SECOND AND THIRD TERM APPROXIMATIONS OF SIEVE-GENERATED SEQUENCES[1]

BY

M. C. WUNDERLICH AND W. E. BRIGGS

We will consider sequences of natural numbers generated by the sieve process described in an earlier paper [2] by Wunderlich. In that paper, a criterion was presented which characterized the sieve generated sequences $\{a_n\}$ for which $a_n \sim n \log n$. The purpose of this paper is to investigate the nature of $a_n - n \log n$ for sequences which satisfy the above mentioned criterion. It was hoped that the authors could construct a sequence $\{a_n\}$ for which

$$a_n - n \log n \sim p_n - n \log n \sim n \log \log n$$

where p_n is the n-th prime. It is shown that this cannot be achieved by a sieve of this type but the methods employed do suggest a modification of the sieve process which may generate such a prime-like sequence.

For the sake of completeness, the sieve method and the related functions will be defined.

$$A = \{a_k\} = \bigcap_{k=1}^{\infty} A^{(k)}$$

where the $A^{(k)} = \{a_1^{(k)}, a_2^{(k)}, \cdots\}$ are sequences of natural numbers defined inductively as follows. $A^{(1)} = \{2, 3, 4, \cdots\}$, and $A^{(k+1)}$ is obtained from $A^{(k)}$ as follows: For each integer $t \geq 0$, choose one element

$$\alpha_t^{(k)} \in \{a_{k+ta_k+1}^{(k)}, \cdots, a_{k+ta_k+a_k}^{(k)}\}$$

where $a_k = a_k^{(k)}$. Delete these $\alpha_t^{(k)}$ from $A^{(k)}$ to form $A^{(k+1)}$. The following functions will be used:

(a) $R_n(x)$ is the number of elements in $A^{(n)}$ not exceeding x.
(b) $\sigma_n = \prod_{k=1}^{n} (1 - 1/a_k)$.
(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 largest k for which $f_k(a_n) \geq 2$.
(*f*) $d(n) = n/(n + l(n))$.

The following two lemmas from [2] will be used in this paper.

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

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

Received August 11, 1965.

[1] This research was sponsored by a National Science Foundation grant awarded to the University of Colorado.

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

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

We will begin by considering those a_n for which $a_n \sim n \log n$. Letting $x = a_n + 1$ in [2, Lemma 1.1], we obtain

$$\sigma_n a_n = n - E_n(a_n + 1). \tag{1}$$

We now proceed to estimate $E_n(a_n + 1)$

LEMMA 1. $E_n(a_n + 1) = -l(n) + O\left(\dfrac{n \log \log n}{\log n}\right)$.

Proof. Let c_2 be the constant obtained in [2, Lemma 1.4], and let

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

We split up $E_n(a_n + 1)$ as follows:

$$\begin{aligned} E_n(a_n + 1) &= \textstyle\sum_{k=2}^{s(n)} E(k, n) \\ &\quad + \textstyle\sum_{k>s(n), f_k(a_n)=1} E(k, n) + \textstyle\sum_{k>s(n), f_k(a_n)=0} E(k, n) \\ &= S_1 + S_2 + S_3 \end{aligned} \tag{2}$$

where $s_n = [c_2 n/\log n]$.

We first observe that since for all n and k, $E(k, n)$ is bounded, it follows that

$$S_1 = O(n/\log n). \tag{3}$$

Since $f_k(a_n) = 1$ for all k in the range of the second summation, we have that

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

If $\varepsilon_k = 1$, then $(R_k(a_n + 1) - k)/a_k < 1$ so that

$$E(k, n) = \frac{\sigma_n}{\sigma_k}\left(\frac{R_k(a_n + 1)}{a_k} - 1\right). \tag{4}$$

