Proceedings of the Edinburgh Mathematical Society (2021) **64**, 234–246
doi:10.1017/S0013091521000110

MOST PERMUTATIONS POWER TO A CYCLE OF SMALL PRIME LENGTH

S. P. GLASBY[1], CHERYL E. PRAEGER[1] AND W. R. UNGER[2]

[1] *Centre for the Mathematics of Symmetry and Computation, University of Western Australia, 35 Stirling Highway, Perth 6009, Australia* (stephen.glasby@uwa.edu.au; cheryl.praeger@uwa.edu.au)
[2] *School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia* (william.unger@sydney.edu.au)

(Received 31 March 2020; first published online 24 May 2021)

Abstract We prove that most permutations of degree n have some power which is a cycle of prime length approximately $\log n$. Explicitly, we show that for n sufficiently large, the proportion of such elements is at least $1 - 5/\log\log n$ with the prime between $\log n$ and $(\log n)^{\log\log n}$. The proportion of even permutations with this property is at least $1 - 7/\log\log n$.

Keywords: permutations; cycles; prime; length; density

2020 Mathematics subject classification: Primary 05A05; 68Q17;
Secondary 11N05

1. Introduction

The symmetric and the alternating groups S_n and A_n of degree n have been viewed as probability spaces with the uniform distribution since the seminal work of Gruder [9], Gončarov [7] and Erdös-Turán [4]. There is an analogy between the disjoint cycle decomposition of a permutation and the prime factorization of an integer whereby the cycles correspond to prime numbers (see [5, 8] for a description). The probability that a permutation $g \in \mathrm{S}_n$ is an n-cycle is $1/n$, while the probability that a number $p \leqslant x$ is prime is $1/\log x$, and according to this analogy, n corresponds to $\log x$. Counting integers without small prime factors is like counting permutations without small cycles: both lead to limits reminiscent of Mertens' third theorem:

$$\lim_{n\to\infty}\left(\log n \prod_{p\leqslant n}\left(1-\frac{1}{p}\right)\right) = e^{-\gamma}, \quad \text{where } \gamma = 0.5772\cdots$$

is the Euler–Mascheroni constant. Statistical methods have been invaluable for proving theoretical results. For example, the number of cycles of a uniformly distributed random element $g \in \mathrm{S}_n$ (or $g \in \mathrm{A}_n$) behaves as $n \to \infty$ like a normal distribution $N(\mu, \sigma^2)$ (c.f. [6, Theorem 1]) with mean and variance $\mu = \sigma^2 = \log n$, see [7, 15]. This paper focuses on proving the following theorem. We abbreviate $\log(n)$ and $\log(\log(n))$ as $\log n$ and $\log\log n$, respectively. All logarithms are to the natural base $e = 2.718\cdots$.

Theorem 1. *Suppose that $G \in \{\mathrm{A}_n, \mathrm{S}_n\}$. Let ρ_G be the proportion of permutations in G for which some power is a cycle of prime length p, where p lies in the open interval $(\log n, (\log n)^{\log\log n})$. Then, for n sufficiently large, $\rho_G \geqslant 1 - c/\log\log n$ where $c = 5$ when $G = \mathrm{S}_n$, and $c = 7$ when $G = \mathrm{A}_n$.*

A permutation $g \in \mathrm{S}_n$ having a single cycle of length greater than 1 is called a k-cycle, where the cycle has length k. Their theoretical importance has long been recognised: the presence of a k-cycle in a primitive subgroup G of S_n was shown to imply that G is A_n or S_n by Jordan in 1873 in the case where k is prime and $k \leqslant n-3$ ([12], or see [19, Theorem 13.9]). The same conclusion also holds for arbitrary $k < n/2$ by a result in Marggraff's dissertation [14], see also [11, Corollary 1.3], [19, Theorem 13.5] and P. M. Neumann's Mathematical Review MR0424912.

To use Jordan's result for deciding whether a given primitive subgroup G of S_n is indeed A_n or S_n, one needs to locate a p-cycle with p a prime, say by choosing random elements from G. It is inefficient to do this directly since the proportion of such elements in A_n or S_n is $O(n^{-1})$ (see § 1.1). Instead one searches for a 'pre-p-cycle', a permutation which powers to a p-cycle. It is shown in [17, Lemma 10.2.3] that the proportion of elements in A_n or S_n that power to a p-cycle with $n/2 < p \leqslant n-3$ is asymptotically $\log 2/\log n$, while an application of the main result Theorem 1 of [1] shows that considering only pre-p-cycles with p bounded, say $p \leqslant m$, produces a proportion $c(m)/n^{1/m}$. Thus, to approach Seress's asymptotic proportion, the primes p must be allowed to grow unboundedly with n.

Quite decisively, and perhaps surprisingly, the third author recently showed [18, Theorem 2] that *almost all* permutations in $G \in \{\mathrm{A}_n, \mathrm{S}_n\}$ are pre-p-cycles for some prime p. The purpose of this paper is to prove an even stronger statement: namely that, asymptotically, almost all permutations in G power to a p-cycle where the prime p is roughly $\log n$ (Theorem 1) and we may derive, from Proposition 6 and the proof of Theorem 1, explicit values for n_G such that the bounds of Theorem 1 hold for all $n \geqslant n_G$. On the one hand, we quantify the asymptotic results [18], giving a precise analysis with explicit bounds rather than asymptotics. In addition, we show (which came as a surprise to the authors) that the prime p can be chosen in a very small interval of length logarithmic in n. Theorem 1 relies on Proposition 6 which strengthens a key technical result [4, Theorem VI] of Erdös and Turán. Theorem 10 shows that the proportion of pre-p-cycles in G is at least $1 - c' \log\log n/\log(n-3)$ for p in the range $2 \leqslant p \leqslant n-3$. We prove in Remark 11 that for all $n \geqslant 5$, the proportion of pre-p-cycles in S_n with $2 \leqslant p \leqslant n-3$ is at least $1/19$. In the remainder of this section, we comment on uses and proofs of these results.

