Primes in arithmetic progressions and bounded gaps

Julia Stadlmann

Stanford University

June 2025

Bounded gaps between primes

Given h, do there exist ∞-many pairs of primes $(p, p+h)$?

Weaker: How big is $H_1 = \liminf_{n\to\infty}(p_{n+1} - p_n)$?

- Goldston, Pintz, Yıldırım (2005):
  - $\liminf_{n\to\infty} \frac{(p_{n+1}-p_n)}{\log p_n} = 0$
  - Conditional on improving Bombieri-Vinogradov theorem:

$$H_1 = \liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$$

- Zhang (2013): $H_1 \leq 70000000$
- Polymath (2014): $H_1 \leq 246$

Bounded gaps between primes

How big is $H_m = \liminf_{n\to\infty}(p_{n+m} - p_n)$?

Author	Bound
Maynard (2013), Tao	$H_m \ll m^3 \exp(4m)$
Polymath (2014)	$H_m \ll \exp(3.822m)$
Baker, Irving (2017)	$H_m \ll \exp(3.815m)$
S. (2023)	$H_m \ll \exp(3.8075m)$

What about gaps between 3 primes? Conjecture: $H_2 \le 6$.

Author	Bound
Polymath (2014)	$H_2 \le 398130$
S. (2023)	$H_2 \le 396506$

Results on bounded gaps are obtained via the GPY method.

Key ingredient: Various equidistribution estimates for primes in arithmetic progressions.

Primes in APs

How many primes are contained in the arithmetic progression

$$(a, a+q, a+2q, \dots)?$$

If $(a, q) = 1$, then

$$\#\{p < x : p \equiv a \bmod q\} \sim \frac{\#\{p < x\}}{\phi(q)} \sim \frac{x}{\phi(q)\log(x)}.$$

The primes are **equidistributed** in APs with modulus q.

Primes in APs (Unrestricted moduli)

Bombieri-Vinogradov theorem (1965):

$$\sum_{q \leq \frac{x^{1/2}}{\log(x)^{B(A)}}} \max_{(a,q)=1} \left| \sum_{\substack{p \leq x \\ p \equiv a(q)}} 1 - \frac{1}{\phi(q)} \sum_{p \leq x} 1 \right| \ll_A \frac{x}{\log(x)^A}.$$

"The primes have exponent of distribution $\frac{1}{2}$."

Elliott–Halberstam conjecture:

$$\sum_{q \leq x^{1-\varepsilon}} \max_{(a,q)=1} \left| \sum_{\substack{p \leq x \\ p \equiv a(q)}} 1 - \frac{1}{\phi(q)} \sum_{p \leq x} 1 \right| \ll_{A,\varepsilon} \frac{x}{\log(x)^A}.$$

Primes in APs (Smooth moduli)

Statement: Let $P(z) = \prod_{p<z} p$, $\delta > 0$ and $\omega > 0$. Then

$$\sum_{\substack{q \leq x^{1/2+2\omega} \\ q | P(x^\delta),\ (q,a)=1}} \left| \sum_{\substack{p \leq x \\ p \equiv a (q)}} 1 - \frac{1}{\phi(q)} \sum_{p \leq x} 1 \right| \ll_{A,\omega,\delta} \frac{x}{\log(x)^A}.$$

- Polymath: True for $\frac{600\omega}{7} + \frac{180\delta}{7} < 1$.
- S.: Also true for $80\omega + \frac{80\delta}{3} < 1$.

In particular, letting $\delta \to 0$, we have:

Zhang (2013): Exp. of dist. $\theta = \frac{1}{2} + \frac{1}{584}$ to smooth moduli.

Polymath (2014): Exp. of dist. $\theta = \frac{1}{2} + \frac{7}{300}$ to smooth moduli.

S. (2023): Exp. of dist. $\theta = \frac{1}{2} + \frac{1}{40}$ to smooth moduli.

Admissible tuples

($\star$): Consider a k-tuple $(h_1, \ldots, h_k)$. Are there infinitely many n so that all entries of $(n + h_1, \ldots, n + h_k)$ are prime?

- The k-tuple $(h_1, \ldots, h_k)$ is admissible if $(h_1, \ldots, h_k)$ avoids at least one residue class mod p for each prime p.
- Hardy-Littlewood conjecture:
  If $(h_1, \ldots, h_k)$ is an admissible k-tuple, then ($\star$) is true.
- Weaker: Can we show that at least $m + 1$ entries of $(n + h_1, \ldots, n + h_k)$ are prime infinitely often?

