Limitations to the Equi-distribution of Primes III

John Friedlander * **and Andrew Granville** **

Abstract: In an earlier paper [FG] we showed that the *expected* asymptotic formula $\pi(x;q,a) \sim \pi(x)/\phi(q)$ does not hold uniformly in the range $q < x/\log^N x$, for any fixed $N > 0$. There are several reasons to suspect that the expected asymptotic formula might hold, for large values of q, when a is kept fixed. However, by a new construction, we show herein that this fails in the same ranges, for a fixed and, indeed, for almost all a satisfying $0 < |a| < x/\log^N x$.

1. Introduction.

For any positive integer q and integer a coprime to q, we have the asymptotic formula

$$\pi(x;q,a) \sim \frac{\pi(x)}{\phi(q)} \tag{1.1}$$

as $x \to \infty$, for the number $\pi(x;q,a)$ of primes $p \le x$ with $p \equiv a \pmod q$, where $\pi(x)$ is the number of primes $\le x$, and ϕ is Euler's function. In fact (1.1) is known to hold uniformly for

$$q < \log^N x \tag{1.2}$$

and all $(a,q) = 1$, for every fixed $N > 0$ (*the Siegel–Walfisz Theorem*), for almost all $q < x^{1/2}/\log^{2+\varepsilon} x$ and all $(a,q) = 1$ (*the Bombieri–Vinogradov Theorem*) and for almost all $q < x/\log^{2+\varepsilon} x$ and almost all $(a,q) = 1$ (*the Barban–Davenport–Halberstam Theorem*). It is widely believed that (1.1) should hold in a far wider range than (1.2) and, partly because of the large number of applications that would follow, this question has received much attention.

Recently, however, the error term

$$\left|\pi(x;q,a) - \frac{1}{\phi(q)}\,\pi(x)\right| \tag{1.3}$$

has been given lower bounds (in [FG] and [FGHM]) that are larger than had been expected ([Mo]), provided that q is fairly large. These bounds even suffice to show that the asymptotic formula (1.1) cannot hold uniformly in the range

$$q < x \,/\, \log^N x \tag{1.4}$$

* Partially supported by NSERC grant A5123
** Partially supported by NSF grant DMS-8610730

for any fixed N. However, in those arithmetic progressions constructed in [FG] and [FGHM], the value of a grows with x so that one can not use them to disprove the asymptotic formula $\pi(x;q,1) \sim \pi(x)/\phi(q)$ in the range (1.4). By a different method, we are now able to do this and indeed much more:

Theorem. *For any given real number $N > 2$ there exist positive constants γ_N, δ_N and Q_N, such that, for all $Q > Q_N$, and for all non–zero integers a with $|a| \le Q$ and having fewer than $(\log Q)^{\gamma_N}$ distinct prime factors, there are at least $Q^{1-1/\log\log Q}$ integers $q_\pm$ with*

$$Q \ < \ q_\pm \ \le \ 2Q \ , \quad (q_\pm, a) \ = \ 1$$

for which

$$\pi(q_+ \log^N q_+;\ q_+, a) \ > \ (1+\delta_N)\ \frac{\pi(q_+ \log^N q_+)}{\phi(q_+)} \tag{1.5}$$

and

$$\pi(q_- \log^N q_-;\ q_-, a) \ < \ (1-\delta_N)\ \frac{\pi(q_- \log^N q_-)}{\phi(q_-)}. \tag{1.6}$$

In fact, we shall only give the proof of (1.5) as the modifications required to prove (1.6) are minor. It is possible to extend this result so as to provide strong lower bounds in (1.3) for much larger values of x (indeed this is why we give our proof of Proposition 2 rather than the shorter proof indicated by the remark at the end of Section 4); however this would be rather complicated, and we do not pursue it here. We shall actually prove the theorem for all non–zero integers a satisfying $|a| \le Q$ and

$$\sum_{p|a} \frac{\log p}{p} \ \le 2\gamma_N \log\log Q; \tag{1.7}$$

all the values of a satisfying the hypothesis of the theorem clearly also satisfy (1.7). Actually the theorem implies that (1.1) cannot hold uniformly in the range (1.4), for *almost all* integers a with $|a| \le Q$ — see the remark after the proof of the theorem. Moreover, an immediate consequence is

Corollary. *For any fixed integer $a \ne 0$ and real $N > 0$ the asymptotic formula (1.1) cannot hold uniformly in the range (1.4).*

2. A discussion of the main ideas.

It had long been believed that an estimate such as

$$\pi(x+y) \;-\; \pi(x) \;\sim\; y/\log x \tag{2.1}$$

holds uniformly as $x \longrightarrow \infty$, for all $y \le x$, with $y/\log^2 x \longrightarrow \infty$. Not only does this follow from the heuristic assumption that a 'randomly chosen' integer n is prime with probability $1/\log n$, but Selberg [Se] had even shown that such a result is 'almost always' true. It thus came as a surprise when, in 1985, Helmut Maier [Ma] introduced a simple, but effective, new idea to show that (2.1) is false for y equal to any fixed power of $\log x$.

