# Lecture 1

## Infinitely many primes

Prime number theory begins with the following famous result from antiquity:

**Theorem 1:** There are infinitely many primes.

**Euclid's proof** ($\sim$ 300 BC)

"prime numbers are more than any assigned multitude of primes"

Suppose $p_1, \ldots, p_k$ is any finite list of primes. Let $P := \prod_{j=1}^{k} p_j$ & consider $P+1$. Since $P+1 \equiv 1 \bmod p_j$ for each $1 \le j \le k$, none of the $p_j$ divide $P+1$. But since $P+1 > 1$ it must have a prime divisor. It follows that there is always a prime missing from any finite list. $\square$

**Exercises**

1. (Open!) Let $p_j$ denote the $j$th prime. Are there infinitely many $n$ for which $p_1 p_2 \cdots p_n + 1$ is prime?
2. Prove that there are arbitrarily large gaps in the primes.

**Notation** $\boxed{\pi(x) = \#\{p \le x : p \text{ is prime}\}.}$

We have of course just showed that $\lim_{x \to \infty} \pi(x) = \infty$. For more than 23 centuries, mathematicians have been concerned with providing quantitative versions of this qualitative relation.

One aim of this course (at least the first half) is to describe in detail the various methods which have been invented and implemented to achieve this.

Exercises

(3) Show that Euclid's proof gives the following (weak) quantitative information:

(i) The $n^{th}$ prime $p_n \leq 2^{2^n}$

(ii) There exists a constant $c>0$ such that

$$(*) \quad \pi(x) \geq c \log\log x \quad \text{for all sufficiently large } x.$$

Remarks on Notation (Landau & Vinogradov Notation)

We remind the reader that "$A = O(B)$" indicates that $|A| \leq c|B|$ for some constant $C>0$; an equivalent notation is "$A \ll B$". The notation "$A \gg B$" means $B \ll A$, and we write "$A \asymp B$" if $A \ll B$ & $A \gg B$. If $A$ & $B$ are functions of a single real variable $x$, we often speak of an estimate holding "as $x \to a$", which means that the estimate is valid in some (deleted) neighorhood of $a$.

Ex: $(*) \Leftrightarrow \pi(x) \gg \log\log x \quad (x \to \infty)$.

Subscripts on any of these symbols indicates parameters on which the implied constants may depend.

The notation "$A \sim B$" means $A/B \to 1$ while "$A = o(B)$" means $A/B \to 0$.

The lower bound on $\pi(x)$ in Exercise 3 is very far from being optimal. After having been conjectured for more than a century (notably by Legendre & Gauss) the prime number theorem, viz.

$$\pi(x) \sim \frac{x}{\log x} \qquad (x \to \infty)$$

was established independently in 1896 by Hadamard and de La Vallée-Poussin. Their methods rest on techniques of complex analysis, which we will discuss later in the course.

(one had to wait until 1949 for the appearance of the first elementary proofs of the prime number theorem, due to Erdős & Selberg.)

The first serious work on the function $\pi(x)$ is due to Chebyshev. In 1851 and 1852 he proved the following results:

1. If $\displaystyle\lim_{x\to\infty} \frac{\pi(x)}{x/\log x}$ exists, then that limit is 1.
2. For $x \geq 2$, $\pi(x) \asymp \dfrac{x}{\log x}$
3. (Bertrand's postulate) For all sufficiently large $x$, there is a prime in the interval $(x, 2x]$.

We will discuss these results and estimates of Mertens in the next lecture.

Exercise (4) (Open) Prove that there is a prime between any 2 squares.

# Further proof of the infinitude of primes

## Erdős' proof (1938)

Recall that a number is said to be squarefree if it is not divisible by any square greater than 1.

It is easy to see that

(i) # squarefree integers less than $N \leq 2^{\pi(N)}$ (extremely weak!)

(ii) # squares less than $N \leq \sqrt{N}$.

Since every natural number $n$ can be written as $rs^2$ where $r, s \in \mathbb{N}$ and $r$ is squarefree, it follows that

$$2^{\pi(N)}\sqrt{N} \geq N \iff \pi(N) \geq \log N / \log 4 \gg \log N \quad \square.$$

With this idea we can prove even more, namely

**Theorem 2**: $\sum_p \frac{1}{p}$ diverges.

*Proof*: Suppose $\sum_p \frac{1}{p}$ converges, then $\exists\, M > 0$ s.t. $\sum_{p>M} \frac{1}{p} < \frac{1}{2}$. (**)

Keep this $M$ fixed.

