KONINKL. NEDERL. AKADEMIE VAN WETENSCHAPPEN – AMSTERDAM
Reprinted from Proceedings, Series A, **61**, No. 1 and Indag. Math., **20**, No. 1, 1958

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 I

0. *Introduction*

The Sieve of Erathostenes is an algorithm which yields the sequence of all primes. This paper deals with a family of somewhat similar algorithms for creating sequences of integers.

These algorithms depend on an initial integer λ and on an auxilary sequence B of integers b_k $(k=1, 2, \ldots)$ with $b_k \geqslant 2$. A family of intermediary sequences $A^{(i)}$ $(i=1, 2, \ldots)$ is formed, $A^{(i)}$ consisting of the integers $a_k^{(i)} (k=1, 2, \ldots)$. The sequence $A^{(1)}$ is defined by:

$$a_k^{(1)} = \lambda + k.$$

The sequence $A^{(i+1)}$ is obtained from the sequence $A^{(i)}$ by striking out all the terms of the form $a_{1+mb_i}^{(i)}$ $(m=0, 1, \ldots)$ and by renaming the remaining terms: $a_1^{(i+1)}, a_2^{(i+1)}, \ldots$. Finally, the sequence A consisting of integers a_k $(k=1, 2, \ldots)$ is defined by:

$$a_k = a_1^{(k)}.$$

Two examples of sieves will be considered: in the first the sequence B will be given in advance while in the second it will be determined by the sieving process itself.

In the first example we will take $b_k = k+1$. If we choose $\lambda = 0$, the first terms of the first several sequences $A^{(i)}$ are (the numbers in bold are those to be struck out in the next sequence):

$A^{(1)}(b_1=2)$:	**1**	*2*	**3**	*4*	**5**	*6*	**7**	*8*	**9**	*10*	**11**	*12*	**13**	*14*	**15**	*16*	**17**	*18*	**19**	*20*	**21**	*22*	**23**	*24*	**25**	*26*	**27**	*28*	**29**	*30*
$A^{(2)}(b_2=3)$:		**2**		*4*		*6*		**8**		*10*		*12*		**14**		*16*		*18*		**20**		*22*		*24*		**26**		*28*		*30*
$A^{(3)}(b_3=4)$:				**4**		*6*				*10*		*12*				**16**		*18*				*22*		*24*				**28**		*30*
$A^{(4)}(b_4=5)$:						**6**				*10*		*12*						*18*				*22*		**24**						*30*
$A^{(5)}(b_5=6)$:										**10**		*12*						*18*				*22*								*30*

Therefore, in this case $a_1 = 1$, $a_2 = 2$, $a_3 = 4$, $a_4 = 6$, $a_5 = 10$. Also $a_6 = 12$, $a_7 = 18$, $a_8 = 22$ and $a_9 = 30$ because these numbers will no longer be struck out in the following sequences until each one of them reaches the head of the line.

In the second example we shall take $b_k = a_k$. If we choose $\lambda = 1$, the first terms of the first several sequences are:

$A^{(1)}(b_1=2)$:	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$A^{(2)}(b_2=3)$:		3		5		7		9		11		13		15		17		19		21		23		25		27		29	
$A^{(3)}(b_3=5)$:				5		7				11		13				17		19				23		25				29	
$A^{(4)}(b_4=7)$:						7				11		13				17						23		25				29	

Therefore, in this case: $a_1 = 2$, $a_2 = 3$, $a_3 = 5$, $a_4 = 7$ and also $a_5 = 11$, $a_6 = 13$, $a_7 = 17$, $a_8 = 23$, $a_9 = 25$ and $a_{10} = 29$.

The following are the principal results obtained.

1. General explicit formulas giving a_k in terms of the $b_i (i = 1, 2, \ldots, k-1)$ are found and various more or less precise estimates of a_k are given under different restrictive assumptions as to the nature of the sequence B. The various estimates of a_k will be known respectively as the zero-step, the one-step and the multi-step estimate.

