# An Introduction to Large Gaps Between Primes 2023 Fall Analytic Number Theory

Kai Zhu

Chern Class Nankai University

December 17, 2023

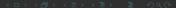
## Summary

Introduction

2 Erdös' Proof

3 Further Results

## Introduction

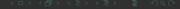


#### Introduction

The small gap between two consecutive primes is an well-known and interesting open problem, for instance, the twin prime conjecture. And the known best result about the small gap between primes is

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \le 246.$$

Hence a natural question is, how large can two consecutive primes have?



A easy and well-known result is

## Theorem (1)

For any M>0, there exists  $n\in\mathbb{N}$  such that  $p_{n+1}-p_n\geq M$ .

#### Proof.

Let m=[M]+2, here [x] denotes the largest integer that is smaller than x. Note that for any  $2 \le k \le m$ , we have m!+k is divided by k, hence  $m!+2, \cdots, m!+m$  are all composite numbers. Then there exists  $n_m=\pi(m!+2)$  such that

$$p_{n_m+1} - p_{n_m} \ge m \ge M, (1.1)$$

where  $\pi(x)$  denotes the number of primes that less than x.

From theorem (1), we know that

$$\lim_{n \to \infty} \sup(p_{n+1} - p_n) = \infty. \tag{1.2}$$

Then a natural question is that, how can we sharpen the estimate above? From Bertrand theorem, we know that

$$\frac{1}{2}(m!+2) \le p_{n_m} \le m!+2,$$

hence we take logarithm at the same time, then

$$\log(m! + 2) - \log 2 \le \log p_{n_m} \le \log(m! + 2).$$

from Stirling formula, we have

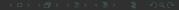
$$\log p_{n_m} = \log(m!) + O(1)$$

$$= m \log m - m + O(\log m),$$

$$\log \log p_{n_m} = \log m + O(\log \log m).$$

Then we subsititute m by  $p_{n_m}$  in (1.1) , we have

$$p_{n_m+1} - p_{n_m} > [1 + o(1)] \frac{\log p_{n_m}}{\log \log p_{n_m}}.$$



Then we substitute m by  $p_{n_m}$  in (1.1), we have

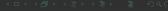
$$p_{n_m+1} - p_{n_m} > [1 + o(1)] \frac{\log p_{n_m}}{\log \log p_{n_m}}.$$

In other words, we have

#### Theorem (2)

For any  $\varepsilon > 0$ , there are infinite many  $n \in \mathbb{N}$  such that

$$p_{n+1} - p_n > (1 - \varepsilon) \frac{\log p_n}{\log \log p_n}.$$



Actually, from the prime theorem, we can deduce a stronger result. Note that for each X>0, there are  $\pi(X)$  primes in the interval [1,X]. Hence there will must exist two primes  $p_n,p_{n+1}$  such that

$$p_{n+1} - p_n \ge \frac{X}{\pi(X)} = [1 + o(1)] \log X \ge [1 + o(1)] \log p_n.$$

Thus we have

#### Theorem (3)

For any  $\varepsilon > 0$ , there are infinite many  $n \in \mathbb{N}$  such that

$$p_{n+1} - p_n > (1 - \varepsilon) \log p_n.$$

#### Historical Reuslts

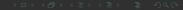
Brauer and Zeitz showed that  $1-\varepsilon$  in theorem 3 could be replaced by  $4-\varepsilon$ . Westzynthius proved that there are infinite n such that

$$p_{n+1} - p_n \ge \frac{2\log p_n \log \log \log p_n}{\log \log \log \log \log p_n},$$

and Ricci then proved that this can be improved to

$$p_{n+1} - p_n > c \log p_n \log \log \log p_n$$

for a certain constant c. Then Erdös showed that it can be further improved, which is

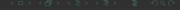


#### Erdös ' resut

#### Theorem (4)

There exist a certain positive constant  $c_1$  and infinite many  $n \in \mathbb{N}$  such that

$$p_{n+1} - p_n \ge \frac{c_1 \log p_n \log \log p_n}{(\log \log \log p_n)^2}.$$
(1.3)



## Erdös' Proof



## Reduction to a equivalent statement

We reduce our problem to the proof of the following theorem.

## Theorem (5)

For a certain positive constant c, we can find  $cp_n \log p_n/(\log \log p_n)^2$  consecutive integers so that no one of them is relatively prime to the product  $p_1p_2 \cdots p_n$ , i.e. each of these integers is divisible by at least one of the primes  $p_1, p_2, \cdots, p_n$ .