Maier started by crossing out those integers that are divisible by a 'small' prime ($\le z$) from the interval $(x, x+y]$, leaving b integers. Now, as a 'randomly chosen' integer is divisible by a given prime p with probability $1/p$, the probability of a 'randomly chosen' integer n being prime, given that it has no prime factors $\le z$, is $1 \Big/ \left\{\prod_{p\le z}\left(1-\frac{1}{p}\right)\right\}\log n$. Thus the 'expected' number of primes in $(x, x+y]$ is $b \Big/ \left\{\prod_{p\le z}\left(1-\frac{1}{p}\right)\right\}\log x$, and this agrees with (2.1) only if

$$b \;\sim\; y \prod_{p\le z}\left(1-\frac{1}{p}\right). \tag{2.2}$$

Maier used a result of Buchstab to find intervals $(x, x+y]$ where (2.2) does not hold (with z a small, fixed power of $\log x$). Then he was able to show that (2.1) cannot hold in some of these intervals, by invoking a deep theorem of Gallagher on the distribution of primes.

We note here Buchstab's result: Define $\Phi(y,z)$ to be the number of integers $\le y$ that are free of prime factors $\le z$. There exists a continuous function ω such that, for any fixed $u > 0$,

$$\Phi(y,z) \;\sim\; e^{\gamma}\,\omega(u)\,y \prod_{p\le z}\left(1-\frac{1}{p}\right) \qquad \text{for } y = z^u, \text{ as } y \longrightarrow \infty.$$

Maier took x to be divisible by the product of the 'small' primes ($\le z$) and so $b = \Phi(y,z)$. If (2.2) were true then we should expect $\omega(u) = e^{-\gamma}$; however, in truth, $\omega(u) - e^{-\gamma}$ oscillates, crossing zero either once or twice in every interval of length 1, though it does tend to zero as $u \to \infty$. (We also note here that $1 \ge \omega(u) \ge 1/2$ for $u > 1$.)

In [FG] we modified Maier's idea to study the distribution of primes in arithmetic progressions. Just like (2.1), the estimate (1.1) had been widely believed to hold uniformly for the range (1.4) for any fixed $N > 2$. However, by constructing arithmetic progressions that do not contain the expected number of terms free of 'small' prime factors, we were able to show that for *almost all* q in the range (1.4), there is some a, with $(a,q) = 1$, for

which (1.1) fails. Like Maier, we used Buchstab's result, although now with a divisible by the product of the 'small' primes.

Actually these results were obtained with sufficient uniformity to establish that, for arbitrarily large values of x, for certain values of a and for $Q = x/\log^N x$,

$$\sum_{\substack{Q<q\le 2Q\\(q,a)=1}} \left|\pi(x;q,a) - \frac{\pi(x)}{\phi(q)}\right| \gg_N \sum_{\substack{Q<q\le 2Q\\(q,a)=1}} \frac{\pi(x)}{\phi(q)}\ . \tag{2.3}$$

The proofs of the above results gave values of a that grow larger as $x \longrightarrow \infty$, and so did not resolve whether (1.1) could hold uniformly for fixed a. There are, perhaps, good reason to guess that (1.1) may hold with larger q than otherwise in the case that a is kept fixed. First, although the Bombieri–Vinogradov Theorem has not been extended beyond $x^{1/2}$, the estimate (1.1) has been shown [BFI] to be true for any fixed a and almost all $q < x^{1/2+o(1)}$ coprime to a. Second, it follows from the Barban–Davenport–Halberstam Theorem that, for almost all a, (1.1) cannot be false for as many arithmetic progressions $a \pmod q$ as it is in (2.3), where a grows in a certain way with x.

In this paper we give a new construction that allows us to prove that (1.1) cannot hold uniformly, for fixed a, in the range (1.4) (although not sufficiently often to give (2.3)). We again use Buchstab's Theorem to construct arithmetic progressions that do not have the expected number of terms free of 'small' prime factors. However, a is now fixed and so cannot be divisible by many 'small' primes; instead, we ensure that this is true of $a + q$, and so consider only those integers q that belong to the arithmetic progression $-a$ modulo the product of the 'small' primes not dividing a.

In the next two sections we prove results needed for the proof of the main Theorem. In Proposition 1 we construct a suitable analogue of Buchstab's result. In Proposition 2 we are careful to minimize the effect of possible Siegel zeros, so as to efficiently apply Gallagher's Theorem.

