SIEVE-GENERATED SEQUENCES

M. C. WUNDERLICH

We shall consider a generalization of the sieve process introduced by W. E. Briggs **(1)** in 1963. Let $A^{(1)}$ be the sequence $\{a_k^{(1)}\}$, where $a_k^{(1)} = k + 1$, so that $A^{(1)} = \{2, 3, 4, \ldots\}$. Suppose inductively that $A^{(1)}, A^{(2)}, \ldots, A^{(n)}$ has been defined. $A^{(n+1)}$ will be defined from $A^{(n)} = \{a_1^{(n)}, a_2^{(n)}, a_3^{(n)}, \ldots\}$ in the following manner: For each integer $t \geqslant 0$, choose an arbitrary element $\alpha_t^{(n)}$ from the set $\{a^{(n)}_{n+ta_n+1}, a^{(n)}_{n+ta_n+2}, \ldots, a^{(n)}_{n+ta_n+a_n}\}$, where $a_n = a_n^{(n)}$, and delete the elements $\alpha_t^{(n)}$ from $A^{(n)}$ to form $A^{(n+1)}$. The sequence A is defined to be the sequence $\{a_n\}$. It is also the set-theoretic intersection of all the sequences $A^{(n)}$, $n = 1, 2, \ldots$. Let $\mathfrak{R}$ be the class of all sequences that can be generated by this sieve process.

Sequences of this nature have been studied by S. Ulam **(3)**, P. Erdös **(2)**, D. Hawkins **(5)**, B. Lachapelle **(6)**, and most recently by W. E. Briggs **(1)**. The principal purpose of these studies was to determine whether or not $a_n \sim n \log n$ (as is the case with the sequence of prime numbers). In this paper, the author presents a criterion characterizing all the sequences in $\mathfrak{R}$ for which $a_n \sim n \log n$.

For the remainder of this paper, $\{a_n\}$ is considered to be an arbitrary sequence in $\mathfrak{R}$, and $A^{(1)}, A^{(2)}, \ldots$ are the successive sequences obtained in the sieve process generating $\{a_n\}$.

Definition 1.

(a) $R_n(x)$ is the number of elements in $A^{(n)}$ not exceeding x.

(b) $\sigma_n = \prod_{k=1}^{n} \left(1 - \frac{1}{a_k}\right)$.

(c) $f_k(x) = R_k(x) - R_{k+1}(x)$.

(d) $l(n)$ is the number of k for which $f_k(a_n) = 1$.

(e) $t(n)$ is the greatest k such that $f_k(a_n) \geqslant 2$.

(f) $d(n) = n/(n + l(n))$.

THEOREM 1. *$a_n \sim n \log n$ if and only if*

$$\sum_{k=2}^{n} \frac{d(k)}{k} \sim d(n) \log n. \tag{1}$$

The proof of this theorem is contained in a sequence of lemmas.

Received November 13, 1964. The research for this paper was supported in part by N.S.F. contract GP-1823.

LEMMA 1.1. *If* $x < a_n$, $R_{n+1}(x) = R_n(x)$. *If* $x \geqslant a_n$,

$$R_{n+1}(x) = \sigma_n R_1(x) + \sum_{k=1}^{n} \frac{\sigma_n}{\sigma_k}\left(\left\{\frac{R_k(x) - k}{a_k}\right\} + \frac{k}{a_k} - \epsilon_k\right), \tag{2}$$

where ϵ_k *is either* 0 *or* 1, *and* $\{x\}$ *refers to the fractional part of* x.

Proof. The first part is obvious from the definition of the sieve process. To prove the second assertion, note that when $x \geqslant a_n$,

$$f_n(x) = \left[\frac{R_n(x) - n}{a_n}\right] + \epsilon_n,$$

where $\epsilon_n = 0$ or 1, and $[x]$ means the greatest integer in x. Hence

$$\begin{aligned} R_{n+1}(x) &= R_n(x) - \frac{R_n(x) - n}{a_n} + \left\{\frac{R_n(x) - n}{a_n}\right\} - \epsilon_n \\ &= R_n(x)\left(1 - \frac{1}{a_n}\right) + \frac{n}{a_n} + \left\{\frac{R_n(x) - n}{a_n}\right\} - \epsilon_n. \end{aligned} \tag{3}$$

The lemma then follows by iteration in (3).

