2 Years of Maths Society

Maths Society 2023-25

The Story of Maths Society

In year 11, I realised that I would soon be able to run a society of my own at the school and I noticed that even though there were some STEM related societies, a strictly maths society was missing. And so I began to think about starting one in the following year, as I was sure there would be a great appetite for a maths society at Latymer. Therefore I excitedly looked out at the start of year 12 for information on how to start a society, and shortly thereafter, maths society came into existence.

You could tell that we weren't exactly the most experienced society when we made our entrance to the societies fair- we hadn't set up a stand! Thankfully Dominic was a musician and a quick thinker, so we quickly rushed to the library and grabbed some A4 paper and a music stand and set up shop in front of the entrance with our slightly less glamorous display. Nevertheless, the enthusiasm of the people of Latymer for maths prevailed and we managed to secure many emails for the maths society weekly email (including upset year 7-9s who didn't have the same lunchtime as us and thus couldn't attend).

The next issue was which room we would be located in. There was a meeting in the English corridor at break one day designating rooms to each society, which I had completely forgotten about but thankfully Saik was there and called me so I eventually showed up a tad late and our room had been secured. Despite the slightly rocky start, we now had a society which had a room and had interested members- we would not need anything else! And so, the dive into mathematics and its mysteries and secrets began...

Year 1

In the first year of maths society, we were able to use Mr Lunn's room 56 without fail as he didn't work on Mondays. With this stability, maths society was able to explore chaos and began its journey by looking at dynamical systems. The first ever talk on the Collatz conjecture was packed (probably the highest attendance ever with close runners up being Akshayan and Danica's talk on the golden ratio and my talk on the Chinese remainder theorem). Even though I was using the small whiteboard next to the digital one because I didn't realise we could use the digital one, the talk went down well and whetted everyone's appetites for further discussions on dynamical systems. Therefore we dived in, talking about the famous Mandelbrot set and then Newton's fractal and how it relates to the Mandelbrot set in the most unexpected way. These talks were followed up by me on Youtube in a video entitled "The Mandelbrot set is Universal", which is based on a paper with the same title.

The next saga was, in my opinion, the peak of maths society. We talked about the Fibonacci numbers and the golden ratio in great depth, beginning with a discussion on continued fractions that beautifully set up some "shocking connections" (I remember this being a hit at the time, especially with Mithush) between the golden ratio and the Fibonacci numbers and also between the Pell numbers and the quadratic $x^2 - 2x - 1 = 0$. This was a wonderful place to stop and develop a general formula for the nth Fibonnaci number, F_n, in terms of the golden ratio. This set us up wonderfully to show that x is a Fibonacci numbers iff $5x^2 + 4$ or $5x^2 - 4$ is a perfect square, which is a neat test. To finish this saga off, we generalised our results not only to Pell numbers but also to any sequence of the form $M_i = nM_{i-1} + M_{i-2}$.

The third saga was unfortunately nowhere near as good, with a half baked exposition of transcendental numbers being provided. I delivered a talk on the Liouville numbers, but fumbled around with the technical details of the proof and got mixed up mid talk, correcting myself later in the notes. The saga unfortunately didn't get any better with the next "talk" being us watching a YouTube video by mathologer on why e and π are transcendental. This was really a low point of the whole two years in maths society.

I knew this couldn't run, and thankfully the next saga was back to the highest standards. We dived into some more philosophical and fundamental questions about mathematics, by talking about set theory and logic. To start off with, we deconstructed the notion that naive set theory is a consistent model for set theory and thus talked about the ZFC axioms, which are what mathematicians use today and assume as true. To cap off the talk we looked at what every set looks like in this model of set theory, by looking at the Von Neumann universe. Next we zoomed in on the natural numbers, examining them carefully and defining exactly what we meant by addition, and indeed what exactly the numbers themselves mean. Building upon this, we were able to show that 1+1=2 and that $1a = a$. However, this was too nice and comforting so we had to shake things up with an existential threat to the foundations of mathematics- Gödel's Incompleteness Theorems, which could be a threat to the consistency of our current model of set theory.

After this deep and scary saga, we moved back to some number theoretical considerations, this time in analytic number theory. We began with Euler's classic proof that $\zeta(2) = \frac{\pi^2}{6}$, before Prasan came along and gave an extremely original proof of the Euler product for the Zeta function using probability theory. To finish off this saga, I briefly discussed the Gamma function and how one can make sense of $\left(\frac{1}{2}\right)!$ (and why it turns out to be $\frac{\sqrt{\pi}}{2}$).

Of course the first year would not have been anywhere near as good as it was without the valuable contributions of the guest talks. To get the ball rolling, Wren delivered an eye opening exposition of the Zeta function and what the Riemann hypothesis is, giving us a method for extending the Riemann function to $\mathrm{Re}(s) > 0$ from $\mathrm{Re}(s) > 1$ and updating us on how far people are from solving the mystical problem, which has a value of £1000000 for solving it. We didn't know it then, but that would go on to be the first of many excellent talks by Wren.

Next up we had one of maths society's all time classic talks that still gets mentioned to this day- Prasan's Taylor series talk. Although his talk on the Euler product featured arguably more interesting mathematics itself, the delivery of this talk was inch perfect and a great inspiration for future guest speakers (and myself). In this talk Prasan gave us some fun history about Brook Taylor and then explained Taylor series using the example of e^x (he let the audience choose) and then extended it to any (sufficiently nice) function to give us the general Taylor series formula.

Next up, Wren delivered another (highly underrated) talk on differentiation under the integration sign. In this talk we covered the approach on how the method works and then applied it to evaluate the famous integral $\int_0^\infty \frac{\sin(x)}{x}\mathrm{d}x$ and then the integral $\int_0^1 \frac{x-1}{\ln(x)}\mathrm{d}x$.

The next guest talk marks the time that I missed maths society. I was over in Bielefeld on a German exchange trip and I needed someone to do a talk. Thankfully the perfect man for the job stepped up- Alex Dowling. Alex Dowling is one of the school's most renowned mathematicians and when he stepped up to deliver the talk, boy did it deliver. He spoke about Fourier transforms, providing good physical intuition via sound waves and constantly showering us with diagrams to hammer home the intuition behind the Fourier transform formula. For those of you wondering how I know that he talked about this, I'd like to answer by thanking Saik for recording the talk on his phone and sending me the voice note. Originally the plan was for me to watch the talk via a call but unfortunately I was still at the primary school that I was working at so I could feel my phone buzzing constantly and when I went to check what was causing it to blow up I sadly put the phone back in my pocket and pondered what I was missing out on. Nevertheless, it must be said that this was another one of maths society's all time great talks, right up there with Prasan's dissection of Taylor series.

Next up we had a talk from the man himself Rowan Oliveira. Before this talk I was really looking forward to this talk because I had heard word on the street that it was going to be about the logistic map, and also because it was Rowan presenting it. The logistic map is an iterative map that has many interesting properties. In this talk we looked at what happens to the logistic map (defined by $x_{n+1} = rx_n(1 - x_n)$, with x_0 being between 0 and 1) as r varies and uncovered a shocking connection with the Mandelbrot set which greatly pleased me.

Rounding up the guest talks from this year we had Saik, one of the most fundamental members of maths society, delivering a talk. Saik is quite a physically minded person (his favourite area of A level maths is mechanics, and he wants to do engineering) and so we looked at something more of that flavour- more specifically we investigated the derivatives and integrals of displacement. In this talk we learnt about weird concepts like jerk, crackle and pop as well as absement, alongside their physical applications too.

Finally I should also mention one of maths society's most notorious talks- Danica and Akshayan's presentation on the golden ratio. Although the talk is not in the notes, (it was done via a powerpoint presentation, the only talk at maths society to boast such an accolade) it was still one of maths society's most well attended talks and brought quite a fresh energy to maths society with fun activities and sweets on offer too! It would also be an injustice not to mention the role that wikirace played in the second half of this year, as we enjoyed many games of wikirace if the talks were over before lunch had ended as a way to amuse ourselves.

Year 2

Year 2 of maths society began with a tumultuous time. Firstly, our staff sponsor Ms Edwards was running a problem solving society on Mondays and so in order to ensure that neither us nor the problem solving society lost members, maths society moved to Wednesdays. Furthermore room 56 was no longer available for us. Once we had filled out the necessary form to restart maths society for our second year, we were assigned Room 55, but a mix-up with the communication occurred and Ms Smith hadn't been informed that we were to be using her room during Wednesday lunch. Thankfully we managed to move next door to room 54, Ms Perrin's room that week, although the status of which room we were going to be in in the future was very much up in the air. A couple of days later, Ms Smith informed me that she had discussed with the teachers in the maths department and that we could use Mr Jeevagan's room on week 1s and Ms Perrin's room on week 2s. And so, even though the room situation was less secure, maths society was back up and running.

The first saga was on modular arithmetic, a surprisingly powerful tool in mathematics which studies numbers by considering their remainders when divided by other numbers (for example $6 \equiv 1 \equiv -4 \mod 5$ because they all have the same remainder when divided by 5). In the first talk we saw some cool applications like how to determine the last digit of 2^{1000} or how to determine if 224678946 is a square number or not. Having now seen the motivation for why modular arithmetic was so powerful, we then looked at some properties in modular arithmetic that we could perhaps use later. Next came the Chinese remainder theorem which is an extremely powerful theorem and can give some surprising (almost magic) results, and finally looked at how equivalence classes can be useful.

The next saga was a brief but insightful look at the tools of algebraic number theory and more specifically the Gaussian integers. We saw how we can look at Gaussian integers (numbers of the form $a+bi$, $a, b \in \mathbb{Z}$) and then get results about $\mathbb{Z}$, which is very surprising.

The next saga was on Fourier series, which I was sure would go down a treat as Taylor series was so popular, and Fourier series are very similar to Taylor series just with trig functions instead of polynomials. Giving examples all along the way, we saw how to construct both the sine and cosine Fourier series, before seeing the complete Fourier series and the extremely powerful piece of machinery: Parseval's theorem, which allowed us to find the much less known sum of the reciprocals of the 4th powers. Finally, Wren wrapped up the saga with a great talk on the Poisson summation formula which is very beautiful and an example of it in action which was helpful in seeing why it's so useful. Personally, this saga was definitely my favourite of year 2 and has an argument for being one of the best sagas in all of maths society history.

The final saga was on set theory, mainly focusing on the Cantor set. The Cantor set is so interesting because it challenges lots of common misconceptions about infinite sets which can come from extending intuition from finite sets to infinite sets. In the discussion of the Cantor set we took a brief tour through other interesting topics: ternary decimals and the set theoretical way of defining whether one set is "bigger" than the other.

As for the guest talks this year, we had Kathleen giving a talk on Taylor series in which we saw how Taylor series can be used to approximate $\cos x$ and very shortly after how to derive Euler's formula $e^{ix} = \cos x + i \sin x$ using Taylor series which was interesting. But the highlight of the guest talks this year was definitely Vincent and his exposition of his new statistical methods of number, clump and spread. He gave us the motivation using his timetable and then carefully defined number, clump and spread in order to study what was really happening on his timetable. For those interested, don't bother trying to look up number, clump and spread up online: this is a maths society exclusive!

We must also mention the introduction of the problems of the week. These were a good addition to the society, because in year 1 after a talk was finished there wasn't much left to do so we would normally end up on Wikirace. However, with the introduction of the problem of the week, such problems didn't exist as there was always something interesting to chew on even after the talk was over, ensuring that we had a whole lunchtime filled with mathematics.

Acknowledgements

Firstly, I would like to thank the teachers that helped this society to run: firstly Ms Edwards who was the staff sponsor for the society both in years 1 and 2; secondly I'd like to thank Dr KC, who helped us in the first year by always being on hand to give us the keys for room 56 when it was locked and occasionally popping in and making amusing comments that made the soci-

ety more memorable (for example her quip on the gender disparity or when she tried to force Thapisan to join); thirdly I'd like to thank Ms Perrin who kindly gave up her room once a fortnight for us and was also one of the most enthusiastic recipients of the weekly email (definitely the most enthusiastic of the teachers!); fourth I'd like to thank Mr Jeevagan for also giving us his room once a fortnight during the second year and I'd also like to thank Mr Billington for giving us the keys to Mr Jeevagan's room when it was locked; fifth I'd like to thank Ms Smith for occasionally giving us her room in year 2 when we didn't have one for whatever reason. Without these teachers, maths society would not have run nearly as smoothly, so thank you to all of you.

Now, onto the students who made maths society what it was. First and foremost, I'd like to thank the (often absent!) vice president Saik, who may not have been in school very often but certainly made sure his maths society attendance (given that he was in school at the time) was 100%. His contributions to the tech at maths society were invaluable (many talks given on the digital board may not have occurred without him) and he helped to get people in the door which was greatly appreciated. Next, this acknowledgements section would be incomplete without mentioning the one and only Dominic Kamel. To those who attended maths society, this may be a mysterious inclusion in the list, but Dominic was helping maths society from day one. Without Dominic nowhere near as many people would even know that maths society existed, because he is the reason we even had a stand at the societies fair in year 1, and in year 2 when we had a stall, Dominic put together one of the most incredible presentations I've ever seen which certainly boosted maths society attendance by a lot (especially with the younger years). Thirdly, I have to thank Henry, another very regular attendee (probably the most regular one), who filled in on the tech side when Saik wasn't in and often provided amusing commentary before, during and after the talks which helped to give the society the entertainment factor. Fourth, I'd like to thank Kathleen who was easily the best year 12 member of maths society; she was a very regular attendee, gave a talk (something no other non year 13 can boast!) and is probably the reason for almost all of my other year 12 attendance. She is the reason that I have any hope for the future of maths society and if she decides to take it over next year then maths society is in very safe hands. Fifth I'd like to thank all of the guest speakers: Wren for providing some very deep insights on calculus and complex analysis which were extremely interesting. Prasan: I'm lucky that guy didn't start his own maths society because it almost certainly would've been more popular than mine: his talks were extremely well presented and still get mentioned to this day in the discussions of the all time great talks. Alex was a legendary mathematician among our year, so hearing his insights into the Fourier transform was extremely enlightening. Then of course thee is the legendary Rowan, a renowned scientific mind in year 13, his talk was one of the most interesting things I've ever heard and that didn't even surprise me, I'm just sad we only got the one talk from him. Of course there's Saik, whose contributions have already been mentioned. Kathleen, the only year 12 speaker- her talk on Taylor series helped to provide a new perspective from Prasan's talk the previous year. Finally, Vincent who treated us to his own statistical methods of number, clump and spread which were motivated by his pursuit of studying the school timetable, which was incredible to see as it was exclusive to maths society! Finally, I'd like to thank all of the regular attendees (you're likely in the image at the start) and all of the regular readers of the maths society weekly email: without you I may have stopped sending the emails.

Contents

Part I

Year 1

Chapter 1

Saga 1- Dynamical Systems

1.1 Collatz Conjecture

1.1.1 Introduction

The Collatz conjecture is an infamous problem which was posed in 1937. It's simplicity has lured many mathematicians into thinking they can solve it, yet nobody has done so so far (there was even a conspiracy that the Soviets posed this problem to the West to deliberately slow down their mathematicians' progress!). The problem is as follows: Pick a number n.

- If n is odd, multiply by 3 and add 1.
- If n is even, divide it by 2.

For example, if we start with $n = 5$ we would get the sequence:

$$5 \to 16 \to 8 \to 4 \to 2 \to 1.$$

Conjecture 1.1.1 (Collatz Conjecture)**.** No matter which number n you start with, your sequence will always go down to 1.

1.1.2 Arguments for the conjecture being true

1. The conjecture has been tested for the first 2^{68} numbers, and all of those numbers have satisfied the conjecture.

2. It could be the case that there is another loop of numbers in the collatz conjecture other than $1 \to 4 \to 2 \to 1$, which would disprove it, but any such loop would have to be at least 17,000,000 numbers long!

3. The sequence where you multiply odd numbers by 3 and add 1 seems to be special. For example, when we add 5 instead of 1 (sending odd n to $3n + 5$) we get all sorts of other loops of integers, for example $5 \to 20 \to 10 \to 5$, yet when you multiply the odd numbers by 3 and add 1 we still haven't found any counterexamples for the first 2^{68} digits.

4. In 2003, Krasikov and Lagarias proved that for any set of numbers $1, 2, 3, \ldots, x$, at least $x^{0.84}$ of them will satisfy the collatz conjecture.

5. **"The probabilistic argument"**. Every even number will obviously be decreased by the sequence (because it is divided by 2). However it seems that odd numbers increase, so possibly they could diverge off to infinity. However "on average" this is not true either- it will be multiplied by a factor of $\frac{3}{4}$ "on average". More precisely, an odd integer n will be sent to $3n+1$. This will always be even, so this goes to $\frac{3n+1}{2}$. Now this has a 50% chance of being even so 25% of the time the odd integer will be sent to $\frac{3n+1}{4}$ and so 12.5% of the time it will be sent to $\frac{3n+1}{8}$. Thus the expected growth of this odd integer n will end up being:

$$\left(\frac{3}{2}\right)^{\frac{1}{2}} \left(\frac{3}{4}\right)^{\frac{1}{4}} \left(\frac{3}{8}\right)^{\frac{1}{8}} \cdots = \frac{3}{4}.$$

Don't worry if that last one went a bit over your head, essentially all it was saying is that odd numbers get multiplied by $\frac{3}{4}$ on average so they should decrease to 1- I just wanted to add the precise detail for those who were interested.

1.1.3 Arguments Against

1. They still haven't proven it! It's nearly been 90 years since the problem was posed, and nobody has solved it yet so perhaps that is simply because it is not true; you can't prove something that is actually false.

2. Despite the fact that they have computed the first 2^{68} numbers, that is still not convincing- some conjectures have broken only at the 2^{1000}th number!

3. Even though it would be at least $17,000,000$ numbers long, the $\frac{3}{4}$ argument doesn't counter the possibility of there being a loop.

1.1.4 Further Reading

- In 2019, mathematician Terrence Tao made huge progress in the problem, which you can look into if interested.

1.2 Introduction to Dynamical Systems and the Mandelbrot Set

1.2.1 Introduction

In today's talk we talk about dynamical systems. Roughly speaking, this is the study of what happens when you iterate some rational function ϕ (meaning it is of the form $\phi(z) = \frac{F(z)}{G(z)}$ where F, G are polynomials) a bunch of times. For example, the Collatz conjecture is a question about a dynamical system, which should already show how interesting this field of maths can get very quickly. Today we look in more depth into some interesting properties, and some more unsolved questions in this mysterious, but beautiful field of maths.

1.2.2 A simple example

Let us begin by looking at the iterations of $\phi(z) = z^2$. We can ask if any numbers will go in a loop, and of course the answer is yes.

- The orbit of z is the set $\{\phi(z), \phi^2(z), \phi^3(z), \dots\}$

- For example 0 will go in the sequence $\{0, 0, 0, \dots\}$ and similarly 1 will continue off as $\{1, 1, 1, \dots\}$. These numbers are called fixed points, because they satisfy the equation $\phi(a) = a$.

- The number -1 is called preperiodic, because it will eventually end up in a loop which doesn't actually have the number -1 as follows: $\{-1, 1, 1, 1, \dots\}$.
- The complex number $\frac{-1+\sqrt{-3}}{2}$ is called periodic, because it goes in a loop as follows: $\{\frac{-1+\sqrt{-3}}{2}, \frac{-1-\sqrt{3}}{2}, \frac{-1+\sqrt{-3}}{2}, \dots\}$.
- Points such as 2 or $\frac{1}{3}$ are called wandering points, because they will never go in a loop; the number 2 diverges off to infinity and the number $\frac{1}{3}$ gets closer and closer to 0.
- Note that we can have "points at infinity". For example if we study the function $\phi(z) = \frac{z^2+1}{z^2-1}$, then if we allow the point ∞, 1 is periodic with the following orbit: $\{1, \infty, 1, \infty, \dots\}$.

1.2.3 Chaos and the Julia Set

If we look again at the function $\phi(z) = z^2$, we notice something interesting. For numbers α between 1 and 0, the orbits will go towards zero and for any number β that is close to α, their orbits will stay very similar too. Similarly, for numbers greater than 1, if α and β are close to each other then the orbits will also stay close to each other. However, when $\alpha = 1$, because it is a fixed point, no matter what β we pick, no matter how close it is to 1, will always have an orbit that goes further and further away from 1. This is an example of a chaotic point.

Definition 1.2.1. We can define the Fatou set $F(\phi)$ and Julia set $J(\phi)$ as follows:

$$J(\phi) = \{\alpha \text{ such that } \alpha \text{ is chaotic.}\}$$
$$F(\phi) = \{\alpha \text{ such that } \alpha \text{ is not chaotic}\} = J(\phi)'$$

1.2.4 A closer look at periodic points

Recall that a periodic point of a rational function ϕ is a number z such that $\phi^n(z) = z$ for some integer n. We will spend the rest of the talk examining rational periodic points specifically. It is known that a rational function will have infinitely many complex periodic points, and many times infinitely many real ones as well. However Northcott's theorem states the following:

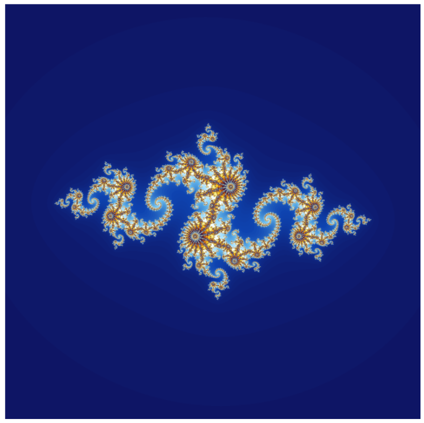


Figure 1.1: Julia set of $\phi(z) = z^2 - 0.8 + 0.156i$

Theorem 1.2.1 (Northcott, 1949)**.** *A rational function ϕ has only finitely many rational periodic points.*

Let us look at an example. $\phi_c(z) = z^2 + c$. When $c = 0$, this has a couple of rational numbers that have period 1 (aka fixed points). Those are $z = 0$ and $z = 1$. Now let's take $c = \frac{1}{4}$. Then $z = \frac{1}{2}$ is also a fixed point. What about larger loops? Well when $c = -1$, then $z = 0$ will be a point of period 2 as follows:

$$-1 \to 0 \to -1 \to 0 \to \dots$$

Can we go higher? Yes. Take $c = \frac{-29}{16}$ and $z = \frac{-7}{4}$. Then the loop becomes:

$$\frac{-7}{4} \to \frac{5}{4} \to \frac{-1}{4} \to \frac{-7}{4}.$$

At this point you might be thinking that we can always keep making loops that are longer and longer, but here comes a shock:

- Morton proved that ϕ_c can't have a rational periodic point of period 4.
- Flynn, Poonen and Schaefer proved that ϕ_c can't have a rational periodic point of period 5.
- Poonen has conjectured that there are no rational periodic points of φ_c that have a period greater than 5 either.

1.2.5 Mandelbrot Set

Now we will look at a famous set that has come out of the study of dynamic systems: the Mandelbrot set. We look again at the function $\phi_c(z) = z^2 + c$. The Mandelbrot set is defined as the set of all complex numbers c such that the function doesn't go off to infinity when we start at $z = 0$. For example, $c = 1$ is not in the Mandelbrot set because the sequence will be:

$$1 \to 1^1 + 1 = 2 \to 2^2 + 1 = 5 \to 26 \to \dots$$

But -2 is clearly not going to explode to infinity because the sequence is:

$$-2 \to 2 \to 2 \to 2 \to \dots$$

To finish the talk we played around with the Mandelbrot set on geogebra here and watched a cool animation of it, because it is simply so aesthetically pleasing.

1.2.6 Summary

In this talk we went through a tour in the basics of dynamical systems. We started by looking at the basic example of $\phi(z^2)$ to define some of the important terms in dynamical systems, and then went through an interesting conjecture about rational periodic points, before finally ending by looking at the Mandelbrot set, which is widely considered one of the most beautiful mathematical objects in the world.

1.3 Newton's Fractal

1.3.1 Introduction

This talk can be split roughly into three parts:

1. First, look at the Newton-Raphson method, which is a method for approximating solutions to polynomials.

2. Look at how we can make it into a fractal on the complex plane (known as the Newton fractal).

3. Third, look at the connection between the Newton fractal and the Mandelbrot set (a set which was introduced in the last talk) and generalise it (based on the paper entitled “The Mandelbrot set is universal”), which is a truly remarkable result.

1.3.2 The Newton-Raphson Method

First we look at the Newton-Raphson method. For solving quadratics, we have an easy method: simply use the quadratic formula. For cubics, it’s not so simple but there is a cubic formula one can use if one wishes to get an exact solution. For quartics, there is an absolute mess of a formula that will give you solutions. However, there is no quintic (or above) formula, so we can’t get the exact solutions in this way. Thus, we approximate it instead (and also we do the same for quartics and cubics, because their formulas are so clunky it’s better to just use this method instead). Since it is not the focus of the talk, I shall simply state the formula used to approximate the solutions without saying why it works. For a polynomial $p(z)$, we guess an answer z_0 and then we use the recursive formula to get a more and more accurate estimate:

$$z_{n+1} = z_n - \frac{p(z)}{p'(z)}.$$

Example 1.3.1. If we want to approximate the square root of 2, we can use the Newton-Raphson method on the equation $z^2 - 2$. If we guess $z_0 = 1.5$ then we have:

$$\begin{aligned} z_0 &= 1.5 \\ z_1 &= 1.5 - \frac{1.5^2 - 2}{2 \times 1.5} = 1.42 \\ z_2 &= z_1 - \frac{z_1^2 - 2}{2z_1} = 1.41 \\ z_3 &= \ldots \end{aligned}$$

here we see how quickly the Newton-Raphson approximated $\sqrt{2}$, which is a good sign that it is a good way to approximate roots. Now we will move to the complex plane and see how to make a fractal out of it.

1.3.3 The Newton Fractal

Now we will look at how this makes a fractal. Clearly, since there are multiple roots of most polynomial equations, different starting guesses will be attracted to different roots. Thus, we will colour the points in the plane that are attracted to different roots in different colours and this process will produce some beautiful fractals. It also highlights the sensitivity of the Newton-Raphson method in certain areas of the complex plane.

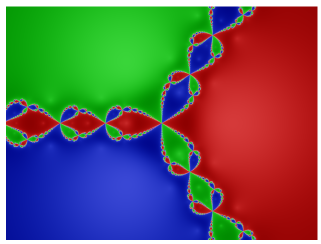


Figure 1.2: Newton's Fractal for $p(z) = z^3 - 1$

1.3.4 The Connection to the Mandelbrot Set

Now we look at a shocking connection to the Mandelbrot set that Newton's fractal reveals. We will now look at the cubic polynomial

$$p(z) = (z+1)(z - \frac{1}{2} - \lambda)(z - \frac{1}{2} + \lambda).$$

Now start off the Newton-Raphson method for this polynomial with $z_0 = 0$. If the Newton-Raphson method converges to a certain root, colour that λ a certain colour. But here's where it gets exciting: when you colour the points λ such that it doesn't converge to any root, we get a picture which looks **exactly like the Mandelbrot set!** I cannot exaggerate enough how incredible this is: a seemingly unrelated fractal given by the Newton-Raphson method gives us a fractal (already surprising), which also reveals the Mandelbrot set too! This is a prime example of the beauty of mathematics. This connection was further explored in a paper entitled "The Mandelbrot Set is Universal" by McMullen, if you are curious about how deep this goes.

