ACTA ARITHMETICA
LXXVII.2 (1996)

Irregularities in the distribution of primes in an arithmetic progression

by

Ti Zuo Xuan (Beijing)

1. Introduction. For $x \ge 2$ real, and q and a coprime positive integers, set

$$\theta(x;q,a) = \sum_{\substack{p \le x \\ p \equiv a \,(\mathrm{mod}\, q)}} \log p = \frac{x}{\varphi(q)}(1 + \Delta(x;q,a)),$$

where φ is Euler's function.

The prime number theorem for arithmetic progressions is equivalent to the statement that $\Delta(x;q,a) = o(1)$ as $x \to \infty$, for fixed q and a. The Siegel–Walfisz theorem gave a uniform upper estimate for the function Δ, and the Bombieri–Vinogradov theorem gave a mean value estimate for Δ.

Montgomery conjectured that if $(a,q) = 1$ then

$$|\Delta(x;q,a)| \ll_\varepsilon (q/x)^{1/2-\varepsilon} \log x \tag{1}$$

uniformly for $q \le x$, for any given $\varepsilon > 0$.

Recently, Friedlander and Granville [1] disproved Montgomery's conjecture (1). They showed that for any $A > 0$ there exist arbitrarily large values of x and integers $q \le x/(\log x)^A$ and a with $(a,q) = 1$ for which $|\Delta(x;q,a)| \gg 1$.

Then Friedlander, Granville, Hildebrand and Maier [2] further showed that (1) fails to hold for almost all moduli q as small as $x \exp\{-(\log x)^{1/3-\delta}\}$, for any fixed $\delta > 0$, if the parameter ε in (1) is sufficiently small.

They also showed the following

Theorem A [2]. *Let $\varepsilon > 0$. There exist $N(\varepsilon) > 0$ and $q_0 = q_0(\varepsilon) > 0$ such that for any $q > q_0$ and any x with*

$$q(\log q)^{N(\varepsilon)} < x \le q \exp\{(\log q)^{1/3}\},$$

Project supported by the National Natural Science Foundation of the People's Republic of China.

there exist numbers $x_\pm$ *with* $x/2 < x_\pm \le 2x$ *and integers* $a_\pm$ *coprime to* q *such that*

$$\Delta(x_+; q, a_+) \ge (\log x)^{-5} y^{-(1+\varepsilon)\delta_1(x,y)}, \tag{2}$$

$$\Delta(x_-; q, a_-) \le -(\log x)^{-5} y^{-(1+\varepsilon)\delta_1(x,y)}, \tag{3}$$

where $y = x/q$ *and* $\delta_1(x,y) = 3\log(\log y/\log_2 x)/\log(\log x \log y)$. (*Here* $\log_2 x = \log\log x$.)

It follows from Theorem A that (1) fails to hold for all moduli q with

$$x/(\log x)^{N(\varepsilon)} \ge q > x\exp\{-(\log x)^{1/5-\delta}\}.$$

In this note, our purpose is to extend the above result by showing the following

THEOREM. *For* $\varepsilon > 0$, *there exists* $q_0(\varepsilon) > 0$ *such that for any* $q > q_0(\varepsilon)$ *and any* x *with*

$$q(\log q)^{1+\varepsilon} < x \le q\exp\{(\log q)^{1/3}\}, \tag{4}$$

there exist numbers $x_\pm$ *with* $x/2 < x_\pm \le 2x$ *and integers* $a_\pm$ *coprime to* q *such that*

$$\Delta(x_+; q, a_+) \ge (\log x)^{-3} y^{-(1+\varepsilon)\delta_2(x,y)}, \tag{5}$$

$$\Delta(x_-; q, a_-) \le -(\log x)^{-3} y^{-(1+\varepsilon)\delta_2(x,y)}, \tag{6}$$

where $y = x/q$ *and* $\delta_2(x,y) = 2\log_2 y/\log_2 x$.

