AN ASYMPTOTIC FORMULA FOR PRIMES OF THE FORM 4n + 1

Robert Breusch

Introduction. The prime number theorem for arithmetic progressions asserts that if $(a, b) = 1$ and $\pi(x; a, b)$ represents the number of primes of the form $an + b$ $(an + b \leq x)$, then

$$\pi(x; a, b) = \frac{1}{\phi(a)} \int_2^x \frac{dt}{\log t} (1 + o(1)),$$

where ϕ is Euler's function. With $o(1) = O((\log x)^{-1-\delta})$ $(\delta > 0)$, it follows immediately that

$$\text{(A)} \qquad \sum_{\substack{p \leq x \\ p \equiv b \pmod a}} \frac{\log p}{p} = \frac{\log x}{\phi(a)} + O(1),$$

just as

$$\text{(B)} \qquad \sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$$

follows from the corresponding form of the basic prime number theorem. However, it is well known that (B) can also be derived directly, in very elementary ways, from the relation $\pi(x) = O(x/\log x)$. In this paper, (A) will be derived by elementary reasoning, but only for the very special case $a = 4$. A similar proof can be given for $a = 3$, but the method employed here does not seem to be applicable if $\phi(a) > 2$.

THEOREM. *Let* p *stand for primes of the form* $4n + 1$. *Then*

$$\sum_{p \leq x} \frac{\log p}{p} = \frac{1}{2} \cdot \log x + O(1).$$

Proof. For a given positive integer n, let

$$P_n = \prod_{r=1}^{n} \prod_{s=1}^{n} (r^2 + s^2).$$

Certainly, $P_n < \prod\prod(2n^2) = 2^{n^2} n^{2n^2}$, and

$$P_n > \prod\prod(s^2) = (n!)^{2n} > (n^n \cdot e^{-n})^{2n} = n^{2n^2} \cdot e^{-2n^2}.$$

Thus

Received April 13, 1964.

$$\text{(I)} \qquad \log P_n = n^2 \cdot \log n^2 + O(n^2).$$

Let $P_n = 2^{\alpha_2} \cdot \prod_{q<2n^2} q^{\alpha_q} \cdot \prod_{p<2n^2} p^{\alpha_p}$, where q represents primes of the form $4n+3$; then $\log P_n = \alpha_2 \log 2 + \sum_{q<2n^2} \alpha_q \log q + \sum_{p<2n^2} \alpha_p \log p$. For convenience, we split the last sum into three parts and write

$$\text{(II)} \qquad \begin{aligned} \log P_n &= \alpha_2 \log 2 + \sum_{q<2n^2} \alpha_q \log q \\ &\quad + \left(\sum_{p<4n} + \sum_{4n<p<n^2} + \sum_{n^2<p<2n^2} \right) \alpha_p \log p. \end{aligned}$$

We now consider separately each of the five parts in II.

1) $2^{2t} \mid (r^2+s^2)$ if and only if $2^t \mid r$ *and* $2^t \mid s$. This occurs in $[n/2^t]^2$ of the factors (r^2+s^2). An additional factor 2 is contained in (r^2+s^2) if r and s contain precisely the same number of factors 2. The total contribution of this effect is clearly fewer than n^2 factors 2. Thus $\alpha_2 < 2\Sigma_t [n/2^t]^2 + n^2 = O(n^2)$, and therefore

$$\alpha_2 \log 2 = O(n^2).$$

2) $q^{2t} \mid (r^2+s^2)$ if and only if $q^t \mid r$, $q^t \mid s$. Therefore

$$\alpha_q = 2\sum_t [n/q^t]^2 < 2n^2/(q^2-1),$$

and thus

$$\sum \alpha_q \log q = O(n^2).$$

3) For a given p, and for every r, there exist precisely two numbers s_i $(i = 1, 2)$ such that $0 < s_i \le p$ and $p \mid (r^2+s^2)$. $[s_1 = s_2 = p$ if $p \mid r]$. Precisely the numbers $s = s_i + mp$, and no others, are divisible by p; the number of such $s \le n$ is $2[n/p] + O(1)$. An additional factor p is contained in those (r^2+s^2) that are divisible by p^2; their number is $2[n/p^2] + O(1)$. Continuing with higher powers of p, we see that p^t can be a divisor of (r^2+s^2) only as long as $p^t \le 2n^2$; thus $t = O((\log n)/(\log p))$. It follows that the number of factors p in $\prod_{s=1}^n (r^2+s^2)$ is

$$2\sum_t [n/p^t] + O\left(\frac{\log n}{\log p}\right) = \frac{2n}{p} + O\left(\frac{n}{p^2}\right) + O\left(\frac{\log n}{\log p}\right).$$

