Analytic Number Theory

Andrew Granville

1 Introduction

What is number theory? One might have thought that it was simply the study of numbers, but that is too broad a definition, since numbers are almost ubiquitous in mathematics. To see what distinguishes number theory from the rest of mathematics, let us look at the equation $x^2 + y^2 = 15\,925$, and consider whether it has any solutions. One answer is that it certainly does: indeed, the solution set forms a circle of radius $\sqrt{15\,925}$ in the plane. However, a number theorist is interested in *integer* solutions, and now it is much less obvious whether any such solutions exist.

A useful first step in considering the above question is to notice that 15 925 is a multiple of 25: in fact, it is 25×637. Furthermore, the number 637 can be decomposed further: it is 49×13. That is, $15\,925 = 5^2 \times 7^2 \times 13$. This information helps us a lot, because if we can find integers a and b such that $a^2 + b^2 = 13$, then we can multiply them by $5 \times 7 = 35$ and we will have a solution to the original equation. Now we notice that $a = 2$ and $b = 3$ works, since $2^2 + 3^2 = 13$. Multiplying these numbers by 35, we obtain the solution $70^2 + 105^2 = 15\,925$ to the original equation.

As this simple example shows, it is often useful to decompose positive integers multiplicatively into components that cannot be broken down any further. These components are called *prime numbers*, and the FUNDAMENTAL THEOREM OF ARITHMETIC states that every positive integer can be written as a product of primes in exactly one way. That is, there is a one-to-one correspondence between positive integers and finite products of primes. In many situations we know what we need to know about a positive integer once we have decomposed it into its prime factors and understood those, just as we can understand a lot about molecules by studying the atoms of which they are composed. For example, it is known that the equation $x^2 + y^2 = n$ has an integer solution if and only if every prime of the form $4m + 3$ occurs an even number of times in the prime factorization of n. (This tells us, for instance, that there are no integer solutions to the equation $x^2 + y^2 = 13\,475$, since $13\,475 = 5^2 \times 7^2 \times 11$, and 11 appears an odd number of times in this product.)

Once one begins the process of determining which integers are primes and which are not, it is soon apparent that there are many primes. However, as one goes further and further, the primes seem to consist of a smaller and smaller proportion of the positive integers. They also seem to come in a somewhat irregular pattern, which raises the question of whether there is any formula that describes all of them. Failing that, can one perhaps describe a large class of them? We can also ask whether there are infinitely many primes? If there are, can we quickly determine how many there are up to a given point? Or at least give a good estimate for this number? Finally, when one has spent long enough looking for primes, one cannot help but ask whether there is a quick way of recognizing them. This last question is discussed in COMPUTATIONAL NUMBER THEORY; the rest motivate this article.

Now that we have discussed what marks number theory out from the rest of mathematics, we are ready to make a further distinction: between *algebraic* and *analytic* number theory. The main difference is that in algebraic number theory (which is the main topic of ALGEBRAIC NUMBERS) one typically considers questions with answers that are given by exact formulas, whereas in analytic number theory, the topic of this article, one looks for *good approximations*. For the sort of quantity that one estimates in analytic number theory, one does not expect an exact formula to exist, except perhaps one of a rather artificial and unilluminating kind. One of the best examples of such a quantity is one we shall discuss in detail: the number of primes less than or equal to x.

Since we are discussing approximations, we shall need terminology that allows us to give some idea of the quality of an approximation. Suppose, for example, that we have a rather erratic function $f(x)$ but are able to show that, once x is large enough, $f(x)$ is never bigger than $25x^2$. This is useful because we understand the function $g(x) = x^2$ quite well. In general, if we can find a constant c such that $|f(x)| \leqslant cg(x)$ for every x, then we write $f(x) = O(g(x))$. A typical usage occurs in the sentence "the average number of prime factors of an integer up to x is $\log\log x + O(1)$"; in other words, there exists some constant $c > 0$ such

that $|\text{the average} - \log\log x| \leqslant c$ once x is sufficiently large.

We write $f(x) \sim g(x)$ if $\lim_{x\to\infty} f(x)/g(x) = 1$; and also $f(x) \approx g(x)$ when we are being a little less precise, that is, when we want to say that $f(x)$ and $g(x)$ come close when x is sufficiently large, but we cannot be, or do not want to be, more specific about what we mean by "come close."

It is convenient for us to use the notation $\sum$ for sums and $\prod$ for product. Typically we will indicate beneath the symbol what terms the sum, or product, is to be taken over. For example, $\sum_{m\geqslant 2}$ will be a sum over all integers m that are greater than or equal to 2, whereas $\prod_{p \text{ prime}}$ will be a product over all primes p.

2 Bounds for the Number of Primes

Ancient Greek mathematicians knew that there are infinitely many primes. Their beautiful proof by contradiction goes as follows. Suppose that there are only finitely many primes, say k of them, which we will denote by $p_1, p_2, \dots, p_k$. What are the prime factors of $p_1p_2\cdots p_k + 1$? Since this number is greater than 1 it must have at least one prime factor, and this must be p_j for some j (since *all* primes are contained amongst $p_1, p_2, \dots, p_k$). But then p_j divides both $p_1p_2\cdots p_k$ and $p_1p_2\cdots p_k+1$, and hence their difference, 1, which is impossible.

Many people dislike this proof, since it does not actually exhibit infinitely many primes: it merely shows that there cannot be finitely many. It is more or less possible to correct this deficiency by defining the sequence $x_1 = 2$, $x_2 = 3$ and then $x_{k+1} = x_1x_2\cdots x_k+1$ for each $k \geqslant 2$. Then each x_k must contain at least one prime factor, q_k say, and these prime factors must be distinct, since if $k < \ell$, then q_k divides x_k which divides $x_\ell - 1$, while q_ℓ divides x_ℓ. This gives us an infinite sequence of primes.

In the seventeenth century EULER gave a different proof that there are infinitely many primes, one that turned out to be highly influential in what was to come later. Suppose again that the list of primes is $p_1, p_2, \dots, p_k$. As we have mentioned, the fundamental theorem of arithmetic implies that there is a one-to-one correspondence between the set of all integers and the set of products of the primes, which, if those are the only primes, is the set $\{p_1^{a_1}p_2^{a_2}\cdots p_k^{a_k} : a_1, a_2, \dots, a_k \geqslant 0\}$. But, as Euler observed, this implies that a sum involving the elements of the first set should equal the analogous sum involving the elements of the second set:

$$\begin{aligned}
&\sum_{\substack{n\geqslant 1\\ n \text{ a positive integer}}} \frac{1}{n^s}\\
&= \sum_{a_1,a_2,\dots,a_k\geqslant 0} \frac{1}{(p_1^{a_1}p_2^{a_2}\cdots p_k^{a_k})^s}\\
&= \Big(\sum_{a_1\geqslant 0} \frac{1}{(p_1^{a_1})^s}\Big)\Big(\sum_{a_2\geqslant 0} \frac{1}{(p_2^{a_2})^s}\Big)\cdots\Big(\sum_{a_k\geqslant 0} \frac{1}{(p_k^{a_k})^s}\Big)\\
&= \prod_{j=1}^{k}\Big(1 - \frac{1}{{p_j}^s}\Big)^{-1}.
\end{aligned}$$

The last equality holds because each sum in the second-last line is the sum of a geometric progression. Euler then noted that if we take $s = 1$, the right-hand side equals some rational number (since each $p_j > 1$) whereas the left-hand side equals ∞. This is a contradiction, so there cannot be finitely many primes. (To see why the left-hand side is infinite when $s = 1$, note that $(1/n) \geqslant \int_n^{n+1}(1/t)\,\mathrm{d}t$ since the function $1/t$ is decreasing, and therefore $\sum_{n=1}^{N-1}(1/n) \geqslant \int_1^N (1/t)\,\mathrm{d}t = \log N$ which tends to ∞ as $N \to \infty$.)

During the proof above, we gave a formula for $\sum n^{-s}$ under the false assumption that there are only finitely many primes. To correct it, all we have to do is rewrite it in the obvious way without that assumption:

$$\sum_{\substack{n\geqslant 1\\ n \text{ a positive integer}}} \frac{1}{n^s} = \prod_{p \text{ prime}} \Big(1 - \frac{1}{p^s}\Big)^{-1}. \quad (1)$$

Now, however, we need to be a little careful about whether the two sides of the formula converge. It is safe to write down such a formula when both sides are absolutely convergent, and this is true when $s > 1$. (An infinite sum or product is *absolutely convergent* if the value does not change when we take the terms in any order we want.)

Like Euler, we want to be able to interpret what happens to (1) when $s = 1$. Since both sides converge and are equal when $s > 1$, the natural thing to do is consider their common limit as s tends to 1 from above. To do this we note, as above, that the left-hand side of (1) is well approximated by

$$\int_1^\infty \frac{\mathrm{d}t}{t^s} = \frac{1}{s-1},$$

so it diverges as $s \to 1^+$. We deduce that

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p}\right) = 0. \tag{2}$$

Upon taking logarithms and discarding negligible terms, this implies that

$$\sum_{p \text{ prime}} \frac{1}{p} = \infty. \tag{3}$$

So how numerous are the primes? One way to get an idea is to determine the behaviour of the sum analogous to (3) for other sequences of integers. For instance, $\sum_{n \geqslant 1} 1/n^2$ converges, so the primes are, in this sense, more numerous than the squares. This argument works if we replace the power 2 by any $s > 1$, since then, as we have just observed, the sum $\sum_{n \geqslant 1} 1/n^s$ is about $1/(s-1)$ and in particular converges. In fact, since $\sum_{n \geqslant 1} 1/n(\log n)^2$ converges, we see that the primes are in the same sense more numerous than the numbers $\{n(\log n)^2 : n \geqslant 1\}$, and hence there are infinitely many integers x for which the number of primes less than or equal to x is at least $x/(\log x)^2$.

