Large gaps in sets of primes and other sequences

Kevin Ford

University of Illinois at Urbana-Champaign

October, 2018

Large gaps between primes

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

$$2, 3, 5, 7, \ldots, 109, 113, 127, 131, \ldots, 9547, 9551, 9587, 9601, \ldots$$

Upper bound: $G(x) \ll x^{0.525}$ (Baker-Harman-Pintz, 2001).

Large gaps between primes

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

$$2, 3, 5, 7, \ldots, 109, 113, 127, 131, \ldots, 9547, 9551, 9587, 9601, \ldots$$

Upper bound: $G(x) \ll x^{0.525}$ (Baker-Harman-Pintz, 2001).

Rankin, 1938. $G(x) \geqslant c(\log x)\dfrac{\log_2 x \log_4 x}{(\log_3 x)^2}$, $c = \frac{1}{3}$.

Improvements to the constant c until 1997 (Pintz, $c = 2e^{\gamma}$).

Large gaps between primes

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

$$2, 3, 5, 7, \ldots, 109, 113, 127, 131, \ldots, 9547, 9551, 9587, 9601, \ldots$$

Upper bound: $G(x) \ll x^{0.525}$ (Baker-Harman-Pintz, 2001).

Rankin, 1938. $G(x) \geqslant c(\log x)\frac{\log_2 x \log_4 x}{(\log_3 x)^2}$, $c = \frac{1}{3}$.

Improvements to the constant c until 1997 (Pintz, $c = 2e^\gamma$).

Erdős Conjecture. ($10,000) Rankin's bound is true, for *any* $c > 0$.
Solved: Ford-Green-Konyagin-Tao and Maynard (arXiv,Aug-2014).

Theorem (Ford-Green-Konyagin-Maynard-Tao (2018))

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

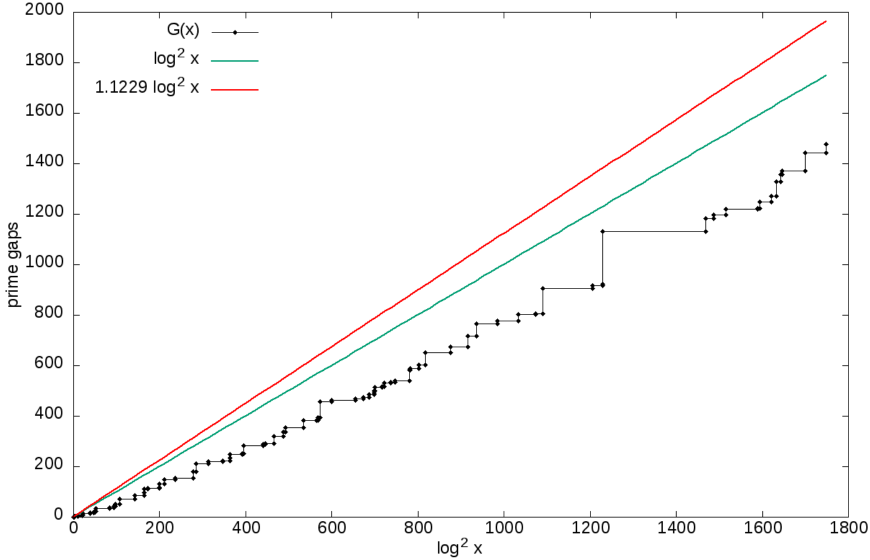
Conjectures

Cramér (1936):

