MATHEMATICS

ON SEQUENCES OF INTEGERS GENERATED BY A SIEVING PROCESS

BY

PAUL ERDÖS AND ERI JABOTINSKY

(Communicated by Prof. J. POPKEN at the meeting of June 29, 1957)

PART II

4. *The second term of the asymptotic expansion for a_k (for $b_k = a_k$ and any $\lambda \geqslant 1$)*

Using formula (4) and (26), we shall prove that:

$$\frac{a_k}{k} = \prod_{a_i < k} \frac{a_i}{a_i - 1} + O(1). \tag{27}$$

Indeed, the number Q in (4) is defined as the smallest integer for which $a_{k-Q} < Q+1$, whence, by (26):

$$k - Q = [1 + o(1)] \frac{k}{\log k}. \tag{28}$$

Formula (4) now becomes:

$$\frac{a_k - \lambda}{k} \left[\prod_{a_i \leqslant Q} \left(\frac{a_i}{a_i - 1} \right) \right] \left[1 - \frac{\theta}{\log k} \right] \quad \text{with } 0 < \theta < 2. \tag{28'}$$

But it can be seen by using (26) and (28) that:

$$\prod_{Q \leqslant a_i < k} \left(\frac{a_i}{a_i - 1} \right) = 0 \left[\prod_{k - \frac{k}{\log k} < r \log r < k} \left(1 + \frac{1}{r \log r} \right) \right] = 1 + o\left(\frac{1}{(\log k)^2} \right), \tag{29}$$

and further from (21) and (26):

$$\prod_{a_i < Q} \frac{a_i}{a_i - 1} < \prod_{a_i \leqslant a_k} \frac{a_i}{a_i - 1} = (1 + o(1)) \log k. \tag{29'}$$

Thus (27) follows from (28′), (29) and (29′).

Now we want to prove that:

$$a_n = n \log n + (\tfrac{1}{2} + o(1))\, n\, (\log \log n)^2. \tag{30}$$

We will omit some of the details. Put:

$$a_n = n \log n + \tfrac{1}{2} n (\log \log n)^2 + n f(n) \log \log n. \tag{31}$$

To prove (30) we must prove that:

$$f(n)=o\,(\log\log n). \tag{32}$$

First we show that for every $\varepsilon>0$ and $n>n_0(\varepsilon)$:

$$f(n)<\varepsilon \log\log n. \tag{33}$$

The proof of $f(n)>-\varepsilon\log\log n$ would be similar.

If (33) would not hold, a simple argument shows that there would exist two infinite sequences n_k and m_k satisfying:

$$(34)\quad \begin{cases} m_k^{1/2}<n_k<m_k,\ f(m_k)\geq f(n_k)+\varepsilon, \\ \quad f(m_k)>f(u),\ 1\leq u\leq m_k,\ f(n_k)<f(n_k+v),\ 0<v<m_k-n_k. \end{cases}$$

By (26) and (27) we have:

$$(35)\quad \begin{cases} \frac{a_m}{m}=\frac{a_n}{n}\prod_{n\leq a_i<m}\left(1-\frac{1}{a_i}\right)^{-1}+O(1)=\frac{a_n}{n}\Big[\prod_{n\leq a_i<m}\left(1+\frac{1}{a_i}\right)+O\left(\frac{1}{n}\right)\Big]+O(1)= \\ \qquad\qquad =\frac{a_n}{n}\prod_{n\leq a_i<m}\left(1+\frac{1}{a_i}\right)+O(1). \end{cases}$$

Hence from (31) by putting $m=m_k$ and $n=n_k$ in (35), for some $c\geq\varepsilon$:

$$(36)\quad \begin{cases} \log m+\tfrac{1}{2}(\log\log m)^2+(f(n)+c)\log\log m= \\ \qquad =[\log n+\tfrac{1}{2}(\log\log n)^2+f(n)\log\log n]\prod_{n\leq a_i<m}\left(1+\frac{1}{a_i}\right)+O(1). \end{cases}$$

Now we show that for $n\leq a_i<m$:

$$a_i/i\geq\log i+\tfrac{1}{2}(\log\log i)^2+(f(n)+o(1))\log\log i. \tag{37}$$

For $i\geq n$ this follows from the definition of n_i. For the a_i satisfying $n\leq a_i\leq a_n$ it follows from (35) and $a_i=(1+o(1))\,i\log i$ by a simple computation.

Suppose now that (33) does not hold. Then, from (35) and (36), we have $f(m)=f(n)+c$ and:

$$(38)\quad \begin{cases} \log m+\tfrac{1}{2}(\log\log m)^2+[f(n)+c]\log\log m< \\ \qquad <\log n+\tfrac{1}{2}(\log\log n)^2+f(n)\cdot\log\log n\cdot\prod_{n\leq a_k<m}\left(1+\frac{1}{g(k)}\right), \end{cases}$$

where:

$$g(k)=k[\log k+\tfrac{1}{2}(\log\log k)^2+\{f(k)+o(1)\}\log\log k].$$

Put $m=n^{1+\delta}$. In the computation which follows we will neglect terms which are $o(\log\log n)$, or in estimating the product on the right side of (38) we can neglect terms which are $o(\log\log n/\log n)$.