1.1. Recognising finite symmetric and alternating groups

We call a permutation $g \in \mathrm{S}_n$ a *pre-k-cycle* if it powers to a k-cycle where $k > 1$. For example, $g = (1,2,3,4)(5,6)(7,8,9)$ is a pre-3-cycle as $g^4 = (7,8,9)$, but g is not a pre-2-cycle. A simple argument (Lemma 3) shows that the disjoint cycle decomposition of a

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

pre-k-cycle contains exactly one k-cycle and all other cycles (if any) have lengths coprime to k. The number of k-cycles in S_n equals $n!/c(k)$ where $c(k) = k(n-k)!$. Hence the proportion ρ_n of cycles in S_n equals $\sum_{k=2}^{n} 1/c(k) = \sum_{j=0}^{n-2}(1/(n-j)j!)$. For n even,

$$n\rho_n - \sum_{j=0}^{n-2}\frac{1}{j!} = \sum_{j=0}^{n-2}\frac{j}{(n-j)j!} = \sum_{j=0}^{n/2}\frac{j}{(n-j)j!} + \sum_{j=n/2}^{n-2}\frac{j}{(n-j)j!} < \frac{2}{n}\sum_{j=0}^{n/2}\frac{j}{j!} + \sum_{j=n/2}^{\infty}\frac{j}{j!}.$$

Thus, $|n\rho_n - e| \to 0$ as $n \to \infty$, and so $\rho_n = O(n^{-1})$. By contrast, the proportion of permutations in S_n which are pre-k-cycles (for $2 \leqslant k \leqslant n$) approaches 1 as $n \to \infty$.

Let $X \subseteq \mathrm{S}_n$ and suppose that the subgroup $\langle X\rangle$ of S_n generated by X is primitive. Then a probabilistic algorithm for testing whether $\langle X\rangle$ contains A_n involves a random search for a pre-p-cycle, where p is a prime less than $n-2$, and it turns out that such elements have density 1 as $n \to \infty$ (Theorem 1). If the subgroup $\langle X\rangle$ does not contain A_n, we want to limit the (fruitless) search for pre-p-cycles, and for this we need explicit lower bounds on their density in both A_n and S_n, see Remark 2.

There are two important tools in our proof. The first is to quantify the sum $\sum_{a<p\leqslant b} 1/p^2$ as a function of the real numbers a and b. (Here, and henceforth, p denotes a prime.) The second is to estimate the number of permutations whose cycle lengths do not have certain 'forbidden lengths'. Erdös-Turán [4, Theorem VI] proved that the probability ρ that the cycle lengths of $g \in \mathrm{S}_n$ do not lie in a subset $\mathcal{A} \subseteq \{1,\ldots,n\}$ is at most $1/\mu$ where $\mu = \sum_{a\in\mathcal{A}} 1/a$. Applying a result of Ford [5, Theorem 2.9] gives the bound $\rho \leqslant e^{1-\mu}$ (take $r = 1$, $T_1 = \mathcal{A}$ and $k_1 = 0$ in Theorem 2.9). We prove in Proposition 6 that $\rho < e^{\gamma-\mu}$, where $\gamma = 0.5772\cdots$. Our bound is less than 2/3 the size of bounds of Erdös-Turán and Ford. This small improvement is needed to obtain practically useful bounds in Theorem 10. Manstavičius [13, p. 39] claimed that a slightly weaker bound, *viz.* $\rho < e^{\gamma-\mu}(1 + 1/n)$, followed from a pre-print of his. Unfortunately, we could not locate his pre-print.

After describing the cycle structure of pre-k-cycles in § 2, we prove our bound on forbidden cycle lengths in § 3. The sum $\sum_{a<p\leqslant b} 1/p^2$ is estimated in § 4, and the main theorems are proved in § 5.

Remark 2. Suppose that the proportion of pre-p-cycles in A_n or S_n is at least $c_0 = c_0(n)$. Given a primitive permutation group $G \leqslant \mathrm{S}_n$, then the probability that $\mathrm{A}_n \leqslant G$ and m independent random selections from G do *not* find a pre-p-cycle for some p with $2 \leqslant p \leqslant n-3$, is at most $(1-c_0)^m$. This upper bound is at most a prescribed 'failure probability' ε if and only if $m \geqslant \log(\varepsilon)/\log(1-c_0)$. Using Theorem 1, we can take $c_0(n) = 1 - c/\log\log n$, and a larger value is given in Theorem 10. Therefore, we may take c_0 to be an absolute constant thus requiring only $m = m(\varepsilon)$ random selections. We prove in Remark 11 that $c_0 = 0.05$ works for S_n and this gives $m(\varepsilon) \geqslant 20\log(\varepsilon^{-1})$. By contrast, the standard analysis of this algorithm [17, pp. 226–227] requires $m = O(\log(n)\log(\varepsilon^{-1}))$ random selections (because the proportion of pre-p-cycles in the range $n/2 < p < n-2$ is approximately $\log 2/\log n$). Thus, our analysis gives an algorithm that is faster by a factor of $\log(n)$.

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

2. Precycles

We begin with some notation. Fix $g \in \mathrm{S}_n$ and let $\lambda = \lambda(g)$ be the partition of n induced by the (disjoint) cycle lengths of g. We write $\lambda \vdash n$ and $\lambda = \langle 1^{m_1} 2^{m_2} \cdots n^{m_n} \rangle$ where m_k is the multiplicity of the part k, so $n = \sum_{k=1}^{n} k m_k$. The vector $(m_1, \ldots, m_n)$ is called the *cycle type* of g. A pre-k-cycle $g \in \mathrm{S}_n$ can be recognized by its *cycle partition* $\lambda(g)$.

