James Maynard: (Oxford)

Large gaps between primes in subsets

w/ K. Ford, S. Konyagin, C. Pomerance, T. Tao

Qn: Given $A \subseteq \mathbb{Z}_{>0}$, how ~~large~~ can gaps between primes in A be?

$\boxed{\text{Qn. of Pomerance}}$

Ex: $A = \mathbb{N}$, $\{n^2+1 \mid n \in \mathbb{Z}\}$, $\{f(n) \mid n \in \mathbb{Z}\}$ any $f \in \mathbb{Z}[x]$ (irred.)

$\{p+2 \mid p \text{ prime}\}$

Twin prime Conj. + qn. of Heath-Brown

• Also converse question: How many consecutive composite values in A?

Let $\{\text{primes in } A\} = \{p_1, p_2, \ldots\}$

How large is $G(X) := \sup_{p_j \le X} (p_{j+1} - p_j)$?

Sieve methods give upper bound

$$\#\{p = f(n) \le X\} \ll_f \frac{X^{1/d}}{\log X} \qquad d = \deg f$$

$\Longrightarrow$ Strings of $\gg \log X$ consecutive composite vals. of f

• Probabilistic guess: $(\log X)^2$

Realistic goal: Exceed $\log X$ by any amount

Thm (FKMPT) Given $f \in \mathbb{Z}[X]$, there exists

$$\gg_f \log \underline{X} \cdot (\log\log \underline{X})^{\varepsilon_f} \text{ consec. composite vals. in } f(1), \dots, f(\underline{X}).$$

with $\varepsilon_f > 0$

Notation: Small $\underline{X}$ will be the "small" primes that are used for sieving

Main technique: Use CRT to turn into sieving problem

Lemma: If $\exists$ residue classes $a_p \pmod p$ $\forall\ p \le X$

s.t. $n \in [1, y] \Rightarrow \exists\ p \le X$ s.t. $f(a_p + n) \equiv 0 \pmod p$,

then $\exists$ y consecutive composite vals. f with

$$n \ll \prod_{p \le X} p$$

Pf: CRT. If $u \equiv a_p \pmod p$, then $f(u+j)$ composite $\forall\ j \le y$. ∎

Goal is to choose a_p as "efficiently" as possible so that $a_2 \bmod 2$, $a_3 \bmod 3, \dots$ cover all of ~~$f(1), \dots, f(y)$~~ $1, 2, \dots, y$

Ex: For $f(n) = n^2+1$, cross out a_2 (2)

$a_5 \pm 2 \pmod 5$

$\vdots$

$a_p \pm \sqrt{-1} \pmod p$ for $p \equiv 1$ (4)

Trivial attempt: Pick a_p greedily for each $p = 2, 3, \dots$

- Small primes $p \le X/10$ Remove $\frac{n_p}{p}$, $n_p = \#$ roots $f(x) \bmod p$
- Large $\frac{X}{10} < p \le X$ Remove at least 1 a_p $\forall\ p$

Leaves $y \prod_{p \le \frac{x}{10}} \left(1 - \frac{n_p}{p}\right)$ after small primes

$\sim y/\log x$, so if $y = C_f x$, There are enough large primes remove the rest.

<u>Erdös-Rankin</u> Choose $a_p = 0$ for $z \le p \le \frac{x}{10}$ ("medium-sized")

This leaves just z-smooth #'s and $n = k \cdot p$

z-smooth, $p \ge x/10$

w/ $k \le \frac{10y}{x}$

Total survivors $\sim \dfrac{\log \frac{y}{x} \cdot y}{\log x}$ survivors

$\left(\text{instead of } \dfrac{\log z \cdot y}{\log x}\right)$

Now choose greedily.

- For $f(n) = n^2 + 1$, choosing $a_p = 0$ for medium p leaves $n \le y$ s.t. $f(n)$ has no prime factors in $[z, x/10]$.

  But since $n^2 + 1 = \text{prime} \cdot z\text{-smooth}$ anyway, this is not an improvement.

- Consider $f(n) = n$ and modify Erdös-Rankin.

$$a_p = \begin{cases} 0 & p \le z^{10} & 1) \\ \text{i.i.d. random} & z^{10} \le p \le \cancel{z} \; x/z & 2) \\ \text{Conditional order} & x/z \le p \le x/10 & 3) \\ \text{greedily} & x/10 \le p \le x & 4) \end{cases}$$

After 1), 2), left w/ random set of integers in $[1,y]$ 80

size $\sim y \prod_{p \le x/z} (1 - 1/p)$ w/ high prob.

Consider residue class $n \bmod q$:

Initially y/q elements.

After 1), $\sim \frac{y}{q \log y/q}$ elts.

After 2), typically $\frac{y}{q \log y/q} \prod_{z^{10} \le p \le x/z} (1 - 1/p) < 1$ already done!

Occasionally, many more

w/ positive prob., still $\frac{y}{q \log y/q}$

$$\left(\prod_{2)} 1 - 1/p \right)^{\frac{y}{q \log y/q}} \sim (\log x)^{-y/q}$$

This might still be sizable compared to y, which is why 3) can be used to repair.

In particular, if many residues mod q remain, then step 3) is more likely to contribute & eliminate more

Remark: "2) is to (on average) cancel out undetected structure"

Idea (for 3)): Independently for each $\frac{x}{z} \le p \le x/10$, choose uniformly amongst "good" residue classes.

$$P(n \text{ survives } 3)) = \prod_{x/z \le q \le x/10} (1 - P(n \text{ sieved by } q))$$

$$\sim \exp\left(- \sum_{x/z \le q \le x/10} P(n \text{ sieved by } q) \right)$$

So if $\mathbb{E}\#\{n \text{ sieved out}\} \geq t$, $\forall n$ after 1), 2), $\boxed{81}$

then $\mathbb{E}\#\{n \text{ surviving } 1), 2), 3)\} \leq \frac{y}{\log x} e^{-t}$

This is good for $y \simeq x e^{t}$

- Choosing uniformly $\Rightarrow$ even sieving, just need average case behavior

Guess $\mathbb{E}\#\{n \text{ sieved in } 3)\} \simeq \sum_{x/z \leq q \leq x/10} \frac{y}{q \log y/q} \Big/ \, y/\log x$

Average survivors mod q after 1)

Total Average after 2)

$$\simeq \sum_{\frac{x}{Z} \leq Q = 2^j \leq x/10} \frac{1}{\log y/Q} \simeq \log\log \cancel{Z}$$

if y/x small vs. Z.

Dyadic intervals are needed for technical reasons

<u>Remark</u>: In the end, thm roughly has $x \sim \log X$