Since this holds for every r $(1 \le r \le n)$,

$$\alpha_p = \frac{2n^2}{p} + O\left(\frac{n^2}{p^2}\right) + O\left(\frac{n \cdot \log n}{\log p}\right),$$

and

$$\sum_{p<4n} \alpha_p \log p = 2n^2 \cdot \sum_{p<4n} \frac{\log p}{p} + n^2 \cdot O\left(\sum_{p<4n} \frac{\log p}{p^2}\right) + n(\log n)\, O\left(\sum_{p<4n} \frac{\log p}{\log p}\right).$$

The content of the first O-term is O(1), and that of the second O-term is $\pi(4n) = O(n/\log n)$. Thus

$$\sum_{p<4n} \alpha_p \log p = 2n^2 \sum_{p<4n} \frac{\log p}{p} + O(n^2).$$

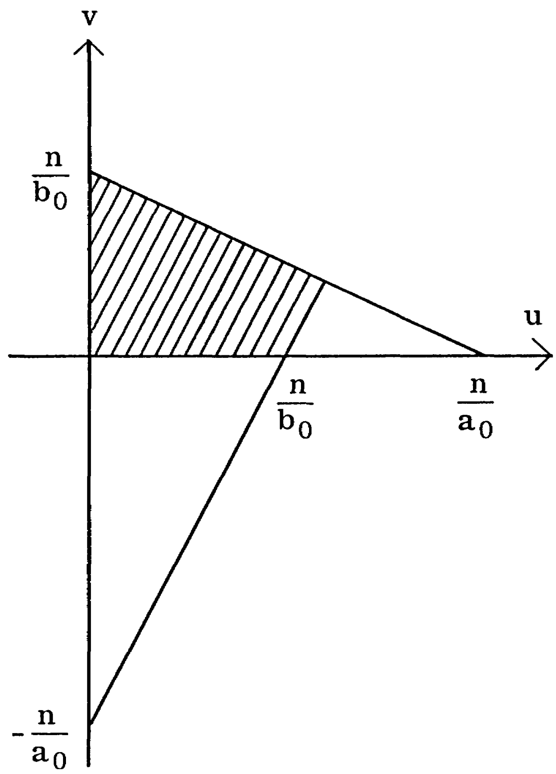
4) Every $p < n^2$ may be written uniquely as $p = a_0^2 + b_0^2$, with $0 < a_0 < b_0 < n$ and $(a_0, b_0) = 1$. If $p > 4n$, then no single factor $(r^2 + s^2)$ of P can contain p^2. And if such a factor is divisible by p, then certainly $r \neq s$, because $r = s$ would imply $p \mid 2s^2$, thus $p \mid s$, contrary to the relation $s \leq n < p$. Thus, if $p \mid (r^2 + s^2)$, and if we call a the smaller and b the larger of r and s, then $0 < a < b \leq n$. We are interested in the number of such pairs (a, b).

LEMMA. *Let* C_1 *be the class of pairs of integers* (a, b) *with* $0 < a < b \leq n$ *and with the property that* $a^2 + b^2$ *is divisible by a fixed* $p = a_0^2 + b_0^2$ $[0 < a_0 < b_0$, $4n < p < n^2]$. *Let* C_2 *be the class of ordered pairs of integers* (u, v) *with*

$$u > 0, \qquad v \geq 0, \qquad a_0 u + b_0 v \leq n, \qquad |b_0 u - a_0 v| \leq n.$$

Then C_1 *and* C_2 *contain the same number of elements.*

A proof of the lemma will be given at the end of the paper.



The set C_2 is represented by the lattice points in the (u, v)-plane that lie within the quadrilateral formed by the straight lines $a_0 u + b_0 v = n$, $b_0 u - a_0 v = n$, $u = 0$, $v = 0$, not counting the lattice points on the left-hand edge. [For all the points within this quadrilateral, the last condition, $a_0 v - b_0 u \leq n$, is automatically satisfied.] The area of the quadrilateral is $A = n^2/(a_0^2 + b_0^2) = n^2/p$; its perimeter L is less than $4n/b_0$, and since $b_0^2 > p/2$, $L < 8n/\sqrt{p}$. Thus the number of lattice points within is

$$A + O(L) = n^2/p + O(n/\sqrt{p}).$$

By the lemma, this is also the number of pairs (a, b) in C_1. Since P_n contains every $(a^2 + b^2)$ with $(a, b) \in C_1$ twice,

$$\alpha_p = 2n^2/p + O(n/\sqrt{p}),$$

and thus

$$\sum_{4n<p<n^2} \alpha_p \log p = 2n^2 \cdot \sum_{4n<p<n^2} \frac{\log p}{p} + O\left(n \cdot \sum_{p<n^2} \frac{\log p}{\sqrt{p}}\right).$$

Summation by parts shows that the sum in the O-term is O(n). Thus

$$\sum_{4n<p<n^2} \alpha_p \log p = 2n^2 \cdot \sum_{4n<p<n^2} \frac{\log p}{p} + O(n^2).$$

