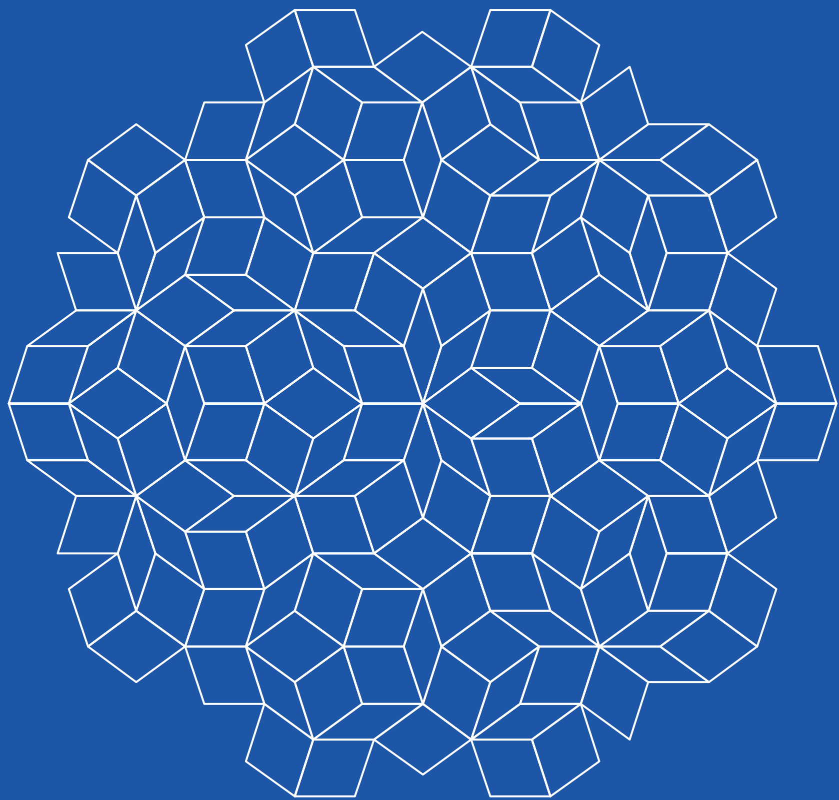


the invariant

Contents

Editor's letter **2**

Hilbert Space Approaches to the Riemann Hypothesis
Aditya Ghosh **4**

Descartes' Rule of Signs
Jaymin Shah **10**

When Randomness Isn't Random
Jacob Mercer **14**

The Success Paradox
Xingzhe Li **20**

Associative Metrics
Andres Klene **30**

Optimal dice for the game of Ludo
Max French **39**

Reflection Groups and Uniform Polytopes
Gavin Jared Bala **46**

A War of Words and Numbers: Exploring Controversy in Maths
Lauren Aitken **55**

Data and the stories it tells
Maria Tașcă **59**

Asymptotic hopes
Felix Stokes **62**

Sorry, I don't do maths
Siddiq Islam **63**

Proof by contradiction
Aleksandra Bozovic **64**

Galois' life
Eva Xu **65**

Editor's letter

Our society boasts an illustrious history. The Invariants was founded by Whitehead in 1936 and, since then, a long list of distinguished lecturers has adorned the society's term cards, including such names as Hardy, Mandelbrot, and Conway. The *Invariant*, the society's magazine and flagship, has been published for almost as long. It is with great honour, therefore, that I present to you this new edition.

In this issue you will find plenty of fascinating mathematics. Aditya Ghosh gives a reformulation, through functional analytic methods, of the Riemann Hypothesis as a simple (to state!) problem in analysis. Jaymin Shah's elegant exposition of a proof by Descartes provides a way to estimate the number of positive real roots of a polynomial just by looking at the signs of its coefficients. Gavin Bala takes us on a tour of the wonderful geometric world of Coxeter groups. These are just a few of the delightful pieces you will come across in these pages. I hope you enjoy reading them.

Hopefully, this magazine will go some way towards fulfilling the society's mission: to spread ideas and provide inspiration to the community of Oxford mathematicians that has been so active and creative for so long. An article might plant the seed of an idea that will later grow into the solution of a problem. An essay might spark a conversation on an issue you hadn't considered before. Perhaps a poem will bring a smile to your lips.

We want you to take part in this project! There are a myriad of ways to get involved: attend our talks and events and meet your fellow mathematicians (there is often food and games provided), or tell us about speakers you'd like to see lecture (think big – last year we had Roger Penrose!). Most importantly, we need your ideas – the continuation of this magazine depends on you. It could be a slick solution to a tricky problem, a summer project you've worked on, or an unusual proof of your favourite theorem. You could also write a more discursive piece, perhaps a historical, philosophical, or biographical essay. You could even submit a poem or illustrations for the magazine. Whatever your idea, and however unfinished, email it to us at *editor@invariants.org.uk* and we'll read it with great interest.

I urge you, then, to take inspiration from this issue and write! As the contributors of the present edition will attest, working on an article and polishing it to perfection is one of the most rewarding projects you can undertake. Our goal is to put a copy of the Invariant in the pigeonhole of every Oxford mathematician, and we need your passion and creativity to achieve it. This is *the* magazine for Mathematics at Oxford. It ought to be a good one.

I am indebted to the Invariants committee, especially president Otillia Cășuneanu, for their invaluable advice, to Eva Xu for her beautiful illustrations, and to my friend Tian-Long Lee for his tireless help in reviewing the drafts. Above all, I must thank the writers and contributors, without whose enthusiasm and originality the magazine couldn't exist.

Invariably yours,
Diego Vurgait
Editor.

Hilbert Space Approaches to the Riemann Hypothesis

Aditya Ghosh

Introduction

It is hard to understate the importance of the Riemann Hypothesis (RH). It remains one of the greatest unsolved problems in modern mathematics. It is closely related to the distribution of primes. Since its conception by Bernhard Riemann in 1859 there have been numerous attempts made over the centuries through vastly different approaches. In this article I shall highlight one such approach, which makes the Riemann Hypothesis equivalent to a closure problem - essentially how well can you *approximate* an element by other elements.

Let us start by recalling what the Zeta Function is. It was first introduced by Euler in 1737, which is now defined for $\mathrm{Re}(s) > 1$ by:

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

You might be wondering what this has to do with prime numbers. The Zeta Function can actually be shown to be equal to an infinite product called the *Euler Product*:

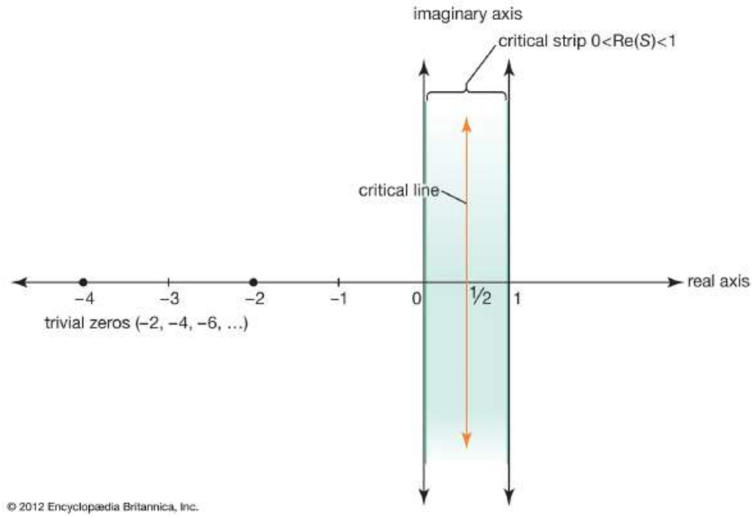
$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1} \qquad \text{for } \mathrm{Re}(s) > 1 \tag{2}$$

I encourage you to go ahead and prove this, without thinking too much about convergence, just manipulating the symbols.

Euler was primarily interested in $s \in \mathbb{N}$. Riemann further showed that $\zeta(s)$ can be extended to $s \in \mathbb{C} \setminus \{1\}$. The location of zeroes of the Zeta Function is of, pardon the pun, prime importance. There are a few *trivial zeroes* at $s = -2, -4, -6, \ldots$, which can be easily spotted. It can be shown that all the other zeroes, called the *non-trivial zeroes* lie in the *Critical strip* $0 < \mathrm{Re}(s) < 1$. Also they are symmetric about the *Critical Line* $\mathrm{Re}(s) = \frac{1}{2}$.

We can now state the Riemann Hypothesis:

Riemann Hypothesis All the non-trivial zeros of $\zeta(s)$ lie on the critical line $\{\mathrm{Re}(s) = \frac{1}{2}\}$.



A link between $\zeta(s)$ and $L^2(0, 1)$

RH has profound consequences on the distribution of primes and number theory in general. So you might wonder how primes, which are very simple algebraic objects, can be related to a closure problem, which is in the domain of analysis. It can be seen from the following result:

Lemma. *For every* $t \geq 1$, $\mathrm{Re}(s) > 0$, *we have*

$$\frac{1}{t(s-1)} - \frac{\zeta(s)}{t^s s} = \int_0^1 \rho\left(\frac{1}{tx}\right) x^{s-1}\, \mathrm{d}x \tag{3}$$

where $\rho(x)$ *is the fractional part of* $x \in \mathbb{R}$.

You can go ahead and prove this yourself (*Hint*: Try substituting $u = \frac{1}{tx}$ and exploit the fact that $\rho(u)$ is a periodic function).

If we evaluate equation (3) at $t = 1$, we get:

$$\frac{1}{(s-1)} - \frac{\zeta(s)}{s} = \int_0^1 \rho\left(\frac{1}{x}\right) x^{s-1}\, \mathrm{d}x \tag{4}$$

We can now obtain a simplified version of equation (3) by subtracting $\frac{1}{t} \times (4)$ from it:

Corollary. *For every $t \geq 1$, $\Re(s) > 0$, we have*

$$-\frac{\zeta(s)}{s}(t^{-s} - t^{-1}) = \int_0^1 \underbrace{\rho\left(\frac{1}{tx}\right) - \frac{1}{t}\rho\left(\frac{1}{x}\right)}_{g_t(x)} x^{s-1}\,\mathrm{d}x \tag{5}$$

where $\rho(x)$ is the fractional part of $x \in \mathbb{R}$.

Closure Problem

The above equation (5) establishes a relationship between $\zeta(s)$ and the elements g_t in $L^2(0, 1)$, the square integrable functions on $[0, 1]$.

Suppose RH is false and we have a non-trivial zero s_0 of $\zeta(s)$ such that $Re(s_0) > \frac{1}{2}$. Plugging $s = s_0$ into equation (5), we obtain:

$$0 = \int_0^1 g_t(x)\, x^{s_0-1}\,\mathrm{d}x$$

As integration operation is linear, we can replace $g_t(x)$ in the above equation by any element $g(x)$ in their linear span ν. Also, in $L^2(0, 1)$, this integration operation (integrating with x^{s_0-1}) is *continuous*. So if $g(x)$ is a limit of functions in ν, the above equation still holds.

Question Is the constant function **1** a limit of functions in ν

The answer, if we assume RH is false, is **no**.

Suppose it were true. Then we have:

$$0 = \int_0^1 \mathbf{1}\, x^{s_0-1}\,\mathrm{d}x = \frac{1}{s_0}$$

A contradiction! [1]

Hence we have proved a very nice result:

Theorem. **1** *is a limit point of ν in $L^2[0, 1]$ $\implies$ **RH***

[1] You might wonder if we can apply the same logic to s_0 with real part $\frac{1}{2}$. After all, we do know there are zeroes on this *critical line*. However, we can't do that because our (x^{s_0-1}-weighted) integration operation is not continuous for $\mathrm{Re}(s_0) = \frac{1}{2}$.

In fact the converse statement is also true. However the proof of it requires quite a lot of operator theory and is far too difficult to include here. The closure problem be stated as:

Theorem (Beurling, 1955). *Given the functions* $g_t \in L^2(0,1),\ t \geq 1$ *defined as before, let* $\nu := \operatorname{span}\{g_t \mid t \geq 1\}$ *the following statements are equivalent:*

1. *RH*
2. $\mathbf{1} \in \operatorname{clos}_{L^2(0,1)}(\nu)$
3. $\operatorname{clos}_{L^2(0,1)}(\nu) = L^2(0,1)$*, that is,* ν *is dense in* $L^2(0,1)$

We can do even better. We can restrict our attention from uncountable set $\{t \geq 1\}$ to simply $t \in \mathbb{N}$. This was proven by Báez-Duarte as recently as 2003. So the Riemann Hypothesis simply depends on whether one can *approximate* the constant function $\mathbf{1}$ using $g_1, g_2, g_3, \ldots$

A first-year analysis problem

The functions $g_1, g_2, g_3, \ldots$ are not too complicated. Here's a little exercise: prove that for $k = 1, 2, 3 \ldots$, the function g_k is constant on each interval $(\frac{1}{n+1}, \frac{1}{n})$ and the value is $\rho(\frac{n}{k})$. So essentially these functions g_k add and subtract just like the sequences $s_k = \left(\rho(\frac{n}{k})\right)_{n\in\mathbb{N}}$. Let's write down some of them:

$$s_1 = \left(\rho\left(\frac{n}{1}\right)\right) = (0, 0, 0, 0, 0, 0, 0 \ldots)$$
$$s_2 = \left(\rho\left(\frac{n}{2}\right)\right) = (\frac{1}{2}, 0, \frac{1}{2}, 0, \frac{1}{2}, 0, \ldots)$$
$$s_3 = \left(\rho\left(\frac{n}{3}\right)\right) = (\frac{1}{3}, \frac{2}{3}, 0, \frac{1}{3}, \frac{2}{3}, 0, \ldots)$$

Even the sequences s_k themselves are periodic, repeating themselves every k terms. So **RH** is equivalent to whether we can *approximate* the sequence $(1, 1, 1, 1, 1, \ldots)$, which corresponds to the constant function $\mathbf{1}$, using linear combinations of these s_k sequences.

Isn't that remarkable, that something as difficult as the Riemann Hypothesis has such a simple reformulation? One can even write a formula or code to calculate how close the constant sequence $(\mathbf{1})$ is from the linear subspace spanned by $\{s_1, s_2, \ldots s_k\}$. Call this distance d_k. The Riemann Hypothesis follows if $d_k \to 0$ as $k \to \infty$! The proof is left as a margin-worthy exercise to the reader.

Positive Root
p(x) = a0 + a1x + a2x2 + ... anxn
GEOMETRY
ALGEBRA
By Eva Xu
RENÉ DESCARTES MAR 31, 1596 - FEB 11, 1650

Descartes' Rule of Signs

Jaymin Shah

How many positive real roots does the polynomial $x^2 + 1$ have? How about $4x^3 - 2x - 1$? How about $a_n x^n + \cdots + a_1 x + a_0$? Now before I get your hopes too high, I won't be deriving a formula that takes in a polynomial and spits out the number of positive roots it has. Instead, I'll be looking for a non-trivial[1] upper bound to the number of positive roots a polynomial can have. If we think about polynomials with lots of positive real roots, we think of the signs of the coefficients changing a lot (after all, if most of the coefficients were positive, say, the polynomial would be increasing most of the time, and not have the ability to fluctuate much between positive and negative values). This suggests that there should be a relationship between these two quantities - and it turns out that there is one! And it's simple to state and elegant to prove!

