



π in the Sky

ISSUE 20, 2017



Pacific Institute *for the*
Mathematical Sciences

Pi in the Sky

Issue 20, 2017

Editorial Board

John Bowman (University of Alberta)

Murray Bremner (University of Saskatchewan)

John Campbell (Edmonton)

Florin Diacu (University of Victoria)

Krista Francis (Edmonton)

Gordon Hamilton
(Masters Academy and College, Calgary)

Klaus Hoechsmann (University of British Columbia)

Dragos Hrimiuc (University of Alberta)

Michael Lamoureux (University of Calgary)

Fok Leung (University of British Columbia)

David Leeming (University of Victoria)

Patrick Maidorn (University of Regina)

Managing Editor

Anthony Quas (University of Victoria)
Tel: (250) 721-7463, E-mail: aquas@uvic.ca

Production Coordinator

Miguel Eichelberger, PIMS Communications
Manager

Designer

Anthony Santiago

With thanks to SpikedMath.com for permission to reproduce the cartoons.

Contact Information

Pi in the Sky

PIMS University of Victoria Site Office,
SSM Building Room A418b
PO Box 3060 STN CSC, 3800 Finnerty Road,
Victoria, BC, V8W 3RT

(250) 472-4271
F/(250) 721-8958
pims@math.uvic.ca

Table of Contents

Welcome	2
Mathematics to Scale	3
Jeffrey S. Rosenthal	
On Assessing Real World Prediction Skill	6
David Aldous	
Snakes & Ladders Playing Time	9
Doddy Kastanya	
Big Data	13
Chris Budd	
Worlds in Your Pocket: 2D Screens to 3D Touch	18
Harrison Cui	
Simple Set Game Proof Stuns Mathematicians	22
Erica Klarreich for Quanta Magazine	
The Three Radii Problem	26
Tamar Barabi	
2017 Math Quickies	28
2015 Math Quickies Solutions	29

Submission Information

For details on submitting articles for our next edition of Pi in the Sky, please see: <http://www.pims.math.ca/resources/publications/pi-sky>

Pi in the Sky is a publication of the Pacific Institute for the Mathematical Sciences (PIMS). PIMS is supported by the Natural Sciences and Engineering Research Council of Canada, the Province of Alberta, the Province of British Columbia, the Province of Saskatchewan, Simon Fraser University, the University of Alberta, the University of British Columbia, the University of Calgary, the University of Lethbridge, the University of Regina, the University of Saskatchewan, the University of Victoria and the University of Washington.

Pi in the Sky is aimed primarily at high school students and teachers, with the main goal of providing a cultural context/landscape for mathematics. It has a natural extension to junior high school students and undergraduates, and articles may also put curriculum topics in a different perspective.

Welcome to Pi in the Sky!

The Pacific Institute for the Mathematical Sciences (PIMS) sponsors and coordinates a wide assortment of educational activities for the K-12 level, as well as for undergraduate and graduate students and members of underrepresented groups. PIMS is dedicated to **increasing public awareness** of the importance of mathematics in the world around us. We want young people to see that mathematics is a subject that opens doors to **more than just careers in science**. Many different and exciting fields in industry are eager to recruit people who are well prepared in this subject.

PIMS believes that **training the next generation of mathematical scientists** and **promoting diversity within mathematics** cannot begin too early. **We believe numeracy is an integral part of development and learning.**

Pi in the Sky is available online at www.pims.math.ca/resources/publications

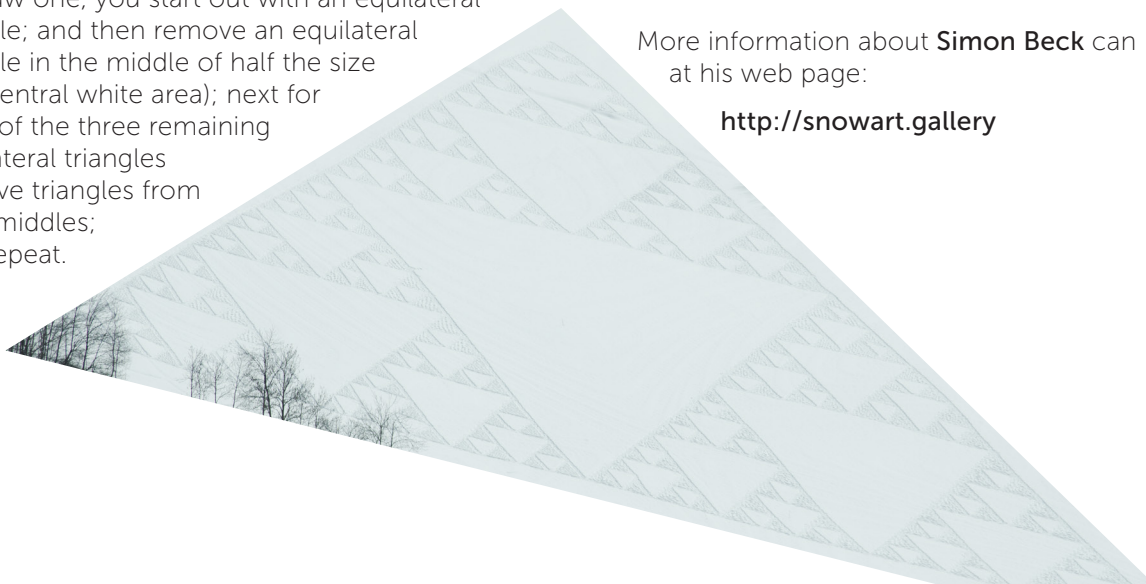
Solutions to Math Challenges at the end of this issue will be published Pi in the Sky Issue 20. See details on page 28 for your chance to win \$100!

