Periodica Mathematica Hungarica Vol. 1 (1), (1971), pp. 5—13.

ON SOME PROBLEMS OF A STATISTICAL GROUP THEORY V

by

P. ERDŐS (Budapest) and P. TURÁN (Budapest)

To the memory of CATHERINE RÉNYI

1. In the second paper of this series (see [2]) we dealt with statistical theorems concerning the arithmetical structure of $O(P)$ the order of the element P in symmetric group S_n of n letters. If $\omega(x) \nearrow \infty$ with x arbitrarily slowly and assigning to the phrase "for almost all P's" the meaning "for all but $o(n!)$ P's" the theorems in question run as follows.

THEOREM A. *For almost all P's the order $O(P)$ is divisible by all prime powers not exceeding*

$$\frac{\log n}{\log\log n}\left\{1 + 3\frac{\log\log\log n}{\log\log n} - \frac{\omega(n)}{\log\log n}\right\}. \tag{1.1}$$

The theorem is best possible in the strong sense that the number of P's whose order $O(P)$ is divisible by all primes not exceeding

$$\frac{\log n}{\log\log n}\left\{1 + 3\frac{\log\log\log n}{\log\log n} + \frac{\omega(n)}{\log\log n}\right\}. \tag{1.2}$$

is only $o(n!)$.

THEOREM B. *The maximal prime factor of $O(P)$ is for almost all P's between*

$$n\exp\left(-\omega(n)\sqrt{\log n}\right) \quad \text{and} \quad n\exp\left(-\frac{1}{\omega(n)}\sqrt{\log n}\right). \tag{1.3}$$

Though these theorems reveal surprisingly simple statistical laws, they still leave a big playground for the prime factors of $O(P)$. We have found that formulating the problem in a different form the quantity in (1.1) can be replaced by a much bigger one and those in (1.3) by a much smaller one. Observing namely that the P's in a fixed conjugacy class H of S_n have the same order $O(H)$ it is plausible to consider the order rather as a function of the conjugacy class than that of the single P's. Since the total number of conjugacy classes in S_n is $p(n)$, the number of partitions of n, it is plausible to mean by the phrase "for almost all classes H" the one "for all but $o(p(n))$

classes H". Then we are going to prove in this paper the following analogon of Theorem B.

THEOREM. *Denoting by $\varrho(H)$ the maximal prime factor of $O(H)$ the inequality*

$$\left| \varrho(H) - \left(\frac{\sqrt{6}}{2\pi}\sqrt{n}\log n - \frac{\sqrt{6}}{\pi}\sqrt{n}\log\log n\right)\right| < \sqrt{n}\,\omega(n) \tag{1.4}$$

holds for almost all classes H if only $\omega(x) \nearrow \infty$ with x arbitrarily slowly.[1]

This is indeed much smaller than the quantity in (1.3). In the next paper of this series we are going to prove that $O(H)$ is for almost all classes divisible by *all* prime powers not exceeding

$$\frac{2\pi}{\sqrt{6}}\frac{\sqrt{n}}{\log n}\left\{1 + 5\frac{\log\log n}{\log n} - \frac{\omega(n)}{\log n}\right\} \tag{1.5}$$

with the same convention for $\omega(n)$ and this is again best possible in the sense of (1.2). The proof of (1.5) is rather deep. The quantities in (1.4) resp. (1.5) are surprisingly close to each other.

2. Next we turn to the proof of our theorem. First we have to investigate $p_q(n)$, the number of such partitions of n where no summand is divisible by the fixed q. Since for $y > 0$ we have

$$\begin{aligned}\sum_{n=0}^{\infty} p_q(n)\, e^{-ny} &= \prod_{q \nmid \nu} \frac{1}{1-e^{-\nu y}} = \prod_{\nu=1}^{\infty}\frac{1}{1-e^{-\nu y}} \cdot \prod_{\mu=1}^{\infty}(1-e^{-\mu q y}) = \\ &= \left\{\sum_{m=1}^{\infty} p(m)\, e^{-my}\right\}\prod_{\mu=1}^{\infty}(1-e^{-\mu q y})\end{aligned} \tag{2.1}$$

and owing to the "Pentagonalzahlsatz" of EULER—LEGENDRE

$$\prod_{\mu=1}^{\infty}(1-e^{\mu r}) = \sum_{k=-\infty}^{\infty}(-1)^k \exp\left(-\frac{3k^2+k}{2}r\right) \tag{2.2}$$

