A tale of two primes

By Hayden Tronnolone

1. Primes apart

There are few opportunities to pair the phrase “by only” with the number “70 million”; however, a recent result in number theory has allowed just such an abuse of scale and language. (Only in maths, right?) The cause of all the fuss is a notoriously difficult open problem, along with its proposed solution: the twin prime conjecture.

A twin prime is a pair of prime numbers that differ by 2; that is, two prime numbers \( p \) and \( p + 2 \). The first few twin primes are (3, 5), (5, 7) and (11, 13), with the twin primes designated sequence A077800 in the Online Encyclopedia of Integer sequences (OEIS, 2013). Since there is only one even prime there is only one set of consecutive primes: 2 and 3. Aside from this singular example, twin primes represent the most compact groupings of primes. Given how close together the first few twin primes are, we may start to wonder if they continue to occur as frequently; however, at first it may appear that twin primes should not continue to occur at all due to a result known as the prime number theorem.

**Theorem 1** (prime number). Let \( \pi(x) \) represent the number of primes less than or equal to \( x \). Then

\[
\lim_{x \to \infty} \frac{\pi(x)}{x/\ln(x)} = 1.
\]

This result may be written as \( \pi(x) \sim x/\ln(x) \), which means that \( \pi(x) \) is “like” \( x/\ln(x) \) for large values of \( x \). From this, we can show that the average gap between prime numbers less than some integer \( n \) is approximately \( \ln(n) \): on average, as \( n \) increases the primes get farther and farther apart. This is bad news for twin-prime enthusiasts since it seems to suggest that the primes spread apart rather than cluster in pairs, leaving a few twins found only amongst the smallest primes. Luckily for us, this is not the case. Exhaustive computational searches have shown that there are 808,675,888,577,436 twin primes less than \( 10^{18} \) (e Silva, 2013), while the largest-known twin prime is \( 3,756,801,695,685 \times 2^{666,669} \pm 1 \), found in 2011 by PrimeGrid’s Sophie Germain Prime Search (PrimeGrid, 2011). Much like the small Indian village of Kodinhi (look it up), it would appear that there is no shortage of twins.

The primes grow farther apart on average but still occasionally occur in pairs, at least as far as we have been able to check.

Long before such searches were possible mathematicians have mused on and debated exactly how many twin primes there are. Such efforts led to the following proposition, known as the twin prime conjecture (TPC):

**Conjecture 1** (twin prime). There are infinitely many primes \( p \) such that \( p + 2 \) is also prime.

While it has often appeared in mathematical literature it is not clear exactly when this conjecture was first made. Prime numbers have been studied since at least the time of Euclid (around 300 BC) and it is likely that the occurrence of twin primes has been noted and questioned by many mathematicians since. In the 19th century de Polignac (1849) proposed a more-general version of the TPC, which is as follows:

**Conjecture 2** (Polignac). For every positive even number \( N \) there are infinitely many prime gaps of size \( N \).

For the special case \( N = 2 \) this is just the TPC. A related idea arises in a seminal paper by
Brun’s constant

Brun’s theorem says that the sum of the reciprocals of twin primes converges. Despite this, it must be emphasised that it is still only a conjecture: a statement based upon intuition and partial results but lacking a rigourous proof. (Before you get too excited at the prospect of conjecturing all your problems away it is also important to note that, while conjectures are an accepted part of mathematics, they are not suitable for assignments.) Contrary to the apparent simplicity of the problem, its longevity and its fame, we can say very little about the number of twin primes outside of what large computations can tell us. It perhaps goes without saying that mathematicians find this sub-optimal.

While the conjecture has remained just that, it has inspired many interesting yet unsuccessful efforts in the quest for a proof, such as those of Erdős (1940), or the computational studies of Bohman (1973) and, more recently, the Twin Prime Search (2013) and aforementioned PrimeGrid (2013). A particularly noteworthy and altogether rather astounding result was discovered by Brun (1919), which is as follows:

Theorem 2 (Brun’s). The sum of the reciprocals of the twin primes converges.

