain (since $L - \log(v/u) \leq L$):

$$\leq L \sum_{M \leq u \leq N} \frac{1}{R!} \left\{ \left(\sum_{u \leq j \leq ue^L} \frac{1}{j} \right)^R - \left(\sum_{u < j \leq ue^L} \frac{1}{j} \right)^R \right\}$$

The internal term $= RL^{R-1}/u(1 + O(R/uL))$ when $R \leq uL$, and is $O((L + O(1/u))^R)$ otherwise, So the sum is

$$\ll \begin{cases} R \cdot \frac{L^R}{R!}(\log(N/M) + 1) & \text{if } R \leq ML \\ R \cdot \frac{L^R}{R!}(\log(N/(R/L)) + (1 + O(1/ML))^R) & \text{if } ML < R \leq NL \\ LN \cdot \frac{L^R}{R!}(1 + O(1/ML))^R & \text{if } NL < R \end{cases}$$

Therefore the terms with $R \geq m$ in (4.3) sum up to

$$\begin{aligned} &\ll \sum_{R \geq m} \binom{R}{r} R \cdot \frac{L^R}{R!} + \sum_{R \geq ML} \frac{L^R}{R!} \min\{R, NL\} \exp(O(R/ML)) \\ &\ll \frac{L^r}{r!} \sum_{R \geq m} R \frac{L^{R-r}}{(R-r)!} + \frac{L^{[ML]}}{([ML]-1)!} \leq \frac{L^r}{r!} \frac{1}{e^{m-r}} \end{aligned}$$

by suitably modifying Lemma 1 since $m - r > 9m/10 > 9L$ (if $L < m/10$). Combining the above gives (4.1) in the range (4.2).

5 Permutations with a given number of cycles

Assuming $k \to \infty$ with $k \leq n^{1/2-\epsilon}$, let $M = k(\log\log n)^2$ and $N = n/(2k + \log^2 n)$, and take $\lambda := kL/v$. We will show that

$$\frac{1}{|S_{n,k}|} \sum_{\sigma \in S_{n,k}} \left| \mu_{r,L}(\sigma) - \frac{e^{-\lambda}\lambda^r}{r!} \right| \ll e^{-\lambda} \frac{\lambda^r}{r!} \frac{1}{2^m} \ . \tag{5.1}$$

holds in the range

$$r, \ \lambda, \ 1/\lambda \leq \min \left\{ \frac{\log\log n}{(\log\log\log n)^2}, \frac{\log k}{\log\log k} \right\}.$$

This, together with Proposition 1, implies Theorem 3.

To prove this we essentially follow the calculation of section 4, with minor modifications: We let $m := \min\{[k^{1/4}], [\log\log N/(\log\log\log N)^2]\}$ and assume below that $R \leq m$. Throughout that argument we have

$$\sum_i d(C_i) \leq 2RN \leq Rn/k = o(n) \text{ and } \ll nv/k.$$

Instead of (4.5) we consider

$$\frac{1}{|S_{n,k}|}\sum_{\sigma\in S_{n,k}} A_{R,L}(\sigma) \tag{5.2}$$

replacing $(n-\sum_{i=1}^{R} d(C_i))!/n!$ there by

$$S\left(n-\sum_{i=1}^{R} d(C_i), k-R\right)\Big/ S(n,k).$$

This ratio of Stirling numbers of the first kind gives, by (3.2) the analogy to the right side of (4.5) times a factor

$$\left(\frac{k}{\nu}\right)^R\left\{1+O\left(\frac{R^2}{k}+\frac{RN}{\frac{n}{k}\log\left(\frac{n}{k}\right)}+\frac{1}{\log\left(\frac{n}{k}\right)}\right)\right\} . \tag{5.3}$$