LEMMA 1.2. $\sigma_n a_n = n - E_n(a_n + 1)$, *where*

$$E_n(x) = \sum_{k=1}^{n} \frac{\sigma_n}{\sigma_k}\left(\left\{\frac{R_k(x) - k}{a_k}\right\} + \frac{k}{a_k} - \epsilon_k\right).$$

Proof. Let $x = a_n + 1$ in (2), and note that since a_1 is always 2, $a_n + 1 \neq a_{n+1}$, and therefore $R_{n+1}(a_n + 1) = n$.

LEMMA 1.3. *There exists a constant* c_1 *such that for all* n *sufficiently large*

$$a_n > c_1 n \log n.$$

Proof. The argument used to prove this lemma is completely analogous to the proof by W. E. Briggs of (**1**, formula (9)). Since $a_2 \geqslant 3$, $a_k \geqslant 3k/2$. Hence since $0 \leqslant \sigma_n/\sigma_k \leqslant 1$,

$$-1 < \frac{\sigma_n}{\sigma_k}\left(\left\{\frac{R_k(x) - k}{a_k}\right\} + \frac{k}{a_k} - \epsilon_k\right) < \frac{5}{3}$$

and hence

$$-n < E_n < 5n/3.$$

Therefore, from Lemma 1.2, $\sigma_n a_n < 2n$ for $n > 1$. Noting that

$$\frac{1}{\sigma_k} - \frac{1}{\sigma_{k-1}} = \frac{1}{a_k \sigma_k},$$

we obtain by summing from 2 to n,

$$\frac{1}{\sigma_n} - \frac{1}{\sigma_1} > \sum_{k=2}^{n} \frac{1}{2k} > \frac{1}{2}\log n - 2,$$

or, for sufficiently large n,

$$(4) \qquad 1/\sigma_n > \tfrac{1}{2}\log n.$$

Now clearly for any p and q, $p \leqslant R_q(a_p + 1)$ so that noting that

$$R_{n+1}(x) = \sigma_n([x] - 1) + E_n(x),$$

we have

$$2n \leqslant R_{n+1}(a_{2n} + 1) = \sigma_n a_{2n} + E_n(a_{2n} + 1) \leqslant \sigma_n a_{2n} + 5n/3$$

and

$$2n - 1 \leqslant R_{n+1}(a_{2n-1} + 1) = \sigma_n a_{2n-1} + E_n(a_{2n-1}) \leqslant \sigma_n a_{2n-1} + 5n/3.$$

The lemma then follows from these inequalities and (4).

LEMMA 1.4. *There exists a constant* c_2 *such that* $t(n) < c_2 n/\log n$.

Proof. Let $k = t(n) - 1$ so that $f_k(a_n) \geqslant 2$. Since for all $k' > k, f_{k'}(a_n) \leqslant 1$, $R_k(a_n) < n + (n - k) < 2n$. Also, $R_k(a_n) > a_k + k > a_k$ so that $a_k < 2n$. Hence, applying Lemma 1.3, $2n > \frac{1}{2}c_1 k \log k$, so that $t(n) - 1 < cn/\log n$ for some constant c. Hence there exists a constant c_2 for which $t(n) < c_2 n/\log n$.

LEMMA 1.5. $E_n(a_n + 1) = -l(n) + o(n)$.

Proof. If we let

$$E(k, n) = \frac{\sigma_n}{\sigma_k}\left(\left\{\frac{R_k(a_n + 1) - k}{a_k}\right\} + \frac{k}{a_k} - \epsilon_k\right),$$

we can write

$$(5) \qquad E_n(a_n + 1) = \sum_k{}' E(k, n) + \sum_k{}'' E(x, n) + \sum_{k \leqslant c_2 n/\log n} E(k, n)$$

where the first sum is taken over those k for which $k > c_2 n/\log n$ and $f_k(a_n) = 0$, and the second sum is taken over those k for which $k > c_2 n/\log n$ and $f_k(a_n) = 1$.

