Large gaps in sets of primes and other sequences
II. New bounds for large gaps between primes
Random methods and weighted sieves

Kevin Ford

University of Illinois at Urbana-Champaign

May, 2018

Large gaps between primes

Def: $G(x) := \max_{p_n \leqslant x}(p_n - p_{n-1})$, p_n is the n^{th} prime.

Theorem (F-Green-Konyagin-Maynard-Tao, 2018)

$$G(x) \gg \log x \frac{\log_2 x \log_4 x}{\log_3 x}.$$

Chains of gaps:

$G_k(x) := \max_{p_{n+k} \leqslant x} \min(p_{n+1} - p_n, \ldots, p_{n+k} - p_{n+k-1})$

Theorem (F-Maynard-Tao, 2018+)

For every k,

$$G_k(x) \gg_k \log x \frac{\log_2 x \log_4 x}{\log_3 x}$$

Proving large gaps: Jacobsthal's function

$$\mathcal{S}_T = \{n \in \mathbb{Z} : (n, Q_T) = 1\}, \quad Q_T = \prod_{p \leqslant T} p.$$

Main goal: Find $J(T)$, the largest gap in $\mathcal{S}_T$.

Lower bound (FGKMT, 2018). $J(T) \gg T \frac{\log T \log_3 T}{\log_2 T}$.

Covering: $J(T)$ is the largest y so that there are $a_2, a_3, a_5, \ldots$ with

$$\{a_p \mod p : p \leqslant T\} \supseteq [0, y]$$

Least prime in an arithmetic progression

Let $p(k,l) = \min\{p : p \equiv l \pmod{k}\}$, $M(k) = \max_{(l,k)=1} p(k,l)$.

Upper bounds

Linnik, 1944. $M(k) \ll k^L$. **(Xylouris -** $L = 5.18$**).**
ERH: $L = 2 + \varepsilon$;   Chowla conjecture: $L = 1 + \varepsilon$.

Lower bounds

Trivial: $M(k) \gg \phi(k) \log k$.
Prachar; Schinzel - 1961/62. For infinitely many k,

$$M(k) \gg \phi(k) \log k \frac{\log_2 k \log_4 k}{(\log_3 k)^2}. \tag{1}$$

Wagstaff (1978) - (1) holds for all prime k.
Pomerance (1980) - (1) holds for almost all k, in fact all k with at most $\exp(\log_2 k / \log_3 k)$ prime factors.

Least prime in an arithmetic progression, II

Pomerance: $M(k) \gg \phi(k) \log k \frac{\log_2 k \log_4 k}{(\log_3 k)^2}$ for almost all k.

Lemma (Pomerance): Let $j(m)$ be the maximal gap between numbers comprime to m. If $0 < m \leqslant k/j(k)$ and $(m, k) = 1$ then $M(k) > kj(m)$.

Take $m = \prod_{\substack{p \leqslant (1-\delta) \log k \\ p \nmid k}} p$ need a lower bound on $j(m)$.

Corollary (FGKMT, 2018). If k has no prime factor $\leqslant \log k$, then

$$M(k) \gg \phi(k) \log k \frac{\log_2 k \log_4 k}{\log_3 k}. \tag{2}$$

Theorem (J. Li-K. Pratt-G. Shakan, 2017)

Inequality (2) holds for all k with at most $\exp\{(1/2 - \varepsilon)\frac{\log_2 k \log_4 k}{\log_3 k}\}$ prime factors.

Least Prime in an A.P. – conjectures

Conjecture (folklore): $M(k) \ll k \log^{2+\varepsilon} k$.

Conjecture (Wagstaff, 1979): $M(k) \sim \phi(k) \log^2 k$ for "most k"

Wagstaff's heuristic:Given $l < k(\log k)^3$, $(l, k) = 1$, the "probability" that l is prime is $\approx \frac{k/\phi(k)}{\log k}$. So

$$\mathbb{P}(l, l+k, \ldots, l + \lfloor m \log k \rfloor k \text{ all composite}) \sim \left(1 - \frac{k/\phi(k)}{\log k}\right)^{m \log k}$$
$$\sim e^{-m/\phi(k)}.$$