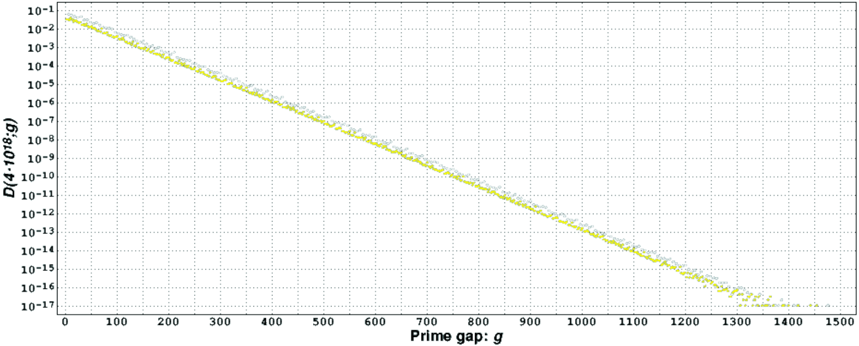
- $\displaystyle\limsup_{x\to\infty} \frac{G(x)}{\log^2 x} = 1.$
- $\displaystyle\frac{p_n - p_{n-1}}{\log p_n}$ has approximate exponential distribution.

Granville (1995): $\displaystyle\limsup_{x\to\infty} \frac{G(x)}{\log^2 x} \geqslant 2e^{-\gamma} = 1.1229\ldots$

Computational evidence, up to 10^{18}



Exponential distribtion of gaps



Prime gap statistics, $p_n < 4 \cdot 10^{18}$

Gallagher, 1976. Prime $k-$tuples conjecture $\Rightarrow$ exponential prime gap distribution

Cramér's model defects: small gaps

Cramér's model produces a set $\mathscr{P} \in \mathbb{N}$ of "random primes":

$$\mathbb{P}(n \in \mathscr{P}) = 1/\log n \quad (n \geqslant 3).$$

Theorem. With probability 1, $\#\{n : n, n+1 \in \mathscr{P}\} = \infty$

This does not hold for real primes!

Cramér’s model defects: small gaps

Cramér’s model produces a set $\mathscr{P} \in \mathbb{N}$ of “random primes”:

$$\mathbb{P}(n \in \mathscr{P}) = 1/\log n \quad (n \geqslant 3).$$

Theorem. With probability 1, $\#\{n : n, n+1 \in \mathscr{P}\} = \infty$

This does not hold for real primes!

Theorem. With probability 1,

$$\#\{n \leqslant x : n, n+2 \in \mathscr{P}\} \sim \frac{x}{\log^2 x}.$$

Conjecture (Hardy-Littlewood, 1923).

$$\#\{n \leqslant x : n, n+2 \text{ prime}\} \sim C\frac{x}{\log^2 x},$$

where $C = 2\prod_{p>2}(1 - 1/(p-1)^2) \approx 1.3203$

General Cramér type model

Theorem (classical, 1960s)

Choose N random points in $[0, x]$. With high probability, the max. gap is $\sim \frac{\log N}{N}\, x$.

Conjecture (Hardy-Littlewood;Bateman-Horn)

Let $f_1, \ldots, f_k$ be distinct, irreducible polynomials $f_i : \mathbb{Z} \to \mathbb{Z}$ with pos. leading coeff., and $f_1 \cdots f_k$ has no fixed prime factor. Then

$$\#\{n \leqslant x : f_1(n), \ldots, f_k(n) \text{ all prime}\} \sim C \frac{x}{(\log x)^k},$$

where $C = C(f_1, \ldots, f_k) > 0$ is constant.

General Cramér type model

Theorem (classical, 1960s)

Choose N random points in $[0, x]$. With high probability, the max. gap is $\sim \frac{\log N}{N}\, x$.

Conjecture (Hardy-Littlewood;Bateman-Horn)

Let $f_1, \ldots, f_k$ be distinct, irreducible polynomials $f_i : \mathbb{Z} \to \mathbb{Z}$ with pos. leading coeff., and $f_1 \cdots f_k$ has no fixed prime factor. Then

$$\#\{n \leqslant x : f_1(n), \ldots, f_k(n) \text{ all prime}\} \sim C \frac{x}{(\log x)^k},$$

where $C = C(f_1, \ldots, f_k) > 0$ is constant.

For $\{n \leqslant x : f_1(n), \ldots, f_k(n) \text{ all prime}\}$, the model prediction:

average gap $\sim \frac{(\log x)^k}{C}$, maximal gap $\sim \frac{(\log x)^{k+1}}{C}$.

