The sum property of the binomial coefficients also explains the presence of some interesting numbers in Pascal's triangle.

11.1.3 Explain why the third diagonal from the left in the triangle, namely 1, 3, 6, 10, 15, 21, ..., consists of the triangular numbers.

11.1.4 The numbers on the next diagonal, namely 1, 4, 10, 20, 35..., can be called "tetrahedral numbers." Why is this an apt description?

11.2 Fermat's Little Theorem

The best-known theorem actually proved by Fermat (1640a), and known as his "little" or "lesser" theorem to distinguish it from his "last" or "great" theorem (next section), is the following.

If p is prime and n is relatively prime to p, then

$$n^{p-1} \equiv 1 \pmod{p}.$$

Equivalent statements of the conclusion, which avoid using the "congruent mod p" language unknown in Fermat's time, are

$$n^{p-1} - 1 \text{ is divisible by } p$$

or

$$n^{p} - n \text{ is divisible by } p.$$

The latter holds because $n^p - n = n(n^{p-1} - 1)$ is divisible by p only if $n^{p-1} - 1$ is, since p is prime and does not divide n.

Fermat's little theorem has recently become indispensable in certain areas of applied mathematics, such as cryptography, so it is thought-provoking to learn that it originated in one of the least applied problems in mathematics, the construction of perfect numbers. As we saw in Section 3.2, this depends on the construction of prime numbers of the form $2^m - 1$, and it was initially for this reason that Fermat became interested in conditions for $2^m - 1$ to have divisors. At the same time (mid-1630s) he was investigating the binomial coefficients, and the combination of these two interests very likely led to the discovery of his little theorem, for $n = 2$.

His actual proof is unknown, but various authors [for example Weil (1984), p. 56] have pointed out that the theorem follows immediately from the fact that $\binom{p}{1}, \binom{p}{2}, \ldots, \binom{p}{p-1}$, for p prime, are divisible by p:

$$2^p = (1+1)^p = 1 + \binom{p}{1} + \binom{p}{2} + \cdots + \binom{p}{p-1} + 1,$$

hence

$$2^p - 2 = \binom{p}{1} + \binom{p}{2} + \cdots + \binom{p}{p-1}$$

is divisible by p, and therefore so is $2^{p-1} - 1$.

But how does one prove that $\binom{p}{1}, \binom{p}{2}, \ldots, \binom{p}{p-1}$ are divisible by p? This follows easily from the Levi ben Gershon formula

$$\binom{p}{k} = \frac{p!}{(p-k)!k!},$$

which shows that the prime p is a factor of the numerator but not of the denominator. The denominator nevertheless divides the numerator, since $\binom{p}{k}$ is an integer, so the factor must remain intact after the division has taken place. Fermat may not have had precisely this result, since he did not yet have Pascal's combinatorial interpretation of the binomial coefficients, but he did have the formula

$$n\binom{n+m-1}{m-1} = m\binom{n+m-1}{m},$$

which implies it and from which the divisibility property may be extracted [see Weil (1984), p. 47].

Thus far we have a proof of Fermat's little theorem for $n = 2$. Weil (1984) suggests two possible routes to the general theorem from this point. One is by iteration of the binomial theorem, a method that was used in the first published proof of Fermat's theorem by Euler (1736). The other is by direct application of the *multinomial theorem*, the method of the earliest known proof, which is in an unpublished paper of Leibniz from the late 1670s [see Weil (1984), p. 56].