Hence

$$\mathbb{P}\left(M(k) \leqslant mk \log k\right) \sim \left(1 - e^{-mk/\phi(k)}\right)^{\phi(k)}$$
$$\sim \exp\{-\phi(k) e^{-mk/\phi(k)}\}.$$

Threshhold value $m \sim \frac{\phi(k)}{k} \log \phi(k)$

Least prime in AP: Refined conjectures

Conjecture (Li-Pratt-Shakan, 2017)

$$\liminf_{k\to\infty} \frac{M(k)}{\phi(k)\log^2 k} = 1, \qquad \limsup_{k\to\infty} \frac{M(k)}{\phi(k)\log^2 k} = 2.$$

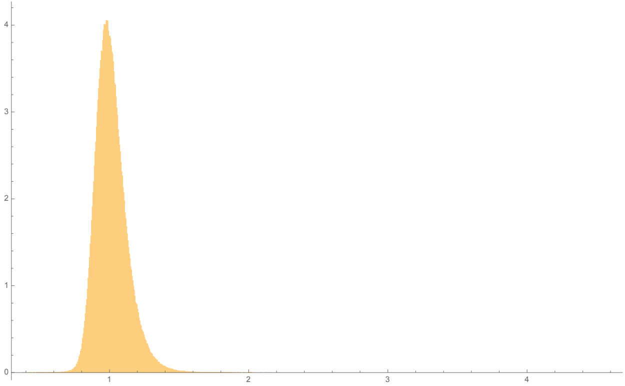


Figure: Histogram for $M(k)/\phi(k)\log(\phi(k))\log k$ for $k \leqslant 10^6$

Least prime in AP: Li-Pratt-Shakan conjecture

Rough heuristic argument: "coupon collectors problem"
p_n - n-th prime, m_k - a param, $a \in \mathbb{Z}/k\mathbb{Z}$

E_a - the event that $p_1, p_2, \ldots, p_{m_k} \not\equiv a \pmod{k}$

A_k - the event $\{M(k) > p_{m_k}\} = \bigcup_a E_a$.

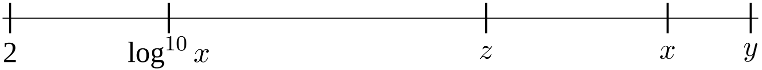
We have $\mathbb{P}(A_k) \sim \sum_a \mathbb{P}(E_a) \sim \phi(k) e^{-m/\phi(k)}$

If $m = \lambda \phi(k) \log \phi(k)$, this is $\sim \phi(k)^{1-\lambda}$.

(1) If $\lambda \approx 1$, threshhold for being "small". Justifies Wagstaff and $\liminf$.

(2) When $\lambda \approx 2$, threshhold for $\mathbb{P}(A_k)$ holding for infinitely many k (using Borel-Cantelli). Justifies $\limsup$.

New lower bounds on $J(T)$: outline



$y = cx\frac{\log x \log_3 x}{\log_2 x}$, $z = x^{c\frac{\log_3 x}{\log_2 x}}$ Want $\{a_p \bmod p : p \leqslant x\} \supseteq [0, y]$

1. $a_p = 0$ for $p \in (z, x/4] \cap [2, \log^{10} x]$. Uncovered: z-smooth numbers and primes;
2. Random, uniform choice of a_p, $\log^{10} x < p \leqslant z$.
3. Strategic choice of a_p, $x/4 < p \leqslant x/2$ to cover *many* reminaing elements.
4. (trivial) Use single a_p for each $x/2 < p \leqslant x$ to cover each remaining uncovered element.

Stage 2: random, uniform choice of a_p

$\mathcal{Q}_1$ - the set of uncovered elements after stage 1 (mainly primes).

$$\mathcal{Q}_2(\mathbf{a}) := \text{ the set of uncovered elements after stage 2}$$
$$= \mathcal{Q}_1 \setminus \bigcup_{p \in \mathcal{P}} (a_p \mod p),$$

where $\mathcal{P}$ is the set of primes in $(\log^{10} x, z]$.

