

## 16. LONG GAPS BETWEEN PRIMES.

There are  $\sim x/\log x$  primes up to  $x$ , so the average gap between two consecutive primes of size  $x$  is around  $\log x$ . Of course we believe that gaps can be much smaller and much larger (as small as 2, and as large as  $c(\log x)^2$  for some constant  $c \geq 1$ ). In this section we shall prove that the gap between two consecutive primes can be much larger than the average:

Our goal is to show that if  $p_1 = 2 < p_2 = 3 < \dots$  is the sequence of prime numbers then

$$\limsup_{n \rightarrow \infty} \frac{p_{n+1} - p_n}{\log p_n} = \infty.$$

The only method known that succeeds in showing that there are gaps between consecutive primes which are far larger than average is to show that there are long sequences of consecutive integers which each have a small prime factor. A neat way to approach this was given by Erdős:

**Lemma 16.1.** *For any given  $z$  suppose that we can find arithmetic progressions  $a_p \pmod{p}$  for each prime  $p \leq z$  such that every integer  $n$  in the range  $1 \leq n \leq y = y(z)$  belongs to one of these arithmetic progressions. Then there exists an integer  $x \leq 2 \prod_{p \leq z} p$  such that there are no primes in the interval  $(x, x + y]$ . In particular we deduce that there exists  $p_n \leq 2 \prod_{p \leq z} p$  for which  $p_{n+1} - p_n > y(z)$ .*

*Proof.* Let  $X = 2 \prod_{p \leq z} p$  and let  $x \in [X + 1, 2X]$  for which  $x \equiv -a_p \pmod{p}$  for each  $p \leq z$ . We deduce that each integer in  $(x, x + y]$  has a prime factor  $\leq z$  and so is not itself prime, since if  $n \equiv a_p \pmod{p}$  then  $x + n \equiv -a_p + a_p \equiv 0 \pmod{p}$ .

Note that  $\log x \sim \log X \sim z$  by the prime number theorem, so our goal becomes to select  $y(z)$  is as large as possible, in particular so that  $y(z)/z \rightarrow \infty$  as  $z \rightarrow \infty$ .

We will need the following lemma:

**Lemma 16.2.** *Let  $\Psi(x, y)$  denote the number of integers up to  $x$  that are free of prime factors  $> y$ . Then*

$$\Psi(x, y) \ll x \left( \frac{e + o(1)}{u \log u} \right)^u \log y$$

where  $x = y^u$ , in the range  $x \geq y \geq (\log x)^{2+\epsilon}$ .

*Proof.* For any  $\sigma > 0$  we have

$$\Psi(x, y) = \sum_{\substack{n \leq x \\ p|n \implies p \leq y}} 1 \leq \sum_{\substack{n \geq 1 \\ p|n \implies p \leq y}} \left( \frac{x}{n} \right)^\sigma = x^\sigma \prod_{p \leq y} \left( 1 - \frac{1}{p^\sigma} \right)^{-1}.$$

Suppose that  $\frac{1}{2} + \epsilon < \sigma = 1 - \delta < 1$  so that the above is  $\ll x^\sigma \exp(\sum_{p \leq y} p^{-(1-\delta)})$ . Using exercise 16.1a with the choice  $\delta = \log(u \log u) / \log y$ , we obtain our result.

### Exercises

16.1a. Prove that if  $0 < 1 - \delta < 1$  then

$$\sum_{p \leq y} \frac{1}{p^{1-\delta}} \leq \frac{y^\delta}{\log(y^\delta)} + \log(1/\delta) + O\left(\frac{y^\delta}{\delta(\log y)^2} + 1\right).$$

(Hint: Split the sum at  $p \leq e^{2/\delta}$  and consider the two parts separately.)

**The Erdős-Rankin construction.** We select the values of  $a_p$  according to the size of the prime  $p$ : Select  $2 < z_1 < z_2 < z < y \leq z_1 z_2$ .

• Let  $a_p = 0$  for each prime  $p$  in  $(z_1, z_2]$ . Let  $S_1$  be the set of integers  $\leq y$  that are not divisible by any prime in  $(z_1, z_2]$ . If  $n \in S_1$  then either it is  $z_1$ -smooth, or it equals a prime  $q$  in  $(z_2, y]$  times an integer  $\leq y/q$ . Therefore

$$|S_1| = \Psi(y, z_1) + \sum_{z_2 < q < y} \left[ \frac{y}{q} \right] \leq y \left( \frac{e + o(1)}{u \log u} \right)^u \log z_1 + y \log \left( \frac{\log y}{\log z_2} \right) + O\left( \frac{y}{\log z_2} \right)$$

where  $y = z_1^u$ , using Lemma 16.2 and the prime number theorem. Therefore, if we have  $u \geq 2 \log \log y / \log \log \log y$  and  $z_2 = o(y)$  with  $\log z_2 \sim \log y$  then

$$|S_1| \leq (1 + o(1))y \frac{\log(y/z_2)}{\log y}.$$

• For each prime  $p \leq z_1$ , we select  $a_p$  “greedily”; that is, so that the largest number of remaining integers are  $\equiv a_p \pmod{p}$ . Evidently at least  $1/p$  of the remaining integers will fall in this congruence class, and so if  $S_2$  is the set remaining after this secondary sieving then

$$|S_2| \leq \prod_{p \leq z_1} \left(1 - \frac{1}{p}\right) N_1 \lesssim \frac{e^{-\gamma}}{\log z_1} y \frac{\log(y/z_2)}{\log y} = e^{-\gamma} y \frac{u \log(y/z_2)}{(\log y)^2}$$

by Mertens’s theorem.

• We assign to each prime  $p \in (z_2, z]$  a distinct elements of  $S_2$ , which we call  $a_p$ . This will sieve out all of the remaining elements of  $S_2$  provided  $\pi(z) - \pi(z_2) \geq |S_2|$ . Assuming that  $z_2 = o(z)$  we therefore require that

$$e^{-\gamma} y \frac{u \log(y/z_2)}{(\log y)^2} \leq (1 - \epsilon) \frac{z}{\log y}.$$

To maximize how large we can take  $y$ , we select  $u = 2 \log \log y / \log \log \log y$ , and  $z_2 = z / \log \log z$  so that  $\log(y/z_2) \sim \log(y/z)$ . We then solve the resulting equation to deduce that

$$y \leq \left( \frac{e^\gamma}{2} - \epsilon \right) \frac{z \log z \log \log \log z}{(\log \log z)^2}.$$

We deduce from Lemma 16.1 the following result:

**Theorem 16.3.** *There exist infinitely pairs of consecutive primes  $p_n < p_{n+1}$  for which*

$$p_{n+1} - p_n \geq \left( \frac{e^\gamma}{2} - o(1) \right) \frac{\log p_n \log \log p_n \log \log \log p_n}{(\log \log \log p_n)^2}.$$

A lot of effort have gone into improving the constant  $e^\gamma/2$ . Paul Erdős offered a prize of \$ 10,000 for anyone who could prove such a result with the  $e^\gamma/2$  replaced by an arbitrarily large constant

*Future versions of the chapter will discuss:*

Konyagin's improvement.

Maier's multiple gaps.

Hensley-Richards.

Is this the place for Daniel Shiu's thesis?