Modular Forms & Diophantine Questions

Kenneth A. Ribet

Mathematics Department
University of California
Berkeley, CA 94720-3840
USA
e-mail: ribet@math.berkeley.edu

Introduction

I visited Singapore in March, 2000 at the invitation of the International Conference on Fundamental Sciences. During that conference, I spoke to students who were assembled at Victoria Junior college. My theme was a set of questions that had been sent to me by students and amateur mathematicians. My article [22] was based on the transparencies that I showed during my talk. At the same conference, I gave a "Public Lecture" at the National University of Singapore on the connection between Fermat's Last Theorem, and the conjecture — now a theorem! — to the effect that elliptic curves are related to modular forms. This article is based on the second lecture.

The intention of this article is to offer a glimpse of some of the mathematics that is associated with Fermat's Last Theorem. It might plausibly be read in conjunction with other articles that I have written about Fermat's Last Theorem: My article [23] with Brian Hayes in *American Scientist* focuses on the connection between Fermat's equation and elliptic curves. It was written in 1994, when the proof that Andrew Wiles announced in 1993 was not yet complete. My exposition [21] is intended for professional mathematicians who are not necessarily specialists in number theory. The introduction [25] by Simon Singh and me will be useful to readers who seek a summary of Singh's book [24] and to the documentary on Fermat's Last Theorem that Singh directed for the BBC [15].

I thank the National University of Singapore and the Isaac Newton Institute for Mathematical Sciences at Cambridge for organizing the ICFS and for their kind hospitality in Singapore. My research was partially supported by the US National Science Foundation during the preparation of this article.

Background

Arguably the single most famous statement in mathematics is the assertion that Fermat's equation

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

has no solutions in positive integers a, b, and c when n is an integer greater than 2. According to his son Samuel, Pierre de Fermat wrote this assertion in the margin of his copy of Diophantus's *Arithmetic*, roughly in 1637.

Although Fermat may have believed in the 1630's that he had a proof of what came to be known as "Fermat's Last Theorem," we can only speculate as to what Fermat had in mind. It is widely believed that the argument that Fermat had mapped out for himself ran into unexpected difficulties. Indeed, when he was a mature mathematician, Fermat detailed a proof that

$$a^4 + b^4 = c^2$$

has no solution in positive integers, thus proving in particular that a perfect fourth power is not the sum of two others. Had Fermat been able to treat

$a^n + b^n = c^n$ for all n, he probably would not have been interested in the special case $n = 4$.

It is worth pointing out that Fermat made at least one other mathematical assertion that proved to be incorrect: Fermat believed that the "Fermat numbers" $F_n := 2^{2^n} + 1$ are all prime. The first few of them—3, 5, 17, 257 and 65537—are indeed prime numbers. The next number in the series, $F_5 = 2^{32} + 1 = 4294967297$, is *not* a prime: it's the product of 641 and 6700417. Incidentally, there is no known n bigger than 4 for which F_n is prime. On the other hand, the numbers $F_6, \ldots, F_{32}$ are known currently to be composite (i.e., non-prime); see `http://www.prothsearch.net/fermat` for information of this type, including a list of known factors of specific numbers F_n.

Early History

Fermat's Last Theorem has a long history, beginning with Fermat's work on the case $n = 4$ and Euler's 18th century study of $a^3 + b^3 = c^3$.

The techniques used in the 17th and 18th centuries are now included in the curriculum of undergraduate courses in number theory. For example, the work of Fermat and of Euler is discussed in the first two chapters of [9] and at various junctures in [12]. (The latter book is one of my favorite introductions to number theory. I recommend it enthusiastically to Berkeley students who seek an introduction to modern methods in number theory.)

After thinking about the first cases $n = 3$ and $n = 4$ of Fermat's equation, one turns naturally to exponents larger than 4. In fact, a simple remark shows that one need treat only the case where n is a prime number bigger than 2. Indeed, it is clear that Fermat's assertion, when true for a given exponent n, is true for all exponents that are multiples of n. For example, knowing the assertion for $n = 3$ allows us to conclude that there are no counterexamples to Fermat's assertion when n is 6, 9, 12, and so on. This remark follows from the simple observation that any perfect sixth power is in particular a perfect cube, and so forth.