3. A poorly sifted interval.

For given positive real numbers x and $z > 1$ and integer b, we define

$$\Phi_b(x,z) \;=\; \#\{n \le x \;:\; p|n \;\Rightarrow\; p > z \text{ or } p|b\},$$

$$\sigma(b) \;=\; \sum_{p|b} \frac{\log p}{p}\;,$$

$$P(z) \;=\; \prod_{p\le z} p, \quad P_b(z) \;=\; \prod_{\substack{p\le z\\ p\nmid b}} p, \quad \text{and} \quad b_z \;=\; \prod_{\substack{p\le z\\ p| b}} p \;(= P(z)/P_b(z)).$$

Our first result provides us with an interval 'poorly sifted' by the primes $\le z$:

Proposition 1. *Fix $M > 2$ and $\varepsilon > 0$. For all sufficiently large z, for y satisfying $z^{2+\varepsilon} < y < z^{M-\varepsilon}$, and for integers a and ℓ where $b(:= a\ell)$ has $\le 3z\log^2 z$ distinct prime factors, there exists an integer h, coprime to $P_b(z)$, for which*

$$G(h) :=\; \#\{r : 1 \le r \le y, (r,a) = (r+h, P_b(z)) = 1\}$$

satisfies

$$G(h) \;\ge\; \frac{\ell_z}{\phi(\ell_z)}\;\frac{y}{\log z}\;\{\omega(M) + O(\theta)\}\,, \tag{3.1}$$

where $\theta := \frac12\;\left(\frac{\sigma(b)+\log\log z}{\log z}\right)^{1/2}$.

To prove Proposition 1 we shall need two lemmas. We start by quoting a consequence of the "Fundamental Lemma" of sieve theory (cf. [HRi, Theorem 2.5]), which we shall use repeatedly:

The estimate

$$\#\{n \in (t, t+x] :\; (n,m) = 1\} \;=\; \frac{\phi(m)}{m}\;x\;\left\{1 + O\left(\frac{\log z}{\log x}\right)\right\} \tag{3.2}$$

holds uniformly for any $t \ge 1$ and $x \ge z$, where z denotes the largest prime divisor of m.

Lemma 1. *For M, z, b and θ as in Proposition 1, we have*

$$\Phi_b(z^M, z) \;=\; \frac{b_z}{\phi(b_z)}\;\frac{z^M}{\log z}\;\{\omega(M) + O(\theta)\}. \tag{3.3}$$

Proof: In [FG,Lemma 5] the estimate

$$\Phi_r(z^M, z) \;=\; \frac{r_z}{\phi(r_z)}\;\frac{z^M}{\log z}\;\left\{\omega(M) + O\left(\frac{\log\nu + \log\log z}{\log z}\right)\right\}$$

is proved for integers r with $\nu(= z^{o(1)})$ distinct prime factors (see also the end of §3 of [FGHM] where a gap in the proof was filled); in fact, that proof applies, verbatim, to give this estimate for all integers r with $\nu \le z$. Therefore, if d is the product of the prime divisors of b that are $\le z^\alpha$, where $\alpha = \frac{1}{2}\ (\sigma(b)/\log z)^{1/2}$, then

$$\Phi_d(z^M, z) \ = \ \frac{d}{\phi(d)}\ \frac{z^M}{\log z}\ \left\{\omega(M) + O\left(\alpha \ + \ \frac{\log\log z}{\log z}\right)\right\}.$$

Now

$$\sum_{p|b, p>z^\alpha} \frac{1}{p} \ \le \ \frac{\sigma(b)}{\log z^\alpha} \ = \ 4\alpha \ (\ll \theta)\ , \tag{3.4}$$

and so $\frac{d}{\phi(d)} \ = \ \frac{b_z}{\phi(b_z)}\ e^{O(\alpha)}$; thus the right hand side of (3.3) is an estimate for $\Phi_d(z^M, z)$. Finally, by (3.2),

$$\begin{aligned}\Phi_b(z^M, z) \ - \ \Phi_d(z^M, z) \ &\ll \sum_{p|b, z\ge p>z^\alpha} \Phi_b\left(\frac{z^M}{p}, z\right)\\ &\ll \frac{b_z}{\phi(b_z)}\ \frac{z^M}{\log z} \sum_{p|b, z\ge p>z^\alpha} \frac{1}{p}\ ,\end{aligned}$$

and the result then follows from (3.4).

We also require

Lemma 2. *There exists a ν_0 such that, for any integer m with no more than ν distinct prime factors, the estimate*

$$\#\{n \in (t, t+x] :\ (n, m) = 1\} \ = \ \frac{\phi(m)}{m}\ x\ \{1 + O(\alpha)\} \tag{3.5}$$

holds uniformly in the range $\nu \ge \nu_0,\ x \ge \nu^2,\ t \ge 1$, where $\alpha = \frac{1}{2}\ (\sigma(m)/\log\nu)^{1/2}$.