Lemma 3. *Let $g \in \mathrm{S}_n$ be a pre-k-cycle with $\lambda(g) = \langle 1^{m_1} 2^{m_2} \cdots n^{m_n} \rangle$. Then $m_k = 1$ and for $i \neq k$, $m_i > 0$ implies $\gcd(k, i) = 1$. That is, g has a unique k-cycle, and its other cycles (if any) have lengths coprime to k.*

Proof. Suppose g has disjoint cycle decomposition $g_1 \cdots g_r$ where g_i is a cycle of length λ_i and $\lambda_1 + \cdots + \lambda_r = n$. Note that $g^\ell = g_1^\ell \cdots g_r^\ell$ and g_i^ℓ is a product of $\gcd(\lambda_i, \ell)$ cycles each of length $\lambda_i / \gcd(\lambda_i, \ell)$. Suppose g^ℓ is a k-cycle. Then there exists an i for which $\lambda_i / \gcd(\lambda_i, \ell) = k$ and $\gcd(\lambda_i, \ell) = 1$. Also, $\lambda_j / \gcd(\lambda_j, \ell) = 1$ for $j \neq i$. Thus, $\lambda_i = k$, $\gcd(\lambda_i, \ell) = 1$. For $j \neq i$, $\lambda_j \mid \ell$. Hence $\gcd(k, \lambda_j) = 1$. $\square$

Corollary 4. *For a prime p, the set of pre-p-cycles in $X \subseteq \mathrm{S}_n$ equals*

$$\left\{ g \in X \;\middle|\; m_p(g) = 1 \text{ and } \sum_{i \geqslant 2} m_{ip}(g) = 0 \right\}.$$

A group element whose order is coprime to p, for a prime p, is called a p'*-element*. The density of p'-elements in S_n is $\sigma_n = \prod_{i=1}^{\lfloor n/p \rfloor} (1 - 1/(ip))$ by [4, Lemma I]. We are grateful to John Dixon for pointing out to us that the density of pre-p-cycles in S_n is $(1/p)\sigma_{n-p}$. The density in A_n can be calculated similarly using [2, Theorem 3.3].

3. Forbidden cycle lengths

Given a set $\mathcal{A} \subseteq \{1, \ldots, n\}$ of 'forbidden' cycle lengths, let ρ be the proportion of elements of S_n having no cycle length in $\mathcal{A}$. Erdös and Turán prove $\rho \leqslant \mu^{-1}$ in [4, Theorem VI] where $\mu = \sum_{a \in \mathcal{A}} 1/a$. To see that his bound is not optimal, consider the case when $\mathcal{A} = \{1\}$ and $\mu = 1$. In this case, ρ is the proportion of derangements and the Erdös-Turán bound is unhelpful. However, an inclusion–exclusion argument shows that $\rho = \sum_{k=0}^{n} (-1)^k / k!$. As $e^{-1} < \rho$ when n is even and $\rho < e^{-1}$ otherwise, we expect for large n a bound of the form c/e where $c = 1 + o(1)$. We prove in Proposition 6 that there is a bound of the form $\rho \leqslant c e^{-\mu}$ where $c = 1.781 \cdots$. The constant c equals e^γ where $\gamma = 0.5772 \cdots$ is the Euler–Mascheroni constant. Our proof uses exponential generating functions, and has similarities to calculations used by Gruder [9], 15 years before Erdös and Turán [4].

If $\mu < 1$, then the Erdös–Turán bound gives no information. On the other hand, if $\mu \geqslant 1$, then we will show that the upper bound we obtain in Proposition 6, namely $e^{\gamma - \mu}$,

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

is strictly less than the bound μ^{-1} of Erdös and Turán. We use the fact that $e^\mu \geqslant e\mu$, since the tangent to the curve $y = e^x$ at $x = 1$ is $y = ex$. Hence

$$e^{\gamma-\mu} = \frac{c}{e^\mu} < \frac{1.782}{e^\mu} \leqslant \frac{1.782}{e\mu} < \frac{2}{3\mu}.$$

Similarly, our bound improves the bound $e^{1-\mu}$ of Ford [5, Theorem 2.9] by a factor of $e^{1-\gamma} < 2/3$. Note that Ford's bound can be sharpened using estimations in the proof of Proposition 6.

The centralizer of g in S_n has order $C(\lambda(g)) := \prod_{k=1}^n k^{m_k} m_k!$, and hence the conjugacy class g^{S_n} has size $n!/C(\lambda(g))$. Since the conjugacy classes in S_n are parameterized by the partitions of n, it follows that $\sum_{\lambda\vdash n} n!/C(\lambda) = n!$, or equivalently that $\sum_{\lambda\vdash n} 1/C(\lambda) = 1$.

The reader when trying to compare our analysis with that in [4] may wish to know that the quantity $L(P)$ defined on [4, p. 159] for a permutation P should be the number of cycles of P with length equal to some $a_\nu \in \mathcal{A}$, and not the number of $a_\nu \in \mathcal{A}$ with the property that P has a cycle of length a_ν.

Remark 5. Given a set X and a property P, let $\rho(X)$ denote the proportion of elements $x \in X$ that have property P. If $Y \subseteq X$, then $|Y|\rho(Y) \leqslant |X|\rho(X)$. To see this let $P(X)$ be the subset of $x \in X$ that have property P. Since $|P(X)| = |X|\rho(X)$ and $P(Y) \subseteq P(X)$, we have $|Y|\rho(Y) \leqslant |X|\rho(X)$.