We can then follow through the same argument to get the following right side for the analogy to (4.7):

$$\frac{\lambda^R}{R!}\left\{\log(N/M)\left\{1+O\left(\frac{R^2}{k}+\frac{R}{v}\right)\right\}+O\left(\frac{1}{M}+R\left(\frac{R}{eLM}+\left(\frac{R}{eLM}\right)^{R-1}\right)\right)\right\}. \tag{5.4}$$

To determine the mean square we similarly multiply each term of the right side of (4.8) through by the relevant factor, analogous to (5.3). Thus we obtain the analogy to (4.4) though with new error terms arising from (5.3); namely

$$\ll \frac{\lambda^{2R}}{R!^2}\left\{\frac{1}{(\log N)^{1/2}}+\frac{R^2}{k}\right\} ,$$

which holds for $\lambda/k \ll R \ll \lambda v$. Therefore, by the same method,

$$\begin{aligned}\sum_{r\le R\le m}\binom{R}{r}\frac{1}{|S_{n,k}|}\sum_{\sigma\in S_{n,k}}\left|\frac{A_{R,L}(\sigma)}{\log(N/M)}-\frac{\lambda^R}{R!}\right| &\ll m^{1/2}e^{O(\lambda)}\left(\frac{1}{(\log N)^{1/2}}+\frac{1}{k^{1/2}}\right)^{1/2}\\ &\ll \frac{1}{(\log N)^{1/4-o(1)}}+\frac{1}{k^{1/2-o(1)}} .\end{aligned}$$

assuming $\lambda = o(\log k)$. For those terms with $R > m$, we use (3.4) and the argument of just above (4.1) to get an analogous bound: The term given by (3.4) is $< (k/v)^R$ times what we have just above (4.1) if $R > ML$ (which happens so long as $k = o(MLv)$). For smaller R the term given by (3.4) is $< (ke^{o(1)}/v)^R$ times what we have just above (4.1) which will lead to a similar bound (though in terms of λ) using Lemma 1.

Combining the above, in analogy to the argument of section 4, implies (5.1).

6 Orders

A classical theorem of Erdős and Turán [5] states that the order $\mathcal{O}(\sigma)$ is given by $\exp(\{\frac{1}{2}+o(1)\}\log^2 n)$ for almost all $\sigma \in S_n$. Note that $\mathcal{O}(\sigma) = \mathrm{lcm}[d_1(\sigma), d_2(\sigma), \ldots, d_\ell(\sigma)]$. We will give a simple proof of this by comparing $\mathcal{O}(\sigma)$ to $P_y(\sigma) = \prod_{d_i>y} d_i(\sigma)$ where $y = n^{o(1)}$. By the results of section 2, the expected size of $\log P_y(\sigma)$ is

$$\sum_{y<m\leq n} \frac{\log m}{m} \sim \frac{1}{2}\log^2 n.$$

Moreover the expected size of $(\log P_y(\sigma))^2$ is

$$\sum_{y<m\leq n} \frac{\log^2 m}{m} + \sum_{\substack{y<r,s\\ r+s\leq n}} \frac{\log r \log s}{rs} = \left(\sum_{y<m\leq n} \frac{\log m}{m}\right)^2 + O(\log^3 n);$$

and so $\log P_y(\sigma) \sim \frac{1}{2}\log^2 n$ for almost all $\sigma \in S_n$.

Define $\mathcal{O}_y(\sigma)$ to be the product of the prime powers p^e dividing $\mathcal{O}(\sigma)$ with $p > y$. Since each $p^e \leq d_i \leq n$ for some i, thus $1 \leq \mathcal{O}(\sigma)/\mathcal{O}_y(\sigma) \leq n^{\pi(y)}$. Define $g_y(d_i, d_j)$ to be the product of the prime powers p^e dividing (d_i, d_j) with $p > y$, so that

$$1 \leq P_y(\sigma)/\mathcal{O}_y(\sigma) \leq n^{\pi(y)} \prod_{y<d_i<d_j} g_y(d_i, d_j).$$