2. For $b_k = k+1$ it is shown using the more precise multi-step estimates that:

$$a_k = \frac{k^2}{\pi} + O(k^{4/3}).$$

3. For $b_k = a_k$ it is shown that $a_k \sim k \log k$. The a_k are in this case (for every λ) asymptotic to the primes. The proof has some similarity to that of the prime number theorem and makes use of the Tauberian self regulation of the sieving process. However, because of the greater regularity of the present process as compared to the Erathostenes method, the asymptotic formula for a_k is obtained much more easily than that for the primes.

4. Again for $b_k = a_k$, using the one-step formulas, it is shown that:

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

and hence it will follow that $a_k = k \log k + \frac{1}{2} k (\log \log k)^2 + o(k \log \log k)^2$. Thus it is seen that for large k we have $a_k > p_k$. It was surmised by Eri Jabotinsky in a paper read to the 1953-meeting of the Israel Mathematical Society, on heuristic grounds, that $a_k \sim p_k$ and that a_k oscillates around p_k. The second surmise is thus proven to be wrong.

5. Again for $b_k = a_k$ using the more precise multi-step estimates for a_k it is shown that (γ Euler's constant)

$$\frac{a_k}{k} = \prod_{a_i < k} \frac{a_i}{a_i - 1} - (1-\gamma) + o(1).$$

From this it is deduced that

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

1. *General Explicit Formulas*

We shall denote by $\lceil X \rceil$ the smallest integer $\geqslant X$.

Considering the generation of the sequence $A^{(i+1)}$ from the sequence $A^{(i)}$, we see that:

$$a_k^{(i+1)} = a^{(i)}_{\lceil \frac{b_i}{b_i-1} k \rceil}.$$

Repeating this reasoning i times we find that:

$$a_k^{(i+1)} = a_{k'}^{(1)} \text{ with } k' = \left\lceil \frac{b_1}{b_1-1} \left\lceil \frac{b_2}{b_2-1} \left\lceil \cdots \left\lceil \frac{b_i}{b_i-1} k \right\rceil \cdots \right\rceil \right\rceil \right\rceil.$$

Using the fact that $a_{k'}^{(1)} = \lambda + k'$ and $a_1^{(i+1)} = a_{i+1}$ we get, putting $i+1=k$:

$$(1) \qquad \begin{cases} a_1 = \lambda + 1, \\ a_k = \lambda + \left\lceil \frac{b_1}{b_1-1} \left\lceil \frac{b_2}{b_1-1} \left\lceil \cdots \left\lceil \frac{b_{k-1}}{b_{k-1}-1} \right\rceil \cdots \right\rceil \right\rceil \right\rceil, \end{cases}$$

which is the first of the announced explicit formulas for a_k. We have:

$$\frac{b_j}{b_j-1} m \leqslant \left\lceil \frac{b_j}{b_j-1} m \right\rceil < 1 + \frac{b_j}{b_j-1} m.$$

Applying this to (1) we deduce the following estimate for $a_k (k \geqslant 2)$:

$$\lambda + \prod_{i=1}^{k-1} \left(\frac{b_i}{b_i-1} \right) \leqslant a_k < \lambda + \prod_{i=1}^{k-1} \left(\frac{b_i}{b_i-1} \right) + \sum_{s=1}^{k-2} \left[\prod_{i=1}^{s} \left(\frac{b_i}{b_i-1} \right) \right],$$

or, because $\frac{b_i}{b_i-1} > 1$:

$$(2) \qquad \lambda + \prod_{i=1}^{k-1} \left(\frac{b_i}{b_i-1} \right) \leqslant a_k < \lambda + (k-1) \prod_{i=1}^{k-1} \left(\frac{b_i}{b_i-1} \right).$$

This formula will be called the zero-step estimate for a_k.