Proposition 6. *The proportion $\rho(\mathrm{S}_n)$ of elements of S_n having no cycle length in $\mathcal{A} \subseteq \{1, \ldots, n\}$ satisfies $\rho(\mathrm{S}_n) < e^{\gamma-\mu}$ where $\mu = \sum_{a\in\mathcal{A}} 1/a$ and $e^\gamma \approx 1.781$. The proportion $\rho(\mathrm{A}_n)$ of elements of A_n with no cycle length in $\mathcal{A}$ satisfies $\rho(\mathrm{A}_n) < 2e^{\gamma-\mu}$.*

Proof. For $k \in \{1, \ldots, n\}$, let p_k equal 0 if $k \in \mathcal{A}$, and 1 otherwise. Set

$$P(z) = \sum_{k=1}^{n} \frac{p_k z^k}{k} \quad \text{and} \quad Q(z) = e^{P(z)}.$$

As $P(z)$ is a polynomial, $Q(z)$ is a differentiable function of the complex variable z, so its Taylor series $\sum_{m=0}^\infty q_m z^m$ at $z = 0$ converges to $Q(z)$ for all z. We consider the coefficient q_n. Since

$$Q(z) = \prod_{k\notin\mathcal{A}, k\leqslant n} e^{z^k/k} = \prod_{k\notin\mathcal{A}, k\leqslant n} \sum_{m_k=0}^{\infty} \frac{(z^k/k)^{m_k}}{m_k!},$$

the term $q_n z^n$ is a sum whose summands have the form

$$\prod_{k\notin\mathcal{A}, k\leqslant n} \frac{z^{km_k}}{k^{m_k}(m_k)!} = \frac{z^n}{C(\lambda)} \quad \text{for each partition } \lambda = \langle k^{m_k}\rangle_{k\notin\mathcal{A}} \text{ with } n = \sum\nolimits_{k\notin\mathcal{A}} km_k.$$

It follows that $q_n = \sum_{\lambda\vdash n} 1/C(\lambda)$ where the parts of λ do not lie in $\mathcal{A}$ (equivalently $m_k = 0$ for all $k \in \mathcal{A}$). This is precisely the proportion $\rho(\mathrm{S}_n)$ of elements of S_n none of whose cycle lengths lie in $\mathcal{A}$. Alternatively, [9, Eq. (18)] shows that $q_n = \rho(\mathrm{S}_n)$.

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

We now compute an upper bound for q_n. Define the nth harmonic number to be

$$H_n = \sum_{k=1}^{n} \frac{1}{k}, \quad \text{and let} \quad \gamma = \lim_{n\to\infty} (H_n - \log n) = 0.5772\cdots$$

be the Euler–Mascheroni constant. The error term $E(n) = H_n - \log n - \gamma$ satisfies $0 < E(n) < 1/(2n)$ by [10, p. 75]. Since $e^x < 1 + x + x^2 + \cdots = 1/(1-x)$ for $0 < x < 1$, it follows that $e^{1/(2n)} < 1 + 1/(2n-1) \leqslant 1 + 1/n$. Hence

$$1 < e^{E(n)} < 1 + \frac{1}{n} \quad \text{and} \quad P(1) = \sum_{k\notin\mathcal{A}, k\leqslant n} \frac{1}{k} = H_n - \mu = \log n + \gamma - \mu + E(n).$$

The previous display implies

$$Q(1) = e^{P(1)} = e^{\log n} e^{\gamma-\mu} e^{E(n)} < n e^{\gamma-\mu}\left(1 + \frac{1}{n}\right) = (n+1)e^{\gamma-\mu}. \tag{1}$$

Differentiating the equation $Q(z) = e^{P(z)}$ gives

$$Q'(z) = P'(z)Q(z), \quad \text{that is,} \quad \sum_{k=0}^{\infty} k q_k z^{k-1} = \left(\sum_{k=0}^{n-1} p_{n-k} z^{n-1-k}\right)\left(\sum_{k=0}^{\infty} q_k z^k\right).$$

Equating the coefficients of z^{n-1} on both sides gives

$$n q_n = \sum_{k=0}^{n-1} p_{n-k} q_k \leqslant \sum_{k=0}^{n-1} q_k \leqslant \left(\sum_{k=0}^{\infty} q_k\right) - q_n = Q(1) - q_n.$$

It follows from (1) that

$$(n+1)q_n \leqslant Q(1) < (n+1)e^{\gamma-\mu}.$$

Cancelling $n+1$ gives $q_n < e^{\gamma-\mu}$, that is, $\rho(\mathrm{S}_n) < e^{\gamma-\mu} < 1.782e^{-\mu}$. Finally, Lemma 5 implies that $\rho(\mathrm{A}_n) \leqslant 2\rho(\mathrm{S}_n) < 2e^{\gamma-\mu} < 3.563e^{-\mu}$. □

4. Estimating sums

In this section, we bound the quantity μ in Proposition 6. One approach is to use the Stieltjes integral [16, p. 67]. Instead, we take an elementary approach using finite sums, which we now briefly review. In the sequel, k denotes an integer and p denotes a prime. The *backward difference* of a function $f(k)$ is defined by $(\nabla f)(k) = f(k) - f(k-1)$. An

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

easy calculation shows

$$\sum_{k=a+1}^{b} (\nabla f)(k) = f(b) - f(a) \quad \text{and} \tag{2}$$

$$\nabla(f(k)g(k)) = (\nabla f)(k)g(k-1) + f(k)(\nabla g)(k). \tag{3}$$

Rearranging (3) and using (2) gives

$$\sum_{k=a+1}^{b} f(k)(\nabla g)(k) = f(b)g(b) - f(a)g(a) - \sum_{k=a+1}^{b} (\nabla f)(k)g(k-1). \tag{4}$$

