Small gaps between primes

James Maynard

Magdalen College, Oxford

Barcelona Mathematical Days, Barcelona
November 2014
The main problem in prime number theory is to understand the distribution of the primes.

Today we want to understand the gaps between primes.
The main problem in prime number theory is to understand the distribution of the primes.

Today we want to understand the gaps between primes.

**Theorem (prime number theorem)**

\[
\#\{\text{primes } \leq x\} \approx \frac{x}{\log x}.
\]

This means that for \( p_n \leq x \), the **average** gap \( p_{n+1} - p_n \approx \log x \), so the primes get sparser.
The main problem in prime number theory is to understand the distribution of the primes.

Today we want to understand the gaps between primes.

**Theorem (prime number theorem)**

\[ \#\{\text{primes } \leq x\} \approx \frac{x}{\log x}. \]

This means that for \( p_n \leq x \), the average gap \( p_{n+1} - p_n \approx \log x \), so the primes get sparser.

**Question**

Are prime gaps always this big?
Question

Are prime gaps always this big?

- $(2, 3)$ is the only pair of primes which differ by 1. (One of $n$ and $n + 1$ is a multiple of 2 for every integer $n$).
- There are lots of pairs of primes which differ by 2: $(3, 5), (5, 7), (11, 13), \ldots, (1031, 1033), \ldots, (1000037, 1000039), \ldots, (100000007, 100000009), \ldots$
Question

Are prime gaps always this big?

- $(2, 3)$ is the only pair of primes which differ by 1. (One of $n$ and $n + 1$ is a multiple of 2 for every integer $n$).
- There are lots of pairs of primes which differ by 2: $(3, 5), (5, 7), (11, 13), \ldots, (1031, 1033), \ldots, (1000037, 1000039), \ldots, (1000000007, 1000000009), \ldots$

Conjecture (Twin prime conjecture)

There are infinitely many pairs of primes $(p, p')$ which differ by 2.

As we all know, this is one of the oldest problems in mathematics, and is very much open!
More generally, we can look for triples (or more) of primes.

- \((2, 3, 5), (2, 3, 7), (2, 5, 7), (3, 5, 7)\) are the only triples contained in an interval of length 5.  
  (At least one of \(n, n + 2, n + 4\) is a multiple of 3.)

- There are lots of triples of primes in an interval of length 6.  
  \((5, 7, 11), (11, 13, 17), \ldots, (1091, 1093, 1097), \ldots, (1000033, 1000037, 1000039), \ldots\)
More generally, we can look for triples (or more) of primes.

- \((2, 3, 5), (2, 3, 7), (2, 5, 7), (3, 5, 7)\) are the only triples contained in an interval of length 5.
  (At least one of \(n, n + 2, n + 4\) is a multiple of 3.)
- There are lots of triples of primes in an interval of length 6.
  \((5, 7, 11), (11, 13, 17), \ldots, (1091, 1093, 1097), \ldots, (1000033, 1000037, 1000039), \ldots\)
- All such triples are of the form \((n, n + 2, n + 6)\) or \((n, n + 4, n + 6)\), and we find lots of both types.
- In fact, we find lots of triples \((n, n + h_1, n + h_2)\) if one of the triple doesn’t have to be a multiple of 2 or 3.

It is natural to generalize to look for patterns \(n + h_1, \ldots, n + h_k\) of primes.
It is natural to generalize to look for patterns $n + h_1, \ldots, n + h_k$ of primes.

**Definition (admissibility)**

\( \{h_1, \ldots, h_k\} \) is **admissible** if \( \prod (n + h_i) \) has no fixed prime divisor.

**Conjecture (prime k-tuples conjecture)**

Let \( \{h_1, \ldots, h_k\} \) be admissible. Then there are infinitely many integers \( n \) such that all of \( n + h_1, \ldots, n + h_k \) are primes.
This conjecture tells us a huge amount about the ‘small scale’ structure of the primes.

These questions are difficult because they ask additive questions about multiplicative objects.

**Corollary**

Assume the prime $k$-tuples conjecture. Then

\[
\lim \inf_n (p_{n+1} - p_n) = 2.
\]

\[
\lim \inf_n (p_{n+m} - p_n) \leq (1 + o(1))m \log m.
\]

Therefore we believe that occasionally primes come clumped closely together. (Despite becoming sparser on average.)
In the RSA algorithm one wants to choose $N = pq$ which is hard to factor.