Definition. *Let $p(x) \in \mathbb{R}[x]^*$. Write $p(x) = a_0 x^{b_0} + a_1 x^{b_1} + \cdots + a_n x^{b_n}$ where $b_0 < b_1 < \cdots < b_n$ and $a_i \neq 0 \; \forall \; i = 1, \cdots, n$. We define $s(p)$ to be the number of sign changes of the coefficients of $p(x)$. That is, the number of i for which $a_i a_{i+1} < 0$.*

Definition. *Let $p(x) \in \mathbb{R}[x]^*$. We define $z(p)$ to be the number of strictly positive real zeroes of p counted with multiplicity.*

Lemma. *Let $p(x) \in \mathbb{R}[x]^*$. Write $p(x) = a_0 x^{b_0} + a_1 x^{b_1} + \cdots + a_n x^{b_n}$ where $b_0 < b_1 < \cdots < b_n$ and $a_i \neq 0 \; \forall \; i = 1, \cdots, n$. Then $z(p)$ is even if and only if $a_0 a_n > 0$.*

Proof. We begin by noting that without loss of generality, we may assume that $b_0 = 0$ as factoring out x^{b_0} does not impact $z(p)$ or the sign of $a_0 a_n$. Furthermore, we may assume without loss of generality that $a_0 > 0$ as if it were negative we could consider $-p(x)$ instead and, again, this would not impact $z(p)$ or the sign of $a_0 a_n$. *The remainder of the proof won't be too rigorous for the sake of brevity. To make the proof rigorous is a simple exercise.*

Consider the case where $a_n > 0$. So we have $p(0) = a_0 > 0$ and $\lim_{x \to \infty} p(x) = \infty$. Hence at $x = 0$ the graph of $p(x)$ is above the x-axis and for sufficiently large x we

[1] Any polynomial of degree n has n roots in $\mathbb{C}$ and thus $\leq n$ roots in $\mathbb{R}^{\geq 0}$. Any bound better than this is considered non-trivial.

remain above the x-axis. Hence, if $p(x)$ was to cross the x-axis at some positive root of p, it must cross back at some point, which will contribute 2 to $z(p)$. Hence $z(p)$ must be even. A similar argument proves that if $a_n < 0$, then $z(p)$ is odd. □

We now proceed to the main theorem of finding a non-trivial upper bound to $z(p)$. In fact, the result we'll be interested in will be an obvious corollary to what we'll prove. We'll prove a stronger result by induction, and this illustrates the fact that sometimes it can be easier to prove stronger statements than weaker statements by induction. The reason it's easier (in some cases) to prove a stronger statement this way is because when we assume the inductive hypothesis, we can assume more than in the case of the weaker statement.

Theorem. *Let $p(x) \in \mathbb{R}[x]^*$. Then $z(p) \leq s(p)$ and $z(p)$ and $s(p)$ have the same parity, ie $z(p) \equiv s(p) \pmod 2$.*

Proof. As mentioned, we'll prove this result by (strong) induction: we'll induct on $k := \deg(p)$. The base case is $k = 0$, in which case $z(p) = s(p) = 0$ and so we have both $z(p) \leq s(p)$ and $z(p) \equiv s(p) \pmod 2$. Now fix k and suppose the result holds true for all polynomials of degree $\leq k - 1$.

Let $p(x) \in \mathbb{R}[x]^*$ be of degree k. Write $p(x) = a_0x^{b_0}+a_1x^{b_1}+\cdots+a_nx^{b_n}+\alpha x^k$ where $b_0 < b_1 < \cdots < b_n < k$ and $\alpha, a_i \neq 0 \ \forall\, i = 1, \cdots, n$. Now, if $b_0 \neq 0$, we consider $\tilde{p}(x) := \frac{p(x)}{x^{b_0}} \in \mathbb{R}[x]$ and observe that $z(p) = z(\tilde{p})$ and $s(p) = s(\tilde{p})$. Furthermore, since $\tilde{p}$ is a polynomial of degree $\leq k - 1$, by the inductive hypothesis it would follow that $z(p) \leq s(p)$ and $z(p) \equiv s(p) \pmod 2$. Hence, without loss of generality, we assume $b_0 = 0$.

Case 1. We first consider the case where $a_0a_1 > 0$. Then this implies that a_0 and a_1 have the same sign so $s(p) = s(p')$. But, also, by observing that $p'(x) = b_1a_1x^{b_1-1} + \cdots + b_na_nx^{b_n-1} + k\alpha x^{k-1}$ and using the fact that a_0 and a_1 have the same sign, we deduce that

$$a_0\alpha > 0 \iff a_1\alpha > 0 \iff b_1a_1k\alpha > 0$$

And so by the lemma, we have that $z(p) \equiv z(p') \pmod 2$. Using the inductive hypothesis, we have $z(p') \equiv s(p') \pmod 2$ and so $z(p) \equiv z(p') \equiv s(p') \equiv s(p) \pmod 2$. Finally, by Rolle's theorem we know that between any two roots of p there exists at least one point for which $p' = 0$. Hence, we must have that $z(p') \geq z(p) - 1$, but since they have the same parity, it must be that $z(p') \geq z(p)$. Hence $z(p) \leq z(p') \leq s(p') = s(p)$.

Case 2. We now consider the case where $a_0 a_1 < 0$. In this case, a_0 and a_1 have different signs and so $s(p) = s(p') + 1$. Like in the first case, we can use the lemma, except this time to prove that $z(p) \equiv z(p') + 1 \pmod 2$. Using the inductive hypothesis we then have $z(p) \equiv z(p') + 1 \equiv s(p') + 1 = s(p) \pmod 2$. Then $z(p) \leq s(p) \leq \deg(p)$. Again, using Rolle's theorem we have that $z(p') \geq z(p) - 1$ and so $z(p) \leq z(p') + 1 \leq s(p') + 1 = s(p)$.

□

Corollary (Descartes' Rule of Signs). *Let $p(x) \in \mathbb{R}[x]^*$. Then $z(p) \leq s(p) \leq \deg(p)$.*

Proof. Immediate from the theorem. □

In my opinion, this theorem is beautiful and uses only a little bit of analysis to prove, but in fact, most of the time the use of techniques from analysis was overkill, and was only stated to make things slightly more rigorous.

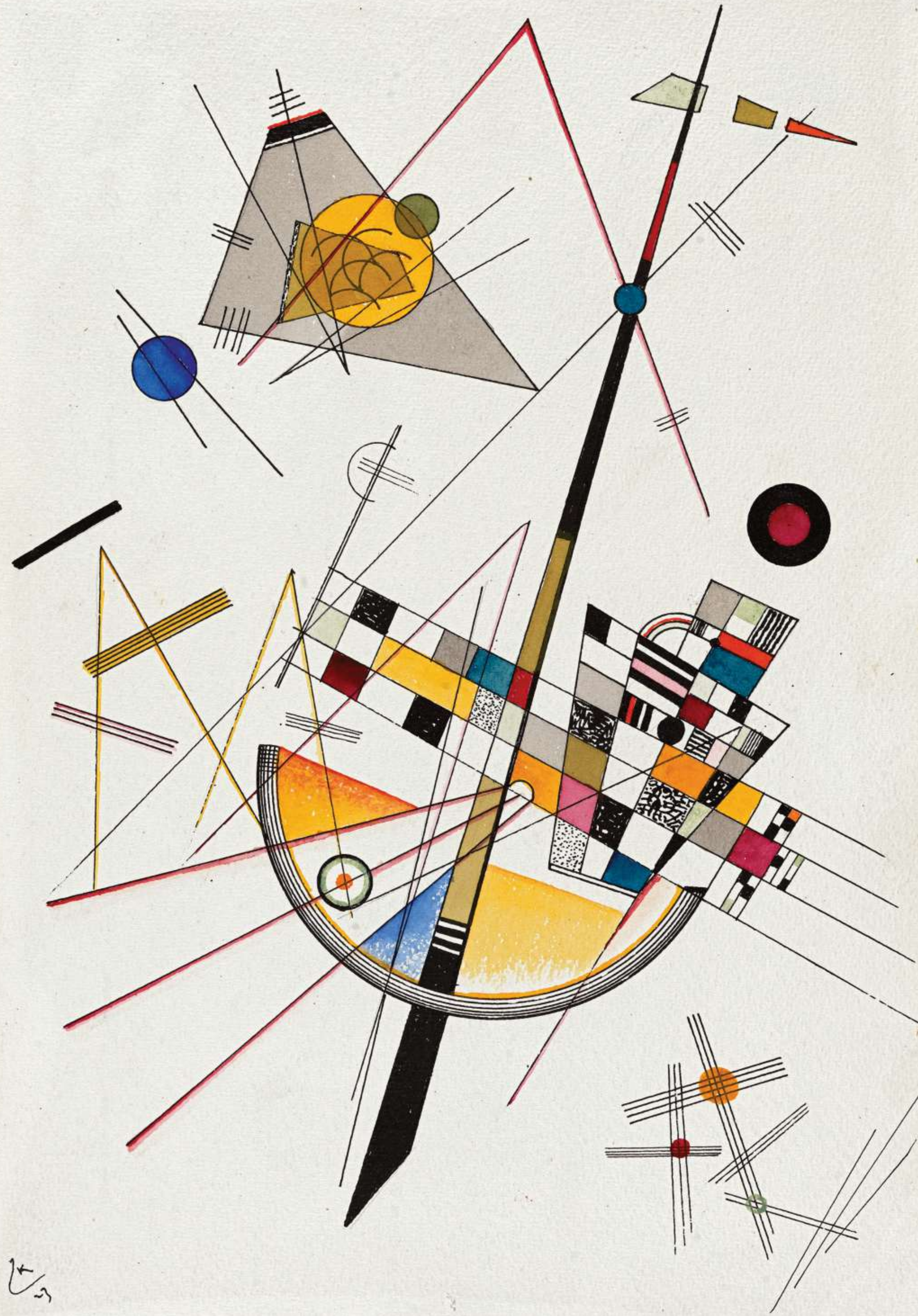
There is another similar theorem which may interest the reader. Sturm's theorem allows one to calculate the exact number of real roots certain polynomials have in intervals of the form $(a, b]$ based on the number of sign changes in a particular sequence of numbers.

Definition. *Let $p(x) \in \mathbb{R}[x]^*$. We define the Sturm sequence of p as follows. $p_0 = p$, $p_1 = p'$ and $p_{i+1} = -R(p_{i-1}, p_i)$ where $R(p_{i-1}, p_i)$ is the remainder upon Euclidean division of p_{i-1} by p_i.*

Note that the Sturm sequence of p terminates after at most $\deg(p)$ terms. For any $x \in \mathbb{R}$ define $s_p(x)$ as the number of sign changes in the sequence $p_0(x), p_1(x), p_2(x), \cdots$.

Theorem (Sturm's Theorem). *Let $p(x) \in \mathbb{R}[x]^*$ be square-free. Then the number of distinct real roots in the interval $(a, b]$ is $s_p(b) - s_p(a)$.*

To see more maths from me, please do check out my YouTube channel where I solve fun maths problems and prove cool theorems (just like this one): `https://www.youtube.com/jpimaths`.



When Randomness Isn't Random

Jacob Mercer

A monkey and a typewriter...

Probability is, in essence, the study of randomness. However, 'random' means different things in common parlance and in mathematics. For example, one might say that rolling an integer on a fair dice isn't a random event because it always happens! Semantics aside, mathematicians certainly have no problem with events taking probability zero or one. It's axiomatic even, since $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$. And in true mathematical fashion there are many theorems dedicated to when an event has probability zero or one, sometimes called zero-one laws. Naturally, such theorems inhabit a narrow space in between being too specific to be useful and/or non-trivial, and having enough conditions to ensure that the probabilities can only take two values.

The aim of this article is to give a brief overview of some of the elegant zero-one laws out there, outline their similarities and uses, and sketch their proofs.

Throughout the article we will call events with probability zero or one *trivial* and sets of trivial events we will also call *trivial*. To understand Blumenthal's zero-one law, it will be helpful to be familiar with the definition of a Brownian motion (a continuous Markovian random walk).

Lemma (The Borel-Cantelli Lemmas). *Fix a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and let* $(A_n)_{n\in\mathbb{N}}$ *be a family of events in* $\mathcal{F}$*.*

1. *If* $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ *then* $\mathbb{P}(A_n \textit{ occurs infinitely often}) = 0$.
2. *If* (A_n) *are independent and* $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ *diverges then* $\mathbb{P}(A_n \textit{ occurs infinitely often}) = 1$

A famous thought experiment serves as an illustration: a monkey at a typewriter will eventually produce Shakespeare's complete works. Perhaps the mathematician would clarify that the monkey must hit each key with non-zero probability, but otherwise this is simply an application of lemma 1. In fact lemma 1 goes further: the monkey will write the entirety of Shakespeare's works *infinitely many times*.

Some mathematical background

We will take some time before the following theorems to introduce the mathematical language we need to talk about them.

Throughout the rest of the article we will talk about random variables set in a filtered probability space, that is a probability space with a *filtration* which may be indexed by $\mathbb{N}$ or $\mathbb{R}^{\geq 0}$: write it $(\mathcal{F}_t)_{t \in I}$ for some index set I. A filtration is simply an increasing sequence of sets of events, so for all $s < t \in I$, $\mathcal{F}_t$ is a family of events just like $\mathcal{F}$, and $\mathcal{F}_s \subseteq \mathcal{F}_t$. Given a random/stochastic process, we might consider its *natural filtration*. This is, as it sounds, a very natural choice of filtration in which $\mathcal{F}_t := \sigma(X_s : s \leq t)$ is the σ-algebra generated by the set $\{X_s : s \leq t\}$. In layman's terms, $\mathcal{F}_t$ is the smallest family of events which can be determined by $\{X_s : s \leq t\}$. So at time t, the natural filtration contains all the events you could know by time t and nothing more.

Definition. *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables. Then the tail algebra is the defined as:*

$$\mathcal{G} := \bigcap_{n=0}^{\infty} \sigma(X_n, X_{n+1}, \ldots)$$

In layman's terms, since $\sigma(X_n, X_{n+1}, \ldots)$ is roughly what you can know given variables $X_n, X_{n+1}, \ldots$, so $\mathcal{G}$ is roughly what you can know given only the tail of the sequence.

Zero-One Laws

Theorem (Kolmogorov's Zero-One Law). *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of indepedent random variables. Then the tail algebra is trivial.*

This theorem is surprisingly powerful. In essence it says that *any event determined by all but finitely many of the initial X_i has probability 0 or 1.* For example, $X_1 + X_2 + \ldots$ converges with probability 1 or 0. The proof has an elegant simplicity as well: to prove that some number x can only take values 0 or 1, one only needs to show that it solves $x(1 - x) = 0$. And that's precisely how the proof of Kolmogorov's law works, by showing that any event $G \in \mathcal{G}$ is independent of itself, and therefore that $\mathbb{P}(G) = \mathbb{P}(G \cap G) = \mathbb{P}(G)^2$.

As an example, consider a set S which contains each integer n, independently of other integers, with probability $p_n = \mathbb{P}(n \in S)$. Then the event that S contains an infinite sequence of consecutive integers is trivial. Of course whether this probability is zero or one requires a little more information, but the power of this theorem is that we have reduced the problem to two possible values.

Kolmogorov's zero-one law is also useful in percolation theory in which one studies graphs (often lattices like $\mathbb{Z}^d$) in which each edge (or alternatively each vertex) is present independently with probability p. In this setting, any event determined by all but finitely many edges is trivial. For example: does the graph contain an infinite component? Does the graph contain infinitely many triangles?

Very similar in form to Kolmogorov's zero-one law is the zero-one law of Hewitt and Savage.

Theorem (Hewitt-Savage Zero-One Law). *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables. Then any event whose occurrence is invariant under finite permutations is trivial.*

This theorem, you will notice is very similar to Kolmogorov's zero-one law. Like Kolmogorov's, the Hewitt-Savage zero-one law can tell you that $\mathbb{P}(\sum_i X_i \text{ converges}) \in \{0, 1\}$. Or it can tell you about other events concerning the random walk $(S_n) = (\sum_{i=1}^n X_i)$ like that the event $\{S_n = 0 \text{ infinitely often}\}$ is trivial.

Another example, for those familiar with the definition of Brownian motion (or willing to look it up), is the following: consider the case where the i.i.d. random variables (X_n) are $X_n = B_n - B_{n-1}$, where (B_n) is the standard Brownian motion. Then we can apply the Hewitt-Savage zero-one law to an event like $\{B_n > c\sqrt{n} \text{ infinitely often}\}$, and conclude that the probability of such an event is zero or one, and we can rule out zero by Fatou's lemma. The remarkability of this result is that we can thus arrive at a result like: $\limsup B_n/\sqrt{n} = \infty$ — a result about a nowhere-differentiable stochastic process in continuous time — simply by using a theorem about a countable sequence of random variables and finite permutations thereon.

The next theorem also has applications to continuous random processes and Brownian motion and may be familiar to students who have studied Part B Continuous

Martingales and Stochastic Calculus.

Theorem (Blumenthal's Zero-One law). *Let $(X_t)_{t\geq 0}$ be a right-continuous Feller process (for ease we might just think of Brownian motion) with $\mathbb{P}(X_0 = x_0) = 1$. Let $(\mathcal{F}_t^X)$ be the natural filtration, and let $\mathcal{F}_{0+}^X := \bigcap_{t>0} \mathcal{F}_t^X$. Then $\mathcal{F}_{0+}$ is trivial.*

Blumenthal's Zero-One law is another powerful result, and can prove fundamental facts about Brownian motion, like the fact that Brownian motion (started from 0) always goes above zero at some point in any arbitrarily small interval. That is that, for any $\varepsilon > 0$, we have $\sup_{s<\varepsilon} B_s > 0$. Indeed we can also prove the result above, that $\limsup B_n/\sqrt{n} = \infty$.