Let $N$ be an arbitrary natural number, it follows from (**) that more than half of the integers up to $N$ factor completely over primes $\leq M$.

Since (**) $\Rightarrow \sum_{M<p\leq N} \frac{N}{p} \leq \frac{N}{2}$. ↯ for $N > (2^{\pi(M)+1})$ since there are at most $2^{\pi(M)}\sqrt{N}$ numbers $\leq N$ with all prime factors $\leq M$. $\square$.

# of numbers $\leq N$ divisible by $p$.

## Euler's Proof of Theorem 1 (1737)

If there are finitely many primes, then

$$P_0 := \prod_p \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) = \prod_p \left(1 - \frac{1}{p}\right)^{-1} < \infty$$

Let

$$P_0(x) := \prod_{p \le x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) = \sum_{\substack{n \text{ whose prime} \\ \text{factors are all } \le x}} \frac{1}{n} \geq \sum_{n \le x} \frac{1}{n} \quad \left(\geq \log(x+1)\right)$$

Since $P_0(x) \le P_0 \ \forall x \Rightarrow \sum_{n \le x} \frac{1}{n} \le P_0 \ \forall x$ $\lightning$ $\square$

## Euler's Proof of Theorem 2

Suppose that $\sum_p \frac{1}{p} < \infty$. We observed above that $\prod_{p \le x} \left(1 - \frac{1}{p}\right)^{-1} \geq \log(x+1)$. (***)

It is easy to verify that $e^{y+y^2} \geq (1-y)^{-1} \ \forall\ 0 \le y \le \frac{1}{2}$ & hence

$$\prod_{p \le x} e^{\frac{1}{p} + \frac{1}{p^2}} \geq \log(x+1)$$

$$\Rightarrow \quad \sum_{p \le x} \frac{1}{p} + \sum_{p \le x} \frac{1}{p^2} \geq \log\log(x+1)$$

Since $\sum_{p \le x} \frac{1}{p^2} \le \sum_{n=2}^{\infty} \frac{1}{n} < 1$, it follows that

$$\sum_{p \le x} \frac{1}{p} \geq \log\log(x+1) - 1 \quad \left(= \log\log x + O(1) \text{ as } x \to \infty\right)$$

$\square$

This is actually close to the truth! (See one of "Mertens' Theorems" next time.)

We conclude this lecture with the observation that we can now prove that while the primes are infinite and "substantial", in the sense that $\sum_p \frac{1}{p}$ diverges, there are infact "not too many primes".

<u>Theorem 3</u>: The primes have asymptotic density 0, that is

$$\lim_{x\to\infty} \frac{\pi(x)}{x} = 0$$

This is of course immediate from the prime number theorem or even Chebyshev's estimates. <u>But</u>, it is more elementary than tha

<u>Proof</u> Fix $q\in\mathbb{N}$.

For any prime $p$ that does not divide $q$

$$p\equiv a \bmod q \quad \text{for some } (a,q)=1.$$

Since the number of natural numbers $n\leq x$ that fall into a given residue class mod $q$ is at most $\frac{x}{q}+1$ it follows that

$$\#\{n\leq x : (n,q)=1\} \leq \frac{\varphi(q)}{q}x + \varphi(q)$$

$\#\{1\leq a\leq q : (a,q)=1\}$

and hence (as only finitely many $p$ divide $q$, certainly $\leq q$) that

$$\pi(x) \leq \frac{\varphi(q)}{q}x + 2q .$$

It thus suffice to show that $\varphi(q)/q$ can be made arbitrarily small.

For each $z>0$, let $q := q_z = \prod_{p\leq z} p$.

$$\text{Since } \underbrace{\varphi(q_z) = \prod_{p\leq z}\varphi(p) = \prod_{p\leq z}(p-1)}_{\text{Exercise 5}} \Rightarrow \frac{\varphi(q_z)}{q_z} = \prod_{p\leq z}\left(1-\frac{1}{p}\right) \overset{\text{by }(***)}{\leq} \frac{1}{\log(z+1)} \to 0$$

as $z\to\infty$

$\square$