"Mathematicians have tried in vain to this day to discover some order in the sequence of prime numbers, and we have reason to believe that it is a mystery into which the mind will never penetrate." - Euler. How successful has mathematics been in finding 'order in the sequence of prime numbers'?

Often described as the 'atoms' of mathematics, primes have fascinated and inspired mathematicians throughout history. We consider the extent to which we can find 'order' in the primes to be a measure of our understanding of patterns in the prime sequence. Finding regularities in this seemingly random and chaotic list is one of the greatest mathematical challenges. To quote Marcus du Sautoy, 'The list of primes is the heartbeat of mathematics, but it is a pulse wired by a powerful caffeine cocktail.' Unlike the sequence of squares, for example, where the nth square is given by n^2, we have no usable formula allowing us to explicitly determine the nth prime. We address the successes (and failures) of mathematics in understanding the distribution, spacings and properties of the primes, in a quest to find order.

A prime is a positive integer with exactly two positive factors: one and itself. The Ancient Greek mathematician, Euclid, was the first to really establish the centrality of the primes in mathematics. In his ground-breaking work, 'Elements', he offered two proofs, which are still at the heart of number theory[1] today. Firstly, he demonstrated that there are infinitely many primes, with this exquisite proof by contradiction:

Assume (for contradiction) that there are finitely many primes, which we label $p_1, p_2, p_3, \ldots, p_n$. Then we can consider the number $N = p_1 p_2 p_3 \cdots p_n + 1$. When we divide N by each of $p_1, p_2, p_3, \ldots, p_n$ we will be left with remainder 1, so N is not divisible by any of these primes. Therefore N must itself be prime, or its prime factorisation contains only primes which are not in our original list. This contradicts our assumption that $p_1, p_2, p_3, \ldots, p_n$ contained all the primes. Therefore, there are infinitely many primes.

Euclid also presented the Fundamental Theorem of Arithmetic, which states: *Every positive integer greater than 1 can be written uniquely as the product of primes.*

Both theorems are still at the core of all work on primes today. These weapons in hand, we can now explore the primes more deeply.

0.1 A composite tool-box

Whilst mathematicians have a limited understanding of the primes, the composite numbers are much more predictable. We know, for instance, that multiples of m occur every m numbers in the list of integers, and that the nth multiple of m is mn. It is easy to establish order and patterns in the composites, which makes it surprising and even infuriating when primes are so evasive. Therefore, the most effective 'tools' mathematicians use to understand the primes often invoke the composites. We introduce two key ideas here.

(A) Sieve Theory:

Sieve theory is a set of techniques used to investigate the primes, involving filtering out composites. Primes can be thought of as blocks, which can be stacked to form composites. Constructing a sieve with holes small enough for only the single blocks (the primes) to pass through allows us to remove all the composites. The most famous example is the *sieve of Eratosthenes*[2]. It gives a systematic way to determine all the primes in a given range, by crossing out all their multiples; any uncrossed numbers are then prime. Modern sieves are typically much more complex, giving weights to different numbers, and are particularly useful for finding bounds and approximations.

[1] Number theory is the study of the integers and their properties.

[2] The sieve of Eratosthenes was the work of the Ancient Greek, Eratosthenes (276-194 BCE).

(B) The 'almost' primes:
The almost primes are integers with very few prime factors. P_n is often used to denote the set of positive integers with n prime factors (counted with multiplicity), and the study of almost primes focuses on small values of n. Importantly, the primes are a subset of the almost primes, so taking the almost primes in their entirety, it is possible to derive properties about the primes themselves.

These concepts have provided a solid foundation for deciphering the primes. In 1966, Chinese mathematician, Chen Jingrun, used these ideas to make immense progress on Goldbach's conjecture, a famous unsolved problem which asks if every even integer greater than 2 is the sum of two primes[3]. Chen proved that every 'large' even number (greater than $e^{e^{36}}$) may be written as the sum of a prime and an integer with at most two prime factors (an almost prime).

0.2 Spacings in the primes

Mathematicians have been puzzling over the spacings between the primes for millennia, and we have certainly had several advancements in this area. Bertrand's postulate[4] is of fundamental importance, as it bounds these spacings. It states that *for any $n > 1$ we can find a prime between n and $2n$*. Many interesting results can be derived from this. For example, for any natural number k, we can find a k-digit prime, as there will always be a prime between 10^k and 2×10^k. In fact, it is an essential breakthrough in establishing order. Importantly, it gives us an (albeit loose) upper bound on the distance between consecutive primes.

