What can Calculus tell us about prime numbers?

Introduction

The problem of determining the distribution of the primes is a question mathematicians have worked on for centuries. Beginning with Euclid's proof of the infinitude of primes, a natural question to ask is about how often do primes turn up? One way of quantifying this is through the lens of the harmonic series, the fact that

$$\sum_{k=1}^{n} \frac{1}{k} \to \infty \text{ as } n \to \infty \tag{1}$$

Does the same hold for $\sum_p \frac{1}{p}$? We can answer this question using an object known as the Riemann zeta function, which is the subject of the Riemann Hypothesis, the most famous open problem in mathematics. This approach turns out to be very useful and can be used to answer other questions on the distribution of primes and highlights the relevance of functions like the Riemann zeta function to prime numbers.

The Integral Test

The Harmonic series is the series

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \tag{2}$$

The prereading contained the first known proof of the divergence of the harmonic series, by Nicole Oresme in the 14^{th} century. Here we'll use an alternative approach known as the integral test which will allow us to analyse a wider range of series.

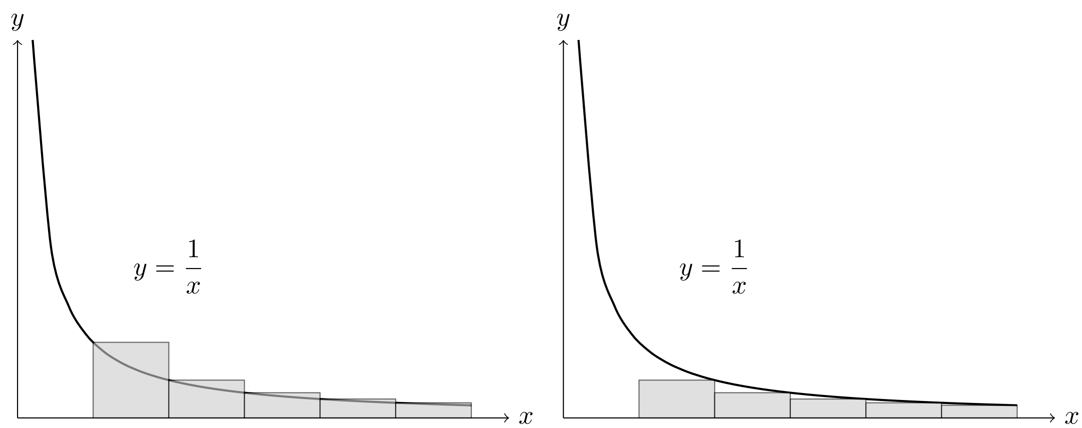

Figure 1: Overestimating and underestimating the area under the curve with rectangles

In Figure 1, rectangles are used to approximate the area under the curve. The rectangles are of width 1 - at $x = k$, the overestimate rectangle has height $\frac{1}{k}$ and the underestimate rectangle has height $\frac{1}{k+1}$. Summing this from $x = 1$ to $x = n$ and comparing it to the area under the curve, we have

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_1^n \frac{1}{x} \mathrm{d}x = [\ln x]_1^n = \ln n \leq \sum_{k=1}^{n} \frac{1}{k}$$

So we've bounded the partial sums of the harmonic series i.e

$$\ln n \leq \sum_{k=1}^{n} \frac{1}{k} \leq \ln n + 1$$

Since we know that $\ln n \to \infty$ as $n \to \infty$, we have the divergence of the harmonic series and the knowledge that the harmonic series grows like $\ln n$. This is very slow growth, for example, after summing 1 million terms, we have $13.8 < H_{1000000} < 14.8$. The difference between the rectangles and the area under the curve is always a non zero positive number so we know that $H_n - \ln n$ is an increasing sequence which is bounded above by 1 so it converges to some number γ:

$$\lim_{n\to\infty} H_n - \ln n = \gamma \approx 0.577\ldots \tag{3}$$

This constant γ is known as the Euler Mascheroni constant and appears throughout mathematics and physics in areas such as analytic number theory, statistics, quantum information and quantum field theory. Despite its ubiquity like that of π and e, unlike those two, its not known whether γ is irrational.