Although maybe after all this connection isn't too surprising. It is well known that if B_t is a Brownian motion, then so is the stochastic process $W_t := tB_{1/t}$, hence knowing about the natural filtration of B_t as t goes to zero is like knowing about the natural filtration of W_t as t goes to infinity. In the case of Brownian motion, then, Kolmogorov's zero-one law and Blumenthal's zero-one law are just two sides of the same coin, connected by the fact that B_t and W_t are both Brownian motions.

The next zero-one law we will see is Levy's zero-one law. This theorem again walks the narrow line between being completely trivial — indeed it almost seems so — and actually being a powerful result. I hope to convince you that it is the latter. In this theorem we will need the notion of the family of events $\mathcal{F}_\infty^X$ which consists of all the events which can be determined by knowing the entirety of some process $(X_n)_{n\in\mathbb{N}}$, ie $\mathcal{F}_\infty^X := \sigma((X_n)_{n\in\mathbb{N}})$.

Theorem (Levy's Zero-One Law). *Suppose $\mathcal{F}_n$ increases to $\mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$. Then $\mathbb{E}[1_A|\mathcal{F}_n]$ tends to 1_A almost surely.*

It seems almost trivial or obvious in the sense that of course $\mathbb{E}[1_A|\mathcal{F}_\infty] = 1_A$, so of course if $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then it would make sense that $\mathbb{E}[1_A|\mathcal{F}_n] \to 1_A$. Now when one comes to proving the result rigorously one sees that it is not quite so trivial. In fact Levy's zero-one law is powerful enough to prove Kolmogorov's zero-one law as a corollary:

Corollary. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables. Then the tail algebra is trivial.*

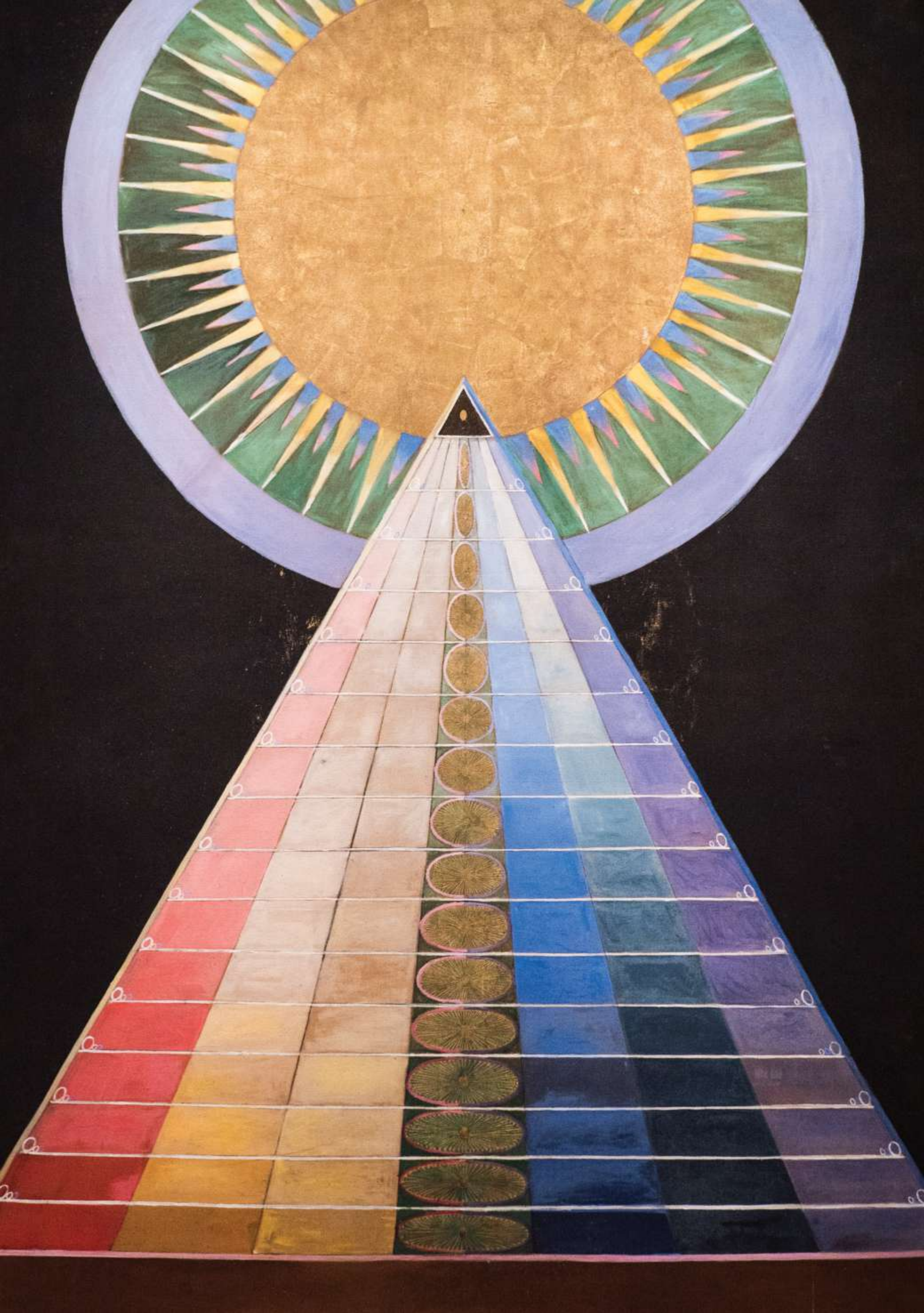
Proof. Let A be an event in the tail algebra. This means that, for all integers n, A is independent of $\mathcal{F}_n$, so $\mathbb{P}(A) = \mathbb{E}[1_A|\mathcal{F}_n] \to 1_A$ almost surely. □

Levy's zero-one law is therefore definitely non-trivial since it can prove Kolmogorov's zero-one law immediately. Another corollary is a dominated convergence theorem (the dominated convergence theorem for conditional expectations).

Conclusion

Looking back at the results above we see that they reach, by their nature, a balance between being trivial and needing conditions to be strong enough. This is evident as well by the fact they all require infinity. Somehow we can't expect to find any interesting statement about probabilities being trivial with only finite objects in play, but when limits to infinity are involved we can find subtleties which leave events trivial and/or independent of themselves.

As we have started to see through the examples in this article, these theorems produce a number of interesting results of which we can be certain.



The Success Paradox

Xingzhe Li

Introduction

What makes someone successful? Is it grit, persistence, determination, and hard work? No doubt you cannot be successful without these qualities. However, what is often neglected is the element of luck. This paper is inspired by a YouTube video from the popular channel Veritasium titled 'Is Success Luck or Hard Work'. The aim is to determine the role of luck in success.

In the video by Veritasium, he made a simulation where 11 astronauts are selected from 18,300 applicants. Specifically, each applicant is given a luck score and a merit score to generate a total score with 0.05 weighting from the luck score and 0.95 weighting from the merit score. In this paper, we will abstract this experiment to ask what is the expected luck score of someone that beat a proportion c of the population when luck and merit score have a and $1-a$ weighting respectively. While the focus is on highly competitive situations with low weighting of luck (high value of c and low value of a), general results will be deduced as well.

Notation

We let T, H, and L be the random variables for total score, merit score and luck score respectively. We assume H and L are independent. Next, let a be weight of luck, and $1 - a$ be weight of 'merit' (essentially packaging all non-luck elements). That is, $T = aL + (1 - a)H$. Valuing merit over luck, I will assume $0 < a \leq 0.5$. For a random variable X, we let f_X be its probability density function, and F_X its cumulative distribution function. Furthermore, let $P(A)$ denote the probability of an event A. Finally, we let $0 < c < 1$ be the proportion of unsuccessful candidates. We aim to find the expected luck score of a successful candidate, that is $E(L|T > \alpha)$, where $\alpha = F_T^{-1}(c)$. ($E(X)$ being the expected value of X).

Uniform Distribution

We will start by analysing the scenario where H and L follow a uniform distribution, as was assumed in the video. That is, $H, L \sim U(0, 1)$. This implies $0 \leq T \leq 1$.

First note that for $T > \alpha$ to be possible, $L > \frac{1}{a}(\alpha - (1-a)H) \geq \frac{\alpha+a-1}{a}$. Let $m = max(\frac{\alpha+a-1}{a}, 0)$, we have $E(L|T > \alpha) = \int_m^1 x f_{L|T>\alpha}(x)dx$. Moreover, $f_{L|T>\alpha}(x) = \frac{d}{dx}(F_{L|T>\alpha}(x))$.

We have $P(T > \alpha) = 1 - P(T < \alpha) = 1 - c$, but it is useful to find this in terms of α. This will also help in identifying $F_T(t)$ and subsequently $F_T^{-1}(c)$ in terms of c. (The reason we pursue this approach instead of directly finding F_T is because the definition of m makes this approach cleaner to execute, as shown below.)

$$\begin{aligned} P(T > \alpha) &= P(aL + (1-a)H > \alpha) \\ &= P(L > m, H > max(0, \tfrac{\alpha - aL}{1-a})) \\ &= \int_m^1 \int_{max(0, \frac{\alpha-aL}{1-a})}^1 1dhdl \end{aligned}$$

Case 1: $\alpha < a$ Note since we assumed $a < 0.5$, we know that $\alpha < 1 - a$, so $m = 0$.

$$\begin{aligned} P(T > \alpha) &= \int_0^1 \int_{max(0, \frac{\alpha-al}{1-a})}^1 1dhdl \\ &= \int_0^{\alpha/a} 1 - \frac{\alpha-al}{1-a}dl + \int_{\alpha/a}^1 1dl \\ &= 1 - \frac{\alpha^2}{2a(1-a)} \end{aligned}$$

Case 2: $a \leq \alpha \leq 1 - a$ In this case, we still have $m = 0$. Also, since $0 \leq l \leq 1$, $al \leq a \leq \alpha$, $\frac{\alpha-al}{1-a} > 0$.

$$\begin{aligned} P(T > \alpha) &= \int_0^1 \int_{\frac{\alpha-al}{1-a}}^1 1dhdl \\ &= \frac{2-2\alpha-a}{2(1-a)} \end{aligned}$$

Case 3: $\alpha > 1 - a$ In this case, $m = \frac{\alpha+a-1}{a}$, $\frac{\alpha-al}{1-a} > 0$.

$$\begin{aligned} P(T > \alpha) &= \int_{\frac{\alpha+a-1}{a}}^1 \int_{\frac{\alpha-al}{1-a}}^1 1dhdl \\ &= \frac{(1-\alpha)^2}{2a(1-a)} \end{aligned}$$

Now we can write:

$$P(T > \alpha) = \begin{cases} 1 - \frac{\alpha^2}{2a(1-a)}, & 0 \leq \alpha < a \\ \frac{2-2\alpha-a}{2(1-a)}, & a \leq \alpha \leq 1-a \\ \frac{(1-\alpha)^2}{2a(1-a)}, & 1-a < \alpha \leq 1 \end{cases}$$

$$F_T(t) = \begin{cases} \frac{t^2}{2a(1-a)}, & 0 \leq t < a \\ \frac{2t-a}{2(1-a)}, & a \leq t \leq 1-a \\ 1 - \frac{(1-t)^2}{2a(1-a)}, & 1-a < t \leq 1 \end{cases}$$

$$F_T^{-1}(u) = \begin{cases} \sqrt{2a(1-a)u}, & 0 \leq u < \frac{a}{2(1-a)} \\ \frac{2(1-a)u+a}{2}, & \frac{a}{2(1-a)} \leq u \leq \frac{2-3a}{2(1-a)} \\ 1 - \sqrt{2a(1-a)(1-u)}, & \frac{2-3a}{2(1-a)} < u \leq 1 \end{cases}$$

Now, we evaluate $P(T > \alpha, L < x)$ for x between 0 and 1 in much the same way.

$$\begin{aligned} P(T > \alpha, L < x) &= P(aL + (1-a)H > \alpha, L < x) \\ &= P(x > L > max(0, \alpha + a - 1/a), H > max(0, \tfrac{\alpha - aL}{1-a})) \\ &= \int_m^x \int_{max(0, \frac{\alpha-al}{1-a})}^1 1dhdl \\ &= \begin{cases} x - \frac{\alpha^2}{2a(1-a)}, & 0 \leq \alpha < ax \\ \frac{x(ax-2a-2\alpha+2)}{2(1-a)}, & ax \leq \alpha \leq 1-a \\ \frac{(1-\alpha-a+ax)^2}{2a(1-a)}, & ax < 1-a < a \end{cases} \end{aligned}$$

A good sanity check is that when $x = 1$, we have $P(T > \alpha, L < x) = P(T > \alpha)$ as one would expect since $L \sim U(0, 1)$. Another more geometric method to evaluate $P(T > \alpha)$ is to consider the $\mathbb{R}^2$ plane with L and H axis. Since, (L, H) is distributed over the unit square evenly, the proportion of area above the line $aL + (1-a)H = \alpha$ in the unit square is $P(T > \alpha)$. The three cases refers to the scenarios where the line intersects both axes between 0 and 1, one axis between 0 and 1, and neither axis between 0 and 1.

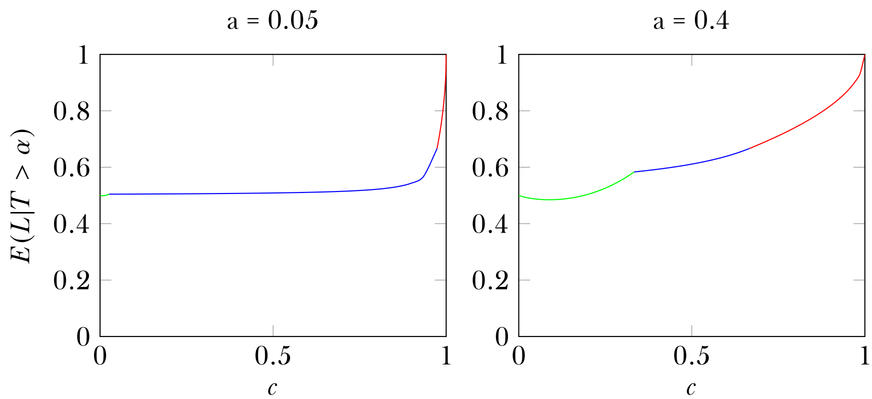
To evaluate $P(T > \alpha, L < x)$, we could use the same geometric approach to evaluate $P(T > \alpha | L < x)$ (by considering the rectangle bounded by $L = 0, L = x, H = 0, H = 1$ instead of the unit square) then use the fact that $P(L < x) = x$ in the relevant range.

In any case, we now know that:

$$F_{L|T>\alpha}(x) = \begin{cases} \frac{2ax(1-a)-\alpha^2}{2a(1-a)-\alpha^2}, & \alpha < ax \\ \frac{ax(ax-2a-2\alpha+2)}{2a(1-a)-\alpha^2}, & ax \le \alpha < a \\ \frac{x(ax-2a-2\alpha+2)}{2-2\alpha-a}, & ax \le a \le \alpha \le 1-a \\ \frac{(1-a-\alpha+ax)^2}{(1-\alpha)^2}, & ax \le 1-a < \alpha \end{cases}$$

By taking $x = \alpha/a$ and considering the three cases $\alpha < a$, $a \le \alpha \le 1-a$, and $1-a < \alpha$, one can see that $F_{L|T>\alpha}$ is in fact continuously differentiable. (It is helpful to notice in the second case, $ax = a = \alpha$ so $x = 1$ and in the third case, $ax = 1-a = \alpha$ implying $a = 0.5 = \alpha, x = 1$.) Thus, we can apply integration by parts to find $E(L|T > \alpha)$.

$$\begin{aligned} E(L|T > \alpha) &= \int_m^1 x f_{L|T>\alpha}(x)dx \\ &= xF_{L|T>\alpha}(x)\Big|_m^1 - \int_m^1 F_{L|T>\alpha}(x)dx \\ &= \begin{cases} 1 - \frac{1}{2(1-c)} + \frac{2c\sqrt{2a(1-a)c}}{3a(1-c)}, & c < \frac{a}{2(1-a)} \\ \frac{1}{2} + \frac{a}{12(1-a)(1-c)}, & \frac{a}{2(1-a)} \le c \le \frac{2-3a}{2(1-a)} \\ 1 - \frac{\sqrt{2a(1-a)(1-c)}}{3a}, & \frac{2-3a}{2(1-a)} < c \end{cases} \end{aligned}$$



Above is the graph of $E(L|T > \alpha)$ as a function of c with $a = 0.05$ and $a = 0.4$ respectively. First, it is worth noting that the graph increases more gradually (with less sharp of a turn) for larger values of a. This means the smaller the role of luck, the longer the function is stable for, that is the higher threshold you can achieve (greater value of c) before you are expected to be significantly 'lucky'. In the case where $a = 0.05$, the expected luck score only reaches 0.6 when $T > 0.9561$ (roughly). However, to achieve highly competitive/challenging success like in the astronaut example, you are required to be somewhat lucky and expected to be incredibly lucky. When 11 astronauts are chosen from 18300 applicants, the applicants need to be in the top 0.06%, which in our analogy means $c = 0.9994$. This means that the minimum luck score you can have is $\frac{\alpha+a-1}{a} = 0.850$, and the expected luck score of the successful applicants are about 0.9497. In other words, very lucky. In fact, the expected luck score does not fall under 0.8 until $a < 0.003$.