(2.1) gives at once

$$p_q(n) = \sum_{\frac{3k^2+k}{2} \le \frac{n}{q}} (-1)^k\, p\left(n - \frac{3k^2+k}{2}q\right). \tag{2.3}$$

Denoting by $r_q(n)$ the number of partitions of n such that at least one summand is divisible by q we get from (2.3)

$$r_q(n) = \sum_{\frac{3k^2+k}{2} \le \frac{n}{q},\ k \neq 0} (-1)^{k+1}\, p\left(n - \frac{3k^2+k}{2}q\right). \tag{2.4}$$

[1] We assume throughout this paper $\omega(x) = o(\log\log x)$.

We shall use that classical formula due to HARDY and RAMANUJAN (see [1])

$$p(n) = (1 + o(1)) \frac{1}{4n\sqrt{3}} \exp\left(\frac{2\pi}{\sqrt{6}}\sqrt{n}\right). \tag{2.5}$$

Let now q be an arbitrary prime divisor of $O(H)$. If

$$n = m_1 n_1 + m_2 n_2 + \ldots + m_k n_k \tag{2.6}$$

$$1 \leq n_1 < n_2 < \ldots < n_k \ (\leq n)$$

then as is well known the conjugacy classes of S_n are in one to one correspondence with the partitions in (2.6) and

$$O(H) = [n_1, n_2, \ldots, n_k]. \tag{2.7}$$

Hence

$$q \leq n.$$

Let first

$$100 \sqrt{n} \log n \leq q \leq n. \tag{2.8}$$

Since evidently

$$r_q(n) < 4 \sqrt{\frac{n}{q}}\, p(n-q)$$

we get using (2.4)[2]

$$r_q(n) < c \sqrt{\frac{n}{q}} \exp\left(\frac{2\pi}{\sqrt{6}} \sqrt{n - 100\sqrt{n}\log n}\right) < \tag{2.9}$$

$$< c\sqrt{n} \exp\left\{\frac{2\pi}{\sqrt{6}}\sqrt{n}\left(1 - \frac{50\log n}{\sqrt{n}}\right)\right\} < c\, p(n) n^{-50}$$

and hence

$$\sum_{\substack{100\sqrt{n}\log n \leq \\ \leq q \leq n}} r_q(n) = o(p(n)). \tag{2.10}$$

Next let

$$\lambda_1 \sqrt{n} \log n \leq q < 100 \sqrt{n} \log n, \tag{2.11}$$

λ_1 positive, to be determined later. First we remark that the contribution of k's with $|k| \geq \log n$ to $r_q(n)$ in (2.4) is, as before, $< cp(n)n^{-50}$ and hence

$$\sum_{\lambda_1 \leq \frac{q}{\sqrt{n}\log n} \leq 100} r_q(n) \leq o(p(n)) + \sum_{\lambda_1 \leq \frac{q}{\sqrt{n}\log n} \leq 100} \ \sum_{1 \leq |k| < \log n} p(n - k^2 q). \tag{2.12}$$

[2] The c's without indices mean unspecified positive constants.