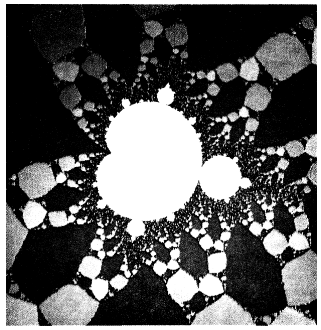


Figure 1.3: Zooming in on the points that fail to converge for $p(z) = (z + 1)(z - \lambda - \frac{1}{2})(z + \lambda - \frac{1}{2})$, taken from "On the Dynamics of Polynomial-Like Mappings"

1.3.5 Summary

In conclusion, we looked at the Newton-Raphson method, and how it surprisingly generates fractals which produces some of the best images in mathematics. Then, we proceeded to look at an extremely interesting connection between this fractal and the Mandelbrot set, which highlights the beauty of mathematics.

Chapter 2

Saga 2- Generalising Connections Between the Golden ratio and Fibonacci Numbers

2.1 Continued Fractions

2.1.1 Introduction

In this talk we looked at the idea of <u>continued fractions</u>. We begin with a simple example

$$\begin{aligned}
\frac{43}{19} &= 2 + \frac{5}{19} \\
&= 2 + \frac{1}{\frac{19}{5}} = 2 + \frac{1}{3 + \frac{4}{5}} \\
&= 2 + \frac{1}{\frac{19}{5}} = 2 + \frac{1}{3 + \frac{1}{\frac{5}{4}}} \\
&= 2 + \frac{1}{3 + \frac{1}{1 + \frac{1}{4}}}
\end{aligned}$$

Thus, we see that even a simple fraction admits an interesting continued fraction expansion. But the theory of continued fractions, as we shall see provides us with some deep connections to number theory...

2.1.2 Application to Quadratics

Now consider a quadratic equation, for example $x^2 - 5x - 1 = 0$. Normally, to solve this we would use the quadratic formula, but here we shall use the language of continued fractions:

$$\begin{aligned} x^2 &= 5x + 1 \\ x &= 5 + \frac{1}{x} \\ x &= 5 + \frac{1}{5 + \frac{1}{x}} \\ x &= 5 + \frac{1}{5 + \frac{1}{5 + \frac{1}{5 + \frac{1}{\cdots}}}} \end{aligned}$$

And so we've expressed the solution to this quadratic as an infinite continued fraction! But now let us dive deeper and look at a connection to the golden ratio. Recall that the golden ratio is the solution to the equation

$$\varphi^2 = \varphi + 1.$$

And so we shall rearrange it:

$$\begin{aligned} \varphi &= 1 + \frac{1}{\varphi} \\ \varphi &= 1 + \frac{1}{1 + \frac{1}{\varphi}} \\ \varphi &= 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{\cdots}}} \end{aligned}$$

Or in other notation, $\varphi = [1; 1, 1, 1, \ldots]$. It turns out that this continued fraction has a shocking connection to the Fibonacci numbers...

2.1.3 Two Shocking Connections

Definition 2.1.1. Remember that the Fibonacci numbers F_n are defined by the relation $F_n = F_{n-1} + F_{n-2}$ with $F_0 = 0$ and $F_1 = 1$. The first few numbers are $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$

We observe that the convergents of the golden ratio is precisely the ratio of the Fibonacci numbers. Let us the first few incidents:

$$1, 1+\frac{1}{1}, 1+\frac{1}{1+\frac{1}{1+1}}, 1+\frac{1}{1+\frac{1}{1+\frac{1}{1+1}}}, \ldots$$

which becomes:

$$\frac{1}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \ldots$$

As we can see, as we go through the convergents of the infinite continued fraction of φ, we keep getting ratios of two Fibonacci numbers. From this we can also infer the cool fact that as we take this process to infinity, we have:

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \varphi$$

We now look at a second shocking connection to number theory which we observed- a connection to Pell numbers. First, we have to find the continued fraction expansion for $\sqrt{2}+1$. This is the solution to $(x-1)^2 = 2$ and so we obtain:

$$\begin{aligned}
x^2 &= 2x+1 \\
x &= 2+\frac{1}{x} \\
x &= 2+\frac{1}{2+\frac{1}{x}} \\
x &= 2+\frac{1}{2+\frac{1}{2+\frac{1}{\cdots}}}
\end{aligned}$$

Now we look at how this connects to Pell numbers.

Definition 2.1.2. The Pell numbers are defined by $P_n = 2P_{n-1} + P_{n-2}$ with $P_0 = 0$, $P_1 = 1$. The first few Pell numbers are: $0, 1, 2, 5, 12, 29, 70, \ldots$.

The convergents of $x = \sqrt{2}+1$ are the following:

$$2, 2+\frac{1}{2}, 2+\frac{1}{2+\frac{1}{2}}, 2+\frac{1}{2+\frac{1}{2+\frac{1}{2}}}, \ldots$$

which become

$$\frac{2}{1}, \frac{5}{2}, \frac{12}{5}, \frac{29}{12}, \ldots$$

and again we see how the convergents of a certain special continued fraction are deeply related to a famous sequence in number theory. Fun fact: the number $1+\sqrt{2}$ is called the silver ratio. Next time we shall look in more depth into the Fibonacci numbers and the golden ratio.

2.2 On the Fibonacci Numbers and Golden Ratio

2.2.1 Introduction

In the last talk we unveiled a shocking connection between the continued fraction of the golden ratio and the Fibonacci numbers (check it out if you haven't seen the notes from that talk). This week we delve deeper into the connection between the golden ratio and the Fibonacci numbers, and uncover some deep results about the Fibonacci numbers.

2.2.2 Deriving the General Formula for F_n

We will now use an ingenious argument to find a general formula for the nth Fibonacci numbers. First recall that the golden ratio satisfies $\varphi^2 = \varphi + 1$ and so we have:

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2}.$$

We also have that the other solution to the quadratic $\psi := 1 - \varphi = \frac{-1}{\varphi}$ satisfies this too. Thus in general, any sequence that satisfies the recursion $U_n = U_{n-1} + U_{n-2}$ will be of the form:

$$U_n = a\varphi^n + b\psi^n$$

since

$$\begin{aligned} U_n &= a\varphi^n + b\psi^n \\ &= a(\varphi^{n-1} + \varphi^{n-2}) + b(\psi^{n-1} + \psi^{n-2}) \\ &= U_{n-1} + U_{n-2} \end{aligned}$$

as required. Now we apply this to F_n. Since $F_0 = 0$ and $F_1 = 1$, we have that:

$$\begin{aligned} a + b &= 0 \\ a\varphi + b\psi &= 1 \end{aligned}$$

And once the algebra is done, you have: $a = \frac{1}{\sqrt{5}}$, $b = \frac{-1}{\sqrt{5}}$ and so we have:

$$\boxed{F_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}} \tag{2.1}$$

(using $\psi = -\varphi^{-1}$).

2.2.3 Developing a Test For Fibonacci Numbers

Using the equation we derived in the previous section, we can develop a test to see if a random integer x is a Fibonacci number, meaning you don't have to go through all of the Fibonacci numbers to test it! First we must prove a preliminary lemma:

Lemma 2.2.1. $\varphi^n = \varphi F_n + F_{n-1}$

Proof. We do this by induction which means we prove it for the first case ($n = 1$) and then we assume that it is true for the nth case and show that this implies that the statement is true for the $(n+1)$th case. So let us proceed:
The base case: $\varphi^1 = \varphi(1) + 0$ which is clearly true.
The inductive step: Let us assume that $\varphi^n = \varphi F_n + F_{n-1}$. Then:

$$\begin{aligned} \varphi^{n+1} &= \varphi^2 F_n + \varphi F_{n-1} \\ &= (\varphi + 1)F_n + \varphi F_{n-1} \\ &= \varphi(F_n + F_{n-1}) + F_n \\ &= \varphi F_{n+1} + F_n \end{aligned}$$

as required. □

Now we can go about making this test. Using 2.1 we have that:

$$\begin{aligned}
&\sqrt{5}F_n = \varphi^n - (-\varphi)^{-n} \\
&\varphi^n\sqrt{5}F_n = \varphi^{2n} - (-1)^n \\
&\varphi^{2n} - \varphi^n\sqrt{5}F_n - (-1)^n = 0 \\
&\varphi^n = \frac{\sqrt{5}F_n + \sqrt{5F_n^2 + 4(-1)^n}}{2},
\end{aligned}$$

where the last step comes from the quadratic formula (if look closely, it's a quadratic of φ^n). Now using the lemma we derived before, we have:

$$\begin{aligned}
&\varphi F_n + F_{n-1} = \frac{\sqrt{5}F_n + \sqrt{5F_n^2 + 4(-1)^n}}{2} \\
&= 2\varphi F_n + 2F_{n-1} = \sqrt{5}F_n + \sqrt{5F_n^2 + 4(-1)^n} \\
&\overset{2\varphi=1+\sqrt{5}}{\Longrightarrow} (1+\sqrt{5})F_n + 2F_{n-1} = \sqrt{5}F_n + \sqrt{5F_n^2 + 4(-1)^n} \\
&\Longrightarrow F_n + 2F_{n-1} = \sqrt{5F_n^2 + 4(-1)^n} \\
&\Longrightarrow (F_n + 2F_{n-1})^2 = 5F_n^2 + 4(-1)^n.
\end{aligned}$$

And here we are at the crux of the argument! Since the right hand side is always a perfect square, we can say that x is a Fibonacci number iff $5x^2 - 4$ is a perfect square or $5x^2 + 4$ is a perfect square. Pretty nifty, eh!

2.3 Pell Numbers, The Silver Ratio and More Generalisations

2.3.1 Introduction

This week we finish the saga on all that we've been looking at in number theory for the last few weeks. And this will hopefully be an exciting end!

2.3.2 Pell Numbers and The Silver Ratio

The silver ratio is defined in a very similar way to the golden ratio. Whilst the golden ratio is defined as the solution to $x^2 - x - 1 = 0$, the silver ratio shall be defined as the positive solution to $x^2 - 2x - 1 = 0$. In other words, the silver ratio is $1+\sqrt{2}$. We can observe that the silver ratio enjoys a very similar continued fraction expansion to the golden ratio:

$$\begin{aligned}
&x^2 - 2x - 1 = 0\\
&x^2 = 2x + 1\\
&x = 2 + \frac{1}{x}\\
&\ = 2 + \frac{1}{2+\frac{1}{x}}\\
&\ = 2 + \frac{1}{2+\frac{1}{2+\frac{1}{2+\dots}}}
\end{aligned}$$

which looks exactly like the golden ratio, except for that one the 2s were replaced with 1s (see the notes from that legendary talk last week for a refresher). However, here the fun has just begun! Because recall that last time we observed a connection between the golden ratio and the Fibonacci numbers (which were defined with the recurrance relation $F_n = F_{n-1}+F_{n-2}$.) So this time, the silver ratio will have a shocking connection with the very similar <u>Pell numbers</u> which are defined by $P_n = 2P_{n-1} + P_{n-2}$ (with $P_0 = 0$ and $P_1 = 1$). So the first few Pell numbers would be: $0, 1, 2, 5, 12, 29, 70, \dots$. Again there is a shocking connection: the convergents of the silver ratio are the following:

$$2, 2+\frac{1}{2}, 2+\frac{1}{2+\frac{1}{2}}, 2+\frac{1}{2+\frac{1}{2+\frac{1}{2}}}, \dots$$

which become

$$\frac{2}{1}, \frac{5}{2}, \frac{12}{5}, \frac{29}{12}, \dots$$

Furthermore, just as we did with the Fibonacci numbers, we can now use the silver ratio to get a formula for the nth Pell number. Because we can simply multiply $x^2 = 2x+1$ by x^{n-2} on both sides, we have that $x^n = 2x^{n-1} + x^{n-2}$ and so since the solutions are $x = 1 \pm \sqrt{2}$, we know the formula for the nth Pell number will be of the form

$$P_n = a(1+\sqrt{2})^n + b(1-\sqrt{2})^n$$

since

$$\begin{aligned} P_n &= a(1+\sqrt{2})^n + b(1-\sqrt{2})^n \\ &= 2a(1+\sqrt{2})^{n-1} + 2b(1-\sqrt{2})^{n-2} + a(1+\sqrt{2})^{n-1} + b(1-\sqrt{2})^{n-2} \\ &= 2P_{n-1} + P_{n-2} \end{aligned}$$

as required. So now we must find a and b. We have set $P_0 = 0$ and $P_1 = 1$, so we just plug these two values in and get simultaneous equations and then we should be fine and dandy. More specifically, we get:

$$\begin{aligned} P_0 &= a(1+\sqrt{2})^0 + b(1-\sqrt{2})^0 = a + b = 0 \\ P_1 &= a(1+\sqrt{2}) + b(1-\sqrt{2}) = 1. \end{aligned}$$

After you do the algebra, you get that $a = \frac{1}{2\sqrt{2}}$ and $b = \frac{-1}{2\sqrt{2}}$. So we obtain a glorious formula for the nth Pell number being:

$$P_n = \frac{(1+\sqrt{2})^n - (1-\sqrt{2})^n}{2\sqrt{2}}.$$

So again we see that there is a connection between these special ratios and some very nice sequences in number theory. Thus, as any mathematician would seek to do, we generalised it and saw the full story!

2.3.3 The Maths Society Ratio

Yes, I know these are called the metallic means, but we generalised it ourselves and the results agreed with what mathematicians had inevitably already done so I called it the maths society ratio: sue me! Anyway, remember that the quadratic equations defining the golden and silver ratios ,respectively, were $x^2 - x - 1 = 0$ and $x^2 - 2x - 1 = 0$. So naturally we then looked at $x^2 - nx - 1 = 0$, and called this the maths society ratio (which we denoted δ because it was Mithush's favourite Greek letter). Anyway, let us observe that $\delta (= \frac{n+\sqrt{n^2+4}}{2})$ enjoys a very similar continued fraction expansion to the golden and silver ratios:

$$\begin{aligned}
x^2 - nx - 1 &= 0\\
x^2 &= nx + 1\\
x &= n + \frac{1}{x}\\
&= n + \frac{1}{n + \frac{1}{x}}\\
&= n + \frac{1}{n + \frac{1}{n + \frac{1}{n + \dots}}}
\end{aligned}$$

We now see how this links to the sequence $M_i = nM_{i-1} + M_{i-2}$, with $M_0 = 0$ and $M_1 = 1$. For example, when $n = 3$ the first few terms would be: $0, 1, 3, 10, 33, 109, \dots$. And the convergents of the continued fraction would be:

$$3, 3 + \frac{1}{3}, 3 + \frac{1}{3 + \frac{1}{3}}, 3 + \frac{1}{3 + \frac{1}{3 + \frac{1}{3}}}, \dots$$

which become

$$\frac{3}{1}, \frac{10}{3}, \frac{33}{10}, \frac{109}{33}, \dots$$

And in general the convergents of the maths society ratio for any n are

$$n, n + \frac{1}{n}, n + \frac{1}{n + \frac{1}{n}}, n + \frac{1}{n + \frac{1}{n + \frac{1}{n}}}, \dots$$

which become

$$\frac{n}{1}, \frac{n^2 + 1}{n}, \frac{n(n^2 + 1) + n}{n^2 + 1}, \frac{n(n(n^2 + 1) + n) + n^2 + 1}{n(n^2 + 1) + n}, \dots$$

which is precisely just the first few terms of our sequence in a fraction. The last thing we did was find a general formula for M_i. Using the same reasoning again it will be of the form $M_i = a\delta^i + b\bar{\delta}^i$, where $\bar{\delta}$ is $\frac{n - \sqrt{n^2+4}}{2}$. And again, doing the algebra on the simultaneous equations

$$\begin{aligned}
a + b &= 0\\
a\delta + b\bar{\delta} &= 1
\end{aligned}$$

yields $a = \frac{1}{\sqrt{n^2+4}}$ and $b = \frac{-1}{\sqrt{n^2+4}}$ and so our general formula is:

$$\boxed{M_i = \frac{\delta^i - \bar{\delta}^i}{\sqrt{n^2+4}}}$$

(fun exercise: check that this agrees with the formulae for the nth Fibonacci and nth Pell number.)

2.3.4 Further Reading

Thank you to Saik for pointing out this link which had some more amazing results on the metallic means- we barley scratched the surface in this saga!

Chapter 3

Saga 3- Transcendental Numbers

3.1 Transcendental Numbers Exist- A Look At Liouville Numbers

3.1.1 Introduction

Last time it was revealed that our next saga would be a look into the world of transcendental numbers, which are numbers that are not a root of any equation $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ where $a_i \in \mathbb{Z}$. In this talk we establish the fact that these numbers exist by looking at the first example that was discovered in 1844- the Liouville constant, which is an example of a class of transcendental numbers known as Liouville numbers

3.1.2 Liouville Numbers

The goal of this talk is to prove that the number $L = \sum_{n=1}^{\infty} \frac{1}{10^{n!}}$ is irrational. We do this in the following way:

1. Prove that all Liouville numbers are transcendental
2. Show that L is a Liouville number.

Sadly, the definition of a Liouville number is pretty technical.

Definition 3.1.1. A Liouville number α is a number such that for all $n \in \mathbb{N}$, there exists a rational number $\frac{a}{b}$ (with $b > 1$) such that:

$$0 < \left|\alpha - \frac{a}{b}\right| < \frac{1}{b^n}.$$

Proposition 3.1.1. If α is a Liouville number, then α is irrational.

Proof. Let us assume that α is a rational Liouville number $\frac{p}{q}$. Then for some rational number $\frac{a}{b} \neq \frac{p}{q}$, we have:

$$0 < \left|\alpha - \frac{a}{b}\right| = \left|\frac{p}{q} - \frac{a}{b}\right| = \left|\frac{pb - aq}{qb}\right|$$

Now we pick a natural number n such that $2^{n-1} > q$. Then we have:

$$\left|\frac{pb - aq}{qb}\right| > \frac{1}{2^{n-1}b} \geq \frac{1}{b^n}$$

where the last inequality sign comes from the fact that $b > 1$. Thus we have shown that for any $\frac{a}{b}$ we try to choose, there will be an $n \in \mathbb{N}$ such that $\left|\alpha - \frac{a}{b}\right| > \frac{1}{b^n}$ which contradicts the assumption that α was a Liouville number. Thus, all Liouville numbers are irrational. $\square$

Now that we have established that all Liouville numbers are irrational, we can now move forward and try to prove that they are also transcendental too.

Theorem 3.1.2. *Liouville numbers are transcendental.*

Proof. Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \dots a_0$ and assume by contradiction that $f(\alpha) = 0$. We now define a few things: $M = \max_{[\alpha-1,\alpha+1]} |f'(x)|$, $A < \{1, \frac{1}{M}, |\alpha - \alpha_1|, \dots |\alpha - \alpha_m|\}$ where $f(\alpha_i) = 0, \forall i \leq m$. Then we pick some $r \in \mathbb{N}$ such that $2^r \geq \frac{1}{A}$. Since α is a Liouville number, we have some $\frac{a}{b} \in \mathbb{Q}$ such that:

$$0 < \left|\alpha - \frac{a}{b}\right| < \frac{1}{b^{n+r}} \leq \frac{1}{2^r b^n} \leq \frac{A}{b^n} < A. \tag{3.1}$$

Now, because $\left|\alpha - \frac{a}{b}\right| < A$, we have that:

1. $\frac{a}{b} \in [\alpha - 1, \alpha + 1]$

2. $f\left(\frac{a}{b}\right) \neq 0$.

Thus we can use the mean value theorem to say that there exists some $x_0 \in \left(\alpha, \frac{a}{b}\right)$ such that

$$f'(x_0) = \frac{f(\alpha) - f\left(\frac{a}{b}\right)}{\alpha - \frac{a}{b}} = \frac{-f\left(\frac{a}{b}\right)}{\alpha - \frac{a}{b}}$$

So we can say that:

$$|f'(x_0)| = \frac{\left|f\left(\frac{a}{b}\right)\right|}{\left|\alpha - \frac{a}{b}\right|} \implies \left|\alpha - \frac{a}{b}\right| = \frac{\left|f\left(\frac{a}{b}\right)\right|}{|f'(x_0)|} \geq \frac{\left|f\left(\frac{a}{b}\right)\right|}{M} \tag{3.2}$$

Now note that we have:

$$\left|f\left(\frac{a}{b}\right)\right| = \left|a_0 + \frac{a_1 a}{b} + \cdots + \frac{a_n a^n}{b^n}\right| = \frac{1}{b^n}|b^n a_0 + a_1 a b^{n-1} + \cdots + a_n a^n| \geq \frac{1}{b^n}.$$

So finally we are done because now we plug this into 3.2 to obtain:

$$\frac{\left|f\left(\frac{a}{b}\right)\right|}{M} \geq \frac{1}{b^n M} > \frac{A}{b^n} > \left|\alpha - \frac{a}{b}\right|. \tag{3.3}$$

But 3.1 and 3.3 imply that $\left|\alpha - \frac{a}{b}\right| > \left|\alpha - \frac{a}{b}\right|$ which is a clear contradiction so there could not have existed some f such that $f(\alpha) = 0$, so α is transcendental as required! □

3.1.3 Liouville Constant

We now show that $L = \sum_{n=1}^{\infty} \frac{1}{10^{n!}}$ is a Liouville number.

Proposition 3.1.3. *L is a Liouville number.*

Proof. Let us write L as $\sum_{n=1}^{m} \frac{1}{10^{n!}} + \sum_{n=m+1}^{\infty} \frac{1}{10^{n!}}$. Then this first part can be collapsed into a single fraction of the form $\frac{a}{10^{n!}}$. We shall pick our b in the fraction $\frac{a}{b}$ to be $10^{n!}$. Then:

$$0 < \left|L - \frac{a}{b}\right| = \left|\sum_{n=m+1}^{\infty} \frac{1}{10^{n!}}\right| < \frac{1}{b^n}\left(\sum_{n=1}^{\infty} \frac{1}{2^n}\right) = \frac{1}{b^n}$$

where the last inequality comes from simply comparing the denominators of the fractions on the sums and observing that they are bigger in the sum on the left and so we have shown that L satisfies the conditions being a Liouville number and so L is transcendental! □

3.2 e and π are transcendental

I was lacking that week, so we just watched a YouTube video. The notes are in this link: here

Chapter 4

Saga 4: Set Theory and Logic

4.1 Russel's Paradox and ZFC Set Theory

4.1.1 Introduction

Today we shall begin our journey through set theory and logic by discussing one of the most famous paradoxes around- Russel's paradox. This paradox sparked a whole branch of discussion about set theory and logic, which then prompted mathematicians Ernst Zermelo and Abraham Fraenkel to come up with a set of axioms to resolve this, which seems to fix Russel's paradox and has worked very well so far. We conclude by looking at the Von Neumann universe, which describes what all of the sets in ZFC set theory should look like.

4.1.2 The Paradox

We shall jump straight into the paradox itself. Firstly, we shall talk about the barber of Seville. The barber of Seville shaves everyone who doesn't shave themselves. Then does he shave himself? Well, if he doesn't then he does, but if he does then he doesn't- a paradox! Whilst this is an interesting paradox, you may be wondering what relation this has to set theory. Well naively, one may say that a set is any definable collection of elements. Well, in that case we can define the set of all sets that don't contain themselves:

$$S = \{X | X \notin X\}$$

Well then the question is: is $S \in S$. Well if that were the case, then it contains itself, so its not in S, but if it doesn't contain itself, then it is in S, which is a paradox!

4.1.3 A look towards ZFC set theory

Since naive set theory was shown to lead to contradictions, mathematicians had to try to find a satisfactory way to rebuild the foundations of mathematics, or else much of mathematics would be under threat. This lead to a set of axioms known as the Zermelo–Fraenkel axioms (ZFC set theory). We shall roughly outline them here, but they will be a topic for further discussion in the following talks.

1. **The axiom of extensionality**
   This states that two sets are equal if they have the same elements. For example, the set of all numbers of the form $2n + 2$ and the set of all numbers of the form $2(n + 1)$ are the same set.

2. **The axiom of regularity**
   This states that every non-empty set A contains a member B such that $A \cap B = \emptyset$. Note that this means that we cannot have a set that contains itself, which eliminates Russel's paradox.

3. **The axiom of seperation**
   Any "subset" of a set is itself a set (we need to define what precisely we mean by subset).

4. **The axiom of pairing**
   Given two sets a, b we can form a new set which contains a and b as elements.

5. **The axiom of union**
   Given some sets $a, b, c, \dots$ we can form their union $a \cup b \cup c \cup \dots$.

6. **The axiom of infinity**
   There is a set X which has infinitely many elements.

7. **The axiom of powerset**
   Given a set a, we can form a set $\mathcal{P}(a)$ which is the set of all subsets of a.

8. **The axiom of replacement**
   The image of a function is itself a set.

9. **The axiom of choice**
   This axiom says that if we're given a bunch of sets, we can take one element of each set and together those elements will form a new set. Whilst this seems controversial, this was the most controversial axiom, but it is now accepted by most mathematicians. ZF set theory refers to just axioms 1 to 8, whilst ZFC set theory includes the axiom of choice (C stands for choice).

4.1.4 Various Interpretations

We now move onto the slightly more philosophical question of how to interpret these axioms. There are many ways of doing this:

1. **Platonic**
   The axioms are describing the collection of all sets, and the properties they have. The problem with this interpretation is that we can't really justify why if we can form this collection of all sets, why shouldn't it be a set itself?

2. **Formalism**
   There's no point in discussing the <u>meaning</u> of the axioms, we should just investigate their consequences and see what happens. The issue with this is if you ignore the meaning of the axioms completely, then why bother studying them in the first place? How do you know that they are interesting at all?

3. **Pragmatism**
   Assume that the axioms are true and just get on with proving theorems without worrying about this stuff in the first place.

4. **Finitism**
   The axioms are nonsense anyway because infinite sets don't actually exist so there's no point studying these axioms because they're not actually describing anything anyway. The problem with this is the question of: why do these axioms work so well? If we simply reject these axioms, then we would lose most of mathematics!