Moreover, $F_T(x) = 1 - \frac{(1-x)^2}{2a(1-a)}$ when $x > 1 - a$. Hence, $f_T(x) = \frac{1-x}{a(1-a)}$, and $E(T|T > \alpha) = \frac{1}{P(T>\alpha)} \int_\alpha^1 \frac{x(1-x)}{a(1-a)} dx = \frac{2\alpha^3-3\alpha^2+1}{3(1-\alpha)^2}$ when $\alpha > 1-a$. Taking the numbers from the astronaut example, $c = 0.9994$ and $\alpha = F_T^{-1}(c) = 0.9925$, gives $E(T|T > \alpha) = 0.995$. This means that $E(H|T > \alpha) = \frac{E(T|T>\alpha)-aE(L|T>\alpha)}{1-a} = 0.9974$. So, unsurprisingly, the merit score of a successful individual is expected to be even higher than their luck score. However, a good way to capture the importance of luck is by noting the fact that $P(T > 0.9925|H = 0.9974) = P(L > 0.8994) = 0.1006$. In other words, if you were an astronaut applicant, even if you worked as hard as the average successful applicant, your chance about being successful is about 1 in 10.

Normal Distribution

Fortunately for the hard-working readers, the above conclusions rely on the assumption that both luck and merit follow a uniform distribution. While at first glance this seems reasonable, you might find it difficult to justify. Instead, we usually consider similar variables, e.g. IQ, to be normally distributed. It could make more sense if both the luck and merit of the population are distributed on a bell-shaped curve, with the majority closer to the population average. So, in this section, we will consider the scenario where $L, H \sim N(0, 1)$.

We maintain $T = aL + (1 - a)H$, where $0 < a \leq 0.5$ is the weight of luck in the total score. Further, c is, as before, the cut-off threshold, the proportion of the population a candidate must beat to be successful. Our objective is still finding $E(L|T > \alpha) = \int_{-\infty}^{\infty} x f_{L|T>\alpha}(x) dx$, where $f_{L|T>\alpha}(x) = \frac{d}{dx}(F_{L|T>\alpha}(x))$.

Because of the beautiful property that linear combinations of normal distributions are still normal, we have $T \sim N(0, a^2 + (1 - a)^2)$, hence $\frac{T}{\sqrt{a^2+(1-a)^2}} \sim N(0, 1)$. Therefore,

$$\begin{aligned} P(T > \alpha) &= 1 - P(T < \alpha) \\ &= 1 - P\left(\frac{T}{\sqrt{a^2+(1-a)^2}} < \frac{\alpha}{\sqrt{a^2+(1-a)^2}}\right) \\ &= 1 - \Phi\left(\frac{\alpha}{\sqrt{a^2+(1-a)^2}}\right) \end{aligned}$$

Φ being the cumulative distribution function of the standard normal distribution. Since, $P(T > \alpha) = 1 - c$ we can conclude $\alpha = \Phi^{-1}(c)\sqrt{a^2 + (1 - a)^2}$.

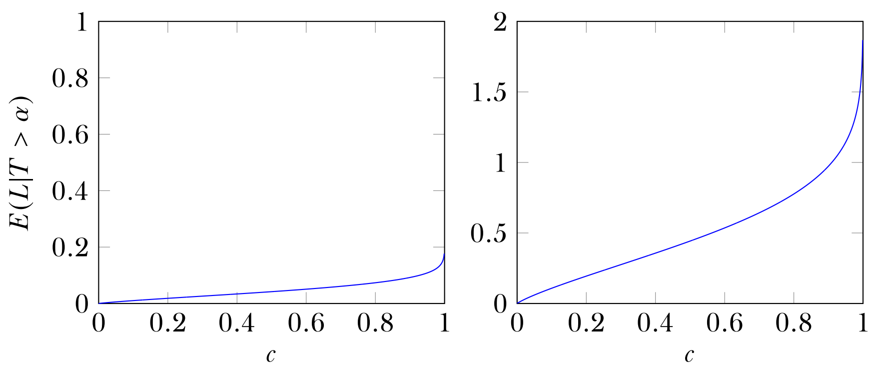
On the other hand, we have $P(T > \alpha, L < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{l^2}{2}} (1 - \Phi(\frac{\alpha - al}{1-a}))dl$, differentiating by x gives $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (1 - \Phi(\frac{\alpha - ax}{1-a}))$. A sanity check is to note that if $a = 0$, we get $P(T > \alpha, L < x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (1 - \Phi(\alpha))$, which is what we would expect since then T and L are independent and $T \sim N(0, 1)$. Another approach to find $P(T > \alpha, L < x)$ is to compute the covariance of T and L to derive the bivariate normal distribution function specific to this situation. In any case, we arrive at

$$f_{L|T>\alpha}(x) = \frac{1}{\sqrt{2\pi}(1-c)} e^{-\frac{x^2}{2}} \left(1 - \Phi(\frac{\alpha - ax}{1-a})\right)$$

Finally, we have

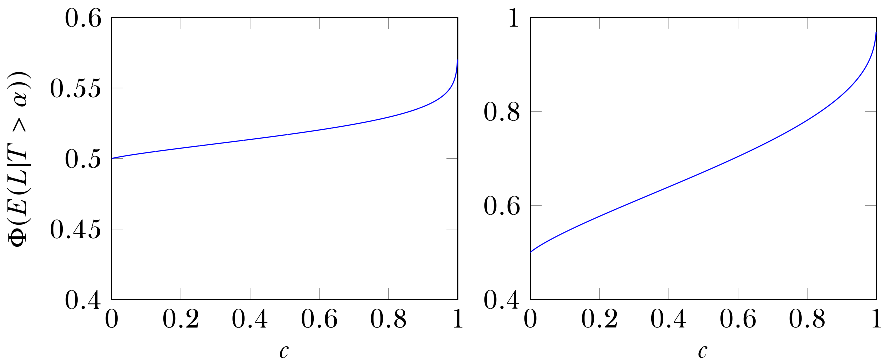
$$\begin{aligned} E(L|T > \alpha) &= \frac{1}{1-c} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(1 - \Phi(\frac{\alpha - ax}{1-a})\right) dx \\ &= \frac{1}{1-c} \frac{a}{\sqrt{2\pi(a^2+(1-a)^2)}} e^{-\frac{\Phi^{-1}(c)^2}{2}} \end{aligned}$$

The graph of this as a function of c when $a = 0.05$ and $a = 0.4$ are given below.



It may not be immediately obvious from the graphs and formula (as both numerator and denominator approach 0) but $E(L|T > \alpha) \to \infty$ as $c \to 1$ for all values of a. This can be shown using L'Hôpital's rule.

The above result tells us the expected value of L when $T > \alpha$ but what is the meaning of this value relative to the population? To find out, we can simply compute $\Phi(E(L|T > \alpha))$. The following are the graphs when $a = 0.05$ and $a = 0.4$ respectively.



Again, it is not clear in the first graph that for when $c = 1$, $\Phi(E(L|T > \alpha) = 1$, but we now know this to be the case.

Surprisingly, the somewhat extreme result in the previous section (when L and H are uniformly distributed) is not reproduced here. First of all, since L and H are unbounded, there is not minimum threshold for L (on a separate note, this is also the reason why the calculations for this case are much more straightforward). Moreover, going back to the astronaut example where $c = 0.9994$ and $a = 0.05$, we get $E(L|T > \alpha) = 0.1843$. Considering the fact that $L \sim N(0, 1)$, the value is quite low. In fact, $\Phi(E(L|T > \alpha)) = 0.5731$ which means that the chance of having a luck score higher than what is expected to be successful is as high as 0.4269, compared to 0.0503 when L and H are uniformly distributed.

Of course, when the weight of luck, a, is higher, $\Phi(E(L|T > \alpha))$ becomes close to 1 at much smaller values of c. This is not unexpected: if luck plays a large role in success, then you would expect successful people to be quite lucky.

Conclusion

Setting aside the two different models for the time being, there are two parameters that determine the expected luck of a successful individual: competitiveness (c) and weight of luck (a). To best interpret our results, I will split all situations into 4 categories: competitive and high weight of luck (1), competitive and low weight of luck (2), not competitive and high weight of luck (3), not competitive and low weight of luck (4).

Situation (1) is characterised by high values of c and a. In this case, both models more-or-less agree that a successful individual needs to have high luck score. This matches our intuition: consider a situation where a few people are selected from a vast group of applicants largely based on luck, of course those selected are fairly lucky.

Situation (3) is characterised by low c and high a. Again, both models agree that successful people are generally fairly lucky. While in this case success is not as selective, the high weight of luck means one would still expect successful individual to have higher than average luck score.

Situation (4) is characterised by low c and low a. Both models indicate that successful individuals do not need to have significantly higher than average luck, which is what we would expect. Considering applying to a low-barrier, entry-level job where the selection process is relatively deterministic (not much luck involved). Then, you wouldn't expect most successful applicants to have a noticeably high luck score.

The situation that I am most interested in and the situation where the two models show significant difference is situation (2), with high c and low a. This situation best reflects a lot of crucial processes in life, such as applications to highly competitive opportunities (university, career, etc) where the selection process is mostly merit-based. With a bit of imagination and hand-waving, it can also be used to estimate more abstract processes like starting a firm or searching for love. Depending on what you deem to be the more appropriate distribution for luck and merit, the expected luck of successful individuals varies greatly. If you believe that the distribution for luck and merit is uniform, then not only are you to expect successful people to be absurdly lucky, you can also conclude that a certain threshold of luck is required to be successful. On the other hand, if you believe that the distribution for luck and merit is normal, then there is no minimum luck score for successful individuals and whilst we do expect them to be luckier than average, the difference is not significant.

I would also like to make it clear that this paper is not meant as a comment or assertion about the reality of luck and its role in success, only the theoretical aspect. In fact, the result of this paper is not restricted to luck. We could have substituted any other factor instead. The main point is that a factor with a small weight can still potentially play a decisive role in a process. The magnitude of this phenomenon is dependent on the way the factor is distributed in the population. To apply the theory discussed in this paper to the real world would require deep understanding of the factors that determine success, their effects on each other, and their distribution in the population.

For those interested in the reality of the success paradox, I strongly recommend checking out the original YouTube video. It mentions some interesting statistical evidence. For example, it has been found that 40% of the hockey players in top leagues have a birthday in the first quarter of the year while only 10% have a birthday in the last quarter of the year. The presumed reason for this is that the cut-off date for kids hockey leagues is January 1st, which means kids with an early birthday would be up to a year older and thus more developed than the other kids in the league. This gives them more game time, attention from the coaches, etc, which makes them better players. The effect compounds through the years, resulting in the statistical disparity of birthdays in pro hockey league.

Another example is the eight Olympics world records in track and field (men's and women's 100m spring, 100m hurdle, long jump, and triple jump). When the records were set, in 7 of these occurrences, the athlete who set the record had a tailwind. While each of these athletes were of course extremely skilled in their respective field, a little luck helped them when setting the world records.

For further exploration, the book *Success and Luck: Good Fortune and the Myth of Meritocracy* by Robert H. Frank was credited as the inspiration of the original video and contains more fascinating examples.

Associative Metrics

Andres Klene

Throughout, the naturals include 0 (as they always should!).

Introduction

Groups and metric spaces are quite different mathematical objects in spirit; the former are abstract sets of symmetries, while the latter are sets with a rigid geometry. Of course, an object X can be both a metric space and a group at the same time (for example, X = a normed real vector space). However, the group operation and the metric are normally completely different functions. One one hand, a group operation is a binary operation $X \times X \to X$ which is associative, has an identity and has inverses — on the other, a metric is a positive definite, symmetric function $X \times X \to [0, \infty)$, which satisfies the triangle inequality.

There is one obvious similarity: both are functions from $X \times X$. Furthermore, if X is a subset of $[0, \infty)$ then the metric might corestrict[1] to a function $X \times X \to X$ (this is the case, for example, when $X = \mathbb{N}$ and the metric is the standard one). Given a set $X \subseteq [0, \infty)$, one can therefore reasonably ask: is there a binary operation $X \times X \to X$ which satisfies both the group axioms and the metric axioms?

It will turn out that associative metrics are group operations, so an equivalent question is: *does there exist an associative metric on X?* The late mathematician John Conway asked this in the case $X = [0, \infty)$; we will answer his question in due time.

Associative metrics are group operations

As the name suggests, an *associative metric* on X is a function $d\colon X \times X \to X$ which satisfies the metric axioms and is associative, i.e. we have

$$d(x, d(y, z)) = d(d(x, y), z)$$

[1]Corestricting refers to restricting the codomain. If $f\colon A \to B$ is a function and $f(A) \subseteq C$, then $f|^C\colon A \to C$ is the same mapping with a smaller codomain.

for all $x, y, z \in X$. This looks a bit clunky; to avoid writing things in terms of nested functions, we will use infix notation $x*y = d(x, y)$ (as is common for binary operations in general). The above condition now looks more familiar:

(A0) associativity: $x * (y * z) = (x * y) * z$.

To get comfortable with this notation, here are the metric axioms. For all $x, y, z \in X$,

(M1) definiteness: $x * y = 0 \iff x = y$,

(M2) symmetry: $x * y = y * x$,

(M3) the triangle inequality: $x * z \leq x * y + y * z$.

First we quickly show that adding associativity to the metric axioms is enough to guarantee a group structure.

Proof. Associativity is given, so we just need to check the existence of an identity and inverses. Using **(A0)** and **(M1)**,

$$0 \overset{\textbf{(M1)}}{=} 0 * 0 \overset{\textbf{(M1)}}{=} 0 * (x * x) \overset{\textbf{(A0)}}{=} (0 * x) * x \overset{\textbf{(M1)}}{\Longrightarrow} 0 * x = x.$$

So 0 is the identity element. Since $0 = x * x$, every element is self-inverse. □

This result is useful, but the discusson has not yet addressed an obvious concern: do associative metrics even exist?

Examples and non-examples

Most metrics are not associative. For example, consider the standard metric on $\mathbb{N}$: we have

$$\big||1-2|-3\big| = |1-3| = 2,$$

but

$$\big|1-|2-3|\big| = |1-1| = 0.$$

The discrete metric on $\mathbb{N}$, given by

$$(x, y) \mapsto \begin{cases} 0 & x = y \\ 1 & x \neq y, \end{cases}$$

is not associative either, for almost exactly the same reason. There aren't a lot of other metrics one might immediately try, so let's back up and look at finite subsets of $\mathbb{N}$ first. Note that since 0 is the identity, any set which admits an associative metric must contain 0.

($|X| = 1$) The case $X = \{0\}$ is trivial. There is exactly one binary operation on X; if you are really bored, you are invited to check that it satisfies both the group axioms and the metric axioms. Here is the (sad-looking) Cayley table:

$*$	0
0	0

($|X| = 2$) The case $X = \{0, 1\}$ is similar: there are four binary operations, but only one metric, which happens to be associative. The Cayley table is

$*$	0	1
0	0	1
1	1	0

and we can see $X \cong \mathbb{Z}_2$.

($|X| = 3$) Recall that 0 is the identity, and every element has order 2. This forces the following Cayley table, which leads to a contradiction.

$*$	0	1	2
0	0	1	2
1	1	0	
2	2		0

$\rightsquigarrow \#.$

Moreover, this contradiction does not depend on the actual values 1 and 2; every set with 3 elements does not admit an associative metric.