Proof: Let m_1, m_2, m_3 be the product of the prime factors of m in the ranges $[1, \nu^\alpha]$, $(\nu^\alpha, \nu\log\nu]$, $(\nu\log\nu, \infty)$, respectively. Then

$$\sum_{\substack{t<n\le t+x\\(n,m)=1}} 1 \ = \ \sum_{\substack{t<n\le t+x\\(n,m_1)=1}} 1 \ + \ O\left(\sum_{\substack{t<n\le t+x\\(n,m_1)=1,\ (n,m_2)>1}} 1\right) \ + \ O\left(\sum_{\substack{t<n\le t+x\\(n,m_3)>1}} 1\right). \tag{3.6}$$

Now $m_1/\phi(m_1) \le \prod_{p\le\nu^\alpha}\ p/(p-1) \ \ll \ \alpha\log\nu$ and so the last error term of (3.6) is

$$\ll \sum_{p|\, m_3} \left(\frac{x}{p} + 1\right) \ \le \ x\ \frac{\nu}{\nu\log\nu} \ + \ \nu \ \ll \ \frac{x}{\log\nu} \ \ll \ \frac{\phi(m_1)}{m_1}\ x\ \alpha.$$

As $x \geq p\nu^\alpha$ for each prime p dividing m_2, we can use (3.2) to show that the first error term in (3.6) is

$$\ll \sum_{p|m_2} \sum_{\substack{t/p<n\leq(t+x)/p \\ (n,m_1)=1}} 1 \ll \sum_{p|m,p>\nu^\alpha} \frac{\phi(m_1)}{m_1}\frac{x}{p} \leq \frac{\phi(m_1)}{m_1}\, x\, \frac{\sigma(m)}{\log(\nu^\alpha)} = 4\frac{\phi(m_1)}{m_1}\, x\, \alpha.$$

By again invoking (3.2), we see that the main term on the right hand side of (3.6) is

$$\frac{\phi(m_1)}{m_1}\, x\, \left\{1+O\left(\frac{\log \nu^\alpha}{\log x}\right)\right\} = \frac{\phi(m_1)}{m_1}\, x\, \{1+O(\alpha)\}\,,$$

and so, collecting the estimates above, we have

$$\#\{n\in(t,t+x]:\ (n,m)=1\} = \frac{\phi(m_1)}{m_1}\, x\, \{1+O(\alpha)\}. \tag{3.7}$$

Now

$$0 \leq \log\left(\frac{m}{\phi(m)}\Big/\frac{m_1}{\phi(m_1)}\right) \ll \sum_{p|m,p>\nu^\alpha}\frac{1}{p} \leq \frac{\sigma(m)}{\log(\nu^\alpha)} = 4\alpha\,,$$

which implies that $\frac{\phi(m_1)}{m_1} = \frac{\phi(m)}{m}\{1+O(\alpha)\}$, and the result follows from the substitution of this into (3.7).

Proof of Proposition 1: Let $H=[z^M], \nu = 4z\log^2 z$ and $\zeta = 2^{\sigma(b)}\log^3 z$. By taking $j=r+h$ in the first sum below, we get

$$\begin{aligned}\sum_{\substack{h=0\\(h,P_b(\zeta))=1}}^{H} G(h) &= \sum_{\substack{j=0\\(j,P_b(z))=1}}^{H} \sum_{\substack{r\leq y\\(r,a)=(j-r,P_b(\zeta))=1}} 1 + O(y^2) \\ &= \sum_{\substack{j=0\\(j,P_b(z))=1}}^{H} \sum_{\substack{n=t_j+1\\(n,aP_b(\zeta))=1}}^{t_j+y} 1 + O(y^2) \end{aligned} \tag{3.8}$$

where t_j is any integer such that $t_j \equiv 0 \pmod a$ and $t_j \equiv -j \pmod{P_b(\zeta)}$; such an integer exists by the Chinese Remainder Theorem. Applying Lemma 2 (note that $aP_b(\zeta)$ has $\leq \nu$ distinct prime factors and that $y>\nu^2$) and then Lemma 1, we get

$$\begin{aligned}\sum_{\substack{h=0\\(h,P_b(\zeta))=1}}^{H} G(h) &= \Phi_b(z^M,z)\,\frac{\phi(aP_b(\zeta))}{aP_b(\zeta)}\, y\, \{1+O(\theta)\} + O(y^2) \\ &= \prod_{\substack{p|a\\p>z}}\left(1-\frac{1}{p}\right)\, \frac{\phi(P_b(\zeta))}{P_b(\zeta)}\, H\, \frac{\ell_z}{\phi(\ell_z)}\, \frac{y}{\log z}\, \omega(M)\, \{1+O(\theta)\}\,.\end{aligned}$$