Thus, there seem to be primes in abundance, but we would also like to verify our observations, made from calculations, that the primes constitute a smaller and smaller proportion of the integers as the integers become larger and larger. The easiest way to see this is to try to count the primes using the "sieve of Eratosthenes." In the sieve of Eratosthenes one starts with all the positive integers up to some number x. From these, one deletes the numbers 4, 6, 8 and so on—that is, all multiples of 2 apart from 2 itself. One then takes the first undeleted integer greater than 2, which is 3, and deletes all *its* multiples—again, not including the number 3 itself. Then one removes all multiples of 5 apart from 5, and so on. By the end of this process, one is left with the primes up to x.

This suggests a way to guess at how many there are. After deleting every second integer up to x other than 2 (which we call "sieving by 2") one is left with roughly half the integers up to x; after sieving by 3, one is left with roughly two-thirds of those that had remained; continuing like this we expect to have about

$$x \prod_{p \leqslant y} \left(1 - \frac{1}{p}\right) \tag{4}$$

integers left by the time we have sieved with all the primes up to y. Once $y = \sqrt{x}$ the undeleted integers are 1 and the primes up to x, since every composite has a prime factor no bigger than its square root. So, is (4) a good approximation for the number of primes up to x when $y = \sqrt{x}$?

To answer this question, we need to be more precise about what the formula in (4) is estimating. It is supposed to approximate the number of integers up to x that have no prime factors less than or equal to y, plus the number of primes up to y. The so-called *inclusion–exclusion principle* can be used to show that the approximation given in (4) is accurate to within 2^k, where k is the number of primes less than or equal to y. Unless k is very small, this error term of 2^k is far larger than the quantity we are trying to estimate, and the approximation is useless. It is quite good if k is less than a small constant times $\log x$, but, as we have seen, this is far less than the number of primes we expect up to y if $y \approx \sqrt{x}$. Thus it is not clear whether (4) can be used to obtain a good estimate for the number of primes up to x. What we *can* do, however, is use this argument to give an upper bound for the number of primes up to x, since the number of primes up to x is never more than the number of integers up to x that are free of prime factors less than or equal to y, plus the number of primes up to y, which is no more than 2^k plus the expression in (4).

Now, by (2), we know that as y gets larger and larger the product $\prod_{p \leqslant y}(1 - 1/p)$ converges to zero. Therefore, for any small positive number ε we can find a y such that $\prod_{p \leqslant y}(1 - 1/p) < \varepsilon/2$. Since every term in this product is at least $1/2$, the product is at least $1/2^k$. Hence, for any $x \geqslant 2^{2k}$ our error term, 2^k, is no bigger than the quantity in (4), and therefore the number of primes up to x is no larger than twice (4), which, by our choice of y, is less than εx. Since we were free to make ε as small as we liked, the primes are indeed a vanishing proportion of all the integers, as we predicted.

Even though the error term in the inclusion–exclusion principle is too large for us to use that method to estimate (4) when $y = \sqrt{x}$, we can still hope that (4) is a good approximation for the number of primes up to x: perhaps a different argument would give us a much smaller error term. And this turns out to be the case: in fact, the error never gets much bigger than (4). However, when $y = \sqrt{x}$ the number of primes up to x is actually about 8/9 times (4). So why does (4) not give a good approximation? After sieving

with prime p we supposed that roughly 1 in every p of the remaining integers were deleted: a careful analysis yields that this can be justified when p is small, but that this becomes an increasingly poor approximation of what really happens for larger p; in fact (4) *does not* give a correct approximation once y is bigger than a fixed power of x. So what goes wrong? In the hope that the proportion is roughly $1/p$ lies the unspoken assumption that the consequences of sieving by p are independent of what happened with the primes smaller than p. But if the primes under consideration are no longer small, then this assumption is false. This is one of the main reasons that it is hard to estimate the number of primes up to x, and indeed similar difficulties lie at the heart of many related problems.

One can refine the bounds given above but they do not seem to yield an asymptotic estimate for the primes (that is, an estimate which is correct to within a factor that tends to 1 as x gets large). The first good guesses for such an estimate emerged at the beginning of the nineteenth century, none better than what emerges from Gauss's observation, made when studying tables of primes up to three million, at 16 years of age, that "the density of primes at around x is about $1/\log x$." Interpreting this, we guess that the number of primes up to x is about

$$\sum_{n=2}^{x} \frac{1}{\log n} \approx \int_2^x \frac{\mathrm{d}t}{\log t}.$$

Let us compare this prediction (rounded to the nearest integer) with the latest data on numbers of primes, discovered by a mixture of ingenuity and computational power. Table 1 shows the actual numbers of primes up to various powers of 10 together with the difference between these numbers and what Gauss's formula gives. The differences are far smaller than the numbers themselves, so his prediction is amazingly accurate. It does seem always to be an overcount, but since the width of the last column is about half that of the central one it appears that the difference is something like $\sqrt{x}$.

In the 1930s, the great probability theorist, Cramér, gave a probabilistic way of interpreting Gauss's prediction. We can represent the primes as a sequence of 0s and 1s: Putting a "1" each time we encounter a prime, and a "0" otherwise, we obtain, starting from 3, the sequence $1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, \ldots$. Cramér's idea is to suppose that this sequence, which represents the primes, has the same properties as a "typical" sequence of 0s and 1s, and to use this principle to make precise conjectures about the primes. More precisely, let $X_3, X_4, \ldots$ be an infinite sequence of RANDOM VARIABLES taking the values 0 or 1, and let the variable X_n equal 1 with probability $1/\log n$ (so that it equals 0 with probability $1 - 1/\log n$). Assume also that the variables are independent, so for each m knowledge about the variables other than X_m tells us nothing about X_m itself. Cramér's suggestion was that any statement about the distribution of 1s in the sequence that represents the primes will be true if and only if it is true with probability 1 for his random sequences. Some care is needed in interpreting this statement: for example, with probability 1 a random sequence will contain infinitely many even numbers. However, it is possible to formulate a general principle that takes account of such examples.

Table 1 Primes up to various x, and the overcount in Gauss's prediction.

x	$\pi(x) = \#\{\text{primes} \leqslant x\}$	Overcount: $\int_2^x \frac{\mathrm{d}t}{\log t} - \pi(x)$
10^8	5 761 455	753
10^9	50 847 534	1 700
10^{10}	455 052 511	3 103
10^{11}	4 118 054 813	11 587
10^{12}	37 607 912 018	38 262
10^{13}	346 065 536 839	108 970
10^{14}	3 204 941 750 802	314 889
10^{15}	29 844 570 422 669	1 052 618
10^{16}	279 238 341 033 925	3 214 631
10^{17}	2 623 557 157 654 233	7 956 588
10^{18}	24 739 954 287 740 860	21 949 554
10^{19}	234 057 667 276 344 607	99 877 774
10^{20}	2 220 819 602 560 918 840	222 744 643
10^{21}	21 127 269 486 018 731 928	597 394 253
10^{22}	201 467 286 689 315 906 290	1 932 355 207

Here is an example of a use of the Gauss–Cramér model. With the help of the CENTRAL LIMIT THEOREM one can prove that, with probability 1, there are

$$\int_2^x \frac{\mathrm{d}t}{\log t} + O(\sqrt{x}\log x)$$

1s among the first x terms in our sequence. The model tells us that the same should be true of the sequence

representing primes, and so we predict that

$$\#\{\text{primes up to } x\} = \int_2^x \frac{\mathrm{d}t}{\log t} + O(\sqrt{x}\log x), \quad (5)$$

just as the table suggests.

The Gauss–Cramér model provides a beautiful way to think about distribution questions concerning the prime numbers, but it does not give proofs, and it does not seem likely that it can be made into such a tool; so for proofs we must look elsewhere. In analytic number theory one attempts to count objects that appear naturally in arithmetic, yet which resist being counted easily. So far, our discussion of the primes has concentrated on upper and lower bounds that follow from their basic definition and a few elementary properties—notably the fundamental theorem of arithmetic. Some of these bounds are good and some not so good. To improve on these bounds we shall do something that seems unnatural at first, and reformulate our question as a question about complex functions. This will allow us to draw on deep tools from analysis.

3 The "Analysis" in Analytic Number Theory

These analytic techniques were born in an 1859 memoir of RIEMANN, in which he looked at the function that appears in the formula (1) of Euler, but with one crucial difference: now he considered *complex* values of s. To be precise, he defined what we now call the *Riemann zeta function* as follows:

$$\zeta(s) := \sum_{n\geqslant 1} \frac{1}{n^s}.$$

It can be shown quite easily that this sum converges whenever the real part of s is greater than 1, as we have already seen in the case of real s. However, one of the great advantages of allowing complex values of s is that the resulting function is holomorphic, and we can use a process of analytic continuation (these terms are discussed in Section **??** of SOME FUNDAMENTAL MATHEMATICAL DEFINITIONS) to make sense of $\zeta(s)$ for every s apart from 1. (A similar but more elementary example of this phenomenon is the infinite series $\sum_{n\geqslant 0} z^n$, which converges if and only if $|z|<1$. However, when it does converge, it equals $1/(1-z)$, and this formula defines a holomorphic function that is defined everywhere except $z=1$.) Riemann proved the remarkable fact that confirming Gauss's conjecture for the number of primes up to x is equivalent to gaining a good understanding of the zeros of the function $\zeta(s)$—that is, of the values of s for which $\zeta(s)=0$. Riemann's deep work gave birth to our subject, so it seems worthwhile to at least sketch the key steps in the argument linking these seemingly unconnected topics.