($|X| = 4$) The case $X = \{0, 1, 2, 3\}$ is not as immediate. By the same reasoning as above, the following Cayley table is forced.

$*$	0	1	2	3
0	0	1	2	3
1	1	0		
2	2		0	
3	3			0

$\rightsquigarrow$

$*$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

This operation satisfies the triangle inequality, so it is the unique associative metric on $\{0, 1, 2, 3\}$. Additionally, we can see $X \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

The last example highlights a subtle issue, which is that X cannot be too "unevenly spaced". For example, the set $\{0, 1, 2, 10^{100}\}$ does *not* admit an associative metric, because the group law would force $1 * 2 = 10^{100}$ which does not satisfy the triangle inequality $1 * 2 \leq 1 * 0 + 0 * 2$. So in general there is some dependence on the actual values in the set, other than its cardinality.

Metrics on infinite sets

What links the three examples that worked? Surprisingly, they are all restrictions of the same function — note how each Cayley table is contained in the next. It turns out that this function is the binary bitwise exclusive-or operation (denoted XOR), which is defined as follows: given two natural numbers, write them in binary and perform the logical exclusive-or operation on each pair of bits. In the following examples, a subscript of 2 indicates a number is in binary, and no subscript indicates it is in base-10.

$$\begin{array}{rrl} & 1_2 & = 1 \\ \text{XOR} & 1_2 & = 1 \\ \hline & 0_2 & = 0 \end{array} \qquad \begin{array}{rrl} & 10_2 & = 2 \\ \text{XOR} & 11_2 & = 3 \\ \hline & 1_2 & = 1 \end{array} \qquad \begin{array}{rrl} & 111_2 & = 7 \\ \text{XOR} & 101_2 & = 5 \\ \hline & 10_2 & = 2 \end{array}$$

Hence 1 XOR $1 = 0$ (anything XOR itself is 0), 2 XOR $3 = 1$ and 7 XOR $5 = 2$.

This idea works on more than just the naturals. Denote by $\mathbb{N}[\frac{1}{2}]$ the set of non-negative fractions with denominator equal to a power of 2. In other words, these are precisely the numbers with finite binary expansions. There is an obvious extension of XOR from $\mathbb{N}$ to $\mathbb{N}[\frac{1}{2}]$, which is also given by comparing bits in the binary representation. For example:

$$\begin{array}{rrl} & 10.1_2 & = 2.5 \\ \text{XOR} & 11.1_2 & = 3.5 \\ \hline & 1.0_2 & = 1 \end{array} \qquad \begin{array}{rrl} & 111.10_2 & = 7.5 \\ \text{XOR} & 101.11_2 & = 5.75 \\ \hline & 10.01_2 & = 2.25 \end{array}$$

Here is the point: XOR is an associative metric on both $\mathbb{N}$ and $\mathbb{N}[\frac{1}{2}]$! Proving this is just checking axioms, so...

Exercise. *Check that* XOR *is an associative metric on* $\mathbb{N}[\frac{1}{2}]$.

Now we are in a position where Conway's question — "does there exist an associative metric on $[0, \infty)$?" — seems like it is within our grasp. After all, real numbers also have binary representations. Generically the representations will be infinite, but the same principle might still apply. So, does XOR extend further from $\mathbb{N}[\frac{1}{2}]$ to $[0, \infty)$?

As it turns out, no. The obstacle is that when we allow infinite binary strings, representations of elements in $\mathbb{N}[\frac{1}{2}]$ are no longer unique (for example, $0.\dot{1} = 1$). Naively using bitwise XOR then suggests

$$1_2 \text{ XOR } 1_2 = 0.\dot{1}_2 \text{ XOR } 1_2 = 1.\dot{1}_2 = 2,$$

despite the fact that we expect 1_2 XOR $1_2 = 0$. We could try to fix this by artificially selecting which representations to pick, but this fails as well. Suppose we select only the finite ones, so that we don't allow expansions with a tail of 1s (so for example, $0.\dot{1}$ is not allowed). Then the expansions for $1/3$ and $2/3$ are $0.\dot{0}\dot{1}$ and $0.\dot{1}\dot{0}$ respectively, which become $0.\dot{1}$ when XOR'd together. On the other hand, if we don't allow tails of 0's then $0.\dot{1}$ XOR $1.\dot{1} = 1.0$ which is not allowed. In short, whichever set of allowable representations we pick, XOR is not closed as a binary operation on those representations. It seems all hope is lost...

However, in 2010 it was shown (on a MathOverflow thread) that there actually *does* exist an associative metric on $[0, \infty)$. Unfortunately, this is purely an existence result; the proof is completely non-constructive and relies on the Axiom of Choice[2]. This means we can't write down an explicit expression or description for this function in the same way that we can for XOR on $\mathbb{N}$ and $\mathbb{N}[\frac{1}{2}]$.

So, Conway's question has been answered, but the non-explicitness makes the solution a bit disappointing. Luckily, there are still ways of explicitly constructing associative metrics. The next section describes one such method.

[2]They use a transfinite induction argument, after well-ordering $[0, \infty)$. We refer the interested reader to `https://mathoverflow.net/questions/16214/is-there-an-associative-metric-on-the-non-negative-reals`.

More associative metric spaces via the symmetric difference

Another way of thinking about binary representation is as follows: every finite subset of $\{2^n : n \in \mathbb{Z}\}$ can be uniquely identified with a number in $\mathbb{N}[\frac{1}{2}]$ given by adding together all the elements of the set. For example,

$$\{2^{-1}, 1, 2, 4\} \mapsto 2^{-1} + 1 + 2 + 4 = 7.5.$$

This means there is a bijective correspondence $\sum\colon \mathscr{P}_{\text{finite}}(\{2^n : n \in \mathbb{Z}\}) \to \mathbb{N}[\frac{1}{2}]$ called 'sum the elements'. (Here, $\mathscr{P}_{\text{finite}}$ denotes the collection of finite subsets. This notation follows closely that for the collection of *all* subsets, which is often denoted by $\mathscr{P}$ for 'power set').

Now, here is the key idea: when pulled back under this bijection, the XOR operation becomes the *symmetric difference* $\triangle$, defined by $A \triangle B := (A \cup B) \setminus (A \cap B)$. Explicitly, we have

$$\textstyle\sum(A) \text{ XOR } \sum(B) = \sum(A \triangle B)$$

for all $A, B \in \mathscr{P}_{\text{finite}}(\{2^n : n \in \mathbb{Z}\})$. For example,

$$\begin{array}{ccccc} \{2^{-1}, 1, 2, 4\} & \triangle & \{4^{-1}, 2^{-1}, 1, 4\} & = & \{4^{-1}, 2\} \\ \downarrow{\scriptstyle\Sigma} & & \downarrow{\scriptstyle\Sigma} & & \downarrow{\scriptstyle\Sigma} \\ 7.5 & \text{XOR} & 5.75 & = & 2.25 \end{array}$$

It is worth thinking about exactly what's going on here. On each pair of binary bits, the XOR operation outputs 1 iff exactly one of them is 1; otherwise, it outputs 0. This corresponds exactly to the "union minus intersection" behaviour of $\triangle$. In this sense, the symmetric difference is the exclusive-or analogue in the world of sets.

This perspective lets us work backwards, with the purpose of defining associative metrics on more general sets. Suppose we are given a set $\mathscr{B} \subset [0, \infty)$ with the property that all finite sums of subsets of B are unique. What this means is that

$$\textstyle\sum(A) = \sum(B) \iff A = B \quad \text{for all } A, B \in \mathscr{P}_{\text{finite}}(\mathscr{B}) \tag{$\star$}$$

Write $span(\mathscr{B}) := \{\sum(A) : A \in \mathscr{P}_{\text{finite}}(\mathscr{B})\}$ for the collection of all sums of finite subsets of $\mathscr{B}$, and define $*$ to be the unique binary operation satisfying the following identity:

$$\sum(A) * \sum(B) = \sum(A \triangle B).$$

The algebraically inclined reader might enjoy the corresponding commutative diagram:

$$\begin{tikzcd} \mathscr{P}_{\text{finite}}(\mathscr{B})^2 \arrow[r, "\triangle"] \arrow[d, "\Sigma^2"] & \mathscr{P}_{\text{finite}}(\mathscr{B}) \arrow[d, "\Sigma"] \\ span(\mathscr{B})^2 \arrow[r, "*"] & span(\mathscr{B}) \end{tikzcd}$$

You might be able to guess what we claim: that $*$ is an associative metric on $span(\mathscr{B})$.

Proof of the claim. Associativity follows from the associativity of $\triangle$. It remains to show the metric axioms hold. For all $A, B \in \mathscr{P}_{\text{finite}}(\mathscr{B})$,

(M1): We have $\sum(A) * \sum(B) = 0 \iff \sum(A \triangle B) = 0 \iff A \triangle B = \varnothing \iff A = B \iff \sum(A) = \sum(B)$, so **(M1)** holds.

(M2): Symmetry holds by commutativity of $\triangle$.

(M3): $\sum(A) * \sum(B) = \sum(A \triangle B) \leq \sum(A) + \sum(B)$. This is enough because $x * z \leq x + z$ implies $x * z = x * y * y * z \leq x * y + y * z$, which is **(M3)**.

So, $*$ is an associative metric on $span(\mathscr{B})$. $\square$

When $\mathscr{B} = \{2^n : n \in \mathbb{Z}\}$, we recover the motivating example, where $span(\mathscr{B}) = \mathbb{N}[\frac{1}{2}]$ and $* = \text{XOR}$. For a novel example, take $\mathscr{B} = \{2^q : q \in \mathbb{Q}\}$ instead. It is not immediately clear that this choice of $\mathscr{B}$ satisfies ($\star$); we leave this as an exercise[3]. The construction outlined above automatically defines an explicit associative metric on $span(\mathscr{B})$, which can alternatively be expressed as

$$span(\mathscr{B}) = \mathbb{N}[\tfrac{1}{2}, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt[3]{2}}, \ldots].$$

[3]Solution: https://mathoverflow.net/questions/382784/representing-finite-sums-of-rational-powers-of-2

Conclusion

I want to leave you with something to think about. In general, we know how to define an associative metric on a set X, given that X is of the form $span(\mathscr{B})$ for some $\mathscr{B}$ with property ($\star$). Try to come up with your own set which is of this form. What does the associative metric look like? Can you find such a set which is uncountable? Can you prove that the set of non-negative rationals $\mathbb{Q}_{\geq 0}$ is of this form? What about $[0, \infty)$?

Optimal dice for the game of Ludo

Max French

Introduction

For most dice-based board games, the number of sides on the dice is negatively correlated with the mean number of moves in a game. Ludo is an exception to this rule: the rules of Ludo disadvantage both small and large numbers of sides on the dice. This means that the question of what dice minimises the number of throws to finish a game of Ludo does not have a trivial answer.

The Rules of Ludo

Ludo is a dice-based board game played by up to four players. Each player has four pieces and the game ends when one player completes a lap of the board with all of their pieces. Pieces start out of play in their player's "yard" and can only enter the board when the player rolls a maximum score. The piece is then placed onto that player's starting square and has to travel exactly 44 squares to finish, where each roll of the dice determines how many squares the piece must move.

For ease of calculation I consider a single player playing with a single piece. This is done with a series of three increasingly accurate models.

Model One

In model one we calculate the mean number of rolls of an x-sided dice for a piece to travel 44 tiles.

Let R_a be the mean roll for a dice with x sides. The dice roll has uniform distribution over {1,2,...,x}. Hence R_a is simply the expectation of the uniform distribution and we have

$$R_a = \frac{x+1}{2}.$$

Let S be the square that the piece finishes on. Then S can take any integer value s_1 where $44 \leq s_1 \leq 43 + x$ and for each s_1 there are $k = x + 44 - s_1$ possible final dice rolls. We can find the probability of finishing on some square s_1, $\mathbb{P}(S = s_1)$, by taking a weighted average, so

$$\mathbb{P}(S = s_1) = \frac{k}{\sum_{s_1=44}^{43+x} k}.$$

We can simplify $\sum_{s_1=44}^{43+x} k$. We know that $k = x + 44 - s_1$, so

$$\sum_{s_1=44}^{43+x} k = \sum_{s_1=44}^{43+x} x + 44 - s_1 = \sum_{i=0}^{x-1} x - i = \sum_{j=1}^{x} j = \frac{x(x+1)}{2}.$$

Therefore

$$\mathbb{P}(S = s_1) = \frac{x + 44 - s_1}{\left(\frac{x(x+1)}{2}\right)} = \frac{2x + 88 - 2s_1}{x(x+1)}.$$

Let R_{s_1} be the mean number of rolls to finish on the square s_1. Then

$$R_{s_1} = \frac{s_1}{R_a} = \frac{s_1}{\left(\frac{x+1}{2}\right)} = \frac{2s_1}{x+1}.$$

Finally, let y be the mean number of rolls for a single piece to finish. Then

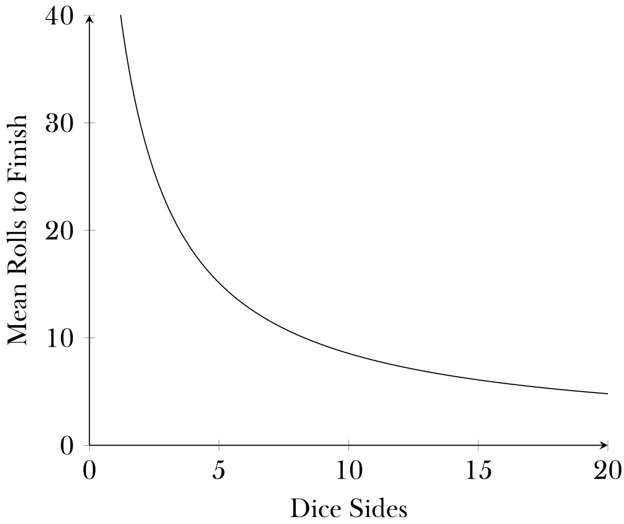
$$y = \sum_{s_1=44}^{43+x} \mathbb{P}(S = s_1) R_{s_1} = \sum_{s_1=44}^{43+x} \left(\frac{2x + 88 - 2s_1}{x(x+1)}\right)\left(\frac{2s_1}{x+1}\right).$$

This can be simplified to

$$y = \frac{4}{x(x+1)^2} \sum_{s_1=44}^{43+x} s_1(x + 44 - s_1).$$

This equation is plotted in Figure 1. As would be expected, the mean number of rolls to finish is inversely proportional to the number of sides on the dice.

Figure 1



Model Two

In model two we introduce the rule that a maximum score must be rolled for a piece to enter play.

Let R_b be the average number of rolls before a maximum score is rolled. The number of rolls before a maximum score is rolled has geometric distribution with parameter $1/x$. Thus R_b is simply the expectation of this distribution and we have

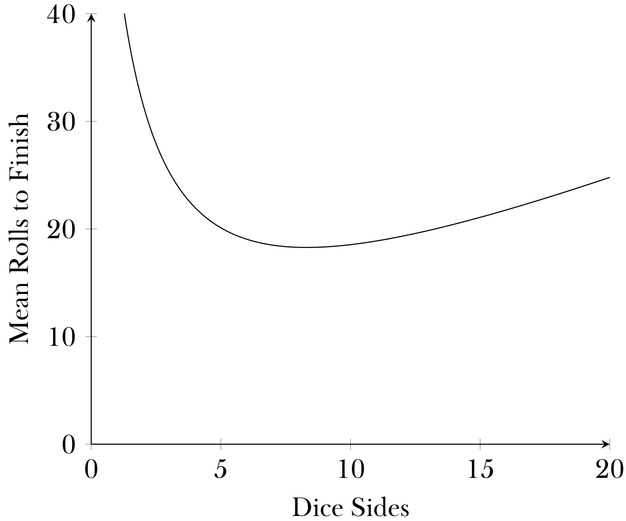
$$R_b = \frac{1}{\left(\frac{1}{x}\right)} = x.$$

The mean number of rolls to finish is the sum of the mean number of rolls for a piece to enter play and the mean number of rolls to travel at least 44 tiles. Therefore

$$y = x + \frac{4}{x(x+1)^2} \sum_{s_1=44}^{43+x} s_1(x + 44 - s_1).$$

This equation is plotted in Figure 2. It has a distinct minimum, and for an integer number of dice sides the minimum number of mean rolls occurs at $x = 8$.

Figure 2



Model Three

In model three we introduce the rule that a piece must finish exactly on tile 44.