We have (the equality sign is to be understood to mean that terms which are $o(\log\log n/\log n)$ have been neglected):

$$\prod_{n\leq a_k<n^{1+\delta}}\left(1+\frac{1}{g(k)}\right)=\prod_{n\leq k\log k<m^{1+\delta}}\left[1+\frac{1}{g(k)}\right]=\exp\Big[\sum_{n\leq k\log k<m}\frac{1}{g(k)}\Big]. \tag{39}$$

Henceforth it is to be understood that in all the products and sums $n<k\log k<n^{1+\delta}$. We have:

$$\sum\frac{1}{g(k)}=\sum\frac{1}{k\log k}-\tfrac{1}{2}\sum\frac{(\log\log k)^2}{k(\log k)^2}-f(n)\sum\frac{\log\log k}{k\log\log k)^2}.$$

Further clearly by the integral test:

$$\sum\frac{1}{k\log k}=\log(1+\delta)+\frac{\log\log n}{\log n}\frac{\delta}{1+\delta},$$

$$\sum\frac{(\log\log k)^2}{k(\log k)^2}=\frac{\delta(\log\log n)^2-2\log(1+\delta)\log\log n}{(1+\delta)\log n}+\frac{2\delta}{1+\delta}\frac{\log\log n}{\log n},$$

$$\sum\frac{\log\log k}{k(\log k)^2}=\frac{\delta}{1+\delta}\frac{\log\log n}{\log n}.$$

Thus:

$$\sum\frac{1}{g(k)}=\log(1+\delta)-\frac{\delta}{2(1+\delta)}\frac{(\log\log n)^2}{\log n}+ \\ +\frac{\log(1+\delta)\log\log n}{(1+\delta)\log n}-f(n)\frac{\delta}{1+\delta}\frac{\log\log n}{\log n}.$$

Hence from (39):

$$(40)\quad\begin{cases}\prod\left[1+\frac{1}{g(k)}\right]=\exp\left[\sum\frac{1}{g(k)}\right]=(1+\delta)-\frac{\delta}{2}\frac{(\log\log n)^2}{\log n}+ \\ \qquad +\frac{\log(1+\delta)\log\log n}{\log n}-\delta f(n)\frac{\log\log n}{\log n}.\end{cases}$$

Thus if we put $m=n^{1+\delta}$ in (38) we obtain from (40):

$$(1+\delta)\log n+\tfrac{1}{2}[\log\log n+\log(1+\delta)]^2+[f(n)+c]\cdot[\log\log n+\log(1+\delta)]< \\ <[\log n+\tfrac{1}{2}(\log\log n)^2+f(n)\log\log n]\cdot\Big[1+\delta-\frac{\delta}{2}\frac{(\log\log n)^2}{\log n}+ \\ +\frac{\log(1+\delta)\log\log n}{\log n}-\delta f(n)\frac{\log\log n}{\log n}\Big],$$

which is easily seen to be false because of the uncancelled term $c\log\log n$ on the left side of the inequality (since the coefficient of $f(n)$ is greater on the left side than on the right side).

5. *The third term of the asymptotic expansion of a_k (for $b_k=a_k$ and any $\lambda>1$)*

We note that formula (27) was obtained by using only step one for the computation of a_k, that is by using formula (4) and not formula (6). It is not possible to get the next term without using steps of all orders. To do this, we have to calculate successively for $m=1,2,\ldots$ all the q_m occurring in (6). q_m is defined as the smallest integer for which:

$$m(a_{k-q_m}-1)<mq_m-\sum_{i=0}^{m-1}q_i.$$

Because of (26) it can be seen that:

$$q_m=k-\frac{k}{m\log k}+o\left(\frac{k}{\log k}\right),$$

and (6) becomes:

$$a_k-\lambda=\Big[\prod_{i<\frac{k}{m\log k}+o\left(\frac{k}{\log k}\right)}\left(\frac{a_i-1}{a_i}\right)\Big]\Big[k+\left(-1+\frac{1}{1}+\frac{1}{2}+\dots+\frac{1}{m-1}+\theta_m\right)\frac{k}{\log k}+$$
$$+o\left(\frac{k}{\log k}\right)\Big],$$

with $0<\theta<1$. Writing θ_m instead of θ and using (26) this becomes:

$$(41)\quad \begin{cases}\dfrac{a_k}{k}=\Big[\prod\limits_{a_i<\frac{k}{m\log k}}\left(\dfrac{a_i}{a_i-1}\right)\Big]\Big[1+\left(-1+\frac{1}{1}+\frac{1}{2}+\dots+\frac{1}{m-1}+\frac{\theta_m}{m}\right)\dfrac{1}{\log k}\Big]\\ \qquad\qquad\Big[1+o\left(\dfrac{1}{\log k}\right)\Big],\end{cases}$$

with $0<\theta_m<1$.