The existence of such consecutive integers is from Chinese reminder theorem. But before we use Chinese reminder theorem, we need some lemmas to find appropriate congruence equation.

#### Lemma (6)

For large T we have

$$\int_{1/T}^{1} \frac{e^{y}}{y} dy = \log T + O(1);$$

$$\int_{1}^{T} \frac{e^{y}}{y^{2}} dy = \frac{e^{T}}{T^{2}} + O\left(\frac{e^{T}}{T^{3}}\right);$$

$$\int_{1/T}^{1} \frac{e^{y}}{y^{2}} dy = T + \log T + O(1).$$

#### Lemma (7)

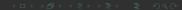
If  $N(e^u)$  is the number of positive integers not exceeding  $e^u$  which contain no prime factor greater than

$$\exp\left(\frac{u\log\log u}{a\log u}\right),\,$$

where a > 0, then

$$N(e^u) < \frac{e^u}{u^{a-1-c_2}} \tag{2.1}$$

for any fixed  $c_2 > 0$  and u large.



## The proof of lemma 7-1

Put  $x = \exp(u \log \log u / (a \log u))$  and take a number  $\eta > 0$ . Let  $k = \pi(x)$ , then

$$N(e^u) = \bigoplus_{v \le e^u} 1 < \bigoplus_{v \le e^u} \left(\frac{e^u}{v}\right)^{\eta} = e^{u\eta} \bigoplus_{v \le e^u} \frac{1}{v^{\eta}} < e^{u\eta} \bigoplus_{v=1}^{\infty} \frac{1}{v^{\eta}},$$

here we use  $\bigoplus$  denotes the summation over those positive integers v which have no prime factor exceeding x. Therfore, since

$$\bigoplus_{v=1}^{\infty} \frac{1}{v^{\eta}} = \prod_{l=1}^{k} (1 - p_l^{-\eta})^{-1}, \tag{2.2}$$

we have

$$N(e^u) < e^{u\eta} \prod_{l=1}^k (1 - p_l^{-\eta})^{-1}.$$
 (2.3)

## proof of lemma 7-II

Put

$$f(\eta) = \prod_{l=1}^{k} (1 - p_l^{-\eta})^{-1} = \exp\left(-\sum_{l=1}^{k} \log(1 - p_l^{-\eta})\right).$$

Then we have

$$\log f(\eta) = -\sum_{l=1}^{k} \log \left(1 - p_l^{-\eta}\right) = -\sum_{t=1}^{x} \log \left(1 - t^{-\eta}\right) (\pi(t) - \pi(t - 1))$$

$$= -\pi(x) \log \left(1 - x^{-\eta}\right) + \eta \int_{2}^{x} \frac{\pi(t)}{t(t^{\eta} - 1)} dt$$

$$= O\left(\frac{x^{1-\eta}}{\log x}\right) + \eta \int_{2}^{x} \frac{dt}{(t^{\eta} - 1) \log t} + O\left(\eta \int_{2}^{x} \frac{dt}{t^{\eta} \log^{2} t}\right)$$

if  $\eta > 1/2$  (for example), since  $\pi(t) = rac{t}{\log t} + O\left(rac{t}{\log^2 t}
ight).$ 

(2.4)

#### Proof of lemma 7-III

Now take  $1 - \eta = \delta = a \log u / u < 1/2$ . Hence

$$\log f(\eta) = \int_{2}^{x} \frac{\mathrm{d}t}{t^{\eta} \log t} + O(1) + O\left(\int_{2}^{x} \frac{\mathrm{d}t}{t^{\eta} \log^{2} t}\right)$$

$$= \int_{\delta \log 2}^{\delta \log x} \frac{\mathrm{e}^{y}}{y} \mathrm{d}y + O(1) + O\left(\delta \int_{\delta \log 2}^{\delta \log x} \frac{\mathrm{e}^{y}}{y^{2}} \mathrm{d}y\right)$$

$$= \frac{x^{\delta}}{\delta \log x} + \log \frac{1}{\delta} + O\left(\frac{x^{\delta}}{\delta^{2} \log^{2} x}\right)$$

by lemma 6.