Formula (1) can be used for actual computation of the a_k. Thus, in the case $b_k = k+1$, $\lambda = 0$, let us for example compute a_9. We have:

$$a_9 = \left\lceil \frac{2}{1} \left\lceil \frac{3}{2} \left\lceil \frac{4}{3} \left\lceil \frac{5}{4} \left\lceil \frac{6}{5} \left\lceil \frac{7}{6} \left\lceil \frac{8}{7} \left\lceil \frac{9}{8} \right\rceil \cdots \right\rceil.$$

Now:

$$\left\lceil \frac{9}{8} \right\rceil = 2, \ \left\lceil \frac{8}{7} \cdot 2 \right\rceil = 3, \ \left\lceil \frac{7}{6} \cdot 3 \right\rceil = 4, \ \left\lceil \frac{6}{5} \cdot 4 \right\rceil = 5, \ \left\lceil \frac{5}{4} \cdot 5 \right\rceil = 7,$$

$$\left\lceil \frac{4}{3} \cdot 7 \right\rceil = 10, \ \left\lceil \frac{3}{2} \cdot 10 \right\rceil = 15 \text{ and } \left\lceil \frac{2}{1} \cdot 15 \right\rceil = 30.$$

Therefore $a_9 = 30$.

We see that a_9 is the last of the sequence of integers: 2, 3, 4, 5, 7, 10, 15, 30, which are the values of the successive brackets $\lceil \ \rceil$. We note that these integers first increase by 1, then by 2, then by more and more. If we knew how many brackets produce an increase by 1, how many by 2 and so on, we would get much shorter expressions for a_k.

In the general case, *let Q be the smallest integer such that*:

$$b_{k-Q} - 1 < Q.$$

Then the first $Q-1$ brackets give an increase by 1 and we have for every $q \leqslant Q$:

$$a_k = \lambda + \left\lceil \frac{b_1}{b_1-1} \left\lceil \frac{b_2}{b_2-1} \left\lceil \dots \left\lceil \frac{b_{k-q}}{b_{k-q}-1}\, q \right\rceil \dots \right\rceil \right\rceil \right\rceil . \tag{3}$$

Formula (3) takes into account the brackets which increase by steps of 1 and can be called, "the explicit formula for the one-step". By the same token formula (1) is "the explicit formula for the zero-step".

Formula (3) with $q=Q$ leads to a second estimate for a_k. Namely:

$$a_k = \lambda + [k - \theta\,(k-Q)] \prod_{i=1}^{k-Q} \left(\frac{b_i}{b_i-1}\right), \text{ with } 0 < \theta \leqslant 1, \tag{4}$$

which is the one-step estimate for a_k. Explicit formulas for steps 2, 3 and so on are conveniently established under the restrictive assumption that B is *a non-decreasing sequence* ($b_{k+1} \geqslant b_k$ for all k). This assumption happens to hold in the two particular examples considered by us. Then the explicit formula for m-step is (for $m=1, 2, \dots$):

$$a_k = \lambda + \left\lceil \frac{b_1}{b_1-1} \left\lceil \frac{b_2}{b_2-1} \left\lceil \dots \left\lceil \frac{b_{k-q_m}}{b_{k-q_m}-1} \left\{ m q_m - \sum_{i=0}^{m-1} q_i \right\} \right\rceil \dots \right\rceil \right\rceil \right\rceil , \tag{5}$$

where $q_0 = 0$ and q_m is the smallest integer for which:

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

The proof by induction is immediate. We note that $(q_n - q_{n-1})$ is the number of brackets which give increases of n. For $m=1$ formula (5) becomes formula (3) with $Q=q_1$. Formula (5) leads to the m-step estimate of a_k. Indeed, from (5):

$$\lambda + \left[\prod_{i=1}^{k-q_m} \left(\frac{b_i}{b_i-1}\right)\right] (m q_m - \sum_{i=0}^{m-1} q_i) \leqslant a_k < \lambda + \left[\prod_{i=1}^{k-q_m} \left(\frac{b_i}{b_i-1}\right)\right] (m q_m - \sum_{i=0}^{m-1} q_i) + R$$

with:

$$R = \sum_{s=1}^{k-q_m-1} \left(\prod_{i=1}^{s} \frac{b_i}{b_i-1}\right) < (k-q_m) \left[\prod_{i=1}^{k-q_m} \left(\frac{b_i}{b_i-1}\right)\right],$$

and therefore:

$$a_k = \lambda + \left[\prod_{i=1}^{k-q_m} \left(\frac{b_i}{b_i-1}\right)\right] \cdot \left[m q_m - \sum_{i=0}^{m-1} q_i + \theta\,(k-q_m) \right], \text{ with } 0 \leqslant \theta < 1, \tag{6}$$

which is the multi-step estimate for a_k.