We can extend the integral test to a wider range of series. Our argument only relied on the fact that $f(n) = \frac{1}{n}$ is a positive and decreasing function. It links together the sum and integral of a function f, more precisely that for a positive integer N,

$$\sum_{n=N}^{\infty} f(n) \quad \& \quad \int_N^{\infty} f(x)\mathrm{d}x \tag{4}$$

either both converge or both diverge.

These conditions are fulfilled for $f(n) = \frac{1}{n^s}$ where $s \neq 1$ is a real number

$$\int_1^{\infty} \frac{1}{x^s}\mathrm{d}x = \left[\frac{-1}{(s-1)x^{s-1}}\right]_1^{\infty} = \begin{cases} \dfrac{1}{s-1} & \text{if } s > 1 \\ \infty & \text{if } s < 1 \end{cases}$$

This implies by the integral test and our knowledge of the harmonic series that

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \begin{cases} \dfrac{1}{s-1} & \text{if } s > 1 \\ \infty & \text{if } s \leq 1 \end{cases}$$

Prime Numbers

Euclid's proof establishes that there are infinitely many primes. Knowing the harmonic series, as part of the question of how many primes there are, it's an interesting question to wonder whether

$$\sum_{p \text{ prime}} \frac{1}{p} \tag{5}$$

diverges. This will give us an idea of how many primes there are - if the sum diverges then there are 'lots' of primes in this sense.

In the 18th century, Euler introduced an interesting function which has a special relationship with primes. It's known as the zeta function:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{6}$$

We know by the previous work that this converges for $s > 1$ and diverges otherwise so our function is defined only for $s > 1$. He discovered that

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}} \tag{7}$$

which is known as the Euler product for the zeta function.

Proof: If you imagine expanding out the product and collecting the terms, since

$$\prod_{p \text{ prime}} \frac{1}{1-p^{-s}} = \prod_{p \text{ prime}} (1 + p^{-s} + p^{-2s} + \ldots)$$

you'll pick up $\frac{1}{n^s}$ exactly once because of the fundamental theorem of arithmetic. The proof essentially follows this intuition but with a finite product as we don't know about its convergence yet. We have

$$\begin{aligned}\left|\zeta(s) - \prod_{p\leq x}\frac{1}{1-p^{-s}}\right| &= \sum_{n=1}^{\infty}\frac{1}{n^s} - \prod_{p\leq x}(1+p^{-s}+p^{-2s}+\dots) \\ &= \left|\sum_{n\in S} n^{-s}\right| \leq \sum_{n=x}^{\infty} n^{-s} \to 0\end{aligned}$$

as $x \to \infty$. Here, S is the set of numbers which involve a prime factor larger than x and in the last inequality, we used the triangle inequality.

From the prereading, we worked out that

$$\ln(1-x) = -\sum_{n=1}^{\infty}\frac{x^n}{n} \tag{8}$$

Taking the logarithm of the Euler product and using this series would give a leading order term of $\frac{1}{p^s}$ summed over primes. But we know that for $s = 1$ that $\zeta(1)$ diverges as it's the harmonic series and so, the sum of $\frac{1}{p}$ should diverge and behave like $\ln\ln x$. Following this intuition, we can write a proof.

Proof: Taking the logarithm of the Euler product and using the Taylor series of $\ln(1-x)$, we have

$$\begin{aligned}\ln\left(\prod_{p\leq x}\frac{1}{1-\frac{1}{p}}\right) &= -\sum_{p\leq x}\ln\left(1-\frac{1}{p}\right) \\ &= \sum_{p\leq x}\sum_{k=1}^{\infty}\frac{1}{kp^k} \\ &= \sum_{p\leq x}\frac{1}{p} + \sum_{p\leq x}\sum_{k=2}^{\infty}\frac{1}{kp^k}\end{aligned}$$