Since $E(k, n)$ is bounded and $\sigma_n/\sigma_k \leqslant 1$, we have

$$(6) \qquad \sum_{k \leqslant c_2 n/\log n} E(k, n) = o(n).$$

For k in the range of $\sum'$, we have

$$\frac{R_k(a_n + 1) - k}{a_k} < 1$$

so that, by Lemma 1.3 and since $R_k(a_n + 1) < 2n$,

$$E(k, n) = \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n + 1)}{a_k} - \epsilon_k\right) < \frac{2n}{c_1 k \log k}$$

and so

$$\sum_k{}' E(k, n) < \sum_k{}' \frac{2n}{c_1 k \log k} < \sum_{k = c_2 n/\log n}^{n} \frac{2n}{c_1 k \log k}.$$

https://doi.org/10.4153/CJM-1966-031-4 Published online by Cambridge University Press

One can easily verify that

$$(7)\qquad \sum_{k=c_2n/\log n}^{n} \frac{2n}{c_1 k \log k} = O\left(\frac{n \log\log n}{\log n}\right) = o(n).$$

Hence

$$(8)\qquad \sum_k{}' E(k, n) = o(n).$$

Finally, for all k in the range of $\sum''$, we have that

$$\left[\frac{R_k(a_n+1)-k}{a_k}\right] + \epsilon_k = 1.$$

If $\epsilon_k = 1$, we have $(R_k(a_n+1) - k)/a_k < 1$ so that

$$(9)\qquad E(k, n) = \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n+1)}{a_k} - 1\right).$$

If $\epsilon_k = 0$, we have $1 \leqslant (R_k(a_n+1) - k)/a_k < 2$, or

$$\left\{\frac{R_k(a_n+1)-1}{a_k}\right\} = \frac{R_k(a_n+1)-k}{a_k} - 1.$$

Hence

$$(10)\qquad E(k, n) = \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n+1)}{a_k} - 1\right).$$

Since $k > c_2 n/\log n$,

$$\begin{aligned} 1 \geqslant \frac{\sigma_n}{\sigma_k} = \prod_{i=k+1}^{n}\left(1 - \frac{1}{a_i}\right) &> \prod_{i=c_2n/\log n}^{n}\left(1 - \frac{1}{c_1 i \log i}\right) \\ &= \exp\left(\sum_{i=c_2n/\log n}^{n} \log\left(1 - \frac{1}{c_1 i \log i}\right)\right) \\ &= \exp\left(O\left(\sum_{i=n/\log n}^{n} \frac{1}{i \log i}\right)\right) \\ &= \exp\left(O\left(\frac{\log\log n}{\log n}\right)\right) = 1 + o(1). \end{aligned}$$

Hence $\sigma_n/\sigma_k = 1 + o(1)$, and so from (9) and (10),

$$\sum_k{}'' E(k, n) = \sum_k{}'' \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n+1)}{a_k}\right) - \sum_k{}'' 1 + o(1).$$

But

$$\sum_k{}'' \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n+1)}{a_k}\right) < \sum_{k=c_2n/\log n}^{n} \frac{2n}{c_1 k \log k} = O\left(\frac{n \log\log n}{\log n}\right) = o(n)$$

by (7). Also,

$$\sum_k{}'' 1 + o(1) = l(n) + o(n).$$

Hence we get

$$\sum_k{}'' E(k, n) = -l(n) + o(n). \tag{11}$$

Now (5), (6), (8), and (11) prove the lemma.

We can now complete the proof of Theorem 1. From Lemmas 1.2 and 1.5, we have

$$\sigma_n a_n = n + l(n) + o(n) \sim n + l(n) = n/d(n)$$

and since $\frac{1}{2} \leqslant d(n) \leqslant 1$, one can verify that

$$\sum_{k=2}^{n} \frac{1}{\sigma_k a_k} \sim \sum_{k=2}^{n} \frac{d(k)}{k}. \tag{12}$$

Let $c(n) = a_n/n \log n$. Then we have

$$\frac{1}{\sigma_n} \sim \frac{a_n d(n)}{n} = c(n)d(n) \log n. \tag{13}$$

However, since

$$\frac{1}{\sigma_k} - \frac{1}{\sigma_{k-1}} = \frac{1}{\sigma_k a_k},$$

we have, using (4),

$$\sum_{k=2}^{n} \frac{1}{a_k \sigma_k} = \frac{1}{\sigma_n} - \frac{1}{\sigma_1} \sim \frac{1}{\sigma_n}. \tag{14}$$

Hence from (12), (13), and (14),

$$c(n)d(n) \log n \sim \frac{1}{\sigma_n} \sim \sum_{k=2}^{n} \frac{1}{a_k \sigma_k} \sim \sum_{k=2}^{n} \frac{d(k)}{k}. \tag{15}$$

Thus, $c(n) \sim 1$ if and only if

$$d(n) \log n \sim \sum_{k=2}^{n} \frac{d(k)}{k},$$

which is the theorem.