Note that $\prod_{p|a,p>z}\left(1-\frac{1}{p}\right) \;=\; 1+O\left(\frac{\log\log z}{\log z}\right)$, as a has $\le 3z\log^2 z$ distinct prime factors, and so

$$\sum_{\substack{h=0\\(h,P_b(\zeta))=1}}^{H} G(h) \;=\; \frac{\phi(P_b(\zeta))}{P_b(\zeta)}\; H\;\frac{\ell_z}{\phi(\ell_z)}\;\frac{y}{\log z}\;\omega(M)\;\{1+O(\theta)\}. \tag{3.9}$$

Now, by Lemma 2 (with $\nu = z$),

$$\sum_{\substack{h=0\\(h,P_b(\zeta))=1}}^{H} 1 \;=\; \frac{\phi(P_b(\zeta))}{P_b(\zeta)} H\{1+O(\theta)\}\;,$$

and so, comparing the last two estimates, we deduce that there exists an integer $h_1,\ 0\le h_1\le H$, coprime to $P_b(\zeta)$, such that

$$G(h_1)\ge \frac{\ell_z}{\phi(\ell_z)}\frac{y}{\log z}\{\omega(M)+O(\theta)\}. \tag{3.10}$$

The hypothesis of Proposition 1 is almost satisfied by h_1; the only possible problem occurs when h_1 has some illegal prime factors between ζ and z. If so, let $g=(h_1,P_b(z))$. Since $g\le h_1\le z^M$ and every prime factor of g is $\ge\zeta$, thus g has $\le M\log z/\log\zeta$ prime factors. Now let h be chosen to satisfy the congruences

$$h\equiv h_1 (\text{mod } P_b(z)/g),\ h\equiv 1(\text{mod } g);$$

again such a choice is possible by the Chinese Remainder Theorem. Therefore $(h,P_b(z)) = (h_1,P_b(z)/g)=1$, and finally

$$\begin{aligned} G(h) &= \#\{r\le y : (r,a)=(r+h,P_b(z)/g)=1\} + O\left(\sum_{\substack{r\le y\\(r+h,g)>1}} 1\right)\\ &= \#\{r\le y : (r,a)=(r+h_1,P_b(z)/g)=1\} + O\left(\sum_{p|g} y/p\right)\\ &\ge G(h_1)+O(y\log z/\zeta\log\zeta), \end{aligned}$$

so that Proposition 1 follows from (3.10).

Remark: It is possible to obtain essentially the same results based on a somewhat different version of Proposition 1 wherein, rather than specifying z and showing the existence of h, one does the opposite.

4. Good moduli and bad moduli.

We fix $c > 0$ and call a modulus q *good* if the L–function $L(s, \chi)$ has no real zeros β with $\beta > 1 - c/\log q$, for every Dirichlet character $\chi (\mathrm{mod}\ q)$. Landau showed that if c is sufficiently small then for any modulus q there is at most one *exceptional* character and one real zero; we assume that c has been fixed this small. Siegel proved that, for any fixed $\varepsilon > 0$, there exists a constant $c_\varepsilon > 0$, such that the exceptional zero β of any *bad* modulus q satisfies $\beta < 1 - c_\varepsilon q^{-\varepsilon}$ (*bad* means 'not good'). The above results may be found in [Da]. Gallagher [Ga], building on ideas of Linnik, gave a result which immediately implies

$$\pi(X + x; q, a) - \pi(X; q, a) \sim \pi(x)/\phi(q),$$

if q is good and (a, q)=1, provided $\log q = o(\log x)$ and $x \asymp X$. In order to use this estimate, we require a result which allows us to avoid having too many bad moduli:

Proposition 2. *Choose $c > 0$ sufficiently small. For all sufficiently large y and z satisfying $\log y \le z^{1/2}$, and non–zero integers a for which $\sigma(a_z) \le \frac{1}{4}\log z$, there exists an integer $k > z$ such that $(k, a) = 1$ and one of the following conditions holds:*
(i) k divides r for every bad modulus $rP_a(z)$ with $r \le y$.
(ii) k divides $P_a(z)$, $k \le z^2$, $k/\phi(k) = 1 + O(1/\log\log z)$, and k divides r for every bad modulus $rP_{ak}(z)$ with $r \le y$.

We first note a technical lemma, for which the proof is straightforward:

Lemma 3. *Suppose that n and ℓ are positive integers such that every prime $\le z$, which divides n, also divides ℓ.*
(a) If $\ell = 1$ then $\frac{n}{\phi(n)} = 1 + O\left(\frac{1+\log\left(1+\frac{\log n}{z}\right)}{\log z}\right)$.
(b) If $\ell/\phi(\ell) = 1 + O(1/\log\log z)$ and $\log n \le z\log^3 z$ then $n/\phi(n) = 1 + O(1/\log\log z)$.