Riemann's starting point was Euler's formula (1). It is not hard to prove that this formula is valid when s is complex, as long as its real part is greater than 1, so we have

$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

If we take the logarithm of both sides and then differentiate, we obtain the equation

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{p \text{ prime}} \frac{\log p}{p^s - 1} = \sum_{p \text{ prime}} \sum_{m\geqslant 1} \frac{\log p}{p^{ms}}.$$

We need some way to distinguish between primes $p \leqslant x$ and primes $p > x$; that is, we want to count those primes p for which $x/p \geqslant 1$, but not those with $x/p < 1$. This can be done using the *step function* that takes the value 0 for $y<1$ and the value 1 for $y>1$ (so that its graph looks like a step). At $y=1$, the point of discontinuity, it is convenient to give the function the average value, $\frac{1}{2}$. Perron's formula, one of the big tools of analytic number theory. describes this step function by an integral, as follows. For any $c>0$,

$$\frac{1}{2\pi\mathrm{i}} \int_{s:\mathrm{Re}(s)=c} \frac{y^s}{s}\,\mathrm{d}s = \begin{cases} 0 & \text{if } 0<y<1, \\ \frac{1}{2} & \text{if } y=1, \\ 1 & \text{if } y>1. \end{cases}$$

The integral is a *path integral* along a vertical line in the complex plane: the line consisting of all points $c+\mathrm{i}t$ with $t\in\mathbb{R}$. We apply Perron's formula with $y = x/p^m$, so that we count the term corresponding to p^m when $p^m < x$, but not when $p^m > x$. To avoid the "$\frac{1}{2}$," assume that x is not a prime power. In that case we obtain

$$\begin{aligned} &\sum_{\substack{p^m \leqslant x \\ p \text{ prime},\ m\geqslant 1}} \log p \\ &\quad = \frac{1}{2\pi\mathrm{i}} \sum_{p \text{ prime},\ m\geqslant 1} \log p \int_{s:\mathrm{Re}(s)=c} \left(\frac{x}{p^m}\right)^s \frac{\mathrm{d}s}{s} \\ &\quad = -\frac{1}{2\pi\mathrm{i}} \int_{s:\mathrm{Re}(s)=c} \frac{\zeta'(s)}{\zeta(s)} \frac{x^s}{s}\,\mathrm{d}s. \end{aligned} \quad (6)$$

We can justify swapping the order of the sum and the integral if c is taken large enough, since everything then converges absolutely. Now the left-hand side of the above equation is not counting the number of primes up to x but rather a "weighted" version: for each prime p we add a weight of $\log p$ to the count. It turns out, though, that Gauss's prediction for the number of primes up to x follows so long as we can show that x is a good estimate for this weighted count when x is large. Notice that the sum in (6) is exactly the logarithm of the lowest common multiple of the integers less than or equal to x, which perhaps explains why this weighted counting function for the primes is a natural function to consider. Another explanation is that if the density of primes near p is indeed about $1/\log p$, then multiplying by a weight of $\log p$ makes the density everywhere about 1.

If you know some complex analysis, then you will know that *Cauchy's residue theorem* allows one to evaluate the integral in (6) in terms of the "residues" of the integrand $(\zeta'(s)/\zeta(s))(x^s/s)$, that is, the poles of this function. Moreover, for any function f that is analytic except perhaps at finitely many points, the poles of $f'(s)/f(s)$ are the zeros and poles of f. Each pole of $f'(s)/f(s)$ has order 1, and the residue is simply the order of the corresponding zero, or minus the order of the corresponding pole, of f. Using these facts we can obtain the *explicit formula*

$$\sum_{\substack{p \text{ prime},\ m\geqslant 1\\ p^m\leqslant x}} \log p = x - \sum_{\rho:\zeta(\rho)=0} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)}. \qquad (7)$$

Here the zeros of $\zeta(s)$ are counted with multiplicity: that is, if ρ is a zero of $\zeta(s)$ of order k, then there are k terms for ρ in the sum. It is astonishing that there can be such a formula, an exact expression for the number of primes up to x in terms of the zeros of a complicated function: you can see why Riemann's work stretched people's imagination and had such an impact.

Riemann made another surprising observation which allows us to easily determine the values of $\zeta(s)$ on the left-hand side of the complex plane (where the function is not naturally defined). The idea is to multiply $\zeta(s)$ by some simple function so that the resulting product $\xi(s)$ satisfies the *functional equation*

$$\xi(s) = \xi(1-s) \quad \text{for all } s. \qquad (8)$$

He determined that this can be done by taking $\xi(s) := \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(\frac{1}{2}s)\zeta(s)$. Here $\Gamma(s)$ is the famous GAMMA FUNCTION, which equals the factorial function at positive integers (that is, $\Gamma(n) = (n-1)!$), and is well-defined and continuous for all other s.

A careful analysis of (1) reveals that there are no zeros of $\zeta(s)$ with $\mathrm{Re}(s) > 1$. Then, with the help of (8), we can deduce that the only zeros of $\zeta(s)$ with $\mathrm{Re}(s) < 0$ lie at the negative even integers $-2, -4, \ldots$ (the "trivial zeros"). So, to be able to use (7), we need to determine the zeros inside the *critical strip*, the set of all s such that $0 \leqslant \mathrm{Re}(s) \leqslant 1$. Here Riemann made yet another extraordinary observation which, if true, would allow us tremendous insight into virtually every aspect of the distribution of primes.

The Riemann hypothesis. If $0 \leqslant \mathrm{Re}(s) \leqslant 1$ and $\zeta(s) = 0$, then $\mathrm{Re}(s) = \frac{1}{2}$.

It is known that there are infinitely many zeros on the line $\mathrm{Re}(s) = \frac{1}{2}$, crowding closer and closer together as we go up the line. The Riemann hypothesis has been verified computationally for the ten billion zeros of lowest height (that is, with $|\mathrm{Im}(s)|$ smallest), it can be shown to hold for at least 40% of all zeros, and it fits nicely with many different heuristic assertions about the distribution of primes and other sequences. Yet, for all that, it remains an unproved hypothesis, perhaps the most famous and tantalizing in all of mathematics.

How did Riemann think of his "hypothesis"? Riemann's memoir gives no hint as to how he came up with such an extraordinary conjecture, and for a long time afterwards it was held up as an example of the great heights to which humankind could ascend by pure thought alone. However, in the 1920s Siegel and WEIL got hold of Riemann's unpublished notes and from these it is evident that Riemann had been able to determine the lowest few zeros to several decimal places through extensive hand calculations—so much for "pure thought alone"! Nevertheless, the Riemann hypothesis is a mammoth leap of imagination and to have come up with an algorithm to calculate zeros of $\zeta(s)$ is a remarkable achievement. (See COMPUTATIONAL NUMBER THEORY for a discussion of how zeros of $\zeta(s)$ can be calculated.)

If the Riemann hypothesis is true, then it is not hard to prove the bound

$$\left|\frac{x^\rho}{\rho}\right| \leqslant \frac{x^{1/2}}{|\mathrm{Im}(\rho)|}.$$

Inserting this into (7) one can deduce that

$$\sum_{\substack{p \text{ prime} \\ p \leqslant x}} \log p = x + O(\sqrt{x} \log^2 x); \tag{9}$$

which, in turn, can be "translated" into (5). In fact these estimates hold if and only if the Riemann hypothesis is true.

The Riemann hypothesis is not an easy thing to understand, nor to fully appreciate. The equivalent, (5), is perhaps easier. Another version, which I prefer, is that, for every $N \geqslant 100$,

$$|\log(\mathrm{lcm}[1, 2, \ldots, N]) - N| \leqslant 2\sqrt{N} \log N.$$

To focus on the overcount in Gauss's guesstimate for the number of primes up to x, we use the following approximation, which can be deduced from (7) if, and only if, the Riemann hypothesis is true:

$$\frac{\int_2^x \frac{\mathrm{d}t}{\log t} - \#\{\text{primes} \leqslant x\}}{\sqrt{x}/\log x} \approx 1 + 2 \sum_{\substack{\text{all real numbers } \gamma > 0 \\ \text{such that } \frac{1}{2} + \mathrm{i}\gamma \\ \text{is a zero of } \zeta(s)}} \frac{\sin(\gamma \log x)}{\gamma}. \tag{10}$$

The right-hand side here is the overcount in Gauss's prediction for the number of primes up to x, divided by something that grows like $\sqrt{x}$. When we looked at the table of primes it seemed that this quantity should be roughly constant. However, that is not quite true as we see upon examining the right-hand side. The first term on the right-hand side, the "1", corresponds to the contribution of the squares of the primes in (7). The subsequent terms correspond to the terms involving the zeros of $\zeta(s)$ in (7); these terms have denominator γ so the most significant terms in this sum are those with the smallest values of γ. Moreover, each of these terms is a sine wave, which oscillates, half the time positive and half the time negative. Having the "$\log x$" in there means that these oscillations happen slowly (which is why we hardly notice them in the table above), but they do happen, and indeed the quantity in (10) does eventually get negative. No one has yet determined a value of x for which this is negative (that is, a value of x for which there are more than $\int_2^x (1/\log t)\,\mathrm{d}t$ primes up to x), though our best guess is that the first time this happens is for

$$x \approx 1.398 \times 10^{316}.$$

How does one arrive at such a guess given that the table of primes extends only up to 10^{22}? One begins by using the first thousand terms of the right-hand side of (10) to approximate the left-hand side; wherever it looks as though it could be negative, one approximates with more terms, maybe a million, until one becomes pretty certain that the value is indeed negative.

It is not uncommon to try to understand a given function better by representing it as a sum of sines and cosines like this; indeed this is how one studies the harmonics in music and (10) becomes quite compelling from this perspective. Some experts suggest that (10) tells us that "the primes have music in them" and thus makes the Riemann hypothesis believable, even desirable.

To prove unconditionally that

$$\#\{\text{primes} \leqslant x\} \sim \int_2^x \frac{\mathrm{d}t}{\log t},$$

the so-called "prime number theorem," we can take the same approach as above but, since we are not asking for such a strong approximation to the number of primes up to x, we need to show only that the zeros near to the line $\mathrm{Re}(s) = 1$ do not contribute much to the formula (7). By the end of the nineteenth century this task had been reduced to showing that there are no zeros actually *on* the line $\mathrm{Re}(s) = 1$: this was eventually established by DE LA VALLÉE POUSSIN and HADAMARD in 1896.

Subsequent research has provided wider and wider subregions of the critical strip without zeros of $\zeta(s)$ (and thus improved approximations to the number of primes up to x), without coming anywhere near to proving the Riemann hypothesis. This remains as an outstanding open problem of mathematics.