One can now obtain theorems concerning the order of a_n, which are analogous to Chebychef's theorems regarding the order of $\pi(x)$.

THEOREM 2. *If ϵ is an arbitrary positive real number,*

$$\frac{1}{2} - \epsilon < \frac{a_n}{n \log n} < 2 + \epsilon \qquad \textit{for } n > n_0(\epsilon).$$

Proof. Since $\frac{1}{2} \leqslant d(n) < 1$,

$$\frac{1}{2} \log n \leqslant \sum_{k=2}^{n} \frac{d(k)}{k} < \log n, \qquad n > n_1.$$

Hence, from (15),

$$(\tfrac{1}{2} - \epsilon) \log n \leqslant c(n)d(n) \log n \leqslant (1 + \epsilon) \log n$$

for $n > n_0$, or

$$\tfrac{1}{2} - \epsilon < c(n)d(n) < 1 + \epsilon \qquad \text{for } n > n_0.$$

But since $\frac{1}{2} < d(n) < 1$, we have

$$c(n) > (\tfrac{1}{2} - \epsilon)/d(n) > \tfrac{1}{2} - \epsilon$$

and

$$c(n) < (1 + \epsilon)/d(n) < (1 + \epsilon)2,$$

which proves the theorem.

THEOREM 3.

$$\liminf \frac{a_n}{n \log n} \leqslant 1 \text{ and } \limsup \frac{a_n}{n \log n} \geqslant 1.$$

Proof. To prove the first assertion, suppose again that $c(n) = a_n/(n \log n)$ and suppose that $\liminf c(n) = 1 + \epsilon$, where $\epsilon > 0$. Then from (15)

$$\sum_{k=2}^{n} \frac{d(k)}{k} > (1 + \epsilon)d(n) \log n + o(\log n).$$

Let $\delta = \limsup d(n)$. Clearly $\frac{1}{2} < \delta < 1$. Also

$$\sum_{k=2}^{n} \frac{d(k)}{k} \leqslant \sum_{k=2}^{n} \frac{\delta + o(1)}{k} < \delta \log n + o(\log n).$$

Hence $(1 + \epsilon)d(n) \leqslant \delta + o(1)$ or

$$d(n) < \frac{\delta}{1 + \epsilon} + o(1),$$

which contradicts the choice of δ. The proof of the second assertion is similar.

It is now possible to use Theorem 1 to obtain a subclass of $\mathfrak{R}$ for which $a_n \sim n \log n$ holds and one for which it does not hold. The asymptotic character of $l(n)/n$ is affected only by the first element eliminated at each execution of the sieve process. Therefore in order to produce these subclasses, it is necessary only to specify the element $\alpha_0^{(n)}$ eliminated from the interval

$$\{a_{n+1}^{(n)}, a_{n+2}^{(n)}, \ldots, a_{n+a_n}^{(n)}\}$$

at the nth sieving. We shall define r_n by supposing that

$$\alpha_0^{(n)} = a_{n+r_n}^{(n)}, \qquad \text{where } 1 \leqslant r_n \leqslant a_n.$$

Alternatively, r_n can be defined as $k - n$, where $a_k^{(n)}$ is the smallest element eliminated from $A^{(n)}$ to form $A^{(n+1)}$.

Definition 2. Let the sequences $\{s_n\}$ and $\{t_n\}$ be defined as follows: $s_1 = 3$, $t_k = s_k \log s_k$; $s_{k+1} = t_k \log t_k$ for $k = 1, 2, \ldots$. Define r_k as follows:

$$r_k = \begin{cases} a_k & \text{for } s_j \leqslant k < t_j, \\ 1 & \text{for } t_j \leqslant k < s_{j+1}. \end{cases}$$

THEOREM 4. *If* $\mathfrak{R}_1$ *is the class of sieve-generated sequences such that* r_k *is defined as above, then if* $\{a_n\} \in \mathfrak{R}_1$, $a_n \sim n \log n$ *does not hold.*