Given this we proceed to the

Proof of Proposition 2: If every $rP_a(z)$ is good with respect to some sufficiently small constant $c_1 > 0$, then (i) holds. So assume $r_1 P_a(z)$ is bad with exceptional character χ, of conductor d, with zero β. Note first that we may assume $d \ge z^A$ for any fixed A and $z > z_0(A)$ since, if not, choosing $\varepsilon = 1/2A$ in Siegel's Theorem we would obtain

$$\beta < 1 - c_\varepsilon d^{-1/2A} < 1 - c_\varepsilon z^{-1/2}$$

contradicting the estimate

$$\beta > 1 - c_1/\log(r_1 P_a(z)) > 1 - c_1 z^{-2/3},$$

which holds for all sufficiently large z. (Note that $\sigma(a_z) \le \frac{1}{4}\log z$ implies $\log(P_a(z)) \gg z^B$, for any fixed power $B < 3/4$.)

Next observe that, provided $c_1 \le c/3$, any bad modulus m (with character ψ) of the form $rP_a(z)$ or $rP_{ak}(z)$ with $k \le z^2,\ r \le y$, must be divisible by d. To see this, note that both ψ and χ yield bad characters modulo q, where $q = rr_1P_a(z)$, with real zeros

$$> \ 1 - 3c_1/\log q \ \ge \ 1 - c/\log q.$$

By Landau's Theorem these characters coincide and so the modulus of ψ is divisible by d, which is its conductor.

Now let $g = (d, P_a(z))$ and write $d = gk_1$.

(i) If $k_1 > z$ then $k_1 = d/g$ divides $rP_a(z)/g$ and so $k_1|r$ for any bad modulus $rP_a(z)$. Choose $k = k_1$.

(ii) If $k_1 \le z$ then, taking $A = 4$ above, $g = d/k_1 \ge z^3$. Since g is a squarefree divisor of $P_a(z)$ we may pick k to be the smallest divisor of g that is $> z$ and free of prime factors $\le \log z$; thus $k \le z^2$. Also, as k is free of prime factors $\le \log z$ and $k \le z^2$, we have $k/\phi(k) \ = \ 1 + O(1/\log\log z)$, by Lemma 3(a). Finally, as any bad modulus $rP_{ak}(z)$ is divisible by d which is divisible by k, and as $(k, P_{ak}(z)) = 1$ by definition, thus k divides r. This completes the proof of Proposition 2.

Remark: If we fix any N then, in the smaller range $y \le z^N$, there exists a value of $k \le yz$ such that every modulus $rP_{ak}(z)$ with $r \le y$ is good. For, if there exists a bad modulus $rP_a(z)$ then, as in the proof above, we may show that d, the conductor of the bad character, satisfies $d > z^{2N+2}$ for all sufficiently large z. Thus $g = (d, P_a(z))$ is $\ge z^{N+1}$ else, as $d|rP_a(z)$ for some $r \le y$, so $d|rg$ and then $d \le rg \le z^{2N+1}$, giving a contradiction. Now choose k to be the smallest divisor of g, greater than y; therefore it is $\le yz$ as all the prime divisors of g are $\le z$. But, if $rP_{ak}(z)$ is bad then k divides d which divides $rP_{ak}(z)$, and so k divides r, as $(k, P_{ak}(z)) = 1$. But this implies that $y < k \le r \le y$, giving a contradiction.

5. Proof of the Theorem.

Given Q and N as in the statement of the theorem, choose $M > N$ such that $\omega(M) > e^{-\gamma}$ and let $\delta = \delta_N = (e^\gamma \omega(M) - 1)/7$. Let $y = \log^N Q$, and $z = \log Q/(\log\log Q)^2$. For each non–zero integer a, with $|a| \le Q$ and $\sigma(a_z) \le \frac{1}{4}\log z$, select k as in Proposition 2 and set $\ell = 1$ or k according as to whether (i) or (ii) holds. Let $b = a\ell$ and $P = P_b(z)$; note that $\log P \ll z$.

Now, by Proposition 1, there exists an integer h, coprime to P, for which the interval $(h, h+y]$ has at least

$$\frac{\ell_z}{\phi(\ell_z)}\ \frac{y}{\log z}\ \left\{\omega(M) + O\left(\left(\frac{\sigma(b)+\log\log z}{\log z}\right)^{1/2}\right)\right\}$$

integers j satisfying $(j-h, a) = (j, P) = 1$. Select $\gamma_N > 0$ sufficiently small and Q_N sufficiently large, so that if $\sigma(b) \le 3\gamma_N \log z$, then the above quantity is

$$\ge (1+6\delta)e^{-\gamma}\ \frac{\ell_z}{\phi(\ell_z)}\ \frac{y}{\log z} \tag{5.1}$$

for $Q > Q_N$.