If $\varepsilon_k = 0$, then $1 \leq (R_k(a_n + 1) - k)/a_k < 2$ or

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

Hence (4) is true in either event. For k in this range one uses the method in (23) of [1] to obtain

$$1 \geq \frac{\sigma_n}{\sigma_k} > 1 - \frac{c \log \log n}{\log n}$$

for some constant c. Hence we can now write

$$
\begin{aligned}
S_2 &= \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right) \sum_{k>s(n), f_k(a_n)=1} \left(\frac{R_k(a_n+1)}{a_k} - 1\right) \\
&= \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)\left(\sum_{k>s(n), f_k(a_n)=1} \frac{O(n)}{a_k} - l(n) + O\left(\frac{n}{\log n}\right)\right)
\end{aligned}
$$

since the number of terms in this summation is $l(n) + O(n/\log n)$ and since $R_k(a_n + 1) = O(n)$. Multiplying and using the fact that $a_n \sim n \log n$, we obtain

$$
\begin{aligned}
S_2 &= \left(-l(n) + O\left(\frac{n\log\log n}{\log n}\right)\right) + O\left(\sum_{k=s_n}^{n} \frac{n}{k\log k}\right) \\
&= -l(n) + O\left(\frac{n\log\log n}{\log n}\right).
\end{aligned} \tag{5}
$$

Finally, for k in the range of the third summation,

$$f_k(a_n) = \left[\frac{R_k(a_n+1) - k}{a_k}\right] + \varepsilon_k = 0$$

so that

$$(R_k(a_n + 1) - k)/a_k < 1.$$

Hence

$$E(k, n) = (1 + o(1))\left(\frac{R_k(a_n+1)}{a_k}\right) = \frac{O(n)}{a_k}.$$

One easily proves that

$$S_3 = O\left(\frac{n\log\log n}{\log n}\right). \tag{6}$$

Combining (3), (5), and (6) completes the proof of the lemma.

LEMMA 2.

$$a_n = (n/d(n))\sum_{k=2}^{n} d(k)/k + O(n(\log\log n)^2).$$

Proof. Using (1) and Lemma 1, we obtain

$$
\begin{aligned}
\sigma_n a_n &= n + l(n) + O\left(\frac{n\log\log n}{\mathrm{lon}\, n}\right) \\
&= \frac{n}{d(n)} + O\left(\frac{n\log\log n}{\log n}\right).
\end{aligned} \tag{7}
$$

Hence

$$\frac{1}{\sigma_k a_k} = \frac{d(k)}{k} + O\left(\frac{\log\log k}{k\log k}\right).$$

Now summing this from 2 to n and using the fact that

$$\sum_{k=2}^{n} 1/a_k\sigma_k = \sum_{k=2}^{n} 1/\sigma_k - 1/\sigma_{k-1} = 1/\sigma_n - \tfrac{1}{2}$$

we obtain

$$(8)\qquad \begin{aligned}\frac{1}{\sigma_n} &= \sum_{k=2}^{n}\left(\frac{d(k)}{k} + O\left(\frac{\log\log k}{k\log k}\right)\right) + \frac{1}{2}\\ &= \sum_{k=2}^{n}\frac{d(k)}{k} + O[(\log\log n)^2].\end{aligned}$$

The proof of the lemma is completed by multiplying (7) by (8) and using the fact that $\frac{1}{2} \le d(n) \le 1$.

LEMMA 3. *Suppose* $d(k) = (1 + \delta(k))\cdot d$ *where* $\frac{1}{2} \le d \le 1$ *and* $\delta(k) = o(1)$. *Then*

$$a_n = n\log n + [1 + o(1)]n\textstyle\sum_{k=2}^{n}\delta(k)/k - [1 + o(1)]n\delta(n)\log n + O(n(\log\log n)^2).$$

Proof. First observe that

$$n/d(n) = (n/d)(1 - (1 + o(1))\delta(n)).$$

Hence from Lemma 2,

$$\begin{aligned}a_n &= (n/d)(1 - (1 + o(1))\delta(n))\\ &\quad\cdot\left(\textstyle\sum_{k=2}^{n} d/k + \sum_{k=2}^{n} d\delta(k)/k + O(n(\log\log n)^2)\right)\\ &= n\log n + (1 + o(1))n\textstyle\sum_{k=2}^{n}\delta(k)/k - (1 + o(1))n\log n\,\delta(n)\\ &\quad + O(n(\log\log n)^2).\end{aligned}$$

We are now going to apply this lemma to a number of specific sequences. To do this, we will suppose that $r_n = k - n$ where $a_k^{(n)}$ is the smallest element eliminated from $A^{(n)}$ to form $A^{(n+1)}$. We will further stipulate that r_k/k is asymptotic to a constant, and r_k is non-decreasing. We first of all need a lemma connecting r_k with $\delta(k)$.