Polynomial gaps

Bunyakovsky (1857) Conj: Infinitely many primes $p = n^2 + 1$.

Bateman-Horn Conj: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \sim C\frac{x}{\log x}$.

Cramér type heuristic: max. gap $\sim \frac{(\log x)^2}{C}$.

Sieve methods: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \ll \frac{x}{\log x}$.

Polynomial gaps

Bunyakovsky (1857) Conj: Infinitely many primes $p = n^2 + 1$.

Bateman-Horn Conj: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \sim C \frac{x}{\log x}$.

Cramér type heuristic: max. gap $\sim \frac{(\log x)^2}{C}$.

Sieve methods: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \ll \frac{x}{\log x}$.

Question: Can one *prove* that large strings of consecutive composite values of $n^2 + 1$ exist? i.e., strings longer than $O(\log x)$ below x.

Problem: Methods for prime gaps $G(x)$ do not work!

Polynomial gaps

Bunyakovsky (1857) Conj: Infinitely many primes $p = n^2 + 1$.

Bateman-Horn Conj: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \sim C\frac{x}{\log x}$.

Cramér type heuristic: max. gap $\sim \frac{(\log x)^2}{C}$.

Sieve methods: $\#\{n \leqslant x : n^2 + 1 \text{ prime}\} \ll \frac{x}{\log x}$.

Question: Can one *prove* that large strings of consecutive composite values of $n^2 + 1$ exist? i.e., strings longer than $O(\log x)$ below x.

Problem: Methods for prime gaps $G(x)$ do not work!

Theorem (Ford-Konyagin-Maynard-Pomerance-Tao, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then there is a string of $\gg (\log x)(\log\log x)^c$ integers $n \leqslant x$ for which $f(n)$ is composite. Here c depends only on f.

Proving large prime gaps: Jacobsthal's function

$\mathcal{S}_x = \{n \in \mathbb{Z} : (n, Q_x) = 1\}, \quad Q_x = \prod_{p \leqslant x} p.$
(i.e., sieve of Eratosthenes using primes $p \leqslant x$)

Main goal: Find $J(x)$, the largest gap in $\mathcal{S}_x$; long string of consecutive integers all with a small prime factor.

Cor: $G(2Q_x) \geqslant J(x)$; essentially $G(X) \gtrsim J(\log X)$.

Proving large prime gaps: Jacobsthal's function

$\mathcal{S}_x = \{n \in \mathbb{Z} : (n, Q_x) = 1\}, \quad Q_x = \prod_{p \leqslant x} p.$
(i.e., sieve of Eratosthenes using primes $p \leqslant x$)

Main goal: Find $J(x)$, the largest gap in $\mathcal{S}_x$; long string of consecutive integers all with a small prime factor.

Cor: $G(2Q_x) \geqslant J(x)$; essentially $G(X) \gtrsim J(\log X)$.

Bounding $J(x)$:

- Average gap $Q_x/\phi(Q_x) \sim e^\gamma \log x$;
- (trivial) $J(x) \geqslant x - 2$ since $[2, x] \cap \mathcal{S}_x = \emptyset$;
- (FGKMT, 2018). $J(x) \gg x(\log x)\frac{\log_3 x}{\log_2 x}$.
- (Iwaniec, 1978). $J(x) \ll x^2(\log x)^2$.

Conjecture (Maier-Pomerance, 1990). $J(x) = x(\log x)^{2+o(1)}$.

Random Cramér type model: $J(x) \sim \frac{Q_x \log Q_x}{\phi(Q_x)} \sim e^\gamma x \log x$.

Least prime in an arithmetic progression

Let $p(k,l) = \min\{p \equiv l \pmod{k}, \text{ prime}\}$, $M(k) = \max\limits_{(l,k)=1} p(k,l)$.

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 $(\log k)^{c/\log_3 k}$ prime factors.

