ON A THEOREM CONCERNING THE DISTRIBUTION OF ALMOST PRIMES

By

Saburô UCHIYAMA

By an *almost prime* is meant a positive rational integer the number of prime factors of which is bounded by a certain constant. Let us denote by $\Omega(n)$ the total number of prime factors of a positive integer n. In 1920 Viggo Brun [2] elaborated an elementary method of the sieve of Eratosthenes to prove that for all sufficiently large x there exists at least one integer n with $\Omega(n) \leqq 11$ in the interval $x \leqq n \leqq x + x^{\frac{1}{2}}$. Quite recently W. E. Mientka [4] improved this result of Brun, showing that for all large x there exists at least one integer n with $\Omega(n) \leqq 9$ in the interval $x \leqq n \leqq x + x^{\frac{1}{2}}$. To establish this Mientka makes use of the sieve method due to A. Selberg instead of Brun's method (cf. [3] and [4]). By refining the argument of Mientka [4] we can further improve his result. Indeed, we shall prove in this paper the following

Theorem. *Let $k \geqq 2$ be a fixed integer. Then, for all sufficiently large x, there exists at least one integer n with $\Omega(n) \leqq 2k$ in the interval $x < n \leqq x + x^{1/k}$.*

Thus, in particular, if $k=2$ then for all large x the interval $x < n \leqq x + x^{\frac{1}{2}}$ always contains an integer n such that $\Omega(n) \leqq 4$. Of course, the restriction in the theorem that k be integral may be relaxed without essential changes in the result.

Let us mention that the existence of a prime number p in the interval $x < p \leqq x + x^{1/k}$ for all large x could not be deduced, as is well known, even from the Riemann hypothesis if only $k=2$.

Note. It is possible to generalize our theorem presented above so as to concern with the distribution of almost primes in an arithmetic progression. Thus, if a and b are integers such that $a \geqq 1$, $0 \leqq b \leqq a-1$, $(a, b)=1$, then we can prove the existence of an integer n satisfying

$$x < n \leqq x + x^{1/k}, \quad n \equiv b \pmod{a},$$
$$\Omega(n) \leqq 2k,$$

provided that x be sufficiently large, $k \geqq 2$ being a fixed integer. Here, in particular, in the case of $k=2$, the inequality $\Omega(n) \leqq 4$ may be replaced by $\Omega(n) \leqq 3$: this result is apparently stronger than the above theorem for the corresponding case. Proof is similar to that of our theorem but somewhat more complicated arguments are needed.

1. Let $M>0$ and $N>1$ be integers and let $z\geqq 2$ and $w>0$ be any real numbers such that $w^2\geqq z$. We denote by S the number of those integers n in the interval $M<n\leqq M+N$ which are not divisible by any prime number $p\leqq z$. Then, by making use of the 'lower' sieve of A. Selberg (cf. [3] and [6]) we can show that

$$S\geqq(1-Q)N-R_1,$$

where

$$Q=\sum_{p\leqq z}\frac{1}{pZ_p}\text{ with }Z_p=\sum_{\substack{1\leqq m\leqq w/\sqrt{p}\\ g(m)<p}}\frac{\mu^2(m)}{\phi(m)}$$

and

$$R_1=O\left(w^2\sum_{p\leqq z}\frac{1}{pZ_p^2}\right).$$

Here $g(1)=1$ and for $m>1$ $g(m)$ denotes the greatest prime divisor of m, and the O-constant for R_1 is absolute. It will be shown later that $Z_p>c\log p$ for all $p\leqq z$, where $c>0$ is a constant, so that we have $R_1=O(w^2)$.

Now we take

$$z=(2N)^{\frac{1}{4}},\quad w=(2N)^{\frac{1}{2}-\varepsilon},$$

where $0<\varepsilon<\frac{1}{4}$. If we fix ε sufficiently small then there holds the following

Lemma 1. *For all sufficiently large N we have*

$$S>1.6054\frac{N}{\log N}.$$

Our proof of Lemma 1 runs essentially on the same lines as in [4]; we shall give an outline of the proof of this lemma in § 3.

Throughout in the following the constants implied in the symbol O are all absolute (apart from the possible dependence on the parameter ε), and c represents positive constants not necessarily the same in each occurrence.

2. In order to prove Lemma 1 we require some auxiliary results due to N. G. de Bruijn [1] on the number $\Psi(x,y)$ of integers $n\leqq x$ and free of prime factors $>y$.