Analysing the second term,

$$\begin{aligned}\sum_{p\leq x}\sum_{k=2}^{\infty}\frac{1}{kp^k} &\leq \sum_{p\leq x}\sum_{k=2}^{\infty}\frac{1}{p^k} \\ &= \sum_{p\leq x}\frac{p^{-2}}{1-p^{-1}} = \sum_{p\leq x}\frac{1}{p(p-1)} \leq \sum_{n=1}^{\infty}\frac{1}{n(n-1)} = 1 < \infty\end{aligned}$$

Now we know that the product behaves like $\ln x$ for large x because of the harmonic series so

$$\sum_{p\leq x}\frac{1}{p} \approx \ln\ln x$$

So the sum diverges - we have enough primes that the sum diverges. We never used the fact that there were infinitely many primes in any of these proofs so this establishes that fact too. Similarly to γ, we have another constant known as the Meissel-Mertens constant

$$M = \lim_{n\to\infty}\sum_{p\leq x}\frac{1}{p} - \ln(\ln(n)) \approx 0.261\dots \tag{9}$$

It diverges very slowly, for $x = 10^6$, we have that the sum is approximately 2.63.

The zeta function was generalised by Riemann in 1859 to the Riemann zeta function which is defined on the complex plane. With this generalisation, new relationships of the Riemann zeta function with primes were discovered such as the Riemann Von Mangoldt explicit formula

$$\psi(x) = x - \sum_{\rho}\frac{x^\rho}{\rho} - \ln(2\pi) - \frac{1}{2}\ln\left(1-\frac{1}{x^2}\right) \tag{10}$$

where the sum is over non trivial zeroes ρ of the Riemann zeta function and $\psi(x)$ is a function defined in terms of primes

$$\psi(x) = \sum_{k=1}^{\infty} \sum_{p^k \leq n} \ln p \tag{11}$$

This highlights the relationship between the distribution of primes and the Riemann zeta function and the Riemann hypothesis, which is about whether all the zeroes lie on the line $\mathrm{Re}(s) = \frac{1}{2}$, will give a lot of information about the distribution of primes.

Other sets of primes

We've established the infinitude of primes using calculus. Now more questions arise. Which subsets of primes have their reciprocals diverge? A good example is primes which are $p \equiv a \bmod d$ for a and d coprime i.e are there infinitely many primes in the sequence

$$a + d, a + 2d, a + 3d, a + 4d, \ldots$$

We must take a and d coprime otherwise none of these terms will be prime due to this shared factor. Our intuition is that primes are random so they should be equally distributed among these different residue classes mod$d$. To quantify that in terms of sums, we expect that for

$$\sum_{\substack{p \leq x \\ p \equiv a \bmod d}} \frac{1}{p} \approx \frac{1}{\phi(d)} \sum_{p \leq x} \frac{1}{p} \tag{12}$$

and if we can prove this, we know there are infinitely many primes a mod d. This is what Dirichlet did in 1837 using an approach similar to the above using something known as a Dirichlet L function

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} \tag{13}$$

where χ is known as a Dirichlet character and it serves to filter out the primes we want. The approach is the same as above, to show that this series diverges so that we have infinitely many primes a mod d. This proof is considered the starting point of analytic number theory, the area of mathematics concerned with using methods of analysis to prove theorems about integers. It's too difficult and lengthy to include here though - a proof can be found in Davenport's Multiplicative Number Theory. This introduced the idea of L functions too - functions of number theoretic importance which have properties similar to the zeta function and the Dirichlet L function. This continues to be an active area of research.

Another interesting subset is the twin primes - primes p such that $p + 2$ is a prime. It turns out the sum over twin primes converges to a number known as Brun's constant:

$$\sum_{p, p+2 \text{ prime}} \frac{1}{p} + \frac{1}{p+2} \approx 1.90216...$$

This was proven by Brun in 1919 and also has historical importance as the introduction of sieve methods, an important tool in analytic number theory. The convergence of this sum means we don't learn about whether there are infinitely many twin primes, which is still a famous open problem. But there has been recent progress on this (Yitang Zhang, James Maynard in 2013) - this is an active area of research.

Conclusion

With the introduction of the zeta function, we established the divergence and growth of the sum $\sum_{p \leq x} \frac{1}{p}$ and showed how numerous generalisations could be made which point towards different active avenues of research in analytic number theory.