Now any integer n bigger than 2 is either a power of 2 (2^t with $t \geq 2$) or else is a multiple of some prime number $p > 2$. Since integers of the first kind are divisible by 4—an exponent for which Fermat himself proved Fermat's Last Theorem—it suffices to consider exponents that are odd prime numbers when one seeks to prove Fermat's Last Theorem. In other words, after verifying Fermat's assertion for $n = 4$ and $n = 3$, mathematicians were left with the problem of proving the assertion for the exponents 5, 7, 11, 13, 17, and so on.

Progress was slow at first. The case $n = 5$ was settled by Dirichlet and Legendre around 1825, while the case $n = 7$ was treated by Lamé in 1832.

In the middle of the nineteenth century, E. Kummer made a tremendous advance by proving Fermat's Last Theorem for an apparently large class of prime numbers, the *regular primes*. The definition of this class may be given

quickly, thanks to a numerical criterion that was established by Kummer. Namely, one considers the expression

$$\frac{x}{e^x-1} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} - \frac{x^8}{1209600} + \frac{x^{10}}{47900160} - \frac{691\,x^{12}}{1307674368000} + \cdots$$

and defines the ith Bernoulli number B_i to be the coefficient of $\frac{x^i}{i!}$ in this expansion. Thus B_{12}, for example is $-\frac{691}{2730}$; the denominator is $2\cdot 3\cdot 5\cdot 7\cdot 13$, the product of those primes p for which $p-1$ divides 12. A prime number $p \geq 7$ is regular if p divides the numerator of none of the even-indexed Bernoulli numbers $B_2, B_4, \ldots, B_{p-3}$. The primes $p < 37$ turn out to be regular. On the other hand, 37 is irregular (i.e., not regular) because it divides the numerator of B_{32}: the numerator is $7709321041217 = 37\cdot 683\cdot 305065927$. (We may conclude that 683 and 305065927 are irregular as well.)

A proof of Fermat's Last Theorem for regular primes, along the lines given by Kummer, may be found in [9, Ch. 5]. See also [17] and [2] for alternative discussions. In these books, the reader will find a proof that there are infinitely many irregular prime numbers; see, for example, [17, Ch. VI, §4] or [2, Ch. 5, §7.2]. Although heuristic probabilistic arguments suggest strongly that regular primes should predominate, the set of regular primes is currently not known to be infinite.

Over the years, Kummer's work was refined repeatedly. Aided by machine calculation, mathematicians employed criteria such as those presented in [17] to verify Fermat's Last Theorem for all prime exponents that did not exceed ever increasing bounds. Most notably, four mathematicians proved Fermat's Last Theorem for all prime exponents below four million in an article that was published in 1993 [4]. It is striking that the calculations in that article were motivated by questions involving Bernoulli numbers and the arithmetic of cyclotomic fields; the proof of Fermat's Last Theorem for a large set of prime numbers came almost as an afterthought.

Readers used to dealing with experimental sciences might well now ask why a mathematician would insist on a rigorous proof of a statement, depending on a parameter n, that can be verified by calculation for all $n \leq 4,000,000$. A statement that is true in this range seems very likely to be true for all n. To answer this question, it suffices to point out the logical possibility that an assertion that is true experimentally may have one or more counterexamples that happen to be very large.

In fact, assertions that realize this possibility are not hard to find in number theory. As Fermat himself knew, the first solution to $x^2 - 109y^2 = 1$ in positive integers x and y is given by:

$$x = 158070671986249, \quad y = 15140424455100.$$

(See [27, Ch. II, §XIII] for an illuminating discussion of Fermat's study of $x^2 - Ny^2 = \pm 1$.) If we set out to examine $x^2 - 109y^2 = 1$ with a computer, we might look for solutions with x and y non-zero, find no such solutions, and

conclude incorrectly that this equation has only the trivial solutions $(-1, 0)$ and $(1, 0)$.