5) In the last term of (II), clearly $\alpha_p \leq 1$; thus

$$\sum_{n^2<p<2n^2} \alpha_p \log p = O(n^2).$$

Combining the five parts, we see by (II) that

$$\log P_n = 2n^2 \cdot \sum_{p<n^2} \frac{\log p}{p} + O(n^2).$$

Together with (I), this shows that

$$\sum_{p<n^2} \frac{\log p}{p} = \frac{1}{2} \log n^2 + O(1).$$

If x is any positive number and $n^2 \leq x < (n+1)^2$, then $\Sigma_{p\leq x} (\log p)/p$ lies between $\log n + O(1)$ and $\log (n+1) + O(1)$, both of which differ from $(1/2) \cdot \log x$ by $O(1)$. Thus

$$\sum_{p<x} \frac{\log p}{p} = \frac{1}{2} \log x + O(1).$$

Except for the proof of the lemma, the theorem is thus established.

COROLLARY. $\sum_{q\leq x} \frac{\log q}{q} = \frac{1}{2} \log x + O(1)$, *by formula* (B) *of the Introduction.*

Proof of the lemma. (i) For every $(u, v) \in C_2$, let a be the smaller and b the larger of the two numbers $a_0 u + b_0 v$ and $|b_0 u - a_0 v|$. Then

$$a^2 + b^2 = (a_0^2 + b_0^2)(u^2 + v^2);$$

thus $p|(a^2 + b^2)$, and $0 \leq a \leq b \leq n$; but $0 = a$ and $a = b$ are both impossible, since each would imply that $p | b$, contrary to the relation $b \leq n < p$. Thus to every member of C_2 has been assigned a unique member of C_1.

(ii) If the correspondence just established would assign the same (a, b) to two different members (u_1, v_1) and (u_2, v_2), of C_2, it would follow that either

$$a_0 u_1 + b_0 v_1 = a_0 u_2 + b_0 v_2 \qquad \text{or} \qquad a_0 u_1 + b_0 v_1 = \pm a_0 v_2 \mp b_0 u_2.$$

Since $(a_0, b_0) = 1$, it follows that *either* $b_0 | (u_1 - u_2)$, *or* $b_0 | (u_1 \pm v_2)$. But since $p > 4n$, $b > \sqrt{2n}$, while $u^2 + v^2 < n/2$, and therefore u_i and v_i are all less than $\sqrt{n/2}$. It follows that $b_0 > |u_1 - u_2|$, and $b_0 > |u_1 \pm v_2|$. Thus $b_0 | (u_1 - u_2)$ is possible only if $u_1 = u_2$, in which case also $v_1 = v_2$; and $b_0 | (u_1 \pm v_2)$ is possible only if $u_1 - v_2 = 0$, so that $b_0(v_1 + u_2) = 0$ and hence $v_1 = u_2 = 0$, contrary to the relation $u_2 > 0$.

(iii) If $(a, b) \in C_1$, then the number

$$(b_0 b + a_0 a)(b_0 b - a_0 a) = b^2(a_0^2 + b_0^2) - a_0^2(a^2 + b^2)$$

is divisible by p; thus of the two numbers $(b_0 b + a_0 a)/p$ and $(b_0 b - a_0 a)/p$, one is a positive integer (positive because $b > a$, $b_0 > a_0$). Since

$$\text{(C)} \qquad \begin{cases} b_0(b_0 b \pm a_0 a)/p + a_0(a_0 b \mp b_0 a)/p = b, \\ a_0(b_0 b \pm a_0 a)/p - b_0(a_0 b \mp b_0 a)/p = \pm a, \end{cases}$$

it follows that $(a_0 b \mp b_0 a)/p$ is an integer if $(b_0 b \pm a_0 a)/p$ is an integer. Now we make use of (C), as follows:

(a) If the upper signs hold, and $a_0 b - b_0 a > 0$, set

$$u = (a_0 b - b_0 a)/p, \qquad v = (b_0 b + a_0 a)/p.$$

Then $a_0 u + b_0 v = b$, $|b_0 u - a_0 v| = a$.

(b) If the upper signs hold, and $a_0 b - b_0 a \leq 0$, set

$$u = (b_0 b + a_0 a)/p, \qquad v = (b_0 a - a_0 b)/p.$$

Then $a_0 u + b_0 v = a$, $b_0 u - a_0 v = b$.

(c) If the lower signs hold, set

$$u = (a_0 b + b_0 a)/p, \qquad v = (b_0 b - a_0 a)/p.$$

Then $a_0 u + b_0 v = b$, $b_0 u - a_0 v = a$.

Thus we see that, in each case, to a given member of C_1 corresponds a member of C_2; and by (ii), this member of C_2 is unique. This completes the proof of the lemma.

Amherst College