A simple question like "How many primes are there up to x?" deserves a simple answer, one that uses elementary methods rather than all of these methods of complex analysis, which seem far from the question at hand. However, (7) tells us that the prime number theorem is true *if and only if* there are no zeros of $\zeta(s)$ on the line $\mathrm{Re}(s) = 1$, and so one might argue that it is inevitable that complex analysis must be involved in such a proof. In 1949 Selberg and Erdős surprised the mathematical world by giving an elementary proof of the prime number theorem. Here, the word "elementary" does not mean "easy" but merely that the proof does not use advanced tools such as complex

analysis—in fact, their argument is a complicated one. Of course their proof must somehow show that there is no zero on the line $\mathrm{Re}(s) = 1$, and indeed their combinatorics cunningly masks a subtle complex analysis proof beneath the surface (read Ingham's discussion (1949) for a careful examination of the argument).

4 Primes in Arithmetic Progressions

After giving good estimates for the number of primes up to x, which from now on we shall denote by $\pi(x)$, we might ask for the number of such primes that are congruent to a mod q. (MODULAR ARITHMETIC is discussed in Part III.) Let us write $\pi(x; q, a)$ for this quantity. To start with, note that there is only one prime congruent to 2 mod 4, and indeed there can be no more than one prime in any arithmetic $a, a+q, a+2q, \dots$ if a and q have a common factor greater than 1. Let $\phi(q)$ denote the number of integers a, $1 \leqslant a \leqslant q$, such that $(a, q) = 1$. (The notation (a, q) stands for the highest common factor of a and q.) Then all but a small finite number of the infinitely many primes belong to the $\phi(q)$ arithmetic progressions $a, a+q, a+2q, \dots$ with $1 \leqslant a < q$ and $(a, q) = 1$. Calculation reveals that the primes seem to be pretty evenly split between these $\phi(q)$ arithmetic progressions, so we might guess that in the limit the proportion of primes in each of them is $1/\phi(q)$. That is, whenever $(a, q) = 1$, we might conjecture that, as $x \to \infty$,

$$\pi(x; q, a) \sim \frac{\pi(x)}{\phi(q)}. \tag{11}$$

It is far from obvious even that the number of primes congruent to a mod q is infinite. This is a famous theorem of DIRICHLET. To begin to consider such questions we need a systematic way to identify integers n that are congruent to a mod q, and this Dirichlet provided by introducing a class of functions now known as *(Dirichlet) characters*. Formally, a *character* mod q is a function χ from $\mathbb{Z}$ to $\mathbb{C}$ with the following three properties (in ascending order of interest):

(i) $\chi(n) = 0$ whenever n and q have a common factor greater than 1;
(ii) χ is *periodic* mod q—that is, $\chi(n+q) = \chi(n)$ for every integer n;
(iii) χ is *multiplicative*—that is, $\chi(mn) = \chi(m)\chi(n)$ for any two integers m and n.

An easy but important example of a character mod q is the *principal character* χ_q, which takes the value 1 if $(n, q) = 1$ and 0 otherwise. If q is prime, then another important example is the *Legendre symbol* $(\cdot/q)$: one sets (n/q) to be 0 if n is a multiple of q, 1 if n is a quadratic residue mod q, and -1 if n is a quadratic nonresidue mod q. (An integer n is called a *quadratic residue* mod q if n is congruent mod q to a perfect square.) If q is composite, then a function known as the *Legendre–Jacobi symbol* $(\cdot/q)$, which generalizes the Legendre symbol, is also a character. This too is an important example that helps us, in a slightly less direct way, to recognize squares mod q.

These characters are all real-valued, which is the exception rather than the rule. Here is an example of a genuinely complex-valued character in the case $q = 5$. Set $\chi(n)$ to be 0 if $n \equiv 0 \pmod 5$, i if $n \equiv 2$, -1 if $n \equiv 4$, $-\mathrm{i}$ if $n \equiv 3$, and 1 if $n \equiv 1$. To see that this is a character, note that the powers of 2 mod 5 are $2, 4, 3, 1, 2, 4, 3, 1, \dots$, while the powers of i are $\mathrm{i}, -1, -\mathrm{i}, 1, \mathrm{i}, -1, -\mathrm{i}, 1, \dots$.

It can be shown that there are precisely $\phi(q)$ distinct characters mod q. Their usefulness to us comes from the properties above, together with the following formula, in which the sum is over all characters mod q and $\bar{\chi}(a)$ denotes the complex conjugate of $\chi(a)$:

$$\frac{1}{\phi(q)} \sum_{\chi} \bar{\chi}(a)\chi(n) = \begin{cases} 1 & \text{if } n \equiv a \pmod q, \\ 0 & \text{otherwise.} \end{cases}$$

What is this formula doing for us? Well, understanding the set of integers congruent to a mod q is equivalent to understanding the function that takes the value 1 if $n \equiv a \pmod q$ and 0 otherwise. This function appears on the right-hand side of the formula. However, it is not a particularly nice function to deal with, so we write it as a linear combination of characters, which are much nicer functions because they are multiplicative. The coefficient associated with the character χ in this linear combination is the number $\bar{\chi}(a)/\phi(q)$.

From the formula, it follows that

$$\sum_{\substack{p \text{ prime},\ m \geqslant 1 \\ p^m \leqslant x \\ p^m \equiv a \pmod q}} \log p = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \bar{\chi}(a) \sum_{\substack{p \text{ prime},\ m \geqslant 1 \\ p^m \leqslant x}} \chi(p^m) \log p.$$

The sum on the left-hand side is a natural adaptation of the sum we considered earlier when we were counting all primes. And we can estimate it if we can get good estimates for each of the sums

$$\sum_{\substack{p \text{ prime},\ m\geqslant 1\\ p^m\leqslant x}} \chi(p^m)\log p.$$

We approach these sums much as we did before, obtaining an explicit formula, analogous to (7), (10), now in terms of the zeros of the *Dirichlet L-function*:

$$L(s,\chi) := \sum_{n\geqslant 1}\frac{\chi(n)}{n^s}.$$

This function turns out to have properties closely analogous to the main properties of $\zeta(s)$. In particular, it is here that the multiplicativity of χ is all-important, since it gives us a formula similar to (1):

$$\sum_{n\geqslant 1}\frac{\chi(n)}{n^s} = \prod_{p \text{ prime}}\left(1-\frac{\chi(p)}{p^s}\right)^{-1}. \qquad (12)$$

That is, $L(s,\chi)$ has an *Euler product*. We also believe the "generalized Riemann hypothesis" that all zeros ρ of $L(\rho,\chi)=0$ in the critical strip satisfy $\mathrm{Re}(\rho)=\frac{1}{2}$. This would imply that the number of primes up to x that are congruent to a mod q can be estimated as

$$\pi(x;q,a) = \frac{\pi(x)}{\phi(q)} + O(\sqrt{x}\log^2(qx)). \qquad (13)$$

Therefore, the generalized Riemann hypothesis implies the estimate we were hoping for (formula (11)), provided that x is a little bigger than q^2.

In what range can we prove (11) unconditionally—that is, without the help of the generalized Riemann hypothesis? Although we can more or less translate the proof of the prime number theorem over into this new setting, we find that it gives (11) only when x is very large. In fact, x has to be bigger than an exponential in a power of q—which is a lot bigger than the "x is a little larger than q^2" that we obtained from the generalized Riemann hypothesis. We see a new type of problem emerging here, in which we are asking for a good starting point for the range of x for which we obtain good estimates, as a function of the modulus q; this does not have an analogy in our exploration of the prime number theorem. By the way, even though this bound "x is a little larger than q^2" is far out of reach of current methods, it still does not seem to be the best answer; calculations reveal that (11) seems to hold when x is just a little bigger than q. So even the Riemann hypothesis and its generalizations are not powerful enough to tell us the precise behaviour of the distribution of primes.

Throughout the twentieth century much thought was put in to bounding the number of zeros of Dirichlet L-functions near to the 1-line. It turns out that one can make enormous improvements in the range of x for which (11) holds (to "halfway between polynomial in q and exponential in q") provided there are no *Siegel zeros*. These putative zeros β of $L(s,(\cdot/q))$ would be real numbers with $\beta > 1-c/\sqrt{q}$; they can be shown to be extremely rare if they exist at all.

That Siegel zeros are rare is a consequence of the *Deuring–Heilbronn phenomenon*: that zeros of L-functions repel each other, rather like similarly charged particles. (This phenomenon is akin to the fact that different algebraic numbers repel one another, part of the basis of the subject of Diophantine approximation.)

How big is the smallest prime congruent to a mod q when $(a,q)=1$? Despite the possibility of the existence of Siegel zeros, one can prove that there is always such a prime less than $q^{5.5}$ if q is sufficiently large. Obtaining a result of this type is not difficult when there are no Siegel zeros. If there are Siegel zeros, then we go back to the explicit formula, which is similar to (7) but now concerns zeros of $L(s,\chi)$. If β is a Siegel zero, then it turns out that in the explicit formula there are now two obviously large terms: $x/\phi(q)$ and $-(a/q)x^\beta/\beta\phi(q)$. When $(a/q)=1$ it appears that they might almost cancel (since β is close to 1), but with more care we obtain

$$x-(a/q)\frac{x^\beta}{\beta} = (x-x^\beta)+x^\beta\left(1-\frac{1}{\beta}\right) \sim x(1-\beta)\log x.$$

This is a smaller main term than before, but it is not too hard to show that it is bigger than the contributions of all of the other zeros combined, because the Deuring–Heilbronn phenomenon implies that the Siegel zero repels those zeros, forcing them to be far to the left. When $(a/q)=-1$, the same two terms tell us that if $(1-\beta)\log x$ is small, then there are twice as many primes as we would expect up to x that are congruent to a mod q.