Another result of a different kind which will be needed holds *in the* *particular case when* $\lim_{k\to\infty} \frac{b_k}{k} = \infty$ *(but the assumption that* $b_{k+1} \geqslant b_k$ *for all* k *is not required)*, namely if $\frac{b_k}{k} \to \infty$ then:

(7) $$\frac{a_k}{k} = [1+0(1)] \prod_{i=1}^{k} \left(\frac{b_i}{b_i - 1}\right).$$

Indeed, in this case the smallest Q for which $b_{k-Q} < Q+1$ is such that $k-Q = o(k)$ and inequality (4) becomes:

$$\frac{a_k}{k} = [1+o(1)] \prod_{i=1}^{k-Q} \left(\frac{b_i}{b_i - 1}\right).$$

All that has to be proven is that:

$$\prod_{i=k-Q+1}^{k} \left(\frac{b_i}{b_i-1}\right) = 1+o(1).$$

This will follow if we prove that:

(7′) $$\sum_{i=k-Q+1}^{k} \frac{1}{b_i} = o(1).$$

Choose any small $\varepsilon > 0$ and write:

$$\sum_{i=k-Q+1}^{k} \frac{1}{b_i} = \sum_{i=k-Q+1}^{[\varepsilon k]} \frac{1}{b_i} + \sum_{i=[\varepsilon k]+1}^{k} \frac{1}{b_i}.$$

We have:

$$\sum_{i=k-Q+1}^{[\varepsilon k]} \frac{1}{b_i} < \frac{[\varepsilon k]}{Q} < \frac{\varepsilon k}{k - \varepsilon k} < 2\varepsilon,$$

because by definition, Q is the smallest integer such that $b_{k-Q} < Q+1$ and so $b_{k-j} > j$ for $j < Q$ and here $b_i > k - i \geqslant Q+1$.

Also from $\frac{b_k}{k} \to \infty$ we have:

$$\sum_{i=[\varepsilon k]+1}^{k} \frac{1}{b_i} < \frac{k}{\min\limits_{\varepsilon k < i \leqslant k} b_i} = o(1),$$

which completes the proof of (7′).

2. *The case* $b_k = k+1$, $\lambda = 0$

We shall need an auxiliary sequence α_i $(i = 1, 2, \ldots)$ of numbers defined by the recurrence relation:

(8) $$m\left(1 - \sum_{i=1}^{m} \alpha_i\right) = \sum_{i=1}^{m} i\,\alpha_i \qquad (\text{for } m = 1, 2, \ldots).$$

By a suitable combination of relation (8) for m, $(m+1)$ and $(m+2)$, one sees readily that:

(9) $$\alpha_m = \frac{1}{m \cdot 2^{2m-1}} \binom{2m-2}{m-1},$$

and:

$$(10)\qquad 1-\sum_{i=1}^{m}\alpha_i=\frac{1}{2^{2m}}\binom{2m}{m}.$$

It is well known that:

$$(11)\qquad \frac{1}{2^{2m}}\binom{2m}{m}=\frac{1}{\sqrt{\pi m}}\left[1+O\left(\frac{1}{m}\right)\right].$$

We shall now use formula (5) which becomes:

$$(12)\qquad a_k=\left\lceil\frac{2}{1}\left\lceil\frac{3}{2}\left\lceil\cdots\left\lceil\frac{k-q_m+1}{k-q_m}\left\{mq_m-\sum_{i=0}^{m-1}q_i\right\}\right\rceil\cdots\right\rceil\right\rceil\right\rceil,$$

where the q_m are defined as follows: $q_0=0$ and q_m is the smallest integer for which:

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

Therefore q_m is an integer for which:

$$(13)\qquad mk+1\leqslant 2m\,q_m-\sum_{i=0}^{m-1}q_i<mk+2m.$$

We now use the sequence α_i and define the numbers $\theta_i^{(k)}$ by the relation:

$$(14)\qquad q_i-q_{i-1}=k\alpha_i+\theta_i^{(k)}.$$

From (14) we deduce that:

$$(15)\qquad q_m=k\left(\sum_{i=1}^{m}\alpha_i\right)+\sum_{i=1}^{m}\theta_i^{(k)},$$

and:

$$(16)\qquad mq_m-\sum_{i=1}^{m-1}q_i=k\left(\sum_{i=1}^{m}i\,\alpha_i\right)+\sum_{i=1}^{m}i\,\theta_i^{(k)},$$

or, using (13), (15), (16) and (8) we obtain by a simple computation:

$$(17)\qquad 1\leqslant\sum_{i=1}^{m}(m+i)\,\theta_i^{(k)}\leqslant 2m.$$

We note that the sum in inequality (17) is an integer because the middle term in (13) is an integer. We shall not use this fact.

Formula (12) now leads to the following estimate for a_k analogous to (6):

$$a_k=(k-q_m+1)\left[mq_m-\sum_{i=0}^{m-1}q_i+\theta(k-q_m)\right],\ \text{with}\ 0\leqslant\theta<1,$$

or, using (15), (16), (8) and (10):

$$a_k=\left[k\binom{2m}{m}\cdot\frac{1}{2^{2m}}-\sum_{i=1}^{m}\theta_i^{(k)}+1\right]\cdot\left[k\binom{2m}{m}\frac{m}{2^{2m}}+\sum_{i=1}^{m}i\theta_i^{(k)}\right]+\theta'(k-q_m)^2,$$

or again by (15):

$$(18)\quad \begin{cases} a_k = k^2 \cdot \frac{m}{2^{4m}} \cdot \binom{2m}{m}^2 - k\binom{2m}{m} \cdot \frac{1}{2^{2m}} \sum_{i=1}^{m} (m-i)\,\theta_i^{(k)} + \\ + \sum_{i=1}^{m} i\,\theta_i^{(k)} \left(1 - \sum_{i=1}^{m} \theta_i^{(k)}\right) + \frac{mk}{2^{2m}}\binom{2m}{m} + \theta' \left[k\binom{2m}{m} \cdot \frac{1}{2^{2m}} - \sum_{i=1}^{m} \theta_i^{(k)}\right]^2, \end{cases}$$

with $0 < \theta' < 2$.

We now need estimates for $\sum_{i=1}^{m} \theta_i^{(k)}$, for $\sum_{i=1}^{m} i\,\theta_i^{(k)}$ and for $\sum_{i=1}^{m} (m-i)\,\theta_i^{(k)}$. Inequality (17) yields easily:

$$(19)\quad \begin{cases} 0 < \sum_{i=1}^{m} \theta_i^{(k)} < 2, \\ -2m < \sum_{i=1}^{m} i\,\theta_i^{(k)} < 2m, \\ -2m < \sum_{i=1}^{m} (m-i)\,\theta_i^{(k)} < 4m. \end{cases}$$

Indeed, the first inequality of (19) holds for $m=1$ and by induction for all m and the two other inequalities follow from this and (17).

(All these are not the best possible inequalities.) Using (11) and (19) in (18) we find:

$$a_k = \frac{k^2}{\pi}\left[1 + O\left(\frac{1}{m}\right)\right] + kO(\sqrt{m}) - O(m) + O\left(\frac{k^2}{m}\right) + \frac{mk}{\sqrt{\pi m}} + O(k).$$

Choosing $m = [k^{2/3}]$ we find:

$$(20)\quad a_k = \frac{k^2}{\pi} + O(k^{4/3}).$$

A sharpening of the inequalities (19) should lead to the reduction of the index $\frac{4}{3}$. Numerical evidence indicates that:

$$a_k = \frac{k^2}{\pi} + O(k).$$

3. *The case* $b_k = a_k$ *(and any* $\lambda \geqslant 1$*)*

We shall show that in this case $a_k/k \to \infty$ so that formula (7), which now becomes:

$$(21)\quad \frac{a_k}{k} = [1 + o(1)] \prod_{i=1}^{k} \left(\frac{a_i}{a_i - 1}\right),$$

