ON THE RESIDUE CLASSES OF $\pi(n)$ MODULO t

Ping Ngai Chung
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts
briancpn@mit.edu

Shiyu Li [1]
Dept of Mathematics, University of California - Berkeley, Berkeley, California
jjl2357@berkeley.edu

Abstract

The prime number theorem is one of the most fundamental theorems of analytic number theory, stating that the prime counting function, $\pi(x)$, is asymptotic to $x/\log x$. However, it says little about the parity of $\pi(n)$ as an arithmetic function. Using Selberg's sieve, we give, for any fixed integers r and t, a positive lower bound on the proportion of positive integers n such that $\pi(n)$ is congruent to r modulo t. Moreover, we generalize this to the counting function of any set of primes with positive density.

1. Introduction and Statement of Results

The prime counting function $\pi(x)$ has been a topic of interest for mathematicians throughout history. One of the first main results about this function is the prime number theorem, which states that $\pi(x) \sim \frac{x}{\log x}$. This asymptotic carries deep implications for the distribution of prime numbers. However, there are many problems about primes which remain unsolved in spite of this theorem. Two examples are the Goldbach conjecture, stating that every even number greater than 2 is the sum of two primes, and the twin prime conjecture, which claims that there are infinitely many pairs of primes that differ by 2.

In spectacular fashion, there has been a spate of recent advances concerning the gaps between prime numbers. In 2005, Goldston, Pintz and Yıldırım [4] proved that there exist infinitely many consecutive primes which are much closer than average. Recent work by Zhang [12] and the polymath project [10] shows that there are

---

infinitely many pairs of primes differing by at most an absolute constant. These advances depend critically on sifting techniques, particularly Selberg's sieve.

In this paper, we provide another application of Selberg's sieve, one which does not seem to have been considered previously in the literature. We treat $\pi(n)$ as an arithmetic function, and consider its distribution mod t as n varies. The prime number theorem alone says little about the parity of $\pi(n)$ or the proportion of positive integers n for which $\pi(n)$ is even. Nevertheless, numerical evidence shows that this seems to be the case about half the time:

x	100	1000	10^4	10^5	10^6	10^7
$\#\{n < x \mid \pi(n) \text{ is even}\}/x$	0.510	0.523	0.486	0.502	0.498	0.499

Define $T_{r,t}(x)$ to be the number of positive integers $n < x$ such that $\pi(n) \equiv r$ mod t. In this paper, we prove the following.

Theorem 1.1. *For any $0 \le r < t$, we have that*

$$\liminf_{x \to \infty} \frac{T_{r,t}(x)}{x} \ge \frac{1}{16t^2}.$$

If we assume the Hardy-Littlewood prime k-tuple conjecture, we can replace 16 *by* 2 *above.*

Remark 1. Recall that the Hardy-Littlewood prime k-tuple conjecture states that given $\mathfrak{a} = (\alpha_1, ..., \alpha_k)$, where $\alpha_1, \ldots, \alpha_k$ are distinct integers which do not cover all residue classes to any prime modulus, the number $\pi(x; \mathfrak{a})$ of positive integers $m < x$ for which $m - \alpha_1, ..., m - \alpha_k$ are all prime satisfies $\pi(x; \mathfrak{a}) \sim Bx(\log x)^{-k}$, where B is a certain nonzero infinite product over primes. With this estimate, it is possible to improve Theorem 2.3 by a factor of $2^k k!$. In our application, $k = 2$, explaining the improvement of the lower bound by a factor of 8.

Remark 2. Theorem 1.1 was obtained independently and almost simultaneously by Alboiu [1].

Theorem 1.1 is a special case of a more general theorem. We say that a set $\mathfrak{S}$ of primes has density δ if

$$\frac{\#\{p < x : p \in \mathfrak{S}\}}{\#\{p < x : p \text{ prime}\}} \to \delta \qquad \text{as } x \to \infty.$$

Define $T_{r,t}(x; \mathfrak{S})$ to be the number of positive integers $n < x$ such that $|\mathfrak{S}(n)| \equiv r$ mod t, where $|\mathfrak{S}(n)|$ denotes the number of primes in $\mathfrak{S}$ less than n.