Lemma

w.h.p., $|\mathcal{Q}_2(\mathbf{a})| \sim \sigma \pi(y)$, $\sigma := \prod_{p \in \mathcal{P}} (1 - 1/p)$

Proof. Recall $|\mathcal{Q}_1| \sim \pi(y)$. We calculate 1st, 2nd moments:

$$\mathbb{E}|\mathcal{Q}_2(\mathbf{a})| = \sum_{n \in \mathcal{Q}_1} \mathbb{P}(n \in \mathcal{Q}_2(\mathbf{a}))$$
$$= \sum_{n \in \mathcal{Q}_1} \prod_{p \in \mathcal{P}} \mathbb{P}(n \not\equiv a_p \pmod p) = \sigma |\mathcal{Q}_1|.$$

Lemma

w.h.p., $|\mathcal{Q}_2(\mathbf{a})| \sim \sigma\pi(y)$, $\sigma := \prod_{p\in\mathcal{P}}(1-1/p)$

Proof (continued). For the 2nd moment,

$$\mathbb{E}|\mathcal{Q}_2(\mathbf{a})|^2 = \sum_{n_1,n_2\in\mathcal{Q}_1} \mathbb{P}(n_1,n_2 \in \mathcal{Q}_2(\mathbf{a}))$$

$$= \mathbb{E}|\mathcal{Q}_2(\mathbf{a})| + \sum_{\substack{n_1,n_2\in\mathcal{Q}_1 \\ n_1\neq n_2}} \prod_{p\in\mathcal{P}} \mathbb{P}(n_i \not\equiv a_p \pmod p; i=1,2).$$

Now $\mathbb{P}(n_i \not\equiv a_p \pmod p; i=1,2) = 1-2/p$ unless $p|n_1-n_2$, which occurs for $O(\log x)$ primes p. Get

$$\mathbb{E}|\mathcal{Q}_2(\mathbf{a})|^2 = \sum_{n_1,n_2\in\mathcal{Q}_1} \sigma^2\left(1+O((\log x)^{-9})\right)$$

$$= (\sigma|\mathcal{Q}_1|)^2\left(1+O((\log x)^{-9})\right).$$

The Lemma follows from the 1st, 2nd moment bounds plus Chebyshev's inequality.

Random residues: higher correlations

Define the random sifted set

$$\mathcal{S}(\mathbf{a}) = \mathbb{Z} \setminus \bigcup_{p \in \mathcal{P}} (a_p \mod p).$$

In particular, $\mathcal{Q}_2(\mathbf{a}) = \mathcal{Q}_1 \cap \mathcal{S}(\mathbf{a})$.

Lemma ($\mathcal{S}(\mathbf{a})$ correlations)

Let $n_1, \ldots, n_t$ *be distinct integers in* $[-y, y]$*, with* $t \ll \log x$*. Then*

$$\mathbb{P}(n_1, \ldots, n_t \in \mathcal{S}(\mathbf{a})) = \sigma^t \left(1 + O(t^2/\log^9 x)\right).$$

Stage 3: Strategic choices

We choose a_p, $x/4 < p \leqslant x/2$ to have two properties:

(a) the sets $e_p := (a_p \mod p) \cap \mathcal{Q}_2(\mathbf{a})$ are large (on average) for $x/4 < p \leqslant x/2$;

(b) the collection of sets $\{e_p : x/4 < p \leqslant x/2\}$ covers most of $\mathcal{Q}_2(\mathbf{a})$ *efficiently* (little overlap).

Item (a) is accomplished using a *weighted, prime detecting sieve*. Recall that $\mathcal{Q}_1$, and hence $\mathcal{Q}_2(\mathbf{a})$ consists mainly of primes. The **average** of $|e_p|$, over all choices of a_p is

$$\frac{|\mathcal{Q}_2(\mathbf{a})|}{p} \asymp \frac{|\mathcal{Q}_2(\mathbf{a})|}{x} \sim \frac{\sigma y}{\log x} = o(1).$$

So a random (uniform) choice for a_p is very inefficient!

Item (b) is accomplished using *hypergraph covering methods*.

Primes in sparse A.P.'s: weighted sieves

Admissible k-tuple $h_1, \ldots, h_k$
Prime-detecting weight fcn. $w(n) = w(n; \mathbf{h})$ (GPY-Maynard-Tao)

Goal: Find $w(n)$ which is large when many of the numbers $n + h_i$ are prime, and small otherwise, and such that the sums

$$T_1(N) = \sum_{n \asymp N} w(n), \qquad T_2(N) = \sum_{n \asymp N} \sum_{j=1}^{k} \mathbf{1}(n + h_j \text{ prime}) w(n)$$

can both be evaluated asymptoticlaly. If

$$T_2(N) \geqslant r T_1(N), \tag{w}$$

then there are some values of $n \asymp N$ such that the set $\{n + h_1, \ldots, n + h_k\}$ contains at least r primes.