Let $\pi(x) = \sum_{p\leqslant x} 1$. Observe that $(\nabla\pi)(k)$ equals 1 if k is prime, and 0 otherwise. Thus, $\sum_{a<p\leqslant b} f(p) = \sum_{a<k\leqslant b} f(k)(\nabla\pi)(k)$ and the latter can be estimated using (4). The following bounds hold for *real* $x \geqslant 17$ by [16, Theorem 1]. In order to get a practically useful bounds in Theorem 10, we note that (5) holds for all *integers* $x \geqslant 11$:

$$\frac{x}{\log x} \leqslant \pi(x) \leqslant \frac{x}{\log x}\left(1 + \frac{3}{2\log x}\right) \quad \text{for } x \in \{11, 12, 13, \dots\}. \tag{5}$$

Lemma 7. *Suppose a and b are real numbers with $12 \leqslant a \leqslant b$. Then*

$$\sum_{a<p\leqslant b} \frac{1}{p^2} \leqslant \frac{2.22}{\lfloor a\rfloor \log\lfloor a\rfloor} - \frac{1.61}{\lfloor b\rfloor \log\lfloor b\rfloor}.$$

Proof. Since the primes in the range $a < p \leqslant b$ coincide with the primes in the range $\lfloor a\rfloor < p \leqslant \lfloor b\rfloor$, we henceforth may (and will) assume that a and b are *integers* satisfying $12 \leqslant a \leqslant b$. Also, if $a = b$ the inequality holds, so we assume also that $a + 1 \leqslant b$. Applying (4) and (5) gives

$$\begin{aligned}\sum_{a<p\leqslant b} \frac{1}{p^2} &= \sum_{k=a+1}^{b} \frac{1}{k^2}(\nabla\pi)(k) = \frac{\pi(b)}{b^2} - \frac{\pi(a)}{a^2} - \sum_{k=a+1}^{b} \frac{-2k+1}{k^2(k-1)^2}\pi(k-1) \\ &\leqslant \frac{1}{b\log b}\left(1 + \frac{3}{2\log b}\right) - \frac{1}{a\log a} \\ &\quad + \sum_{k=a+1}^{b} \frac{2k-1}{k^2(k-1)\log(k-1)}\left(1 + \frac{3}{2\log(k-1)}\right).\end{aligned}$$

Since $12 \leqslant a \leqslant k-1$, we have $1 + 3/(2\log(k-1)) \leqslant 1 + 3/(2\log(12)) < 1.61$. As $k - \frac{1}{2} < k$, the Σ-term above is less than

$$\begin{aligned}\frac{1.61}{\log a}\sum_{k=a+1}^{b} \frac{2(k-\frac{1}{2})}{k(k-\frac{1}{2})(k-1)} &= \frac{3.22}{\log a}\sum_{k=a+1}^{b} \frac{1}{k(k-1)} = \frac{3.22}{\log a}\sum_{k=a+1}^{b} \nabla\left(-\frac{1}{k}\right) \\ &= \frac{3.22}{\log a}\left(\frac{1}{a} - \frac{1}{b}\right) \\ &\leqslant \frac{3.22}{a\log a} - \frac{3.22}{b\log b}.\end{aligned}$$

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

Combining the previous two displays gives the desired inequality:

$$\sum_{a+1<p<b} \frac{1}{p^2} < (-1+3.22)\,\frac{1}{a\log a} + (1.61-3.22)\,\frac{1}{b\log b} = \frac{2.22}{a\log a} - \frac{1.61}{b\log b}. \qquad \square$$

Lemma 8. *Suppose a, b are real numbers with $2 \leqslant a \leqslant b$ and p is prime. Then*

$$\log\left(\frac{\log b}{\log a}\right) - \frac{1}{2(\log b)^2} - \frac{1}{(\log a)^2} < \sum_{a<p\leqslant b} \frac{1}{p} < \log\left(\frac{\log b}{\log a}\right) + \frac{1}{(\log b)^2} + \frac{1}{2(\log a)^2}.$$

Proof. The proof uses the following bounds [16, Theorem 5, Corollary]

$$\log\log x + M - \frac{1}{2(\log x)^2} < \sum_{p\leqslant x} \frac{1}{p} < \log\log x + M + \frac{1}{(\log x)^2} \quad \text{for } x > 1,$$

where $M = 0.26149\cdots$ is the Meissel–Mertens constant. Thus,

$$\begin{aligned}
\sum_{a<p\leqslant b} \frac{1}{p} &= \sum_{p\leqslant b} \frac{1}{p} - \sum_{p\leqslant a} \frac{1}{p} \\
&< \left(\log\log b + M + \frac{1}{(\log b)^2}\right) - \left(\log\log a + M - \frac{1}{2(\log a)^2}\right) \\
&= \log\left(\frac{\log b}{\log a}\right) + \frac{1}{(\log b)^2} + \frac{1}{2(\log a)^2}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{a<p\leqslant b} \frac{1}{p} &= \sum_{p\leqslant b} \frac{1}{p} - \sum_{p\leqslant a} \frac{1}{p} \\
&> \log\log b + M - \frac{1}{2(\log b)^2} - \left(\log\log a + M + \frac{1}{(\log a)^2}\right) \\
&= \log\left(\frac{\log b}{\log a}\right) - \frac{1}{2(\log b)^2} - \frac{1}{(\log a)^2}.
\end{aligned} \qquad \square$$

5. Proof of Theorem 1

Suppose $G \in \{\mathrm{A}_n, \mathrm{S}_n\}$. Given certain functions $a = a(n)$ and $d = d(n)$, our strategy is to find a lower bound for the proportion ρ_G of pre-p-cycles in G with $a(n) < p < a(n)^{d(n)}$. It is shown in [18] that $\rho_G \to 1$ as $n \to \infty$. Theorem 1 was motivated by the needed to analyse certain Las-Vegas algorithms for permutation groups, we shall adapt a combinatorial argument in [18] to quantify this result. Henceforth, p denotes a prime, and k denotes an integer.