For model one we calculated the mean number of moves for a piece to first reach some square s_1 with $44 \leq s_1 \leq 43+x$. We can use the same method to find the mean number of rolls for a piece to first reach some square s_2 with $44 - x \leq s_2 \leq 43$. Let us call this number R_c. Then we have

$$R_c = \frac{4}{x(x+1)^2} \sum_{s_2=44-x}^{43} s_2(44 - s_2).$$

Once the piece has entered play and first reached some square s_2 there are three possibilities for the value of each roll, r:

1. $r = 44 - s_2$
2. $r > 44 - s_2$
3. $r < 44 - s_2$

The probability of the first case is $\frac{1}{x}$, and if the first case occurs then the piece finishes. If the second case occurs then s_2 remains constant and therefore the probability of the first case on the next roll remains $\frac{1}{x}$. If the third case occurs then s_2 increases but the probability of the first case on the next roll remains $\frac{1}{x}$.

Let R_d be the mean number of rolls before $44 - s_2$ is rolled. The probability of rolling $44 - s_2$ is the same as the probability of rolling a maximum score so

$$R_d = R_b = x.$$

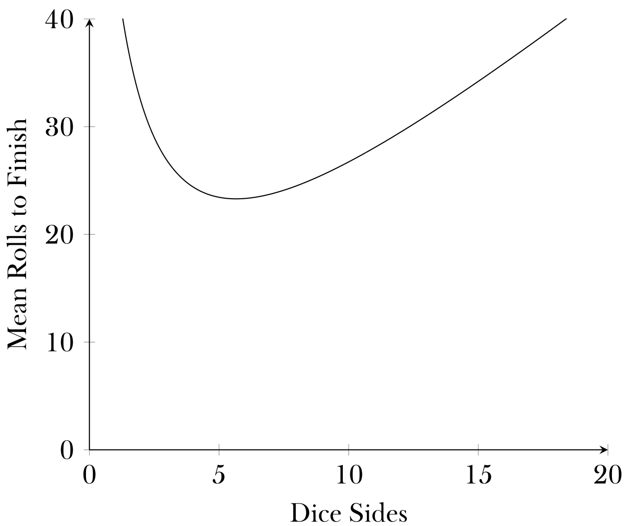
Therefore the mean number of rolls for a piece to finish after first reaching some square s_2 is x.

The mean number of rolls to finish is the sum of the mean number of rolls for a piece to enter play, the mean number of rolls for a piece to first reach some square s_2 with $44 - x \leq s_2 \leq 43$, and the mean number of rolls for a piece to finish after first reaching some square s_2. Therefore

$$y = 2x + \frac{4}{x(x+1)^2} \sum_{s_2=44-x}^{43} s_2(44 - s_2).$$

This equation is plotted in Figure 3. Again there is a distinct minima, and for an integer number of dice sides the minimum mean number of rolls occurs at $x = 6$.

Figure 3



Predictions and Conclusions

A perfectly accurate model would require the introduction of multiple pieces and players. Finding the optimal n-sided dice would then necessitate consideration of different strategies, and so this is not explored here.

For the simplified model considered here, the move optimal number of sides is 6. I predict that in a perfect model this number would increase. After the first piece has entered the game the next three pieces would take fewer "wasted" moves on average to enter the game as non-maximum rolls now have utility in moving a piece already on the board, and so the advantage of dice with fewer sides is somewhat negated. Similarly,the introduction of more players would also favour dice with more sides. Pieces that spend less time on the board have a lower probability of being taken, and so dice with fewer sides are put at a disadvantage.

Reflection Groups and Uniform Polytopes

Gavin Jared Bala

Introduction

The dihedral group $I_2(n)$ and the symmetric group S_n are two famous families of groups that are usually described differently: the dihedral group as the symmetries of a regular n-gon, and the symmetric group as the permutations of the numbers 1 up to n.

But the symmetric group has a natural geometric meaning too. $I_2(3) \cong S_3$, as the symmetries of an equilateral triangle take vertices to vertices, and since the vertices are all equidistant, we can realise every possible permutation of them. This generalises to higher dimensions: the symmetric group S_n is the symmetry group of a regular $(n-1)$-simplex, as its vertices are all equidistant. This can be proved by induction: choose where to send one vertex, then note that an n-simplex is a pyramid based on an $(n-1)$-simplex.

Furthermore, both groups are built out of reflections. The symmetries of the n-gon are rotations and reflections, and successively reflecting in two axes an angle $\alpha/2$ apart gives a rotation by α:

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

Since we can rotate the polygon to make any reflection the one described by $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, taking $\alpha = 2\pi/n$ lets these two matrices generate $I_2(n)$.[1] [6, p. 5]

For the symmetric group, note that (composing permutations right-to-left):

$$(12)(23)(34)\cdots(n-1\ n) = (1234\cdots n)$$

Swapping out the numbers to deal with any cycle, we build up all elements of the symmetric group from the transpositions (2-cycles). These are indeed reflections when we interpret S_n as simplex symmetries!

Reflection groups therefore seem important, so we will analyse them.

[1]This interpretation justifies writing $I_2(1) = \mathbf{Z}/2\mathbf{Z}$ and $I_2(2) = \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, the Klein-four group, even though there aren't regular 1-gons and 2-gons on the plane.

Presentations

Take two reflections in our group, R_α and R_β. (If there's only one, the group is trivially $\mathbf{Z}/2\mathbf{Z}$.) In n dimensions, a reflection leaves an $(n-1)$-plane constant, and switches the half-spaces to either side.

If the group is discrete (that is, it cannot have elements doing arbitrarily close things), the angle between the two hyperplanes of reflection can only be zero (if they are parallel) or a rational multiple of π. (Otherwise their product would create rotations of an irrational fraction of a turn.) As above, *which* multiple it is can be seen from the order of $R_\alpha R_\beta$:

$$R_\alpha^2 = R_\beta^2 = (R_\alpha R_\beta)^{o(\alpha,\beta)} = 1 \qquad o(\alpha,\beta) \in \mathbf{N} \cup \{\infty\}$$

Ernst Witt (1911-1991) proved that *the $o(\alpha,\beta)$ determine the group*, as follows:[2] [2]

Consider not only the hyperplanes of our reflections, but also where they're sent to by all the other reflections. These create a kaleidoscope partitioning hyperspace into a number of congruent regions, and we can take the generating reflections to be those bounding any particular region. Let's mark one region and call it $\mathbf{W}$.

We can then get from $\mathbf{W}$ to any of its neighbouring regions by reflecting in one of its boundary hyperplanes: call $\mathbf{W}R_\alpha$ the region we get to from reflecting in R_α. But $\mathbf{W}R_\alpha$ is congruent to $\mathbf{W}$, and its boundary reflections equally well generate the group, so we can continue this procedure to get to $\mathbf{W}R_\alpha R_\beta$; to $\mathbf{W}R_\alpha R_\beta R_\gamma$; and so on. This creates a path along which we travel from region to region by the boundaries named by the subscripts.

Suppose we had some relation *not* accounted for by the above. As every element is self-inverse, we can move everything to one side: $R_\alpha R_\beta \cdots R_\nu = 1$. That means that $\mathbf{W}R_\alpha R_\beta \cdots R_\nu = \mathbf{W}$, and our path is a loop.

Contract this loop continuously to a point. There are only two things that can happen while we do that. We eliminate dead ends - going into a region and then back out - by the relation $R_\alpha^2 = 1$. And we eliminate vertex-crossings by the relation $(R_\alpha R_\beta)^{o(\alpha,\beta)} = 1$. So those two relation types must generate the group.

[2] If $o(\alpha,\beta) = 1$, then $R_\alpha = R_\beta$ and we can drop one of them from the generators; and if $o(\alpha,\beta) = 2$, then R_α and R_β commute and the group is a direct product. We already saw an example with $I_2(2) = \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z} = I_2(1) \times I_2(1)$.

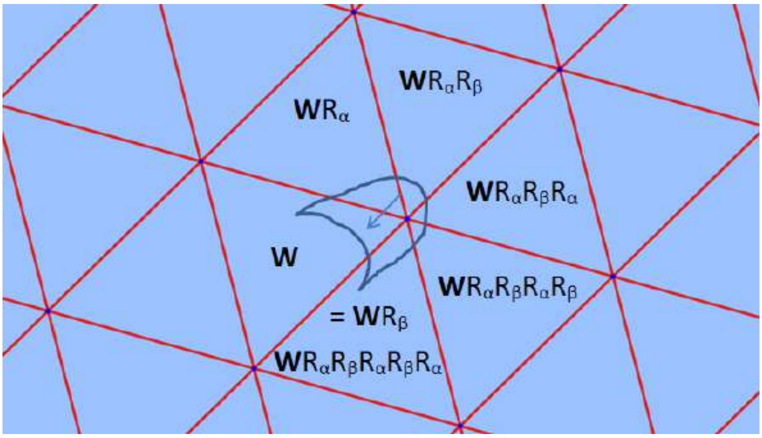


Figure 1: Moving the loop past a vertex invokes the relation $(R_\alpha R_\beta)^{o(\alpha,\beta)} = 1$. Here we illustrate the situation for the group $\tilde{A}_2$, generated by three reflections in the sides of an equilateral triangle: $R_\alpha^2 = R_\beta^2 = R_\gamma^2 = (R_\alpha R_\beta)^3 = (R_\beta R_\gamma)^3 = (R_\alpha R_\gamma)^3 = 1$.

Classification

Linear algebra lets us classify which of the above groups can be exhibited as finite reflection groups in Euclidean space. If we take the normals $\mathbf{n}_i$ to the mirrors generating a finite reflection group, we can define a matrix $\mathbf{A}$ by

$$A_{ij} = \mathbf{n}_i \cdot \mathbf{n}_j = -\cos\frac{\pi}{o(i,j)}$$

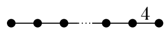
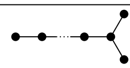
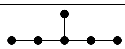
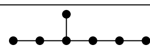
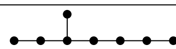
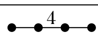
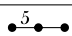
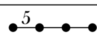
Taking that last expression, we could define a matrix

$$A_{ij} = -\cos\frac{\pi}{o(i,j)}$$

for *any* set of values of $o(i,j)$. Euclidean space admits a positive-definite dot product, i.e. $\mathbf{x}\cdot\mathbf{x} > 0$ when $\mathbf{x} \neq 0$. If the group really can be a finite Euclidean reflection group, then the bilinear form we get from it should be positive-definite too, i.e. $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ when $\mathbf{x} \neq 0$. So the classification boils down to computing eigenvalues. [6, p. 31] [8]

For convenience we introduce here the **Coxeter-Dynkin diagram** to graphically describe the cases that result. This is a graph with one vertex for each generating reflection. If $o(i,j) \geq 3$, we draw an edge between those two vertices, labelling it with $o(i,j)$; since the mark 3 is quite common, it is usually left off. If $o(i,j) = 2$, we leave them disconnected.[3]

The finite reflection groups are then built up as direct products of some building blocks. Those come in four infinite classes and six sporadic cases.[4]

A_n	$B_n = C_n$ 	D_n 
E_6 	E_7 	E_8 
F_4 	H_3 	H_4 
	$I_2(n)$ 	

Most of these are the symmetry groups of regular polytopes. We've already seen A_n as the symmetric group S_{n+1}, which is the symmetry group of the n-simplex;[5] and $I_2(n)$ is the dihedral group, the symmetry of the n-gon. $B_n = C_n$ is the symmetry group of the n-cube.[6] H_3 is the symmetry of the dodecahedron, and H_4 that of its four-dimensional counterpart, the 120-cell. F_4 is the symmetry of the 24-cell, a four-dimensional regular polytope that has no precise analogue in higher or lower dimensions.

The branched Coxeter diagrams don't represent regular polytopes. D_n contains half the symmetries of B_n. It is the symmetry group of the *n-demicube*, which is what you get if you keep every other vertex of the n-cube and delete the others.

The high-dimensional sporadic groups E_6, E_7, and E_8 are difficult to describe, as they lack good low-dimensional analogues. E_8 is in some way the octonions' answer to the Gaussian integers on the complex plane: the units from a natural set of densely packed integers on which unique factorisation theorems hold. (The quaternions' answer is F_4; we could view the Gaussian integers as coming from B_2.) [1, p. 99]

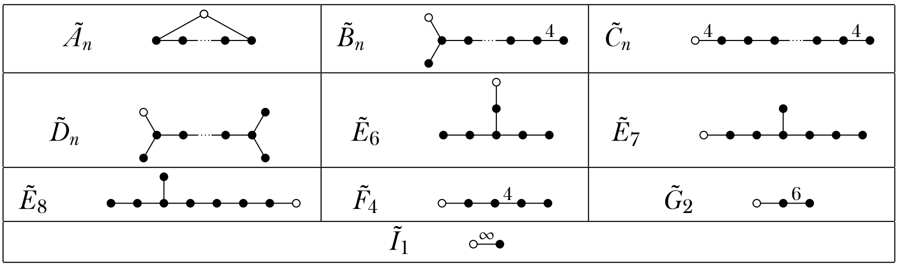
[3]So when a graph breaks into multiple components, the group it represents is a direct product.

[4]The odd notation, with the letter G missing and two letters for one family, comes from Lie theory. G_2 is an alternative name for $I_2(6)$.

[5]Since in this case we've seen the generators, we can confirm the Coxeter-Dynkin diagram directly: two transpositions with no elements in common commute, but when they have a common element, their product is a 3-cycle.

[6]This is also the **signed symmetric group**: the set of permutations π of $\{-n, -(n-1), \cdots, n-1, n\}$ such that $\pi(-i) = -\pi(i)$ for all i. If we have a unit cube centred on the origin, we can permute the axes, and choose an orientation for each one, arbitrarily. [3]

If the matrix **A** is only positive, i.e. $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$, then the group becomes infinite, but still fits in Euclidean space as the symmetry group of a tessellation. The classification is quite similar, except that the groups from polytopes that don't tessellate space disappear: the old diagrams are usually extended by a node, giving a tiling of space by the original polytope (and possibly some others). $\tilde{I}_1$ is the symmetry group of an *apeirogon*, an infinite-sided regular polygon: a line divided into equal-length segments.



If $\mathbf{x}^T\mathbf{A}\mathbf{x}$ can be negative, the group won't fit in Euclidean space. For small cases, it can fit in hyperbolic space; but as the $o(i, j)$ rise higher and higher (and possibly to infinity), the vertices and the fundamental region break through infinity and go beyond it. In general, Coxeter groups can be classified based on the number of positive, zero, and negative eigenvalues **A** has. A hyperbolic Coxeter group naturally fits in Lorentzian space, where one (timelike) basis vector has negative norm, and the others (spacelike) have positive norm. [8]

Other applications

The Coxeter-Dynkin diagram closely resembles the **Schläfli symbol** for regular polytopes. The Schläfli symbol of a p-gon is $\{p\}$, and then it is defined inductively: a regular polyhedron where q p-gons surround each vertex has Schläfli symbol $\{p, q\}$, a regular polychoron where r $\{p, q\}$ surround each edge has Schläfli symbol $\{p, q, r\}$, and so on. And indeed the cube has Schläfli symbol {4,3}, the dodecahedron {5,3}, the 24-cell {3,4,3}, and so on.

This becomes clearer if we consider not the fundamental reflections, but rather the rotations that they multiply to. A cube has fourfold rotation symmetries in each face, because its faces are squares; and threefold rotation symmetry around each vertex, because three squares meet at a corner. (It also has twofold rotation symmetry around each edge.)

This relates to **Wythoff's construction**. Consider the Coxeter diagram that is a triangle with edges marked p, q, and r. Generating reflections are in the sides of a triangle with angles π/p, π/q, and π/r: that could be spherical, Euclidean, or hyperbolic.

Put a generating point in that triangle, and reflect it across the mirrors and their images to fill all of space. By connecting the images appropriately, we obtain a series of seven **uniform** tilings, like the Archimedean solids: all faces are regular polygons (if the generator point is placed appropriately), and the symmetries of the tiling act transitively on the vertices (the vertices are equivalent and thus surrounded alike).