#### Proof of lemma 7-IV

Therefore

$$\log f(\eta) = \log u + O\left(\frac{\log u}{\log \log u}\right). \tag{2.5}$$

Thus

$$N(e^{u}) < e^{u\eta} f(\eta)$$

$$= \exp(u - \delta u + \log f(\eta))$$

$$= \exp\left(u - (a - 1)\log u + O\left(\frac{\log u}{\log\log u}\right)\right)$$

$$< \frac{e^{u}}{u^{a-1-c_{2}}},$$

which is the required result.

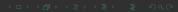
Putting  $e^u = p_n \log p_n$  and a = 5 in (2.1), we have

$$N(p_n \log p_n) = o\left(\frac{p_n}{(\log p_n)^2}\right). \tag{2.6}$$

More precisely, (2.6) shows the lemma below,

#### Lemma (8)

If  $N_0$  is the number of those integers not exceeding  $p_n \log p_n$ , each of whose greatest prime factor is less than  $p_n^{1/(20 \log \log p_n)}$ , then  $N_0 = o\left(p_n/(\log p_n)^2\right)$ .



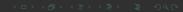
From Brauer, we have the lemma below,

#### Lemma (9)

Let m be any positive integer greater than 1, x and y any numbers such that  $1 \le x < y < m$ , and N the number of primes p less than or equal to m such that p+1 is not divisible by any of the primes P, where  $x \le P \le y$ . Then

$$N < \frac{c_3 m \log x}{\log m \log y},\tag{2.7}$$

where  $c_3$  is a constant independent of m, x and y.



We omit the proof here since it is too technical and not very helpful to the proof of our main theorem. What we need is putting

$$m = \frac{c_4 p_n \log p_n}{(\log \log p_n)^2}, \quad x = \log p_n, \quad y = p_n^{1/(20 \log \log p_n)}.$$

Then we have the lemma below,

#### Lemma (10)

We can find a constant  $c_4$  so that the number of primes p, less than  $c_4p_n/(\log\log p_n)^2$  and such that p+1 is not divisible by any prime between  $\log p_n$  and  $p_n^{1/(20\log\log p_n)}$ , is less than  $p_n/4\log p_n$ .

## Classification of the numbers less than $p_n$

We now return to theorem 5. We denote q,r,s,t the primes satisfying the inequalities

$$1 < q \le \log p_n, \quad \log p_n < r \le p_n^{1/(20 \log \log p_n)}$$
  
 $p_n^{1/(20 \log \log p_n)} < s \le \frac{1}{2} p_n, \quad \frac{1}{2} p_n < t \le p_n.$ 

We denote by  $a_1, a_2, \cdots, a_k$  the two sets of integers not greater than  $p_n \log p_n$ , namely

- In the prime numbers lying between  $\frac{1}{2}p_n$  and  $c_4p_n\log p_n/(\log\log p_n)^2$  and not congruent to -1 to any modulus r,
- 2 the integers not excedding  $p_n \log p_n$  whose prime factors are included only among the r.

Some of the a's may be t's.

Then we have the final lemma below,

### Lemma (11)

The number of the t's is greater than k the number of the a's, if  $p_n$  is large enough.

#### Proof.

From lemma 8 and 10, we have

$$k < \frac{1}{4} \frac{p_n}{\log p_n} + o\left(\frac{p_n}{(\log p_n)^2}\right). \tag{2.8}$$

The number of the t's is greater than  $\frac{1}{3}p_n/\log p_n$  for large  $p_n$ , as is evident from the prime number theorem. This proves the lemma.

#### Proof of Theorem 5-1

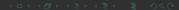
Now we begin the proof of theorem 5. We determine an integer z such that for all q,r,s,

$$0 < z < p_1 p_2 \cdots p_n,$$

and it satisfies the equation below

$$z \equiv 0 \pmod{q}$$
,  $z \equiv 1 \pmod{r}$ ,  $z \equiv 0 \pmod{s}$ ,  $z + a_i \equiv 0 \pmod{t_i}$   $i = 1, 2, \dots, k$ .

By lemma 11, the last congruence is always possible, for, as there are more t's than a's, a case such as  $z+a_1\equiv 0\pmod t$ ,  $z+a_2\equiv 0\pmod t$  cannot occur.



#### Proof of Theorem 5-II

We now show that, if l is any integer such that

$$0 < l < c_2 p_n \log p_n / (\log \log p_n)^2, \tag{2.9}$$

then no one of the integers

$$z, z+1, z+2, \cdots, z+l$$

is relatively prime to  $p_1p_2\cdots p_n$ .