Just as the coefficient of $a^{p-k}b^k$ in $(a+b)^p$ is $p!/(p-k)!k!$, the

coefficient of $a_1^{q_1}a_2^{q_2}\cdots a_n^{q_n}$ in $(a_1 + a_2 + \cdots + a_n)^p$ is $p!/q_1!q_2!\cdots q_n!$,

where $q_1 + q_2 + \cdots + q_n = p$ (Exercise 11.2.4). This *multinomial coefficient* is divisible by p, by the same argument as before, provided no $q_i = p$. Thus the coefficients of all but $a_1^p, a_2^p, \ldots, a_n^p$ in $(a_1 + a_2 + \cdots + a_n)^p$ are divisible by the prime p. It follows, by replacing each of the n terms $a_1, a_2, \ldots a_n$ by 1, that

$$(1 + 1 + \cdots + 1)^p = 1^p + 1^p + \cdots + 1^p + \text{terms divisible by } p,$$

that is, $n^p - n$ is divisible by p. Then if n itself is relatively prime to p (hence not divisible by p), we have $n^{p-1} - 1$ divisible by p, or the general Fermat little theorem.

EXERCISES

The binomial theorem may be iterated to show that p divides $n^p - n$ as follows.

11.2.1 Use the result $2^p = (1+1)^p = 2 +$ terms divisible by p, and its method of proof, to show that

$$3^p = (2+1)^p = 3 + \text{terms divisible by } p.$$

11.2.2 Build on the idea of Exercise 11.2.1 to show that $n^p - n$ is divisible by p for any positive integer n.

11.2.3 Observe the terms divisible by p in the first few rows of Pascal's triangle, computed in the previous section.

Like the binomial theorem, the multinomial theorem can be proved combinatorially by considering the number of ways a term $a_1^{q_1} a_2^{q_2} \cdots a_n^{q_n}$ can arise from the factors of $(a_1 + a_2 + \cdots + a_n)^p$.

11.2.4 Prove the formula for the multinomial coefficient given above by observing that the coefficient equals the number of ways of partitioning p things into disjoint subsets of sizes $q_1, q_2, \ldots, q_n$.

11.3 Fermat's Last Theorem

> On the other hand, it is impossible for a cube to be written as a sum of two cubes or a fourth power to be written as a sum of two fourth powers or, in general, for any number which is a power higher than second to be written as a sum of two like powers. I have a truly marvellous demonstration of this proposition which this margin is too small to contain.
>
> [Fermat (1670), p. 241]

This remark, written in the margin of his copy of Bachet's *Diophantus* when he was studying that work in the late 1630s, is the second item in Fermat's *Observations on Diophantus*, published posthumously in 1670. Fermat was responding to Diophantus' treatment of the problem of expressing a square as a sum of two squares. As we saw in Chapter 1, this is the problem of finding Pythagorean triples (a,b,c) or, equivalently, of finding the rational points $(a/c, b/c)$ on the circle $x^2 + y^2 = 1$.

Fermat's last theorem, the claim that there are no triples (a,b,c) of positive integers such that

$$a^n + b^n = c^n, \quad \text{where } n > 2 \text{ is an integer}$$

became the most famous problem in mathematics. Many mathematicians contributed solutions for particular values of n: Euler for $n = 3$, Fermat himself for $n = 4$ (see next section), Legendre and Dirichlet for $n = 5$, Lamé for $n = 7$, Kummer for all prime $n < 100$ except 37, 59, 67. A thorough account of these early results may be found in Edwards (1977). Of course it is sufficient to prove the theorem for prime exponents, since a counterexample

$$a^n + b^n = c^n$$

for a nonprime exponent $n = mp$, where p is prime, would also be a counterexample

$$(a^m)^p + (b^m)^p = (c^m)^p$$

for the prime exponent p.

After Kummer, not much progress was made until the 1980s, when two new approaches were opened up. Faltings (1983) showed that for each exponent n there were *at most finitely many counterexamples* to Fermat's last theorem. This is a consequence of Faltings' much more general theorem, settling a conjecture of Mordell (1922), that each curve of genus >1 has at most finitely many rational points. The concept of genus is explained in Chapter 15. For the moment we mention only that the "Fermat curve"

$$x^n + y^n = 1$$

has genus 0 when $n = 2$, genus 1 when $n = 3$, and genus >1 otherwise. Thus Faltings' theorem showed that the Fermat curve could have at most finitely many rational points (and hence $a^n + b^n = c^n$ could have at most finitely many integer solutions) in the cases not already settled.