Let s be the least positive residue $a/h \pmod P$. Define

$$\begin{aligned} D :=\ & \{q : Q < q \le 2Q,\ q \equiv s \pmod P\} \\ =\ & \{nP + s : (Q-s)/P < n \le (2Q-s)/P\}. \end{aligned}$$

Therefore

$$\begin{aligned} \sum_{q\in D} \pi(qy; q, a) &= \sum_{Q/P < n \le 2Q/P}\ \sum_{\substack{r\le y \\ a+r(nP+s)\ \text{prime}}} 1 \ +\ O(Q/P) \\ &= \sum_{r\le y} \{\pi(2rQ+\alpha_r; rP, \alpha_r) - \pi(rQ+\alpha_r; rP, \alpha_r)\} \ +\ O(Q/P), \end{aligned} \tag{5.2}$$

where, for convenience, $\alpha_r := a + rs$.

We consider now those values of r for which $r \le y$ and $(rP, \alpha_r) = 1$. For those r for which rP is a good modulus we have, by Gallagher's Theorem,

$$\pi(2rQ+\alpha_r; rP, \alpha_r) - \pi(rQ+\alpha_r; rP, \alpha_r) \sim \frac{rQ}{\phi(rP)\log Q}. \tag{5.3}$$

For those r for which rP is a bad modulus we use the Brun–Titchmarsh estimate

$$\pi(2rQ+\alpha_r; rP, \alpha_r) - \pi(rQ+\alpha_r; rP, \alpha_r) \ll \frac{rQ}{\phi(rP)\log Q}. \tag{5.4}$$

Inserting (5.3) and (5.4) in (5.2) we deduce that

(5.5)
$$\sum_{q\in D}\pi(qy;q,a) \;=\; \{1+o(1)\}\sum_{\substack{r\le y\\(rP,\alpha_r)=1}}\frac{rQ}{\phi(rP)\log Q} \;+\; O\left(\frac{Q}{P} \;+\; \sum_{\substack{r\le y\\ r\equiv 0 (\mathrm{mod}\ k)\\(rP,\alpha_r)=1}}\frac{rQ}{\phi(rP)\log Q}\right)$$

where the condition $r\equiv 0 \pmod k$ in the last sum follows since k was selected as in Proposition 2.

We next study the condition $(rP, a+rs)\ (=(rP,\alpha_r)) \;=\; 1$. This is clearly equivalent to the conditions $(r,a) = (P, a+rs) = 1$. However, $a+rs \equiv s(h+r) \pmod P$ and $(s,P)=1$ so

(5.6) $$(rP, a+rs) = 1 \ \text{ if and only if } (r,a) = (P, h+r) = 1.$$

Note that $r\phi(P)/\phi(rP) \;=\; r'/\phi(r')$, where r' is the largest divisor of r that is coprime to P. Now, for those r satisfying $r\le y$ and $(rP,\alpha_r)=1$, we have $r'\le r\le y$, and $(r,a)=1$ so that all prime factors of r' are $>z$ or divide ℓ. Therefore, using the estimate $\ell/\phi(\ell) = 1+O(1/\log\log z)$ from Proposition 2, we deduce that $r'/\phi(r') \;=\; 1+O(1/\log\log z)$ by Lemma 3(b). Substituting this estimate and (5.1) into (5.5), we get on using Mertens' estimate,

$$\begin{aligned}\sum_{q\in D}\pi(qy;q,a) &\ge \;(1+5\delta)\,\frac{e^{-\gamma}}{\phi(P)}\,\frac{Q}{\log Q}\,\frac{y}{\log z} \;+\; O\left(\frac{Q}{P}+\frac{Q}{\phi(P)\log Q}\frac{y}{k}\right)\\ &\ge \;(1+4\delta)\,\frac{\phi(a_z)}{a_z}\,\frac{yQ}{P\log Q},\end{aligned} \tag{5.7}$$

for sufficiently large Q, since $k>z$ and $y>\log^2 Q$.

Suppose now that there are fewer than $Q^{1-1/\log\log Q}$ values of q satisfying (1.5). Then using the trivial bound $\pi(qy;q,a)\le y+1$ for these q, and

$$\pi(qy;q,a) \;\le\; \pi(q\log^N q;q,a) \;\le\; (1+\delta)\,\frac{\pi(q\log^N q)}{\phi(q)} \;<\; (1+2\delta)\,\frac{q}{\phi(q)}\,\frac{y}{\log Q}$$

for the other q, we have

$$\sum_{q\in D}\pi(qy;q,a) \;\le\; (1+2\delta)\,\frac{y}{\log Q}\sum_{\substack{q\in D\\(q,a)=1}}\frac{q}{\phi(q)} \;+\; O(yQ^{1-1/\log\log Q}). \tag{5.8}$$

Now if $q \in D$ and $(q,a) = 1$ then all prime factors of q are $> z$ or divide ℓ and so, as $\log Q/\log z \sim z \log z$, we get $q/\phi(q) \ = \ 1 + O(1/\log\log z)$ by Lemma 3(b). Also

$$\sum_{\substack{q\in D\\(q,a)=1}} 1 \ \le \sum_{\substack{Q<q\le 2Q\\ q\equiv s \pmod P\\(q,a_z)=1}} 1 \ \sim \ \frac{\phi(a_z)}{a_z}\ \frac{Q}{P}$$

by the "Fundamental Lemma" of sieve theory. Therefore, by (5.8),

$$\sum_{q\in D} \pi(qy;q,a) \ < \ (1+3\delta)\ \frac{\phi(a_z)}{a_z}\ \frac{yQ}{P\log Q},$$

which contradicts (5.7) and thus completes the proof.