It is proved by de Bruijn [1] that we have

$$(1)\qquad \Psi(x,y)=O(xe^{-cu})$$

and more precisely

$$(2)\qquad \Psi(x, y) = x\rho(u) + O(1) + O(xu^2 e^{-c\sqrt{\log y}}) + O\left(\frac{x\rho(u)\log(2+u)}{\log y}\right),$$

where $x>1$, $y\geqq 2$, $u=(\log x)/\log y$, and the function $\rho(u)$ is defined by the following conditions:

$$(3)\qquad \begin{aligned} &\rho(u)=0\ (u<0);\ \rho(u)=1 \qquad (0\leqq u\leqq 1);\\ &u\rho'(u) = -\rho(u-1)\ (u>1);\ \rho(u)\ \text{continuous for}\ u>0. \end{aligned}$$

Lemma 2. *We have for* $t\geqq t_0\geqq 1$

$$\rho(t)\leqq \rho(t_0)e^{-(t-t_0)},$$

so that

$$\int_{t_0}^{\infty}\rho(u)du \leqq \rho(t_0) \qquad (t_0\geqq 1).$$

This is stated and employed without proof in [4] as a lemma of N. C. Ankeny. By integrating by parts we deduce from (3) that for $t\geqq 1$

$$t\rho(t) = \int_0^t \rho(u)du - \int_0^t \rho(u-1)du = \int_{t-1}^{t}\rho(u)du \leqq \rho(t-1),$$

since $\rho(u)$ decreases monotonously for $u\geqq 0$. Honce

$$\frac{\rho'(u)}{\rho(u)} = -\frac{\rho(u-1)}{u\rho(u)} \leqq -1 \qquad (u\geqq 1)$$

and the result follows at once.

3. Following Mientka [4] let us put

$$H_p - \prod_{q<p}\left(1-\frac{1}{q}\right)^{-1} \qquad (p\leqq z),$$

where in the product on the right-hand side q runs through the prime numbers less than p, and

$$T_p = \sum_{\substack{m>w/\sqrt{p}\\ g(m)\leqq p}} \frac{1}{m} \qquad (p\leqq z).$$

Then we have $|H_p - Z_p|\leqq T_p$ and

$$S\geqq N\prod_{p\leqq z}\left(1-\frac{1}{p}\right) - N\sum_{p\leqq z}\frac{T_p}{pH_p(H_p-T_p)} - R_1$$

(cf. [3] and [4]). Since it is well known that

$$\prod_{p\leqq z}\left(1-\frac{1}{p}\right) = \frac{e^{-C}}{\log z} + O\left(\frac{1}{\log^2 z}\right),$$

C being the Euler constant, it remains only to evaluate the middle term on the right-hand side of the above inequality for S.

By partial summation we have

$$T_p = \sum_{m > w/\sqrt{p}} \frac{\Psi(m,p)}{m^2} + O(N^{-\frac{3}{8}+\epsilon}).$$

We find easily that

$$T_p = O\left(\frac{1}{\log^2 N}\right)$$

for every $p \leqq \exp(\log N)^{\frac{2}{3}}$, on taking account of (1). For $\exp(\log N)^{\frac{2}{3}} < p \leqq z$ we have

$$T_p = \sum_{w/\sqrt{p} < m \leqq \exp(\log N)^2} \frac{\Psi(m,p)}{m^2} + O\left(\frac{1}{\log^2 N}\right),$$

where, by (2),

$$\sum_{w/\sqrt{p} < m \leqq \exp(\log N)^2} \frac{\Psi(m,p)}{m^2}$$

$$= \sum_{w/\sqrt{p} < m \leqq \exp(\log N)^2} \frac{1}{m} \rho\left(\frac{\log m}{\log p}\right)\left(1 + O\left(\frac{\log\log N}{\log p}\right)\right)$$

$$+ O\left(\frac{1}{\log^2 N}\right),$$

and this is equal to

$$\left(\int_{w/\sqrt{p}}^{\infty} \frac{1}{x}\rho\left(\frac{\log x}{\log p}\right) dx + O(N^{-\frac{3}{8}+\epsilon})\right)\left(1 + O\left(\frac{\log\log N}{\log p}\right)\right) + O\left(\frac{1}{\log^2 N}\right).$$

Hence

$$T_p = \log p \int_{\log(w/\sqrt{p})/\log p}^{\infty} \rho(u)\, du \left(1 + O\left(\frac{\log\log N}{\log p}\right)\right) + O\left(\frac{1}{\log^2 N}\right)$$

for $\exp(\log N)^{\frac{2}{3}} < p \leqq z$.