There is a close connection between Siegel zeros and *class numbers*, which are defined and discussed in Section **??** of ALGEBRAIC NUMBERS. Dirichlet's *class*

number formula states that $L(1,(\cdot/q)) = \pi h_{-q}/\sqrt{q}$ for $q > 6$, where h_{-q} is the class number of the field $\mathbb{Q}(\sqrt{-q})$ (for more on this topic, see Section 7 of ALGEBRAIC NUMBERS). A class number is always a positive integer, so this result immediately implies that $L(1,(\cdot/q)) \geqslant \pi/\sqrt{q}$. Another consequence is that h_{-q} is small if and only if $L(1,(\cdot/q))$ is small. The reason this gives us information about Siegel zeros is that one can show that the derivative $L'(\sigma,(\cdot/q))$ is positive (and not too small) for real numbers σ close to 1. This implies that $L(1,(\cdot/q))$ is small if and only if $L(s,(\cdot/q))$ has a real zero close to 1, that is, a Siegel zero β. When $h_{-q} = 1$, the link is more direct: it can be shown that the Siegel zero β is approximately $1-6/(\pi\sqrt{q})$. (There are also more complicated formulas for larger values of h_{-q}.)

These connections show that getting good lower bounds on h_{-q} is equivalent to getting good bounds on the possible range for Siegel zeros. Siegel showed that for any $\varepsilon > 0$ there exists a constant $c_\varepsilon > 0$ such that $L(1,(\cdot/q)) \geqslant c_\varepsilon q^{-\varepsilon}$. His proof was unsatisfactory because by its very nature one cannot give an explicit value for c_ε. Why not? Well, the proof comes in two parts. The first assumes the generalized Riemann hypothesis, in which case an explicit bound follows easily. The second obtains a lower bound *in terms of the first counterexample* to the generalized Riemann hypothesis. So if the generalized Riemann hypothesis is true but remains unproved, then Siegel's proof cannot be exploited to give explicit bounds. This dichotomy, between what can be proved with an explicit constant and what cannot be, is seen far and wide in analytic number theory—and when it appears it usually stems from an application of Siegel's result, and especially its consequences for the range in which the estimate (11) is valid.

A polynomial with integer coefficients cannot always take on prime values when we substitute in an integer. To see this, note that if p divides $f(m)$ then p also divides $f(m+p), f(m+2p), \ldots$. However, there are some prime-rich polynomials, a famous example being the polynomial $x^2 + x + 41$, which is prime for $x = 0, 1, 2, \ldots, 39$. There are almost certainly quadratic polynomials that take on more consecutive prime values, though their coefficients would have to be very large. If we ask the more restricted question of when the polynomial $x^2 + x + p$ is prime for $x = 0, 1, 2, \ldots, p-2$, then the answer, given by Rabinowitch, is rather surprising: it happens if and only if $h_{-q} = 1$, where $q = 4p-1$. Gauss did extensive calculations of class numbers and predicted that there are just nine values of q with $h_{-q} = 1$, the largest of which is $163 = 4\times41-1$. Using the Deuring–Heilbronn phenomenon researchers showed, in the 1930s, that there is at most one q with $h_{-q} = 1$ that is not already on Gauss's list; but as usual with such methods, one could not give a bound on the size of the putative extra counterexample. It was not until the 1960s that Baker and Stark proved that there was no tenth q, both proofs involving techniques far removed from those here (in fact Heegner gave what we now understand to have been a correct proof in the 1950s but he was so far ahead of his time that it was difficult for mathematicians to appreciate his arguments and to believe that all of the details were correct). In the 1980s Goldfeld, Gross, and Zagier gave the best result to date, showing that $h_{-q} \geqslant \frac{1}{7700}\log q$ this time using the Deuring–Heilbronn phenomenon with the zeros of yet another type of L-function to repel the zeros of $L(s,(\cdot/q))$.

This idea that primes are well distributed in arithmetic progressions except for a few rare moduli was exploited by Bombieri and Vinogradov to prove that (11) holds "almost always" when x is a little bigger than q^2 (that is, in the same range that we get "always" from the generalized Riemann hypothesis). More precisely, for given large x we have that (11) holds for "almost all" q less than $\sqrt{x}/(\log x)^2$ and for all a such that $(a,q) = 1$. "Almost all" means that, out of all q less than $\sqrt{x}/(\log x)^2$, the proportion for which (11) does not hold for every a with $(a,q) = 1$ tends to 0 as $x \to \infty$. Thus, the possibility is not ruled out that there are infinitely many counterexamples. However, since this would contradict the generalized Riemann hypothesis, we do not believe that it is so.

The *Barban–Davenport–Halberstam theorem* gives a weaker result, but it is valid for the whole feasible range: for any given large x, the estimate (11) holds for "almost all" pairs q and a such that $q \leqslant x/(\log x)^2$ and $(a,q) = 1$.

5 Primes in Short Intervals

The prediction of Gauss referred to the primes "around" x, so it perhaps makes more sense to inter-

pret his statement by considering the number of primes in short intervals at around x. If we believe Gauss, then we might expect the number of primes between x and $x+y$ to be about $y/\log x$. That is, in terms of the prime-counting function π, we might expect that

$$\pi(x+y)-\pi(x)\sim\frac{y}{\log x} \tag{14}$$

for $|y|\leqslant x/2$. However, we have to be a little careful about the range for y. For example, if $y=\frac{1}{2}\log x$, then we certainly cannot expect to have half a prime in each interval. Obviously we need y to be large enough that the prediction can be interpreted in a way that makes sense; indeed, the Gauss–Cramér model suggests that (14) should hold when $|y|$ is a little bigger than $(\log x)^2$.

If we attempt to prove (14) using the same methods we used in the proof of the prime number theorem, we find ourselves bounding differences between ρth powers as follows:

$$\left|\frac{(x+y)^\rho-x^\rho}{\rho}\right|=\left|\int_x^{x+y}t^{\rho-1}\,\mathrm{d}t\right|$$
$$\leqslant\int_x^{x+y}t^{\mathrm{Re}(\rho)-1}\,\mathrm{d}t\leqslant y(x+y)^{\mathrm{Re}(\rho)-1}.$$

With bounds on the density of zeros of $\zeta(s)$ well to the right of $\frac{1}{2}$, it has been shown that (14) holds for y a little bigger than $x^{7/12}$; but there is little hope, even assuming the Riemann hypothesis, that such methods will lead to a proof of (14) for intervals of length $\sqrt{x}$ or less.

In 1949 Selberg showed that (14) is true for "almost all" x when $|y|$ is a little bigger than $(\log x)^2$, assuming the Riemann hypothesis. Once again, "almost all" means 100%, rather than "all," and it is feasible that there are infinitely many counterexamples, though at that time it seemed highly unlikely. It therefore came as a surprise when Maier showed, in 1984, that, for any fixed $A>0$, the estimate (14) fails for infinitely many integers x, with $y=(\log x)^A$. His ingenious proof rests on showing that the small primes do not always have as many multiples in an interval as one might expect.

Let $p_1=2<p_2=3<\cdots$ be the sequence of primes. We are now interested in the size of the gaps $p_{n+1}-p_n$ between consecutive primes. Since there are about $x/\log x$ primes up to x, the average difference is $\log x$ and we might ask how often the difference between consecutive primes is about average, whether the differences can get really small, and whether the differences can get really large. The Gauss–Cramér model suggests that the proportion of n for which the gap between consecutive primes is more than λ times the average, that is $p_{n+1}-p_n>\lambda\log p_n$, is approximately $\mathrm{e}^{-\lambda}$; and, similarly, the proportion of intervals $[x,x+\lambda\log x]$ containing exactly k primes is approximately $\mathrm{e}^{-\lambda}\lambda^k/k!$, a suggestion which, as we shall see, is supported by other considerations. By looking at the tail of this distribution, Cramér conjectured that $\limsup_{n\to\infty}(p_{n+1}-p_n)/(\log p_n)^2=1$, and the evidence we have *seems* to support this (see Table 2).

Table 2 The largest known gaps between primes.

p_n	$p_{n+1}-p_n$	$\dfrac{p_{n+1}-p_n}{\log^2 p_n}$
113	14	0.6264
1 327	34	0.6576
31 397	72	0.6715
370 261	112	0.6812
2 010 733	148	0.7026
20 831 323	210	0.7395
25 056 082 087	456	0.7953
2 614 941 710 599	652	0.7975
19 581 334 192 423	766	0.8178
218 209 405 436 543	906	0.8311
1 693 182 318 746 371	1132	0.9206

The Gauss–Cramér model does have a big drawback: it does not "know any arithmetic." In particular, as we noted earlier, it does not predict divisibility by small primes. One manifestation of this failing is that it predicts that there should be just about as many gaps of length 1 between primes as there are of length 2. However, there is only one gap of length 1, since if two primes differ by 1, then one of them must be even, whereas there are many examples of pairs of primes differing by 2—and there are believed to be infinitely many. For the model to make correct conjectures about prime pairs, we must consider divisibility by small primes in the formulation of the model, which makes it rather more complicated. Since there are these glaring errors in the simpler model, Cramér's conjecture for the largest gaps between consecutive primes must be treated with a degree of suspicion. And in fact, if one corrects the model to account for divis-

ibility by small primes, one is led to conjecture that $\limsup_{n\to\infty}(p_{n+1}-p_n)/(\log p_n)^2$ is greater than $\frac{9}{8}$.

Finding large gaps between primes is equivalent to finding long sequences of composite numbers. How about trying to do this explicitly? For example, we know that $n!+j$ is composite for $2 \leqslant j \leqslant n$, as it is divisible by j. Therefore we have a gap of length at least n between consecutive primes, the first of which is the largest prime less than or equal to $n!+1$. However, this observation is not especially helpful, since the average gap between primes around $n!$ is $\log(n!)$, which is approximately equal to $n\log n$, whereas we are looking for gaps that are *larger* than the average. However, it is possible to generalize this argument and show that there are indeed long sequences of consecutive integers, each with a small prime factor. In the 1930s, Erdős reformulated the question as follows. Fix a positive integer z, and for each prime $p \leqslant z$ choose an integer a_p in such a way that, for as large an integer y as possible, every positive integer $n \leqslant y$ satisfies at least one of the congruences $n \equiv a_p \pmod p$. Now let X be the product of all the primes up to z (which means, by the prime number theorem, that $\log X$ is about z), and let x be the integer between X and $2X$ such that $x \equiv -a_p \pmod p$ for every $p \leqslant z$. (This integer exists, by the *Chinese remainder theorem*.) If m is an integer between $x+1$ and $x+y$, then $m-x$ is a positive integer less than y, so $m-x \equiv a_p \pmod p$ for some prime $p \leqslant z$. Since $x \equiv -a_p \pmod p$, it follows that m is divisible by p. Thus, all the integers from $x+1$ to $x+y$ are composite. Using this basic idea, it can be shown that there are infinitely many primes p_n for which $p_{n+1}-p_n$ is about $(\log p_n)(\log\log p_n)$, which is significantly larger than the average but nowhere close to Cramér's conjecture.