The second approach was initiated by Frey (1986), who made the astonishing suggestion that a counterexample $a^n + b^n = c^n$ to Fermat's last theorem might imply something impossible about the *cubic* curve

$$y^2 = x(x - a^n)(x + b^n).$$

At the time, the property in question—called *nonmodularity*—was only conjectured to be impossible, and it was also not known to be implied by a counterexample to Fermat's last theorem. However, Ribet (1990) proved that a counterexample implies nonmodularity, and in 1994 Andrew Wiles proved that nonmodularity is impossible for cubic curves of the above form. Thus no counterexample to Fermat's last theorem can exist.

There was a dramatic twist to this closing chapter in the story of Fermat's last theorem, because Wiles first announced his result in 1993 (after seven years working on it in seclusion), only to discover within months that there was a serious gap in his proof. However, with the help of Richard Taylor, the gap was filled in 1994, and the completed proof was published in Wiles (1995). The proof is highly sophisticated, but we can at least explain its general setting of cubic curves and elliptic functions; indeed these are important threads throughout the whole of this book.

11.4 Rational Right-angled Triangles

> The area of a right-angled triangle the sides of which are rational numbers cannot be a square number. This proposition, which is my own discovery, I have at length succeeded in proving, though not without much labour and hard thinking. I give the proof here, as this method will enable extraordinary developments to be made in the theory of numbers.
>
> [Fermat (1670), p. 271]

This is number 45 of Fermat's *Observations on Diophantus*, responding to a problem posed by Bachet: to find a right-angled triangle whose area equals a given number. The observation is important not only for the theorem and the method announced, but also because it is followed by the only reasonably complete proof left by Fermat in number theory. As a bonus, the proof implicitly settles Fermat's last theorem for $n = 4$ (see exercises) and is an excellent illustration of his "method" of *infinite descent*, which did indeed lead to extraordinary developments in the theory of numbers. In what follows, the statements that make up Fermat's proof, appearing indented like the quote above, are expanded and expressed in modern notation following the reconstruction of Zeuthen (1903), p. 163. We use the translation of Fermat given by Heath (1910), p. 293, in his version of the reconstruction.

> If the area of a right-angled triangle were a square, there would exist two biquadrates the difference of which would be a square

> number. Consequently there would exist two square numbers the sum and difference of which would be squares.

By choosing a suitable unit of length, we can express the sides of a rational right triangle as a Pythagorean triple of relatively prime integers $p^2 - q^2$, $2pq$, $p^2 + q^2$, as noted in Section 1.2. Since their gcd is 1, $\gcd(p,q) = 1$ also. Therefore, since $2pq$ is even, $p^2 - q^2$ and its factors $p+q$, $p-q$ must be odd. Also, no two of p, q, $p+q$, $p-q$ have a common prime divisor, otherwise p, q would. Then if the area $pq(p+q)(p-q)$ is a square, its factors must all be squares:

$$p = r^2, \quad q = s^2, \quad p+q = r^2+s^2 = t^2, \quad p-q = r^2 - s^2 = u^2. \qquad (1)$$

Thus the sum and difference of the squares r^2, s^2 are also squares, so

$$r^4 - s^4 = (r^2+s^2)(r^2-s^2) = t^2u^2 = v^2.$$

> Therefore we should have a square number which would be equal to the sum of a square and the double of another square, while the squares of which this sum is made up would themselves have a square number for their sum.

From (1) we have

$$t^2 - u^2 = 2s^2, \quad \text{that is,} \quad t^2 = u^2 + 2s^2. \qquad (2)$$

And also from (1),

$$u^2 + s^2 = r^2.$$

> But if a square is made up of a square and the double of another square, its side, as I can very easily prove, is also made up of a square and the double of another square.