An incredible breakthrough was made in 2004, when Terrence Tao and Ben Green proved that there are arbitrarily long arithmetic sequences in the primes. This means that for any positive integer k we can find an increasing sequence of k primes whose consecutive terms each differ by the same amount. To do this, they used Szemerédi's theorem (a part of Ramsey Theory) which asserts that if we take a certain percentage of a large enough set (in this case, the almost primes) containing 'long' arithmetic progressions, then we are guaranteed to find these progressions among the subset (the primes).

Such was this success that Tao was awarded the Fields Medal. However, it is limited by the fact that it gives us no indication of how to find these sequences - only that they exist. We have yet to understand where they occur. Furthermore, the proof of this result is reminiscent of our limited understanding of the primes. It was the consequence of the 'hereditary' property of arithmetic sequences which can be passed from set to subset, not of a fundamental breakthrough in our understanding of the primes themselves.

A famous unsolved problem that mathematicians are still grappling with is the Twin Prime Conjecture: that there are infinitely many pairs of primes that differ by two[5]. However, in recent years, some progress has been made. Activity was triggered in May 2013, when a little known lecturer at the University of New Hampshire, Yitang Zhang, proved that there are infinitely many pairs of primes that differ by at most 70 000 000. Of course 70 million is much larger than 2, but this was the first time that someone had been able to bound the infinite spacings between primes. In turn, this sparked a flurry of blog-posts, as well as the creation of several 'Molymath' projects (huge online collaborations). By the middle of April 2014, the bound had been reduced to just 246. It certainly seems possible that mathematicians will close this gap down to 2 over the coming years, and it is impressive progress in finding order.

---

[3] This was conjectured by Christian Goldbach in 1742, in a letter to Euler, who stated: 'I regard this as a completely certain theorem, although I cannot prove it.'

[4] Conjectured by Joseph Bertrand in 1845, the postulate was proved by Pafnuty Chebyshev in 1850.

[5] It is thought that this was first conjectured around 300 BCE

0.3 The nth prime

The key to finding order in the primes is to find a way to determine where the next prime will be: a formula for the nth prime, denoted p_n. We define the 'prime counting function' $\pi(x)$ as the number of primes less than or equal to x. If we have a formula for $\pi(x)$, then we can determine the list of primes - a new prime, p, appears in our list whenever $\pi(p) = \pi(p-1) + 1$.

The prime number theorem[6] allows us to approximate $\pi(x)$. Considered one of the greatest breakthroughs in number theory, it states:

$$\pi(x) \sim \frac{x}{\ln(x)} \qquad \text{or} \qquad \lim_{x\to\infty} \pi(x) = \frac{x}{\ln(x)}$$

In other words, as x gets very large the number of primes less than or equal to x approaches $\frac{x}{\ln(x)}$, where $\ln(x)$ is the logarithm with base e. This allows us to derive an approximation for p_n.

By definition, $\pi(p_n)$ equals n, as there are n primes up to and including the nth prime. Furthermore, from the prime number theorem we know:

$$\pi(p_n) \sim \frac{p_n}{\ln(p_n)}$$

Therefore:

$$n \sim \frac{p_n}{\ln(p_n)}$$

Multiplying through by $\ln(p_n)$ gives:

$$n\ln(p_n) \sim p_n \qquad\qquad (*)$$

Taking logarithms of both sides gives:

$$\ln(n\ln(p_n)) \sim \ln(p_n)$$

From the laws of logarithms:

$$\ln(n\ln(p_n)) = \ln(n) + \ln(\ln(p_n))$$

Therefore:

$$\ln(n) + \ln(\ln(p_n)) \sim \ln(p_n)$$

Dividing through by $\ln(p_n)$ gives:

$$\frac{\ln(n)}{\ln(p_n)} + \frac{\ln(\ln(p_n))}{\ln(p_n)} \sim 1$$

Using the result that $\frac{\ln(n)}{n} \sim 0$, we have $\frac{\ln(\ln(p_n))}{\ln(p_n)} \sim 0$ and therefore:

$$\frac{\ln(n)}{\ln(p_n)} + \frac{\ln(\ln(p_n))}{\ln(p_n)} \sim \frac{\ln(n)}{\ln(p_n)} \sim 1$$

Multiplying through by $\ln(p_n)$ gives:

$$\ln(p_n) \sim \ln(n)$$

Applying this to our result in $(*)$, we can see that:

$$p_n \sim n\ln(p_n) \sim n\ln(n)$$

Thus we have an inexact formula for the nth prime, $n\ln(n)$, establishing a sense of order in the prime distribution. However, we still do not have a way to pinpoint its intricacies, and it is generally agreed that the answer is deeply intertwined with the notorious Riemann-zeta function.