Using (2.5) the second sum in (2.12) is

$$< c \sum_{\lambda_1 \leq \frac{q}{\sqrt{n}\log n} \leq 100} \sum_{1 \leq |k| \leq \log n} \frac{1}{n - k^2 q} \exp\left(\frac{2\pi}{\sqrt{6}}\sqrt{n - k^2 q}\right) <$$

$$< \frac{c}{n} \exp\left(\frac{2\pi}{\sqrt{6}}\sqrt{n}\right) \sum_{\lambda_1 \leq \frac{q}{\sqrt{n}\log n} \leq 100} \sum_{1 \leq |k| \leq \log n} \exp\left(-\frac{\pi}{\sqrt{6}}\frac{k^2 q}{\sqrt{n}}\right) <$$

$$< c\,p(n) \sum_{q \geq \lambda_1 \sqrt{n}\log n} \exp\left(-\frac{\pi}{\sqrt{6}} \cdot \frac{q}{\sqrt{n}}\right).$$

The last sum is

$$\int_{\lambda_1 \sqrt{n}\log n}^{\infty} \exp\left(-\frac{\pi}{\sqrt{6}}\frac{x}{\sqrt{n}}\right) d\pi(x) = \int_{\lambda_1 \sqrt{n}\log n}^{\infty} \exp\left(-\frac{\pi}{\sqrt{6}}\frac{x}{\sqrt{n}}\right)\frac{dx}{\log x} +$$

$$+ \int_{\lambda_1 \sqrt{n}\log n}^{\infty} \exp\left(-\frac{\pi}{\sqrt{6}}\frac{x}{\sqrt{n}}\right) d(\pi(x) - \mathrm{Li}\,x) < c\,n^{\frac{1}{2} - \frac{\pi}{\sqrt{6}}\lambda_1} (\log n)^{-1}$$

which $\to 0$ if

$$\lambda_1 = \frac{\sqrt{6}}{2\pi} - \frac{\sqrt{6}}{\pi}\frac{\log\log n}{\log n} + \frac{\omega(n)}{\log n}. \tag{2.13}$$

Hence the number of conjugacy classes H for which $O(H)$ is divisible by a prime

$$> \frac{\sqrt{6}}{2\pi}\sqrt{n}\log n\left(1 - 2\frac{\log\log n}{\log n} + \frac{\omega(n)}{\log n}\right)$$

s $o(p(n))$ indeed. This proves the first part of the theorem.

3. Next we turn to the less easy proof of the second part of the theorem which asserts that putting

$$M_1 = \frac{\sqrt{6}}{2\pi}\sqrt{n}\log n\left\{1 - 2\frac{\log\log n}{\log n} - \frac{\omega(n)}{\log n}\right\}$$

(3.1)

$$M_2 = \frac{\sqrt{6}}{2\pi}\sqrt{n}\log n\left\{1 - 2\frac{\log\log n}{\log n} + \frac{\omega(n)}{\log n}\right\}$$

for almost all conjugacy classes H the number $O(H)$ has at least one prime actor between M_1 and M_2. In order to prove it let

$$M_1 \leq q_1 < q_2 < \ldots < q_l \leq M_2 \tag{3.2}$$

be all primes between M_1 and M_2 and we define the "class function" $k(H)$ by

$$k(H) = \sum_{q_i | O(H)}^{(j)} 1 \tag{3.3}$$

empty sum being 0. Let

$$S_1 = \frac{1}{p(n)} \sum_{(H)} k(H). \tag{3.4}$$

Then we need the

LEMMA I. *With* $\omega(n)$ *in* (3.1) *we have*

$$S_1 = (1 + o(1))\, 2 \frac{\sqrt{6}}{\pi} e^{\frac{1}{2}\omega(n)}.$$

4. (3.4)–(3.3)–(3.2) give at once

$$S_1 = \frac{1}{p(n)} \sum_{j=1}^{l} \sum_{\substack{(H) \\ O(H) \equiv 0 \bmod q_j}} 1. \tag{4.1}$$

The inner sum is obviously the number of such partitions of n in which at least one summand is divisible by q and thus $= r_{q_j}(n)$; hence

$$S_1 = \frac{1}{p(n)} \sum_{j=1}^{l} r_{q_j}(n)$$

and using (2.3)

$$S_1 = \frac{1}{p(n)} \sum_{j=1}^{l} \sum_{k \neq 0}^{(k)} (-1)^{k+1} p\left(n - \frac{3k^2+k}{2} q_j\right) \tag{4.2}$$
$$\frac{3k^2+k}{2} \leq \frac{n}{q_j}.$$