If we let \( \mathbb{P} \) be the set of all prime numbers then Brun’s theorem says that

\[
\sum_{p \in \mathbb{P} : p + 2 \in \mathbb{P}} \left( \frac{1}{p} + \frac{1}{p + 2} \right) = \left( \frac{1}{3} + \frac{1}{5} \right) + \left( \frac{1}{5} + \frac{1}{7} \right) + \left( \frac{1}{11} + \frac{1}{13} \right) + \ldots
\]

is equal to some finite number \( B_2 \), which is known as Brun’s constant\(^1\). Brun’s result is remarkable as he was able to prove it without knowing the value of the sum \( B_2 \) or even how many twin primes there were to sum over! A further curiosity arises as both the harmonic series \( \sum_{n=1}^{\infty} \frac{1}{n} \) and the sum of the reciprocals of all the primes \( \sum_{p \in \mathbb{P}} \frac{1}{p} \) diverge, that is, they grow without limit as more terms are added. So, somehow restricting this sum to the twin primes\(^2\) produces a set “small enough” to produce a finite sum. It is tempting to use this to claim that there must be only a finite number of twin primes; however, sums such as these are decidedly nasty and no such conclusion can be drawn using this reasoning. We could, however, infer something about the twin primes from the nature of \( B_2 \). If this number were irrational then it could not be produced from a finite sum of rational numbers, so we would then know that there must be an infinite number of twin primes. (If \( B_2 \) were rational we could draw no conclusion since both finite and infinite series of rational numbers can sum to a rational number.) Alas, such a result has eluded us too.

Nevertheless, large computations have yielded some interesting, if unintended, results. In 1995, Nicely used computers to find all the twin primes up to \( 10^{14} \) and hence produce an estimate for \( B_2 \) of 1.902160577783278. During this process, he compared results produced by different computers and found them to disagree. After an exhaustive verification of his methods he finally discovered that the error was due to the calculation of the reciprocals of the twin primes 82463702442±1, for which the division algorithm in the Intel Pentium P5 floating point unit produced incorrect results. Halfhill (1995) later determined that the problem was due to human error that had resulted in 5 out of 1066 numbers being left out when they were copied to the lookup table used by the chip, estimating that 1 out of every 9 billion divisions would be inaccurate due to this error. This is sometimes referred

\(^1\)The subscript “2” is added to distinguish this from Brun-type constants associated with primes that differ by some even amount \( d \), which are denoted \( B_d \).

\(^2\)1/5 is the only number to occur in two twin primes. To see why, try puzzle 3 in the puzzles section. Because of this, it will occur twice in the sum of the reciprocals of twin primes but only once in the other two series; however, this does not affect the convergence.
to as the Pentium FDIV bug and resulted in Intel recalling the chips at a cost of $475 million. Not only had the TPC withstood the efforts of computers, it had also identified their shortfalls!

More recently, Sebah (2002) has used all the twin primes up to $10^{16}$ to show that $B_2 \approx 1.902160583104$, with this result recorded as sequence A065421 in the On-Line Encyclopedia of Integer Sequences (OEIS, 2013a). While these results have driven the development of computational techniques, piqued our curiosity and caught a computer error, they have brought us no closer to resolving the conjecture. We cannot determine the nature of $B_2$ from numerical results such as these and, hence, are effectively no closer to proving, or disproving, the TPC.

Given that twin primes have been known for at least 2000 years it seemed that a result was not to come anytime soon. That was until the recent publication of a startling new result.