If $p - 1$ has only small prime factors, then there is a way to factor $N$ easily (Bad).

It had been suggested that one could choose $p, q$ such that $(p - 1)/2$ and $(q - 1)/2$ are prime (although this is not recommended).

If there are only 10 (say) 1024-bit primes $p$ such that $(p - 1)/2$ is prime, then this is a VERY bad idea!

A slight generalization of the prime $k$-tuples conjecture predicts there are many such primes, so perhaps you are only wasting CPU cycles.
Goldston, Pintz and Yıldırım developed the ‘GPY method’ for studying small gaps between primes unconditionally.

**Theorem (Zhang)**

\[
\lim \inf_n (p_{n+1} - p_n) \leq 70\,000\,000.
\]

**Theorem (M./Tao)**

1. \[
\lim \inf_n (p_{n+m} - p_n) \leq m^3 e^{4m+8} \text{ for all } m \in \mathbb{N}.
\]
2. \[
\lim \inf_n (p_{n+1} - p_n) \leq 600.
\]

**Theorem (Polymath 8b)**

1. \[
\lim \inf_n (p_{n+m} - p_n) \leq C e^{3.83m} \text{ for all } m \in \mathbb{N} \text{ (some constant } C).\]
2. \[
\lim \inf_n (p_{n+1} - p_n) \leq 246.
\]
Weak k-tuples

These results rely on proving a weak form of the prime $k$-tuples conjecture.

**Conjecture (DHL($k, m$))**

Let $\{h_1, \ldots, h_k\}$ be admissible. Then there are infinitely many integers $n$ such that $m$ of $n + h_1, \ldots, n + h_k$ are primes.
Weak k-tuples

These results rely on proving a weak form of the prime $k$-tuples conjecture.

**Conjecture (DHL($k, m$))**

Let $\{h_1, \ldots, h_k\}$ be admissible. Then there are infinitely many integers $n$ such that $m$ of $n + h_1, \ldots, n + h_k$ are primes.

**Theorem (Zhang)**

$DHL(k, m)$ holds for $m = 2$ and $k \geq 3\,500\,000$.

**Theorem (M.)**

$DHL(k, m)$ holds for $k \geq m^2 e^{4m+6}$, and for $k \geq 105$ if $m = 2$. 

Overview

Primes in A.P.s
   Bombieri-Vinogradov theorem

Sieve Method
   Modified GPY sieve

Optimization problem
   Choice of smooth weight

Small gaps between primes
   Dense admissible sets

Combinatorial problem

Figure: Outline of steps to prove small gaps between primes
One way to view sieve methods is the study of ‘almost-primes’.

- Almost-primes have similar properties to the primes (no small prime factors, distribution in APs)
- The primes have positive density in the almost-primes (Gives upper bounds worse than expected by a constant)
- We can solve additive problems for almost-primes if we know solutions in arithmetic progressions
Look at almost-prime values of $(n + h_i)^k_{i=1}$

We can calculate the **density** of solutions when $n + h_1$ is prime

If this density is bigger than $1/k$ for each $n + h_i$, then more than 1 of the components are prime on average.

By **pigeonhole principle** we deduce that at least $m + 1$ of the components are prime infinitely often if the density greater than $m/k$. 

---

This argument depends on the precise definition of ‘almost-prime’. If we have better knowledge of primes in arithmetic progressions, then we can produce better almost-prime solutions.
The GPY sieve

1. Look at almost-prime values of \((n + h_i)^k \) for each \(i = 1\).
2. We can calculate the density of solutions when \(n + h_1\) is prime.
3. If this density is bigger than \(1/k\) for each \(n + h_i\), then more than 1 of the components are prime on average.
4. By pigeonhole principle we deduce that at least \(m + 1\) of the components are prime infinitely often if the density greater than \(m/k\).

This argument depends on the precise definition of ‘almost-prime’.

If we have better knowledge of primes in arithmetic progressions, then we can produce better almost-prime solutions.
The GPY sieve II

Question

*How do we choose the weights $w_n$ which define almost-primes?*

1 Standard choice: Gives density $\approx \frac{1}{2^k}$. Fails to prove bounded gaps.

2 GPY choice: Gives density $\approx \frac{1}{k}$. Just fails to prove bounded gaps. Zhang's equidistribution results give a density slightly bigger than $1/k$ with this method: bounded gaps!