As in **2.** the contribution of k's with $|k| > \log n$ is $o(1)$ and hence

$$S_1 = o(1) + \frac{1}{p(n)} \sum_{j=1}^{l} \sum_{1 \leq |k| \leq \log n} (-1)^{k+1} p\left(n - \frac{3k^2+k}{2} q_j\right). \tag{4.3}$$

Next we consider the contribution of $k = -1$. Denoting it by S_1' we have using (2.4)

$$S_1' = \frac{1}{p(n)} \sum_{j=1}^{l} p(n - q_j) = (1 + o(1)) \sum_{j=1}^{l} \frac{n}{n - q_j} \exp \frac{2\pi}{\sqrt{6}} (\sqrt{n - q_j} - \sqrt{n}) =$$

$$= (1 + o(1)) \sum_{j=.}^{l} \exp\left(-\frac{\pi}{\sqrt{6}} \cdot \frac{q_j}{\sqrt{n}}\right) = (1 + o(1)) \int_{M_1}^{M_2} \frac{1}{\log x} \exp\left(-\frac{\pi}{\sqrt{6}} \cdot \frac{x}{\sqrt{n}}\right) dx =$$

$$= \frac{1 + o(1)}{\log n} 2 \int_{M_1}^{M_2} \exp\left(-\frac{\pi}{\sqrt{6}} \cdot \frac{x}{\sqrt{n}}\right) dx \tag{4.4}$$

from which using (3.1)

$$S_1' = (1 + o(1))\, 2 \frac{\sqrt{6}}{\pi} \exp\left(\frac{1}{2}\omega(n)\right). \tag{4.5}$$

Finally the contribution of the remaining terms in (4.3) cannot exceed absolutely

$$c\frac{\log n}{p(n)}\sum_{j=1}^{l} p(n-2q_j) < c\log n\sum_{j=1}^{l}\exp\frac{2\pi}{\sqrt{6}}(\sqrt{n-2q_j}-\sqrt{n}) <$$

$$< c\log n\sum_{j=1}^{l}\exp\left\{-\frac{2\pi}{\sqrt{6}}\frac{q_j}{\sqrt{n}}\right\} < c\log n\int_{M_1}^{M_2}\exp\left(-\frac{2\pi}{\sqrt{6}}\cdot\frac{x}{\sqrt{n}}\right)dx =$$

$$= O\left(\frac{\log^2 n}{\sqrt{n}}e^{-\omega(n)}\right) = o(1).$$

This, (4.3) and (4.5) prove Lemma I.

5. Next we are going to investigate

$$S_2 = \frac{1}{p(n)}\sum_{(H)} k(H)^2. \tag{5.1}$$

As to this we state the

LEMMA II. *With the above $\omega(n)$ we have*

$$S_2 = (1+o(1))\left\{2\frac{\sqrt{6}}{\pi}\exp\left(\frac{1}{2}\omega(n)\right)\right\}^2.$$

For the proof we observe first that

$$S_2 = \frac{1}{p(n)}\sum_{(H)}\sum_{\substack{(\mu)\\ q_\mu/O(H)}}\sum_{\substack{(\nu)\\ q_\nu/O(H)}} 1 =$$

$$= S_1 + \frac{1}{p(n)}\sum_{\substack{(H)\\ q_\mu q_\nu/O(H)\\ \mu\neq\nu}} 1 = \tag{5.2}$$

$$= S_1 + \frac{2}{p(n)}\sum_{1\le\mu<\nu\le l}\;\sum_{\substack{(H)\\ q_\mu q_\nu/O(H)}} 1.$$