Now, since either $g_y = 1$ or $g_y > y$ thus the expected size of $\sum_{y<d_i,\ i<j} \log(g_y(d_i, d_j))$

$$\leq \sum_{p>y} \log p \sum_{a\geq 1} \sum_{\substack{r+s\leq n\\ p^a\,|(r,s)}} \frac{1}{rs} \leq \sum_{\substack{p>y,\ a\geq 1\\ p^a\leq n}} \frac{\log p}{p^{2a}} \left(\sum_{R\leq n} \frac{1}{R}\right)^2 \ll \frac{\log^2 n}{y}.$$

Selecting $y = (\log n \log\log n)^{1/2}$ we deduce that the expected value of $|\log(\mathcal{O}(\sigma)/P_y(\sigma))|$ is $\ll (\log n)^{3/2}/(\log\log n)^{1/2}$, and therefore $|\log \mathcal{O}(\sigma) - \log P_y(\sigma)| = o((\log n)^{3/2})$ for almost all $\sigma \in S_n$. Thus we deduce the result of Erdős and Turan.

6.2. And now for $S_{n,k}$.

We shall prove that almost all $\sigma \in S_{n,k}$ have order

$$\exp\left(\left\{\frac{1}{2}+o(1)\right\} k \frac{\log n \log(n/k^2)}{\log(n/k)}\right)$$

proceeding much as above, where $k \to \infty$ and $\log(n/k^2)/\log\log n \to \infty$ as $n \to \infty$. It would be interesting to get results for larger k; evidently the result must then take a different form and require a somewhat different method.

Let $\psi(k) = \log k/\log\log k$. The number of cycles in $S_{n,k}$ of length $> n\psi(k)/k$ is $\leq k/\psi(k)$ and so their total contribution to $\mathcal{O}(\sigma)$ is $\leq (k/\psi(k))\log n = o(k\log(n/k^2))$.

Therefore we now define $P_y(\sigma) = \prod_{n\psi(k)/k \geq d_i > y} d_i(\sigma)$, so that the expected size of $\log P_y(\sigma)$ is

$$\sum_{y<m\leq n\psi(k)/k} \frac{\log m}{m} \frac{S(n-m,k-1)/(n-m)!}{S(n,k)/n!} \sim \frac{k}{\nu} \sum_{y<m\leq n\psi(k)/k} \frac{\log m}{m} \sim \frac{k\log n \log(n/k^2)}{2\log(n/k)},$$

by (3.2). Similarly the expected size of $(\log P_y(\sigma))^2$ is

$$\sim \frac{k}{\nu} \sum_{y<m\leq n\psi(k)/k} \frac{\log^2 m}{m} + \left(\frac{k}{\nu}\right)^2 \sum_{y<r,s\leq n\psi(k)/k} \frac{\log r \log s}{rs} \sim \left(\frac{k\log n \log(n/k^2)}{2\log(n/k)}\right)^2$$

by (3.2), and so $\log P_y(\sigma) \sim k\log n \log(n/k^2)/(2\log(n/k))$ for almost all $\sigma \in S_{n,k}$.

We modify the definitions of $\mathcal{O}_y(\sigma)$ to avoid the cycles of length $> n\psi(k)/k$. We again have $1 \leq \mathcal{O}(\sigma)/\mathcal{O}_y(\sigma) \leq n^{\pi(y)}$; and the expected size of $\sum_{y<d_i\leq n\psi(k)/k,\ i<j} \log(g_y(d_i,d_j))$ is $\ll (k/\nu)^2(1/y)(\log(n/k))^2 \ll k^2/y$. Taking $y = 1 + k(\log\log n/\log n)^{1/2}$ we deduce that the expected value of $|\log(\mathcal{O}(\sigma)/P_y(\sigma))|$ is $\ll k(\log n/\log\log n)^{1/2}$, and therefore $|\log \mathcal{O}(\sigma) - \log P_y(\sigma)| = o(k(\log n)^{1/2})$ for almost all $\sigma \in S_{n,k}$. We therefore deduce the claimed result.