Fix $G \in \{\mathrm{A}_n, \mathrm{S}_n\}$. Recall $\lambda(g) = \langle 1^{m_1(g)} 2^{m_2(g)} \cdots n^{m_n(g)} \rangle$ is the partition of n whose parts are the cycle lengths of g, and part k has multiplicity $m_k(g)$. Define

$$\begin{aligned} \mathcal{P}_n &= \{p \mid a(n) < p \leqslant a(n)^{d(n)}\}, \\ T(G) &= \{g \in G \mid m_p(g) \geqslant 1 \text{ for some } p \in \mathcal{P}_n\}, \\ U_p(G) &= \Big\{g \in G \;\Big|\; m_p(g) \geqslant 1 \text{ and } \sum_{k \geqslant 1} m_{kp}(g) \geqslant 2\Big\}, \text{ and} \\ U(G) &= \bigcup_{p \in \mathcal{P}_n} U_p(G). \end{aligned}$$

Note that $g \in T(G) \setminus U_p(G)$ has $m_p(g) \geqslant 1$ and $\sum_{k \geqslant 1} m_{kp}(g) = 1$. Hence $m_p(g) = 1$ and g is a pre-p-cycle by Lemma 3. Thus, $T(G) \setminus U(G)$ is precisely the set of pre-p-cycles in G for some $p \in \mathcal{P}_n$. In the following proposition, we view $G \in \{\mathrm{A}_n, \mathrm{S}_n\}$ as a probability space with the uniform distribution, and we seek a lower bound for

$$\begin{aligned} \rho_G &= \mathrm{Prob}\,(g \in G \text{ is a pre-}p\text{-cycle for some } p \in \mathcal{P}_n) = \mathrm{Prob}\,(g \in T(G) \setminus U(G)) \\ &= \mathrm{Prob}\Big(g \in G \mid m_p(g) = 1 \text{ and } \sum_{k \geqslant 2} m_{kp}(g) = 0 \text{ for some } p \in \mathcal{P}_n\Big). \end{aligned}$$

Proposition 9. *Let $a(n), d(n)$ be functions satisfying $a(n) \geqslant 12$, $d(n) > 1$ and $a(n)^{d(n)} \leqslant n$ for all n. Using the preceding notation and $\delta = |\mathrm{S}_n : G|$, we have*

$$\rho_G = \frac{|T(G)| - |U(G)|}{|G|} \geqslant 1 - \frac{2.287\delta}{d(n)} - \frac{2.22(\log n - 1)}{\lfloor a(n) \rfloor \log \lfloor a(n) \rfloor} - \frac{4.4\delta \log n}{a(n)(\log a(n))n}.$$

Proof. Let $\mu = \sum_{p \in \mathcal{P}_n} \frac{1}{p}$. Write a and d instead of $a(n)$ and $d(n)$, and set $b = a^d$. As $a < p \leqslant b \leqslant n$, we have $\mathcal{P}_n \subseteq \{1, \ldots, n\}$. Thus, Proposition 6 gives

$$\frac{|T(G)|}{|G|} = 1 - \mathrm{Prob}\,(g \in G \mid m_p(g) = 0 \text{ for all } p \in \mathcal{P}_n) \geqslant 1 - \delta e^{\gamma - \mu}. \tag{6}$$

We seek a lower bound for (6). This means finding a lower bound for μ. Since $a \geqslant 12$, Lemma 8 gives

$$\mu = \sum_{a < p \leqslant a^d} \frac{1}{p} > \log d - \frac{1}{2(d \log a)^2} - \frac{1}{(\log a)^2} > \log d - 0.25. \tag{7}$$

Thus, it follows from (6) and (7) that

$$\frac{|T(G)|}{|G|} \geqslant 1 - \frac{\delta e^{\gamma}}{e^{\mu}} \geqslant 1 - \frac{\delta e^{\gamma}}{e^{\log d - 0.25}} \geqslant 1 - \frac{2.287\delta}{d}. \tag{8}$$

We seek an upper bound for $|U_p(G)|$. We (over)count the number of permutations $g_1 g_2 g_3 \in U_p(G)$ where g_1, g_2, g_3 have disjoint supports, g_1 is p-cycle and g_2 is a kp-cycle for some $k \geqslant 1$. We may choose g_1 in $\binom{n}{p}(p-1)!$ ways, because we take the first element to

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

be the smallest in the cycle. Next, we choose a kp-cycle in $\binom{n-p}{kp}(kp-1)!$ ways. Observe that g_1 is even: p is odd as $p > \log 12 > 2$. Thus, when $G = \mathrm{A}_n$ we must be able to choose g_3 to have the same parity as g_2, so that $g_1g_2g_3$ is even. In the generic case when $(k+1)p \leqslant n-2$ this is possible. The number of choices for (g_1, g_2, g_3) in the generic case is

$$\binom{n}{p}(p-1)!\binom{n-p}{kp}(kp-1)!\frac{(n-p-kp)!}{\delta} = \frac{n!}{\delta}\frac{1}{kp^2}.$$

This is an upper bound for the number of products $g_1g_2g_3$ in the generic case, and we may halve this upper bound if $k = 1$.

Consider the special case when $n-1 \leqslant (k+1)p \leqslant n$ has a solution for k and p. Given p, this happens for at most one value of k, namely $k = n/p - 1$ or $k = (n-1)/p - 1$, where p divides n or $n-1$, respectively. We will show that there are very few primes p for which $(k+1)p$ lies in $\{n-1, n\}$ for some k. Suppose that $(k+1)p = n-1$ and $n-1$ has r (not necessarily distinct) prime divisors greater than a. Then $n-1 > a^r$ and hence there are at most $r < \log_a(n-1)$ choices for p. Similarly, if $(k+1)p = n$ there are less than $\log_a(n)$ choices for p. Thus, the special case has less than $2\log_a(n)$ choices for p. Consequently, the contribution in the special case is small, and our estimations need not be so careful.

