LECTURE 2: PRIMES AND THE FUNDAMENTAL THEOREM OF ARITHMETIC

Definition 0.1. A natural number $p >$ satisfying the condition

$$\text{whenever } d \mid p \text{ with } d > 0, \text{ one has either } d = 1 \text{ or } p,$$

is called a **prime number**. Any integer exceeding 1 which is not a prime number is called a **composite number**.

Theorem 0.2 (Factorisation into primes)**.** *Every integer n exceeding 1 may be written as a product of prime numbers.*

Proof. Suppose that the theorem holds for $1 < n < N$; note that this is vacuously true for $N = 2$. Let p be the least divisor of N greater than 1. Then p must be prime (why?). Now, if $n = N/p$ then $n < N$, so either $n = 1$ or else by hypothesis it is a product of prime numbers. In either case, it follows that $N = pn$ is a product of prime numbers. Therefore, by induction, every integer exceeding 1 is a product of prime numbers. □

Given a factorisation of an integer n into prime numbers, one may collect together like primes and order the primes by size so as to give a factorisation

$$n = \pm \prod_{i=1}^{s} p_i^{r_i},$$

where $p_1 < p_2 < \cdots < p_s$ are prime numbers, and $r_i \in \mathbb{N}$ $(1 \leqslant i \leqslant s)$. We will call this the *canonical prime factorisation* of n. Note that the empty product of (no) primes is equal to 1. If the choice of sign, the primes p_i, and the exponents r_i, are uniquely determined, we say that n has a *unique factorisation* into primes.

Lemma 0.3. *Suppose that p is a prime number, and $p \mid a_1 \dots a_t$. Then $p \mid a_i$ for some i with $1 \leqslant i \leqslant t$.*

Proof. We prove first that if m and n are natural numbers and $p \mid mn$, then $p \mid m$ or $p \mid n$. For if $p \nmid m$, then $(p, m) = 1$, and then it follows that $p \mid n$ (see Lecture 1). Moving now to the general case, the latter argument shows that when $p \mid a_1 \dots a_t$, then either $p \mid a_1$ or $p \mid a_2 \dots a_t$. The conclusion of the lemma therefore follows by induction on t. □

Theorem 0.4 (The Fundamental Theorem of Arithmetic)**.** *Integers $n > 1$ have unique factorisations into primes.*

Proof. Suppose that the theorem holds for $1 < n < N$. Let p be the smallest divisor of N greater than 1. Then p is a prime divisor of N. By Theorem 0.2, N has some factorisation into primes, and it follows from Lemma 0.3 that any such factorisation must contain p as one of the prime factors. If $p = N$ then

N is prime, and consequently it has unique factorisation into primes (why?). Otherwise, the integer $n = N/p$ satisfies $1 < n < N$, and hence possesses a unique factorisation into primes. But then $N = pn$ likewise has a unique factorisation into primes. Therefore, by induction, all integers $n > 1$ have a unique factorisation into primes. □

Aside: This method of proof, including Lemma 0.3, goes back to Euclid's *Elements*, c. 300BC. However, the Fundamental Theorem of Arithmetic itself is not stated in the *Elements.* Euclid and his contemporaries were no doubt aware of the uniqueness of prime factorisation, but it was taken for granted for many centuries. The lofty name and the emphasis of the FTA as an important fact are both due to Gauss in the early 19th century. In retrospect we can see that Gauss was anticipating the development of *algebraic number theory*, which considers more general number rings, including ones in which unique factorisation fails.

The unique factorisation theorem enables one to determine greatest common divisors and least common multiples simply. At least, that is the case when prime factorisations are available, which is computationally expensive data to assemble (the Euclidean Algorithm, on the other hand, is computationally very cheap). Suppose that

$$a = \prod_{i=1}^{s} p_i^{r_i} \quad \text{and} \quad b = \prod_{i=1}^{s} p_i^{t_i},$$

with the p_i distinct prime numbers and the exponents r_i and t_i non-negative integers. Then one has

$$(a,b) = \prod_{i=1}^{s} p_i^{\min\{r_i,t_i\}} \quad \text{and} \quad [a,b] = \prod_{i=1}^{s} p_i^{\max\{r_i,t_i\}}.$$

Moreover, since $\min\{r_i,t_i\} + \max\{r_i,t_i\} = r_i + t_i$, it follows from the latter formulae that $(a,b)[a,b] = |ab|$, as we saw in Lecture 1.

Theorem 0.5 (Euclid)**.** *There are infinitely many prime numbers, and hence also arbitrarily large prime numbers.*

Proof. Suppose that $p_1, \ldots, p_n$ are prime numbers, and put $N = p_1 \cdots p_n + 1$. Since $N > 1$, it has a prime divisor, say p, by Theorem 0.2. However, none of the p_i divide N, so p is a prime different from $p_1, \ldots, p_n$. Hence, no finite list of primes is complete, i.e. there are infinitely many of them. □

Note that, writing p_n for the nth prime number, the expression $p_1p_2 \cdots p_n+1$ is not always prime. Thus, for example, we have $2 \cdot 3 \cdots 13 + 1 = 30031 = 59 \cdot 509$.