4.1.5 The Von Neumann Universe

The Von Neumann universe is a way of describing what all of the sets in ZFC set theory look like. The idea behind the Von Neumann universe is that we shall build it up from the starting point $V_0 = \emptyset$. Let us proceed by taking the powerset (set of all subsets) at each stage:

$$\begin{aligned} &V_0 = \{\} = \emptyset \\ &V_1 = \{\emptyset\} \\ &V_2 = \{\emptyset, \{\emptyset\}\} \\ &V_3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\{\emptyset\}\}\} \\ &V_4 = \dots \end{aligned}$$

One can see that this is already getting complicated, but we've just begun-we now form:

$$V_\omega = V_0 \cup V_1 \cup V_2 \cup \dots$$

But we can still keep going: we take

$$\begin{aligned} &V_{\omega+1} = \mathcal{P}(V_\omega) \\ &V_{\omega+2} = \mathcal{P}(V_{\omega+1}) \\ &\vdots \end{aligned}$$

but we're still not done! We now can form

$$V_{2\omega} = V_{\omega+\omega} = V_\omega \cup V_{\omega+1} \cup V_{\omega+2} \cup \dots$$

and then ,naturally, we can keep going for $V_{2\omega+1}$ and so on. These subscripts are known as ordinal numbers. Then we form the Von Neumann universe:

$$V = \bigcup_\alpha V_\alpha$$

where α ranges through all the ordinal numbers. Note that this isn't a set itself, since if there were a set of all sets we would run into a Russel's paradox type of situation. The interpretation of this is that all of the members of a set are themselves sets which are built up from smaller sets. We shall see next time how this works for the natural numbers.

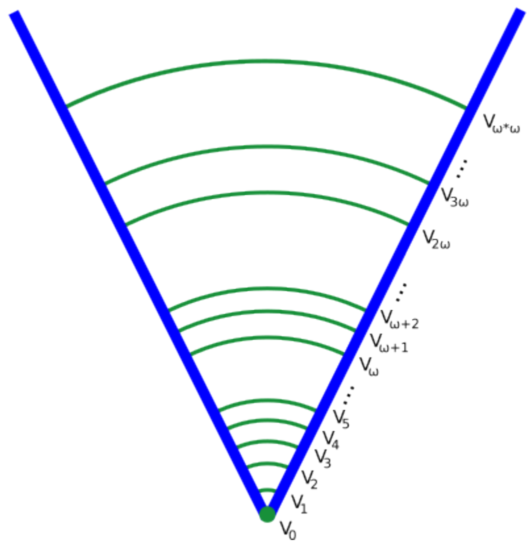


Figure 4.1: A visualisation for the Von Neumann universe, also giving the reason for why the letter "V" is used.

4.2 A Closer Look at the Natural Numbers

4.2.1 Introduction

Last time we looked at the axioms of ZFC set theory and we finished with the Von Neumann universe. Today we look closer at the natural numbers, and the set of rules that we use as axioms for the natural numbers called the "Peano axioms". We finish by seeing the set theoretic definition of the natural numbers, which should hopefully shed some insight into how the Von Neumann universe works, with the idea of building up sets from smaller sets.

4.2.2 The Peano Axioms

The first axioms are to do with the relation "$x = y$". The first Peano axiom (in some literature) is that this is an equivalence relation. This means that:

1. For all $x \in \mathbb{N}$, we have $x = x$. This is called reflexivity.
2. For all $x, y \in \mathbb{N}$, we have $x = y \iff y = x$. This is called symmetry.

3. For all $x, y, z \in \mathbb{N}$. we have $x = y$ and $y = z$ implies $x = z$. This is called transitivity.

With this out of the way, we can move onto the main Peano axioms. The idea is to start with the first natural number 0 and to use a function $\text{succ}(x)$, called the successor function, in order to construct the rest of them. The first two axioms are:

1. $0 \in \mathbb{N}$
2. If $x \in \mathbb{N}$, then $\text{succ}(x) \in \mathbb{N}$.

Now intuitively, we know that $\text{succ}(x) = x+1$, so one might think that we're done. However, we don't even know what "+" is, so as of now we could have $S(0) = 0$ and the set of natural numbers would be $\{0\}$ and this would satisfy the axioms right now. Thus we have to add:

- For all $x \in \mathbb{N}$, $\text{succ}(x) \neq 0$.

Now we run into the issue that we could now define $\mathbb{N} = \{0, 1\}$ where $\text{succ}(0) = 1 = \text{succ}(1)$ and all the axioms are satisfied. Thus we also add:

- For all $x, y \in \mathbb{N}$, if $S(x) = S(y)$, then $x = y$.

Now we're very close, since we have that $\text{succ}(0) = 1$, but then $\text{succ}(1)$ must be some other number, which we call 2 and then $\text{succ}(2)$ must be a different number called 3 so on forever. However, we could still define:

$$\mathbb{N} = \{0, 1, 2, \dots\} \cup \{a, b\}$$

with $S(a) = b$ and $S(b) = a$ and all the axioms are satisfied. Thus we have the last axiom, called the axiom of induction:

- If there is a set S that satisfies the previous axioms, then it contains all of the natural numbers.

So the final list of Peano axioms are the following, given a function succ:

1. $0 \in \mathbb{N}$
2. If $x \in \mathbb{N}$, then $\text{succ}(x) \in \mathbb{N}$
3. For all $x \in \mathbb{N}$, $\text{succ}(x) \neq 0$
4. For all $x, y \in \mathbb{N}$, if $\text{succ}(x) = \text{succ}(y)$, then $x = y$
5. If there is a set S that satisfies the previous axioms, then it contains all of the natural numbers.

4.2.3 Defining Addition and Multiplication

Now, this is all well and good, but we still want to define addition and multiplication in the natural numbers. We define addition as follows:

$$\begin{aligned} &\text{For all } a \in \mathbb{N}, a + 0 = 0 \\ &a + \mathrm{succ}(b) = \mathrm{succ}(a + b) \end{aligned}$$

For example, we have

$$1 + 1 = 1 + \mathrm{succ}(0) = \mathrm{succ}(1 + 0) = \mathrm{succ}(1) = 2.$$

Then we can use this to do

$$1 + 2 = 1 + \mathrm{succ}(1) = \mathrm{succ}(1 + 1) = \mathrm{succ}(2) = 3,$$

and so addition is defined recursively as such.
We also define multiplication recursively by saying:

$$\begin{aligned} &a \cdot 0 = 0 \\ &a \cdot \mathrm{succ}(b) = a + a \cdot b \end{aligned}$$

Then we can show, for example, that $a \cdot 1 = a$, because:

$$a \cdot 1 = a \cdot \mathrm{succ}(0) = a + a \cdot 0 = a.$$

4.2.4 Set Theoretic Definition of Natural Numbers

Again, we define the natural numbers recursively. This time, we define: $0 = \emptyset$ and $\mathrm{succ}(n) = n \cup \{n\}$. Then we have:

$$\begin{aligned} &0 = \emptyset \\ &1 = 0 \cup \{0\} = \emptyset \cup \{\emptyset\} = \{\emptyset\} = \{0\} \\ &2 = 1 \cup \{1\} = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\} = \{0, 1\} \\ &3 = 2 \cup \{2\} = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\} \\ &\vdots \end{aligned}$$

We can see that this interpretation satisfies the Peano axioms. This is very nice, a bit too nice in fact, so I shall throw a spanner in the works next time!

4.3 Gödel's Incompleteness Theorems

4.3.1 Introduction

Now that we have constructed some axioms that seem to work very well in set theory, mathematics seems to be back on steady grounding. In today's talk, we examine a very unnerving set of theorems known as Gödel's incompleteness theorems, which tell us some scary things about mathematics.

4.3.2 What are Gödel's Incompleteness Theorems?

At the start of the 20th century, David Hilbert proposed a set of 23 problems which he thought would be very important in 20th century mathematics. The second one of these problems was to provide a finite list of axioms to start out with and to prove that these axioms don't lead to any contradictions (such as Russel's paradox). Ideally, one would want to be able to prove every true statement in mathematics, building up from these rules.

Theorem 4.3.1 (Gödel's first incompleteness theorem)**.** *No consistent system of axioms is capable of proving all true statements about the arithmetic of natural numbers. That is, no matter what system of rules you work with, there will always be statements about numbers that we cannot prove with those rules.*

This is disappointing! It would be nice to say that all true statements about natural numbers can flow from a certain set of rules, but it turns out that this is not the case. Perhaps the reason that the Collatz conjecture (first ever talk at maths society) or the Riemann hypothesis (Wren's talk) are so difficult is because they literally cannot be proven with the system of rules that we have in place. In fact, weirdly enough, if one showed that it was impossible to prove these statements with our rules, that itself would be a proof because if the statements were false, then our rules would certainly allow us to find a counterexample (just let a computer run for long enough). You may be thinking: wouldn't that be a proof, so surely it can be proven? Well remember that this hypothetical proof strategy would use rules outside of our system.

Theorem 4.3.2 (Gödel's second incompleteness theorem)**.** *No system can prove it's own consistency.*

This means that, given any system of rules that we want to start out with, we cannot prove that no contradictions arise within this system, using the system itself. For example, the Peano axioms for the natural numbers can only be proven to be consistent using ZFC set theory. And ZFC set theory cannot prove that ZFC set theory is consistent; one must assume another rule in order to do this!

Chapter 5

Saga 5: Functions in Analytic Number Theory (Featuring Prasan)

5.1 The Basel Problem

5.1.1 Introduction

The Basel problem is a famous problem in mathematics, originally solved by Euler in 1734. The problem asks what the following sum converges to:

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$$

In this talk we looked at a very clever solution to the problem, but there are many alternative ways to do it, if you are interested.

5.1.2 The Solution

We start by considering the function $\sin(x)$. This function has roots at 0, π, $-\pi$, 2π, -2π, etc. Thus, by the Weierstrass factorisation theorem, we can write:

$$\begin{aligned}\sin(x) &= x\left(1 - \frac{x}{\pi}\right)\left(1 + \frac{x}{\pi}\right)\left(1 - \frac{x}{2\pi}\right)\left(1 + \frac{x}{2\pi}\right)\dots \\ &= x\left(1 - \left(\frac{x}{\pi}\right)^2\right)\left(1 - \left(\frac{x}{2\pi}\right)^2\right)\left(1 - \left(\frac{x}{3\pi}\right)^2\right)\dots\end{aligned}$$

On the other hand, we can look at the Taylor series of $\sin(x)$ (thanks to Prasan's excellent talk on the subject!) which is:

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

Now if we divide both of these by x, we have:

$$\frac{\sin(x)}{x} = \left(1 - \left(\frac{x}{\pi}\right)^2\right)\left(1 - \left(\frac{x}{2\pi}\right)^2\right)\left(1 - \left(\frac{x}{3\pi}\right)^2\right)\dots \tag{5.1}$$

$$\frac{\sin(x)}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \dots \tag{5.2}$$

Now it doesn't look like it, but we're actually now in business! If we expand the top equation, we have:

$$\frac{\sin(x)}{x} = 1 - x^2\left(\frac{1}{\pi^2} + \frac{1}{(2\pi)^2} + \frac{1}{(3\pi)^2} + \dots\right) + O(x^4)$$

where the big O notation roughly means: "junk that we don't care about". Now, all that's left to do is compare the coefficients of x^2 in both series and we obtain:

$$\frac{-1}{3!} = -\left(\frac{1}{\pi^2} + \frac{1}{(2\pi)^2} + \frac{1}{(3\pi)^2} + \dots\right)$$

and so multiplying both sides by $-\pi^2$, we have the result in all of it's glory:

$$\boxed{\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots = \frac{\pi^2}{6}}$$

5.2 The Euler Product for the Zeta Function: Prasan Patel

5.2.1 Introduction

In this talk we looked at a stunning connection between the Zeta function and the prime numbers, known as the Euler product, by considering a special probability distribution.

5.2.2 Setting up the Product

We shall begin by considering a probability distribution given by:

$$P(X = n) = \frac{1}{n^s \zeta(s)}$$

for some $s > 1$. First we have to check that all of the probabilies add up to 1:

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X = n) &= \sum_{n=1}^{\infty} \frac{1}{n^s \zeta(s)} \\
&= \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{n^s} \\
&= \frac{\zeta(s)}{\zeta(s)} = 1
\end{aligned}$$

as required. Now, we look at the probability of picking some number divisible by k. More precisely, this is:

$$\begin{aligned}
&P(X = k) + P(X = 2k) + P(X = 3k) + \dots \\
&= \frac{1}{k^s \zeta(s)} + \frac{1}{(2k)^s \zeta(s)} + \frac{1}{(3k)^s \zeta(s)} + \dots \\
&= \sum_{n=1}^{\infty} \frac{1}{\zeta(s)(nk)^s} \\
&= \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{(nk)^s} \\
&= \frac{1}{k^s \zeta(s)} \sum_{n=1}^{\infty} \frac{1}{n^s} \\
&= \frac{\zeta(s)}{k^s \zeta(s)} \\
&= \frac{1}{k^s}.
\end{aligned}$$

5.2.3 The Euler Product

Now, the probability of being divisible by a prime p is $\frac{1}{p^s}$, so the probability of not being divisible by p (denoted later as p') is $1 - \frac{1}{p^s}$. Now, since being divisible by any two numbers are independent events, we have (denoting the probability of being divisible by n as $P(n)$): $P(n \cap m) = P(n)P(m)$. Therefore the probability of not being divisible by any prime is given by:

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

where the big Π just denotes: "product of".
But what is the only number that is not divisible by any prime? 1. So therefore we have:

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

and thus we have the Euler product:

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

5.3 The Gamma Function: $\left(\frac{1}{2}\right)!$

5.3.1 Introduction

In this talk we ponder a question which at first seems bizarre: what is $\left(\frac{1}{2}\right)!$ Recall that for integers, we define the factorial as follows:

$$n! = n(n-1)(n-2)\dots 2 \cdot 1$$

So for example we have $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. It is unclear, however, how to extend this for non integers like $\frac{1}{2}$. The answer lies within the Gamma function.

5.3.2 The Gamma Function

Notice we can define the factorial function inductively for natural numbers: once we have $0! = 1$, we can use the relation $n! = n(n-1)!$ to do the rest for us. So therefore if we can construct a function such that $f(1) = 1$ and $f(n+1) = nf(n)$, then $f(n) = (n-1)!$ for all natural numbers n. We now introduce the Gamma function, which obeys these properties:

Theorem 5.3.1. *The Gamma function*

$$\Gamma(n) = \int_0^\infty t^{n-1}e^{-t}\mathrm{d}t$$

satisfies:

1. $\Gamma(1) = 1$

2. $\Gamma(n+1) = n\Gamma(n)$

Proof. For the first property, we just plug in $x = 1$ and are left with an elementary integral:

$$\begin{aligned}
&\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}\mathrm{d}t = \int_0^\infty e^{-t}\mathrm{d}t \\
&= \left[-e^{-t}\right]_0^\infty \\
&= -e^{-\infty} - -e^0 = 1
\end{aligned}$$

For the second property we can use integration by parts to obtain:

$$\begin{aligned}
&\Gamma(n+1) = \int_0^\infty t^n e^{-t}\mathrm{d}t \\
&= [-t^n e^{-t}]_0^\infty - \int_0^\infty nt^{n-1}(-e^{-t})\mathrm{d}t \\
&= 0 + n\int_0^\infty t^{n-1}e^{-t}\mathrm{d}t \\
&= n\Gamma(n)
\end{aligned}$$

□

Therefore, we can say that the Gamma function is an extension of $n!$ and so now it makes sense in a way to compute $\left(\frac{1}{2}\right)!$- it's just $\Gamma\left(\frac{3}{2}\right)$.

5.3.3 The Integral

Let's not dilly dally and get straight down to the computation. We wish to compute:

$$\int_0^\infty t^{\frac{1}{2}}e^{-t}\mathrm{d}t.$$

Firstly, by integration by parts we have:

$$\int_0^\infty t^{\frac{1}{2}}e^{-t}\mathrm{d}t = \left[-t^{\frac{1}{2}}e^{-t}\right]_0^\infty + \frac{1}{2}\int_0^\infty t^{-\frac{1}{2}}e^{-t}\mathrm{d}t$$
$$= \frac{1}{2}\int_0^\infty t^{-\frac{1}{2}}e^{-t}\mathrm{d}t$$

We now employ the u substitution $u = t^{\frac{-1}{2}}$ to transform this into the following integral:

$$\int_0^\infty e^{-u^2}\mathrm{d}u.$$

This is exactly half of a very famous integral known as the Gaussian integral, and we will get stuck into a satisfying method for computing it. We say that $I = \int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x$ and so we have:

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x \int_{-\infty}^{\infty} e^{-y^2}\mathrm{d}y$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y.$$

We now have to transform into polar co-ordinates, which is a method of expressing any complex number in the form

$$z = r\cos\theta + ir\sin\theta$$

where r ranges from 0 to ∞ and θ ranges from 0 to 2π. We can see from the formula that the x-coordinate (real part) is $r\cos\theta$ and similarly the y-coordinate (imaginary part) is $r\sin\theta$ and so we employ the substitutions $x = r\cos\theta$ and $y = r\sin\theta$. Now, because we are now working in 2d, doing this substitution is more complicated (for 1d you just see what dx would be in terms of du and plug that in), but since we're in 2d here we have to take all the partial derivatives and put them into a Jacobian matrix and take the determinant of that and in the end we obtain r so overall we get:

$$I^2 = \int_0^{2\pi}\int_0^\infty re^{-((r\cos\theta)^2+(r\sin\theta)^2)}\mathrm{d}r\mathrm{d}\theta$$
$$= \int_0^{2\pi}\int_0^\infty re^{-r^2}\mathrm{d}r\mathrm{d}\theta$$

Now the integral $\int_0^\infty re^{-r^2}\mathrm{d}r$ can be done with elementary calculus: notice that the derivative of r^2 is $2r$, so if you do the reverse of the chain rule in your head (otherwise just do u sub with $u = r^2$), you can obtain that:

$$\int_0^\infty re^{-r^2}\mathrm{d}r = \left[\frac{-1}{2}e^{-r^2}\right]_0^\infty = \frac{1}{2}.$$

So overall our integral becomes:

$$I^2 = \int_0^{2\pi} \frac{1}{2}\mathrm{d}\theta = \pi$$

And so the Gaussian integral $I = \sqrt{\pi}$. And since $\left(\frac{1}{2}\right)!$ was half of that we have:

$$\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}.$$

Neat, eh?

Part II

Year 2

Chapter 6

Saga 6: An Introduction to Modular Arithmetic

6.1 Motivation for Modular Arithmetic

6.1.1 Introduction

In this talk, we shall look at some examples of why modular arithmetic is a powerful tool in number theory, in order to motivate why we will study it for the next few weeks.

6.1.2 What is Modular Arithmetic?

Modular arithmetic is where we do arithmetic, but we consider remainders. More precisely, we say that $a \equiv b \mod n$ iff a and b have the same remainder when you divide them by n. For example, we have $1 \equiv 5 \mod 4$. It turns out that we use modular arithmetic without knowing it in our lives: if I were to ask you what day it will be in 100 days, you would probably not actually count up 100 days. Instead, you would probably observe that since $100 = 14 \times 7 + 2$, 100 days is 14 weeks and 2 days away and so it will be a Wednesday+2=Friday. In this example we have actually used the fact that $100 \equiv 2 \mod 7$ to simplify the problem. Similarly, if I asked you what the time will be in 27 hours, then you would be able to figure that out using the fact that $27 \equiv 3 \mod 24$. We shall now see some examples of the power of modular arithmetic.

6.1.3 Examples

Example 6.1.1. Consider the following question:

Is the number 224678946 a square number?

At first this problem seems unsolvable without a calculator, or lots of tedious calculations, but it turns out that we can show that this is not a square number with ease. We shall use the fact that the remainder of any square number when you divide it by 4 is either 0 or 1. Then, we can figure out what $224678946 \mod 4$ is using the fact that $224678946 = 224678900 + 46 = 2246789 \times 25 \times 4 + 46$ and since $46 \equiv 2 \mod 4$, this means that our number is indeed not a square! Note that if the number did have a remainder of 1 or 0 when divided by 4, then it wouldn't necessarily be a square.

Example 6.1.2. Now let us consider a different question:

What is the last digit of 2^{1000}?

Well if look at the first few powers of two we see the following:

$$2, 4, 8, 16, 32, 64, 128, 256, \ldots$$

One can see that the pattern repeats 4-fold and that for us since 1000 is a multiple of 4, the last digit of 2^{1000} is going to also end in a 6. Let us now look at the last digit of 3^{1000}. If we once again look at the pattern it gives us:

$$3, 9, 27, 81, 243, \ldots$$

and again the pattern repeats 4-fold except this time the last digit of 3^{1000} will be 1. It seems that in both cases however, the last repeats every fourth term. This is a consequence of a powerful theorem in modular arithmetic known as <u>Fermat's little theorem</u>. Fermat's little theorem states that for some positive integer a and prime p such that $\gcd(a, p) = 1$, we have that:

$$a^{p-1} \equiv 1 \mod p.$$

If we apply this for the case of $p = 5$, then we have for all $a \neq 5n$:

$$a^4 \equiv 1 \mod 5.$$

This means that a^4 will always end in either 1 or 6 which just give you the same digit when you multiply them by each other, so every fourth power will end in 1 or 6. We now package this information more nicely together: If a is an even number which isn't a multiple of 10, then a^{1000} will end in a 6 and if a is an odd number which isn't a multiple of 5, then a^{1000} will end in a 1.

6.2 Some Basic Properties in Modular Arithmetic

6.2.1 Introduction

Last time we looked at the basic notion of what modular arithmetic is and what the statement "$a \equiv b \mod n$" actually means and why we care about it. In this talk we delve further into the details of modular arithmetic and bang out some essential properties that modular arithmetic has.

6.2.2 Important Properties

In this talk we shall see some very important properties of addition and multiplication in modular arithmetic, namely the following result:

Lemma 6.2.1. Given that $a \equiv a' \mod n$ and $b \equiv b' \mod n$, we have:

1. $a + b \equiv a' + b' \mod n$
2. $ab \equiv a'b' \mod n$

Proof. Since $a \equiv a' \mod n$ we know that there exists some $k \in \mathbb{Z}$ such that:

$$a = a' + kn$$

Similarly, there exists some $m \in \mathbb{Z}$ such that

$$b = b' + mn.$$

Now we combine these facts to prove the first statement (note this proof also works for subtraction):

$$\begin{aligned} a + b &= a' + kn + b' + mn \\ &= a' + b' + n(l + m) \equiv a' + b' \mod n \end{aligned}$$

Similarly for the second one we have:

$$\begin{aligned} ab &= (a' + kn)(b' + mn) \\ &= a'b' + a'mn + b'kn + kmn^2 \\ &= a'b' + n(a'm + b'k + kmn) \equiv a'b' \mod n \end{aligned}$$

□

Whilst this may seem like a trivial and obvious result, we can use it to our advantage when working with big numbers.

Example 6.2.1. For example, if we wanted to find the remainder of 3^8 when we divide by 13, we can do the following:

$$\begin{aligned} 3^8 = (3^2) = 9^4 &\equiv (-4)^4 \mod 13 \\ &= 16^2 \equiv 3^2 \mod 13 \\ &\equiv 9 \mod 13 \end{aligned}$$

Here we have used the fact that multiplication works nicely in modular arithmetic to say that $9 \times 9 \times 9 \times 9 \equiv (-4) \times (-4) \times (-4) \times (-4)$ and then we used that $16 \times 16 \equiv 3 \times 3$ to simplify the problem even further.

Example 6.2.2. Let's consider another example: what is the remainder when we divide 38×16 by 17. We use the same approach of eliminating the big numbers using modular arithmetic.

$$\begin{aligned} 38 \times 16 &\equiv 38 \times (-1) \equiv -38 \mod 17 \\ &\equiv 19 \times (-2) \equiv 2 \times (-2) \equiv -4 \mod 17. \end{aligned}$$

Saying that the remainder when you divide a number by 17 is -4 is the same thing as saying that the remainder is 13, since $-4 + 17 = 13$.

6.3 On the Chinese Remainder Theorem

6.3.1 Introduction

In this talk we cover a powerful result in modular arithmetic-the Chinese remainder theorem. The name of this theorem originates from the Chinese mathematician Sun Tzu (not the art of war guy) who was a mathematician in the third century. Little is known about the guy sadly, but he was clearly a clever chap.

6.3.2 Multiplicative Inverses

Before we look at the Chinese remainder theorem, we look at the notion of "multiplicative inverses" in modular arithmetic. First recall what we mean by "multiplicative inverse". In the real numbers, the multiplicative inverse of x is the number y such that $xy = 1$ and therefore the multiplicative inverse of x is just $\frac{1}{x}$. We wish to create a similar notion in modular arithmetic, that is, given some integer a we wish to find some x such that:

$$ax \equiv 1 \mod m.$$

By definition, this means that there is some integer b such that:

$$ax = 1 + bm \implies ax - bm = 1.$$

By Bezout's lemma, this means that $\gcd(a, m) = 1$. It turns out that the reverse is also true and so we say that a has a multiplicative inverse $\mod m$ iff $\gcd(a, m) = 1$.

Example 6.3.1. The inverse of $2 \mod 7$ exists, since $\gcd(2, 7) = 1$ and the inverse is 4 since

$$2 \times 4 \equiv 1 \mod 7.$$

However there is no inverse of $3 \mod 12$ since $\gcd(3, 12) = 3$. Therefore there is no x such that

$$3x \equiv 1 \mod 12.$$

We can actually go even further and find an expression for this inverse. First we have to define Euler's totient function. This function is defined as $\varphi(n) =$ number of integers that are coprime to n.

Example 6.3.2. $\varphi(10) = 4$ because there are 4 numbers which are coprime to 10: $1, 3, 7, 9$. One can also note that for a prime p, we have $\varphi(p) = p - 1$ because every number from 1 to $p - 1$ will be coprime to p.

Now that we have defined this function, we are ready to find an expression for a multiplicative inverse in modular arithmetic, due to the following theorem of Euler:

Theorem 6.3.1. *If a and n are coprime, then we have:*

$$a^{\varphi(n)} \equiv 1 \mod n.$$

Therefore we can say that when a and n are coprime, then we have that $a^{-1} \equiv a^{\varphi(n)-1} \mod n$.

6.3.3 Statement of the Theorem

The Chinese remainder theorem tells us the following: if you think of a number from 1 to $N-1$, where $N = n_1 n_2 \dots n_m$, and you tell me the remainders of this number when you divide by each of $n_1, n_2, \dots n_m$, then I can tell you exactly what number you're thinking of (if all of the $n_1, \dots, n_m$ are coprime with each other). In more mathematical terms we have the following:

Theorem 6.3.2 (Chinese remainder theorem)**.** *Given some list of integers $n_1, \dots, n_m$ which are all pairwise coprime, and some integers $a_1, \dots a_m$, the system of equations:*

$$\begin{aligned} x &\equiv a_1 \mod n_1 \\ x &\equiv a_2 \mod n_2 \\ &\vdots \\ x &\equiv a_m \mod n_m \end{aligned}$$

has a unique solution in $\mathbb{Z}/N\mathbb{Z}$ where $N = n_1 \dots n_m$.