3 New choice: Gives density as the ratio of two integrals of an auxiliary function.
Question

How do we choose the weights $w_n$ which define almost-primes?

We choose $w_n$ to mimic ‘Selberg sieve’ weights.

1. **Standard choice:** Gives density $\approx 1/2k$. Fails to prove bounded gaps.
Question

How do we choose the weights $w_n$ which define almost-primes?

We choose $w_n$ to mimic ‘Selberg sieve’ weights.

1. **Standard choice**: Gives density $\approx 1/2k$. Fails to prove bounded gaps.

2. **GPY choice**: Gives density $\approx 1/k$. Just fails to prove bounded gaps.

Zhang’s equidistribution results give a density slightly bigger than $1/k$ with this method: bounded gaps!
Question

*How do we choose the weights $w_n$ which define almost-primes?*

We choose $w_n$ to mimic ‘Selberg sieve’ weights.

1. **Standard choice:** Gives density $\approx 1/2k$. Fails to prove bounded gaps.

2. **GPY choice:** Gives density $\approx 1/k$. Just fails to prove bounded gaps.
   Zhang’s equidistribution results give a density slightly bigger than $1/k$ with this method: bounded gaps!
Question

How do we choose the weights \( w_n \) which define almost-primes?

We choose \( w_n \) to mimic ‘Selberg sieve’ weights.

1. **Standard choice**: Gives density \( \approx 1/2k \). Fails to prove bounded gaps.

2. **GPY choice**: Gives density \( \approx 1/k \). Just fails to prove bounded gaps.
   
   Zhang’s equidistribution results give a density slightly bigger than \( 1/k \) with this method: bounded gaps!

3. **New choice**: Gives density as the ratio of two integrals of an auxiliary function \( F \).
The sieve calculation gives:

**Proposition**

Let \( \{h_1, \ldots, h_k\} \) be admissible. Let

\[
M_k = \sup_F \frac{J_k(F)}{I_k(F)}.
\]

If \( M_k > 4m \) then DHL\((k, m+1)\) holds.

(i.e. there are infinitely many integers \( n \) such that at least \( m + 1 \) of the \( n + h_i \) are primes).

This has reduced our arithmetic problem (difficult) to a smooth optimization (easier).
To show small gaps we need a good lower bound for $M_k$.

**Proposition 1**

$M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.

$M_{10^5} > 4$.
Lower bounds for $M_k$

To show small gaps we need a good lower bound for $M_k$.

**Large $k$:**
- Approach problem from functional analysis viewpoint.
- Use dimensionality to construct good choice of $F$.
- This choice is essentially optimal when $k$ is large.

Proposition 1

$M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.

$M_{105} > 4$. 

James Maynard 
Small gaps between primes
Lower bounds for $M_k$

To show small gaps we need a good lower bound for $M_k$.

**Large $k$:**
- Approach problem from functional analysis viewpoint.
- Use dimensionality to construct good choice of $F$.
- This choice is essentially optimal when $k$ is large.

**Small $k$:**
- Approach problem from numerical analysis viewpoint.
- Reduce optimization to a feasible numerical calculation.
- Gives essentially optimal bounds when $k$ is small.

Proposition 1

$M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.

$M_{10^5} > 4$. 

James Maynard
Small gaps between primes
Lower bounds for $M_k$

To show small gaps we need a good lower bound for $M_k$.

**Large $k$:**
- Approach problem from functional analysis viewpoint.
- Use dimensionality to construct good choice of $F$.
- This choice is essentially optimal when $k$ is large.

**Small $k$:**
- Approach problem from numerical analysis viewpoint.
- Reduce optimization to a feasible numerical calculation.
- Gives essentially optimal bounds when $k$ is small.

**Proposition**

1. $M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.
2. $M_{105} > 4$. 

James Maynard
Small gaps between primes
Putting it all together: large $k$

### Proposition

1. $M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.

2. If $M_k > 4m$ then there are infinitely many integers $n$ such that at least $m + 1$ of the $n + h_i$ are primes.
Proposition

1. \( M_k > \log k - 2 \log \log k - 2 \) if \( k \) is large enough.
2. If \( M_k > 4m \) then there are infinitely many integers \( n \) such that at least \( m + 1 \) of the \( n + h_i \) are primes.