This can naturally be generalised to larger polygons and higher dimensions. One should nevertheless not imagine that the classification is complete even for two-dimensional tilings, because it is not: the Archimedean snub cube and snub dodecahedron cannot be obtained this way, as they do not have the full symmetries of the cube and dodecahedron. They are at least relatively tame examples of failure of the Wythoff construction, as they can be obtained by taking alternate vertices of the even-faced polyhedra in the last column below, which *are* Wythoffian. A worse example exists on the Euclidean plane: the tiling by alternating strips of triangles and squares is uniform, but it cannot be obtained by Wythoff's construction, even applying an alternation procedure after the fact. Actually, the complete listing of uniform tilings is not known except in the cases of $\mathbf{S}^2$, $\mathbf{E}^2$, and $\mathbf{S}^3$! [4]

The classification of reflection groups isn't just important for geometry. One can see it crop up again in Lie theory, catastrophe theory, and even even the symmetries of tiny water droplets. (Yes, even the E_n have a connexion to our three-dimensional world!) [5] We have only scratched the surface!

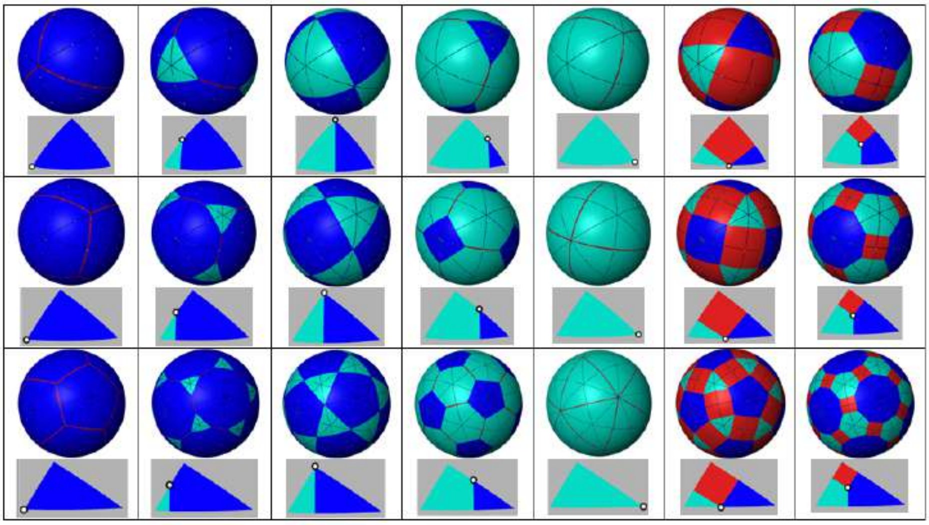


Figure 2: A family portrait of the Wythoffian Platonic and Archimedean solids. Each triangle in the spherical tiling matches the colouring of the generating triangle. The lines dividing the fundamental triangle are perpendiculars dropped from the generating point. The rows are respectively for $(p, q, r) = (3, 3, 2)$ (tetrahedral), $(4, 3, 2)$ (cubic), and $(5, 3, 2)$ (dodecahedral). Made using Jeffrey Weeks' *KaleidoTile* software. [7]

Bibliography

[1] Conway, J. H. & Smith, D. A. (2003). On quaternions and octonions: their geometry, arithmetic, and symmetry. CRC Press.

[2] Coxeter, H. S. M. (1940). Regular and semi-regular polytopes. I. *Mathematische Zeitschrift*, *46*(1), 380–407. https://doi.org/10.1007/bf01181449

[3] Coxeter, H. S. M. (1985). Regular and semi-regular polytopes. II. *Mathematische Zeitschrift*, *188*(4), 559–591. https://doi.org/10.1007/bf01161657

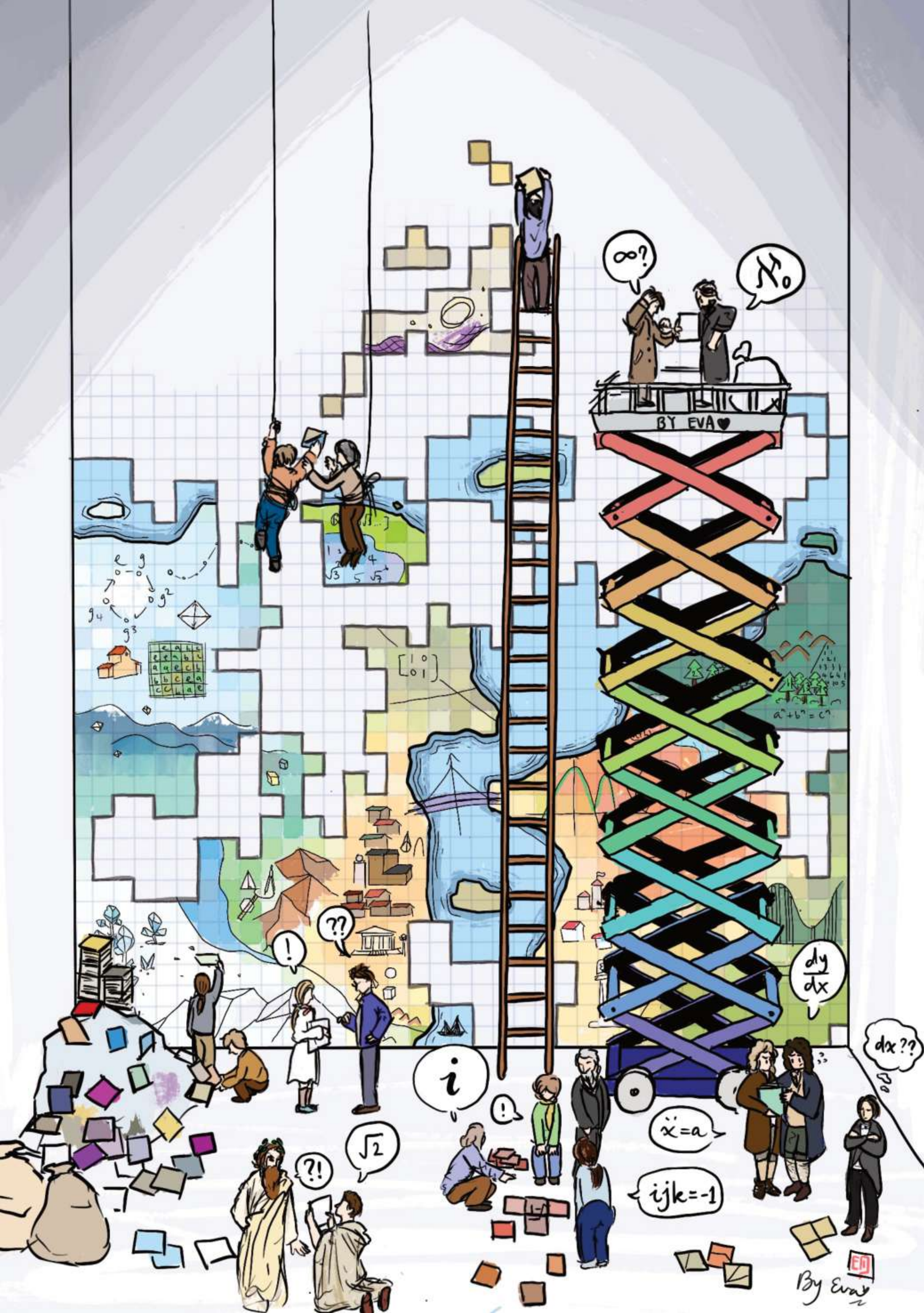
[4] Coxeter, H. S. M. (1988). Regular and semi-regular polytopes. III. *Mathematische Zeitschrift*, *200*(1), 3–45. https://doi.org/10.1007/bf01161745

[5] Fedorets, A. A., Bormashenko, E., Dombrovsky, L. A., & Nosonovsky, M. (2020). Symmetry of small clusters of levitating water droplets. *Physical Chemistry Chemical Physics*, *22*(21), 12239-12244. https://doi.org/10.1039/D0CP01804J

[6] Humphreys, J. E. (1990). Reflection Groups and Coxeter Groups. Cambridge University Press.

[7] KaleidoTile. (n.d.). Www.geometrygames.org. Retrieved September 14, 2022, from https://www.geometrygames.org/KaleidoTile/index.html.en

[8] McCammond, J. (2010). Hyperbolic Coxeter groups and their finite simple cousins. Retrieved September 13, 2022, from https://www.math.lsu.edu/~pdani/conferences/goa2010/SpeakerNotes/mccammond.pdf



∞?
ℵ₀
BY EVA♥
dy/dx
dx??
!
??
i
!
√2
?!
ẍ=a
ijk=-1
aⁿ+bⁿ=cⁿ
By Eva

A War of Words and Numbers: Exploring Controversy in Maths

Lauren Aitken

Tick, cross, tick, cross… At first glance, there is no room for controversy in maths: how can there be disagreement in the strict application of rigour?

Mathematics is commonly viewed as being objective, driven only by logic. This idea is furthered in school, where maths is often taught with an emphasis on exam-taking. Mark schemes are available, questions are designed with specific methods in mind. Most people go through life without ever encountering a maths problem that has not already been solved. Consequently, the possibility of differing interpretations is left to the humanities: what place can controversy hold in mathematics?

Initially, the most famous mathematical controversies seem to confirm this view, with the crux of the problem lying in human quarrels, not the mathematics itself. The (now largely accepted as independent) development of calculus by Newton and Leibniz caused arguments in the mathematical community over which man had scientific priority, eventually drawing the men themselves into accusing each other of plagiarism. The fallout? The 'largely pointless but heated controversy … caused huge damage to English mathematicians' for their refusal to engage with continental work (Stewart, 2012, p. 42). By and large, it was not the idea of infinitesimal calculus itself which raised hackles – although this did arise briefly, in the dismissive remarks of George Berkeley in *The Analyst; or, A Discourse Addressed to an Infidel Mathematician* (1734). Motivated largely by the belief that calculus contradicts Christian thought, Berkeley criticises calculus for deducing 'true Propositions from false Principles' (§20).

The calculus controversy of the early 1700s has become notorious in the history of mathematical disputes, taking on an almost mythological status. As we look further back in time, stories of disagreements intensify rather dramatically. The Pythagoreans were a school ardent about the harmony of natural numbers and their ratios (the rationals). When one of their members, Hippasus, discovered a proof of the existence of irrational numbers, he was allegedly thrown into the sea. Jump forward 2000 years, and the introduction of complex numbers was also regarded with a deep level of scepticism. Descartes' dismissal of them as 'imaginary' reflects the suspicions with which this field was considered for decades, with their existence treated as purely theoretical. With time, their usefulness became undisputed in analysis and, now, complex numbers are found at the heart of quantum theory, with claims that they are not merely a mathematical tool, but instead also hold real physical information about quantum states (Scandolo, 2021).

But what is merely a mathematical tool and what is more than that? What mathematics is worthy of research? Are these 'new' numbers – whether they be negative, irrational, or complex – invented, or discovered? As we begin to delve into the reasons behind debate, underlying philosophies begin to appear. Questions about the philosophical foundations of mathematics re-emerge each time a particularly sensational development appears. A notable example is the reaction to Georg Cantor's treatment of infinity at the end of the 19th century. At the time, mathematics largely stuck to the finite numbers. The notion of infinity was seen as inconsistent and, importantly, in a largely Christian Europe, also violated the belief that the infinite was a condition unique to God (Dauben, 1991). When Cantor started to develop transfinite set theory, he faced intense opposition from mathematicians such as Leopold Kronecker, who famously called Cantor a 'corruptor of youth'. From Cantor's perspective, however, if he could form an internally consistent theory, he was free to do as he liked. He emphasised this point in his *Grundlagen*, stating that pure mathematics, unlike all other sciences is at its core a 'free' discipline. Advancement could not be made, he believed, without mathematicians being allowed to take all reasoning to logical and consistent conclusions, no matter their controversial status.

In a subject so focused on proof, perhaps it is no wonder that the nature of proof itself has become a historical topic of debate. When the four-colour problem was first posed in 1852, it was not expected to have the impact that it eventually had on the mathematical world. However, the question of how many distinct colours are sufficient to colour all possible maps, without adjacent regions being coloured alike, gained in popularity as the answer remained elusive for over a century. Eventually, Appel and Haken proved that four colours suffice, by narrowing the number of fundamental cases and checking each one individually. The catch? There were 1,834 cases and each was checked using a computer, taking over 1,000 hours in total. Controversy developed: is a proof that relies on computation beyond reasonable human capacity a 'true' mathematical proof? Tymoczko (1979) claimed that while the four-colour theorem could conceivably be said to have been proven mathematically (the proof was eventually accepted after several rounds of finding and correcting the algorithm's errors), the method used required a philosophical shift to the concept of 'proof', arguing that 'experimental methods', akin to that of the natural sciences, had been introduced.

Ultimately, controversy does occur among mathematicians. Historically, these controversies often reflect ideological beliefs about mathematics. From the Pythagoreans' alleged dispatch of Hippasus, Berkeley's pamphlet, or Kronecker's disapproval, it is now clear that maths does not exist within a logical bubble, but instead in a divided and belief-driven world. How its development is viewed varies wildly also: compare Cantor's distinction between maths and the natural sciences, to Tymoczko's claims of the emergence of empiricism in the field.

Benford's law of controversy, from his novel *Timescape*, comes to mind: 'passion is inversely proportional to the amount of real information available'. Mathematical controversies often die down as the field or theorem develops. Many historical controversies now have a 'correct' answer, with one interpretation becoming standard: not many contemporary mathematicians will scoff at the existence of irrational numbers, even if their field does not directly require them. Moreover, different areas of maths can coexist while using different foundational philosophies: for example, proofs are published with and without assuming the axiom of choice. As time goes on, new mathematical controversies will inevitably arise and, undoubtedly, many too will be laid to rest in the history books.

Data and the stories it tells

Maria Tașcă

This summer I had the privilege to be part of the van der Schaar lab, being supervised by prof. Mihaela van der Schaar and funded by the Cambridge Open Mathematics Internships Programme, whom I must thank for their support. The project I took upon focused on developing new interpretability methods in Machine Learning, an incredibly exciting research area that focuses on deciphering a black box model. And thus started the first chapter of my story: **Interpretability**.

Most of the time, a Machine Learning model will output a prediction for a given data instance, without explaining why that decision was taken: this is what we call a *black box model*. Intricate and mysterious, black box models are used in high-stake fields, such as healthcare, the criminal justice system or the financial lending system. Thus, it's important for the stakeholders to understand how the model works in order to trust and use it.

One may ask why we don't create an explicit model from the beginning so that we know exactly how everything works and how each decision is taken. These models are called *white box models* and it turns out that they are significantly less accurate than black box models, which is why we're interested in investigating how the latter work.

Moreover, another reason interpretability is such an exciting area of research is that it has the power to unravel the inner workings of the model, which could lead to ground-breaking discoveries in various fields, such as pharmacology or medicine. Imagine a machine that takes the inputs and the outputs of an equation and then comes up with the equation itself.

When reading a paper that proposed a method to interpret the model's prediction using examples from a data set chosen by the user[1], it occurred to me that the quality of the data is extremely important: a Machine Learning model is only as accurate as its training data. This is how the second chapter starts: **Data Centric AI**.

While model centric AI rests on the assumption that the data is fixed and only the model can be improved, Data Centric AI focuses on improving the quality of the data itself. Some approaches consist in understanding which instances are forgotten

[1] `https://arxiv.org/pdf/2110.15355.pdf`

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
people_ID	Region	Gender	Desig nation	Name	Married	Children	Occupation	Mode_ transport	Cases/1M	Deaths/1M	Comorbidity	Age	Coma Score	Pulmonary Score
1	Bhuban	Female	Mrs.	mansi	YES	1	Farmer	Public	2	0	Hypertension	68	8	<400
2	Bhuban	Female	Mrs.	riya masi	YES	2	Farmer	Walk	2	0	Diabeties	64	15	<100
3	Bhuban	Female	Mrs.	sunita	NO	1	Cleaner	Public	2	0	None	19	13	<300
4	Bhuban	Female	Mrs.	anjali	YES	1	Driver	Car	2	0	Coronary Hear	33	9	<200
5	Bhuban	Female	Mrs.	champa k	NO	2	Manufacture	Car	2	0	Diabeties	23	7	<400

	15	16	17	18	19	20	21	22	23	24	25	26	
people_ID	Cardio Logica	Diuresis	Platelets	HBB	d-dimer	Heart Rate	HDL Choles	Charlson	Blood Glucose	Insurance	salary	FT/month	
1	Normal	441	154	96	233	82	58	27	7	3600000	1E + 06	2	
2	Stage-0		121	56	328	89	68	5	6	1600000	400000	1	
3	Elevate	416	124	137	213	77	43	40	6	3400000	900000	1	
4	Stage-0	410	98	167	275	64	60	27	7	700000	2E + 06	1	
5	Normal	390	21	153	331	71	64	32	7	3200000	1E + 06	1	

Figure 1: A dataset used in the study of the COVID-19 virus

during the training process[2], and which are uncertain or inconsistent[3], in order to eliminate unreliable points from the data set and reduce processing times. However, if the data were fair and unbiased from the outset, would such issues arise? The third chapter commences: **Bias in Data**.