Theorem 1.2. *If we assume the notation above and that $\mathfrak{S}$ has positive density δ, then*

$$\liminf_{x \to \infty} \frac{T_{r,t}(x; \mathfrak{S})}{x} \ge \frac{\delta^2}{16t^2}.$$

If we assume the Hardy-Littlewood prime k-tuple conjecture, we can replace 16 *by* 2 *above.*

Examples. Two examples of sets of primes with positive density follow.

1. Let $K/\mathbb{Q}$ be a finite Galois extension with Galois group G, and let C be a conjugacy class in G. Then the set of unramified primes p for which the Frobenius automorphism σ_p is in C has density $\#C/\#G$ by the Chebotarëv density theorem.

2. Given a positive irrational $\alpha > 1$, the set of primes p of the form $\lfloor \alpha n \rfloor$ for some integer n has density α^{-1} by [8, Thm. 2].

In particular, we may consider $T_{r,t}(x; q, a)$ to be the number of positive integers $n < x$ such that $\pi(n; q, a) \equiv r \bmod t$, where $\pi(x; q, a)$ is the counting function of the primes congruent to $a \bmod q$. In this case, we have the following stronger result.

Theorem 1.3. *For any $0 \leq r < t$, $0 \leq a < q$, where $\gcd(a, q) = 1$, we have that*

$$\liminf_{x \to \infty} \frac{T_{r,t}(x; q, a)}{x} \geq \frac{1}{16\phi(q)t^2},$$

where $\phi(q)$ is Euler's totient function.

Remark. Results similar to those in Theorems 1.1, 1.2, and 1.3 can be obtained by making use of Brun's sieve. However, the constants will be weaker.

To prove our theorem, we will rely on a judicious application of Selberg's sieve. In particular, using either Selberg's sieve or the prime k-tuple conjecture, we bound the number of primes which have a particular "small" gap, thus also bounding the number of primes which have any small gap. This lets us say that not all gaps occurring at primes p_{nt+r} can be small, and thus $T_{r,t}(x)$ cannot be too small.

2. Selberg's Sieve

We now introduce the main tool, Selberg's sieve. In this, as well as many other sieve methods (see [2], [7, Ch. 6]), we are given a finite set $\mathcal{A}$ and a set of primes $\mathcal{P}$, and we wish to estimate the size of the set

$$S(\mathcal{A}, \mathcal{P}) := \{a \in \mathcal{A} \mid (a, p) = 1 \text{ for all } p \in \mathcal{P}\}.$$

To do this, we use the inclusion-exclusion principle. For each squarefree integer d, let $\mathcal{A}_d$ be the set of elements in $\mathcal{A}$ divisible by d. Then

$$|S(\mathcal{A}, \mathcal{P})| = \sum_{d | P} \mu(d) |\mathcal{A}_d|,$$

where P is the product of all the primes in $\mathcal{P}$. Sieve methods are most useful when the elements of $\mathcal{A}_d$ distribute approximately evenly in $\mathcal{A}$, in which case one

would expect $|\mathcal{A}_d| = g(d)X + R_d$, where X is approximately the size of $\mathcal{A}$, $g(d)$ is a multiplicative function that is the density of $\mathcal{A}_d$ in $\mathcal{A}$, so that $0 < g(d) < 1$, and R_d is a small error term. Typically, a sieve method gives an estimate for $|S(\mathcal{A}, \mathcal{P})|$ in the following form

$$|S(\mathcal{A}, \mathcal{P})| = X \prod_{p \in \mathcal{P}} (1 - g(p)) + \text{ error term.}$$

The main term is what one would expect if sifting by the primes in $\mathcal{P}$ are pairwise independent events, and the error term varies from different sieve methods. However, when the primes in $\mathcal{P}$ are fairly large, this independence condition fails. Hence, most sieves can only obtain an upper bound for the sifted quantity in question.

The key idea of Selberg's sieve is to replace the Möbius function in the formula of $|S(\mathcal{A}, \mathcal{P})|$ by an optimally chosen quadratic form so that the resulting estimates are minimal. Formally, it is stated in the following theorem.