6. Let at fixed $\mu\neq\nu$ $r_{\mu\nu}^{(1)}(n)$ stand for the number of partitions of n with the property no summands being divisible neither by q_μ nor by q_ν. Then we have obviously putting for $y>0$

$$f(y) = \prod_{\nu=1}^{\infty}\frac{1}{1-e^{-\nu y}} \tag{6.1}$$

the relation

$$F_1(y) \overset{\text{def}}{=} \sum_{n=0}^{\infty} r_{\mu\nu}^{(1)}(n)e^{-ny} = \prod_{\substack{(l)\\ q_\mu\nmid l\\ q_\nu\nmid l}}\frac{1}{1-e^{-ly}} = \frac{f(y)f(q_\mu q_\nu y)}{f(q_\mu y)f(q_\nu y)}. \tag{6.2}$$

Next let $r^{(2)}_{\mu\nu}(n)$ be the number of partitions of n with the property that

a) either no summand is divisible by q_μ

b) or no summand is divisible by q_ν.

Then we have, using also (6.2), for $y > 0$

$$F_2(y) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} r^{(2)}_{\mu\nu}(n) e^{-ny} = \frac{f(y)}{f(q_\mu y)} + \frac{f(y)}{f(q_\nu y)} - \tag{6.3}$$
$$- \frac{f(y) f(q_\mu q_\nu y)}{f(q_\mu y) f(q_\nu y)}.$$

Finally let $r^{(3)}_{\mu\nu}(n)$ stand for the number of partitions of n with the property that at least one summand is divisible by q_μ and at least one by q_ν. Then we have, using (6.3), for $y > 0$

$$F_3(y) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} r^{(3)}_{\mu\nu}(n) e^{-ny} = f(y) - \frac{f(y)}{f(q_\mu y)} - \frac{f(y)}{f(q_\nu y)} + \tag{6.4}$$
$$+ \frac{f(y) f(q_\mu q_\nu y)}{f(q_\mu y) f(q_\nu y)}.$$

7. Returning to (5.2) we see at once (owing to (2.7)) that

$$\sum_{\substack{(H)\\ q_\mu q_\nu / O(H)}} 1 = r^{(3)}_{\mu\nu}(n) = \text{coeffs } e^{-ny} \text{ in} \tag{7.1}$$
$$\left\{1 - \frac{1}{f(q_\mu y)} - \frac{1}{f(q_\nu y)} + \frac{f(q_\mu q_\nu y)}{f(q_\mu y) f(q_\nu y)}\right\} f(y).$$

Since for $n > c$ we have owing to (3.1)

$$q_\mu q_\nu > n.$$

(7.1) gives also

$$\sum_{\substack{(H)\\ q_\mu q_\nu / O(H)}} 1 = \text{coeffs } e^{-ny} \text{ in } f(y) \left(1 - \frac{1}{f(q_\mu y)}\right) \left(1 - \frac{1}{f(q_\nu y)}\right) \tag{7.2}$$

and hence from (5.2)

$$S_2 = S_1 + \frac{2}{p(n)} \sum_{1 \le \mu < \nu \le l} \text{coeffs } e^{-ny} \text{ in } f(y)\left(1 - \frac{1}{f(q_\mu y)}\right)\left(1 - \frac{1}{f(q_\nu y)}\right) =$$
$$= S_1 + \frac{2}{p(n)} \text{coeffs } e^{-ny} \text{ in } f(y) \sum_{1 \le \mu < \nu \le l} \left(1 - \frac{1}{f(q_\mu y)}\right)\left(1 - \frac{1}{f(q_\nu y)}\right) = \tag{7.3}$$
$$= S_1 + \frac{1}{p(n)} \text{coeffs } e^{-ny} \text{ in}$$
$$f(y) \cdot \left\{\left(\sum_{1 \le \mu \le l} \left(1 - \frac{1}{f(q_\mu y)}\right)\right)^2 - \sum_{1 \le \mu \le l} \left(1 - \frac{1}{f(q_\mu y)}\right)^2\right\}.$$