6.3. Large k – a reasoned guess.

Suppose that $k > \sqrt{n}(\log n)^A$ for some large A. By Proposition 1 we see that if $m \ll (n/k)\log(n/k)$ then almost all $\sigma \in S_{n,k}$ have $\gg \log n$ cycles of length m. For larger m recall that the expected number of cycles of length m in $\sigma \in S_{n,k}$ is

$$\frac{1}{m}\frac{S(n-m,k-1)/(n-m)!}{S(n,k)/n!}$$

(as in the proof of Proposition 1). To estimate this in the range $e^v < m = o(e^{2v})$ we follow through the proof of (3.2) to obtain

$$\sim \frac{k}{mv} \prod_{j=1}^{\tau-1} \left(\frac{n-m+j}{n+j}\right) \sim \frac{k}{mv}\exp(-mk/vn),$$

which is negligible once $m \geq (1+\epsilon)(n/k)\log(n/k)\log(k/m)$. Since this ratio is even smaller for larger m, almost all $\sigma \in S_{n,k}$ have order

$$\exp\left(\{1+o(1)\}\frac{n}{k}\log(n/k)\log(k^2/n)\right),$$

at least if $k \ll n/(\log n)^A$.

6.4. Summary.

Pushing the above methods to the edge of their range of validity (and involving quite a bit more number theory), one can show that

• if $(k/v)/\sqrt{n} \to 0$ then almost all $\sigma \in S_{n,k}$ have order

$$\exp\left(\{1+o(1)\}\frac{k}{2v}\log n \log\left(\frac{n}{(k/v)^2}\right)\right);$$

• if $(k/v)/\sqrt{n} \to \infty$ then almost all $\sigma \in S_{n,k}$ have order

$$\exp\left(\{1+o(1)\}\frac{n}{k/v}\log\left(\frac{(k/v)^2}{n}\right)\right).$$

If $k/v \asymp \sqrt{n}$ then there is some interesting transition function (for the size of the normal order) which needs to be determined.

References

[1] R. Arratia, A.D. Barbour and S. Tavaré, *Random combinatorial structures and prime factorizations*, Notices Amer. Math. Soc. **44** (1997), 903–910.

[2] P. Billingsley, *On the distribution of large prime divisors*, Period. Math. Hungar. **2** (1972), 283–289.

[3] E. Rodney Canfield, *Central and local limit theorems for the coefficients of polynomials of binomial type*, J. Comb. Theory, A **23** (1977), 275–290.

[4] J.M. DeLaurentis and B.G. Pittel, *Random permutations and Brownian motion*, Pac. J. Math. **119** (1985), 287–301.

[5] P. Erdős and P. Turán, *On some problems of a statistical group-theory, I*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **4** (1965), 175–186

[6] W. Feller, *An Introduction to Probability Theory and Its Application* (3rd ed), New York, Wiley, 1968

[7] A. Granville, *Prime divisors are Poisson distributed*, Internat. J. of Number Theory (to appear).

[8] A. Granville, *The anatomy of integers and permutations*, (in preparation).

[9] C. Jordan, *The calculus of finite differences (2nd ed)*, Chelsea, New York, 1947.

[10] D.E. Knuth and L. Trabb Prado, *Analysis of a simple factorization algorithm*, Theoret. Comput. Sci. **3** (1976), 321–348.

[11] H. Maier, *Primes in short interval*, Michigan Math. J. **32** (1985), 221–225.

[12] L. Moser and M. Wyman, *Asymptotic development of the Stirling numbers of the first kind*, J. London Math. Soc. **33** (1958), 133–146.

[13] L.A. Shepp and S.P. Lloyd, *Ordered cycle lengths in a random permutation*, Trans. Amer. Math. Soc. **121** (1966), 340–357.