

two factorizations). Without loss of generality, we may suppose that  $p' < p$ . Hence

$$p' < p \leq q \leq r \leq \dots \quad (*)$$

Since  $n$  is not prime,  $n \geq p^2$ , and hence  $n > pp'$ . By minimality of  $n$ ,  $n - pp'$  has a unique factorization. Both  $p$  and  $p'$  are factors of  $n - pp'$  (since  $n - pp' = p(qr \cdots - p') = p'(q'r' \cdots - p)$ ) and hence, for some positive integer  $z$ ,  $n - pp' = pp'z$ . This gives  $qr \cdots - p' = p'z$ , so that  $p'$  is a factor of  $qr \cdots$ . Since  $qr \cdots < n$ ,  $qr \cdots$  has a unique factorization into primes. Thus  $p'$  is one of  $q, r, \dots$ . But this contradicts  $(*)$  above. For another proof, see Part II, Chapter 15.

Like Euclid, Eratosthenes of Cyrene (230 BC) worked at the University of Alexandria. He suggested a method for making a list of all prime numbers, which is called the 'sieve of Eratosthenes'. His method is as follows: write down all the positive integers greater than 1; cross out all multiples of 2 other than 2, cross out all multiples of 3 other than 3 which have not been crossed out yet, etc. In the end, the numbers not crossed out form a complete list of primes.

People often wonder whether there is a simple formula representing prime numbers. For example,  $f(x) = x^2 - x + 41$  is prime for all integer values of  $x$  from 0 to 40. While this might convince a physicist that  $f(x)$  is always prime, unfortunately  $f(41) = 41^2$ .

In 1743, Christian Goldbach observed that a polynomial

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

with integer coefficients  $a_0, a_1, \dots, a_n$  cannot represent primes only, that is, the integers  $f(0), f(1), f(2), \dots$  are not all prime.

Indeed, if  $f(0) = p$ , then  $f(kp)$  is clearly a multiple of  $p$  for all integers  $k$ . But, as  $k$  tends to infinity, so does the absolute value of  $f(kp)$ . Hence, for some value of  $k$ ,  $f(kp)$  will be a proper multiple of  $p$  and therefore not prime.

It therefore came as a great surprise to the mathematical community when, in 1970, Yuri Matiyasevič formed a polynomial  $f(x, y, z, \dots)$  with integer coefficients, but in several variables, such that, when positive integers are chosen for  $x, y, z, \dots$ , one gets all the prime numbers and only the prime numbers as positive values of the polynomial. We shall say more about this in Chapter 21 on Hilbert's Tenth Problem in Part II.

In 1830 (in *Théorie des Nombres* Vol. II, p. 65), A. M. Legendre noted that, if  $\pi(x)$  is the number of primes less than or equal to  $x$ , then  $\pi(x)$  is approximately equal to  $x/(\log_e x - 1.08366)$ , where  $e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$  is the base of the natural logarithm (Chapter 26). We shall write  $\log x$  and assume the base to be  $e$ . He was not able to prove this. In 1896, two mathematicians working independently proved that

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x/\log x} = 1.$$

These two mathematicians were the Frenchman Jacques Hadamard (1865–1963) and the Belgian Charles Jean de la Vallée Poussin (1866–1962). The result they proved is called the *Prime Number Theorem*. It implies that the  $n$ th prime is approximately equal to  $n \log n$ . For, if we let  $p_n$  be the  $n$ th prime, the equation implies that  $n$  is roughly equal to  $p_n / \log p_n$ , so that

$$p_n \approx n \log p_n \approx n \log(n \log p_n) \approx n \log n,$$

since  $n \log \log p_n$  can be neglected in comparison with  $n \log p_n \approx p_n$ .

It was Goldbach who conjectured that every even number greater than 2 is a sum of two primes. This conjecture has not yet been proved or disproved. However, in 1937, the Russian mathematician I. M. Vinogradov made some progress towards proving Goldbach's Conjecture, by showing that every odd integer greater than, say,  $10^{10^{10}}$  (or some similar bound) is a sum of three prime numbers. Some progress in the Goldbach conjecture was recently made by the Chinese mathematician Chen Jing-Run. He proved that every sufficiently large (say  $> 10^{10^{10}}$ ) even number has the form  $p + q$ , where  $p$  is prime and  $q$  is either prime or the product of two primes. During the so-called 'cultural revolution' in the sixties this kind of mathematics was frowned upon in China for being far removed from any conceivable application to industry or agriculture. Because he stubbornly stuck to his esoteric research at the risk of neglecting his teaching, Chen Jing-Run was discriminated against during the reign of the so-called 'gang of four' and may have lost his academic position. After the overthrow of the gang of four, he was rehabilitated and even declared a 'hero of the revolution'.

At the moment (1995), one of the 'hot topics' in prime number theory is cryptography. In its simplest form, the idea is this: the cipher key is a product  $n = pq$  of two large primes, typically having 50 to 80 digits each. Knowing  $n$  is enough to encode messages, but decryption requires knowledge of the factorization. The integer  $n$  is made public (hence the term 'public key') so that everyone can use the code to encipher messages. Security is maintained, because only the intended recipient knows the key, namely, the factorization  $pq$ , necessary to carry out the decryption. The basis for this scheme is that it takes a very long time to factor products of large primes and the war may well be over before the enemy succeeds in doing so. (Try to factor the relatively small product 1,315,685,447, and you will see that the enemy does not have an easy task.)

At the moment, much research is being done to find refinements of the above idea, refinements that are at once economical and secure for those who want to send secret messages. Much research is also being done to find ways of using computers to factor very large numbers, and thus break the codes based on the above idea.