[6] The theorem was first conjectured by Adrien-Marie Legendre in 1798, however it was not proved until 1896 when both Belgian mathematician Charles de la Vallée Poussin (1866-1962) and French mathematician Jacques Hadamard (1865-1963) independently found proofs.

0.4 A complex sequence

First discovered by Euler in 1737, the zeta function offers incredible insight into the behaviour of the primes. It is defined as:

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Although not initially obvious, there is a beautifully direct link to the primes, known as Euler's product formula[7]:

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

The proof is an excellent example of a sieve. We start by taking the first prime, 2. Then we multiply $\zeta(s)$ by $\frac{1}{2^s}$ and subtract from the original function:

$$\zeta(s) - \frac{1}{2^s} \times \zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots - (\frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \cdots)$$

Therefore:

$$(1 - \frac{1}{2^s}) \times \zeta(s) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \cdots$$

Essentially, we remove all the 'even' powers from the denominators. Then we do the same with the next prime, 3, removing all remaining multiples of 3:

$$\left((1 - \frac{1}{2^s}) \times \zeta(s)\right) - \frac{1}{3^s} \times \left((1 - \frac{1}{2^s}) \times \zeta(s)\right) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \cdots - (\frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \cdots)$$

Therefore:

$$(1 - \frac{1}{3^s})(1 - \frac{1}{2^s}) \times \zeta(s) = 1 + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{11^s} + \cdots$$

We know from the Fundamental Theorem of Arithmetic that all natural numbers larger than 1 are the products of primes. Therefore, repeating the above steps for all primes leaves us with 1.

$$(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s})(1 - \frac{1}{7^s}) \cdots \times \zeta(s) = 1$$

We can divide through by $(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s})(1 - \frac{1}{7^s}) \cdots$ to obtain Euler's product formula. In this way, Euler disposed of all the composites, leaving a product entirely involving primes.

So how does the zeta function help us? The breakthrough in understanding its potential came when Bernhard Riemann used analytic continuation[8] to extend its inputs to the complex plane, defining the Riemann-zeta function. This opened up a radical new perspective on the primes. Riemann had peered into a new world, and glimpsed the possibility of order. From here, he conjectured the Riemann Hypothesis, one of the most notorious unsolved problems in all of mathematics.

A complex number has both a real part and an imaginary component (a multiple of $\sqrt{-1}$, denoted i). To visualise complex numbers it is helpful to think geometrically: we can put the imaginary numbers on the y-axis and the real numbers on the x-axis. Then some point (x, y) has real part x and imaginary part y, so it represents $x + yi$.

The Riemann hypothesis concerns the convergences of the Riemann-zeta function. A series is said to 'converge' if it tends to a specific limit when we add on infinitely many terms; otherwise it 'diverges'. When $\zeta(s)$ is extended to the complex plane, for some inputs it converges to zero ('vanishes'). These inputs, known as 'zeros of the zeta function', are divided into two categories: 'trivial' and 'non-trivial'. All non-trivial zeros have real

[7]Note that in this formula, Π refers to the infinite product over all primes.

[8]Analytic continuation is a technique used to extend the domain of a function.

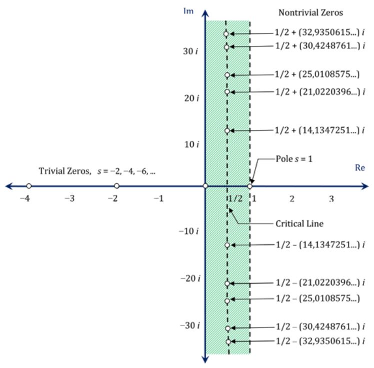

Figure 1: Geometrically, the Riemann hypothesis is equivalent to all non-trivial zeros lying on the line at $\frac{1}{2}$, known as the critical line.

part between 0 and 1, whilst trivial zeros lie outside this range. The Riemann hypothesis conjectures that all non-trivial zeros have real part $\frac{1}{2}$.