Theorem 2.1. *(Selberg's sieve) [7, Thm. 6.5] With notation as above, let $h(d)$ be the multiplicative function given by $h(p) = g(p)(1 - g(p))^{-1}$ and set*

$$H(D) = \sum_{d < \sqrt{D}, d|P} h(d)$$

for any $D > 1$. If we assume that

$$|R_d| \leq g(d)d, \tag{1}$$

$$g(d)d \geq 1 \qquad \qquad \text{if } d|P, \tag{2}$$

$$\sum_{y \leq p \leq x} g(p) \log p \ll \log(2x/y) \qquad \text{for all } 2 \leq y \leq x, \tag{3}$$

then

$$S(\mathcal{A}, \mathcal{P}) \leq \frac{X}{H(D)} + O\left(\frac{D}{\log^2 D}\right).$$

With further restrictions on $g(d)$, one may obtain an estimate for $H(D)$ which is frequently useful in practice.

Lemma 2.2. *[6, Lemmas 5.3, 5.4] Suppose the conditions in Theorem 2.1 hold. Moreover, suppose there exist positive real numbers κ, A_1, A_2, L such that*

$$0 \leq g(p) \leq 1 - \frac{1}{A_1}; \tag{4}$$

$$-L \leq \sum_{w \leq p \leq z} g(p) \log p - \kappa \log \frac{z}{w} \leq A_2 \qquad \text{for all } 2 \leq w \leq z. \tag{5}$$

If P is the product of all primes $p < \sqrt{D}$, then

$$H(D) = C(\log \sqrt{D})^\kappa \left(1 + O\left(\frac{L}{\log D}\right)\right),$$

where

$$C = \frac{1}{\Gamma(\kappa+1)} \prod_p (1 - g(p))^{-1} \left(1 - \frac{1}{p}\right)^{\kappa}.$$

The implied constant depends only on A_1, A_2, and κ. In particular, it is independent of L.

Remark. This is a combination of Lemmas 5.3 and 5.4 of Halberstam and Richert's book [6]. Note that exactly this formulation appears in [5].

In our work, we will find the following theorem, which is proved as a direct application of Selberg's sieve, to be useful.

Theorem 2.3. *[7, Thm. 6.7] Let $\mathfrak{a} = (\alpha_1, \ldots, \alpha_k)$ be distinct integers which do not cover all residue classes to any prime modulus and $\alpha_i \leq c \log x$ for some absolute constant c for all i. Then the number $\pi(x; \mathfrak{a})$ of positive integers $m \leq x$ for which $m - \alpha_1, \ldots, m - \alpha_k$ are all prime satisfies*

$$\pi(x; \mathfrak{a}) \leq 2^k k! B x (\log x)^{-k} \left(1 + O\left(\frac{\log \log x}{\log x}\right)\right),$$

where

$$B = \prod_p \left(1 - \frac{\nu(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}$$

and $\nu(p)$ is the number of roots of the polynomial $f(m) = (m - \alpha_1) \cdots (m - \alpha_k)$ modulo p. The implied constant depends only on k and c.

Proof. We apply Selberg's sieve (Theorem 2.1). Take $\mathcal{A}$ to be the set of polynomial values

$$f(m) = (m - \alpha_1) \cdots (m - \alpha_k)$$

for positive integers $m \leq x$. Take $g(p) = \nu(p)p^{-1}$ and $X = x$. Note that $\nu(p) \leq k$ for all primes p and $\nu(p) = k$ for sufficiently large p; thus, conditions (1) - (5) hold for $\kappa = k$, $A_1 = k + 1$, $A_2 = 5k$, $L = (k - 1) \log(\max \alpha_i) = O_{k,c}(\log \log x)$. By Lemma 2.2, we have that

$$H(D)^{-1} = k! \prod_p \left(1 - \frac{\nu(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \left(\frac{\log D}{2}\right)^{-k} \left(1 + O_k\left(\frac{L}{\log D}\right)\right)$$

$$= 2^k k! B (\log D)^{-k} \left(1 + O_k\left(\frac{L}{\log D}\right)\right),$$

hence, by Theorem 2.1 with $D = x \log^{-(k-1)} x$, we have

$$\pi(x; \mathfrak{a}) \leq S(\mathcal{A}, \mathcal{P}) + \sqrt{D} \leq 2^k k! B x (\log x)^{-k} \left(1 + O_k\left(\frac{\max\{L, \log \log x\}}{\log x}\right)\right).$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 1. The prime k-tuple conjecture predicts that $\pi(x; \mathfrak{a}) \sim Bx(\log x)^{-k}$.