Here's another example: Euler conjectured in the eighteenth century that a perfect fourth power cannot be the sum of three perfect fourth powers. Noam Elkies [10] found the first counterexample to Euler's conjecture in 1988:

$$2682440^4 + 15365639^4 + 18796760^4 = 20615673^4.$$

These examples illustrate the fact that numerical evidence in number theory can be misleading.

Modern History

The proof of Fermat's Last Theorem at the end of the last century hinges on a connection between putative solutions of Fermat's equation and cubic equations with integer coefficients (elliptic curves). To have a solution to Fermat's equation is to have positive integers a and b for which $a^n + b^n$ is a perfect nth power. (We shall suppose that n is at least 5 and that n is a prime number. The results of Fermat and Euler imply that these assumptions are harmless.) Given a and b, we consider the equation

$$E : y^2 = x(x - a^n)(x + b^n),$$

in which x and y are new variables. This equation defines an elliptic curve.

The connection between Fermat and elliptic curves was noticed by several mathematicians, including Yves Hellegouarch and Gerhard Frey. In a recent book [11], Hellegouarch recounts the history of this connection. It was Frey who had the decisive idea that E could not possibly satisfy the *Shimura–Taniyama conjecture*, which states that elliptic curves are *modular*. (We shall discuss this crucial property below.)

Frey's suggestion became known to the mathematical community in the mid 1980s. In 1986, I proved that elliptic curves associated to solutions of Fermat's equation are non-modular, thereby showing that Fermat's Last Theorem is a consequence of the Shimura–Taniyama conjecture [19], [20]. Said differently: each solution to Fermat's Last Theorem gives a counterexample to the Shimura–Taniyama conjecture. Thus if that conjecture is true, so is Fermat's Last Theorem.

As the reader is no doubt aware, Andrew Wiles worked in his Princeton attic from 1986 to 1993 with the goal of establishing the Shimura–Taniyama conjecture. Although the conjecture per se was a central problem of number theory, Wiles has stated that he was drawn to this problem because of the link with Fermat's Last Theorem. In June, 1993, Wiles announced that he could prove the Shimura–Taniyama conjecture for a wide class of elliptic curves, including those coming from Fermat solutions. This announcement implied that the proof of Fermat's Last Theorem was complete.

After a short period of celebration among mathematicians, Wiles's colleague Nicholas Katz at Princeton found a "gap" in Wiles's proof. Because

the gap's severity was not appreciated at first, it was months before the existence of the gap was known widely in the mathematical community. By the end of 1993, however, the fact that Wiles's proof was incomplete was reported in the popular press.

The proof announced by Wiles remained in doubt until October, 1994, when Richard Taylor and Andrew Wiles released a modified version of the proof that circumvented the gap. The new proof was divided into two articles, one by Wiles alone and one a collaboration by Taylor and Wiles [28], [26]. The two articles were published together in 1995. The proof presented in those articles was accepted quickly by the mathematical community.

As a result of his work, Wiles has been honored repeatedly. For example, in December, 1999, he was knighted by the Queen: he received the "KBE/DBE" along with Julie Andrews, Elizabeth Taylor and Duncan Robin Carmichael Christopher, Her Majesty's ambassador to Jakarta[1].

After the manuscripts by Wiles and Taylor–Wiles were written in 1994, the technology for establishing modularity became increasingly more sophisticated and more general. The class of curves to which the technology can be applied was enlarged in three stages [8], [5], [3]. In the last stage, four mathematicians — Christophe Breuil, Brian Conrad, Fred Diamond and Richard Taylor — announced in June, 1999 that they had proved the full Shimura–Taniyama conjecture, i.e., the modularity of all elliptic curves (that are defined by equations with integer coefficients). Although their proof is not yet published, it is available from `http://www.math.harvard.edu/~rtaylor/`, Richard Taylor's Web site at Harvard. In addition, the proof has been the subject of a substantial number of oral presentations. In particular, the proof was explained by the four authors in a series of lectures at a conference held at the Mathematical Sciences Research Institute in Berkeley, California in December, 1999.

The Shimura–Taniyama Conjecture

The conjecture hinges on the notion of "arithmetic mod p," p being a prime number. When working mod p, we ignore all integers that are multiples of p. In other words, when we interact with an integer m, we care only about the remainder when m is divided by p. This remainder is one of the numbers 0, 1, 2,..., $p-1$. For example, the integers mod 5 are 0, 1, 2, 3 and 4.

Suppose that we are given an equation with integer coefficients. Then for each prime number p, we can use the equation to define a relation mod p. As an illustration, the simple equation $x^2 + y^2 = 1$ gives rise to a relation mod 2, mod 3, mod 5, and so on.

This type of relation is best illustrated by a concrete example. Suppose that we take $p = 5$, so that the numbers mod 5 are the five numbers that we listed above. There are thus 25 pairs of numbers (x, y) mod 5. For each pair (x, y), we can ask whether $x^2 + y^2$ is the same as 1 mod 5. For $(0, 4)$, the

[1] `http://files.fco.gov.uk/hons/honsdec99.shtml`

answer is "yes" because 16 and 1 are the same mod 5. For $(2,2)$, the answer is "no" because 8 and 1 are not the same mod 5. After some calculation, one finds that there are four pairs of numbers mod 5 for which the answer is in the affirmative. These pairs are $(0,1)$, $(0,4)$, $(1,0)$ and $(4,0)$. After we recognize that 4 is that same as -1 mod 5, we might notice that the four solutions that we have listed have analogues for every prime number $p > 2$. There are always the four systematic solutions $(0,1)$, $(0,-1)$, $(1,0)$ and $(-1,0)$ for each such prime.

We can make a similar calculation mod 7. It is fruitful to begin by listing the squares of the seven numbers mod 7:

a	0	1	2	3	4	5	6
a^2	0	1	4	2	2	4	1.

In which ways can we write 1 as the sum of two numbers in the bottom row (possibly the sum of a number and itself)? We can write 1 as $0+1=1+0$, and we can also write 1 as $4+4$ (since 8 is the same as 1 mod 7). We end up with the four "new" solutions $(\pm 2, \pm 2)$ in addition to the four systematic solutions that we listed in connection with the case $p = 5$. As a consequence, there are eight solutions to $x^2 + y^2 \equiv 1 \bmod 7$.

After experimenting with other primes ($p = 11$, $p = 13$, etc.), you will have little trouble guessing the general formula for the number of solutions to $x^2 + y^2 \equiv 1 \bmod p$. When $p = 2$, there are the two solutions $(0,1)$ and $(1,0)$. When p is bigger than 2, there are either $p+1$ or $p-1$ solutions to $x^2+y^2 \equiv 1 \bmod p$, depending on whether p is 1 less than or 1 more than a multiple of 4. This simple recipe was known centuries ago. It can be established in various ways; perhaps I should leave its proof as an exercise for the interested reader.

The equation $x^2 + y^2 = 1$ was intended as a warm-up; we shall now consider the superficially analogous equation $x^3 + y^3 = 1$. Here again we study the number of solutions to $x^3 + y^3 \equiv 1 \bmod p$ and seek to understand how this number varies with p. It turns out that the quantity p mod 3 plays an important role here—just as the behavior of p mod 4 was significant for $x^2+y^2 = 1$. When $p = 3$, the quantity x^3 mod p coincides with x mod 3; this is a special case of what is called "Fermat's Little Theorem" in textbooks. Hence the solutions to $x^3 + y^3 \equiv 1 \bmod 3$ are the same as the solutions to $x + y = 1$; there are three solutions, because x can be taken arbitrarily, and then y is $1 - x$ mod 3. If p is 2 mod 3, i.e., if p is 1 less than a multiple of 3, one shows by an elementary argument that there are again p solutions. (If p is 2 mod 3, then every number mod p has a unique cube root.)

The interesting case for this equation is the remaining case where p is 1 more than a multiple of 3. This case was resolved by Gauss in the nineteenth century. To see what is going on, we should look at a few examples:

First of all, let us take $p = 7$. The cubes mod 7 are 0, 1 and $6 \equiv -1$. If two cubes sum to 0, one is 1 and the other is 0. Also, 1 has three cube roots: 1, 2 and 4. Thus there are six solutions to $x^3 + y^3 \equiv 1 \bmod 7$, namely $(0,1)$, $(0,2)$, $(0,4)$ and the analogous pairs with x and y reversed.

When $p = 13$, the cubes are 0, 1, -1, 5 and 8. There are again only six

solutions because the only way to write 1 as a sum of two cubes is to take $0+1$ as before.

Now try $p = 19$. It turns out that there are 24 solutions here—the six that we knew about already, together with 18 unexpected ones arising from the equation $1 = 8+(-7)$ and the fact that 8 and -7 are both cubes mod 19. (Since $4^3 = 64 \equiv 7 \bmod 19$, -7 is the cube of -4.) We get 18 solutions by taking x to be one of the 3 cube roots of 8 and y to be one of the 3 cube roots of -7, or vice versa.

When $p = 31$, there are 33 solutions. (Note that $33 = 6+18+9$.) There are 6 solutions coming from $0+1=1$, 18 coming from $2+(-1)=1$ and 9 from $16+16 \equiv 1$. Summary:

p	7	13	19	31	$\cdots$
# solns.	6	6	24	33	$\cdots$.

How does this table continue? What is the number of solutions that we get when p is, say, 103? It is hard to imagine the rule that expresses the number of solutions in terms of p.

Gauss found an expression for the number of solutions that we can view as a "generalized formula" [12, p. 97]. Namely, when $p \equiv 1 \bmod 3$, Gauss showed that one has

$$4p = A^2 + 27B^2$$

for some integers A and B. These integers are uniquely determined except for their signs. We can and do choose A so that $A \equiv 1 \bmod 3$. Then Gauss's formula states:

$$\#\text{ solns.} = p - 2 + A.$$

For example, if $p = 13$, then $4p = 52 = 5^2 + 27 \cdot 1^2$. Thus $A = -5$. We have $p - 2 + A = 6$.

When $p = 31$, $4p = 124 = 4^2 + 27 \cdot 2^2$. Thus $A = 4$ and $p - 2 + A = 33$.

When $p = 103$, $4p = 13^2 + 27 \cdot 3^2$, so $A = 13$ and the number of solutions is 114.

The equation $x^3 + y^3 = 1$ defines one of the simplest possible elliptic curves. Gauss's explicit recipe shows in particular that $x^3 + y^3 = 1$ defines a *modular* elliptic curve.

The Shimura–Taniyama conjecture states that there's an analogous "formula" for *every* elliptic curve. Because this formula involves modular forms, the Shimura–Taniyama conjecture is usually paraphrased as the statement that elliptic curves are modular.

For a random elliptic curve, the formula provided by the associated modular form is not as explicit as Gauss's formula for $x^3 + y^3 = 1$. Here is a famous example that *begins* to give the flavor of the general case: We consider first the formal power series with integral coefficients $\sum a_n X^n$ that is obtained by expanding out the product

$$X \prod_{m=1}^{\infty} (1 - X^m)^2 (1 - X^{11m})^2.$$

For all $n \geq 1$, a_n is an integer. In fact, the numbers a_n are the coefficients of the Fourier expansion of a well known modular form.

At the same time, we consider the elliptic curve defined by the equation $y^2 + y = x^3 - x^2$. Then a theorem of M. Eichler and G. Shimura states that, for each prime p (different from 11), the number of solutions to this equation mod p is $p - a_p$. The connection between the number of solutions and the pth coefficient of a modular form shows that the elliptic curve defined by $y^2 + y = x^3 - x^2$ is a modular elliptic curve. (A coffee mug that celebrates this relation is currently available from the Mathematical Sciences Research Institute. Go to `http://www.msri.org/search.html` and search for "coffee cup.")

Another Formula of Gauss

For a third example, we look at the elliptic curve defined by the equation $y^2 = x^3 - x$. Although its equation recalls the equation $y^2 + y = x^3 - x^2$ of the second example, this third example is much more analogous to the first example. To explain the analogy, it is important to recall a theorem of Fermat about sums of squares. Namely, suppose that p is a prime number and that we seek to write p is the form $r^2 + s^2$, where r and s are integers. If p is 2, we can write $p = 1^2 + 1^2$. If p is congruent to 3 mod 4, then it is *impossible* to write p as $r^2 + s^2$. Indeed, squares are congruent to either 0 or 1 mod 4; it is therefore impossible that a sum of two squares be congruent to 3 mod 4.

The interesting case is that where p is congruent to 1 mod 4, i.e., where p is 1 plus a multiple of 4. Fermat proved in that case that p may be written as a sum of two squares: we have $p = r^2 + s^2$ with r and s whole numbers. The pair (r, s) is clearly not unique because we can exchange r and s and we can change the signs of either or both of these integers. However, there is no more ambiguity than that: the integers r and s become unique up to sign after we require that r be odd and that s be even. Accordingly, r and s are determined completely if we require that r be odd, that s be even and that both integers be positive.

This theorem of Fermat is proved in most elementary number theory books; see, e.g., [12, Ch. 8] for one proof. (The uniqueness is left as an exercise at the end of the chapter.) A beautiful proof of the existence of r and s, due to D. Zagier, is presented in "Proofs from the Book" [1], a volume celebrating Paul Erdös's idea that there is frequently an optimally beautiful proof of a given proposition in mathematics.

Following Gauss, we will now adjust the sign of r (if necessary) to ensure that the sum $r+s$ is congruent to 1 mod 4. For example, suppose that $p = 5$, so that $(r, s) = (1, 2)$ under the initial choice that has both r and s positive. With this choice, $r + s = 3$ is *not* 1 mod 4. Accordingly, we change the sign of r and put $r = -1$. The sum $r + s$ is then 1, which of course is 1 mod 4. For another example, we take $p = 13$, so that $(r, s) = (3, 2)$ with the initial choice. Here $r + s = 3 + 2 = 5$, which is already 1 mod 4. We therefore leave

r positive in this case. It is perhaps enlightening to tabulate the values of r and s for the first ten primes that are 1 mod 4. In doing so, we write "r_p" instead of "r" and "s_p" instead of "s" to stress that r and s depend on p:

p	5	13	17	29	37	41	53	61	73	89	$\cdots$
r_p	-1	3	1	-5	-1	5	7	-5	-3	5	$\cdots$
s_p	2	2	4	2	6	4	2	6	8	8	$\cdots$.

We return now to $y^2 = x^3 - x$ with the idea of calculating the number of solutions to the mod p congruence defined by this equation. If $p = 2$, there are two solutions: $(0, 0)$ and $(1, 0)$. If p is congruent to 3 mod 4, it turns out that there are exactly p solutions. More precisely, if x is 0, 1 or -1 (i.e., $p - 1$) mod p, then $y = 0$ is the one value of y for which (x, y) is a solution. For each value of x different from 0 and ± 1, there are either two values of y or no values of y for which (x, y) is a solution mod p. (If a non-zero number mod p has a square root, it has exactly two square roots, which are negatives of each other.) An elementary argument shows that if there are two y for a given x, then there are no y for $-x$, and vice versa. The point here is that a non-zero number mod p has a square root mod p if and only if its negative does not; this observation is valid when p is 3 mod 4 but fails to be true when p is 1 mod 4. The end result is that, on average, there is one value of y that works for each x. Thus the number of solutions is p, as was stated.

The interesting case for $y^2 = x^3 - x$ is that where p is congruent to 1 mod 4. We assume now that this is the case. To get a feel for the situation, we can calculate the number of solutions mod 5 and mod 13; these are the first two primes that are 1 mod 4.