We shall dive into the specific case where we have 3 congruences so that we can construct a solution for x, and then look at an example of how this works.

Example 6.3.3. Consider a system of congruences (which satisfy the necessary properties of the Chinese remainder theorem)

$$\begin{aligned} x &\equiv a_1 \mod n_1 \\ x &\equiv a_2 \mod n_2 \\ x &\equiv a_3 \mod n_3. \end{aligned}$$

We wish to construct a solution for x. If we can find some w_1, w_2, w_3 such that:

	mod n_1	mod n_2	mod n_3
w_1	1	0	0
w_2	0	1	0
w_3	0	0	1

Then $x = a_1 w_1 + a_2 w_2 + a_3 w_3$ would be a solution to the system of congruences. Therefore, we want to find such w_i's. We begin by defining

$z_1 = n_2 n_3$, $z_2 = n_1 n_3$, $z_3 = n_1 n_2$. This gives us the appropriate zeros in the table. In order to get the ones, we must use our multiplicative inverses: let $y_1 = z_1^{-1} \mod n_1$, $y_2 = z_2^{-1} \mod n_2$, $y_3 = z_3^{-1} \mod n_3$. Then:

$$
\begin{aligned}
w_1 &= z_1 y_1 \\
w_2 &= z_2 y_2 \\
w_3 &= z_3 y_3
\end{aligned}
$$

will satisfy the properties in the table as required.

Example 6.3.4. Let's apply this to a specific number. I asked Muad to think of a number between 1 and 100 and tell me the remainders when he divided by 3, 5 and 7. What he told me was:

$$
\begin{aligned}
x &\equiv 2 \mod 3 \\
x &\equiv 3 \mod 5 \\
x &\equiv 2 \mod 7.
\end{aligned}
$$

We shall use our construction, step by step. Firstly, we have $z_1 = 5 \times 7 = 35$, $z_2 = 3 \times 7 = 21$ and $z_3 = 3 \times 5 = 15$. Since $35 \equiv 2 \mod 3$, we have $y_1 = 2^{-1} \mod 3 = 2$ (since $2 \times 2 \equiv 1 \mod 3$). We have $y_2 = 21^{-1} \mod 5$ but since $21 \equiv 1 \mod 5$, this means that 21^{-1} will also just be 1 and so $y_2 = 1$. Lastly we have $y_3 = 15^{-1} \mod 7$. But again $15 \equiv 1 \mod 7$ and so $y_3 = 1$. Therefore:

$$
\begin{aligned}
w_1 &= z_1 y_1 = 70 \\
w_2 &= z_2 y_2 = 21 \\
w_3 &= z_3 y_3 = 15.
\end{aligned}
$$

Therefore, a solution for this congruence is:

$$
\begin{aligned}
x &= a_1 w_1 + a_2 w_2 + a_3 w_3 \\
&= 2 \times 70 + 3 \times 21 + 2 \times 15 = 233.
\end{aligned}
$$

However, this isn't between 1 and 100. This can be remedied by the fact that the numbers that satisfy this congruence are all the same mod $N = 3 \times 5 \times \times 7 = 105$. So the final step is to compute $233 \mod 105$, which we can see is 23 which was the number that Muad had in mind!

6.4 Equivalence Classes in Modular Arithmetic and Applications

6.4.1 Introduction

Last time in maths society we saw that multiplication and addition were well defined in modular arithmetic. In this talk we press on further with the theme of modular arithmetic, with the main theme this time being the notion of equivalence classes.

6.4.2 Equivalence Classes

Before we define equivalence classes, let us first look into what an equivalence relation. That is, what are the necessary conditions for some relation R to be a notion of equivalency? The three properties are as follows:

Definition 6.4.1. An relation R between two numbers [1] a and b (denoted aRb) is said to be an equivalence relation if:

1. The relation is reflexive: aRa.
2. The relation is symmetric: $aRb \iff bRa$.
3. The relation is transitive: If aRb and bRc then aRc.

An easy example of this is the $=$ relation, since it's obvious that $a = a$ for all a and $a = b \iff b = a$ and $a = b$ and $b = c$ does indeed imply that $a = c$. A counter example would be the relation $\leq$ because it is not symmetric: $a \leq b \not\Longrightarrow b \leq a$ or another counter example would be "has a greatest common factor greater than 1", because it's not transitive: for example 2 and 10 would be equivalent under this relation and so would 10 and 5, but 2 and 5 aren't. Anyway, it should be relatively clear that the following theorem is true:

Theorem 6.4.1. *The relation $a \equiv b \mod n$ is an equivalence relation.*

With an equivalence relation, we can form sets called equivalence classes, which consider all the elements that are equivalent to each other under an

[1] although generally these could be members of any set

equivalence relation. For example, if we are working mod 5, the equivalence class of zero would be given by:

$$[0] = \{n \in \mathbb{Z} : n \equiv 0 \mod 5\} = \{\dots, -5, 0, 5, 10, \dots\}.$$

Similarly the equivalence classes of 1,2,3 and 4 are given as follows:

$$\begin{aligned}
[1] &= \{\dots, -4, 1, 6, 11, \dots\} \\
[2] &= \{\dots, -3, 2, 7, 12, \dots\} \\
[3] &= \{\dots, -2, 3, 8, 13, \dots\} \\
[4] &= \{\dots - 1, 4, 9, 14, \dots\}
\end{aligned}$$

and all of a sudden, through just five equivalence classes, we have covered all of the integers because we have $[0] = [5]$ and $[1] = [6]$ and so on. Given this, we can define the integers mod 5:

Definition 6.4.2. The integers mod 5 are given by:

$$\mathbb{Z}/5\mathbb{Z} = \{[0], [1], [2], [3], [4]\}$$

And in general we say that the integers mod n are given by:

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], [2], \dots [n-1]\}.$$

We shall look at an example of why splitting up the integers into these finite sets can be powerful by looking at an example from $\mathbb{Z}/7\mathbb{Z}$. Before we move onto our main result, we require the following lemma:

Lemma 6.4.2. For any integer n, $n^3 = 0, 1$ or -1 mod 7.

Proof. We shall prove this by using the fact that we only need to consider 7 cases (due to the results from last week). Those are what happen to $0, 1, 2, 3, 4, 5, 6$ when we cube them? To make things easier we shall consider -3 instead of 4, -2 instead of 5 and -1 instead of 6 (because these are all equivalent mod 7). We have:

$$
\begin{aligned}
&(-3)^3 \equiv 1 \mod 7\\
&(-2)^3 \equiv -1 \mod 7\\
&(-1)^3 \equiv -1 \mod 7\\
&0^3 \equiv 0 \mod 7\\
&1^3 \equiv 1 \mod 7\\
&2^3 \equiv 1 \mod 7\\
&3^3 \equiv -1 \mod 7
\end{aligned}
$$

Since these seven computations tell us about the whole of the integers, the proof is complete. □

Now we move to the main theorem we wish to prove:

Theorem 6.4.3. *The equation*

$$x^3 + 500 = y^3$$

has no integer solutions.

Proof. Let's assume that there exist integers (x, y) that satisfy the equation. Then the two sides would also be equivalent mod 7. However, we have

$$
\begin{aligned}
500 &\equiv 5 \times 10^2 \equiv 5 \times 3^2 \mod 7\\
&\equiv 5 \times 9 \equiv 5 \times 2 \equiv 10 \equiv 3 \mod 7.
\end{aligned}
$$

So therefore our integer pair would satisfy the following equation in $\mathbb{Z}/7\mathbb{Z}$

$$x^3 + 3 = y^3.$$

However since x^3 and y^3 must be 0 or ± 1, this statement cant be true and so our original assumption that such an integer pair existed must've been wrong, as required. □

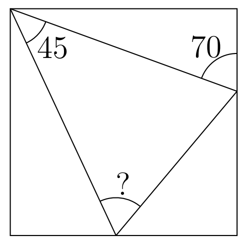
In this proof we can see how powerful our notion of modular arithmetic can be. At first the theorem we just showed with ease seems unapproachable, but by using some more abstract machinery in splitting up the integers into finitely many different equivalence classes, we were able to break down the problem into a much easier one.

6.5 Problems of the Week

Question 6.5.1. How many perfect cubes lie between $2^8 + 1$ and $2^{18} + 1$, inclusive?

Question 6.5.2. If I roll 10 fair dice and sum up all the numbers I get, what is the probability that that number is a multiple of 6?

Question 6.5.3. Given that the quadrilateral is a square, find the angle:



Solution 6.5.1. Firstly, we know that $2^8 + 1 = 257$ and so the first perfect cube after that is $7^3 = 343$. Now, we can cleverly extract a cube number from 2^{18} since $2^{18} = (2^6)^3 = 64^3$. Thus the biggest cube number that is less than $2^{18} + 1$ is 64^3 and the smallest cube number in the range is 7^3 and thus there are $64 - 6 = \boxed{58}$ perfect cubes between $2^8 + 1$ and $2^{18} + 1$.

Solution 6.5.2. The trick to this problem is to realise that the first 9 rolls of the dice don't actually matter for our problem. This is because whatever number you get after the first 9 rolls, there's always a $\frac{1}{6}$ chance that the 10th dice will make the total a multiple of 6. Therefore the probability ends up being $\frac{1}{6}$.

Remark. It has dawned upon me that this actually has a modular arithmetic interpretation although I admit that this was not my original intention. I shall sketch it here. Using the notion of modular arithmetic, we can form equivalence classes for every $n \in \mathbb{Z}$. We shall use the example of mod 6 to help us with our problem. We define the equivalence class of n to be: $[n]_6 = \{k \in \mathbb{Z} : k \equiv n \mod 6\}$. For example, $[1]_6 = \{\dots, -5, 1, 7, 13, \dots\}$ since it is the set of all numbers that have a remainder of 1 when divided by 6. Therefore we can define "the integers mod 6" as:

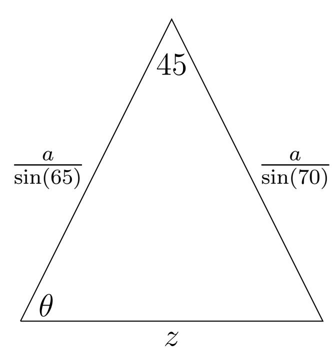
$$\mathbb{Z}/6\mathbb{Z} = \{[0], [1], [2], [3], [4], [5]\},$$

since $[6] = [0]$ and $[7] = [1]$ and so on, these 6 equivalence classes are all we need. Furthermore, we drop the equivalence class notation normally. Now, the result about addition that we showed at the start of the talk allows us to nicely define addition as follows:

$$[a] + [b] = [a + b]$$

For every equivalence class in $\mathbb{Z}/6\mathbb{Z}$ we have exactly one other equivalence class that we can add to it to give us $[0]$ (for example $[4] + [2] = [0]$. Since the dice is precisely just $\mathbb{Z}/6\mathbb{Z}$ (we just replace the number 6 on the dice with 0), there will always be one number on the dice that we can add to the number that takes us to $[0]$. This notion of inverses can lead one later to groups.

Solution 6.5.3. My solution that I will write up here will be much more ugly than Vincent's original solution, which was through constructing a congruent triangle (I'll let you figure out how to do that). Instead we shall use the cosine rule and the sine rule to do it. From now on, the length of the square is called a. We can use the sine rule to show that the triangle in the middle has the following lengths where z is just some unknown variable:



We may then use the cosine rule:

$$\begin{aligned} z^2 &= \frac{a^2}{\sin^2(65)} + \frac{a^2}{\sin^2(70)} - 2\frac{a^2}{\sin(65)\sin(70)}\cos(45) \\ z &= a\sqrt{\frac{1}{\sin^2(65)} + \frac{1}{\sin^2(70)} - 2\frac{\cos(45)}{\sin(65)\sin(70)}}. \end{aligned}$$

and then we can use the sine rule to obtain:

$$\frac{\sin\theta}{\left(\frac{a}{\sin 70}\right)} = \frac{\sin 45}{z}$$

$$\boxed{\theta = 65^\circ}$$

Chapter 7

Saga 7: A Brief Look at Algebraic Number Theory

7.1 An Application of the Gaussian Integers

7.1.1 Introduction

This talk is motivated by an interesting observation about prime numbers. Whenever we consider a prime p which has a remainder of 1 when you divide by 4 (for example 5, 13, 17 etc.), it seems to always be the case that it can be written as the sum of two squares. The first few examples of this are:

$$\begin{aligned} 5 &= 2^2 + 1^2 \\ 13 &= 3^2 + 2^2 \\ 17 &= 4^2 + 1^2 \\ 29 &= 5^2 + 2^2 \\ &\vdots \end{aligned}$$

Whilst we may use brute force to show that this always works for the first million such primes, proving that this works <u>for all such primes</u> seems like an extremely daunting task. The method of attack for proving this is an example of what is studied in the field of algebraic number theory and is really quite ingenious. The plan is as follows:

1. Instead of working in the integers $\mathbb{Z}$, we shall now study the Gaussian integers $\mathbb{Z}[i]$.

2. We define the notion of what it means to be prime in our new system of integers and study what properties they satisfy

3. We define a map which takes a Gaussian integer $a + bi$ and outputs an integer.

4. Since the integers are a subset of the Gaussian integers, we shall then consider our primes of the form $p = 4n + 1$, except in the context of the Gaussian integers and see how this changes things.

5. We then take our results from the Gaussian integers back home to the land of regular integers, via our map (which will be called the norm map) and then with this new information that we've gathered, we can finally prove the original result.

This seems like a solid plan, so we need to get to work!

7.1.2 The Gaussian Integers

The Gaussian integers are defined as:

$$\mathbb{Z}[i] = \{a + bi : a \in \mathbb{Z} \text{ and } b \in \mathbb{Z}\}.$$

Example 7.1.1. A few examples of Gaussian integers would be $5, 1+3i, 2i, 3+2i$ and a few examples of numbers that are not would be $5.5, 1+3.2i, 9.1i, 9.1+i$.

Another vital map which we must define is called the norm, which assigns to every Gaussian integer a positive integer in the following way:

$$N(x + iy) = x^2 + y^2.$$

Notice that the norm is multiplicative: $N(\alpha\beta) = N(\alpha)N(\beta)$ since, writing

$\alpha = a + bi$ and $\beta = c + di$,

$$\begin{aligned} N(\alpha\beta) &= N((a+bi)(c+di)) \\ &= N(ac + bci + adi - bd) \\ &= N(ac - bd + (bc + ad)i) \\ &= (ac - bd)^2 + (bc + ad)^2 \\ &= (ac)^2 + (bd)^2 - 2abcd + 2abcd + (bc)^2 + (ad)^2 \\ &= (ac)^2 + (bd)^2 + (bc)^2 + (ad)^2 \\ &= (a^2 + b^2)(c^2 + d^2) \\ &= N(\alpha)N(\beta). \end{aligned}$$

Therefore, we may now use this norm to our advantage. Firstly, we say that some Gaussian integer α is invertible if there is some Gaussian integer β such that $\alpha\beta = 1$. We can use the norm as follows:

$$N(\alpha\beta) = N(\alpha)N(\beta) = 1.$$

But since the norms are positive integers, this means that $N(\alpha) = 1$. The only Gaussian integers with this property are $\{\pm 1, \pm i\}$ and we shall call them units. Note that this is very similar to ± 1 in the integers, so in the Gaussian integers, being off by a unit is basically the same as being off by a sign in the integers. Now we consider the notion of what a prime is, but in the set of Gaussian integers. In the integers, we say a number p is prime if $p = xy \implies x = \pm 1$. Therefore, we define primes as follows:

Definition 7.1.1. We define a Gaussian integer $p \in \mathbb{Z}[i]$ to be prime if

$$p = \alpha\beta \implies N(\alpha) = 1.$$

That is, if we factor p as the product of two Gaussian integers, one of them must have been a unit. Note that not all regular primes are also Gaussian primes; for example $2 = (1 + i)(1 - i)$ is not a Gaussian prime.

Gaussian primes have the following very important (you will see why soon) property, which regular primes also have:

Lemma 7.1.1. If some Gaussian prime p divides $\alpha\beta$, then p divides α or p divides β.[1]

[1] In more concise notation, $p|\alpha\beta \implies p|\alpha$ or $p|\beta$.

7.1.3 Bringing it All Together

We have now developed enough machinery in the setting of the Gaussian integers to prove our original result. That is the following (incredible) result:

Theorem 7.1.2. *A prime p can be written as the sum of two squares if and only if $p \equiv 1 \mod 4$.*

Proof. ($\Rightarrow$) First, let's quickly remark that if $p = a^2 + b^2$ then since p is odd (for $p > 2$), then one of a and b is odd and the other is even. Since an odd number squared has a remainder of 1 when divided by 4 and an even number is divisible by 4, then this becomes:

$$p \equiv a^2 + b^2 \equiv 0 + 1 \equiv 1 \mod 4.$$

($\Leftarrow$) Now we turn to the amazing part of this theorem. That is, if we're given a prime $p = 4n + 1$, then it can be written as $a^2 + b^2$ for some integers a, b. Firstly, we shall show that our prime $p = 4n+1$ is not a Gaussian prime. To do this, we shall use the fact that there exists a solution to the congruence

$$x^2 \equiv -1 \mod p. \tag{7.1}$$

This implies that p divides $x^2 + 1$. However, since we're working in the Gaussian integers we may factor this as:

$$p \text{ divides } (x + i)(x - i).$$

However, p does not divide $x + i$ or $x - i$, and so by 7.1.1, this means that p is not a Gaussian prime. Before we proceed with the proof, we need to show that such an x exists. We use a theorem of Wilson which states that for any prime p we have:

$$(p - 1)! \equiv -1 \mod p.$$

Since we have said that $p = 4n + 1$, we have $p - 1 = 4n$ and so:

$$\begin{aligned}(p-1)! &\equiv ((p-1)(p-2)\dots(p-2n))((2n)(2n-1)\dots 1) \mod p\\ &\equiv ((-1)(-2)\dots(-2n))\,((2n)(2n-1)\dots 1) \mod p\\ &\equiv (-1)^{2n}(2n)!(2n)! \equiv ((2n)!)^2 \mod p.\end{aligned}$$

So by Wilson's theorem we have:

$$((2n)!)^2 \equiv -1 \mod p$$

and so a solution to 7.1 is $x = (2n)!$ and so p is indeed not a Gaussian prime. This is excellent because now, like a phoenix from the ashes, the norm function comes back to save us. Since p is not a Gaussian prime, we may write

$$p = \alpha\beta$$

for some Gaussian integers that are not units α, β. Taking norms on both sides, we have:

$$N(p) = N(\alpha)N(\beta)$$
$$p^2 = N(\alpha)N(\beta).$$

Since p is a prime (in the regular integers), this implies that either $N(\alpha) = N(\beta) = p$ or one of the norms are 1 and the other is p^2. However, since p is not a Gaussian prime, neither α or β are units and so neither of them have a norm of 1. Thus, writing $\alpha = a + bi$, we have:

$$N(\alpha) = a^2 + b^2 = p$$

as required. □

Chapter 8

Saga 8: Fourier Series (Featuring Wren)

8.1 Sine Fourier Series

8.1.1 Introduction

In today's talk we began our exploration of Fourier series. Just as with Taylor series, we are aiming to express a function as an infinite sum. With the Taylor series we aim to take a function and write it as an infinite polynomial. However the Fourier series aims to take a function and write it as an infinite sum of sines and cosines. This series representation can be extremely helpful, especially when considering differential equations as the trig functions have particularly nice derivatives to work with. The original example of this is with the heat equation when Fourier realised in 1822 that if he could write functions in this form that the problem would be much easier. In today's talk we hold back on writing functions in terms of both sin and cos, but stick to just considering functions that can be written as an infinite sums of sines.

8.1.2 Orthogonality of the Sine Function

In this talk our goal is to be able to write a function $f(x)$ in the form

$$f(x) = \sum_{n=1}^{\infty} a_n \sin(nx).$$

Before I proceed, it is worth noting that this immediately implies that our function is periodic ($f(x + 2\pi) = f(x)$) and that it is odd $f(-x) = -f(x)$. Now, what we need is to find some sort of method for finding the coefficients a_n. To do this we shall use the following property:

Lemma 8.1.1 (Sine is Orthogonal)**.** For integers n, k such that $n \neq k$, we have:

$$\int_{-\pi}^{\pi} \sin(nx)\sin(kx)\mathrm{d}x = 0$$

Proof. Recall that

$$\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$$
$$\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b).$$

Subtracting these two equations gives:

$$\sin(a)\sin(b) = \frac{1}{2}(\cos(a - b) - \cos(a + b))$$

Therefore we may rewrite our integral in the following form:

$$\int_{-\pi}^{\pi} \frac{1}{2}(\cos((n - k)x) - \cos((n + k)x))\mathrm{d}x.$$

This can be much more easily integrated as:

$$\frac{1}{2}\left(\left[\frac{\sin((n-k)x)}{n-k}\right]_{-\pi}^{\pi} - \left[\frac{\sin((n+k)x)}{n+k}\right]_{-\pi}^{\pi}\right)$$
$$= 0$$

because sin is always zero for all integer multiples of π. □

Now we consider the case when $n = k$. In this case we use the identity $\sin^2 x = \frac{1}{2} - \frac{\cos(2x)}{2}$ to write:

$$\int_{-\pi}^{\pi} \sin^2(kx)\mathrm{d}x = \int_{-\pi}^{\pi} \frac{1}{2} - \frac{\cos(2kx)}{2}\mathrm{d}x$$
$$= \left[\frac{1}{2}\right]_{-\pi}^{\pi} - \left[\frac{\sin(2kx)}{4k}\right]_{-\pi}^{\pi}$$
$$= \pi.$$

8.1.3 Finding the Coefficients

Now we shall use the orthogonality condition that we proved in the previous section in order to determine the coefficients of our Fourier series. If we write

$$f(x) = \sum_{n=1}^{\infty} a_n \sin(nx)$$

Then we have:

$$\int_{-\pi}^{\pi} f(x)\sin(kx)\mathrm{d}x = \int_{-\pi}^{\pi} (a_1 \sin(x) + a_2 \sin(2x) + a_3 \sin(3x) + \dots)\sin(kx)\mathrm{d}x$$
$$= \int_{-\pi}^{\pi} a_1 \sin(x)\sin(kx)\mathrm{d}x + \int_{-\pi}^{\pi} a_2 \sin(2x)\sin(kx)\mathrm{d}x + \cdots + \int_{-\pi}^{\pi} a_k \sin^2(kx)\mathrm{d}x + \dots$$
$$= \int_{-\pi}^{\pi} a_k \sin^2(kx)\mathrm{d}x$$
$$= \pi a_k$$

Therefore we may write:

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\mathrm{d}x.$$

So in contrast to how with Taylor series you must compute the kth derivative, for Fourier series you need to compute an integral to find the coefficients of your series.

8.1.4 An Example: Square-wave function

We will now see an example of Fourier series in action. Consider the square wave function, which is 0 at multiples of π and alternates between 1 and -1.

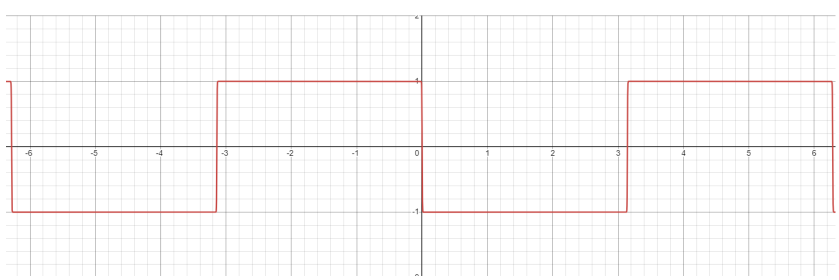


Figure 8.1: Square Wave Function

We will denote this function by $S(x)$ and we shall attempt to expand it via Fourier series. Writing

$$S(x) = \sum_{n=1}^{\infty} a_n \sin(nx),$$

we have that the coefficients are given by

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} S(x) \sin(kx) \mathrm{d}x.$$

Since both $S(x)$ and $\sin(kx)$ are both odd functions, that means that their product is even and so we may write this integral as

$$\frac{2}{\pi} \int_{0}^{\pi} S(x) \sin(kx) \mathrm{d}x.$$

However our square function is precisely 1 between 0 and π and so our coefficients are simply just

$$\begin{aligned} a_k &= \frac{2}{\pi} \int_0^{\pi} \sin(kx) \mathrm{d}x \\ &= \frac{2}{\pi} \left[\frac{-\cos(kx)}{k} \right]_0^{\pi} \end{aligned}$$

For even k, $\cos(k\pi) = \cos(0) = 1$, but for odd multiples of k, $\cos(k\pi) = -1$ and so we have that for even k our coefficients are

$$\frac{2}{\pi} \left(\frac{-1}{k} - \left(\frac{-1}{k} \right) \right) = 0$$

and for odd k our coefficients are

$$\frac{2}{\pi} \left(\frac{1}{k} - \left(\frac{-1}{k} \right) \right) = \frac{4}{k\pi}.$$

With this information, we plug these coefficients back into our sum and see that:

$$\begin{aligned} S(x) &= a_1 \sin(x) + a_2 \sin(2x) + a_3 \sin(3x) + a_4 \sin(4x) + a_5 \sin(5x) + \ldots \\ &= \frac{4}{\pi} \sin(x) + 0 + \frac{4}{3\pi} \sin(3x) + 0 + \frac{4}{5\pi} \sin(5x) + \ldots \\ &= \frac{4}{\pi} \left(\sin(x) + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \frac{\sin(7x)}{7} + \ldots \right) \end{aligned}$$

When we plot these terms we can observe that they become closer and closer to becoming a perfect square wave, as desired.

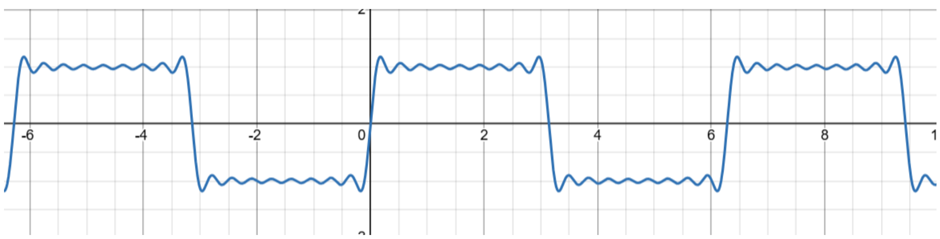


Figure 8.2: Plotting the Fourier series expansion, up to $\frac{\sin(19x)}{19}$

A bonus result that one can obtain from this is by plugging in $\frac{\pi}{2}$ into our Fourier series which gives us:

$$S\left(\frac{\pi}{2}\right) = 1 = \frac{4}{\pi}\left(\sin\left(\frac{\pi}{2}\right) + \frac{\sin\left(\frac{3\pi}{2}\right)}{3} + \frac{\sin\left(\frac{5\pi}{2}\right)}{5} + \frac{\sin\left(\frac{7\pi}{2}\right)}{7} + \dots\right)$$
$$= \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots\right)$$
$$\implies \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots$$

which is a nice series expansion for π.

8.2 Cosine Fourier Series

8.2.1 Introduction

Today we continued on from last week's talk on the sine Fourier series naturally by now considering the cosine Fourier series. Last time we wanted to look at a function as a sum of sines, but now we are considering functions as a sum of cosines. That is we want to take a function and write it as:

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx).$$

8.2.2 Orthogonality of Cosine

Almost exactly like last time, the key property of cosine that we will be using is the following:

Theorem 8.2.1. *For all integers n, k where $n \neq k$ we have that:*

$$\int_{-\pi}^{\pi} \cos(nx)\cos(kx)\mathrm{d}x = 0.$$

Proof. Because sums are much easier to work with than products, we want to decompose this as a sum of trigonometric functions instead. To do this, remember that:

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$
$$\cos(A-B) = \cos(A)\cos(B) + \sin(A)\sin(B).$$

When we add these two together, we get that:

$$\frac{1}{2}(\cos(A+B) + \cos(A-B)) = \cos(A)\cos(B).$$

Thus we can write our integral as

$$\begin{aligned}
&\frac{1}{2}\int_{-\pi}^{\pi} \cos((n+m)x) + \cos((n-m)x)\mathrm{d}x \\
&= \frac{1}{2}\left[\frac{-\sin((n+m)x)}{n+m} - \frac{\sin((n-m)x)}{n-m}\right]_{-\pi}^{\pi} \\
&= 0
\end{aligned}$$

because sin of any integer multiple of π evaluates to zero. Conversely, when $n = k$ we have:

$$\begin{aligned}
&\int_{-\pi}^{\pi} \cos^2(nx)\mathrm{d}x \\
&= \frac{1}{2}\int_{-\pi}^{\pi} \cos(2nx) + 1\mathrm{d}x \\
&= \frac{1}{2}\left[\frac{\sin(2nx)}{2n} + x\right]_{-\pi}^{\pi} \\
&= \pi.
\end{aligned}$$

□

8.2.3 Finding Coefficients

If we want to write our function as

$$f(x) = a_0 + a_1 \cos(x) + a_2 \cos(2x) + \cdots + a_n \cos(nx) + \ldots$$

then to find the coefficient a_n we times every term by $\cos(nx)$ and integrate from $-\pi$ to π and everything will disappear except for our a_n term as follows:

$$\begin{aligned}
&\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x \\
&= \int_{-\pi}^{\pi} a_0\cos(nx) + a_1\cos(x)\cos(nx) + \cdots + a_n\cos^2(nx) + \ldots \mathrm{d}x \\
&= 0 + 0 + \cdots + \pi a_n + 0 + \ldots \\
&\implies \boxed{a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x.}
\end{aligned}$$

The only exception is a_0 because

$$\begin{aligned}
\int_{-\pi}^{\pi} a_0 \mathrm{d}x &= 2\pi a_0 \\
\implies a_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x.
\end{aligned}$$

8.2.4 Example: Repeating Ramp Function

For today's talk we'll be considering the repeating ramp function which is essentially just $|x|$ with a period of $[-\pi, \pi]$, so it looks like:

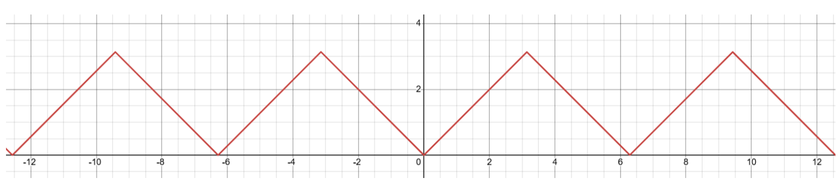


Figure 8.3: Repeating Ramp Function

We shall now just get directly into it and plug in our formula for the cosine Fourier series:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x.$$

Now recall that $f(x)$ and $\cos(nx)$ are both even, so their product is too and thus we can write:

$$\begin{aligned}
a_n &= \frac{2}{\pi}\int_0^\pi f(x)\cos(nx)\mathrm{d}x \\
&= \frac{2}{\pi}\int_0^\pi x\cos(nx)\mathrm{d}x \\
&= \frac{2}{\pi}\left(\left[\frac{x\sin(nx)}{n}\right]_0^\pi - \int_0^\pi \frac{\sin(nx)}{n}\right) \\
&= \frac{2}{\pi}\left(0 - \left[\frac{-\cos(nx)}{n^2}\right]_0^\pi\right) \\
&= \frac{2}{\pi}\left[\frac{\cos(nx)}{n^2}\right]_0^\pi
\end{aligned}$$

Where the second step came from the fact that $f(x) = x$ between 0 and π and the third step came by applying integration by parts. Let's now examine this final expression we have here. When n is even, we will have $\cos(n\pi) = \cos(0)$ and so when we plug in the bounds they will cancel each other out and give us 0 and thus when n is even we have $a_n = 0$. However when n is odd, we have that the coefficient is:

$$\begin{aligned}
&\frac{2}{\pi}\left(\frac{-1}{n^2} - \frac{1}{n^2}\right) \\
&= \frac{-4}{\pi n^2}.
\end{aligned}$$

To conclude, for $n \geq 1$ we have that $a_n = 0$ for even n and $a_n = \frac{-4}{\pi n^2}$ for odd n. The only thing left to calculate is a_0 which is given by:

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^\pi f(x)\mathrm{d}x \\
&= \frac{1}{\pi}\int_0^\pi x\mathrm{d}x \\
&= \frac{\pi}{2}.
\end{aligned}$$

Therefore our function is given by:

$$\begin{aligned}
f(x) &= \frac{\pi}{2} - \frac{4}{\pi}\cos(x) - \frac{4}{9\pi}\cos(3x) - \frac{4}{25\pi}\cos(5x) - \dots \\
&= \frac{\pi}{2} - \frac{4}{\pi}\left(\cos(x) + \frac{\cos(3x)}{9} + \frac{\cos(5x)}{25} + \dots\right)
\end{aligned}$$

Plotting the first 11 terms of this gives us the following very accurate approximation of the repeating ramp function:

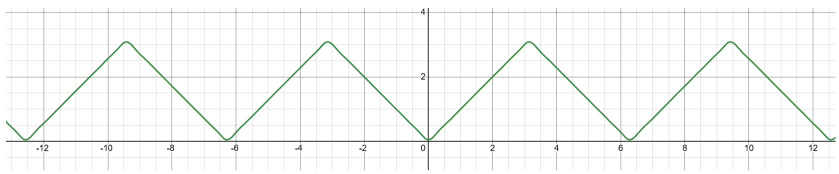


Figure 8.4: Approximation of Repeating Ramp Function

8.3 Parseval's Theorem

8.3.1 Introduction

In today's talk we shall discuss a very powerful theorem known as Parseval's theorem, and we shall use it to compute a lesser known analogue to the Basel problem:

$$\sum_{n=1}^{\infty} \frac{1}{n^4}.$$

8.3.2 The Complete Fourier Series

Before we dive into Parseval's theorem, we shall first combine the last two talks to form the complete Fourier series. That is, when our function at hand is neither odd nor even then we simply write it as:

$$a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx).$$

Our rule for finding the coefficients is still the same because we can use the fact that

$$\int_{-\pi}^{\pi} \sin(nx)\cos(mx)\mathrm{d}x = 0$$

This simply follows from the fact that $\sin(nx)\cos(mx)$ is an odd function and since we're integrating from $-a$ to a (for some real a), the integral will evaluate to zero. Thus the rules for our coefficients remain as follows:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x$$
$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\mathrm{d}x$$

8.3.3 Statement of the Theorem and an Application

Let us write

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n\cos(nx) + \sum_{n=1}^{\infty} b_n\sin(nx).$$

Then we have the following theorem:

Theorem 8.3.1 (Parseval's Theorem)**.**

$$\int_{-\pi}^{\pi}(f(x))^2\mathrm{d}x = 2\pi a_0^2 + \pi\sum_{n=1}^{\infty} a_n^2 + b_n^2.$$

Proof. If we look at the Fourier series of $f(x)$, then when we expand $(f(x))^2$ we will get a bunch of terms of the form $a_m b_n\cos(mx)\sin(nx)$ which will get removed by integrating from $-\pi$ to π. We will also get a bunch of terms of the form $a_m a_n\cos(mx)\cos(nx)$ which also get sent to zero for all $n \neq m$. Similarly all terms of the form $b_n b_m\sin(nx)\sin(mx)$ get sent to zero too, where $n \neq m$. The only other terms to consider are those of the form $a_n^2\cos^2(nx)$ and $b_n^2\sin^2(nx)$. When we integrate these from $-\pi$ to π, we get πa_n^2 and πb_n^2 except for when we integrate a_0^2 in which case we get $2\pi a_0^2$. Collecting all these terms together gives us the sum we desire. □

Now we move to showing why this theorem is so powerful, with the aim of using it to compute $\sum_{n=1}^{\infty}\frac{1}{n^4}$. We begin by considering the function $f(x) = \pi^2 - x^2$ where $x \in [-\pi, \pi]$. Since this function is even its Fourier series will just be a cosine Fourier series. We know that the coefficients are given by

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi}(\pi^2 - x^2)\cos(nx)\mathrm{d}x.$$

Now using the fact that $\cos(nx)$ integrates to zero between $-\pi$ and π we really just have to do the following computation

$$\begin{aligned}
&\frac{-1}{\pi}\int_{-\pi}^{\pi} x^2\cos(nx)\mathrm{d}x \\
&= \frac{-1}{\pi}\left(\left[\frac{x^2\sin(nx)}{n}\right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi}\frac{2x\sin(nx)}{n}\mathrm{d}x\right) \\
&= \frac{-1}{\pi}\left(\left[\frac{2x\cos(nx)}{n^2}\right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi}\frac{2\cos(nx)}{n^2}\mathrm{d}x\right) \\
&= \frac{-1}{\pi}\left[\frac{2x\cos(nx)}{n^2}\right]_{-\pi}^{\pi}
\end{aligned}$$

Where I have done integration by parts twice. Now we know that $\cos(n\pi) = \cos(-n\pi) = (-1)^n$ and so when we plug in the bounds we get that the coefficient a_n is given by:

$$\begin{aligned}
&\frac{-1}{\pi}\left(\frac{2\pi(-1)^n}{n^2} - \frac{-2\pi(-1)^n}{n^2}\right) \\
&= \frac{-1}{\pi}\left(\frac{2\pi(-1)^n}{n^2} + \frac{2\pi(-1)^n}{n^2}\right) \\
&= \frac{-1}{\pi}\left(\frac{4\pi(-1)^n}{n^2}\right) \\
&= \frac{4(-1)^{n+1}}{n^2}.
\end{aligned}$$

Now this is good news because we have an n^2 on the bottom, so when we apply Parseval's theorem we will get an n^4 on the bottom of the fraction and the alternating signs will just always be positive. The only matter left to tend to before we dive in is to find a_0. This is given by

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\pi^2 - x^2\mathrm{d}x \\
&= \frac{1}{2\pi}\left(2\pi^3 - \frac{2\pi^3}{3}\right) \\
&= \frac{2\pi^2}{3}.
\end{aligned}$$

Thus we have that

$$f(x) = \frac{2\pi^2}{3} + \sum_{n=1}^{\infty} \frac{4(-1)^{n+1}}{n^2} \cos(nx).$$

Now we have all the information we need to go in for the kill and apply Parseval's theorem:

$$\begin{aligned}
\int_{-\pi}^{\pi} (\pi^2 - x^2)^2 \mathrm{d}x &= 2\pi \left(\frac{2\pi^2}{3}\right)^2 + \pi \sum_{n=1}^{\infty} \left(\frac{4(-1)^{n+1}}{n^2}\right)^2 \\
\int_{-\pi}^{\pi} \pi^4 - 2x^2\pi^2 + x^4 \mathrm{d}x &= \frac{8\pi^5}{9} + 16\pi \sum_{n=1}^{\infty} \frac{1}{n^4} \\
2\pi^5 - \frac{4\pi^5}{3} + \frac{2\pi^5}{5} &= \frac{8\pi^5}{9} + 16\pi \sum_{n=1}^{\infty} \frac{1}{n^4}
\end{aligned}$$

Now we know that we're almost in business, it's just a matter of rearranging:

$$\begin{aligned}
\frac{16\pi^5}{15} &= \frac{8\pi^5}{9} + 16\pi \sum_{n=1}^{\infty} \frac{1}{n^4} \\
\frac{16\pi^4}{15} &= \frac{8\pi^4}{9} + 16 \sum_{n=1}^{\infty} \frac{1}{n^4} \\
&\boxed{\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}}
\end{aligned}$$

Just like that we have managed to use Parseval's theorem to get the sum that we wanted. Interestingly the power on the bottom of the fraction seems to match up with the power of π in the final sum formula.

8.4 Poisson Summation Formula: Wren Shakespeare

8.4.1 The Result

Say we want to evaulate a sum S, such that

$$S = \sum_{n=-\infty}^{\infty} f(n)$$

for some suitable $f(n)$. To do this, we use the old trick of inserting a new variable - an obvious choice is:

$$S(a) = \sum_{n=-\infty}^{\infty} f(n+a)$$

Examining this sum, it should be fairly clear that $f(a+1) = f(a)$, as this just amount to shifting all the sum indexes by 1. Therefore, $S(a)$ is periodic, and admits a Fourier series expansion on $[0, 1]$. So, we can just use the normal formula for the Fourier series coefficients:

$$a_n = \int_0^1 e^{-2i\pi na} S(n+a)da = \int_0^1 e^{-2i\pi na} \left(\sum_{n=-\infty}^{\infty} f(n+a) \right) da$$

$$= \sum_{n=-\infty}^{\infty} \int_0^1 e^{-2i\pi na} f(n+a)da = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2i\pi na} f(a)da$$

But this is exactly the Fourier *transform* of f! So we have:

$$\sum_{n=-\infty}^{\infty} f(n+a) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2i\pi na}$$

And specifically if we let $a = 0$, we obtain the beautiful result:

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

which is the Poisson summation formula.

8.4.2 An Example

We have already shown (in my last talk, actually!) that

$$\int_{-\infty}^{\infty} \frac{\sin(x)}{x} dx = \pi$$

Now, we can consider the discrete analogue of this integral, and evaluate

$$S = \sum_{n=-\infty}^{\infty} \frac{\sin(n)}{n}$$

To do this we shall find a much simpler function whose Fourier transform is the function that we would like to sum and then apply Poisson summation formula. We're going to begin by evaluating a seemingly unrelated integral:

$$\int_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}} e^{-2i\pi nx} dx = -\left.\frac{e^{-2i\pi nx}}{2i\pi n}\right|_{-\frac{1}{2\pi}}^{\frac{1}{2\pi}}$$
$$= \frac{1}{2i\pi n}\left(e^{in} - e^{-in}\right) = \frac{\sin(n)}{\pi n}$$

We clearly have something related to what we want to sum here, and our integral sort of looks like a Fourier transform - in fact it is a Fourier transform, of the function defined as follows:

$$b(x) = \begin{cases} 1 & -\frac{1}{2\pi} < x < \frac{1}{2\pi} \\ 0 & \text{otherwise} \end{cases}$$

Wrapping this all up, we can now invoke the Poisson summation formula to get to our result:

$$\sum_{n=-\infty}^{\infty} b(n) = \sum_{n=-\infty}^{\infty} \frac{\sin(n)}{\pi n} \text{ (PSF)}$$
$$\cdots + 0 + 0 + 1 + 0 + 0 \cdots = \sum_{n=-\infty}^{\infty} \frac{\sin(n)}{\pi n}$$
$$\sum_{n=-\infty}^{\infty} \frac{\sin(n)}{\pi n} = 1$$
$$\sum_{n=-\infty}^{\infty} \frac{\sin(n)}{n} = \pi$$

which is our desired result, and exactly equal to the integral of the same function.

8.5 Problems of the Week

Question 8.5.1. Show that, for all x,

$$\arcsin(x) + \arccos(x) = \frac{\pi}{2}.$$

Question 8.5.2. Using the expansion of the repeating ramp function, show that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

Solution 8.5.1. Let us pick some x and y such that $\sin(y) = x$. Then we have that

$$\cos\left(\frac{\pi}{2} - y\right) = x.$$

Combining these facts, we have that:

$$\begin{aligned} \arcsin(x) &= y \\ \arccos(x) &= \frac{\pi}{2} - y. \end{aligned}$$

Adding these two equations yields the desired result:

$$\arccos(x) + \arcsin(x) = \frac{\pi}{2}$$

Solution 8.5.2. We begin by plugging in zero into our expansion:

$$\begin{aligned} 0 &= \frac{\pi}{2} - \frac{4}{\pi}\left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \dots\right) \\ \frac{4}{\pi}\left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \dots\right) &= \frac{\pi}{2} \\ 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \dots &= \frac{\pi^2}{8}. \end{aligned}$$

Now we have found $\sum_{n=1}^{\infty} \frac{1}{(2n-1)^2}$, but the missing terms in our sum that we want to find are the even squares:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{n^2} &= \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} + \sum_{n=1}^{\infty} \frac{1}{(2n)^2} \\ &= \frac{\pi^2}{8} + \frac{1}{4}\sum_{n=1}^{\infty} \frac{1}{n^2} \\ \frac{3}{4}\sum_{n=1}^{\infty} \frac{1}{n^2} &= \frac{\pi^2}{8} \\ \sum_{n=1}^{\infty} \frac{1}{n^2} &= \frac{\pi^2}{6} \end{aligned}$$

Chapter 9

Saga 9: Mind boggling Set Theory Considerations

9.1 Introduction to the Cantor Set

9.1.1 Introduction

In this talk, we discussed quite a mind-blowing set known as the Cantor set. This set is defined in quite an odd way and challenges some preconceptions that one may have when thinking about sets and infinity in general. Before we begin our dive into the definition of the Cantor set, we are going to ponder something else- that is converting fractions into decimals in ternary.

9.1.2 On Ternary Decimals

In this section we think about how to convert fractions to decimals, but in base 3. To do this we do a very similar version of the long division algorithm, but we just consider everything in terms of powers of 3 instead of 10. Let us consider, for example, the fraction $\frac{1}{4}$. To figure out the decimal expansion, we multiply our fraction by 3 and see what the integer part is. In this case $3 \times \frac{1}{4} = \frac{3}{4}$ which has integer part zero. Then, repeat: $\frac{3}{4} \times 3 = \frac{9}{4}$ which has integer part 2 with a remainder of $\frac{1}{4}$. Then we multiply the remainder by 3, but this is back to where we started so this process goes on forever giving us the decimal expansion:

$$\frac{1}{4} = 0.0202020\ldots_3$$

(with the zero coming from the fact that originally when we times by 3 the integer part is zero and the 2 coming from the fact that when we times by 3 again the integer part is 2). For another example, note that $\frac{1}{3} = 0.1_3$ because when we multiply $\frac{1}{3}$ by 3 we get 1 with 0 remainder and so we stop there. Similarly $\frac{1}{27} = 0.01_3$.

9.1.3 Defining the Cantor Set

With our pondering on ternary decimals complete, we are now equipped to enter into the realm of the Cantor set. We begin by considering the number line between 0 and 1. Then, we remove all the numbers in the middle third (that is, between $\frac{1}{3}$ and $\frac{2}{3}$). Then, in the remaining two pieces of the number line we remove the middle third of those two lines. Then we are left with four different lines and we get rid of the middle third of those. We keep doing this infinitely and the points that are left form the Cantor set.

Figure 9.1: The Cantor set after seven iterations

The first question we can ask is, what proportion of the number line is left after we do this. To figure this out we can simply add up all of the lengths of the segments that we have removed:

$$\begin{aligned}&\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\dots\\&=\sum_{n=0}^{\infty}\frac{2^n}{3^{n+1}}\\&=\frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n\\&=\frac{1}{3}\times\frac{1}{1-\frac{2}{3}}\\&=1.\end{aligned}$$

But this is the same length as the number line from 0 to 1 itself. Therefore we can say that the Cantor set makes up 0% of the number line from 0 to 1. However, we can't conclude that the Cantor set contains no elements, because it is clear that the endpoints stay each time we remove an interval, for example $\frac{1}{3}$ and $\frac{2}{3}$ will never be removed. Indeed, since we do this removal process infinitely many times, we can actually say that the Cantor set is infinite, even though it is 0% of the number line! But the next question is: are there numbers in the Cantor set that aren't just the endpoints of the intervals that we remove? This question is what brings back our ternary decimals conversation back. If we look at what numbers are removed after the first removal, it is the numbers between $\frac{1}{3}$ and $\frac{2}{3}$, which in ternary form means the numbers between 0.1_3 and 0.2_3, that is, all numbers of the form $0.1xxxx\dots$. Then next we remove numbers between $\frac{1}{9}$ and $\frac{2}{9}$ and numbers between $\frac{7}{9}$ and $\frac{8}{9}$. Putting this into ternary form these are the numbers between 0.01_3 and 0.02_3 and the numbers between 0.21_3 and 0.22_3. Those are the numbers of the form $0.01xxx\dots$ and $0.21xxx\dots$. Continuing this process forever simply removes all numbers which have a 1 in their ternary expansion. Therefore numbers like $\frac{1}{4}$ are also in the Cantor set.

9.1.4 A Set Theory Aside on Size

To wrap up this discussion on the Cantor set we will finish with an incredible fact. However, to set up this fact we need to first look at some set theory definitions.

Definition 9.1.1. Given two sets X and Y, a function $f : X \to Y$ is said to be an injection if, for all $x_1, x_2 \in X$ we have that

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$

For example the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$ is an injection because if $f(x) = f(y)$ then $2x = 2y$ which necessarily means that $x = y$. However the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not an injection because if $f(x) = f(y)$ then $x^2 = y^2$ which does not imply that $x = y$ (since it could be the case that $x = -y$). If we can find a function between two sets $f : X \to Y$ that is an injection, then it means that $|X| \leq |Y|$ (meaning that there are fewer (or the same amount of) elements in X than there are in Y). To see this, imagine that X has more elements than Y, then a function between the two would look like this:

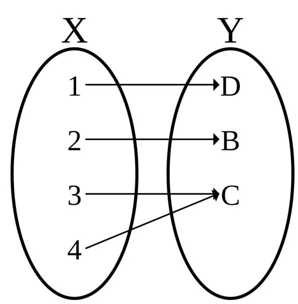


Figure 9.2: Attempting to find an injection $X \to Y$ where $|X| > |Y|$

We can see here that if we didn't have the fourth element in X then our map would be an injection. However, because X has an extra element, we have nowhere to map it to that would preserve the injectivity of the function. Next, we look at surjectivity.

Definition 9.1.2. A function $f : X \to Y$ is said to be surjective if for all $y \in Y$ there exists an $x \in X$ such that:

$$f(x) = y.$$

If we can construct a surjection $f : X \to Y$ then this implies that $|X| \geq |Y|$. To see this, imagine that X has less elements than Y.

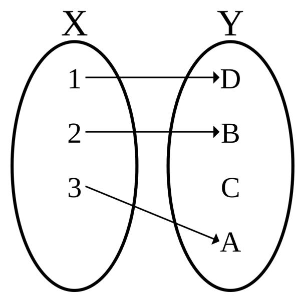


Figure 9.3: Attempting to find an surjection $X \to Y$ where $|X| < |Y|$

Here we can see that our function is not an injection because there is no element of X so that $f(x) = C$. However, if we removed C from Y to give the sets the same amount of elements, suddenly our function would be a surjection. Notice also that this map is an injection and that our failed injection in 9.2 is a surjection. To conclude we can say the following:

1. If there is an injection $f : X \to Y$ then $|X| \leq |Y|$
2. If there is a surjection $f : X \to Y$ then $|X| \geq |Y|$

9.1.5 A Mind-blowing Fact About the Cantor Set

Having considered how we can use set theory to precisely define what it means for one set to be "bigger than" another set, we now claim that the Cantor set has the same amount of elements that the line from 0 to 1 has, despite being 0% of it! Note that this is different to saying that the Cantor set has infinitely many elements, because there are infinitely many integers but there are actually less integers than there are elements of the line between 0 and 1.[1] In order to show this we have to construct a surjection $f : \mathcal{C} \to (0, 1)$, because if we can show that the Cantor set is either bigger than or equal to the line, then we know that it must be the same size as it is a subset so it can't be bigger. To construct this map, remember that the Cantor set is the set of all decimals in ternary form which contain only 0s and 2s. However every number in the line between 0 and 1 can be written in binary form as a decimal with just 0s and 1s. Thus we define our function as follows:

{Write number as ternary decimal} → {Replace 2s with 1s} → {Consider as binary decimal}

[1] For an elaboration on this, look up Cantor's diagonal argument.

We can see that this is a surjection because any number between 0 and 1 can be written as a binary decimal, for instance

$$\frac{1}{8} = 0.001_2$$

Since $\frac{2}{27} = 0.002_3$, we have that

$$f\left(\frac{2}{27}\right) = \frac{1}{8}.$$

In a similar way, any number between 0 and 1 can be written as $f(c)$ for some c in the Cantor set. Therefore we have constructed a surjection between $\mathcal{C}$ and $(0,1)$ meaning that

$$|\mathcal{C}| \geq |(0,1)|.$$

But since the Cantor set is a subset of the line between 0 and 1, it cannot be of a greater cardinality (fancy set theory term for size) and so the Cantor set has the same number of elements as the line between 0 and 1, despite making up 0% of the line! Pretty nifty if I do say so myself.

Chapter 10

Miscellaneous Talks

10.1 On Hyperbolic Trig Functions

10.1.1 Introduction

In A level further maths we learn about the hyperbolic trig functions, but we dive straight into the definition of them before looking at where they come from (understandable, given that we have to get through the content, but still unfortunate). In this talk I endeavoured to give an insight into where they came from, and how we can interpret these functions in a similar way to trig functions. In this talk I look at two different interpretations:

1. The analytic side
2. Comparison to regular trig functions

10.1.2 The Analytic Interpretation

The first way in which we can look at the hyperbolic functions are as the odd and even parts of the exponential function. Given a function $f(x)$, we say that it is even if $f(-x) = f(x)$ (for example $\cos(x)$ or x^2). In contrast, we say $f(x)$ is odd if $f(-x) = -f(x)$. It transpires that we can form the "odd" and "even" parts of any function as follows:

$$f_{\text{even}}(x) = \frac{f(x) + f(-x)}{2}$$

$$f_{\text{odd}}(x) = \frac{f(x) - f(-x)}{2}$$

Notice that

$$f_{\text{even}}(-x) = \frac{f(-x) + f(x)}{2} = f_{\text{even}}(x)$$

and

$$f_{\text{odd}}(-x) = \frac{f(-x) - f(x)}{2} = -f_{\text{odd}}(x)$$

as we were originally desiring. We also have $f_{\text{even}}(x) + f_{\text{odd}}(x) = f(x)$. Therefore we can just say that cosh and sinh are just the even and odd parts of the exponential function respectively, and we study them because the exponential function is interesting, so they could also have some cool properties (for example instead of just differentiating to themselves like e^x differentiates to itself they differentiate to each other). Furthermore if we look back at the Taylor series of e^x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

we can get the Taylor series of sinh and cosh for free by just taking the odd and even parts separately:

$$\sinh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots$$

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots$$

10.1.3 The Complex Interpretation

Now the analytic argument may provide an explanation of why the definitions of the hyperbolic trig functions are potentially interesting, but it still doesn't explain why they share the same names as our original trig functions. Well, through Euler's identity, it turns out that the two concepts are very much linked. More precisely, let us write:

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}.$$

Then if we plug in ix for x, we obtain:

$$\cos(ix) = \frac{e^{-x} + e^{x}}{2} = \cosh(x).$$

Similarly if we say

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$$

Then we have

$$i\sin(ix) = \frac{e^{x} - e^{-x}}{2} = \sinh(x).$$

Therefore we can say:

$$\cos(ix) = \cosh(x)$$
$$i\sin(ix) = \sinh(x).$$

This explains why the hyperbolic trig functions have such similar properties to the original trig functions. Furthermore it also explains the rule that we learn at A level that states that the hyperbolic trig identities are the same, except if you see a $\sin^2$ (explicitly or implicitly, for example seeing a $\tan^2$ also counts) then you have to change the sign. For example we still have identities like:

$$\sinh(2x) = 2\sinh(x)\cosh(x)$$
$$\cosh(2x) = 2\cosh^2(x) - 1.$$

But now we have identities such as:

$$\cosh^2(x) - \sinh^2(x) = 1$$
$$\cosh(2x) = 1 + 2\sinh^2(x).$$

10.1.4 Problem of the Week

Question 10.1.1. Evaluate

$$\int_{-1}^{1} x\sin(x)\sinh(\sqrt{1-x^2})\mathrm{d}x.$$

Solution 10.1.1. We begin by employing the substitution $\cos\theta = x$ in order to get rid of that ugly square root. Doing this gives us that $-\sin\theta \mathrm{d}\theta = \mathrm{d}x$ and our new bounds become π and 0 so we get:

$$-\int_\pi^0 \cos\theta \sin\theta \sin(\cos\theta)\sinh(\sin\theta)\mathrm{d}\theta = \int_0^\pi \cos\theta\sin\theta\sin(\cos\theta)\sinh(\sin\theta)\mathrm{d}\theta$$

Using the double angle formula for sin we rewrite this as:

$$\frac{1}{2}\int_0^\pi \sin(2x)\sin(\cos\theta)\sinh(\sin\theta)\mathrm{d}\theta.$$

Now we need to use an absolute stroke of brilliance to notice the following:

$$\sum_{n=1}^\infty \frac{(-1)^{n-1}\sin(2nx)}{(2n)!} = \mathrm{Im}\left(\sum_{n=1}^\infty \frac{(-1)^{n-1}(e^{ix})^{2n}}{(2n)!}\right)$$

$$= \mathrm{Im}\big(-\cos(e^{ix}) - 1\big)$$

$$= -\mathrm{Im}\left(\cos(e^{ix})\right)$$

$$= -\frac{1}{2}\mathrm{Im}\left(e^{ie^{ix}} + e^{-ie^{ix}}\right)$$

$$= -\frac{1}{2}\mathrm{Im}\left(e^{i\cos(x)}e^{-\sin(x)} + e^{-i\cos(x)}e^{\sin(x)}\right)$$

$$= -\frac{1}{2}\mathrm{Im}\left(e^{-\sin(x)}(\cos(\cos(x)) + i\sin(\cos(x))) + e^{\sin(x)}(\cos(\cos(x)) - i\sin(\cos(x)))\right)$$

$$= -\frac{1}{2}(e^{-\sin(x)}\sin(\cos(x)) - e^{\sin(x)}\sin(\cos(x)))$$

$$= \sin(\cos(x))\frac{e^{\sin(x)} - e^{-\sin(x)}}{2}$$

$$= \sin(\cos(x))\sinh(\sin(x))$$

which was exactly what was in our integral! Therefore we need to evaluate

$$\frac{1}{2}\sum_{n=1}^\infty \frac{(-1)^{n-1}}{(2n)!}\int_0^\pi \sin(2nx)\sin(2x)\mathrm{d}x.$$

The only time this integral is not zero is when $n = 1$, when the integral evaluates to $\frac{\pi}{2}$ (see the Fourier series talk for an explanation of this). Therefore the integral is:

$$\frac{1}{2}\left(\frac{1}{2}\times\frac{\pi}{2}\right) = \frac{\pi}{8}$$

Thank you to Wren for this problem.

Part III

Guest Talks

Chapter 11

Year 1

11.1 The Riemann Zeta Function, or, How to Win £1000000: Wren Shakespeare

11.1.1 On the Real Line

We begin by considering two related problems from the earlier days of maths. The first is the harmonic series:

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \dots$$

This series was proved to equal ∞ by Nicole Oresme in 1350. The second is the Basel problem:

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} \dots$$

This series has a finite value that is now quite well-known:

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} \dots = \frac{\pi^2}{6}$$

This was proven, if very unrigorously, by Leonhard Euler, who is also credited with the first use of the modern-day Riemann Zeta function, which is a generalisation from both of these ideas:

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

This infinite series only produces a finite value in a traditional sense for $s > 1$, as can be seen by comparison with the harmonic series - $s < 1$ means all the terms will be bigger than the harmonic series, so its value will be greater than the harmonic series, which is infinite. Euler also proved, again slightly unrigorously in some cases, two interesting properties of the zeta function:

$$\zeta(2n) = k\pi^{2n}, k \in \mathbb{Q}, n \in \mathbb{N}$$
$$\zeta(s) = \frac{1}{1-2^{-s}} \times \frac{1}{1-3^{-s}} \times \frac{1}{1-5^{-s}} \times \frac{1}{1-7^{-s}} \ldots, s > 1$$

By contrast, very little is known about the zeta function on odd numbers - the most we know is that $\zeta(3)$ is irrational.

11.1.2 Complex Numbers

The zeta function was most famously considered as a complex function by Bernhard Riemann, which is why it bears his name today. It can be extended to a complex function for $\mathrm{Re}(s) > 1$ very easily. However, what if we want to define it for other numbers?

$$\text{Taking } \eta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} :$$
$$\text{We can show that } \zeta(s) = \frac{\eta(s)}{1-2^{1-s}}, \mathrm{Re}(s) > 1$$

However, η also converges for $0 \leq \mathrm{Re}(s) \leq 1$, giving us a handy way to extend the zeta function to this region, which is exactly the region we want to study.
The Riemann Hypothesis asserts that if $\zeta(s) = 0$, then $\mathrm{Re}(s) = \frac{1}{2}$. Initially discarded as a mere curiousity by Riemann himself, it became increasingly apparent that RH implies a number of other things in various fields of maths. Notably, it implies a fairly regular distribution of prime numbers.

So, how far have we gotten with proving it? We can prove there are infinitely many zeroes of the form $\mathrm{Re}(s) = \frac{1}{2}$. We also know, if there are zeroes with other real parts, they cannnont be too close to 0 or 1 (how close depends on their exact location). The strongest result that we currently have is that 40% of the zeroes are of the form $\mathrm{Re}(s) = \frac{1}{2}$. Interestingly, the monetary prize is not for a resolution of the Riemann Hypothesis one way or another, it's for a proof (so if it was false, whoever showed that would not be able to collect the prize).

11.2 Introduction to Taylor Series: Prasan Patel

11.2.1 Introduction

In this talk we were introduced to the concept of a Taylor series, which was named after English mathematician Brook Taylor who was born in 1685 here in Edmonton! So we were blessed to see the magic of Taylor series delivered in its hometown. The idea behind Taylor series is to convert functions that are not polynomials into "infinite polynomials", so that they are easier to compute; a calculator doesn't actually know what e^2 is- it is just plugging in values into the Taylor series to approximate it.

11.2.2 Example: e^x

Our goal is to write e^x in the form $p(x) = c_0 + c_1 x + c_2 x^2 + \dots$. To do this we shall look at the derivatives. It is well known that $\frac{d}{dx} e^x = e^x$ and so when $x = 0$, the derivative will be 1, no matter how many times we take the derivative. Thus we need:

$$\begin{aligned} p(0) &= 1 \\ p'(0) &= 1 \\ p''(0) &= 1 \\ &\vdots \end{aligned}$$

Plugging this information in gives us that $c_0 = 1$, since all of the rest of the terms vanish when plugging in 0 into the polynomial. Furthermore $p'(x) = c_1 + 2c_2 x + 3c_3 x^2 + \dots$ so plugging in 0 into this and setting it equal to one leaves us with $c_1 = 1$. Taking the second derivative leaves us with $2c_2 + 6c_3 x + 12c_4 x^2 + \dots$ and so, again, when $x = 0$ we are left with just the c_2 term and so we have $2c_2 = 1 \implies c_2 = \frac{1}{2}$. In general, we see that because of the power rule and the fact that when $x = 0$ all higher terms cancel, we have that

$$p^{(n)}(0) = n!c_n$$

and so since we want $p^{(n)}(0) = 1$, we have that $c_n = \frac{1}{n!}$ and so our function actually looks like the following:

$$\begin{aligned} e^x &= \sum_{n=0}^{\infty} c_n x^n \\ &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \end{aligned}$$

We observe on desmos how well this can approximate e^x. The first three terms are $1 + x + \frac{x^2}{2}$, the first three terms are $1 + x + \frac{x^2}{2} + \frac{x^3}{6}$.

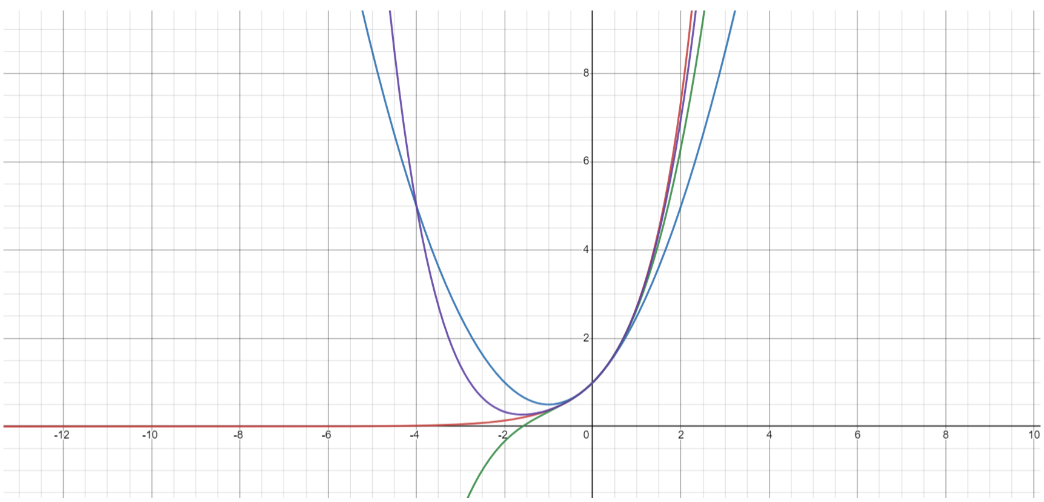


Figure 11.1: The approximations get closer and closer! (Red curve is e^x)

The purple (really close) line is the first 5 terms of the polynomial, the green (still pretty close) is the first 4 terms and so on. So really your calculator is just adding up the first 10-15 terms of this polynomial when you plug in e^x for some x in your calculator. Pretty neat, eh.

11.2.3 Generalising This Process

Now, considering that this is maths society, we want a general formula to do this process with any function f (which isn't a polynomial). Again, let us say that we want to write it in the form $p(x) = \sum_{n=0}^{\infty} c_n x^n$. Then again we would take derivatives and set them equal to the derivative of the original function, except we need not take the derivative at 0, we can pick any point a (for example if we were to do this process with $\ln(x)$, we couldn't pick $x = 0$ to take our derivative on since it isn't defined there). However, like in the e^x example, we want that when we take the nth derivative at a, all of the higher terms cancel out so instead we shall write:

$$p(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots = \sum_{n=0}^{\infty} c_n(x-a)^n.$$

Now, just like last time we take the nth derivative of $p(x)$ at a and we find that:

$$p^{(n)}(a) = n!c_n$$

since we've set $p^{(n)}(a) = f^{(n)}(a)$, we find that $c_n = \frac{p^n(a)}{n!}$ and so we have:

$$f(x) = \sum_{n=0}^{\infty} \frac{p^{(n)}(a)}{n!}(x-a)^n$$

which is called the Taylor series of f.

11.3 Differentiation under the integral sign: Wren Shakespeare

Here we consider a genius method for integrating functions which have no elementary anti derivative. This method, popularised by Feynman, goes as follows:

1. Introduce a new variable inside the integral a and call the integral $I(a)$. We want there to be a value of a for which we know what $I(a)$ is, and also we want there to be some a such that $I(a)$ gives us our original integral.

2. Then, differentiate with respect to a, hopefully making the integral now easier to solve.

3. Once you have found $I'(a)$, integrate that to find out $I(a)$ and then plug in your value of a that gives you back the original integral.

We shall, of course, consider some examples.

$\int_0^\infty \frac{\sin(x)}{x}\mathrm{d}x$

Example 11.3.1.

$$\int_0^\infty \frac{\sin(x)}{x}\mathrm{d}x.$$

The thing in this integral that's really annoying us is the x on the bottom of that fraction. Therefore, we shall introduce a new variable as follows:

$$I(a) = \int_0^\infty e^{-ax}\frac{\sin(x)}{x}\mathrm{d}x.$$

Note that $I(0)$ is our original integral. Now, we shall differentiate both sides with respect to a:

$$\begin{aligned} I'(a) &= \int_0^\infty \frac{\partial}{\partial a}e^{-ax}\frac{\sin(x)}{x}\mathrm{d}x \\ &= \int_0^\infty e^{-ax}\frac{-x\sin(x)}{x}\mathrm{d}x \\ &= -\int_0^\infty e^{-ax}\sin(x)\mathrm{d}x. \end{aligned}$$

We shall use the fact that $\sin(x) = \frac{e^{ix}-e^{-ix}}{2i}$ to make this integral easier, although it can also be done with integration by parts.

$$\begin{aligned} I'(a) &= -\frac{1}{2i}\int_0^\infty e^{-ax}(e^{ix} - e^{-ix})\mathrm{d}x \\ &= \frac{i}{2}\int_0^\infty e^{(i-a)x} - e^{-(a+i)x}\mathrm{d}x \\ &= \frac{i}{2}\left[\frac{1}{i-a}e^{(i-a)x} + \frac{1}{a+i}e^{-(a+i)x}\right]_0^\infty \\ &= \frac{-i}{2}\left(\frac{1}{i-a} + \frac{1}{i+a}\right) \\ &= \frac{-i}{2}\left(\frac{i+a+i-a}{1-a^2}\right) \\ &= \frac{-i}{2}\left(\frac{2i}{-1-a^2}\right) \\ &= \frac{-1}{a^2+1}. \end{aligned}$$

Now that we have $I'(a) = \frac{-1}{a^2+1}$, we can use standard results from calculus to deduce that $I(a) = -\arctan(x) + c$. Since $I(a) \to 0$ as $a \to \infty$, we have: $0 = -\frac{\pi}{2} + c \implies c = \frac{\pi}{2}$. Therefore:

$$I(a) = -\arctan(x) + \frac{\pi}{2}$$

Thus:

$$\boxed{I(0) = \int_0^\infty \frac{\sin(x)}{x} = \frac{\pi}{2}}$$

$\int_0^1 \frac{x-1}{\ln(x)} \mathrm{d}x$

Example 11.3.2.

$$\int_0^1 \frac{x-1}{\ln(x)} \mathrm{d}x.$$

How do we get rid of the $\ln(x)$ on the denominator? Well we know that the derivative of a^x is $\ln(a)a^x$, so we shall utilise this to our full advantage!

$$\begin{aligned}
I(a) &= \int_0^1 \frac{x^a - 1}{\ln(x)} \mathrm{d}x \\
I'(a) &= \int_0^1 \frac{\partial}{\partial a} \frac{x^a - 1}{\ln(x)} \mathrm{d}x \\
&= \int_0^1 \frac{\ln(x) x^a}{\ln(x)} \mathrm{d}x \\
&= \int_0^1 x^a \mathrm{d}x \\
&= \left[\frac{1}{a+1} x^{a+1} \right]_0^1 \\
&= \frac{1}{a+1}.
\end{aligned}$$

Success! We have found a nice form for $I'(a)$ and we can easily deduce that $I(a) = \ln(a+1) + c$. Since we also know that $I(0) = 0$, we can also say that $0 = \ln(1) + c$ and so $c = 0$. Therefore we have $I(a) = \ln(a+1)$ and so:

$$\boxed{I(1) = \int_0^1 \frac{x-1}{\ln(x)} \mathrm{d}x = \ln(2)}$$

11.4 Introduction to Fourier Transforms: Alex Dowling

11.4.1 Introduction

In this talk, we are going to look at Fourier transforms. The Fourier transform finds its origin in studying frequencies of sound waves. Let us first consider two different sound waves $g(t)$ and $h(t)$. If you then add these two sound waves together they give you a new function where it is unclear to see what frequencies made up the new function. And as you add in more and more sound waves it just becomes less and less clear what the frequencies were. The idea of the Fourier transform is that we input a function $f(t)$ and it tells us what the frequencies were that made up the function.

11.4.2 How to Decompose Signals

If we look at just one sound wave, we can simply count the number of peaks that occur in one second, and that will tell us the frequency. Unfortunately, when we add more sound waves into our function, this approach will be unhelpful. Therefore we need a different method of looking at a wave and seeing what the frequency is. We begin by winding our function around a circle. When we wind our function around the circle, an important variable arises: how long should it take for our function to wrap around the circle? (a notion that we will call the winding frequency) For example, if we were considering a sound wave with a frequency of 3Hz and we chose our winding frequency to be 0.5 cycles per second (i.e. it takes 2 seconds to wrap all the way around), we would end up with a circle that has 6 "petal" shapes around it.

11.4.3 The Mathematical Formalism

Now this is all well and good, but how are we going to make this mathematically precise? Well firstly, recall that the formula that gives us a circle in the complex plane is e^{ix}, where x tells us how far along the circle we are, in radians. If we want to incorporate winding frequency into the mix our equation would become something like:

$$e^{2\pi i f t}$$

where the 2π is inserted because that represents going around the circle once and t is time, meaning that f tells us precisely how many cycles are done per second. Now if we scale this by our sound wave $g(t)$ (and adhere to the convention that we are going clockwise instead of anticlockwise) we obtain:

$$g(t)e^{-2\pi i f t}$$

This formula gives us our sound wave g at time t and its point on the circle. In this way we have gotten our formula for "winding a function around the circle" to be mathematically precise. Now the final step is to get a formula for the "centre of mass". To do this we can just take the average of our points along the circle:

$$\frac{1}{N}\sum_{n=1}^{N} g(t_n)e^{-2\pi i f t_n}.$$

As we make this more and more accurate by adding in more and more points to our sum, this becomes an integral over our time period from t_1 to t_2:

$$\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} g(t)e^{-2\pi i f t}\mathrm{d}t.$$

The Fourier transform scales up by the time period and so we obtain our Fourier transform equation

$$\int_{-\infty}^{\infty} g(t)e^{-2\pi f t}\mathrm{d}t$$

[THIS TALK HAD MORE DETAIL AND DEPTH WHICH HAS BEEN LOST TO THE SANDS OF TIME, CHECK OUT 3B1B ON YOUTUBE TO FIND A VIDEO THAT SHOULD EXPLAIN EVERYTHING IN THESE NOTES AND MORE!]

11.5 The logistic map: Rowan Oliveira

11.5.1 Introduction

In this talk we look at a fascinating iterative map known as the logistic map, given by:

$$x_{n+1} = r x_n (1 - x_n),$$

where we take x_n to be between 0 and 1. This map originally finds its applications in biology in the paper "Simple mathematical models with very complicated dynamics" by Robert M. May, due to the fact that for varying values of r it can capture different types of phenomena.

11.5.2 Studying what happens as r varies and the Feigenbaum Constant

When r is between 0 and 1, the iterative sequence will always just die off and get closer to zero, for example if we have $r = 0.5$ and we pick our starting point to be $x_0 = 0.25$, then our sequence becomes:

$$0.25 \to 0.5(0.25)(0.75) = 0.09375 \to 0.04248 \to \dots$$

Now, when r is between 1 and 3, the sequence will converge to $\frac{r-1}{r}$, but the map sequence will fluctuate more when r is between 2 and 3. Now, things get interesting: when r is between 3 and $1+\sqrt{6}$, the sequence will eventually start bouncing between two separate points, which are given by some complicated surd which isn't important. Then when r is between $1+\sqrt{6}$ and approximately 3.54409, the sequence will bounce between four different points. Now this pattern continues, with the number of points doubling very rapidly until we reach $r \approx 3.56995$, when we have most points above this point becoming completely chaotic. However, the beauty is that there still exist islands of stability above 3.56995 where the r value gives us stable behaviour. We can see this in a bifurcation diagram which nicely shows exactly what's going on.

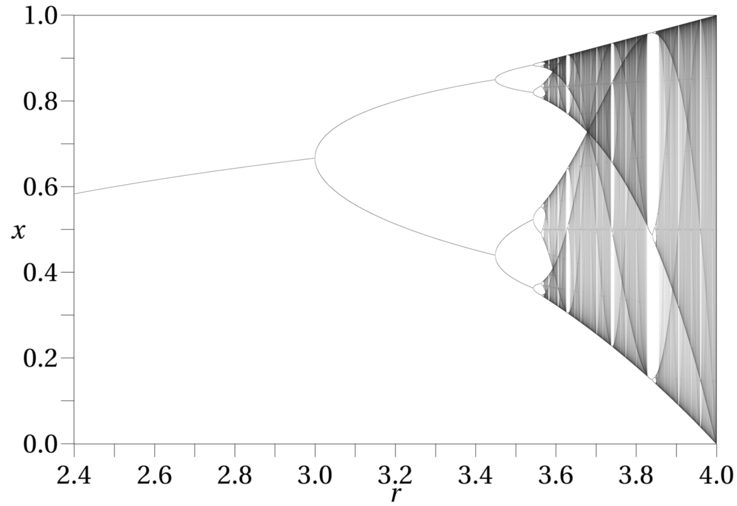


Figure 11.2: The little white strips in the midst of the chaos are the islands of stability.

Furthermore in the area before the chaos, the ratio of the distances between each new r value that doubles the amount of points that the sequence converges to approaches a number δ which is known as Feigenbaum's constant, which is given by $4.669201609\ldots$. We can see this with a table:

n	period	r_n	$\delta \to \frac{r_n - r_{n-1}}{r_{n-1} - r_{n-2}}$
1	2	3	
2	4	$1+\sqrt{6}$	
3	8	3.5440903	4.7514
4	16	3.5644073	4.6562
5	32	3.5687594	4.6683
6	64	3.5696916	4.6686

11.5.3 Connection with the Mandelbrot Set

It turns out that (this is hardly even surprising anymore!) the Mandelbrot set is connected with the logistic map as well. Indeed, it turns out that the "islands of stability" in the logistic map are deeply connected to the Mandelbrot set. Recall that the equation that the Mandelbrot concerns itself with is:

$$z_{n+1} = z_n^2 + c.$$

When we make the change of variables $z = r\left(\frac{1}{2} - x\right)$, $c = \frac{r}{2}\left(1 - \frac{r}{2}\right)$, we can observe a stunning connection between the real parts of the Mandelbrot and the logistic map.

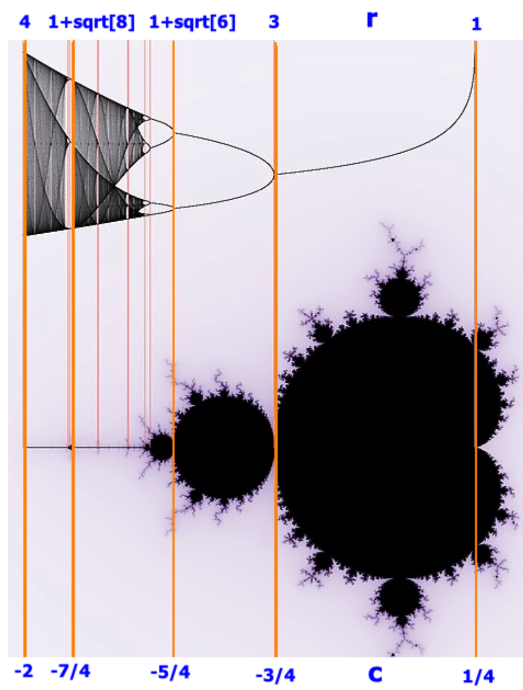


Figure 11.3: The islands of stability line up perfectly with the Mandelbrot set, under the change of variables

This actually makes sense when you think about it, because the Mandelbrot set measures exactly where the iterative sequence is stable and both the logistic map and the Mandelbrot set are given by quadratics, so it is feasible that they should have some connection under a change of variables. That being said, it doesn't make the result any less remarkable!

11.6 Derivatives and Integrals of Displacement: Saik Rudad

11.6.1 Introduction

In this talk we looked at the various concepts in physics that arise when you take repeated integrals and derivatives of displacement with respect to time.

11.6.2 The Derivatives

Of course, when we take the first derivative of displacement $\frac{ds}{dt}$, we get velocity v in ms^{-1}, which measures how fast (and the direction) something is going. Next, we take the derivative of velocity to obtain acceleration, a in ms^{-2}, which measures the rate of change of velocity. Now here things get interesting: when we take the third derivative of displacement is called jerk, j, measured in ms^{-3}. Jerk has many applications in engineering, for example it is considered when designing lifts and cars. This is because one must limit not only the maximum force, but the rate at which the force changes. Since $F = ma$, this means that we not only need to look at acceleration, but also the rate of change of acceleration- that "jerky" feeling- which engineers spend a lot of time to try to minimise when designing different vehicles. The fourth derivative of displacement (equivalently, the derivative of jerk) is called snap. Snap is minimised in the design of railway tracks and is measured in ms^{-4}. The fifth and sixth derivatives of displacement are called crackle and pop respectively and are measured in ms^{-5} and ms^{-6} respectively.

11.6.3 The integral

We shall look at the integral of displacement $\int s\mathrm{d}t$, which is called absement and is measured in ms. What this measures is the sustained displacement of an object from its initial position. Absement has been used in studying fluid flow, for example in music instruments like the hydraulophone.