Now we have to investigate the power series expansion of the function in the curly bracket. Using the Pentagonalzahlsatz in (2.2) we have

$$\sum_{1\le\mu\le l}\left(1-\frac{1}{f(q_\mu y)}\right)=\sum_{1\le\mu\le l}\sum_{m\neq 0}(-1)^{m+1}\exp\left(-\frac{3m^2+m}{2}q_\mu y\right)$$

and its square

$$\sum_{1\le\mu_1\le l}\sum_{1\le\mu_2\le l}\sum_{m_1\neq 0}\sum_{m_2\neq 0}(-1)^{m_1+m_2}\exp\left\{-\left(\frac{3m_1^2+m_1}{2}q_{\mu_1}+\frac{3m_2^2+m_2}{2}q_{\mu_2}\right)y\right\},$$

further

$$\sum_{1\le\mu\le l}\left(1-\frac{1}{f(q_\mu y)}\right)^2=$$

$$=\sum_{1\le\mu\le l}\sum_{m_1\neq 0}\sum_{m_2\neq 0}(-1)^{m_1+m_2}\exp\left\{-\left(\frac{3m_1^2+m_1}{2}+\frac{3m_2^2+m_2}{2}\right)q_\mu y\right\}.$$

These give together with (7.3)

$$S_2=S_1+$$

$$\frac{2}{p(n)}\sum_{1\le\mu_1<\mu_2\le l}\sum_{m_1\neq 0}\sum_{m_2\neq 0}(-1)^{m_1+m_2}\cdot p\left(n-\frac{3m_1^2+m_1}{2}q_{\mu_1}-\frac{3m_2^2+m_2}{2}q_{\mu_2}\right).$$

From this we can finish the proof of Lemma II quickly. The contribution of the terms apart from the one corresponding to $m_1=m_2=-1$ is $o(1)$ as before. The remaining term results owing to (2.5)

$$\frac{2}{p(n)}\sum_{1\le\mu_1<\mu_2\le l}p(n-q_{\mu_1}-q_{\mu_2})=$$

$$=(1+o(1))\,2\sum_{1\le\mu_1<\mu_2\le l}\exp\left\{\frac{2\pi}{\sqrt{6}}\left(\sqrt{n-q_{\mu_1}-q_{\mu_2}}-\sqrt{n}\right)\right\}=$$

$$=(1+o(1))\sum_{\substack{1\le\mu_1,\mu_2\le l\\ \mu_1\neq\mu_2}}\exp\left\{-\frac{q_{\mu_1}+q_{\mu_2}}{\sqrt{n}}\cdot\frac{\pi}{\sqrt{6}}\right\}=$$

$$=(1+o(1))\left\{\sum_{1\le\mu\le l}\exp\left(-\frac{\pi}{\sqrt{6}}\cdot\frac{q_\mu}{\sqrt{n}}\right)\right\}^2$$

which proves Lemma II owing to (4.5) .

8. The proof of the theorem follows now easily from the investigation of

(8.1) $$Z\overset{\text{def}}{=}\frac{1}{p(n)}\Sigma_{(n)}\left\{k(H)-2\frac{\sqrt{6}}{\pi}\exp\left(\frac{1}{2}\omega(n)\right)\right\}^2.$$

Lemma I and II give at once that

$$Z = o(1) \exp \omega(n). \tag{8.2}$$

If $k(H) = 0$ would be for more than $\alpha p(n)$ classes (α positive constant) then we had

$$Z > \frac{1}{p(n)} \alpha p(n) 4 \frac{6}{\pi^2} \exp \omega(n)$$

which contradicts to (8.2). Hence the theorem is proved.

REFERENCES

1] G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, *Proc. of the London Math. Soc.* **16** (1918), 75—115.

[2] P. Erdős and P. Turán, On some problems of a statistical group theory II, *Acta Math. Acad. Sci. Hung.* **18** (1967), 151—163.

(Received October 1, 1969)

MTA MATEMATIKAI KUTATÓ INTÉZETE,
BUDAPEST, V., REÁLTANODA U. 13–15.

EÖTVÖS LORÁND TUDOMÁNYEGYETEM,
ALGEBRA ÉS SZÁMELMÉLETI TANSZÉK,
BUDAPEST, VIII., MÚZEUM KRT. 6–8.
HUNGARY