Suppose that $p = 5$. The values $x = 0$, $x = 1$ and $x = 4$ make $x^3 - x$ congruent to 0, so that they give rise to exactly one solution each; y must be 0. If $x = 2$, then $x^3 - x$ is congruent to 1, a number that has two square roots mod 5, namely ± 1. Thus $x = 2$ gives rise to two solutions. Similarly, if $x = 3$, then $x^3 - x$ is congruent to 4 mod 5, and 4 has two square roots. Thus $x = 2$ also gives rise to two solutions. As a result, there are seven solutions to $y^2 \equiv x^3 - x \bmod 5$.

Suppose now that $p = 13$. The three values $x = 0, 1, -1$ give rise to a single solution each, as before; in each case, y is again 0. The ten remaining values of x (namely, $x = 2, 3, \ldots, 11$) each give rise either to two or to no solutions: the quantity $x^3 - x$ is non-zero mod p and we have to decide in each case whether or not it is a square (i.e., a number with square roots mod p). The quantities are respectively 6, 11, 8, 3, 2, 11, 10, 5, 2 and 7 mod 13. On the other hand, the non-zero squares mod 13 are 1, 3, 4, 9, 10 and 12. It happens, then, that only two of the numbers x between 2 and 11 are such that $x^3 - x$ is a square. Thus we find — one again — that there are seven solutions to $y^2 \equiv x^3 - x \bmod p$.

One could easily guess from these two examples that there are always seven solutions to $y^2 \equiv x^3 - x \bmod p$ when p is 1 mod 4, but these two examples are misleading.

Theorem 1 (Gauss) *Suppose that p is a prime that is 1 mod 4. Then the number of solutions to $y^2 \equiv x^3 - x \bmod p$ is $p - 2r_p$, where r_p is chosen as above.*

The theorem is compatible with the two examples that we presented. When $p = 5$, we have $r_p = -1$, so that $p - 2r_p = 7$. When $p = 13$, r_p is 3, and $13 - 2 \cdot 3 = 7$. Since $r_{73} = -3$, the number of solutions to $y^2 \equiv x^3 - x \bmod 73$ is $73 + 6 = 79$. Here is an example with a p that is considerably larger than the primes that have appeared thus far: Suppose that p is the prime number 144169. We can write p as the sum $315^2 + 212^2$. It follows that $r_p = \pm 315$. Since $315 + 212 = 527$ is 3 mod 4, we must take $r_p = -315$. Gauss's formula then asserts that the number of solutions to $y^2 \equiv x^3 - x \bmod p$ is $144169 + 2 \cdot 315 = 144799$.

A variant of Gauss's formula is proved in [12, Ch. 11, §8]. The connection between the variant given there and the formula of Theorem 1 is made in Exercise 13 at the end of [12, Ch. 11].

Further Reading

During the course of this article, I have mentioned some of my favorite articles and books about number theory, especially those that touch on Fermat's Last Theorem. Here are a few more references that I have not yet had occasion to cite. First, a summary of "elementary" approaches to Fermat's Last Theorem is provided by P. Ribenboim in his book [18]. Secondly, an interesting discussion of elliptic curves and modular forms is contained in A. van der Poorten's book [14]. Next, the recent "diary" by C. J. Mozzochi [13] contains photos of the mathematicians who participated in the proof of Fermat's Last Theorem, along with detailed descriptions of lectures and other events that are associated strongly with the proof. Finally, several accounts of the details of the proof of Fermat's Last Theorem have been written for professional mathematicians [7], [16], [6]. What is missing from the literature, at least so far, is an extended account of the proof that is accessible to a scientifically literate lay reader and does justice to the mathematics behind the proof.

References

[1] M. Aigner and G. M. Ziegler, *Proofs from the Book.* New York-Berlin-Heidelberg: Springer-Verlag, 1998.

[2] Z. I. Borevich and I. R. Shafarevich, *Number Theory.* New York: Academic Press, 1966.

[3] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q**. To appear.

[4] J. Buhler, R Crandall, R. Ernvall, and T. Metsänkylä, *Irregular primes and cyclotomic invariants to four million*, Math. Comp. **61** (1993), 151–153.