In this special case, arguing as above, the number of choices of (g_1, g_2) is $n!/(kp^2)$, and the number n_3 of choices for g_3 is 1, unless $G = \mathrm{A}_n$ and k is even, in which case $n_3 = 0$. Thus, the number of products $g_1g_2g_3$ is at most $n!/(kp^2)$ and we bound the denominator as follows:

$$kp^2 > kap \geqslant \frac{ka(n-1)}{k+1} \geqslant \frac{a(n-1)}{2}.$$

In the special case, the number of choices for $g_1g_2g_3$ is at most

$$\frac{n!}{kp^2} = \frac{\delta|G|}{kp^2} \leqslant \frac{2\delta|G|}{a(n-1)} = \varepsilon|G| \qquad \text{where } \varepsilon = \frac{2\delta}{a(n-1)}.$$

For each prime p, let ε_p be ε if $n-1 \leqslant (k+1)p \leqslant n$ has a solution for k, and 0 otherwise. In the generic case we have $(k+1)p \leqslant n-2$ and $k \leqslant m := \lfloor (n-2)/p \rfloor - 1$. The bound $\sum_{k=1}^{m} 1/k < 1 + \int_1^m \mathrm{d}t/t = 1 + \log m$ is problematic if $m = 0$, so we replace m with $m+1$ to get

$$\frac{|U_p(G)|}{|G|} \leqslant \left(\sum_{k=1}^{m+1}\frac{1}{kp^2}\right) + \varepsilon_p < \frac{1+\log(m+1)}{p^2} + \varepsilon_p.$$

However, $1 + \log(m+1) \leqslant 1 + \log(\lfloor n/p \rfloor) < \log n - 1$ as $\log p > \log 12 > 2$. Therefore

$$\frac{|U_p(G)|}{|G|} \leqslant \frac{\log n - 1}{p^2} + \varepsilon_p.$$

As $\varepsilon_p = \varepsilon \neq 0$ for less than $2\log_a(n)$ choices of p, we have

$$\frac{|U(G)|}{|G|} \leqslant \sum_{a<p\leqslant b}\frac{|U_p(G)|}{|G|} \leqslant (\log n - 1)\left(\sum_{a<p\leqslant b}\frac{1}{p^2}\right) + 2\varepsilon\log_a(n).$$

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

Applying the bound for $\sum_{a<p\leqslant b} 1/p^2$ in Lemma 7 gives

$$\frac{|U(G)|}{|G|} \leqslant (\log n - 1)\left(\frac{2.22}{\lfloor a\rfloor \log\lfloor a\rfloor} - \frac{1.61}{\lfloor b\rfloor \log\lfloor b\rfloor}\right) + \frac{4\delta \log_a(n)}{a(n-1)}.$$

Since $12 \leqslant a < n$ and $4/(n-1) < 4.4/n$ holds for $n \geqslant 13$, we have

$$\frac{|U(G)|}{|G|} \leqslant (\log n - 1)\left(\frac{2.22}{\lfloor a\rfloor \log\lfloor a\rfloor} - \frac{1.61}{\lfloor b\rfloor \log\lfloor b\rfloor}\right) + \frac{4.4\delta \log_a(n)}{an}. \tag{9}$$

Now (8) and (9) give

$$\rho_G = \frac{|T(G)| - |U(G)|}{|G|} \geqslant 1 - \frac{2.287\delta}{d} - \frac{2.22(\log n - 1)}{\lfloor a\rfloor \log\lfloor a\rfloor} - \frac{4.4\delta \log n}{a(\log a)n}. \tag{10}$$

□

Proof of Theorem 1. Set $a = \log n$ and $d = \log\log n$. Suppose that $n \geqslant e^{12}$. Then $a \geqslant 12$, and also $a^d = (\log n)^{\log\log n} < n$. Using Proposition 9 and the inequalities $a - 1 < \lfloor a\rfloor$ and $4.4/\log a < 2$ gives

$$\begin{aligned}\rho(G) &\geqslant 1 - \frac{2.287\delta}{d} - \frac{2.22(a-1)}{\lfloor a\rfloor \log\lfloor a\rfloor} - \frac{4.4\delta a}{a(\log a)n}\\ &> 1 - \frac{2.287\delta}{d} - \frac{2.22}{\log(a-1)} - \frac{2\delta}{n}.\end{aligned}$$

However,

$$\log(a-1) = \log a + \log\left(1 - \frac{1}{a}\right) \geqslant \log a + \log\left(\frac{11}{12}\right) \geqslant c' \log a,$$

where $c' = 1 - \log(12/11)/\log(12) = \log(11)/\log(12)$ and $2/n < 10^{-3}/\log\log n$, so

$$\rho_G > 1 - \frac{2.287\delta}{\log\log n} - \frac{2.22}{c'\log\log n} - \frac{0.001\delta}{\log\log n} \geqslant 1 - \frac{2.288\delta}{\log\log n} - \frac{2.301}{\log\log n}.$$

This is at least $1 - c/\log\log n$ where $c = 4.6$ if $\delta = 1$ and $c = 6.9$ if $\delta = 2$. □

Allowing pre-p-cycles with larger p gives us a sharper lower bound for ρ_G.