Aside (the largest prime number): At the time of writing, the largest known prime is $2^{74207281} - 1$, a number with $22,338,618$ decimal digits. The primality of this number was established through the efforts of GIMPS (see Great Internet Mersenne Prime Search, at `http://www.mersenne.org/`). One can check that the integer $2^n - 1$ can be prime only when n is prime (why?). The integers $2^p - 1$ with

p a prime number are known as Mersenne numbers, and an industry of efficient primality tests for these special numbers is reflected in the GIMPS effort. On the other hand, it is conjectured that there are only finitely many Fermat primes, that is to say, integers of the shape $2^{2^n}+1$ which are prime numbers. These integers are known to be prime for $n = 0, 1, 2, 3, 4$, and known to be composite for $5 \leqslant n \leqslant 32$.

The method of the proof of Theorem 0.5 can be also use to give a quantitative estimate on the sequence of prime numbers p_n.

Theorem 0.6. *The nth largest prime number p_n satisfies $p_n \leqslant 2^{2^{n-1}}$.*

Proof. Suppose that N is a natural number and that the conclusion holds for $1 \leqslant n < N$; note that this is vacuously true for $N = 1$. Then by the argument of the proof of Theorem 0.5, one finds that

$$p_N \leqslant p_1 p_2 \cdots p_{N-1} + 1 \leqslant 2^{2^0} 2^{2^1} \cdots 2^{2^{N-2}} + 1 = 2^{2^{N-1}-1} + 1 \leqslant 2^{2^{N-1}}.$$

Then the inequality holds also for N, and so the desired conclusion follows by induction. □

Now define the function $\pi(x)$ for positive numbers x by putting

$$\pi(x) = \sum_{\substack{p \leqslant x \\ p \text{ prime}}} 1.$$

Thus one has $\pi(2) = 1$, $\pi(3) = 2$, $\pi(\sqrt{10}) = 2$, and so on.

Corollary 0.7. *One has $\pi(x) > \log_2 \log_2 x$ for $x > 1$.*

Proof. Let $x > 1$ and put $n = \pi(x) + 1$. Then $x < p_n \leqslant 2^{2^{n-1}} = 2^{2^{\pi(x)}}$, by Theorem 0.6. Taking logarithms establishes the corollary. □

Exercise 6* of Problem Sheet 2 shows that there are positive constants c_1 and c_2 with $c_1 < c_2$ such that for each number x with $x \geqslant 2$, one has

$$c_1 \frac{x}{\log x} \leqslant \pi(x) \leqslant c_2 \frac{x}{\log x}.$$

Given an interesting sequence such as the prime numbers, number theorists are interested in analysing features of their distribution. We begin with arithmetic progressions, about which we will say more as the course progresses.

Theorem 0.8. *There are infinitely many prime numbers of the shape $4k + 3$ ($k \in \mathbb{N}$).*

Proof. Suppose that $p_1, \ldots, p_n$ are primes of the form $4k + 3$. Consider the number $N = 4p_1 \cdots p_n - 1$. The integer N is odd, and of the shape $4k + 3$, so cannot be divisible exclusively by primes of the shape $4k + 1$. Moreover, none of the primes $p_1, \ldots, p_n$ divide N. Thus N is divisible by a new prime of the shape $4k + 3$ not amongst $p_1, \ldots, p_n$. Hence there are infinitely many such primes. □

See Question 4 on the second problem sheet for a proof that there are infinitely many prime numbers of the shape $4k+1$ ($k \in \mathbb{N}$). More generally, when a and b are natural numbers with $(a,b)=1$, Dirichlet proved in 1837 that $an+b$ is prime for infinitely many integers n.

Now we consider gaps between consecutive prime numbers.

Theorem 0.9. *There are arbitrarily large gaps between consecutive prime numbers.*

Proof. Consider the sequence $n!+2$, $n!+3$, $\ldots$, $n!+n$ of $n-1$ consecutive integers. The first of these integers is divisible by 2, the second by 3, and so on, with the last divisible by n. None of these integers can be prime, therefore, and so there are gaps of length at least n, for any natural number n, between consecutive prime numbers. □

Aside (gaps between primes): This theorem shows that one can find gaps between consective primes p_n and p_{n+1} at least as large as $C \log p_n / \log\log p_n$, for a suitable positive constant C, infinitely often. It was shown in 2014 by Ford, Green, Konyagin, Maynard and Tao that there is a positive number C with the property that the gaps can be as large as

$$C\frac{(\log p_n)(\log\log p_n)(\log\log\log\log p_n)}{\log\log\log p_n}$$

infinitely often. On the other hand, as highlighted in the first lecture, stunning recent progress has established that $p_{n+1}-p_n \leqslant 246$ infinitely often.

A natural question is whether there are simple ways to produce prime numbers. The next theorem shows that polynomials, at least, cannot take prime values all the time.

Theorem 0.10. *There is no non-constant polynomial which takes only prime values.*

Proof. Suppose that

$$f(x)=\sum_{k=0}^{d} c_k x^k$$

is a polynomial with integer coefficients c_k and degree $d \geqslant 1$. If $f(0)=c_0$ is not prime then there is nothing to prove, so assume that c_0 is prime. Then, for any $n \in \mathbb{N}$, we have

$$f(c_0 n)=c_0+\sum_{k=1}^{d} c_k (c_0 n)^k.$$

Hence $c_0 \mid f(c_0 n)$, and since $d \geqslant 1$, $f(c_0 n) \neq c_0$ for sufficiently large n. Therefore, $c_0 > 1$ is a proper divisor of $f(c_0 n)$, so $f(c_0 n)$ is not prime. □

Aside: Matiyasevich showed in 1970 that there exist polynomials $f(n_1,\ldots,n_k)$ such that the set of *positive* values assumed by f, as $n_1,\ldots,n_k$ vary through all natural numbers, is exactly the set of prime numbers.