Remark 2. For our purpose, it is important that the implied constant depends only on k and c but not the individual values of α_i since we will take α_i to infinity as $x \to \infty$ in the proof of Lemma 3.1.

3. Proof of Theorem 1.2

We are now ready to begin the proof of the main theorem. For each positive integer k, let S_k be the number of positive integers n such that $|\mathfrak{S}(n)| = k$. We label the elements in $\mathfrak{S}$ by $q_1, q_2. \ldots$ in increasing order, and we note that $S_k = q_{k+1} - q_k$. In other words, S_k is just the gap between q_k and q_{k+1}. Observe that for each prime q_m,

$$T_{r,t}(q_m; \mathfrak{S}) = \sum_{\substack{k \le m \\ k \equiv r \,(\mathrm{mod}\ t)}} S_k.$$

By a simple application of the prime number theorem, we see that, for all $A > 1$,

$$T_{r,t}(x, \mathfrak{S}) = \sum_{\substack{q_k < x, \\ k \equiv r \,(\mathrm{mod}\ t)}} (q_{k+1} - q_k) + O_t\left(\frac{x}{\log^A x}\right), \tag{6}$$

where O_t indicates that the implied constant may depend on t.

The proof of Theorem 1.2 relies on the following key lemma, which gives an upper bound on the number of small gaps.

Lemma 3.1. *Let $f(x)$ tend to infinity as $x \to \infty$ and $f(x) = O(\log x)$. Then the number of pairs of primes (p_i, p_j) with $p_i, p_j < x$ such that*

$$0 < p_i - p_j < f(x)$$

is at most

$$\frac{8x}{\log^2 x} f(x)(1 + o(1)).$$

To prove this lemma, we start with an auxiliary estimate, which follows from Theorem 2.3.

Lemma 3.2. *Let a be a positive integer such that $a \le c \log x$ for some positive constant c. Then the number of prime solutions $p_i, p_j < x$ to $p_i - p_j = a$ is at most*

$$\frac{8x}{\log^2 x} B(a) \left(1 + O\left(\frac{\log\log x}{\log x}\right)\right)$$

where

$$B(a) := \prod_{p|a} \left(1 - \frac{1}{p}\right)^{-1} \prod_{p \nmid a} \left(1 - \frac{2}{p}\right) \left(1 - \frac{1}{p}\right)^{-2}$$

and the implied constant depends only on c. Assuming the prime k-tuple conjecture, the upper bound can be improved by a factor of 8.

Proof. In Theorem 2.3, take $k = 2$, $\mathfrak{a} = (0, -a)$. Note that $\nu(p) = 1$ if $p|a$ and $\nu(p) = 2$ if $p \nmid a$. Then the result follows. $\quad\square$

We also need the following estimate.

Lemma 3.3. *Assuming the notation in Lemma 3.2, we have that*

$$\sum_{\substack{a \leq f(x), \\ 2|a}} B(a) = f(x)(1 + o(1)).$$

Proof.

$$\sum_{\substack{a \leq f(x), \\ 2|a}} B(a) = \sum_{\substack{a \leq f(x), \, p|a \\ 2|a}} \prod \left(1 - \frac{1}{p}\right)^{-1} \prod_{p \nmid a} \left(1 - \frac{2}{p}\right) \left(1 - \frac{1}{p}\right)^{-2}$$

$$= 2 \prod_{p \geq 3} \left(1 - \frac{2}{p}\right) \left(1 - \frac{1}{p}\right)^{-2} \sum_{\substack{a \leq f(x), \\ 2|a}} \psi(a),$$

where

$$\psi(a) := \prod_{p|a, p \geq 3} \left(1 + \frac{1}{p - 2}\right).$$

Note that $\psi(a)$ is a multiplicative function, so

$$L(s, \psi) = \sum_{n=1}^{\infty} \frac{\psi(n)}{n^s}$$

$$= \prod_p \left(1 + \frac{\psi(p)}{p^s} + \frac{\psi(p^2)}{p^{2s}} + \dots\right)$$

$$= \prod_p \left(1 + \psi(p) \left(\frac{1}{p^s} + \frac{1}{p^{2s}} + \dots\right)\right)$$

$$= \left(1 - \frac{1}{2^s}\right)^{-1} \prod_{p \geq 3} \left(1 + \frac{p - 1}{(p^s - 1)(p - 2)}\right).$$