LEMMA 4. *If* $r_n = n\cdot r(1 + \rho(n))$ *where* r *is a positive constant and* $\rho(n) = o(1)$, *then*

$$\delta(n) = \frac{r}{(r+1)(r+2)}(1 + o(1))\rho(k) + O[|\rho(k+1) - \rho(k)| + \rho^2(k+1) + \rho^2(k) + 1/\log n],$$

where k *is defined by*

$$(9)\qquad r_k + k \le n < r_{k+1} + k + 1.$$

Proof. Using (9) and the fact that $r_k = k\cdot r(1 + \rho(k))$ we obtain

$$k((r+1)/n) = 1 - (r/(r+1))\rho(k) + O[|\rho(k+1) - \rho(k)| + \rho^2(k+1) + \rho^2(k) + 1/n].$$

Since $t(n) < c_2 n/\log n$, we have

$$k = l(n) + O(n/\log n)$$

and so

$$l(n) = (n/(r+1))[1 - (r/(r+1))\rho(k) + O(|\rho(k+1) - \rho(k)| + \rho^2(k+1) + \rho^2(k)] + O(n/\log n).$$

Now using the fact that $d(n) = n/(n + l(n))$ a routine calculation completes the proof of the lemma.

THEOREM 1. *If* $r_k = k \cdot r(1 + \rho(k))$ *where* $\rho(k) = O\left(\frac{\log \log k}{\log k}\right)$, *then*

$$a_n - n \log n = O(n(\log \log n)^2).$$

Proof. Since the k in Lemma 4 is asymptotic to $n/(r + 1)$, Lemma 4 yields

$$\delta(k) = O\left(\frac{\log \log k}{\log k}\right).$$

Hence

$$\sum_{k=2}^{n} \frac{\delta(k)}{k} = O\left(\sum_{k=1}^{n} \frac{\log \log k}{k \log k}\right) = O(\log \log k)^2.$$

The theorem then follows from Lemma 3.

THEOREM 2. *If* $r_k = kr(1 + \rho(k))$ *where* r *is a positive constant and* $\rho(k) = o(1)$ *and furthermore has the property that* $\rho(k \cdot \alpha \cdot (1 + o(1)) \sim \rho(k)$ *where* α *is a constant, then*

$$a_n = n \log n + [1 + o(1)] \frac{nr}{(r+1)(r+2)} \sum_{k=2}^{n} \frac{\rho(k)}{k} - (1 + o(1)) \frac{rn \log n}{(r+1)(r+r)} \rho(n) + O(n(\log \log n)^2)$$

Proof. For the k in Lemma 4,

$$\rho(k) \sim \rho(n) \quad \text{and} \quad |\rho(k+1) - \rho(k)| = o(\rho(k)).$$

Also, $\rho^2(k) + \rho^2(k + 1) = o(\rho(k))$. Hence

$$\delta(n) = \frac{r}{(r+1)(r+2)} (1 + o(1))\rho(k) + O\left(\frac{1}{\log n}\right).$$

Lemma 4 completes the proof of the theorem.

Theorem 2 can now be used to construct sieve-generated sequences $\{a_n\}$ for which $a_n - n \log n$ is asymptotic to any given function lying between $n \log n$ and $O(n(\log \log n)^2)$. To demonstrate this, we will produce one for which

$$a_n - n \log n \sim cn(\log n)^{1-\varepsilon}$$

for any given $1 > \varepsilon > 0$. To do this, we let

$$r_k = k\{1 + (1 - \varepsilon)/(\log k)^{\varepsilon}\}$$

and apply Theorem 2:

This yields

$$a_n = n \log n + [1 + o(1)] \frac{n}{6} \sum_{k=2}^{n} \frac{(1-\varepsilon)}{k(\log k)^{\varepsilon}}$$
$$- [1 + o(1)] \frac{n \log n(1-\varepsilon)}{6(\log n)^{\varepsilon}} + O(n(\log\log n)^2)$$
$$= n \log n + [1 + o(1)]n(\log n)^{1-\varepsilon}/6$$
$$- [1 + o(1)](1-\varepsilon)n(\log n)^{1-\varepsilon}/6 + O(n(\log\log n)^2)$$
$$= n \log n + [1 + o(1)]\varepsilon n(\log n)^{1-\varepsilon}/6.$$