2. Primes (almost) together

On April 17, 2013, mathematician Yitang Zhang released a paper entitled “Bounded gaps between primes” (Zhang, 2013). This work focuses on proving a single and deceptively esoteric result: if $p_n$ is the $n^{th}$ prime, then

$$\lim inf_{n \to \infty} (p_{n+1} - p_n) < 7 \times 10^7.$$ 

What this means is that for some integer $N$ less than 70 million there are infinitely many pairs of primes that differ by $N$. This result is incredible, ground-breaking, magnificent, awe-inspiring and just about every other positive adjective you can think of. For the first time we can say that there are infinitely many primes that differ by only 70 million, perhaps much less! This may seem like an absurdly large number. For example: the dinosaurs became extinct about 70 million years ago; at its farthest Mercury is about 70 million km from the Sun; while Pinterest has about 70 million users (Semio cast, 2013). Disappointingly, 70 million milliseconds leaves you an agonising 37/3 ≈ 12.34 minutes short of being able to watch all eight Harry Potter movies (you’d need to skip the credits) (Wikipedia, 2013). Despite this cinematic shortfall, 70 million would generally be considered a sizeable number.

The reason this result is impressive is that it represents the first time we can get some handle on the spacing of pairs of primes; previously it was not known whether any bound should exist at all. We finally know that, although the primes become less dense, there are always pairs that stay relatively close together.

Just as remarkable as this result is the story of Zhang himself. He completed his PhD at Purdue University, taking seven years to complete his thesis, which was on the Jacobian conjecture (Moh, 2013), another difficult open problem. After gaining his PhD, he struggled to find a job and spent time working as both an accountant and at a Subway restaurant, amongst other jobs (Klarreich, 2013). He began his work on the TPC after reading a paper from Goldston, Pintz and Yildirim (2009), originally released on the arXiv in 2005 (Goldston, Pintz and Yildirim, 2005) and referred to as GPY after the authors’ initials, on the distribution of small gaps between primes. This work, one of a number by the same group, showed that there are infinitely many primes less than the predicted average spacing; however, the authors could not demonstrate conclusively that there is some upper bound on this gap. (With some assumptions they were able to show that there are infinitely many primes that differ by 16, but this relies on other unproven results.) Their method was based on the idea of sieving, which is used to select numbers that satisfy certain properties. A simple example of this technique is the Sieve of Eratosthenes, which is taught as early as primary school and is used for finding primes. Commenting on the existence of infinitely many bounded prime pairs, Goldston, Pintz and Yildirim noted

---

*I hope my supervisors are reading and appreciate my relative swiftness.*
that their first theorem “would appear to be within a hair’s breadth of obtaining this result”; however, no significant further progress was made.

After GPY was first communicated Zhang began to think he could extend the method (Klarreich, 2013). Working in solitude, he developed new methods that eventually led to his 2013 result. He considers a set $\mathcal{H} = \{h_1, h_2, \ldots, h_{k_0}\}$ of distinct non-negative integers, where $k_0$ is a parameter to be determined. Under certain conditions, too complicated to mention here\(^4\), $\mathcal{H}$ can be said to be admissible. Zhang’s main result is to show that for an admissible $\mathcal{H}$ and a particular value of $k_0$ there are infinitely many integers $n$ such that $\{h_1 + n, h_2 + n, \ldots, h_{k_0} + n\}$ contains at least two primes. This infinite number of translated sets guarantees that there are infinitely many pairs of primes with the two primes in each pair some set distance apart that is at most the difference between the smallest and largest elements of $\mathcal{H}$. To ensure that $\mathcal{H}$ is admissible Zhang takes this to be a set of $k_0$ distinct primes all greater than $k_0$. He notes that

$$\pi(7 \times 10^7) - \pi(3.5 \times 10^6) > 3.5 \times 10^6,$$

which means that there are at least $3.5 \times 10^6$ primes between $3.5 \times 10^6$ and $7 \times 10^7$. Thus, by taking $k_0 \geq 3.5 \times 10^6$ we are able to find $k_0$ distinct primes greater than $k_0$ that can be used to construct $\mathcal{H}$. Since this means that all the numbers in $\mathcal{H}$ are less than $7 \times 10^7$, we arrive, in a somewhat crude fashion, at the upper bound in Zhang’s theorem.