Theorem 10. *Suppose that $n \geqslant e^{12}$ and $A_n \leqslant G \leqslant S_n$. Let ρ_G be the proportion of permutations in G that power to a cycle with prime length $p \leqslant n - 3$. Then*

$$\rho_G \geqslant 1 - \frac{(4.58\delta + 0.17)\log\log n}{\log(n-3)} \qquad \text{where } \delta = \frac{n!}{|G|} \in \{1, 2\}.$$

https://doi.org/10.1017/S0013091521000110 Published online by Cambridge University Press

Proof. Set $a(n) = (\log n)^2$ in Proposition 9 and suppose that $a(n)^{d(n)} = n-3$. (The hypotheses $a(n) \geqslant 12$, $d(n) > 1$ and $a^d \leqslant n$ of Proposition 9 clearly hold.) Then $d(n) = \log(n-3)/\log a = \log(n-3)/2\log\log n$ and Proposition 9 gives

$$\rho_G \geqslant 1 - \frac{2.287\delta}{d(n)} - \frac{2.22(\log n - 1)}{\lfloor(\log n)^2\rfloor \log\lfloor(\log n)^2\rfloor} - \frac{4.4\delta \log n}{a(\log a)n}$$
$$> 1 - \frac{4.574\delta \log\log n}{\log(n-3)} - \frac{2.22(\log n - 1)}{\lfloor(\log n)^2\rfloor \log\lfloor(\log n)^2\rfloor} - \frac{2.2\delta}{\log(n-3)(\log\log n)n}.$$

Note that $\lfloor(\log n)^2\rfloor > (\log n)^2 - 1 = (\log n - 1)(\log n + 1)$ and since $n \geqslant e^{12}$ we have

$$\frac{2.22(\log n - 1)}{\lfloor(\log n)^2\rfloor} < \frac{2.22}{\log n + 1} < \frac{2.05}{\log(n-3)} \quad \text{and} \quad \frac{2.2\delta}{(\log\log n)n} \leqslant \frac{\delta \log\log n}{10^3}.$$

Using these inequalities, and combining the δ terms, shows that

$$\rho_G \geqslant 1 - \frac{4.575\delta \log\log n}{\log(n-3)} - \frac{2.05}{\log(n-3)\log(12^2)} \geqslant 1 - \frac{(4.58\delta + 0.17)\log\log n}{\log(n-3)}. \qquad \square$$

Remark 11. Suppose that $n \geqslant 5$. We prove that the proportion π_n of elements of S_n that are pre-p-cycles for some p with $2 \leqslant p \leqslant n-3$ is at least $1/19$. We know that $\pi_n \geqslant \pi_0$ where $\pi_0 := \sum_{n/2 < p \leqslant n-3} 1/p$. A simple computation with Magma [3] shows that $\pi_0 \geqslant 1/19$ for all n satisfying $5 \leqslant n \leqslant 400{,}000$. For $n > 400{,}000 > e^{12}$ we have $\pi_n \geqslant 1 - 4.75 \log\log n/\log(n-3) > 1/19$ by Theorem 10. Precise computations of π_n for $n \leqslant 50$ suggest that $\pi_n > 1/3$ may even hold for all $n \geqslant 5$.

Acknowledgments. We thank the referee for some very helpful suggestions. SPG and CEP gratefully acknowledge support from the Australian Research Council (ARC) Discovery Project DP190100450, and WRU gratefully acknowledges the support from ARC Discovery Project DP160104626.

References

1. J. Bamberg, S. P. Glasby, S. H. Harper and C. E. Praeger, *Permutations with orders coprime to a given integer*, *Electronic J. Combin.* **27** (2020), P1.6.
2. R. Beals, C. R. Leedham-Green, A. C. Niemeyer, C. E. Praeger and Á. Seress, Permutations with restricted cycle structure and an algorithmic application, *Combin. Probab. Comput.* **11**(5) (2002), 447–464.
3. W. Bosma, J. Cannon and C. Playoust, The Magma algebra system. I. The user language, *J. Symbolic Comput.* **24**(3–4) (1997), 235–265. Computational algebra and number theory (London, 1993).
4. P. Erdös and P. Turán, On some problems of a statistical group-theory. II, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 151–163.
5. K. Ford, *Anatomy of integers and random permutations course lecture notes*, available at https://faculty.math.illinois.edu/ford/anatomyf17.html, 25 March 2020.
6. W. M. Y. Goh and E. Schmutz, The expected order of a random permutation, *Bull. London Math. Soc.* **23**(1) (1991), 34–42.
7. V. Gončarov, On the field of combinatory analysis, *Am. Math. Soc. Transl. (2)* **19** (1962), 1–46.

8. A. Granville, Cycle lengths in a permutation are typically Poisson, *Electron. J. Combin.* **13**(1) (2006), Research Paper 107, 23.
9. O. Gruder, Zur Theorie der Zerlegung von Permutationen in Zyklen, *Ark. Mat.* **2** (1952), 385–414. (German).
10. J. Havil, *Gamma* (Princeton Science Library, Princeton University Press, Princeton, NJ, 2009). Exploring Euler's constant; Reprint of the 2003 edition.
11. G. A. Jones, Primitive permutation groups containing a cycle, *Bull. Aust. Math. Soc.* **89**(1) (2014), 159–165.
12. C. Jordan, Sur la limite de transitivité des groupes non alternés, *Bull. Soc. Math. France* **1** (1872/73), 40–71.
13. E. Manstavičius, On random permutations without cycles of some lengths, *Period. Math. Hungar.* **42**(1–2) (2001), 37–44.
14. B. Marggraff, *Über primitive Gruppen mit transitiven Untergruppen geringeren Grades*, Univ. Giessen, Giessen, circa 1890, Jbuch Volume 20, p. 141.
15. W. Plesken and D. Robertz, The average number of cycles, *Arch. Math. (Basel)* **93**(5) (2009), 445–449.
16. J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–94.
17. Á. Seress, *Permutation group algorithms*, Cambridge Tracts in Mathematics, **Volume 152** (Cambridge University Press, Cambridge, 2003).
18. W. R. Unger, *Almost all permutations power to a prime length cycle*, arXiv:1905.08936 (2019).
19. H. Wielandt, *Finite permutation groups* (Academic Press, New York–London, 1964).