GPY method (Goldston-Pintz-Yıldırım)

- $(h_1, \dots, h_k)$ admissible k-tuple
  $\nu : \mathbb{N} \to [0, \infty)$ weight function
  $$N(x) = \sum_{x \le n \le 2x} \nu(n) \left(\sum_{i=1}^{k} 1_{\mathbb{P}}(n + h_i) - m \right)$$
- $N(x) > 0 \implies \exists n \in [x, 2x]$ so that $n + h_i$ is prime for at least $m + 1$ choices of $i \in \{1, \dots, k\}$
- What is a good choice for the weight function ν?

Maynard–Tao sieve weights

- Goal: $\nu(n)$ should only be large when $n+h_i$ is prime or a product of few primes.
- "Smoothed out indicator function for primes and products of few primes":

$$\lambda_F(n) = \sum_{d|n} \mu(d) F(\log_x(d)).$$

- Choose $J \in \mathbb{N}$, $c_j \in \mathbb{R}$, $F_{j,i} : [0,\infty) \to \mathbb{R}$ smooth.

$$\boxed{\nu(n) = \Big(\sum_{j=1}^{J} c_j \lambda_{F_{j,1}}(n+h_1) \dots \lambda_{F_{j,k}}(n+h_k)\Big)^2.}$$

The role of equidistribution estimates

We want to estimate $\sum_{x\le n\le 2x} \nu(n) 1_{\mathbb{P}}(n+h_{i_0})$.

$$\boxed{\begin{aligned}&\nu(n) = \Big(\sum_{j=1}^{J} c_j \lambda_{F_{j,1}}(n+h_1)\dots\lambda_{F_{j,k}}(n+h_k)\Big)^2\\ &\text{with}\quad \lambda_F(n) = \sum_{d\mid n}\mu(d)F(\log_x(d)).\end{aligned}}$$

Substituting the definition of $\nu(n)$ and rearranging, we get:

$$\sum_{\substack{d_1,\dots,d_k\\ d_1',\dots,d_k'}} \Big\{\sum_{j,j'} c_j c_{j'} \prod_{i\ne i_0} \mu(d_i)\mu(d_i')F_{j,i}(\log_x(d_i))F_{j',i}(\log_x(d_i'))\Big\} \sum_{\substack{m\sim x\\ m\equiv h_{i_0}-h_i\,(d_i)\\ m\equiv h_{i_0}-h_i\,(d_i')}} 1_{\mathbb{P}}(m)$$

The role of equidistribution estimates

- We must estimate the following expression:

$$\sum_{\substack{d_1,\dots,d_k\\ d_1',\dots,d_k'}} \left\{ \sum_{j,j'} c_j c_{j'} \prod_{i \neq i_0} \mu(d_i)\mu(d_i') F_{j,i}(\log_x(d_i)) F_{j',i}(\log_x(d_i')) \right\} \sum_{\substack{m \sim x\\ m \equiv h_{i_0} - h_i (d_i)\\ m \equiv h_{i_0} - h_i (d_i')}} 1_{\mathbb{P}}(m)$$

Recall: $\sum_{m \asymp x, m \equiv a(d_1 \dots d_k')} 1_{\mathbb{P}}(m) \approx \frac{x}{\phi(d_1 \dots d_k') \log(x)}$.

- Asymptotic estimates are available when:

$$\sum_{\substack{d_1,\dots,d_k,d_1',\dots,d_k'\\ \log_x(d_i) \in \mathrm{supp}(F_{j,i})\\ \log_x(d_i') \in \mathrm{supp}(F_{j',i})}} \left| \sum_{\substack{p \leq x\\ p \equiv a(d_1 \dots d_k')}} 1 - \frac{1}{\phi(d_1 \dots d_k')} \sum_{p \leq x} 1 \right| \ll_A \frac{x}{\log(x)^A}.$$

Summary (Steps so far)

- Goal: Bound $H_m = \liminf_{n\to\infty}(p_{n+m} - p_n)$.
- WTS: If $(h_1, \dots, h_k)$ is admissible, $(n + h_1, \dots, n + h_k)$ has $m + 1$ prime entries infinitely often.
- GPY method: Suffices to find weight function $\nu(n)$ with

$$\frac{\sum_{i=1}^{k} \sum_{x\le n\le 2x} \nu(n) 1_{\mathbb{P}}(n + h_i)}{\sum_{x\le n\le 2x} \nu(n)} > m. \qquad (\star)$$

- Maynard-Tao sieve weights + Equidistribution estimates:
  $\longrightarrow$ Asymptotics for $(\star)$ (depending on $F_{j,i}$ which define ν)
  $\longrightarrow$ Optimization problem.