A dataset is biased when it doesn't accurately represent the situation to be studied. This discrepancy often arises from the human relationships at play between the people involved. Despite being filled with numbers and cryptic notation, datasets tell stories of people and how they interact and see each other, from a doctor treating their patient to an employer trying to find their next employee. The datasets I studied are usually tabular and they contain information that the collector of the data (e.g. the healthcare system) deems necessary for the purpose of the study (e.g. age, gender, weight, blood pressure), but nothing about the agent that acts on the data (e.g. the doctor). Surely, we can agree that certain groups of people are more impacted by bias than others, either positively (someone is favoured unfairly) or negatively (someone is opposed unfairly). Maybe it is time to think about the source of the bias by establishing relationships between groups of people, characterised by some specific features. How can bias be addressed if the root of problem seems to be everyone, and thus no one in particular? How can we find the source of bias if all spreadsheets look like this?[4]

[2] `https://arxiv.org/pdf/1812.05159.pdf`

[3] `https://arxiv.org/pdf/2202.08836.pdf`

[4] `https://www.researchgate.net/figure/Sample-healthcare-dataset-Singh-2020_fig3_346868725`

This is why I'm proposing another way to look at and gather data sets, that considers the features of not only the ones who are impacted by the actions of the decision makers, but the features of the decision makers themselves, as well as the features of the environment where they interact. I call my method The 5Ws of Machine Learning: Who, What, Where, When and by Whom? I consider the data set to be complete if it contains the features (that might relate to bias) of both the assessed and the assessors e.g. race, gender, socio economic status (**Who? By Whom?**), enough information for the scope of the study (**What?**), the place and time of the interaction (**Where? When?**). To prove why this is necessary, I analysed a popular data set in ML literature: Heart Failure Predictions[5] and here are my conclusions:

There is no record of features that might be related to potential bias, making it much harder to trace. Furthermore, over a long period of time conditions which could affect the severity and the outcomes of the cases might change (e.g. environmental), thus making earlier records less relevant, which is why recording the times a patient is treated is important. Additionally, writing down the hospital in which the patient was treated allows for the tracking of good and bad practices.

If we were to collect data in such a way, not only could we have a clearer view of the bias present in data sets and thus increase the accuracy of the models, but we could quantify both positive and negative bias, which is the aim of my report. There are many ML models that work towards eliminating bias in data sets[6], as well as other ways of collecting data in a healthier way, reasons why I am optimistic about the future of data science and machine learning.

[5] https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

[6] https://proceedings.neurips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

Asymptotic hopes

Felix Stokes

This poem takes inspiration from Catullus' poem 51.

it seems to me that he inhabits four-space,
or really, if I squint, it's even more space:
he's x and you're one over him, while I
am struggling, i'm zero = y

which means that I am here. eternally
reaching at you and your proximity
but asymptotic hopes have never met
you. it's only me within my set.

and Zeno, he was right: Achilles' heel
approaches - never meets - I'll never feel
the touch of all your curves upon my skin.
instead, I sit and stoke the flames within.
the only thing that's left and positive
is $\frac{\mathrm{d(heartbeat)}}{\mathrm{d}t}$, I'm closeted

by night, my tongue is looping, möbius,
I'm inside out and not, my flesh erupts
and folds. I am the shadow of a klein
bottle, my temperature's looping like sines;
I try to find my roots, I factorise,
but I must not have factored in your eyes

for suddenly you are transcendental:
I have no roots. I spiral out. your pull
does nothing. e's got you into this mess:
you're exponential, man. take logs. I guess
these are my logs. I add. I multiply.

let him become your ex, and I'll be why.

Sorry, I don’t do maths

Siddiq Islam

They’d try and convince me with "Go on, just try it"s,
With "Don’t be a wuss"s and "It’s fine, no one’s looking"s.
I’d always tell them, "Sorry, I don’t do maths."

They’d crowd round on soggy park benches,
Applying Green’s theorem to a circle.
I kept my path well clear of them.

They’d offer at parties,
They’d roll up papers on number theory.
"It’s all natural," they’d assure me.
"Sorry, I don’t do maths."

A friend tried to get me into linear maps:
Injections, surjections and bijections,
When I’d never even touched adjoint.

People got addicted.
I saw a man doing lines of complex analysis in a public bathroom.
I only caught a glimpse but his "i"s were read.

And now they’re all lost in thought,
Wasting their minds pondering imaginary numbers.
I don’t do maths, and I never will.

Proof by contradiction

Aleksandra Bozovic

To say it is so, truth lent to a lie,
This treacherous gift, given free of guilt,
I'll use to track you down, and, as a spy,
To break into the world that you have built.

That strange, collapsing world I shall explore,
The inconceivable I shall understand,
These wondrous dreams never seen before,
That reason's piercing gaze cannot withstand…

A rotten tree you are, that bears bad fruit.
Now I shall make your falsity laid bare.
You try to run from my frenzied pursuit,
But all paths lead you right into my snare.

If I affirmed, it was but to deny,
For what is void of truth must surely die.

Galois' Story.
A highly historically inaccurate artistic doodle.
By Eva
Trinity Term 2022.
Évariste Galois
Father: Nicolas-Gabriel Galois
Sister: Nathalie
Mother: Adélaide-Marie Demante
Brother: Alfred
Little Evariste
Galois was born during a tumultuous period of French history: the 19th century. His father was a revolutionary and a mayor. He was taught by his mother.
Little Galois was obsessed with the geometric world.
He devoured math books like novels.
And had a peaceful childhood...
Well... Almost.
MY DEAR FATHER!!
Don't leave us!!
His father committed suicide,
Leaving the young children alone with their mother.


Poor Galois sought comfort in geometry.
Caress my heart, O beautiful symmetry!...
Shine upon the darkness of my soul...
And amid the chaotic mist...
he heard a voice...
..that called him forward.
By Eva
Enchanted, he followed, and glimpsed a land none had seen before.


P(5)
By 15, he had already made significant progress in the theory of polynomial equations.
École Polytechnique
And he intended to enter the prestigious École Polytechnique.
Just a few days after his father's death,
he sat the oral exam ... for the second time.
...
"I'm sorry Mr. Galois, but could you please explain your reasoning step by step?"
By Gray
Why do you think — ?
Why is — ?
...
But why?
...
...!
Why?
I'm sorry, I don't follow
Enough!! I have no further explanation!
Sob
Sob
"Mr. Galois, your ideas are correct but you cannot present them clearly. I'm sorry to tell you that you have failed the examination."


Why... why can't they see ... what I can...?
Galois' ideas were too new for his contemporaries to grasp.
20°
$G \subset G_1 \subset G_2$ ...
My father...
I see it...
I see it! Don't run away! I...
What is it!
Ah!! I see!! ...
He went to study at the École Normale instead.
There, his professors understood his genius and offered their help.
Encouraged by them, he submitted essays to the Academy of Science.
Polynomials roots ...
His essays were reviewed by the most prominent mathematicians of the time.
Cauchy
Fourier
Poisson
Cauchy encouraged him to write new essays for competition. But these essays were sent to Fourier and were lost after his death.
Other than that, his essays were not well-received...
"Incomprehensible",
But interesting... You should publish a more complete account.
- Poisson

Galois, like his father, was an ardent republican and immediately got involved in the rebellion. He was expelled from school, taken to the court several times, and kept in prison for months.
At the same time, like a prophet, he had a mission to fulfill.

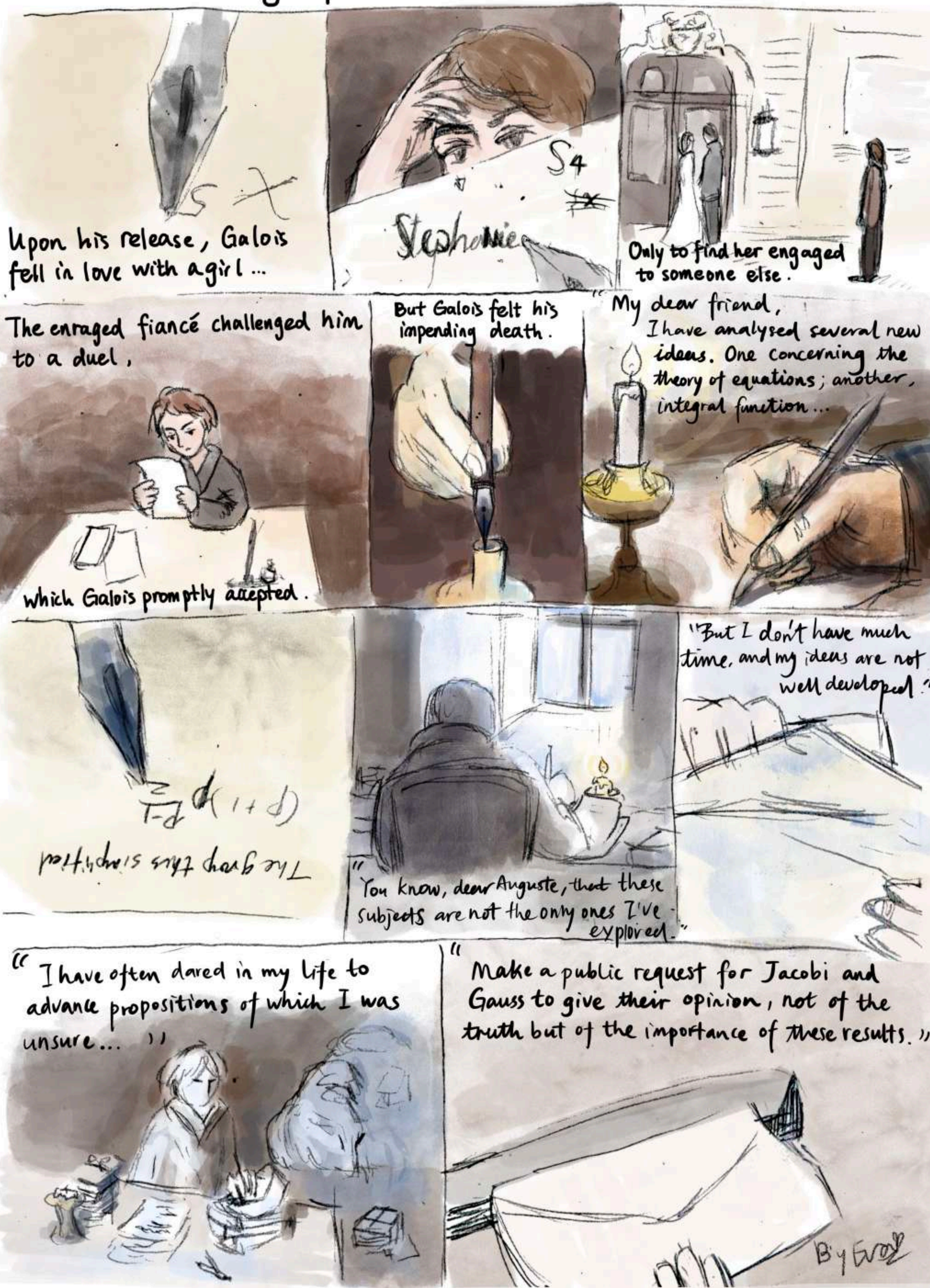
His life was divided in halves: during the day, he was a passionate revolutionary. At night, he walked into the gardens of groups and equations, going further, and further...

He seemed to sense his tragic fate: with increasing frenzy he scribbled his ideas down...

Even...
In prison.

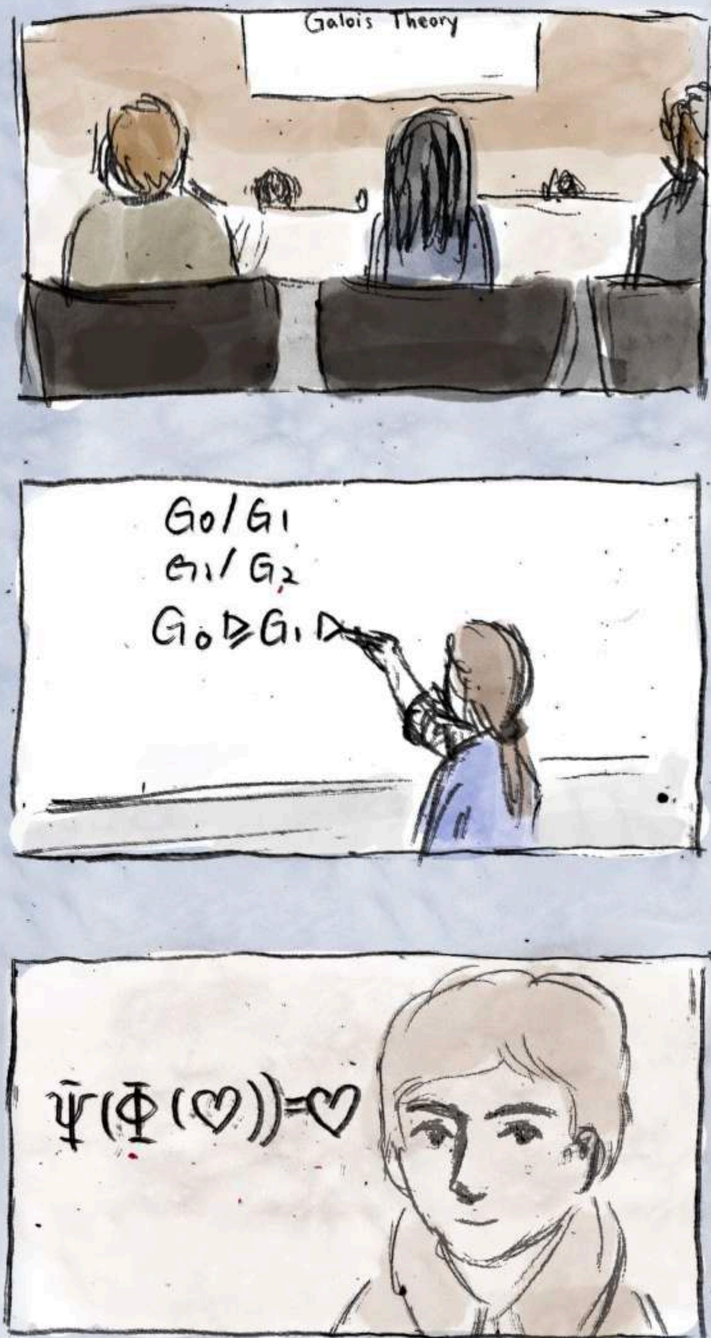

Once, in prison, in a drunken state,
Galois revealed his true heart...
Do you know what I lack, my friend? I can confide it only to you: It's someone whom I can love and love only in spirit. I've lost my father and no one has ever replaced him, do you hear me...?
And I tell you, I will die in a duel on the occasion of some coquette de bas Étage. Why? Because she will invite me to avenge her honor which another has compromised!
Évariste!!
How silly is that!
No! Leave me!... leave...
Please don't do that again!
You frightened me!
"Galois! Galois! For your unfinished world, you must live!"

But tragedy loomed over the horizon ...




"After all this, I hope there will be people who will profit by deciphering ALL THIS MESS."
?
"Don't cry for me. I need all my courage to die at 20."
- Galois to his brother

MMXXII...



Since fate did not give me enough of a life to be remembered by my country, please remember me.

I die your friend,
É. Galois

Painted By Evolz

The Invariants committee 2022-23

Otilia Căşuneanu: President
Vlad Tuchilus: Secretary
Ben Gardner: Treasurer
Deepak Ganger: Social secretary
Maria Taşcă: Social secretary
Leyi Jiang: Publicity secretary
Tobias Bretschneider: IT officer
Aleksander Bosek: Puzzle master
Oliver Hayman: Puzzle master
Yuchen Fan: Puzzle master
Diego Vurgait: Editor

The Invariants are sponsored by:


FERMAT'S
LAST
TANGO
an Empty Set Productions musical
Week 7 Hilary Term
Mathematics Institute, Oxford

Write for the Invariant!

Write for *the* Oxford mathematics magazine! We are looking for all sorts of pieces for our next issue: articles, essays, poems, illustrations, as well as puzzles and games. If you're interested in writing for the *Invariant*, have suggestions or want to write a letter to the editor, email us at *editor@invariants.org.uk*.