6 Gaps between Primes that are Smaller than the Average

We have just seen how to show that there are infinitely many pairs of consecutive primes whose difference is much bigger than the average: that is $\limsup_{n\to\infty}(p_{n+1}-p_n)/(\log p_n)=\infty$. We would now like to show that there are infinitely many pairs of consecutive primes whose difference is much smaller than the average: that is $\liminf_{n\to\infty}(p_{n+1}-p_n)/(\log p_n) = 0$. Of course, it is believed that there are infinitely many pairs of primes that differ by 2, but this question seems intractable for now.

Until recently researchers had very little success with the question of small gaps; the best result before 2000 was that there are infinitely many gaps of size less than one-quarter of the average. However, a recent method of Goldston, Pintz, and Yildirim, which counts primes in short intervals with simple weighting functions, proves that $\liminf_{n\to\infty}(p_{n+1}-p_n)/(\log p_n)=0$, and even that there are infinitely many pairs of consecutive primes with difference no larger than about $\sqrt{\log p_n}$. Their proof, rather surprisingly, rests on estimates for primes in arithmetic progressions; in particular, that (11) holds for almost all q up to $\sqrt{x}$ (as discussed earlier). Moreover, they obtain a conditional result of the following kind: if in fact (11) holds for almost all q up to a little larger than $\sqrt{x}$, then it follows that there exists an integer B such that $p_{n+1}-p_n \leqslant B$ for infinitely many primes p_n.

7 Very Small Gaps between Primes

There appear to be many pairs of primes that differ by two, like 3 and 5, 5 and 7, ..., the so-called *twin primes*, though no one has yet proved that there are infinitely many. In fact, for every even integer $2k$ there seem to be many pairs of primes that differ by $2k$, but again no one has yet proved that there are infinitely many. This is one of the outstanding problems in the subject.

In a similar vein is Goldbach conjecture's from the 1760s: is it true that every even integer greater than 2 is the sum of two primes? This is still an open question, and indeed a publisher recently offered a million dollars for its solution. We know it is true for almost all integers, and it has been computer tested for every even integer up to 4×10^{14}. The most famous result on this question is due to Jing-Run Chen (1966) who showed that every even integer can be written as the sum of a prime and a second integer that has *at most two* prime factors (that is, it could be a prime or an "almost-prime").

In fact, Goldbach never asked this question. He asked Euler, in a letter in the 1760s, whether every integer greater than 1 can be written as the sum of at most three primes, which would imply what we now

call the "Goldbach conjecture." In the 1920s Vinogradov showed that every sufficiently large odd integer can be written as the sum of three primes (and thus every sufficiently large even integer can be written as the sum of four primes). We actually believe that every odd integer greater than 5 is the sum of three primes but the known proofs only work once the numbers involved are large enough. In this case we can be explicit about "sufficiently large"—at the moment the proof needs them to be at least e^{5700}, but it is rumored that this may soon be substantially reduced, perhaps even to 7.

To guess at the precise number of prime pairs q, $q+2$ with $q \leqslant x$ we proceed as follows. If we do not consider divisibility by the small primes, then the Gauss–Cramér model suggests that a random integer up to x is prime with probability roughly $1/\log x$, so we might expect $x/(\log x)^2$ prime pairs q, $q+2$ up to x. However, we do have to account for the small primes, as the q, $q+1$ example shows, so let us consider 2-divisibility. The proportion of random pairs of integers that are both odd is $\frac{1}{4}$, whereas the proportion of random q such that q and $q+2$ are both odd is $\frac{1}{2}$. Thus we should adjust our guess $x/(\log x)^2$ by a factor $(\frac{1}{2})/(\frac{1}{4}) = 2$. Similarly, the proportion of random pairs of integers that are both not divisible by 3 (or indeed by any given odd prime p) is $(\frac{2}{3})^2$ (and $(1-1/p)^2$, respectively), whereas the proportion of random q such that q and $q+2$ are both not divisible by 3 (or by prime p) is $\frac{1}{3}$ (and $(1-2/p)$, respectively). Adjusting our formula for each prime p we end up with the prediction

$$\#\{q \leqslant x : q \text{ and } q+2 \text{ both prime}\} \sim 2 \prod_{p \text{ an odd prime}} \frac{(1-2/p)}{(1-1/p)^2} \frac{x}{(\log x)^2}.$$

This is known as the "*asymptotic twin-prime conjecture*." Despite its plausibility there do not seem to be any practical ideas around for turning the heuristic argument above into something rigorous. The one good unconditional result known is that the number of twin primes less than or equal to x is never more than four times the quantity we have just predicted. One can make a more precise prediction replacing $x/(\log x)^2$ by $\int_2^x (1/(\log t)^2)\,\mathrm{d}t$, and then we expect that the difference between the two sides is no more than $c\sqrt{x}$ for some constant $c > 0$, a guesstimate that is well supported by computational evidence.

A similar method allows us to make predictions for the number of primes in any polynomial-type patterns. Let $f_1(t), f_2(t), \ldots, f_k(t) \in \mathbb{Z}[t]$ be distinct irreducible polynomials of degree greater than or equal to 1 with positive leading coefficient, and define $\omega(p)$ to be the number of integers $n \pmod p$ for which p divides $f_1(n)f_2(n)\cdots f_k(n)$. (In the case of twin primes above we have $f_1(t) = t$, $f_2(t) = t+2$ with $\omega(2) = 1$ and $\omega(p) = 2$ for all odd primes p.) If $\omega(p) = p$ then p always divides at least one of the polynomial values, so they can be simultaneously prime just finitely often (an example of this is when $f_1(t) = t$, $f_2(t) = t+1$, in which case $\omega(2) = 2$). Otherwise we have an *admissible set* of polynomials for which we predict that the number of integers n less than x for which all of $f_1(n), f_2(n), \ldots, f_k(n)$ are prime is about

$$\prod_{p \text{ prime}} \frac{(1-\omega_f(p)/p)}{(1-1/p)^k} \times \frac{x}{\log|f_1(x)|\log|f_2(x)|\cdots\log|f_k(x)|} \qquad (15)$$

once x is sufficiently large. One can use a similar heuristic to make predictions in Goldbach's conjecture, that is, for the number of pairs of primes p, q for which $p+q = 2N$. Again, these predictions are very well matched by the computational evidence.

There are just a few cases of conjecture (15) that have been proved. Modifications of the proof of the prime number theorem give such a result for admissible polynomials $qt+a$ (in other words, for primes in arithmetic progressions) and for admissible $at^2 + btu + cu^2 \in \mathbb{Z}[t,u]$ (as well as some other polynomials in two variables of degree two). It is also known for a certain type of polynomial in n variables of degree n (the admissible "norm-forms").

There was little improvement on this situation during the twentieth century until quite recently, when, by very different methods, Friedlander and Iwaniec broke through this stalemate showing such a result for the polynomial t^2+u^4, and then Heath-Brown did so for any admissible homogenous polynomial in two variables of degree three.

Another truly extraordinary breakthrough occurred recently with a result of Green and Tao, proved in 2004, which states that for every k there are infinitely many k-term arithmetic progressions of primes: that is, pairs of integers a, d such that $a, a+d, a+2d, \ldots, a+(k-1)d$ are all prime. Green and Tao are currently

hard at work attempting to show that the number of four-term arithmetic progressions of primes is indeed well approximated by (15). They are also extending their results to other families of polynomials.

8 Gaps between Primes Revisited

In the 1970s Gallagher deduced from the conjectured prediction (15) (with $f_j(t) = t + a_j$) that the proportion of intervals $[x, x + \lambda \log x]$ which contain exactly k primes is close to $\mathrm{e}^{-\lambda}\lambda^k/k!$ (as was also deduced, in Section 5 above, from the Gauss–Cramér heuristics). This has recently been extended to support the prediction that, as we vary x from X to $2X$, the number of primes in the interval $[x, x + y]$ is normally distributed with mean $\int_x^{x+y}(1/\log t)\,\mathrm{d}t$ and variance $(1 - \delta)y/\log x$, where δ is some constant strictly between 0 and 1 and we take y to be x^δ.

When $y > \sqrt{x}$ the Riemann zeta function supplies information on the distribution of primes in intervals $[x, x+y)$ via the explicit formula (7). Indeed when we compute the "variance"

$$\frac{1}{X}\int_X^{2X}\Bigg(\sum_{p \text{ prime},\ x<p\leqslant x+y} \log p - y\Bigg)^2 \mathrm{d}x$$

using the explicit formula we obtain a sum of terms of the form $\int_X^{2X} x^{\mathrm{i}(\gamma_j-\gamma_k)}\,\mathrm{d}x$. Here we are assuming the Riemann hypothesis and writing the zeros of $\zeta(s)$ as $1/2 \pm \mathrm{i}\gamma_n$ with $0 < \gamma_1 < \gamma_2 < \cdots$. This sum is dominated by the terms corresponding to those pairs γ_j, γ_k for which $|\gamma_j - \gamma_k|$ is small (in which case there is little cancellation in the integral). Therefore, in order to understand the variance for the distribution of primes in short intervals we need to understand the distribution of the zeros of $\zeta(s)$ in short intervals. In 1973 Montgomery investigated this and suggested that the proportion of pairs of zeros of $\zeta(s)$ whose difference is less than α times the average gap between consecutive zeros is given by the integral

$$\int_0^\alpha \left(1 - \left(\frac{\sin \pi\theta}{\pi\theta}\right)^2\right)\mathrm{d}\theta, \tag{16}$$

and he proved an equivalent form of this in a limited range. If the zeros were placed "randomly," then (16) would be replaced by α. In fact (16) is about $\frac{1}{9}\alpha^3$ for small α, which is far smaller than α. This means that there are far fewer pairs of zeros of $\zeta(s)$ that are close together than one might expect, which we express informally by saying that the zeros of $\zeta(s)$ *repel* one another.