Least prime in an arithmetic progression, II

Pomerance: $M(k) \gg \phi(k) \log k \dfrac{\log_2 k \log_4 k}{(\log_3 k)^2}$ 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 (Junxian Li, Kyle Pratt and George Shakan, 2018)

Inequality (2) holds for almost all k; in fact, 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"

Conjecture (Li-Pratt-Shakan, 2018)

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

Heuristic: coupon collectors problem.

Least prime in an AP: data

Conjecture (Li-Pratt-Shakan, 2018)

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

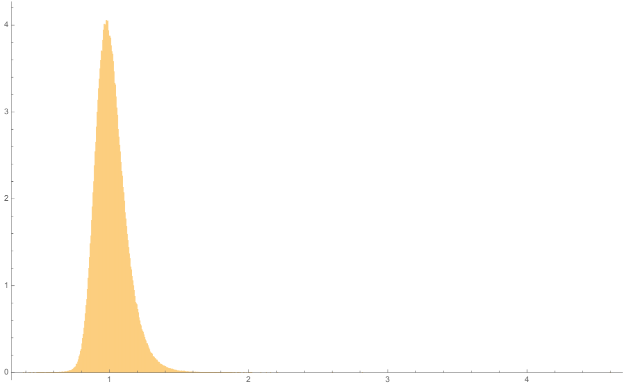


Figure: Histogram for $\frac{M(k)}{\phi(k)\log(\phi(k))\log k}$ for $k \leqslant 10^6$

Covering the gap

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

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

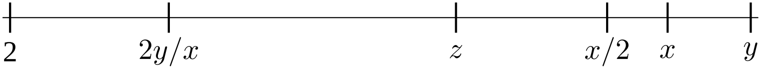
Proof: If $[n, n+y]$ is a gap in S_x, $y = J(x)$, define a_p for $p \leqslant x$ by

$$(-n \mod Q_x) = \bigcap_{p \leqslant x} (a_p \mod p).$$

Goal: succeeed with y a bit larger than x.

Finding large gaps in $\mathcal{S}_x$

Need $\bigcap_{p\leqslant x}(a_p \mod p) \supseteq [0, y]$, with y a bit larger than x.



$y = cx\frac{\log x \log_3 x}{(\log_2 x)^2}$, $z = x^{c\frac{\log_3 x}{\log_2 x}}$ Want $\{a_p \bmod p : p \leqslant x\} \supseteq [0, y]$

Classical 3-stage-process (Westzynthius-Erdős-Rankin)

1. (Key!!) Take $a_p = 0$ for $p \in (z, \frac{x}{2}] \cap [2, \frac{2y}{x}]$. Left uncovered: z-smooth numbers (few for appropriate z) and primes;
   $\sim \frac{y}{\log y}$ numbers uncovered.

   A typical choice of a_p leaves $\sim y\frac{\log z}{\log y}$ uncovered numbers

2. Greedy choice for a_p, $p \in (2y/x, z]$

3. (*trivial*) for $p \in (\frac{x}{2}, x]$, choose a_p to cover one uncovered element from step 2. Success if $\leqslant \pi(x) - \pi(x/2) \sim \frac{x}{2\log x}$ such elements.

New bounds on $J(x)$ [FKMPT, 2018]

$y = cx\frac{\log x \log_3 x}{\log_2 x}$, $z = x^{c\frac{\log_3 x}{\log_2 x}}$ 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; some AP modulo p has many primes in $[0, y]$.
4. (trivial) Use single a_p for each $x/2 < p \leqslant x$ to cover each remaining uncovered element.

Tools: Maynard sieve, efficient hypergraph covering

Analog of $J(x)$ for polynomials

$\mathcal{S}_f(x) = \{n \in \mathbb{Z} : (f(n), Q_x) = 1\}, \quad Q_x = \prod_{p \leqslant x} p.$

Gaps: Let $J_f(x)$ be the largest gap in $\mathcal{S}_f(x)$.

Analog of $J(x)$ for polynomials

$\mathcal{S}_f(x) = \{n \in \mathbb{Z} : (f(n), Q_x) = 1\}, \quad Q_x = \prod_{p \leqslant x} p.$

Gaps: Let $J_f(x)$ be the largest gap in $\mathcal{S}_f(x)$.

Covering problem: Let $I_p = \{n \mod p : f(n) \equiv 0 \pmod p\}$.
$J_f(x)$ is the largest y so that $[0, y]$ is covered by

$$\{b_p + \nu_p \mod p : p \leqslant x, \nu_p \in I_p\}$$

for some residues $b_p \mod p$.

Analog of $J(x)$ for polynomials

$\mathcal{S}_f(x) = \{n \in \mathbb{Z} : (f(n), Q_x) = 1\}, \quad Q_x = \prod_{p \leqslant x} p.$

Gaps: Let $J_f(x)$ be the largest gap in $\mathcal{S}_f(x)$.

Covering problem: Let $I_p = \{n \mod p : f(n) \equiv 0 \pmod p\}$.
$J_f(x)$ is the largest y so that $[0, y]$ is covered by

$$\{b_p + \nu_p \mod p : p \leqslant x, \nu_p \in I_p\}$$

for some residues $b_p \mod p$.

Difficulty: For a set p of positive density, $I_p = \emptyset$ (unused primes).
For $f(n) = n^2 + 1$, $I_p = \emptyset$ for $p \equiv 3 \pmod 4$.

This means that Step 1 in the usual method for large prime gaps (the smooth number estimate) cannot be used. Without it, the other steps give only the trivial bound $J_f(x) \gg x$.

New estimate for $J_f(x)$

Theorem (FKMPT, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then $J_f(x) \gg x(\log x)^c$, where c depends on f.

New estimate for $J_f(x)$

Theorem (FKMPT, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then $J_f(x) \gg x(\log x)^c$, where c depends on f.

Corollary (FKMPT)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a non-constant polynomial. Then $\exists\, G_f \geqslant 2$ such that for any $k \geqslant G_f$ there are infinitely many integers $n \geqslant 0$ so that none of $f(n+1), \ldots, f(n+k)$ is coprime to all the others.

Previously, this was known only for degree $\leqslant 3$.

New estimate for $J_f(x)$

Theorem (FKMPT, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then $J_f(x) \gg x(\log x)^c$, where c depends on f.

Corollary (FKMPT)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a non-constant polynomial. Then $\exists\, G_f \geqslant 2$ such that for any $k \geqslant G_f$ there are infinitely many integers $n \geqslant 0$ so that none of $f(n+1), \ldots, f(n+k)$ is coprime to all the others.

Previously, this was known only for degree $\leqslant 3$.

Proof of Corollary. WLOG $f \in \mathbb{Z}[x]$, irreducible. If $x \in \mathbb{N}$ is large, then $J_f(x) \geqslant 2x+1$. Let $k = 2x$ or $k = 2x+1$. Then $\mathbb{N}$ has infinitely many strings of k consecutive numbers, each having $p|f(n)$ for some $p \leqslant x$. But $p \leqslant k/2$, so $p|f(n \pm p)$ also, and one of $n \pm p$ is in the same interval.

Conjectures for $J_f(x)$

Theorem (FKMPT, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then $J_f(x) \gg x(\log x)^c$, where c depends on f.

Conjecture: $J_f(x) = x(\log x)^{1+o(1)}$.
(based on considering $S_f(x)$ as a random subset of $[1, Q_x]$)

Conjectures for $J_f(x)$

Theorem (FKMPT, 2018+)

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a monic,irreducible polynomial with no fixed prime factor. Then $J_f(x) \gg x(\log x)^c$, where c depends on f.

Conjecture: $J_f(x) = x(\log x)^{1+o(1)}$.
(based on considering $S_f(x)$ as a random subset of $[1, Q_x]$)

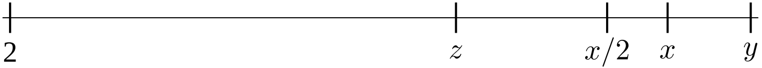
Recall:

Conjecture (Maier-Pomerance, 1990). $J(x) = x(\log x)^{2+o(1)}$.

Why the difference? The smooth number bound gives an "arithmetic boost" to $J(x)$, but not to $J_f(x)$.

Method for showing that $J_f(x)$ is large, I

Main input: $|I_p|$ is 1 on average (Prime Ideal Theorem).



$$y = x(\log x)^c,\ z = x/(\log x)^{1/100}$$

Want $\{b_p + \nu_p \bmod p : p \leqslant x, \nu_p \in I_p\} \supseteq [0, y]$

1. (Random) Pick b_p at random, $p \leqslant z$ (uniformly, independently)
2. (Random-Greedy) For $z < p \leqslant \frac{x}{2}$, choose b_p at random, but only from "rich" residue classes (those covering many uncovered numbers from Step 1. Dependent on Step 1, non-uniform.
3. (Trivial) Same as prime case. Use $b_p \mod p$ for $\frac{x}{2} < p \leqslant x$, $|I_p| \geqslant 1$ to cover anything left over ($\gg \frac{x}{\log x}$ such primes).

Method for showing that $J_f(x)$ is large, II

$$y = x(\log x)^c, \qquad z = \frac{x}{(\log x)^{1/100}}$$

1. (Random) Pick b_p at random, $p \leqslant z$ (uniformly, independently)
2. (Random-Greedy) For $z < p \leqslant \frac{x}{2}$, take b_p from rich classes

Heuristic for Step 2:

- For fixed $q \in (z, x/2]$, let

$$S_1(r,q) = [0,y] \cap (r \mod q) \setminus \bigcup_{p \leqslant (y/q)^{100}} \bigcup_{\nu_p \in I_p} (b_p + \nu_p \mod p),$$

$$S_2(r,q) = [0,y] \cap (r \mod q) \setminus \bigcup_{p \leqslant z} \bigcup_{\nu_p \in I_p} (b_p + \nu_p \mod p).$$

  There are many r for which $S_2(r,q) = S_1(r,q)$;
  (“rich” residue classes.)

- Sieving by primes $< (y/q)^{100}$ always leaves $\asymp \frac{y}{q \log(y/q)}$ elements.

Open Problems

I. Select a residue $a_p \in \mathbb{Z}/p\mathbb{Z}$ for each $p \leqslant x$, let

$$\mathcal{S} = [0, x] \setminus \bigcup_{p \leqslant x} (a_p \mod p).$$

$S = \emptyset$ possible: $a_2 = 1$, $a_p = 0$ $(3 \leqslant p \leqslant \frac{x}{2})$, a_p for $\frac{x}{2} < p \leqslant x$ cover $\{1, 2, 2^2, \ldots\}$

Problem: What is the largest possible $|\mathcal{S}|$?

- A random choice yields $|S| \sim e^{-\gamma} \frac{x}{\log x}$.
- Any choice leaves $|S| \ll \frac{x}{\log x}$ (sieve).

Open Problems

I. Select a residue $a_p \in \mathbb{Z}/p\mathbb{Z}$ for each $p \leqslant x$, let

$$\mathcal{S} = [0, x] \setminus \bigcup_{p \leqslant x} (a_p \mod p).$$

$S = \emptyset$ possible: $a_2 = 1$, $a_p = 0$ $(3 \leqslant p \leqslant \frac{x}{2})$, a_p for $\frac{x}{2} < p \leqslant x$ cover $\{1, 2, 2^2, \ldots\}$

Problem: What is the largest possible $|\mathcal{S}|$?

- A random choice yields $|S| \sim e^{-\gamma} \frac{x}{\log x}$.
- Any choice leaves $|S| \ll \frac{x}{\log x}$ (sieve).

II. For each prime $p \leqslant \sqrt{x}$, choose a residue $a_p \mod p$, and let

$$\mathcal{S} = [0, x] \setminus \bigcup_{p \leqslant \sqrt{x}} (a_p \mod p).$$

I. When $a_p = 0$ for all p, $|S| \sim x/\log x$.
II. A random choice yields $|S| \sim x(2e^{-\gamma}/\log x)$.
Question. Are these the extreme cases?