Chapter 12

Year 2

12.1 Taylor Series: Kathleen Suciu

12.1.1 Introduction

In this talk, we consider the notion of Taylor series, which is a method for approximating non-polynomial functions as polynomials, because they are easier to work with.

12.1.2 Approximating $\cos(x)$

As an example, let us consider the function $f(x) = \cos(x)$. Let's say we would like to approximate this as a quartic, and we'd like to approximate it very accurately for values of x which are close to zero. Then let us write:

$$\cos(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4.$$

Since we would like $\cos(x)$ and our quartic to have the same y-intercept, we shall say $c_0 = 1$. Now, taking the derivative we get:

$$-\sin(x) = c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3.$$

In order to find c_1, we shall just plug in $x = 0$ and we obtain $c_1 = -\sin(0) = 0$. Now we take the second derivative and find:

$$-\cos(x) = 2c_2 + 6c_3 x + 12c_4 x^2.$$

Again, plugging in $x = 0$, we find that $2c_2 = -1 \implies c_2 = \frac{-1}{2}$. Repeating this process again leaves us with $6c_3 = 0 \implies c_3 = 0$ and $24c_4 = 1 \implies c_4 = \frac{1}{24}$. Collecting these coefficients together gives us the following approximation for $\cos(x)$:

$$\cos(x) \approx 1 - \frac{x^2}{2} + \frac{x^4}{24}.$$

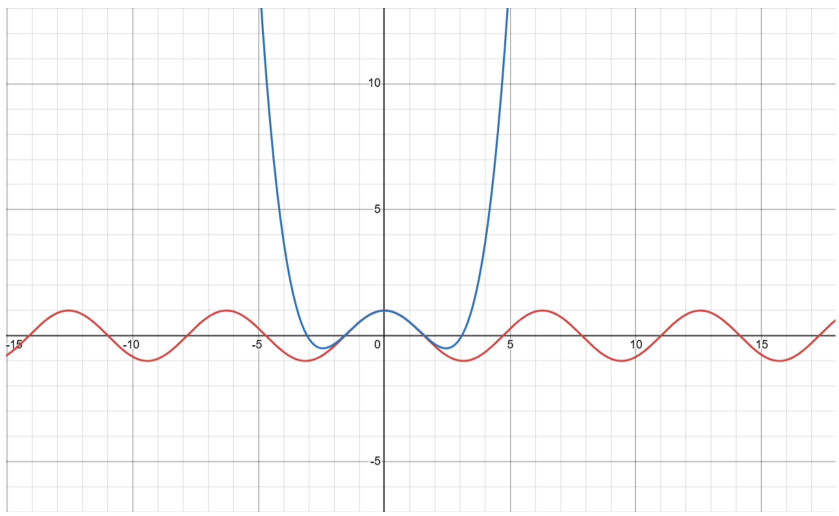


Figure 12.1: We can see that our polynomial (blue) is an excellent approximation for $\cos(x)$ (red) for small values of x.

Notice that if we add more terms to the polynomial and continue doing the process of taking the higher derivatives at zero and finding coefficients we can make our series expansion more and more accurate. If we do this forever, we get the following series expansion for $\cos(x)$:

$$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots$$

or, in neat summation terms, we may also say:

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

12.1.3 The Taylor Series

In general, we use something called the Taylor series to approximate functions as polynomials, for values near some c of our choosing. We set:

$$f(x) = c_0 + c_1(x-c) + c_2(x-c)^2 + c_3(x-c)^3 + c_4(x-c)^4 + \dots$$

Then we can say that $f(c) = c_0$. Taking derivatives, we have:

$$f'(x) = c_1 + 2c_2(x-c) + 3c_3(x-c)^2 + 4c_4(x-c)^3 + \dots$$

Plugging in $x = c$, we have that $c_1 = f'(c)$. Taking the second derivative yields:

$$f''(x) = 2c_2 + 6c_3(x-c) + 12c_4(x-c)^2 + \dots$$

Plugging in $x = c$ again gives us $c_2 = \frac{f''(c)}{2}$. Taking the third derivative gives us that $c_3 = \frac{f'''(c)}{6}$. In general, taking the nth derivative yields:

$$c_n = \frac{f^{(n)}(c)}{n!}$$

and so in general we have that the Taylor series expansion of a function $f(x)$ around c is:

$$f(x) = \sum_{n=0}^{\infty} c_n(x-c)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n.$$

Notice that what we did with $\cos(x)$ was the Taylor series around 0 (hence why it was a good approximation for small values of x). This specific kind of Taylor series when you expand around 0 is also called the McLaurin series. We shall expand e^x and $\sin(x)$ around 0. Since the derivative of e^x is e^x, then we have that for all n:

$$f^{(n)}(0) = e^0 = 1.$$

Therefore:

$$e^x = \sum_{n=0}^{\infty} \frac{f^(n)(0)}{n!}x^n = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Now we move onto the expansion of $\sin(x)$. Let us differentiate $\sin(x)$ a few times:

$$\begin{aligned}
f'(x) &= \cos(x) \\
f''(x) &= -\sin(x) \\
f'''(x) &= -\cos(x) \\
f^{(4)}(x) &= \sin(x).
\end{aligned}$$

Therefore the higher derivatives of sin loop around every fourth derivative. Evaluating these derivatives at 0 gives:

$$1, 0, -1, 0.$$

Therefore the expansion of $\sin(x)$ is:

$$\begin{aligned}
\sin(x) &= 0 + x + \frac{0x^2}{2!} - \frac{x^3}{3!} + \frac{0x^4}{4!} + \frac{x^5}{5!} + \dots \\
&= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots
\end{aligned}$$

12.1.4 Euler's Formula

We shall end this talk with a beautiful application of the work that we have just done.

Theorem 12.1.1. *Given any number θ, we have:*

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Proof. If we consider the Taylor series for e^x we obtain:

$$\begin{aligned}
e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \dots \\
&= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \dots \\
&= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots\right) \\
&= \cos(\theta) + i\sin(\theta)
\end{aligned}$$

where the last line comes from our Taylor series expansions of cos and sin. □

Plugging in the value $\theta = \pi$ gives us one of the most beautiful equations ever:

$$\begin{aligned} e^{i\pi} &= \cos(\pi) + i\sin(\pi) = -1 \\ \implies &\boxed{e^{i\pi} + 1 = 0} \end{aligned}$$

12.1.5 Problem of the Week

Question 12.1.1. By considering the Taylor series for $\sin(x)$ and the sum of roots formulae, show that:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Solution 12.1.1. We want to consider a function whose roots are the square numbers and then use the formula for the sum of the reciprocals of the roots to find our infinite sum. Since we are told to consider the Taylor series of $\sin(x)$, our first guess for the function that we should consider would be $\sin(\sqrt{x})$. This nearly works since the roots are indeed of the form $n^2\pi^2$, but we have a pesky 0 which is annoying us (since it will disallow us to take the sum of the reciprocals of the roots). The solution to this issue is instead to consider the function:

$$\frac{\sin(\sqrt{x})}{\sqrt{x}}.$$

Now this function has precisely the roots that we desire! The only order of business left is to Taylor expand it. Recalling the Taylor series for $\sin(x)$ and then just halving all the exponents (since we have a $x^{\frac{1}{2}}$ inside), we get:

$$\begin{aligned} \sin(\sqrt{x}) &= x^{\frac{1}{2}} - \frac{x^{\frac{3}{2}}}{3!} + \frac{x^{\frac{5}{2}}}{5!} - \dots \\ \frac{\sin(\sqrt{x})}{\sqrt{x}} &= 1 - \frac{x}{3!} + \frac{x^2}{5!} - \dots \end{aligned}$$

Now, since for a polynomial $a + bx + cx^2 + \dots$, it is true that the sum of the reciprocals of the roots are equal to $-\frac{b}{a}$, we apply this to see that:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{(n\pi)^2} &= -\left(\frac{-\frac{1}{3!}}{1}\right) \\
\sum_{n=1}^{\infty} \frac{1}{n^2\pi^2} &= \frac{1}{6} \\
\frac{1}{\pi^2}\sum_{n=1}^{\infty} \frac{1}{n^2} &= \frac{1}{6} \\
\sum_{n=1}^{\infty} \frac{1}{n^2} &= \frac{\pi^2}{6}
\end{aligned}$$

as required! Thus we have used Taylor series to solve the Basel problem! Thanks again to Wren for this neat problem and application of Taylor series.

12.2 Number Clump Spread: Vincent Karthaus

12.2.1 Introduction

When I received my timetable at the end of the summer holidays, I was blown away by the distribution of my German lessons. With one teacher, F (in purple in figure 12.2), we had all our lessons in four consecutive days out of the two-week rota. This included, to the dismay of the class and the teacher, a triple period that filled up the entire space between morning break and lunch.

I allocated my free periods into blocks (in a light grey) and then created a small table of lessons and free period blocks, labelling each with "number", "clump" and "spread". This was for the purpose of planning my time: the vision was that I'd look at the "number", "clump" and "spread" of my four blocks of frees and decide to do one thing (e.g. Physics revision) in one block and another (e.g. writing hand-wavey musings in Overleaf) in another block. Remember, this was at the beginning of the year when I could be bothered with such silliness as time planning

But this did give me something to think about as I sat bored in physics: is there a way to formalise the thoughts I had at the beginning of the year? This entire document is essentially an exploration of my various different hunches, some of which turned out to reveal some interesting results and some of which turned out to be completely made up.

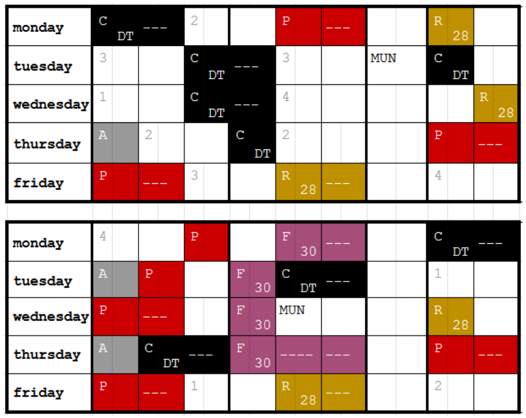


Figure 12.2: My 2024-25 timetable, colour-coded

12.2.2 Some Setup

We begin presently with some definitions:

A sequence $L = L_1, L_2, \ldots$ where $L_i \in \mathbb{L}$ for finite $\mathbb{L}$.

Then we define the "sequence of distances" of L:

$$\Delta(L, l) : l \in \mathbb{L} \text{ is such that } \Delta(L, l)_i = j - k \tag{12.1}$$

where $L_j = L_k (= l)$ are the ith and $(i-1)$th occurrences of $l \in L$

This gives that the number of differences is the same as the number of occurrences of l in L:

$$|\Delta(L, l)| = |L(l)| \tag{12.2}$$

The point of Δ is because we want to be working relatively. Whether one has German on Tuesday or Thursday is a useless fact unqualified: what is more interesting is if, for example, one hasn't had German since the Wednesday before. To put the (poorly notated) definition in English, the sequence $\Delta(L, l)$ is the sequence of differences between the places of two ls in L.

Of course the order of elements in L matters (it is a sequence, after all) but the order of elements in Δ doesn't matter for the purposes of this PDF. It is simply a sequence because Why Not. Hence, whether we choose L_k to be the $(i+1)$th or $(i-1)$th occurrence is irrelevant, to the point where it's changed twice in the writing of this document.

There is one reconciliation that must be made to make the facts given true and useful: although the timetable given in figure 12.2 is laid out starting in one corner and ending in the opposite, it is a cyclic pattern. If $i = 1$, then the $(i-1)$th corresponds to the final, and it is as if we count backwards off the beginning of the timetable and continue counting as we go back onto the end and to the last occurrence of l. Since this is essentially a division of the size of L by its ls, this gives us

$$\sum_{d \in \Delta(L,l)} d = |L| \tag{12.3}$$

The next definition we give is a simplification of $\Delta(L, l)$:

$$\Delta(L) = \bigcup_{l \in \mathbb{L}} \Delta(L, l) (:= \Delta) \tag{12.4}$$

$$\text{thus } |\Delta| = |L| \tag{12.5}$$

This could just as easily be defined from the same angle that was used in (12.1), but we don't have the effort to do that presently. The order is irrelevant for the arithmetic in the present work, but for the sake of examples we order them back into the order in which they were in L. The notation of dropping the argument of Δ is simply to make writing slightly quicker and less messy later on.

(12.5) is also a convenient extension of (12.2). Is it not lovely when maths is consistent?

Since this has all been rather abstract thus far, let us take a concrete example with a "mini-timetable" made up of a single five-period day that repeats eternally, to aid the reader's comprehension (see table 12.1).

L	$\Delta(L,\text{“F”})$	$\Delta(L,\text{“R”})$	$\Delta(L,\text{“C”})$	Δ
F	2			2
R		4		4
R		1		1
F	3			3
C			5	5

Table 12.1: small-scale examples of Δ

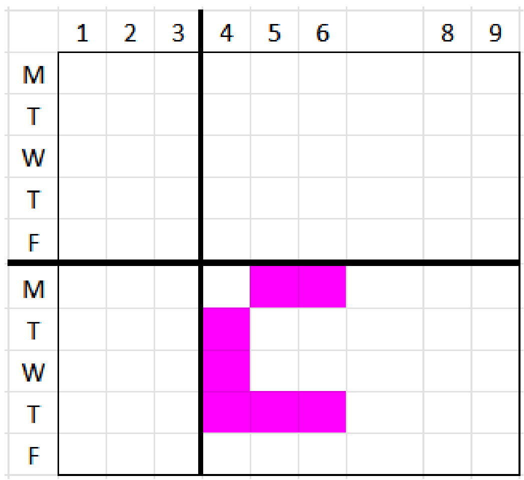


Figure 12.3: My 2024-25 timetable but with only teacher “F”

When it comes to applying these methods to a full-length timetable, there are certain decisions that have to be made one way or another. For the sake of usefulness in the present work, we say the following: when one day ends, the next begins immediately (that is to say, the difference between Period 8 and Period 9 on a Tuesday is the same as the difference between Period 9 on a Tuesday and Period 1 the following Wednesday); lunchtimes, break times and weekends are completely ignored (that is to say, the aforementioned difference is also the difference between Period 9 on a Friday and Period 1 on the following Monday, and between Period 6 and Period 8 on any same day). To some extent, we make these points for the sake of simplicity when it comes to interpreting the results later down the line. Now let us think about defining these three terms that I used no more than qualitatively at the beginning of the term, here in reference to my lessons with one of my German teachers, denoted with “F” (see figure 12.3). Beginning with “Number” because that is trivial to define, we quickly see that some units would help us. The Système international d’unités can’t help us much here: there are a lot of seconds between two lessons. All numbers henceon will instead be given in ”periods” - in reality, forty minutes, but in abstract, any length.

12.2.3 Working It Out

Number

The “Number” of periods with the teacher in question can be counted by an infant. It is 7. Since our units are well-defined (lessons certainly feel like they’re moving slowly enough not to worry about relativistic effects...), we count a “double-” or “triple-period” as 2 or 3 separate periods respectively, which is an attitude that will help us later. By (12.2),

$$N(\Delta(L,l)) = |\Delta(L,l)| = |L(l)| \tag{12.6}$$

Spread

It is no doubt that the lessons of the teacher highlighted in figure 12.3 are “poorly spread”. What do we mean when we say this phrase though? In this phrase, we’re referring to the large timeframes that have absolutely no lessons with this teacher: a frame of 55 periods, no less, which is over two-thirds of the entire timetable. A ”well-spread” distribution of lessons with this teacher would have these gaps minimised.

Thus, we want to perform some sort of iterative calculation on our sequence of gaps $\Delta(L,\text{“F”})$ that puts weight on the longer gaps. We could attempt a sum of Δ since the longer gaps would thusly account for a larger proportion of the sum. However, the reader who reads can read that the sum of Δ, as per (12.3) is just the total number of periods, $|L|$.

Instead, we want to put an extra weight on our longer gaps. Luckily, larger values have an in-built tool to weight them favourably: themselves. We can weigh the sum by the elements themselves, which is known to the common mathematician as “squaring” it. Thus, we’re dealing with some variation of

$$\sum_{d \in \Delta(L,l)} d^2 \tag{12.7}$$

However, the same readers who happen to be reading closely will realise that this is almost what we wanted: a poorly-spread distribution will have larger gaps, and thus this sum will be disproportionately *larger*. Thus, the more useful definition to take is a reciprocal of (12.7).

There is one final amendment to this definition: currently, a sequence of twenty lessons distributed evenly throughout the week gets an unfavourable score in comparison to a sequence of two lessons distributed similarly evenly, by simple virtue of there being more lessons. This is an explicable advantage (it's unfair to expect the latter sequence to populate the timetable as evenly as the former), and so a combatable one: we divide through by the number of lessons (which, by (12.2) is also the number of gaps). This finally gives way to our definition of "spread":

$$S(\Delta(L,l)) = \frac{1}{|\Delta|} \left(\sum_{d \in \Delta(L,l)} d^2 \right)^{-1} \tag{12.8}$$

Clump

Let us now whisk through the same process for the word "clump". This is in fact a poor choice of word as one's immediate thought may be that "clump" is the antithesis to "spread", as you can see that the lessons highlighted in figure 12.3 are "not very spread" and "very clumped". However, let us treat it as a separate phenomenon and instead we will find that a distribution of lessons can in fact be both "well-spread" and "clumped" or "poorly-spread" and "not clumped".

When we say that the lessons in figure 12.3 are "clumped", we are this time saying that the majority are in close proximity to one another. Specifically, we refer to the fact that most distances in Δ are very very small. In the case of the "triple-period" on Thursday, there are two elements of $\Delta(L, \text{"F"})$ that are just 1! (Interpret that as either surprise or factorial, because it is, by chance, unambiguous.) In a similar fashion to how we defined "spread", we use the reciprocal squares to place weight on these shorter values. This

L	$\Delta(L, \text{“F”})$	$\Delta(L, \text{“R”})$	$\Delta(L, \text{“C”})$	Δ
F	2			2
R		4		4
R		1		1
F	3			3
C			5	5

Table 12.2: table 12.1 repeated for your convenience

$l \in \mathbb{L}$	$N(\Delta(L,l))$	$\sum_{d\in\Delta(L,l)} d^2$	$S(\Delta(L,l))$	$\sum_{d\in\Delta(L,l)} d^{-2}$	$C(\Delta(L,l))$
F	2	13	$\frac{1}{26}$	$\frac{13}{36}$	$\frac{13}{72}$
R	2	17	$\frac{1}{34}$	$\frac{17}{16}$	$\frac{17}{32}$
C	1	25	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$

Table 12.3: small-scale examples of N, S and C

also means that we no longer have to deal with larger values mapping to “less clumped” distributions, as reciprocating our smaller distances gives us a larger “clump”. This gives us the following definition of “clump”:

$$C(\Delta(L,l)) = \frac{1}{|\Delta|} \sum_{d\in\Delta(L,l)} d^{-2} \tag{12.9}$$

12.2.4 Interpreting the Values

Relatively

Let us use the same “mini-timetable” as in table 12.1 to give some sample values for C and S, as it's easy to see where values come from on such a small scale: see table 12.3.

Perusing table 12.3 and comparing to what we would qualitatively say about the “mini-timetable” (reproduced in table 12.2), we can see that it's pretty spot-on: teacher “C” has the most spread-out lessons, but teacher “F” is very close behind, for example, while “R”'s “clumpedness” is definitely coming to light. There is a small failure of using a small sample size, in that the scaling suffers a little, and small discrepancies arisen by the number of lessons don't hide themselves. Ignoring that, we can see how this metric has potential.

(Somewhat) Absolutely

However, this is only good for comparing different lessons within one timetable. Since we have a Δ-sequence for the entire timetable, performing "NCS-analysis" on that is, in its current state, not very useful as we'd have nothing to compare it against, or we may find ourselves comparing apples to pears. We presently briefly restate our N, S and C functions slightly more neatly and using the Δ-sequence of the whole of L. It is now that we remind the reader that "Δ" unqualified represents this complete sequence, $\Delta(L)$.

$$C(\Delta) = \frac{1}{N(\Delta)} \sum_{d \in \Delta} d^{-2} \tag{12.10}$$

$$N(\Delta) = \sum_{d \in \Delta} d^{0} \tag{12.11}$$

$$S(\Delta) = \frac{1}{N(\Delta)} \Big(\sum_{d \in \Delta} d^{2} \Big)^{-1} \tag{12.12}$$

Unfortunately for the reader, we're unable to align the equations exactly as we'd like while still giving them different reference numbers. The reader will have to cope. However, this formatting does demonstrate the relationship between the three functions: if one ignores the reciprocating adjustment we to S, they're in a neat series. The reader with a particularly tenacious memory will remember that (12.3) fits into this progression, but this is just at a superficial level as (12.3) references $\Delta(L, l)$ rather than Δ. However, we will return to this insight later.

(12.11) can be rewritten as the following, and thus by (12.5) as

$$N(\Delta) = |\Delta| = |L|$$

Using the same order of thinking as with section 3, we begin with regularising values of S. S is formed from the sum of squares: if we take a step back, we have the average of squares,

$$\frac{1}{|\Delta|}\sum_{d\in\Delta} d^2 \tag{12.13}$$

The average of squares, which we henceon notate as $\hat{\mu}_S^2$ to highlight that it's weird, can be square-rooted to give a plain "average": we're simply undoing what we did to construct $\hat{\mu}_S^2$ to get an ordinary average of gaps but "skewed" in favour of longer gaps, as per the processes in section 3. This gives

$$\hat{\mu}_S = \sqrt{\frac{\sum_{d\in\Delta} d^2}{|\Delta|}} \tag{12.14}$$

And similarly,

$$\hat{\mu}_C = \sqrt{\frac{|\Delta|}{\sum_{d\in\Delta} d^{-2}}} \tag{12.15}$$

Simple dimensional analysis reveals that these retain the same units as the elements of Δ to begin with: thus, we can interpret the values of $\hat{\mu}_C$ and $\hat{\mu}_S$ to be the average number of periods between a lesson and the next time that lesson occurs, weighted in favour of clump and spread respectively.

For example, the "mini-timetable" in table 12.2 yields the following values:

$$\hat{\mu}_C = 1.848\ldots$$
$$\hat{\mu}_S = 3.317\ldots$$

It was roughly at this point in the present article's draft that a friend pointed out the similarity between the formula for "spread-weighted mean" $\hat{\mu}_S$ in (12.14) and the formula for "standard deviation" σ of a dataset. As far as we are concerned in the present work, this is purely coincidental. However, any further insight from the reader is welcomed.

(Somewhat) Better

The reader is forgiven if the end of the last subsection felt anticlimactic. The numbers and methods outlined therein were developed after the methods in the proceeding subsection herein, but were presented first due to slightly more elegance and slightly less complexity. We do not see much use in $\hat{\mu}_{S|C}$, hence why they were notated with a hat (such that we can justify ignoring them because they are ugly).

Instead, we take a step back to $\Delta(L, l)$ and attempt to use the same means to evaluate an equivalent of $\hat{\mu}_{S|C}$ for each l. Most of our thinking work has, however, been done in the preceding subsection, with just one major addition.

The perceptive reader with an English Literature qualification will recognise that the sum from (12.3) remains unused. Now that we are dealing with $\Delta(L, l)$ we can complete the pattern if not only because it is pleasing to typeset:

$$\sum_{d\in\Delta(L,l)} d^{-2} \quad \sum_{d\in\Delta(L,l)} d^{-1} \quad \sum_{d\in\Delta(L,l)} d^{0} \quad \sum_{d\in\Delta(L,l)} d^{1} \quad \sum_{d\in\Delta(L,l)} d^{2} \tag{12.16}$$

(12.16.-2) and (12.16.2) are used in the definitions for C and S respectively, while (12.16.0) is trivially $N(\Delta(L, l)) = |\Delta(L, l)|$. If anyone would like to come forward with a use-case or theoretical explanation of (12.16.-1), they would be more than welcome.

We use (12.16.1) to define the conventional "mean" of $\Delta(L, l)$. This is hopefully not a new idea to the reader, although we do denote it with a hat to continue the idea that it is the "imperfect mean" or the "mean" of $\Delta(L, l)$ for some l:

$$\hat{\mu} = \frac{\sum_{d\in\Delta(L,l)} d}{|\Delta(L, l)|} \tag{12.17}$$

This gives us something to compare $\hat{\mu}_{C|S}$ against. Rather than succumb ourselves to the "mini-timetable" again, this property is demonstrated with the timetable represented in figure 12.2: see figure 12.5.

A brief explanation of the layout of figure 12.5 may be in order. The rows are headed by the name of each block ("b" meaning "blank" or "unlabelled", and "a" going unrepresented due to there only being three "a"

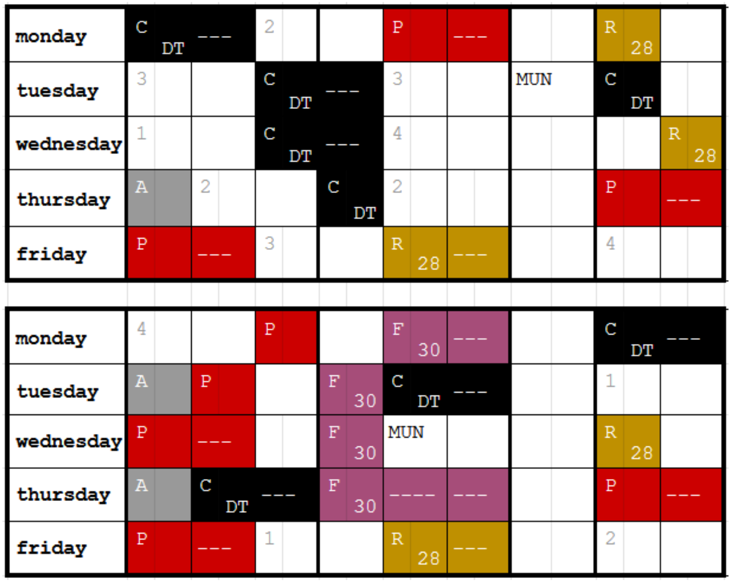


Figure 12.4: My 2024-25 timetable, repeated from figure 12.2 for your convenience

periods in the fortnight), corresponding (almost) exhaustively to the values of $l \in \mathbb{L}$. Columns B through F correspond to the five expressions in (12.16) respectively, while columns G, H and I correspond to $C(\Delta(L,l))$, $N(\Delta(L,l))$ and $S(\Delta(L,l))$ respectively. Columns J, K and L are, as one would expect, $\hat{\mu}_C$, $\hat{\mu}$ and $\hat{\mu}_S$, despite the lack of hat in the column heading (which proved confusing later down the line). Columns G, I, J, K and L have each been rounded to a number of decimal places for brevity and legibility.

Some numbers suddenly make sense when looking at a table. For example,

$$\forall \mathfrak{L} \in \mathbb{L}, \sum_{d \in \Delta(L,\mathfrak{L})} d^1 = |\Delta| = \sum_{l \in \mathbb{L}} N(\Delta(L,l)) \tag{12.18}$$

We can also begin making observations similar to those we made looking at the sample tables 12.2 and 12.3. For example, all blocks appear mostly in "double-periods", with the exception of "b", which appears (by its nature) mostly in "singles": this is reflected in its $\hat{\mu}_C$ being closer to its $\hat{\mu}$ than any other l. The infamously poorly-spaced-out "f" lessons had the highest $\hat{\mu}_S/\hat{\mu}$ ratio and the lowest $\hat{\mu}_C/\hat{\mu}$ ratio of any lesson blocks, even though the latter was by a smaller margin. This last fact is down to the fact that although

	A	B	C	D	E	F	G	H	I	J	K	L
1		-2	-1	0	1	2	C	N	S	μ_c	μ	μ_s
2	c	6.25	7.22	14	80	902	0.45	14	0.000079	1.496	5.714	8.027
3	p	8.06	8.60	14	80	1102	0.58	14	0.000065	1.317	5.714	8.872
4	r	2.02	2.35	7	80	1362	0.29	7	0.000105	1.858	11.429	13.949
5	f	3.05	3.43	7	80	3192	0.44	7	0.000045	1.513	11.429	21.354
6	b	1.57	2.62	9	80	1252	0.18	9	0.000089	2.39	8.889	11.795
7	1	3.00	3.12	6	80	2174	0.5	6	0.000077	1.413	13.333	19.035
8	2	4.36	4.89	8	80	2902	0.55	8	0.000043	1.354	10	19.046
9	3	3.11	3.39	6	80	3262	0.52	6	0.000051	1.388	13.333	23.317
10	4	4.00	4.07	6	80	3774	0.67	6	0.000044	1.224	13.333	25.08

Figure 12.5: numerical examples of variables given thus far performed on the timetable in figure 12.2

the lessons are highly clumped, it's relatively not to a surprising extent. At my year group, the overwhelming majority of lessons are "doubles" (for better or for worse) (which is also reflected in the artificial segmentation of free periods), and "F" has just a good a good proportion of "singles" and "doubles" as any other block. Perhaps the choice to treat these "doubles" as two "singles" is limiting the scope of this specific implementation.

Now to evaluate the table as a whole: instead of combining the blocks when they were in sequence form (as we did with Δ) we combine their respective $\hat{\mu}_{C|S}$. A weighted average has exactly the properties desirable. Although it produces a clunky formula for our "clump" and "spread", it cancels out for the "mean" due to the relation in (12.18).

$$\mu = |\mathbb{L}| \tag{12.19}$$

(We briefly interject to alternatively offer the following ugly definition of μ, which doesn't require the proceedings to be calculated for *all* $l \in \mathbb{L}$, as was done in the example where "A" was omitted in the calculations. The discrepancies are small either way, but the following expression gives a worse definition of "mean" and a better comparison point for $\mu_{C|S}$.)

$$\mathbb{M} \subset \mathbb{L} : \mu = \frac{|L| \sum_{l \in \mathbb{M}} 1}{\sum_{l \in \mathbb{M}} N(\Delta(L, l))}$$

Thus, in the last few proceeding definitions, any $\mathbb{L}$ can be swapped for some $\mathbb{M}$ (and thus $|L|$ for $|L(\mathbb{L})|$) with limited loss of accuracy as long as $\mathbb{M} \approx \mathbb{L}$ and/or $L(\mathbb{M}) \approx L$.

$$\mu_C = \frac{\sum_{l\in\mathbb{L}} \hat{\mu}_{C(\Delta(L,l))} N(\Delta(L,l))}{|L|} \tag{12.20}$$

$$\mu_S = \frac{\sum_{l\in\mathbb{L}} \hat{\mu}_{S(\Delta(L,l))} N(\Delta(L,l))}{|L|} \tag{12.21}$$

If the reader feels the compulsion to expand (12.20) and (12.21) with the square roots given in (12.15) and (12.14) respectively, they must first remember to change the sequence in question from Δ to $\Delta(L,l)$ and then to regret their life choices.

What finally remains to round off is to give the values that we have been working towards for the duration of this subsection that has frankly gone on too long. With reference to the L reproduced in figure 12.4 and using $\mathbb{M} = \mathbb{L} \setminus \{\text{“A”}\}$,

$$\mu = 9.35\ldots \qquad \mu_C = 1.55\ldots \qquad \mu_S = 14.89\ldots$$

12.2.5 Sample Implementations

We presently present some exemplars on how this “NCS Analysis” could be used to make observations on various topics.

The majority of these examples take data from the author’s personal collection: such personalised data are not to be taken seriously. In fact, this article is written as part of a series of articles to celebrate the end of the fourth year of “a fact a day”, as the annual tradition of doing primary research for once. This year’s New Year’s Eve “fact of the day” is a collection of statistical analyses of various facts about the author themself, of which this section is forming a component.

Timetable *Verschlimmbesserung*

Verschlimmbesserung (literally something along the lines of “worse-bettering”) is a German word that can only be translated roughly to “disimprovement” or "enshittification", but instead it refers specifically to when attempts at improvement just make it worse. This is a phenomenon observed particularly among the bureaucratisation and digitalisation of society, which is exactly what's been happening to the timetabling process at the author's school. Timetabling over a thousand timetables in over a hundred rooms for 90 periods a fortnight is understandably a formidable task, but recently the way it's been done has changed and with it, many have reported a worsening of timetables. This is to what I attributed my surprising distribution of German lessons, as described at the beginning of section 1. However, we presently outline that this is not the case:

Timetables from three years were serialised and "NCS Analysed" with exactly the same methods as demonstrated on my timetable of the academic year 2024-2025 in section 4.3. The data from 2024-2025 is the data as given previously, and the data from 2023-2024 and 2022-2023 are both timetables also from the same year group as the first given one. A small amount of conforming was done to make the timetables as comparable as possible, as different people have different numbers of teachers and similar discrepancies can exist. (Further Mathematics, for example, is taught by four teachers simultaneously, whereas Physics only by one.) The results are surprising (all values rounded to three significant figures) and reproduced in table 12.4.

Timetable of	μ	μ_C	μ_C/μ	μ_S	μ_S/μ
'22-23	11.9	1.50	0.127	21.0	1.77
'23-24	10.7	1.45	0.135	18.8	1.75
'24-25	9.35	1.55	0.166	14.9	1.59

Table 12.4: "NCS Analysis" on timetabling development over three years

These results provide counterexample to the anecdote given in the introduction: that the timetabling has not been getting markedly worse in its distribution of lessons, perhaps even better. This is by no means conclusive, as analysis was performed on an extremely small sample size over an extremely short developmental timeframe. Moreover, this is just one descriptive measurement: one could argue that the "triple period" the German class this year received is enough to describe the German class's timetable this year as *verschlimmbessert*. Similarly, the number of clashes that have been timetabled in this year has gone up from more-or-less zero to a non-negligible number. However, by the measures of "clump" and "spread", there are certainly differences between these almost-randomly chosen timetables from three different points in time.

Focussing

I keep a log of how "focussed" I am for every activity that I do. The data are integers between 1 and 5 and corresponds purely to the ability to focus, dubbed "focability" (so not necessarily on one thing, as this is logged by a separate variable, "monobility"). Each datum, as we use it in this subsection, comprises a starting timestamp, an ending timestamp, an activity label, and a set of values in [5] (for example "focability" and "monobility"). For this analysis, we simply took the "focability" data as a one-dimensional array and ignored the timestamps. This does mean that the results are limited in their scope because there may be unaccounted-for biases when we ignore all durations. (Some "focability" data arguably should've held more weight, as they could last 12 minutes or 12 hours.)

"Focability" data were taken from the Master Log (called "O'Brien", for those that know it well) from a 130-day period between early August and early December 2024, forming a total of 1308 values. To conform the data to our processes, L is the "focability" sequence and $\mathbb{L}$ is the set of integers [5]. Unlike the implementations in sections 3 and 4, where the timetables "looped round", our sequence is now in a completely linear shape. This means that either the first or last instance of each $l \in \mathbb{L}$ cannot be mapped to an element in $\Delta(L, l)$ and thus $|\Delta(L, l)| = |L(l) - 1|$ and by its definition in (12.4), (12.5) changes to $|\Delta| = |L| - |\mathbb{L}|$. These changes are not given their own line because they are entirely ignorable facts, given thought-out notation.

Before performing the analysis, it may prove useful to speculate some hypotheses on what sort of results are informative. The threshold used for inputting a 1 or 5 (the most extreme values) are more stringent than thresholds between other numbers, as it's more remarkable when these inputs are given (as the inputter must be in a rut of un"focability" or a glut of high "focability"). This is useful only to a certain extent: too few 1s or 5s gives little sample size on other information regarding gluts and ruts. We can check the quality of these personal and entirely subjective thresholds by checking whether $N(\Delta(L, l))$ is roughly normally distributed over $l \in \mathbb{L}$.

There is another fact of the nature of inputting that could present itself herein. The data are recorded in a spreadsheet by a computer program that takes a numerical input on command from a panel and appends it to the list. However, if the numerical input is empty but the timestamp inputs are not, then the input is recorded as the same l as the previous value. This is for ease of inputting larger quantities of data, but there could also be the fear that values become "sticky" (where the inputter is likelier to leave the "default" value than change the input, especially when the "focability" is low) and thus not truly reflective of the real-world data. This is difficult to distinguish between ordinary time taken for "focability" to change (humans are, after all, slowly fluctuating beings), but a remarkably high "clump" across the board could point towards this "stickiness" being an issue.

Moreover, as just mentioned, the rate of fluctuation and variation and similar attributions will be shown in the "spread" and "clump" of the analysis, helping to strategically quantify the scale of fluctuation of "focability" and possibly aiding with time-planning in the future. (For example, "clump" holds information on the duration of how long one stays in a particular state, so this could help interleave high-focus tasks with low-focus tasks with the right timings.)

"NCS analysis" was performed on Δ first directly, as in section 4.2 with the $\hat{\mu}$ process, and subsequently on $\Delta L, l$, giving the final μ as in section 4.3. The information given by thus process is then compared and contrasted.

With respect to Δ in its entirety,

$$\hat{\mu}_C = 1.78\ldots \qquad \hat{\mu} = 5 \qquad \hat{\mu}_S = 9.09\ldots$$

Table 12.5 holds the breakdown of values with respect to each $\Delta(L, l)$. Values are rounded to three significant figures, with the exception of values for N.

Completing the process from section 4.3 using the derived data in table 12.5,

$$\mu_C = 1.81\ldots \qquad \mu = 5 \qquad \mu_S = 7.31\ldots$$

Immediately it is no surprise that $\hat{\mu} = \mu$ and that $\hat{\mu}_C \approx \mu_C$, although we must confess that we did not expect $\hat{\mu}_S$ to differ so greatly from μ_S, for which an intuitive explanation cannot be given. It's an. . . exercise for the reader.

l	$N(\Delta(L,l))$	$\hat{\mu}_C$	$\hat{\mu}$	$\hat{\mu}_S$	$\hat{\mu}_C/\hat{\mu}$	$\hat{\mu}_S/\hat{\mu}$
1	108	2.09	12.1	20.7	0.173	1.72
2	369	1.64	3.53	4.73	0.463	1.34
3	288	1.72	3.36	4.32	0.512	1.29
4	312	1.82	4.17	6.62	0.437	1.32
5	125	2.34	10.4	17.0	0.225	1.64

Table 12.5: values for "NCS Analysis" on "focability" data, broken down by $\mathbb{L}$

With regards to the "stickiness" predicted earlier, μ_C is certainly particularly low (corresponding to high "clump"). To put this in words, the average time between one state of "focability" and the next point in time when that state occurred, weighted in favour of "clump", is less than two activities. However, looking at table 12.5, we speculate that this is not due to "stickiness" of inputting, but rather "latency" or "drag" in the natural fluctuations of the state of the human mind. This is because the greatest geometric deviation of $\hat{\mu}_C$ occurred for $l = 1$ and $l = 5$, the most "drastic" inputs and thus the inputs on which one would expect to find the least "input stickiness". Instead, this points to the "dragging" of "ruts" and "gluts" being the primary factor of an unexpectedly low μ_C and $\hat{\mu}_C$. This is a positive affirmation that the collection of data is of sufficiently high quality to reflect the real world, and that it may be a good capture of an inherently subjective quality.

The discussion in the paragraph above was notionally on the value of μ_C, which was very similar to $\hat{\mu}_C$. The primary difference between these two values was that μ_C took around half an hour of spreadsheeting by an expert spreadsheeter to evaluate, whereas $\hat{\mu}_C$ took around five minutes. This is of course ignoring the processes involved in both paths (e.g. collecting, selecting and conforming the data) but there is undoubtedly greater levels of thinking and computation required to perform the "Better" method. Thus, $\hat{\mu}_{C|S}$ is presented as an approximation of $\mu_{C|S}$.

There is however still merit to the longer process. Having more steps involved makes it more laborious, but equally more useful: the intermediate steps (values of $\hat{\mu}_C$ for each $\Delta(L,l)$ in this case) proved useful for resolving the aforementioned discussion.

Socks

Another beautiful datalog in my archives is a log of my outfit each day. In this log, each day a "top", "overtop", "pair of bottoms" and "pair of socks" forms a datum. Before we perform an "NCS Analysis", if the reader would indulge us in validating our nerd credentials with a fact: by careful design and planning of day-to-day outfits, no two data since 2023-05-20 are the same. That is to say, no two "outfits" have been the same, to date and for the foreseeable future. This is less impressive when one considers that the majority heavy-lifting is done by the socks, of which there are at least 40 listed pairs and at least 20 pairs in regular use - one could argue that they form an underwhelming component of the "outfit", but this is not the point.

Due to the large number of socks (33 distinct elements $l \in \mathbb{L}$, as of before Christmas 2024 (one can understand that this number increases rapidly every Christmas)), we discuss the limitations of "NCS Analysis" thereon.

One fact that was regrettably not given more emphasis in the previous section is that with a linear log (as opposed to a circular timetable), the first (or last, depending on your choice of k in (12.1)) occurrence of each l in L can't be mapped meaningfully to a value in Δ. Hence it is omitted, and thus, as discussed previously, by (12.4), (12.2) becomes $|\Delta| = |L| - |\mathbb{L}|$. Previously, this posed no problem as $|\mathbb{L}|$ was negligibly small in comparison to $|L|$, but here it pushes the value of N in an unaccountable direction by around one sixteenth, immediately limiting the accuracy of our results thereto. Multiple pairs of socks were only worn once in this one-and-a-half-year period, and thus have no effect on the result - perhaps this is fair or unfair.

Due to the large size of $\mathbb{L}$, the "approximation" method only was used (the method notated with a hat and dealing with Δ in its entirety). To begin,

$$|L| = 588 \qquad |\mathbb{L}| = 33 \qquad |\Delta| = 555 \qquad \hat{\mu} = 33$$

A sample of the data (just 49 entries, thus representing less than a tenth of the whole dataset) is given in table 12.6 such that the reader can understand at a superficial level what sort of data is being dealt with. For the reader's interest, measures of average and spread for the entirety of Δ are also given, to highlight its peculiar nature. Obviously Δ is strictly positive. The most notable parts are the high standard deviation, and the disparity between non-weighted mean and the value for $\hat{\mu}(= \mu)$ above.

$$\text{mean} = 18.6\ldots \qquad \text{stdev} = 28.3\ldots \qquad \text{mode} = 6 \qquad \text{median} = 12$$

5	9	7	12	23	15	88
9	17	11	106	11	9	6
13	8	5	6	14	8	11
4	10	7	18	6	10	9
2	15	15	1	6	26	16
173	4	4	5	1	1	28
9	1	6	4	40	19	1

Table 12.6: the last 49 elements of the sequence Δ from the "socks log"

.

The following results are derived:

$$\hat{\mu}_C = 4.28\ldots \qquad \hat{\mu} = 33 \qquad \hat{\mu}_S = 33.8\ldots$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.1298\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 1.025\ldots$$

The tiny disparity between $\hat{\mu}_S$ and $\hat{\mu}$ comes from an incredible high value for "spread": by and large, most socks were worn remarkably evenly throughout the timeframe. We can vouch anecdotally that this is not true for a non-negligible number of pairs of socks, which were worn for a small sub-period and then not again for a long time (for example if they became hole-ridden and entered the long queue to be darned). This exposes two facts about the processes at play:

Firstly, that removing the circular nature of it means that pairs of socks that were poorly spread to the point that they never appeared again after a certain point do not have this accounted for. That is to say, there are long periods, sometimes spanning the entire timeframe of the data, where there ought to be a "distance" but due to a lack of element in Δ this is not accounted for in "spread" calculations. This may be desirable, but we believe that this could be improved upon. Possibly treating the linear log as a circular log ("patching it up" from the end to the beginning) inserts these periods back in in a neat manner, but may also give rise to less clear problems that cannot easily be seen in the results at 11pm at night. For example (and this is pure speculation), socks that were only acquired into the log at some point during the timeframe of recording might not want to have that initial "lead time" to be counted towards their value for $\hat{\mu}_S$ as this would be "unfair". The optimally-spread scenario in this case would involve it being locally well-spread but not globally as the new sock cannot be spread into the period before which it was adopted, obviously.

Secondly, the socks with a poorer spread, by virtue of them having become holey or unpreferred, also have a lower number and thus count less significantly towards the value for μ_S and by extension its simulated counterpart $\hat{\mu}_S$. This may be a preferred feature of the analysis, helping to give "fair" treatment to socks that become damaged or forcibly removed. It similarly helps hide the weight of socks whose preferability changed, but then it comes down to the purpose for which one wishes to perform the analysis in the first place, if it is not to gauge the magnitude of changes in preferability and fashion choices.

One alteration to the method that may aid to solve some of these uncertainties could be a time-based regressive analysis. For example, a new $\widetilde{\Delta}(t)$ could be formed from a window of L truncated at t and $t+k$, similar to how moving averages are computed. This exacerbates certain aspects of the problems discussed both in this subsection and elsewhere in the document, but helps to analyse others. The ins-and-outs of how to go about this escape the author at this hour, but the reader is welcome to practice their spreadsheet skills hereon.

Flashcard Reviews: Bins

Anki is a FOSS program to help memorisation, based on a spaced repetition system (SRS) and active recall. I use this program (and associated programs, e.g. AnkiDroid, AnkiWeb) to help learn vocabulary for language-learning, among other things. One of the core tenets of the SRS is that one uses the app on a daily basis to optimise long-term memorisation. The power of the SRS only works when one is consistent. Thus, it may be that an "NCS Analysis" of the review history of an Anki collection could shed some light, as well as revealing some limitations of the analysis process.

There is a component of Anki's culture that involves one's "streak" (how many consecutive days one has reviews flashcards) and one's "heatmap" (the representation of the density of one's reviews). One example of the heatmap is given below (cosmetics may vary) in figure 12.6. The brighter the cell, the more reviews were performed on that day. It is generally the goal to have one's reviews as evenly distributed throughout the heatmap as possible, as this benefits one's long-term memory and reduces "buildup" (where one becomes indebted to the rising number of "due" flashcards due to the SRS algorithm). High "clump" (low $\hat{\mu}_C$) could arise from "cramming", while poor "spread" (high $\hat{\mu}_S$) could arise from poor time-management.

In figure 12.6 it can be seen how the reviews were most reliable during the period of exams (due to a regular morning routine incorporating flashcard reviews) and least reliable during the first term of the current academic year (due to the intensity of schoolwork and the exacerbating effects of the buildup of due cards).

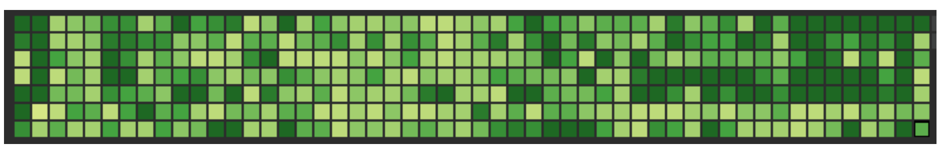


Figure 12.6: an example of Anki's heatmap, showing data between 2024-1-1 and 2024-12-29

As Anki is completely FOSS, the "revlog" (a database of every review ever) is easily accessible. A list of timestamps were extracted, corresponding to every time a flashcard was reviewed. In this case, as of writing, this is a list of 161,999 points in time, given with millisecond-precision. This was extracted, but needs to be conformed in some manner to L. The process behind the heatmap was simulated: timestamps from this calendar year (84,023 reviews) were collected into 366 bins, corresponding to the days of the year, and the size of each bin formed L.

However, because most values of L is in the order of magnitude of hundreds, the vast majority of elements of L are unique (that is to say, $\mathbb{L}$ is very large) and thus Δ is very small (120). Thus, each day is rounded to the nearest hundred, instead giving $|\mathbb{L}| = 11$. The results of the analysis hence are as follows:

$$\hat{\mu}_C = 1.82\ldots \qquad \hat{\mu} = 11 \qquad \hat{\mu}_S = 18.0\ldots$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.165\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 1.64\ldots$$

Performing the same on the entire history, dating back to November 2022, yields

$$\hat{\mu}_C = 1.71\ldots \qquad \hat{\mu} = 13 \qquad \hat{\mu}_S = 28.7\ldots$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.132\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 2.208\ldots$$

Which can be interpreted to affirm that this year's reviews were a remarkable improvement on the baseline, with significantly better "spread". However, if we take just the last academic term (corresponding roughly to the right-hand third of the heatmap in figure 12.6, where Mondays-Fridays are notably unpopulated), then we yield

$$\hat{\mu}_C = 1.75\ldots \qquad \hat{\mu} = 8 \qquad \hat{\mu}_S = 9.56\ldots$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.219\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 1.195\ldots$$

This demonstrates an embarrassing value for "clump" but an unexpectedly passable value for "spread". Contrasting with the heatmap, which offers qualitative and hand-wavey results, we can numericise our hunch to a degree, but equally it is evidently limited as we would have expected such a poor pattern to yield worse "spread" than is given.

Thus, the "bins" approach is limited in its usefulness. However, we offer an alternative process in the subsequent subsection:

Flashcard Reviews: Timestamps

Our data was conformed quite drastically to fit in with the model we developed looking at timetables. However, a completely different model more suited to this type of data could be more prospective.

If we take discard the idea of $\mathbb{L}$, everything suddenly becomes simpler, to the point where we should have begun with this example rather than the timetables. In the case of the timetables, we have points (the beginning of lessons) dotted through a medium (other lessons), and thus the difference between two lessons is given in terms of other lessons. This was also the case in the two other examples. However, in the case of the Anki review history, we have points (timestamps) dotted through a medium (the passage of time). This gives a much more absolute and sensible derivation for Δ:

$$\Delta_i = L_{i+1} - L_i$$

As previously, there is an offset due to the fact that L_{i+1} runs out and there is one element (the last/first, depending on how one counts it) of L that does not correspond to an element Δ. This is even less of a problem than previously, as this only applies to one L_i, in spite of $|\mathbb{L}| = \infty$ by some twisted interpretation.

As one would expect, this new approach devalidates the definition of "mean" in (12.19), as this would give $\mu = 1 \forall L$. Instead, we give a more intuitive definition of "mean":

$$\hat{\mu} = \frac{\sum_{d \in \Delta} d}{|\Delta|} = \frac{L_{\text{final}} - L_1}{|\Delta|}$$

Performing this updated process on the 2024 dataset of 84,023 datapoints gives

$$\hat{\mu}_C = 3.04\ldots \qquad \hat{\mu} = 371.6\ldots \qquad \hat{\mu}_S = 4348.0\ldots$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.00819\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 11.7\ldots\ldots$$

This at first appears truly shocking, but there is a rational explanation: when one says that one has reviewed n cards in a day, one often can also say that one has reviewed n cards in a small kn time period during that day. For me, that k is usually around 6 seconds. When one revises less in a day, there is still an extraordinarily high clump and poor spread because that revision occurs in a very small period of time in comparison to the day as a whole. This could be seen as fair treatment (perhaps one would learn more if one woke up every ten minutes during the night to review one (1) flashcard) but for this specific case, we perform a "smear":

This is a programmatic method of redistributing reviews throughout a day. For example, if a total of twelve reviews were executed on a date, but within two minutes of one another, the timestamps for these reviews are spread out evenly across the date in question, instead becoming Midnight, 2am, 4am, etc.. This could arguably lose a lot of precision with respect to *when* in the day the the reviews were carried out, but it also more loyally analyses the heatmap reproduced in figure 12.6. Unfortunately for the author, these subsections exceed the capabilities of spreadsheeting and enter the realm of Shoddy Python Code.

After this "smearing" (which is significantly more computationally intensive, unfortunately) was performed on the 2024 dataset, "NCS Analysis" was performed on the new set of timestamps (all averages rounded to the nearest integer):

$$\hat{\mu}_C = 180 \qquad \hat{\mu} = 374 \qquad \hat{\mu}_S = 1282$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.6575\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 4.674\ldots$$

Since these are all values with units of seconds, they need to be translated to "number of cards per day", which is proportional to the reciprocal of these values. $\hat{h}_{C|S}$ is used to denote these values.

$$\hat{h}_C = 480 \qquad \hat{h} = 231 \qquad \hat{h}_S = 67$$

These are very revealing values and much more useful than those given before. An average of 231 cards per day is good and a stone's throw from my target, and the fact that these stray comfortingly little when weighed by "clump" and "spread" is equally reassuring, as 67 or 480 cards in a day points to no failure, either by neglect or mismanagement.

This year, an active effort was made as a component of my New Year's Resolution to improve the usage of Anki and consistency thereof. It's reassuring that, in spite of the cliff-fall in the last third, good usage was made.

To aid these comparisons and evidence the goodness and success of 2024, we give the same analyses for the other two periods also outlined in the previous subsection, with all integers given having been rounded.

For the entire 26-month history:

$$\hat{\mu}_C = 199 \qquad \hat{\mu} = 420 \qquad \hat{\mu}_S = 4043$$

$$\hat{h}_C = 434 \qquad \hat{h} = 206 \qquad \hat{h}_S = 21$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.474\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 9.64\ldots\ldots$$

And for the last third of 2024:

$$\hat{\mu}_C = 217 \qquad \hat{\mu} = 510 \qquad \hat{\mu}_S = 2173$$

$$\hat{h}_C = 398 \qquad \hat{h} = 231 \qquad \hat{h}_S = 40$$

$$\frac{\hat{\mu}_C}{\hat{\mu}} = 0.426\ldots \qquad \frac{\hat{\mu}_S}{\hat{\mu}} = 4.26\ldots\ldots$$

Here's to another year, taking all the successes of 2024 and building upon the failures! May 2025 treat us all better.