In a now-famous conversation that took place at the Institute for Advanced Study in Princeton, Montgomery mentioned his ideas to the physicist Freeman Dyson. Dyson immediately recognized (16) as a function that comes up in modelling energy levels in quantum chaos. Believing that this was unlikely to be a coincidence, he suggested that the zeros of the Riemann zeta function are distributed, *in all aspects*, like energy levels, which are in turn modelled on the distribution of EIGENVALUES of random HERMITIAN MATRICES. There is now substantial computational and theoretical evidence that Dyson's suggestion is correct and can be extended to Dirichlet L-functions, as well as other types of L-functions, and even to other statistics about L-functions.

One note of caution. Few of the conjectured consequences of this new "random matrix theory" have been unconditionally proved, or seem likely to be in the foreseeable future. It simply provides a tool to make predictions where that was too difficult to do before. However, there is at least one key question about which we still cannot make a well-substantiated prediction: how big does $\zeta(s)$ get on the $\frac{1}{2}$-line? One can show that $\log|\zeta(\frac{1}{2} + \mathrm{i}t)|$ gets larger than $\sqrt{\log T}$ for values of t close to T, and that it gets no larger than $\log T$. However, it is unclear, even if we do not insist on a rigorous proof, whether the true maximal order is nearer the upper or lower bound.

9 Sieve Methods

Almost all of our discussion so far has been about developments of Riemann's approach to counting primes. This approach is very delicate and not as adaptable as one might wish to many natural questions (such as counting k-tuples of primes $n + a_1, n + a_2, \ldots, n + a_k$). However, one can go back to *sieve methods*, which are modifications of the sieve of Eratosthenes, and at least get upper bounds. For example, suppose we want to find an upper bound for the number of prime pairs n, $n + 2$ with $N < n \leqslant 2N$. One possibility would be to fix a number y and determine for how many pairs n, $n + 2$ with $N < n \leqslant 2N$

it is the case that neither n nor $n+2$ has a prime factor less than y. If we took y to be $(2N)^{1/2}$, then this method would exactly count the twin primes, but it seems to be far too difficult to implement. But it turns out that if instead we take y to be a small power of N, then the calculations become much easier and there are ways of obtaining good bounds. (However, these bounds become less accurate as the power gets closer to $\frac{1}{2}$.)

In the 1920s Brun showed how to make the principle of inclusion–exclusion into a useful tool in this type of question. This principle is best exhibited when counting the number of integers n in a set S that are coprime to given integer m. We begin with the number of integers in S, which is obviously more than the quantity we seek. Next, we subtract, for each prime p dividing m, the number of integers in S that are divisible by p. If $n \in S$ is divisible by exactly r prime factors of m, then we have counted $1 + r \times (-1)$ for the contribution of n so far, which is less than or equal to 0, and less than 0 for $r \geqslant 2$; whereas we wanted to count 0 when $r \geqslant 2$ (since n is not coprime to m). Thus we obtain a number that is less than the quantity we seek. To compensate for that, we add back in the number of integers in S divisible by pq for each pair of primes $p < q$ which divide m. We have now counted $1 + r \times (-1) + \binom{r}{2} \times 1$ for the contribution of n, which is greater than or equal to 0, and greater than 0 for $r \geqslant 3$. Similarly, we subtract the number of integers divisible by pqr, etc.

For each $n \in S$ we end up counting $(1-1)^r$ for n, where r is the number of distinct prime factors of (m, n). Expanding this sum with the binomial theorem we may reexpress this identity as follows. Let $\chi_m(n) = 1$ if $(n, m) = 1$ and 0 otherwise. Then

$$\chi_m(n) = \sum_{d|(m,n)} \mu(d),$$

where $\mu(m)$, the Möbius function, equals 0 if m is divisible by the square of a prime and equals $(-1)^{\omega(m)}$ otherwise, where $\omega(m)$ is the number of distinct prime factors of m.

The inclusion-exclusion inequalities just discussed may be obtained from

$$\sum_{\substack{d|(m,n)\\ \omega(d)\leqslant 2k+1}} \mu(d) \leqslant \chi_m(n) \leqslant \sum_{\substack{d|(m,n)\\ \omega(d)\leqslant 2k}} \mu(d),$$

which holds for any $k \geqslant 0$, by summing over all $n \in S$.

The reason for using these abbreviated sums rather than the complete sum is that there are far fewer terms and thus, when one sums over values of n, there will be far fewer rounding errors (remember that it was rounding errors that sank our attempt to estimate the number of primes up to x using the sieve of Eratosthenes). On the other hand, they have the disadvantage that they cannot possibly give the exact answer, since they are missing many appropriate terms. However, with a judicious choice of k the missing terms do not contribute much to the complete sum and we get a good answer.

Minor variants work well for many questions. In the "combinatorial sieve" one selects which d are part of the upper and lower bound sums, not by counting the total number of prime factors they contain but instead using other criteria, such as the numbers of prime factors of d in each of several intervals. Using such a method Brun showed that there cannot be too many twin primes p, $p+2$; indeed that the sum of $1/p$, over all primes p for which $p+2$ is also prime, converges, in contrast with (3).

In the "Selberg upper bound sieve" one comes up with some numbers λ_d that are nonzero only when $d \leqslant D$ (where D is chosen to be not too large), with the property that

$$\chi_m(n) \leqslant \Big(\sum_{d|n} \lambda_d\Big)^2 \quad \text{for all } n.$$

Summing over the appropriate n one then finds the optimal solution by minimizing the resulting quadratic form. Lower bounds can also be obtained out of Selberg's methods. It was using such methods that Chen was able to prove there are infinitely many primes p for which $p+2$ has at most two prime factors, and that Goldston, Pintz, and Yildirim were able to establish that there are sometimes short gaps between primes. It is also an essential ingredient in the work of Green and Tao. One can also get good upper bounds on the number of primes in arithmetic progressions and short intervals:

- there are never more than $2y/\log y$ primes in any interval of length y;
- there are never more than $2x/\phi(q)\log(x/q)$ primes up to x in an arithmetic progression mod q.

Notice that in each case the log in the denominator is of the number of integers being considered (y and

x/q, respectively), not $\log x$ as expected, though this will only make a significant difference if the number of integers being considered is small. Otherwise these inequalities are bigger than the expected quantity by a factor of 2. Can this "2" be improved? It will be difficult because we showed earlier that if there are Siegel zeros then we get twice as many primes as expected in certain arithmetic progressions. Therefore, if we can improve the "2" in these two formulas, then we can deduce that there are no Siegel zeros!

10 Smooth Numbers

An integer is *y-smooth* if all of its prime factors are less than or equal to y. A proportion $1 - \log 2$ of the integers up to x are $\sqrt{x}$-smooth, and indeed, for any fixed $u > 1$ there exists some number $\rho(u) > 0$ such that if $x = y^u$, then a proportion $\rho(u)$ of the integers up to x are y-smooth. This proportion does not seem to have any easy definition in general. For $1 \leqslant u \leqslant 2$ we have $\rho(u) = 1 - \log u$, but for larger u it is best defined as

$$\rho(u) := \frac{1}{u} \int_0^1 \rho(u-t)\,\mathrm{d}t,$$

an *integral delay equation*. Such an equation is typical when we give precise estimates for questions that arise in sieve theory.

Questions about the distribution of smooth numbers arise frequently in the analysis of algorithms, and have consequently been the focus of a lot of recent research. (See COMPUTATIONAL NUMBER THEORY for an example of the use of smooth numbers.)

11 The Circle Method

Another method of analysis that plays a prominent role in this subject is the so-called *circle method*, which goes back to HARDY and LITTLEWOOD. This method uses the fact that, for any integer n,

$$\int_0^1 \mathrm{e}^{2\mathrm{i}\pi nt}\,\mathrm{d}t = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For example, if we wish to count the number, $r(n)$, of solutions to the equation $p + q = n$ with p and q prime, we can express it as an integral as follows:

$$\begin{aligned} r(n) &= \sum_{\substack{p,q \leqslant n \\ \text{both prime}}} \int_0^1 \mathrm{e}^{2\mathrm{i}\pi(p+q-n)t}\,\mathrm{d}t \\ &= \int_0^1 \mathrm{e}^{-2\mathrm{i}\pi nt} \Bigg(\sum_{p \text{ prime},\ p \leqslant n} \mathrm{e}^{2\mathrm{i}\pi pt} \Bigg)^2 \mathrm{d}t. \end{aligned}$$

The first equality holds because the integrand is 0 when $p + q \neq n$ and 1 otherwise, and the second is easy to check.