holds. Indeed, $a_k - \lambda \geqslant k$ and a_k is an unbounded and increasing function of k. Let Q be the smallest integer such that $a_{k-Q} < Q+1$. Then Q is such that $k - Q - \lambda < Q + 1$ and:

$$\frac{Q}{k} > \frac{1}{2} - \frac{\lambda+1}{2k}.$$

Formula (4) now becomes, by using first $\theta=0$ and then $\theta=1$:

$$(22)\qquad \prod_{i=1}^{k-Q}\left(\frac{a_i}{a_i-1}\right) \geqslant \frac{a_k-\lambda}{k} > \left(\frac{1}{2}-\frac{\lambda+1}{2k}\right)\prod_{i=1}^{k-Q}\left(\frac{a_i}{a_i-1}\right).$$

When $k \to \infty$ both Q and $(k-Q) \to \infty$. If the product in (22) converged, a_k/k would be bounded, but then the product would diverge. Therefore the product diverges, so that a_k/k tends to ∞ and Q/k to 1.

From (21) we easily see that:

$$\frac{a_{k+1}}{k+1}-\frac{a_k}{k}=o\left(\frac{a_k}{k}\right)$$

and thus that a_k/k changes slowly.

We now want to show that:

$$(23)\qquad \liminf \frac{a_k}{k\log k} \leqslant 1.$$

Indeed we would otherwise have:

$$a_k > (1+\delta)\, k \log k \quad \text{for } k > k_0 \quad \text{with } \delta > 0,$$

and so from (21):

$$(1+\delta)\log k < c_1 \prod_{i=1}^{k}\left(1-\frac{1}{(1+\delta)i\log i}\right)^{-1} = c_2\, e^{\sum_{i=1}^{k}\frac{1}{(1+\delta)i\log i}}.$$

Since:

$$\sum_{i=1}^{k}\frac{1}{i\log i} = c_3+\log\log k+o(1),$$

we would have:

$$(1+\delta)\log k < c_4\, e^{\frac{\log\log k}{1+\delta}} = c_4\,(\log k)^{\frac{1}{1+\delta}},$$

which is wrong. Similarly it can be shown that:

$$(24)\qquad \limsup \frac{a_k}{k\log k} \geqslant 1.$$

Suppose now that:

$$\limsup \frac{a_k}{k\log k} = 1+c \quad (c>0).$$

Then, because of the slowness of change of a_k/k and because of (23), for a suitable choice of c_1 and c_2 so that $0<c_1<c_2<c$, there exist two numbers k_1 and k_2 such that:

$$\frac{a_{k_1}}{k_1\log k_1} < 1+c_1 \text{ and } \frac{a_k}{k\log k} > 1+c_1 \text{ for } k_1 < k < k_2,$$

and:

$$\frac{a_{k_2}}{k_2\log k_2} > 1+c_2 \text{ and } \frac{a_k}{k\log k} < 1+c_2 \text{ for } k_1 < k < k_2.$$

We have:

$$\frac{a_{k_2}}{k_2} : \frac{a_{k_1}}{k_1} > \frac{(1+c_2)\log k_2}{(1+c_1)\log k_1}. \tag{25}$$

But from (21) we have:

$$\frac{a_k}{k_2} : \frac{a_{k_1}}{k_1} = [1+o(1)] \prod_{i=k_1}^{k_2} \left(\frac{a_i}{a_i-1}\right) = [1+o(1)]\, e^{\sum_{i=k_1}^{k_2} \frac{1}{a_i}} \leqslant$$

$$\leqslant [1+o(1)]\, e^{\sum_{i=k_1}^{k_1} \frac{1}{(1+c_1)\, i \log i}} = [1+o(1)] \left(\frac{\log k_2}{\log k_1}\right)^{\frac{1}{1+c_1}},$$

which contradicts (25). Therefore:

$$\lim_{k\to\infty} \sup \frac{a_k}{k \log k} = 1.$$

Similarly it can be shown that:

$$\lim_{k\to\infty} \inf \frac{a_k}{k \log k} = 1,$$

and hence:

$$\lim_{k\to\infty} \frac{a_k}{k \log k} = 1, \tag{26}$$

as stated in the introduction.