Proof. For all k in the range $t_{j-1} \leqslant k < s_j$, $f_k(a_{s_j})$ is clearly equal to 1, and since $t_{j-1} \sim s_j/(\log s_j)$, we have

$$l(s_j) \sim s_j, \qquad \text{or } d(s_j) \sim \tfrac{1}{2}. \tag{16}$$

Furthermore, if k is in the range $s_j \leqslant k < t_j$ and $f_k(a_{t_j}) = 1$, then

$$R_k(a_{t_j}) > k + r_k > k + a_k > a_k.$$

But since $R_k(a_{t_j}) < 2t_j$, $2t_j > a_k > c_1 k \log k$. Hence for some constant c,

$$k < \frac{ct_j}{\log t_j} = o(t_j).$$

Hence

$$l(t_j) = o(t_j), \qquad \text{or } d(t_j) \sim 1. \tag{17}$$

Now suppose that $a_n \sim n \log n$. Then

$$\sum_{k=2}^{t_n} \frac{d(k)}{k} = \sum_{k=2}^{t_n} \frac{d(k)}{k} + \sum_{k=s_n+1}^{t_n} \frac{d(k)}{k}.$$

By Theorem 1 and (17)

$$\sum_{k=2}^{t_n} \frac{d(k)}{k} \sim d(t_n) \log t_n \sim \log t_n$$

$$= \log(s_n \log s_n) \sim \log s_n.$$

Also,

$$\sum_{k=2}^{s_n} \frac{d(k)}{k} \sim d(s_n) \log s_n \sim \frac{1}{2} \log s_n.$$

Hence

$$\sum_{k=s_n+1}^{t_n} \frac{d(k)}{k} \sim \frac{1}{2} \log s_n. \tag{18}$$

On the other hand,

$$(19) \qquad \sum_{k=s_n+1}^{t_n} \frac{d(k)}{k} < \sum_{k=s_n+1}^{t_n} \frac{1}{k}$$
$$= O(\log(s_n \log s_n) - \log s_n)$$
$$= O(\log s_n + \log\log s_n - \log s_n)$$
$$= O(\log\log s_n) = o(\log s_n).$$

But (18) and (19) are a contradiction.

Definition 3. Let $\mathfrak{R}_2$ be the class of sequences in $\mathfrak{R}$ for which the finite or infinite $\lim_{k\to\infty} (r_k/k)$ exists. Let this limit be denoted by r, where $r \in [0, \infty]$.

THEOREM 5. *If* $\{a_n\} \in \mathfrak{R}_2$, *then* $a_n \sim n \log n$. *Furthermore,*

$$\frac{1}{\sigma_n} \sim \frac{r+1}{r+2} \log n.$$

Proof. We must first obtain estimates for $l(n)$ and $d(n)$. If $f_k(a_n) = 0$ and $r < \infty$, then $k + r_k > R_k(a_n) \geqslant n$ so that

$$k > \frac{n}{1+r}(1 + o(1)) \quad \text{and} \quad l(n) > \frac{n}{1+r}(1+o(1)).$$

Secondly, if $f_k(a_n) = 1$ but $f_{k'}(a_n) = 0$ for all $k' > k$, then $k + r_k < R_k(a_n) = n + 1$ so that

$$k < \frac{n}{1+r}(1+o(1)).$$

Hence in every case

$$\frac{l(n)}{n} \sim \frac{1}{1+r}$$

or

$$(19) \qquad d(n) \sim \frac{1+r}{2+r}.$$

Thus, using Theorem 1, one can easily show that $a_n \sim n \log n$. Finally using (18) and (13), we obtain

$$\frac{1}{\sigma_n} \sim \frac{1+r}{2+r} \log n.$$

This theorem is certainly not the best possible. It is conjectured that $a_n \sim n \log n$ whenever the function r_k/k has a limiting distribution on the positive real line. The author can show that $d(k)$ is asymptotic to a constant whenever r_k/k has a limiting distribution and the distribution function is a finite-valued step function, but the methods used in the proof are very cumbersome.

References

1. W. E. Briggs, *Prime-like sequences generated by a sieve process*, Duke Math. J., *30* (1963), 297–312.
2. P. Erdös and E. Jabotinsky, *On sequences generated by a sieving process*, Proc. Konink. Nederl. Akad. von Wetensch., *61* (1958), 297–312.
3. V. Gardiner, R. Lazarus, M. Metropolis, and S. Ulam, *On certain sequences of integers defined by sieves*, Math. Magazine, *31* (1957/58), 1–3.
4. D. Hawkins, *The random sieve*, Math. Magazine, *31* (1957/58), 1–3.
5. D. Hawkins and W. E. Briggs, *The lucky number theorem*, Math. Magazine, *31* (1957/58), 277–280.
6. B. Lachapelle, *L'espérance mathématique du nombre de nombres premiers aléatoires inférieurs ou égaux à x*, Can. Math. Bull., *4* (1961), 139–142.

State University of New York, Buffalo