Theorem (Maynard, 2016)

For $k \leqslant (\log N)^{1/5}$, $h_i \ll x^c$, $\exists$ weights s.t. (w) holds with $r \sim \log k$.

Sieve weights

Fix an admissible k-tuple $1 \leqslant h_1 < \cdots < h_k \ll k^2$, $k \sim (\log x)^{1/5}$.
Let $x/4 < p \leqslant x/2$. Then $\mathbf{h}_p := (h_1 p, \ldots, h_k p)$ is admissible.
Define the weight $w(p, n)$ by

$$w(p,n) = w(n; \mathbf{h}_p); \quad (0 \leqslant n \leqslant y).$$

Two crucial estimates (after suitable notmalization)

Theorem (FGKMT)

(a) $$\frac{1}{\pi(y)} \sum_{n \leqslant y} w(p,n) \sim 1 \quad (p \in \mathcal{P});$$

(b) $$\frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \sum_{i=1}^{k} w(p, q - h_i p) \sim \log_2 x \quad (x < q \leqslant y]$$

(a) is a T_1 sum; (b) is a T_2-type sum (with a *different* k-tuple).

Weighted choice of a_p for $x/4 < p \leqslant x/2$

Select a random number in $\mathbf{n}_p \in [0, y]$ with probability proportional to $w(p, n)$; that is

$$\mathbb{P}(\mathbf{n}_p = n) := \frac{w(p,n)}{\sum_l w(p,l)} \quad (0 \leqslant n \leqslant y).$$

Bigger weight when many of $n + h_i p$ are prime.
For each $p \in \mathcal{P}$ and fixed (non-random) vector $\vec{a}$, let

$$X_p(\vec{a}) := \mathbb{P}(\mathbf{n}_p + h_i p \in \mathcal{S}(a) \text{ for all } i = 1, \dots, k).$$

Lemma ($\mathcal{Q}_2(\mathbf{a})$ correlations) + Chebyshev $\Rightarrow X_p(\mathbf{a}) \sim \sigma^k$ w.h.p.
Spse $\vec{\mathbf{a}} = \vec{a}$. Define r.v. $\mathbf{m}_p$ by

$$\mathbb{P}(\mathbf{m}_p = m | \mathbf{a} = \vec{a}) := \frac{Z_p(\vec{a}; m)}{X_p(\vec{a})},$$

$$Z_p(\vec{a}; m) = \begin{cases} \mathbb{P}(\mathbf{n}_p = m) & \text{if } m + h_j p \in \mathcal{S}(\vec{a}) \text{ for } j = 1, \dots, k \\ 0 & \text{otherwise,} \end{cases}$$

weights, II

$$\mathbb{P}(\mathbf{m}_p = m | \mathbf{a} = \vec{a}) := \frac{Z_p(\vec{a}; m)}{X_p(\vec{a})},$$

$$Z_p(\vec{a}; m) = \begin{cases} \mathbb{P}(\mathbf{n}_p = m) & \text{if } m + h_j p \in \mathcal{S}(\vec{a}) \text{ for } j = 1, \dots, k \\ 0 & \text{otherwise,} \end{cases}$$

Let $a_p \equiv \mathbf{m}_p \pmod p$, $x/4 < p \leqslant x/2$. Then

(1) $\mathbf{m}_p + h_i p \in \mathcal{S}(\mathbf{a})$ for all i;

(2) (on avg) $|\mathcal{Q}_2(\mathbf{a}) \cap (a_p \mod p)| \gtrsim \log k \gg \log_2 x$ (the k-tuple contains many primes)

If the sets $e_p = \mathcal{Q}_2(\mathbf{a}) \cap (a_p \mod p)$ have little overlap (efficiently chosen), they cover about $\gg x \log_2 x / \log x$ elements. Good if

$$\sigma \frac{y}{\log x} \leqslant c \frac{x \log_2 x}{\log x}.$$