Finally

Lemma

There is an admissible set of size \( k \) contained in \([0, 2k \log k]\).
Putting it all together: large $k$

**Proposition**

1. $M_k > \log k - 2 \log \log k - 2$ if $k$ is large enough.
2. If $M_k > 4m$ then there are infinitely many integers $n$ such that at least $m + 1$ of the $n + h_i$ are primes.

Finally

**Lemma**

*There is an admissible set of size $k$ contained in $[0, 2k \log k]$.*

These give

**Theorem**

\[
\liminf_n (p_{n+m} - p_n) \leq Cm^3 e^{4m}.
\]
Proposition

1. $M_{105} > 4$.
2. If $M_k > 4$ then there are infinitely many integers $n$ such that at least 2 of the $n + h_i$ are primes.
Putting it together: small $k$

Proposition

1. $M_{105} > 4$.
2. If $M_k > 4$ then there are infinitely many integers $n$ such that at least $2$ of the $n + h_i$ are primes.

Lemma (Engelsma)

There is an admissible set of size $105$ contained in $[0, 600]$.
Proposition

1. $M_{105} > 4$.
2. If $M_k > 4$ then there are infinitely many integers $n$ such that at least 2 of the $n + h_i$ are primes.

Lemma (Engelsma)

There is an admissible set of size 105 contained in $[0, 600]$.

Theorem

$$\lim \inf_n (p_{n+1} - p_n) \leq 600.$$
Other applications

Observation

Since $M_k \to \infty$, this method doesn’t depend too heavily on the strength of equidistribution results.

This makes the method very flexible.

There is hope that this can have applications in many other contexts.
The prime $k$-tuples conjecture is true for a positive proportion of admissible sets.
The prime $k$-tuples conjecture is true for a positive proportion of admissible sets.

There are unusually large gaps between primes.
The prime $k$-tuples conjecture is true for a positive proportion of admissible sets.

There are unusually large gaps between primes.

Bounded gaps between many primes restricted to lie in short interval, an arithmetic progression, or be represented by a quadratic form.
The prime $k$-tuples conjecture is true for a positive proportion of admissible sets.

There are unusually large gaps between primes.

Bounded gaps between many primes restricted to lie in short interval, an arithmetic progression, or be represented by a quadratic form.

There are intervals $[x, x + (\log x)^\epsilon]$ containing $\gg \log \log x$ primes (many more than average).
The prime $k$-tuples conjecture is true for a positive proportion of admissible sets.

There are unusually large gaps between primes.

Bounded gaps between many primes restricted to lie in short interval, an arithmetic progression, or be represented by a quadratic form.

There are intervals $[x, x + (\log x)^\epsilon]$ containing $\gg \log \log x$ primes (many more than average).

We have quantitative estimates which fit well with Cramér’s random model for the primes.
The prime \( k \)-tuples conjecture is true for a positive proportion of admissible sets.

There are unusually large gaps between primes.

Bounded gaps between many primes restricted to lie in short interval, an arithmetic progression, or be represented by a quadratic form.

There are intervals \([x, x + (\log x)^\epsilon]\) containing \(\gg \log \log x\) primes (many more than average).

We have quantitative estimates which fit well with Cramér’s random model for the primes.

There are arbitrarily large sets of primes, with any pair differing in at most 2 decimal places.
Thank you for listening.
Conditional results

If we assume stronger results about primes in arithmetic progressions, then we obtain stronger results.

**Theorem**

Assume the Bombieri-Vinogradov Theorem can be extended to $q < x^{1-\varepsilon}$. Then

\[
\liminf_n (p_{n+1} - p_n) \leq 16 \quad \text{(Goldston-Pintz-Yıldırım)}
\]

\[
\liminf_n (p_{n+1} - p_n) \leq 12. \quad \text{(M.)}
\]

**Theorem (Polymath 8b)**

Assume ‘GEH’. Then we have,

\[
\liminf_n (p_{n+1} - p_n) \leq 6.
\]

There is a barrier to obtaining the twin prime conjecture with this method.
This has the amusing consequence:

**Theorem (Polymath 8b)**

Assume ‘GEH’. Then at least one of the following is true.

1. *There are infinitely many twin primes.*
2. *Every large even number is within 2 of a number which is the sum of two primes.*

Of course we expect both to be true!
Thank you for listening.