Observe that $L(s, \psi) = \zeta(s)A(s)$, where

$$A(s) = \prod_{p \geq 3} \left(1 + \frac{1}{p^s(p-2)}\right)$$

is absolutely convergent in $\mathrm{Re}(s) > 0$. Hence by the Wiener-Ikehara Tauberian theorem [9, Cor. 8.8],

$$\sum_{\substack{a \leq f(x), \\ 2|a}} \psi(a) = \sum_{a \leq f(x)/2} \psi(a) = A(1)\frac{f(x)}{2}(1 + o(1))$$

$$= \prod_{p \geq 3}\left(1 + \frac{1}{p(p-2)}\right)\frac{f(x)}{2}(1 + o(1)).$$

Therefore

$$\sum_{\substack{a \leq f(x), \\ 2|a}} B(a) = 2\prod_{p \geq 3}\left[\left(1 - \frac{2}{p}\right)\left(1 - \frac{1}{p}\right)^{-2}\left(1 + \frac{1}{p(p-2)}\right)\right]\frac{f(x)}{2}(1 + o(1))$$

$$= f(x)(1 + o(1)).$$

<div align="right">□</div>

Using Lemma 3.2 and 3.3, we can give an upper bound on the number of pairs of primes with prime gap in a given range.

Proof of Lemma 3.1. By Lemmas 3.2 and 3.3, the number of such pairs (p_i, p_j) is bounded by

$$\sum_{a \leq f(x)} \frac{8x}{\log^2 x} B(a) \left(1 + O\left(\frac{\log \log x}{\log x}\right)\right)$$

$$= \frac{8x}{\log^2 x}\left(1 + O\left(\frac{\log \log x}{\log x}\right)\right) \sum_{\substack{a \leq f(x), \\ 2|a}} B(a) + O(f(x))$$

$$= \frac{8x}{\log^2 x} f(x)(1 + o(1)).$$

<div align="right">□</div>

If we take $f(x) = \alpha \log x$ with α small, then we can give an upper bound on the number of consecutive primes with an unusually small prime gap. From this regularity result, we will now deduce the main theorem.

Proof of Theorem 1.2. Recall from (6) that

$$T_{r,t}(x; \mathfrak{S}) = \sum_{\substack{q_k < x, \\ k \equiv r (\text{mod } t)}} (q_{k+1} - q_k) + O_t \left(\frac{x}{\log^A x} \right).$$

Given a function $f(x)$ which tends to ∞ as $x \to \infty$ and $f(x) = O(\log x)$, by Lemma 3.1, we know that

$$\#\{p_n < x \mid p_{n+1} - p_n < f(x)\} \leq \frac{8x}{\log^2 x} f(x)(1 + o(1)). \tag{7}$$

Note that

$$\#\{q_n < x \mid n \equiv r (\text{mod } t)\} = \frac{\delta \pi(x)}{t}(1 + o(1)) \geq \frac{\delta x}{t \log x}(1 + o(1))$$

by the prime number theorem and the estimate in [11, Cor. 1, P.69]. Hence

$$\#\{q_n < x \mid q_{n+1} - q_n \geq f(x) \text{ and } n \equiv r (\text{mod } t)\} \geq \left(\frac{\delta x}{t \log x} - \frac{8x}{\log^2 x} f(x) \right)(1 + o(1))$$

In particular, for $f(x) \leq \delta \log x / (8t)$, the main term is nonnegative. Hence

$$T_{r,t}(x; \mathfrak{S}) = \sum_{\substack{q_k < x, \\ k \equiv r (\text{mod } t)}} (q_{k+1} - q_k) + O_t \left(\frac{x}{\log^A x} \right) = \sum_{y=1}^{\infty} \sum_{\substack{q_k < x, \\ k \equiv r (\text{mod } t), \\ q_{k+1} - q_k = y}} \sum_{z=1}^{y} 1 + O_t \left(\frac{x}{\log^A x} \right)$$

$$= \sum_{z=1}^{\infty} \sum_{y=z}^{\infty} \sum_{\substack{q_k < x, \\ k \equiv r (\text{mod } t), \\ q_{k+1} - q_k = y}} 1 + O_t \left(\frac{x}{\log^A x} \right) \geq \sum_{z=1}^{\delta \log x / (8t)} \sum_{\substack{q_k < x, \\ k \equiv r (\text{mod } t), \\ q_{k+1} - q_k \geq z}} 1 + O_t \left(\frac{x}{\log^A x} \right)$$

$$\geq \sum_{z=1}^{\delta \log x / (8t)} \left(\frac{\delta x}{t \log x} - \frac{8x}{\log^2 x} z \right)(1 + o(1)) + O_t \left(\frac{x}{\log^A x} \right)$$

$$= \left(\frac{\delta x}{t \log x} \frac{\delta \log x}{8t} - \frac{8x}{2 \log^2 x} \left(\frac{\delta \log x}{8t} \right)^2 \right)(1 + o(1))$$

$$= \frac{\delta^2 x}{16t^2}(1 + o(1)).$$

Therefore, we have

$$\frac{T_{r,t}(x; \mathfrak{S})}{x} \geq \frac{\delta^2}{16t^2}(1 + o(1)),$$

and so

$$\liminf_{x \to \infty} \frac{T_{r,t}(x; \mathfrak{S})}{x} \geq \frac{\delta^2}{16t^2}.$$

If we assume the prime k-tuple conjecture, the upper bound in Lemma 3.2 can be improved by a factor of 8. Proceeding analogously in our above proof, we obtain

$$\liminf_{x\to\infty} \frac{T_{r,t}(x;\mathfrak{S})}{x} \geq \frac{\delta^2}{2t^2}.$$

$\square$

Proof of Theorem 1.3. Theorem 1.3 follows *mutatis mutandis* by considering only prime gaps divisible by q and primes congruent to a mod q in Lemma 3.2, in which case we obtain an extra factor of $\phi(q)$ in the denominator of the upper bound in (7). We also recall that $\pi(x;q,a) \sim \pi(x)/\phi(q)$ by Dirichlet's theorem on arithmetic progressions. $\square$

References

[1] M. Alboiu, *On the parity of the prime counting function and related problems*. Preprint.

[2] A. Cojocaru and M. Murty, *An introduction to sieve methods and their applications*. London Math. Soc. Stud. Texts, 2006.

[3] S. R. Finch, *Mathematical Constants*. Encyclopedia Math. Appl., vol. 94, Cambridge Univ. Press, 2003.

[4] D. Goldston, J. Pintz and C. Yıldırım, *Primes in tuples. I*. Ann. of Math. (2) 170 (2009), no. 2, 819-862.

[5] D. Goldston, S. Graham, J. Pintz and C. Yıldırım, *Small gaps between products of two primes*. Proc. Lond. Math. Soc. (3) 98 (2009), no. 3, 741-774.

[6] H. Halberstam and H.-E. Richert, *Sieve methods*. London Math. Soc. Monogr. Ser., 1974.

[7] H. Iwaniec and E. Kowalski, *Analytic number theory*. Amer. Math. Soc. Colloq. Publ., 2004.

[8] D. Leitmann and D. Wolke, *Primzahlen der Gestalt* $[f(n)]$. (German) Math. Z. 145 (1975), no. 1, 81-92.

[9] H. Montgomery, R. Vaughan, *Multiplicative Number Theory I. Classical Theory*. Cambridge Stud. Adv. Math., 2007.

[10] Polymath, *Improving the bounds for small gaps between primes*. http://michaelnielsen.org/polymath1/index.php?title=Bounded_gaps_between_primes.

[11] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*. Illinois J. Math. Volume 6, Issue 1 (1962), 64-94.

[12] Y. Zhang, *Bounded gaps between primes*. Ann. of Math., to appear.