Optimization problem

GPY method succeeds if we find $F : [0,\infty)^k \to \mathbb{R}$ with:

$$\frac{\sum_{i=1}^{k} \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty F(t_1,\dots,t_k) \mathrm{d}t_i\right)^2 \prod_{j\neq i} \mathrm{d}t_j}{\int_0^\infty \cdots \int_0^\infty F(t_1,\dots,t_k)^2 \mathrm{d}t_1 \dots \mathrm{d}t_k} > m$$

where the support of F is restricted as follows:

$$\sum_{\substack{d_1,\dots,d_k,d'_1,\dots,d'_k \\ (\log_x(d_1),\dots,\log_x(d_k))\in\mathrm{supp}(F) \\ (\log_x(d'_1),\dots,\log_x(d'_k))\in\mathrm{supp}(F)}} \left| \sum_{\substack{p\leq x \\ p\equiv a(d_1\dots d'_k)}} 1 - \frac{1}{\phi(d_1\dots d'_k)} \sum_{p\leq x} 1 \right| \ll_A \frac{x}{\log(x)^A}.$$

Suitable equidistribution estimates

Optimize over F with: Equidististribution estimates available for moduli $d_1 \dots d_k d'_1 \dots d'_k$ with $(\log_x(d_1), \dots, \log_x(d_k)) \in \operatorname{supp}(F)$ and $(\log_x(d'_1), \dots, \log_x(d'_k)) \in \operatorname{supp}(F)$.

- (1.) $\operatorname{supp}(F) \subseteq \{(t_1, \dots, t_k) \in [0, \infty)^k : \sum_{i=1}^k t_i \leq \frac{1}{4} - \frac{\varepsilon}{2}\}$
  $\implies \sum_{i=1}^k \log_x(d_i) < \frac{1}{4} - \frac{\varepsilon}{2}, \sum_{i=1}^k \log_x(d'_i) < \frac{1}{4} - \frac{\varepsilon}{2}$
  $\implies d_1 \dots d_k d'_1 \dots d'_k \leq x^{1/2-\varepsilon}$
  The Bombieri-Vinogradov theorem is applicable.
- (2.) $\operatorname{supp}(F) \subseteq \{(t_1, \dots, t_k) \in [0, \delta)^k : \sum_{i=1}^k t_i \leq \frac{1}{4} + \omega - \frac{\varepsilon}{2}\}$
  $\implies \sum_{i=1}^k \log_x(d_i), \sum_{i=1}^k \log_x(d'_i) < \frac{1}{4} - \frac{\varepsilon}{2}, \log_x(d_i) < \delta$
  $\implies d_1 \dots d_k d'_1 \dots d'_k \leq x^{1/2+2\omega-\varepsilon}$ and $d_i < x^{\delta}$
  Can use results for primes in APs to smooth moduli.

Results for large m

If the primes have exponent of dist. $\frac{1}{2} + 2\omega$ to smooth moduli:
Then $\frac{\sum_{i=1}^{k} \sum_{x \le n \le 2x} \nu(n) 1_{\mathbb{P}}(n+h_i)}{\sum_{x \le n \le 2x} \nu(n)} > \left(\frac{1}{4} + \omega\right) \log(k) - O_{\omega,\delta}(1)$.

- GPY method works if $k \gg_{\omega,\delta} \exp\left(m/(1/4+\omega)\right)$
- Length of shortest admissible k-tuple $\ll k \log(k)$
- $H_m = \liminf_{n\to\infty}(p_{n+m} - p_n) \ll \exp\left(\frac{(1+\epsilon)m}{(1/4+\omega)}\right)$
- Exponent of distribution $\frac{1}{2} + \frac{1}{40} \implies H_m \ll \exp(3.8075m)$

Result for $H_2 = \liminf_{n\to\infty}(p_{n+2} - p_n)$:

- $k = 35265$ and $(\omega, \delta) \approx (0.006, 0.021) \longrightarrow H_2 \le 396504$.

Considerations for H_1

- Polymath: $k = 50$ and Bombieri-Vinogradov $\longrightarrow H_1 \leq 246$
- $A = \{(t_1, \ldots, t_{50}) \in [0, \infty)^{50} : \sum_{i=1}^{50} t_i \leq \frac{1}{4}\}$
  $B(\omega, \delta) = \{(t_1, \ldots, t_{50}) \in [0, \delta)^{50} : \sum_{i=1}^{50} t_i \leq \frac{1}{4} + \omega\}$

Support of F	Area	Optim. with $\mathcal{S}_7$
A	2.59373×10^{-95}	0.9697
$B(\frac{7}{1200}, 0.02)$	3.00772×10^{-95}	0.7344
$A \cup B(\frac{7}{1200}, 0.02)$	4.54685×10^{-95}	0.9696
$A \cup B(\frac{7}{2160}, 0.028)$	4.46979×10^{-95}	0.9697
$A \cup B(\frac{1}{1680}, 0.036)$	2.92734×10^{-95}	0.9698

$\mathcal{S}_7 = \{(\frac{1}{4} - \sum_{i=1}^{50} t_i)^a p(t_1^2, \ldots, t_{50}^2) : p \text{ symmetric}, 2\deg(p) + a < 14\}$