It follows from the Theorem that (1) fails to hold for all moduli q with

$$x/(\log x)^{6+\varepsilon} \ge q > x\exp\{-(\log x)^{1/4-\delta}\}.$$

The exponent $1/4$ is the best possible, using this method.

Moreover, we note that the estimates (5) and (6) are slightly better than (2) and (3) for $q < x\exp\{-(\log_2 x)^4\}$.

2. Some lemmas. The following two lemmas are Theorem B2 and Proposition 11.1 of [2], respectively.

LEMMA 1 [2]. *For* $z \ge z_0$, $h \le z/2$, $k \ge 1$, *and* P *the product of any* k *primes all of which are in the interval* $(z-h, z]$, *we have*

$$(-1)^{j-1} r_P(y) := (-1)^{j-1}\Big(\sum_{n\le y,(n,P)=1} 1 - \frac{\varphi(P)}{P} y\Big) \ge \frac{1}{4} y \binom{k}{j} z^{-j},$$

for every integer j *with* $1 \le j \le k/5$ *and every real* y *with* $(z-h)^j \ge y \ge 4jz^j/(k-j+1)$.

LEMMA 2 [2]. *Fix* $\varepsilon > 0$. *For any squarefree integer* $n > 1$ *all of whose prime factors are* $\le n^{1-\varepsilon}$, *there exists a divisor* P *of* n, *with* n/P *prime*, *such that if* $(a, P) = 1$, $x \geq P^2$, *and* $x \geq h \geq x\exp(-\sqrt{\log x})$, *then*

$$\theta(x+h; P, a) - \theta(x; P, a) = \frac{h}{\varphi(P)}(1 + O(e^{-c\log x/\log P} + e^{-c\sqrt{\log x}})),$$

where c *is a constant depending only on* ε.

3. Proof of Theorem. For the proof of this result we use combinatorial means. This is a simple modification of the argument in [2]. We only prove (5), the proof of (6) is similar.

Let $y = x/q$. Define v to be the positive solution of the equation

$$(\lambda v \log_2 x \cdot \log x/\log y)^v = y, \tag{7}$$

where $\lambda = 1 + N/\log y, 1 \leq N \leq 9\log y$, and the positive integer N will be given in the latter part of the proof.

We pick $j = [v] - 1$ or $j = [v]$ so that j is odd. Then we take

$$l = y^{1/j}(\log y/\log x), \tag{8}$$

and

$$z = (l + 1/2)\log x/\log y, \qquad h = (1/2)\log x/\log y, \tag{9}$$

so that $(z-h)^j = y$. By the definition of v, we have $v \leq \log y$ and

$$v \geq (\log y/\log_2 x)(1 + O(\log_3 x/\log_2 x)). \tag{10}$$

From this and the definition of v, we deduce

$$v \leq \log y/\log_2 x. \tag{11}$$

Using the estimates (10) and (11), we obtain

$$\lambda \log y(1 + O(\log_3 x/\log_2 x)) \leq l \leq \lambda \log y \exp\{(5/2)\log_2^2 x/\log y\}. \tag{12}$$

Now take $k = 1 + [c\log x/(20j\log_2^2 x)]$, where c is the constant c of Lemma 2. From this, the definition of j, (10), (11) and the first inequality of (12), we deduce

$$(z-j)^j = y \geq 4jz^j/(k-j+1). \tag{13}$$

Let n be the product of any $k+1$ primes in $(z-h, z]$ that do not divide q. By Huxley's theorem (cf. [2]) we have $\pi(z) - \pi(z-h) \sim h/\log z$ as $z \to \infty$. Now we choose N in (7). First we note that the number of distinct prime factors of q does not exceed $(1+\varepsilon)\log x/\log_2 x$. When N runs over $1, 2, \ldots, [9\log y]$, the intervals $(z-h, z]$ do not overlap. Thus, there is at least one N such that the corresponding interval $(z-h, z]$ contains less than $\nu_q = [(1+\varepsilon)\log x/(8\log y \cdot \log_2 x)]$ primes that divide q. By this we see that the interval $(z-h, z]$ contains at least $\nu_q + k + 1$ primes. Moreover, we choose P as in Lemma 2, with $\varepsilon = 1/2$.