These non-trivial zeros hold the key to the primes. In fact, a proof of the Riemann hypothesis would give what has been described as the 'best possible estimate' (with current methods available) for $\pi(x)$. Mathematicians can relate $\pi(x)$ to the coefficients of the series $\frac{\zeta'(s)}{\zeta(s)}$, where $\zeta'(s)$ is the derivative of $\zeta(s)$ with respect to s. In order to estimate these coefficients, we integrate $\frac{\zeta'(s)}{\zeta(s)}$ along a vertical line in the complex plane, with real part equal to r. The best estimate would use the smallest possible value of r. However, the integration is only valid if the half of the complex plane to the right of the vertical line at r does not contain any 'poles' - values at which $\zeta(s)$ equals zero and therefore at which we would be dividing by zero in $\frac{\zeta'(s)}{\zeta(s)}$ - making it undefined. In fact, the prime number theorem is equivalent to not having any zeros with real part greater than or equal to 1; this was actually how it was proven. We have found zeros on the line at $\frac{1}{2}$, so the smallest value that r could take is $\frac{1}{2}$ which would occur if the Riemann hypothesis is true.

A proof of the hypothesis would shed further light on the primes, by pointing to a sort of 'pseudorandomness' in the prime sequence, not dissimilar to that seen when rolling an unbiased dice. We can consider the non-trivial zeros to be this 'dice'. It is unbiased if and only if all non-trivial zeros have the same real part; otherwise they have uneven weights. A 'biased dice' would contribute to a pattern in the primes, whilst an 'unbiased dice' gives a random sequence.

Even more excitingly, Riemann determined and proved an explicit formula for $\pi(x)$, which involves these non-trivial zeros. It uses the function $R(x)$, which is a refinement of the prime number theorem. Like $\frac{x}{\ln(x)}$, the function $R(x)$ tends to $\pi(x)$ as x tends to infinity. Riemann's success was to find an error term, meaning that his function was not an approximation, but exactly equal to $\pi(x)$. The error term involves a sum over the non-trivial zeros of the Riemann-zeta function:

$$\pi(x) = R(x) - \sum_{\rho} R(x^{\rho})$$

Therefore, we start out with $R(x)$, a smooth curve that is asymptotically equivalent to the number of primes up to x, and as we subtract $R(x^{\rho})$ for each non-trivial zero, ρ, the curve begins to indent, forming a wave that ultimately aligns perfectly to our prime counting function. The wave-like function is formed due to the intimate relationship between powers of imaginary numbers and sinusoidal functions, which comes from Euler's formula: $e^{i\theta} = \cos\theta + i\sin\theta$, where θ is in radians.

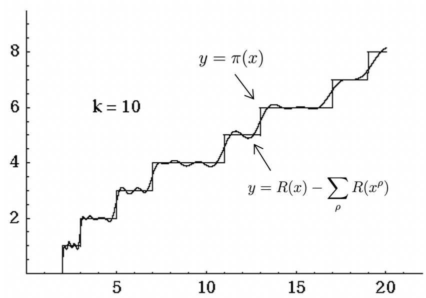

Figure 2: This graph shows the stepwise prime counting function, with Riemann's explicit formula on top. With just ten non-trivial zeros used, we can see the wave already starting to align with $\pi(x)$.

Although we have a limited understanding of these non-trivial zeros and do not have a way to exactly compute the sum required, there is no doubt that this is one of the greatest successes in finding order in the primes.

0.5 Conclusion

As Dr Vicky Neale wrote in her book 'Closing the Gap', *it's easy to ask hard questions.* There is still much we do not know about the primes. We have a very limited understanding of almost all of their properties: their distribution, spacings and even arithmetic properties. Yet after millennia of grappling with the primes, we have certainly had several successes. We have a bound on the spacings between consecutive primes; we know that they contain arbitrarily long arithmetic sequences; and we have even made progress on the Twin Prime Conjecture. We have an accurate approximation for their distribution, and a compelling link between primes and the Riemann-zeta function. Since Euler, we have come a long way in finding order.

However, mathematicians cannot seem to pin down the primes. Powerful computers have generated lists of billions - some with over 20 million digits. We are certainly not suffering from a lack of data. As hinted by the Riemann hypothesis, the primes seem to have a fundamental element of randomness. In recent years, striking similarities have been drawn between the behaviour of Riemann's non-trivial zeros and that of quantum energy levels of chaotic systems. It may even be that these will lead us to the missing piece of the prime puzzle. To quote Paul Erdős, *God may not play dice with the universe, but something strange is going on with the prime numbers.*

Whilst we have made immense progress, we have not yet reached the heart of order in the primes. There is some way to go before we fully understand their eccentricities, and it is a challenge that lies ahead.