Since $(t+u)(t-u) = t^2 - u^2 = 2s^2$ from (2), $(t+u)(t-u)$ is even. Then one of $t+u$, $t-u$ is even, and consequently so is the other. Put

$$t+u = 2w, \quad t-u = 2x. \qquad (3)$$

Then

$$s^2 = (t+u)(t-u)/2 = 2wx.$$

Tracing back through (3), (2), (1) we see that any common divisor of w, x would also be common to t, u, to t^2, u^2, to r^2, s^2 and hence to p, q. Thus

w, x are relatively prime and therefore, since wx is twice a square, we have either

$$w = y^2, \quad x = 2z^2 \qquad \text{or} \qquad w = 2z^2, \quad x = y^2.$$

In either case,

$$t = w + x = y^2 + 2z^2. \tag{4}$$

> From this we conclude that the said side is the sum of the sides about the right angle in a right-angled triangle, and that the simple square contained in the sum is the base, and the double of the other square the perpendicular.

If we let y^2, $2z^2$ be the sides of a right triangle, then the hypotenuse h satisfies

$$\begin{aligned} h^2 = (y^2)^2 + (2z^2)^2 &= \frac{1}{2}\left((y^2+2z^2)^2 + (y^2-2z^2)^2\right) \\ &= \frac{1}{2}(t^2+u^2) && \text{by (3) and (4)} \\ &= r^2. && \text{by (1)} \end{aligned}$$

Hence $h = r$ and the triangle is rational.

> This right-angled triangle will thus be formed from two squares, the sum and difference of which will be squares. But both these squares can be shown to be smaller than the squares originally assumed to be such that both their sum and their difference are squares.

The original squares with sum and difference equal to squares were $p = r^2$, $q = s^2$, which came from the perpendicular sides $p^2 - q^2$ and $2pq$ of the rational right triangle whose area was assumed to be a square. We now have a rational (indeed integral) right triangle with perpendicular sides y^2, $2z^2$ whose area y^2z^2 is also a square. This triangle is smaller, since its hypotenuse r is less than side $2pq$ of the original triangle, and hence it gives a smaller pair of (integer) squares p', q', whose sum and difference are squares.

> Thus, if there exist two squares such that the sum and difference are both squares, there will also exist two other integer squares which have the same property but a smaller sum. By the same reasoning we find a sum still smaller than the last

> found, and we can go on *ad infinitum* finding integer square numbers smaller and smaller with the same property. This is, however, impossible because there cannot be an infinite series of numbers smaller than any given integer we please.

This contradiction means that the initial assumption of a rational right triangle with square area is false. The versions of Zeuthen and Heath proceed more directly to a contradiction than Fermat by observing that the descent from the hypothetical initial triangle to the one with area y^2z^2 can be iterated to give an infinite descending sequence of integer areas. Weil (1984), p. 77, shortens the proof even further.

The logical principle involved in Fermat's method of descent is of course the same as that on which mathematical induction is based: any set of natural numbers has a least member. However, the circumstances in which the two methods can be applied are quite different. With induction, one needs a suitable hypothesis on which to make the induction step; with descent, one needs a suitable quantity on which to descend. In practice, descent is a much more special method, being associated with geometric properties of certain curves: the genus 1 curves we shall meet in section 11.6 and later chapters [see also Weil (1984), p. 140]. The general problem raised by Bachet—deciding which numbers n are the areas of rational right triangles—is in fact intimately connected with the theory of genus 1 curves, and its recent resurgence is beautifully covered by Koblitz (1985).

Exercises

Two of the propositions that arise in the descent from the hypothetical rational right triangle with square area are of independent interest and are also false because they imply the existence of such a triangle.

11.4.1 Show that the existence of squares r^2 and s^2 for which r^2+s^2 and r^2-s^2 are both squares implies the existence of a rational right triangle with square area.

11.4.2 Show that a nonzero integer solution of $r^4-s^4=v^2$ implies the existence of a rational right triangle with square area. (Hint: It's the same triangle as in Exercise 11.4.1.)

11.4.3 From Exercise 11.4.2, deduce Fermat's last theorem for $n=4$.

The impossibility of a nonzero integer solution $r^4-s^4=v^2$ can also be shown by a more direct descent that avoids some of the steps used by Fermat. The main