An interesting sequence can be produced letting

$$r_k = k(1 - (1-\varepsilon)/(\log k)^{\varepsilon}).$$

In this case, one can apply Theorem 2 and get

$$a_n = n \log n - [1 + o(1)]\varepsilon n(\log n)^{1-\varepsilon}/6$$

In view of the fact that $p_n = n \log n + (1 + o(1))n \log\log n$ where p_n is the n-th prime number, this yields a sieve-generated sequence for which $a_n < p_n$ for n sufficiently large. This sequence for $\varepsilon = \frac{1}{2}$ was computed on the I.B.M. 709 computer at the University of Colorado and for $n = 73{,}594$, $a_n > p_n$. ($a_n = 1{,}239{,}993$ and $p_n = 931{,}783$).

If we consider a smaller class of sieve-generated sequences, we can obtain sharper estimates for $a_n - n \log n$. In this sieve, we eliminate those elements in $A^{(n)}$ of the form $a^{(n)}_{n+r_n+ma_n}$ for $m = 0, 1, 2, \cdots$ to form $A^{(n+1)}$. The method described by Briggs in [1] will yield

THEOREM 3. (a) *If the sequence* $\{r_k\}$ *is non-decreasing and* $r_k = o(k/\log k)$, *then*

$$a_n = n \log n + (n/2)(\log\log n)^2 - (\gamma + \log 2)n \log\log n + o(n \log\log n).$$

(b) *If* $c > 0$, $r_1 = 1$, *and* $r_n = [cn/\log n] + 1$, $n > 1$, *then*

$$a_n = n \log n + (n/2)(\log\log n)^2$$
$$- (\gamma + \log 2 - c/2)n \log\log n + o(n \log\log n).$$

(c) *If* $c > 0$ *and* $r_n = [cn] + 1$, *then*

$$a_n = n \log n + (n/2)(\log\log n)^2$$
$$- (\gamma + \log (c+2)/c + 1)n \log\log n + o(n \log\log n).$$

(d) *If* $0 < c \leq 1$ *and* $r_k = [ca_k] + 1$, *then*

$$a_n = n \log n + (n/2)(\log\log n)^2 - \psi(c)n \log\log n + o(n \log\log n)$$

where $\psi(z) = -\Gamma'(z)/\Gamma(z)$.

One notices that all of the estimates of $a_n - n \log n$ contain a term of the form $O(n(\log \log n)^2)$. For sequences generated by this sieve process, this term cannot be eliminated for the second and third summation in (2) both contain a term of the form $\sum R_k(a_n + 1)/a_k$ and hence

$$E_n(a_n + 1) = \sum_{k > s(n)} \frac{R_k(a_n + 1)}{a_k} - l(n) + O\left(\frac{n \log \log n}{\log n}\right)$$
$$\geq \sum_{k > s(n)} \frac{n}{ck \log k} > cn\left(\frac{\log \log n}{\log n}\right).$$

Applying this to the proof of Lemma 2, one sees that the term $O(n(\log \log n)^2)$ cannot be replaced by $o(n(\log \log n)^2)$. Thus, the asymptotic expression

$$p_n = n \log n + (1 + O(1))n \log \log n$$

for the n-th prime cannot be duplicated for sequences of this nature. There is some evidence to support the conjecture that the following modification of the sieve process could eliminate this objectionable term: Let $g(n)$ be a number theoretic function such that $g(n) > n^{1+\varepsilon}$. When obtaining $A^{(n+1)}$ from $A^{(n)}$ one does not sieve out any element of $A^{(n)}$ which is less than $a_{n+g(n)}^{(n)}$. Elements greater than $a_{n+g(n)}^{(n)}$ are sieved out in the usual manner. This sieve method is quite similar in some respects to the sieve of Eratosthenes when $g(n) \sim \frac{1}{2}n^2 \log n$.

The authors would like to thank the referee for clearing up certain difficulties in Lemma 4.

References

1. W. E. Briggs, *Prime-like sequences generated by a sieve process*, Duke Math. J., vol. 30 (1963), pp. 297–312.
2. M. Wunderlich, *Sieve-generated sequences*, Canad. J. Math., vol. 18 (1966), pp. 291–299.

State University of New York
Buffalo, New York
University of Colorado
Boulder, Colorado