As in [2], we consider the matrix $\mathcal{M} = (a_{rs})$, where $a_{rs} = \log(rP + qs)$ if $rP + qs$ is prime, and $a_{rs} = 0$ otherwise, and where r and s run over the values $R < r \le 2R$ and $1 \le s \le y$ with

$$R = (x/P)\exp\{-\sqrt{\log x}\}. \tag{14}$$

Let $|\mathcal{M}|$ denote the sum of the entries of $\mathcal{M}$. For given s, the sum of entries in the sth column equals

$$\theta(2RP + qs; P, qs) - \theta(RP + qs; P, qs).$$

This vanishes if $(qs, P) > 1$. Now we consider the case when s satisfies $(qs, P) = 1$. Applying Lemma 2 with $x = PR + qs$, $h = PR$, $a = qs$ yields

$$|\mathcal{M}| = \sum_{n\le y,(n,P)=1} \frac{RP}{\varphi(P)}(1 + O(y^{-3})),$$

where we have used the inequalities

$$c\log x/\log P \ge c\log x/(k\log z) \ge 3\log y,$$

which follows from (9)–(11) and the second inequality of (12).

By the definition of $r_P(y)$, we further have

$$|\mathcal{M}| = R\{y + (P/\varphi(P))r_P(y)\}(1 + O(y^{-3})). \tag{15}$$

On the other hand, the number of r satisfying $R < r \le 2R$ and $(r, q) = 1$ equals

$$R\varphi(q)/q + O(\tau(q)) = R\varphi(q)/q(1 + O(y^{-3})).$$

Therefore we may choose some such row (say row r_0) such that the sum of the entries in this row is more than

$$(q/\varphi(q))\{y + (P/\varphi(P))r_P(y)\}(1 + O(y^{-3})). \tag{16}$$

Let $x_0 = x_+ = r_0P + qy$ and $a = a_+ = r_0P$, so $(a, q) = 1$. Now, the sum of the entries in row r_0 equals

$$\theta(r_0P + qy; q, r_0P) - \theta(r_0P; q, r_0P) = \theta(x_0; q, a).$$

(Since, by (14), $r_0P \le 2RP < q$, we have therefore $\theta(r_0P; q, r_0P) = 0$.) By the definitions of θ and Δ and (14) we obtain

$$\theta(x_0; q, a) = (qy/\varphi(q))(1 + \Delta(x_0; q, a))(1 + O(y^{-3})). \tag{17}$$

Combining (16) and (17) yields

$$(-1)^{j-1}\Delta(x_0; q, a) \ge (-1)^{j-1}\frac{P}{\varphi(P)} \cdot \frac{r_P(y)}{y} + O(y^{-2}).$$

Thus, by Lemma 1, (9)–(11) and the second inequality of (12) we obtain

$$(-1)^{j-1}\frac{r_P(y)}{y} \geq \frac{1}{4}\binom{k}{j}\frac{1}{z^j} \gg \frac{1}{\sqrt{j}}\left(\frac{ek}{jz}\right)^j \gg \left(\frac{c_1 \log y}{j^2 l \log_2^2 x}\right)^j$$
$$\gg \exp\left\{-(1+\varepsilon)\frac{\log y}{\log_2 x}\left(2\log_2 y + \frac{5\log_2^2 x}{2\log y}\right)\right\}$$

(where $c_1 = ce/30$). From this, the desired estimate (5) follows.

References

[1] J. Friedlander and A. Granville, *Limitations to the equi-distribution of primes I*, Ann. of Math. 129 (1989), 363–382.

[2] J. Friedlander, A. Granville, A. Hildebrand and H. Maier, *Oscillation theorems for primes in arithmetic progressions*, J. Amer. Math. Soc. 4 (1991), 25–86.

Department of Mathematics
Beijing Normal University
Beijing 100875, China

Received on 12.7.1995 (2830)