[5] B. Conrad, F. Diamond, and R. Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), 521–567.

[6] G. Cornell, J. H. Silverman and G. Stevens, eds. *Modular forms and Fermat's last theorem*, Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995. Berlin-Heidelberg-New York: Springer-Verlag, 1997.

[7] H. Darmon, F. Diamond and R. L. Taylor, *Fermat's last theorem.* In "Elliptic curves, modular forms & Fermat's last theorem," Proceedings of the Conference on Elliptic Curves and Modular Forms held at the Chinese University of Hong Kong, Hong Kong, December 18–21, 1993, J. Coates and S. T. Yau, eds., second edition. Cambridge, MA: International Press, 1997.

[8] F. Diamond, *On deformation rings and Hecke rings*, Ann. of Math. (2) **144** (1996), 137–166.

[9] H. M. Edwards, *Fermat's Last Theorem: a genetic introduction to modern number theory.* Graduate Texts in Mathematics, volume 50. New York-Berlin-Heidelberg: Springer-Verlag, 1977.

[10] N. Elkies, *On* $A^4 + B^4 + C^4 = D^4$, Math. Comp. **51** (1988), 825–835.

[11] Y. Hellegouarch, *Invitation aux mathématiques de Fermat-Wiles.* Paris: Masson, 1997.

[12] K. F. Ireland and M. I. Rosen, *A Classical Introduction to Modern Number Theory*, second edition. Graduate Texts in Mathematics, volume 84. New York-Berlin-Heidelberg: Springer-Verlag, 1990.

[13] C. J. Mozzochi, *The Fermat Diary.* Providence: American Math. Soc., 2000.

[14] A. van der Poorten, *Notes on Fermat's Last Theorem.* New York: John Wiley & Sons., 1996.

[15] J. Lynch and S. Singh, *The Proof.* A television documentary written and produced by John Lynch and directed by Simon Singh. See `http://www.pbs.org/wgbh/nova/proof/` for more information, and for a transcript.

[16] V. K. Murty, ed., *Seminar on Fermat's Last Theorem*, Papers from the seminar held at the Fields Institute for Research in Mathematical Sciences, Toronto, Ontario, 1993–1994. Providence, American Math. Soc., 1995.

[17] P. Ribenboim, *13 Lectures on Fermat's Last Theorem.* New York-Berlin-Heidelberg: Springer-Verlag, 1979.

[18] P. Ribenboim, *Fermat's Last Theorem for Amateurs.* New York-Berlin-Heidelberg: Springer-Verlag, 1999.

[19] K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), 431–476.

[20] K. A. Ribet, *From the Taniyama-Shimura conjecture to Fermat's last theorem*, Ann. Fac. Sci. Toulouse Math. (5) **11** (1990), 116–139.

[21] K. A. Ribet, *Galois representations and modular forms*, Bull. Amer. Math. Soc. (N.S.) **32** (1995), 375–402.

[22] K. A. Ribet, *Squares mod p and the Golden Theorem*, Mathematical Medley, **27 (2)** (2000), 69–75.

[23] K. A. Ribet and B. Hayes, *Fermat's Last Theorem and modern arithmetic*, American Scientist (March–April, 1994), 144–156.

[24] S. Singh, *Fermat's Enigma: The epic quest to solve the world's greatest mathematical problem, with a foreword by John Lynch.* New York: Walker and Co., 1997.

[25] S. Singh and K. A. Ribet, *Fermat's Last Stand*, Scientific American **227** (1997), 68–73.

[26] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Annals of Math. **141** (1995), 553–572.

[27] A. Weil, *Number Theory: An approach through history, From Hammurapi to Legendre.* Boston, Mass: Birkhäuser Boston 1984.

[28] A. Wiles, *Modular elliptic curves and Fermat's Last Theorem*, Annals of Math. **141** (1995), 443–551.

Editor's note The above article has appeared as part of the author's paper *Modular Forms and Diophantine Questions* in "Challenges for the 21st Century", World Scientific, 2001. Reprinted herein with kind permission of World Scientific.