At first sight it looks more difficult to estimate the integral than it is to estimate $r(n)$ directly, but this is not the case. For instance, the prime number theorem for arithmetic progressions allows us to estimate $P(t) := \sum_{p \leqslant n} \mathrm{e}^{2\mathrm{i}\pi pt}$ when t is a rational ℓ/m with m small. For in this case,

$$\begin{aligned} P\Big(\frac{\ell}{m}\Big) &= \sum_{(a,m)=1} \mathrm{e}^{2\mathrm{i}\pi a\ell/m} \sum_{\substack{p \leqslant n, \\ p \equiv a \pmod m}} 1 \\ &\approx \sum_{(a,m)=1} \mathrm{e}^{2\mathrm{i}\pi a\ell/m} \frac{\pi(n)}{\phi(m)} = \mu(m) \frac{\pi(n)}{\phi(m)}. \end{aligned}$$

If t is sufficiently close to ℓ/m, then $P(t) \approx P(\ell/m)$; such values of t are called the *major arcs* and we believe that the integral over the major arcs gives, in total, a very good approximation to $r(n)$; indeed we get something very close to the quantity one predicts from something like (15). Thus to prove the Goldbach conjecture we need to show that the contribution to the integral from the other values of t (that is, from the *minor arcs*) is small. In many problems one can successfully do this, but no one has yet succeeded in doing so for the Goldbach problem. Also useful is the "discrete analogue" of the above: using the identity

$$\frac{1}{m} \sum_{j=0}^{m-1} \mathrm{e}^{2\mathrm{i}\pi jn/m}\,\mathrm{d}t = \begin{cases} 1 & \text{if } n \equiv 0 \pmod m, \\ 0 & \text{otherwise} \end{cases}$$

(which holds for any given integer $m \geqslant 1$), we have that

$$\begin{aligned} r(n) &= \sum_{\substack{p,q \leqslant n \\ \text{both prime}}} \frac{1}{m} \sum_{j=0}^{m-1} \mathrm{e}^{2\mathrm{i}\pi j(p+q-n)/m} \\ &= \sum_{j=0}^{m-1} \mathrm{e}^{-2\mathrm{i}\pi jn/m} P(j/m)^2 \end{aligned}$$

provided $m > n$. A similar analysis can be used here but working mod m sometimes has advantages, as it

allows us to use properties of the multiplicative group mod m.

Sums like $P(j/m)$ in the paragraph above, or more simple sums like $\sum_{n\leqslant N} \mathrm{e}^{2\mathrm{i}\pi n^k/m}$ are called "exponential sums." They play a central role in many of the calculations one does in analytic number theory. There are several techniques for investigating them.

(1) It is easy to sum the geometric progression $\sum_{n\leqslant N} \mathrm{e}^{2\mathrm{i}\pi n/m}$. With higher-degree polynomials one can often reduce to this case; for example, by writing $n_1 - n_2 = h$ we have

$$\begin{aligned}\Big|\sum_{n\leqslant N} \mathrm{e}^{2\mathrm{i}\pi n^2/m}\Big|^2 &= \sum_{n_1,n_2\leqslant N} \mathrm{e}^{2\mathrm{i}\pi(n_1^2-n_2^2)/m}\\ &= \sum_{|h|\leqslant N} \mathrm{e}^{2\mathrm{i}\pi h^2/m} \sum_{\substack{\max\{0,-h\}<n_2\\ \leqslant\min\{N,N-h\}}} \mathrm{e}^{4\mathrm{i}\pi h n_2/m},\end{aligned}$$

and the inner sum is now a geometric progression.

(2) The work of Weil and Deligne, which gives very accurate results on the number of solutions to equations mod p, is ideally suited to many applications in analytic number theory. For example, the "Kloosterman sum" $\sum_{a_1a_2\cdots a_k\equiv b \pmod p} \mathrm{e}^{2\mathrm{i}\pi(a_1+a_2+\cdots+a_k)/p}$, where the a_i run over the integers mod p and $(b,q) = 1$, appears naturally in many questions; Deligne showed that it has absolute value less than or equal to $kp^{(k-1)/2}$, an extraordinary amount of cancellation in this sum which has about p^{k-1} summands, each of absolute value 1.

(3) We discussed earlier the fact that the values of $\zeta(s)$ satisfy a symmetry about the line $\mathrm{Re}(s) = \frac{1}{2}$, given by the "functional equation." There are other functions (called "modular functions") that also have symmetries in the complex plane; typically the value of the function at s is related to the value of the function at $(\alpha s+\beta)/(\gamma s+\delta)$, for some integers α, β, γ, δ satisfying $\alpha\delta - \beta\gamma = 1$. Sometimes an exponential sum can be related to the value of a modular function, and subsequently to the value of that modular function at another point, using the symmetry of the function.

12 More *L*-Functions

There are many types of L-functions beyond Dirichlet L-functions, some of which are well understood, some not. The type that have received the most attention recently are a class of L-functions that can be associated with elliptic curves (see p. **??** of ARITHMETIC GEOMETRY). An *elliptic curve* E is given by an equation of the form $y^2 = x^3 + ax + b$, where the *discriminant* $4a^3 + 27b^2$ is nonzero. The associated L-function $L(E,s)$ is most easily described in terms of its Euler product:

$$L(E,s) = \prod_p \left(1 - \frac{a_p}{p^s} + \frac{p}{p^{2s}}\right)^{-1}. \qquad (17)$$

Here a_p is an integer which, for primes p not dividing $4a^3 + 27b^2$, is defined to be p minus the number of solutions $(x,y) \pmod p$ to the equation $y^2 \equiv x^3+ax+b \pmod p$. It can be shown that each $|a_p|$ is less than $2\sqrt{p}$, so the Euler product above converges absolutely when $\mathrm{Re}(s) > \frac{3}{2}$. Therefore, (17) is a good definition for these values of s. Can we now extend it to the whole of the complex plane, as we did for $\zeta(s)$? This is a very deep problem—the answer is yes; in fact, it is the celebrated theorem of Andrew Wiles that implied FERMAT'S LAST THEOREM.

Another interesting question is to understand the distribution of values of $a_p/2\sqrt{p}$ as we range over primes p. These all lie in the interval $[-1,1]$. One might expect them to be uniformly distributed in the interval, but in fact this is never the case. As discussed in ALGEBRAIC NUMBERS one can write $a_p = \alpha_p + \bar{\alpha}_p$, where $|\alpha_p| = \sqrt{p}$, and α_p was called the Weil number. If we write $\alpha = \sqrt{p}\mathrm{e}^{\pm\mathrm{i}\theta_p}$, then $a_p = 2\sqrt{p}\cos(\theta_p)$ for some angle $\theta_p \in [0,\pi]$. We can then think of θ_p as belonging to the upper half of a circle. The surprise is that for almost all elliptic curves the θ_p are not uniformly distributed, which would mean the proportion in a certain arc would be proportional to the length of that arc. Rather, they are distributed in such a way that the proportion of them in any given arc is proportional to the area under that arc. This is a recent result of Richard Taylor.

The correct analogue of the Riemann hypothesis for $L(E,s)$ turns out to be that all the nontrivial zeros lie on the line $\mathrm{Re}(s) = 1$. This is believed to be true. Moreover, it is believed that they, like the zeros of $\zeta(s)$, are distributed according to the rules that govern the eigenvalues of randomly chosen matrices.

These L-functions often have zeros at $s = 1$ (which is linked to the "Birch–Swinnerton-Dyer conjectures") and these zeros repel zeros of Dirichlet L-functions

(which is what was used by Goldfeld, Gross, and Zagier, as mentioned in Section 4, to get their lower bound on h_{-q}).

L-functions arise in many areas of arithmetic geometry, and their coefficients typically describe the number of points satisfying certain equations mod p. The *Langlands program* seeks to understand these connections at a deep level.

It seems that every "natural" L-function has many of the same analytic properties as those discussed in this article. Selberg has proposed that this phenomenon should be even more general. Consider sums $A(s) = \sum_{n \geqslant 1} a_n/n^s$ that

- are well-defined when $\mathrm{Re}(s) > 1$,
- have an Euler product $\prod_p (1 + b_p/p^s + b_{p^2}/p^{2s} + \cdots)$ in this (or an even smaller) region,
- have coefficients a_n that are smaller than any given power of n, once n is sufficiently large,
- satisfy $|b_n| < \kappa n^\theta$ for some constants $\theta < \frac{1}{2}$ and $\kappa > 0$.

Selberg conjectures that we should be able to give a good definition to $A(s)$ on the whole complex plane, and that $A(s)$ should have a symmetry connecting the value of $A(s)$ with $A(1 - s)$. Furthermore, he conjectures that the Riemann hypothesis should hold for $A(s)$!

The current wishful thinking is that Selberg's family of L-functions is precisely the same as those considered by Langlands.

13 Conclusion

In this article we have described current thinking on several key questions about the distribution of primes. It is frustrating that after centuries of research so little has been proved, the primes guarding their mysteries so jealously. Each new breakthrough seems to require brilliant ideas and extraordinary technical prowess. As EULER wrote in 1770:

> Mathematicians have tried in vain to discover some order in the sequence of prime numbers but we have every reason to believe that there are some mysteries which the human mind will never penetrate.

Further Reading

Hardy and Wright's classic book (1980) stands alone amongst introductory number theory texts for the quality of its discussion of analytic topics. The best introduction to the heart of analytic number theory is the masterful book by Davenport (2000). Everything you have ever wanted to know about the Riemann zeta-function is in Titchmarsh (1986). Finally, there are two recently released books by modern masters of the subject (Iwaniec and Kowalski 2004; Montgomery and Vaughan 2006) that introduce the reader to the key issues of the subject.

The reference list below includes several papers, significant for this article, whose content is not discussed in any of the listed books.

Davenport, H. 2000. *Multiplicative Number Theory*, 3rd edn. Springer.

Deligne, P. 1977. Applications de la formule des traces aux sommes trigonometriques. In *Cohomologie Étale* (SGA 4 1/2). Lecture Notes in Mathematics, Volume 569. Springer.

Green, B., and T. Tao. Forthcoming. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, in press.

Hardy, G. H., and E. M. Wright. 1980. *An Introduction to the Theory of Numbers*, 5th edn. Oxford: Oxford Science Publications.

Ingham, A. E. 1949. Review 10,595c (MR0029411). *Mathematical Reviews*. Providence, RI: American Mathematical Society.

Iwaniec, H., and E. Kowalski. 2004. *Analytic Number Theory*. Colloquium Publications, Volume 53. Providence, RI: American Mathematical Society.

Montgomery, H. L., and R. C. Vaughan. 2006. *Multiplicative Number Theory I: Classical Theory*. Cambridge University Press.

Soundararajan, K. Forthcoming. Small gaps between prime numbers: the work of Goldston–Pintz–Yildirim. *Bulletin of the American Mathematical Society*, in press.

Titchmarsh, E. C. 1986. *The Theory of the Riemann Zeta-Function*, 2nd edn. Oxford University Press.