Now any integer b(0 < b < l) can be replaced in one at least of the following classes:

- $b \equiv 0 \pmod{q}$ , for some q;
- $m b \equiv 1 \pmod{r}$ , for some r;
- $\mathbf{m}b \equiv 0 \pmod{s}$ , for some s;
- (b) is an  $a_i$ .

## Proof of Theorem 5-III (why b can be replaced)

For b cannot be divisible by an r and by a prime greater than  $\frac{1}{2}p_n$ , since if this were so we should have

$$b > \frac{1}{2}p_n r > \frac{1}{2}p_n \log p_n > l,$$

for sufficiently large n. Hence, if b does not satisfy (i) or (iii), b is either a product of primes r only, and so satisfies (iv), or b is not divisible by any q, r, s. In the latter case, b must be a prime, for otherwise

$$b > \left(\frac{1}{2}p_n\right)^2 > l,$$

for sufficiently large n. Since, then, b is a prime between

$$\frac{1}{2}p_n \quad \text{and} \quad \frac{c_2p_n\log p_n}{(\log\log p_n)^2},$$

b is either an  $a_i$ , or b satisfies (ii).

#### Proof of Theorem 5-IV

It is now clear that z+b is not relatively prime to  $p_1p_2\cdots p_n$ , if

$$b < \frac{c_2 p_n \log p_n}{(\log \log p_n)^2}. \quad \Box$$

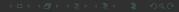
Hence also, if  $p_1, p_2, \cdots, p_n$  are the primes not excedding x, say, z+b is not relatively prime to  $p_1p_2\cdots p_n$ , if  $b < c_5x\log x/(\log\log x)^2$ , where  $c_5$  is an appropriate constant independent of x. This is clear from the first case on noticing that, by Bertrand's theorem,  $p_n \geq \frac{1}{2}x$ .

## Why thm 5 implies thm 4

We return to the main problem. Take  $x=\frac{1}{2}p_n$ . Then the product of the primes not exceeding x is less than  $\frac{1}{2}p_n$  for large  $p_n$  by the prime number theorem. By theorem 5, since now  $b<\frac{1}{2}p_n$ , we can find K consecutive integers less than  $p_n$ , where

$$K = \frac{c_5 \log p_n \log \log p_n}{(\log \log \log p_n)^2},$$

each of which is divisible by a prime less than  $\frac{1}{2} \log p_n$ . Hence there are at least  $K - \frac{1}{2} \log p_n > \frac{1}{2} K$  consecutive integers which are not primes.



## Why thm 5 implies thm 4

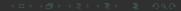
Thus we have proved that at least one of the intervals between successive primes less than  $p_n$  is always of length not less than

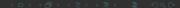
$$c\frac{\log p_n \log \log p_n}{(\log \log \log p_n)^2}$$

for large  $p_n$  and an appropriate constant c. Since this expression is an increasing function n, it follows immediately that for infinity of n,

$$p_{n+1} - p_n \ge \frac{c \log p_n \log \log p_n}{(\log \log \log p_n)^2}.$$

Hence we finish the proof of theorem 4.

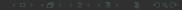




After Erdös, Rakin succeeded in showing that there are infinite many n such that

$$p_{n+1} - p_n \ge (c + o(1)) \frac{\log p_n \log \log \log \log \log \log \log p_n}{(\log \log \log p_n)^2},$$
(3.1)

with the constant c=1/3. Since this result, however, the only improvements have been in the constant c. And finally, Pintz find a better constant  $c=2\mathrm{e}^{\gamma}$  in 1997, where  $\gamma$  denotes the Euler constant.



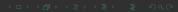
Erdös conjectured that 3.1 holds for arbitrary c>0, and he would like to offer \$5000 for this conjecture. But this conjecture is not been proved until 2014, by the joint work of K. Ford, B. Green, S. Konyagin, J. Maynard and T. Tao , they succeeded in showing that

Theorem (K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao)

We have

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{(\log p_n)(\log_2 p_n)(\log_4 p_n)(\log_3 p_n)^{-2}} = \infty,$$

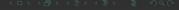
where  $\log_v$  denotes the v-fold logarithm.



Actually, Erdös had also conjectured a stronger result, for arbitrary small  $\varepsilon > 0$ , there exists infinite many n such that

$$p_{n+1} - p_n \ge (\log p_n)^{1+\varepsilon},$$

and he would like to offer \$10000 for the proof of this conjecture. But no known result about this harder conjecture.



## The End