[Word count: 2497]

Bibliography

[1] Bombieri, E. (2006). *Problems of the Millennium: the Riemann Hypothesis.* Eds. J Carlson, A Jaffe and A Wiles, Clay Mathematics Institute and American Mathematical Society. Paper available from web archives: https://www.claymath.org/sites/default/files/official_problem_description.pdf

[2] Caldwell, C. (2020). *Prime Conjectures and Open Questions.* Prime Pages. Available at: https://primes.utm.edu/notes/conjectures/ [Accessed February 2021]

[3] Cojocaru, A. C. and Ram Murty, M. (2006). *An introduction to sieve methods and their applications.* London Mathematical Society Student Texts, 66, xii+224 pp. Cambridge University Press, Cambridge.

[4] Devlin, K. (2001). *How Euler discovered the zeta function.* Available at: https://www.maa.org/external_archive/devlin/Zeta.PDF [Accessed January 2021]

[5] Du Sautoy, M. (2004). *The Music of the Primes.* Harper Perennial, London.

[6] Elwes, R. (2013). *Maths in 100 Key Breakthroughs.* Quercus Editions Ltd, London.

[7] Green, B. and Tao, T. (2008). 'The primes contain arbitrarily long arithmetic progressions,' *Annals of Mathematics*, Volume 167 (2), p. 481-547.

[8] Hoffman, P. (1999). *The Man Who Loved Only Numbers.* Fourth Estate Limited, London.

[9] Kontorovich, A. (2016). *How I Learned to Love and Fear the Riemann Hypothesis.* Quanta magazine. Available at: https://www.quantamagazine.org/how-i-learned-to-love-and-fear-the-riemann-hypothesis-20210104 [Accessed January 2021]

[10] Mazur, B. (2014). *"A Lecture on Primes and the Riemann Hypothesis."* Available at: https://www.youtube.com/watch?v=way0jAWpjZA [Accessed January 2021]

[11] Neale, V. (2017). *Closing the Gap: A quest to understand prime numbers.* Oxford University Press, Oxford.

[12] Tao, T. (2009). *Structure and Randomness in the Prime Numbers, UCLA.* Available at: https://www.youtube.com/watch?v=PtsrAw1LR3E [Accessed February 2021]

[13] Tao, T. (2021). '246B, Notes 4: The Riemann zeta function and the prime number theorem,' *What's new*, 12 February. Available at: https://terrytao.wordpress.com/tag/prime-number-theorem/ [Accessed February 2021]

[14] Veisdal, J. (2016). *The Riemann Hypothesis, explained.* Cantor's Paradise. Available at: https://medium.com/cantors-paradise/the-riemann-hypothesis-explained-fa01c1f75d3f [Accessed February 2021]

[15] Watkins, R. (2017). *The "encoding" of the distribution of prime numbers by the nontrivial zeros of the Riemann zeta function.* Available at: http://empslocal.ex.ac.uk/people/staff/mrwatkin/zeta/encoding1.htm. [Accessed January 2021]

[16] Watkins, R. (2017). *The "encoding" of the distribution of prime numbers by the nontrivial zeros of the Riemann zeta function [elegant approach].* Available at: http://empslocal.ex.ac.uk/people/staff/mrwatkin/zeta/encoding2.htm. [Accessed January 2021]

[17] *Prime Number Theorem.* Wikipedia. Available at: https://en.wikipedia.org/wiki/Prime_number_theorem [Accessed January 2021]

[18] *Quantum physics sheds light on Riemann hypothesis.* School of Mathematics, University of Bristol. Available at: http://www.bristol.ac.uk/maths/research/highlights/riemann-hypothesis/ [Accessed January 2021]

[19] Figure 1 source: Sadler, C. (2015). *The Riemann Hypothesis in Simple Terms.* Available at: http://www.hdd23.com/blog/the-riemann-hypothesis-in-simple-terms/21/1/2015. [Accessed February 2021]

[20] Figure 2 source: Baez, J. (2019). *The Riemann Hypothesis (Part 1).* The n-Category Café. Available at: http://www.hdd23.com/blog/the-riemann-hypothesis-in-simple-terms/21/1/2015. [Accessed February 2021]