Put

$$I_p = \int_{\log(w/\sqrt{p})/\log p}^{\infty} \rho(u)\, du \qquad (p \leqq z).$$

Then it follows immediately from the above results that

$$\sum_{p \leqq z} \frac{T_p}{pH_p(H_p - T_p)}$$

$$= e^{-C} \sum_{\exp(\log N)^{\frac{2}{3}} < p \leqq z} \frac{1}{p \log p} \frac{I_p}{e^C - I_p} + O\left(\frac{(\log\log N)^2}{(\log N)^{4/3}}\right).$$

For p in the interval $(2N)^{\frac{1}{\nu+1}}<p\leqq(2N)^{\frac{1}{\nu}}(\nu\geqq 4)$ we have, by Lemma 2,

$$I_p\leqq\rho\left(\frac{\log(w/\sqrt{p})}{\log p}\right)\leqq\rho(t_\nu),$$

where we have put

$$t_\nu=\left(\frac{1}{2}-\varepsilon\right)\nu-\frac{1}{2}.$$

Therefore

$$\sum_{\exp(\log N)^{\frac{2}{3}}<p\leqq z}\frac{1}{p\log p}\frac{I_p}{e^C-I_p}$$

$$\leqq\sum_{4\leqq\nu<c(\log N)^{\frac{1}{3}}}\left(\sum_{(2N)^{\frac{1}{\nu+1}}<p\leqq(2N)^{\frac{1}{\nu}}}\frac{1}{p\log p}\right)\frac{\rho(t_\nu)}{e^C-\rho(t_\nu)}$$

$$=\frac{1}{\log N}\sum_{\nu=4}^{\infty}(\nu+1)\log\frac{\nu+1}{\nu}\frac{\rho(t_\nu)}{e^C-\rho(t_\nu)}+O\left(\frac{1}{(\log N)^{4/3}}\right).$$

Here we used the relation

$$\sum_{p\leqq x}\frac{1}{p}=\log\log x+c_1+O\left(\frac{1}{\log x}\right),$$

c_1 being a constant. Hence

$$\sum_{p\leqq z}\frac{T_p}{pH_p(H_p-T_p)}$$

$$\leqq\frac{e^{-C}}{\log N}\sum_{\nu=4}^{\infty}(\nu+1)\log\frac{\nu+1}{\nu}\frac{\rho(t_\nu)}{e^C-\rho(t_\nu)}+O\left(\frac{(\log\log N)^2}{(\log N)^{4/3}}\right).$$

Now, by the definition of $\rho(u)$, we have

$$\rho(u)=1-\log u\qquad(1\leqq u\leqq 2).$$

If we take $\varepsilon=10^{-4}$, then we find that

$$\rho(t_4)=\rho(1.4996)<0.5949,$$
$$\rho(t_5)=\rho(1.9995)<0.3072,$$

so that

$$5\log\frac{5}{4}\frac{\rho(t_4)}{e^C-\rho(t_4)}<0.5597,$$

and

$$\sum_{\nu=5}^{\infty}(\nu+1)\log\frac{\nu+1}{\nu}\frac{\rho(t_\nu)}{e^C-\rho(t_\nu)}$$
$$\leqq 6\log\frac{6}{5}\frac{\rho(t_5)}{e^C-\rho(t_5)}\frac{1}{1-e^{-0.4999}}<0.5807\,,$$

on appealing to Lemma 2. We thus have proved that

$$\sum_{p\leqq z}\frac{T_p}{pH_p(H_p-T_p)}<1.1404e^{-C}\frac{1}{\log N}+O\left(\frac{(\log\log N)^2}{(\log N)^{4/3}}\right)$$

and this completes the proof of Lemma 1 since

$$(4-1.1404)e^{-C}>1.6055\,.$$

4. Let q be any prime number in the interval $z<q\leqq z^2$, where, as before, $z=(2N)^{\frac{1}{4}}$. We next estimate the number $S(q)$ of those integers n in $M<n\leqq M+N$ which are multiples of q and are not divisible by any prime number $p\leqq z$. We have by the 'upper' sieve of A. Selberg (cf. [5])