The value of $k_0$, and hence also the bound of 70 million, can be modified through the choice of another parameter $\varpi$. Zhang’s work requires a particular value $1/4 + \varpi$ to be close $1/4$ but neither too far above nor too far below; Zhang eventually settled on $\varpi = 1/1168$. When asked about this choice at a conference at the University of Massachusetts in June 2013, Zhang is said to have replied simply that “I was tired . . . and this number worked” (Aczel, 2013), a sentiment remarkably reminiscent of almost every undergraduate assignment the author of the current article ever worked on. Despite this, it would be unfair to accuse Zhang of laziness: he had worked on the problem for three years without success. The breakthrough came in 2012 while Zhang was visiting a friend and waiting in a backyard to leave for a concert. Discussing this moment, Zhang commented that “I immediately realised that it would work” (Klarreich, 2013). Chalk up another win for taking a holiday.

It must also be noted that, despite the large bound in the final proof, Zhang’s work is a significant breakthrough. The potential for a better bound was not lost on him, however. The introduction to his paper concludes with the following alluring remark: “This result is, of course, not optimal. The condition $k_0 \geq 3.5 \times 10^6$ is also crude and there are certain ways to relax it. To replace the right side of (1.5)\(^5\) by a value as small as possible is an open problem that will not be discussed in this paper.” With this, the challenge to find a better bound had been set.

3. Primes even closer

Since the release of Zhang’s work there has been a burst of activity. Once his results were verified, other (less-tired) mathematicians began to work on reducing the bound of 70 million by modifying his assumptions, curious as to whether this could be reduced to 2, providing a proof the TPC. In particular, it drew the attention of Australian-born mathematician Terry Tao and the Polymath project. Polymath seeks to solve mathematical problems online using many different mathematicians working in collaboration. Results are shared through blogs and wikis with any conclusions generated published under the pseudonym D. H. J. Polymath. On June 4 this problem was set as the 8th Polymath project, imaginatively dubbed Polymath8 (Tao, 2013). 

---

\(^4\)But I will anyway. We need $v_p(\mathcal{H}) < p$ for every prime $p$, where $v_p(\mathcal{H})$ denotes the number of distinct residue classes of the $h_i$ modulo $p$ (Zhang, 2013).

\(^5\)This is the equation number of the bound in Zhang’s theorem.
By the end of May, the bound had dropped to 42,342,946, corresponding to \( k_0 = 2,618,607 \). A week later a new parameter \( \delta \) had been added and it was shown that we could allow \( 828\varpi + 172\delta < 1 \), dropping the bound to 387,534. As of July 20, it has been verified that with \((93 + 1/3)\varpi + (26 + 2/3)\delta < 1 \) and \( k_0 = 720 \) the bound is a mere 5414 (Polymath8 Wiki, 2013).

4. Towards 2?

It remains to be seen just how low this bound can be pushed using Zhang’s methods. Whether or not this will yield the long-sought twin prime nirvana it will forever stand as a remarkable piece of mathematics. Perhaps most interesting are the methods employed, with a striking contrast between Zhang’s isolation and the internet-based Polymath project. If evidence of the internet’s influence over modern mathematics is needed then look no further than the number of references to websites and internet projects contained in this article. There can be no doubt that the way maths is done is changing and that it is for the better, pushing us ever onwards towards new and better results faster than has ever been possible before. Yet, it cannot be ignored that the spark for much of this work came from a single mathematician and a sudden moment of inspiration. It seems the future of maths will involve both approaches working in tandem.

As for Zhang, he is quite content with his role in the matter. “My mind is very peaceful. I don’t care so much about the money, or the honour,” he notes, going on to add that “I like to be very quiet and keep working by myself.” He is said to be already at work on his next project with some promising results to show (Klarreich, 2013). Given his propensity for tackling difficult problems, don’t be surprised to see his name attached to the Riemann hypothesis in a decade from now. For the time being, we can only watch and marvel as this mathematical quest plays out before us.

Hayden is avoiding a PhD in applied maths.

References


Bohman, J. (1973), ‘Some computational results regarding the prime numbers below 2,000,000,000’, Nordisk Tidskr. Informationsbehandling (BIT) 13, 242–244.