We now use the fact that, because of (26):

$$(42)\qquad \prod_{\frac{k}{m\log k}<a_i<\frac{k}{(m-1)\log k}}\left(\frac{a_i}{a_i-1}\right)=1+\frac{\log\frac{m}{m-1}}{\log k}+o\left(\frac{1}{\log k}\right).$$

Rewriting (41) for $(m-1)$ instead of m and comparing with (41), we find, using (42):

$$1+\left(-1+\frac{1}{1}+\frac{1}{2}+\dots+\frac{1}{m-2}+\frac{\theta_{m-1}}{m-1}+\log\frac{m}{m-1}\right)\frac{1}{\log k}=$$
$$=1+\left(-1+\frac{1}{1}+\frac{1}{2}+\dots+\frac{1}{m-1}+\frac{\theta_m}{m}\right)\frac{1}{\log k}+o\left(\frac{1}{\log k}\right),$$

and so:

$$(43)\qquad \frac{\theta_{m-1}}{m-1}+\log\frac{m}{m-1}=\frac{\theta_m}{m}+\frac{1}{m-1}+o(1).$$

Rewriting (43) for $(m-1)$, $(m-2)$, ..., 2 instead of m, summing up and cancelling we find:

$$(44)\qquad \frac{\theta_1}{1}+\log m=\frac{\theta_m}{m}+\frac{1}{1}+\frac{1}{2}+\dots+\frac{1}{m-1}+o(1).$$

But $0<\theta_m<1$ and for large k and m (44) can only hold if $\lim\limits_{k\to\infty}\theta_1=\gamma$. Thus:

$$(45)\qquad \theta_1=\gamma+o(1).$$

Formula (41) now becomes:

$$(46)\qquad \frac{a_k}{k}=\Big[\prod_{a_i<\frac{k}{m}}\left(\frac{a_i}{a_i-1}\right)\Big]+(-1+\gamma+\log m)+o(1),$$

and, for $m=1$:

$$(47)\qquad \frac{a_k}{k}=\Big[\prod_{a_i<k}\left(\frac{a_i}{a_i-1}\right)\Big]+(-1+\gamma)+o(1).$$

We note that (45) and (26) together yield (46), so that (47) contains all the information that results from the use of formula (6) with the estimate $q_m = k - \frac{k}{m \log k} + o\left(\frac{k}{\log k}\right)$ of q_m.

Any improvement of the $o(1)$ term in (47) can only result from an improvement of these estimates of q_m.

Put now:

$$\frac{a_k}{k} = \log k + \tfrac{1}{2} (\log \log k)^2 + (2-\gamma) \log \log k + f(k).$$

Then we can show by the same method as was used in proving (30), but by more laborious computations that $f(k) = o(\log \log k)$: we supress all details. This completes the proof of the result stated at the end of the introduction.

Several further questions can be asked about the a_k all of which have been investigated for the sequence of primes e.g. Is it true that $\lim\inf (a_{k+1} - a_k) < \infty$?

Is it true that $\lim\sup (a_{k+1} - a_k)/\log k = \infty$, [1] we do not know the answer to any of these questions.

After writing our paper we find that the quadruple paper of V. Gardiner, R. Lazarus, N. Metropolis and S. Ulam deals with a slight variant of our case $b_k = a_k$, they make a table of these numbers up to 48600 (Math. Magazine 29 (1956), 117–122). They further conjecture $a_k/p \to 1$. Hawkins proved this conjecture and Chowla proved

$$a_k = k \log k + (\tfrac{1}{2} + o(1))\, k (\log \log k)^2,$$

the proofs of Hawkins and Chowla are not yet published.

Added March 1957. Viggo Brun asked the following question: Put $n = n_1$, $n_{l+1} = n_l - [n_l/l]$. Determine the smallest integer k for which $n_{k+1} = n_k$ (i.e. for which $k+1 > n_k$).

By the methods used in in dealing with the case $b_k = k+1$ we can prove that $k = (1 + o(1))\, (\pi^2/8)\, n^{1/2}$.

David, in a paper to appear in Riveon le Matematika, vol. 11, considers the sequence $u_1 = n$, $u_k = k\left[\frac{u_{k-1}}{k}\right]$ and asks when $u_k = 0$. This reduces to our problem for $b_k = k+1$.

Israël Institute of Technology, Haifa

---

[1] The fact that $\lim\sup (p_{k+1} - p_k)/_{\log p_k} = \infty$ is due to Westzynthius, see P. Erdös, Quarterly Journal of Math. 6, 124–128 (1934). $\lim\inf (p_{k+1} - p_k) < \infty$ has never been proved.