Remark: It is easy to show that the number of integers a, with $|a| \le Q$ and with $\ge \log^\gamma Q$ distinct prime factors, is $\ll Q/\exp(\log^\gamma Q)$: A famous result of Hardy and Ramanujan [HRa] asserts that the number of integers $\le x$ with exactly $k+1$ distinct prime factors, is $\ll \frac{x}{\log x}\ \frac{(\log\log x + c)^k}{k!}$, for some constant $c > 0$. Thus we can deduce that the number of integers $\le x$ with more than k distinct prime factors, for any $k > 30\log\log x$, is $\ll xe^{-2k}$.

6. Concluding remarks.

In Theorem 3 of [FG] we were able to show that (1.1) is false for *almost all* moduli q; the exceptions including those integers q with many small prime factors. Although we were able to obtain larger than expected lower bounds in (1.3) for these exceptional moduli q, in Theorem A2 of [FGHM], it still remains to determine the truth of (1.1) when, say, q is the product of the first k primes.

Similarly, we have shown here that (1.1) is false for *almost all* a, with $|a| \le Q$, for some q in the range $Q < q < 2Q$; the exceptions including those integers a with many small prime factors. On the other hand it is not difficult to modify the method for those exceptional integers a that are the product of the first k primes. The really difficult values of a come from the set of those a with $\sigma(a)$ around $\frac{1}{2}\log\log Q$.

In order to prove a result like (1.5) for all such a, one could suitably modify our method if one could show:

For any fixed $N > 0$ there exists a constant $\delta_N > 0$ such that, for all sufficiently large Q, and all integers a, with $|a| \le Q$, there exist integers P and h, with $(P, ha) = 1$ and

$$\#\{r \le \log^N Q : \ (r,a) = (r+h,P) = 1\} \ \ge \ \{1+\delta_N\}\ \frac{\phi(aP)}{aP}\ \log^N Q.$$

In our theorem, we obtain only $Q^{1-1/\log\log Q}$ moduli q, with $Q < q \le 2Q$, because of the restriction of q to a suitable arithmetic progression. By taking $z = \log^{2\varepsilon} Q$ in the proof (and making suitable alterations throughout) we can improve this to $Q/\exp(\log^{\varepsilon} Q)$ values of q.

It would be interesting to know how often (1.1) fails, for arithmetic progressions $a \pmod q$ with $(a,q) = 1$ in the range (1.4). From our theorem (and the improvement noted above), (1.1) fails for $\gg x^2/\exp(\log^{\varepsilon} x)$ such arithmetic progressions and, by the Barban–Davenport–Halberstam Theorem, for no more than $O(x^2/\log^{N-1} x)$.

References

[BFI] E. Bombieri, J.B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156**, (1986) 203–251; II, Math. Ann. **277** (1987), 361–393; III, J. Amer. Math. Soc. **2** (1989), 215-224.

[Bu] A.A. Buchstab, *On an asymptotic estimate of the number of numbers of an arithmetic progression which are not divisible by relatively small prime numbers* (Russian), Mat. Sb. **28** (70) (1951), 165–184.

[Da] H. Davenport, Multiplicative Number Theory (2nd ed.) Springer-Verlag (New York) 1980.

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

[FGHM] J. Friedlander, A. Granville, A. Hildebrand, and H. Maier, *Oscillation theorems for primes in arithmetic progressions and for sifting functions*, to appear in J. Amer. Math. Soc.

[Ga] P.X. Gallagher, *A large sieve density estimate near* $\sigma = 1$, Invent. Math. **11** (1970), 329–339.

[HRi] H. Halberstam and H.-E. Richert, Sieve Methods, L.M.S. Monographs, Academic Press (London) 1974.

[HRa] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number* n, Quart. J. Math. **48** (1917), 76–92.

[HM] A. Hildebrand and H. Maier, *Irregularities in the distribution of primes in short intervals*, J. Reine Angew. Math. **397** (1989), 162–193.

[Ma] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), 221–225.

[Mo] H.L. Montgomery, Topics in Multiplicative Number Theory, Lecture Notes in Mathematics, Vol. 227, Springer-Verlag (Berlin) 1971.

[Se] A. Selberg, *On the normal density of primes in small intervals and the difference between consecutive primes*, Arch. Math. Naturvid. **47** (1943), 87–105.

Department of Mathematics, University of Toronto, Toronto, Ontario M5S 1A1, CANADA.

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA.