By Proposition 8.1, N is equal to a product of primes, say $N = q_1 \dots q_r$ with all q_i prime. As q_1 is prime, it belongs to the above list of all primes, so $q_1 = p_i$ for some i.

Now q_1 divides N, hence p_i divides N. Also p_i divides $p_1 p_2 \dots p_n$, which is equal to $N-1$. Thus, p_i divides both N and $N-1$. But this implies that p_i divides the difference between these numbers, namely 1. This is a contradiction. ∎

Theorem 13.1 is of course not the end of the story about the primes — it is really the beginning. A natural question to ask that flows from the theorem is: What *proportion* of all positive integers are prime? On the face of it this question makes no sense, as the integers and the primes are both infinite sets. But one can make a sensible question by asking

Given a positive integer n, how many of the numbers $1, 2, 3, \dots, n$ are prime?

Is there any reason to expect to be able to answer this question? On the face of it, no. If you stare at a long list of primes, you will see that the sequence is very irregular, and it is very difficult to see any pattern at all in it. (See, for example, Exercise 6 at the end of the chapter.) Why on earth should there then be a nice formula for the number of primes up to n?

The amazing thing is that there *is* such a formula, albeit an "asymptotic" one (I will explain this word later). The great Gauss, by calculating a lot with lists of primes (and also by having a lot of brilliant thoughts) formed the incredible conjecture (i.e., informed guess) that the number of primes up to n should be pretty close to the formula

$$\frac{n}{\log_e n}.$$

To understand this a little, compare the number of primes up to, say, 10^6 — namely, 78498 — with the value of $\frac{10^6}{\log 10^6}$ — namely, 72382.4. The *difference* between these two numbers, about 6000, appears to be quite large; but their *ratio* is 1.085, quite close to 1. It was on the ratio, rather than the difference, that Gauss concentrated his mind: his conjecture was that the ratio of the number of primes up to n and the expression $\frac{n}{\log_e n}$ should get closer and closer to 1 as n gets larger and larger. (Formally, this ratio *tends to* 1 *as n tends to infinity*.)

Gauss did not actually manage to prove his conjecture. The world had to wait until 1896, when a Frenchman, Hadamard, and a Belgian, de la Vallée-Poussin, both produced proofs of what is now known as the Prime Number Theorem:

THEOREM 13.2

For a positive integer n, let $\pi(n)$ be the number of primes up to n. Then the ratio of $\pi(n)$ and $\frac{n}{\log_e n}$ tends to 1 as n tends to infinity (i.e., the ratio can be made as close as we like to 1 provided n is large enough).

The proof of this result uses some quite sophisticated tools of Analysis. Nevertheless, if you are lucky you might get the chance to see a proof in an undergraduate course later in your studies — in other words, it is not *that* difficult!

You should not think that every question about the primes can be answered (if not by you, then by some expert or other). On the contrary, many basic questions about the primes are unsolved to this day, despite being studied for many years. Let me finish this chapter by mentioning a couple of the most famous such problems.

The Goldbach conjecture If you do some calculations, or program your computer, you will find that any reasonably small even positive integer greater than 2 can be expressed as a sum of two primes. For example,

$$10 = 7+3,\ 50 = 43+7,\ 100 = 97+3,\ 8000 = 3943+4057$$

and so on. Based on this evidence, it seems reasonable to conjecture that *every* even positive integer is the sum of two primes. This is the Goldbach conjecture, and it is unsolved to this day.

The twin prime conjecture If p and $p+2$ are both prime numbers, we call them *twin primes*. For example, here are some twin primes:

$$3,5;\ 5,7;\ 11,13;\ 71,73;\ 1997,1999.$$

If you stare at a list of prime numbers, you will find many pairs of twin primes, getting larger and larger. One feels that there should be infinitely many twin primes, and indeed, that statement is known as the twin prime conjecture. Can one prove the twin prime conjecture using a proof like Euclid's in Theorem 13.1? Unfortunately not — indeed, no one has come up with any sort of proof, and the conjecture remains unsolved to this day.

Exercises for Chapter 13

1. Prove Liebeck's *triplet prime conjecture*: the only triplet of primes of the form $p, p+2, p+4$ is $\{3,5,7\}$.
2. Let n be an integer with $n \geq 2$. Suppose that for every prime $p \leq \sqrt{n}$, p does not divide n. Prove that n is prime.

   Is 221 prime? Is 223 prime?