$$S(q)\leqq\frac{N}{qZ}+R_2\,,$$

where

$$Z=\sum_{1\leqq m\leqq z}\frac{\mu^2(m)}{\phi(m)}$$

and

$$R_2=O\left(\frac{z^2}{Z^2}\right)\,.$$

It is easily verified that

$$Z\geqq\sum_{1\leqq m\leqq z}\frac{1}{m}=\log z+O(1)\,,$$

and therefore

$$S(q)\leqq\frac{4N}{q\log N}+O\left(\frac{N}{q\log^2 N}\right)\,. \tag{4}$$

Lemma 3. *Let U denote the number of those integers n in $M<n\leqq M+N$ which are divisible by no primes $p\leqq z$, by at most two primes q with $z<q\leqq z^2$, and by no integers of the form q^2, q being a prime in $z<q\leqq z^2$. Then, for all sufficiently large N, we have*

$$U>0.6811\frac{N}{\log N}\,.$$

Let N be a sufficiently large positive number. The number of those integers n in $M<n\leqq M+N$ which are not divisible by any prime $p\leqq z$ and are divisible by some q^2, where q is a prime in $z<q\leqq z^2$, does not exceed

$$\sum_{z<q\leqq z^2}\left(\left[\frac{M+N}{q^2}\right]-\left[\frac{M}{q^2}\right]\right)-O(N^{\frac{3}{4}})\,.\,.$$

Now, the number of those integers n with $M<n\leqq M+N$ which are not divisible by any prime $p\leqq z$ and are divisible by at least three (distinct) primes q in $z<q\leqq z^2$ is, by (4), not greater than

$$\frac{1}{3}\sum_{z<q\leqq z^2}S(q)\leqq\frac{4\log 2}{3}\frac{N}{\log N}+O\left(\frac{N}{\log^2 N}\right).$$

It thus follows from Lemma 1 that

$$U>\left(1.6054-\frac{4\log 2}{3}\right)\frac{N}{\log N}+O\left(\frac{N}{\log^2 N}\right),$$

which proves our lemma since $(4/3)\log 2<0.9242$.

5. We can now conclude the proof of our theorem. Let x be a sufficiently large positive real number and put

$$M=[x]\,,\qquad N=[x^{1/k}]\,.$$

Then, by Lemma 3, there exists at least one integer n in the interval $M<n\leqq M+N$, i. e. in the interval

$$x<n\leqq x+x^{1/k}\,,$$

such that it is not divisible by any prime $p\leqq(2N)^{\frac{1}{4}}$ and is divisible by at most two primes q in $(2N)^{\frac{1}{4}}<q\leqq(2N)^{\frac{1}{2}}$ but not divisible by the squares of these q, where

$$(2N)^k>(2(x^{1/k}-1))^k>x+x^{1/k}$$

since $k\geqq 2$. Therefore, according as n has no, one or two prime factors q in $(2N)^{\frac{1}{4}}<q\leqq(2N)^{\frac{1}{2}}$ it has at most $2k-1$, $2k-1$ or $2k-2$ additional prime factors. Hence the total number of prime factors of n is at most $2k$, i. e. $\Omega(n)\leqq 2k$. This completes the proof of the theorem.

References

[1] N. G. de BRUIJN: On the number of positive integers $\leqq x$ and free of prime factors $>y$, Indagationes Math., vol. 13 (1951), pp. 50-60.

[2] VIGGO BRUN: Le crible d'Eratosthène et le théorème de Goldbach, Norske Videnskapselskapets Skrifter, I. Kristiania, 1920, No. 3.
[3] W. E. MIENTKA: Notes on the lower bound of the Selberg sieve method, to appear.
[4] W. E. MIENTKA: An application of the Selberg sieve method, Journ. Indian Math. Soc. (N. S.), vol. 25 (1961), pp. 129–138.
[5] S. UCHIYAMA: A note on the sieve method of A. Selberg, J. Fac. Sci., Hokkaidô Univ., Ser. I, vol. 16 (1962), pp. 189–192.
[6] S. UCHIYAMA: A further note on the sieve method of A. Selberg, J. Fac. Sci., Hokkaidô Univ., Ser. I, vol. 17 (1963), pp. 79–83.

Department of Mathematics,
Hokkaidô University

(Received July 19, 1963)