A Note on the Cover

This is a snow art installation created by **Simon Beck**, an Oxford-educated engineer turned snow artist. Beck creates these images using snow shoes over a 12 hour period. This image was created in the French Alps.

The picture is of a "**Sierpinski triangle**", a famous geometric figure studied by **Waclaw Sierpinski**.

To draw one, you start out with an equilateral triangle; and then remove an equilateral triangle in the middle of half the size (the central white area); next for each of the three remaining equilateral triangles remove triangles from their middles; and repeat.



For more information on our education programs, please contact one of our hardworking Education Coordinators.

Melania Alvarez, UBC, Vancouver, BC
melania@pims.math.ca

Malgorzata Dubiel, SFU, Burnaby, BC
dubiel@cs.sfu.ca

Sean Graves, U of Alberta, AB
sgraves@ualberta.ca

Darja Barr, U of Manitoba, MB
kalajdzi@cc.umanitoba.ca

Indy Lagu, U of Calgary, AB
lLagu@mtroyal.ca

David Leeming, UVic, Victoria, BC
leemingd@uvic.ca

Harley Weston, U of Regina, SK
weston@math.uregina.ca

The **Sierpinski triangle** is a famous example of a **fractal** (its dimension is about 1.585), but has also been found in designs in medieval churches in Rome (where it was created centuries before Sierpinski's life).

More information on the **Sierpinski triangle** can be found at:

https://en.wikipedia.org/wiki/Sierpinski_triangle

More information about **Simon Beck** can be found at his web page:

<http://snowart.gallery>

MATHEMATICS TO SCALE

BY JEFFREY S. ROSENTHAL

Professor in the Department of Statistical Sciences at the University of Toronto.

MATHEMATICS IS A WONDERFUL SUBJECT, but much of it requires years of study to fully appreciate. When I was recently asked to design a course in quantitative reasoning for Humanities students [8], I had to come up with mathematical ideas and examples which were interesting, relevant, exciting, and most importantly, comprehensible without much math background.

Now I do have some experience communicating mathematical ideas to non-mathematicians, including writing a bestselling general-interest book about probabilities [1], helping to uncover a major front-page lottery scandal [6], and writing about such diverse topics as the mathematics of music [2], probability and justice [7], sports statistics [5], the Monty Hall problem [3], and even family relationships [4]. But beyond that, what fundamental mathematical idea could I use to begin the course, to show these Humanities students the interest and power of mathematical thinking, without scaring them off or pushing them towards feelings of math anxiety?

After much consideration, I decided to begin my course with the concept of scaling. That is, how do various quantities change when an object's size is modified?

The key to scaling is that it depends on the *dimension*. Consider first a **one-dimensional object** like a line. If you wanted to draw a second line which was twice as long as your first line, then how many times as much ink would it require? Why, twice as much, of course. Indeed, in one dimension, an object's length or size or amount of ink are all pretty much the same thing, and there is little more to say.

In **two dimensions**, the situation is more interesting. Suppose you draw a square on a page, and then later wish to draw a second square which is twice as big (in all directions). How many times as much ink do you require now? Well, four times as much.

This is because in two dimensions, an object's area is proportional to its length times its height, and if an object is expanded then each of its length and height is multiplied by two, so the product is $2 \times 2 = 4$. Similarly, to draw a square three times as large (in all directions) requires $3 \times 3 = 9$ times as much ink.

The importance of this observation is that it actually **applies to more than just squares**. Indeed, *any* shape drawn on a page can be thought of as being made up of lots and lots of tiny disjoint squares. (Formally, this is justified by *calculus*, upon taking the *limit* of more and more smaller and smaller squares.) So, consider *any* shape drawn on a page (say, a child's drawing of a heart on a mother's day card). If you wanted to draw a new version which was identical except twice as large (in all directions), then it would require $2 \times 2 = 4$ times as much ink. This is true regardless of the chosen shape, and regardless of its actual area (which probably couldn't be computed precisely anyway).

More generally, if any two-dimensional region is stretched by one factor vertically and another factor horizontally, then the ratio of the area is simply the product of both stretch factors. For example, **dimes** are approximately one centimeter by one centimeter. So, how many dimes could be taped flat against my classroom's wall, which is about four meters tall and twelve meters wide? Well, here the vertical scaling factor is 400, and the horizontal factor is 1,200, so the total number of dimes is $400 \times 1,200 = 480,000$, worth a total \$48,000. That's a lot of dimes.

For another example, suppose **flowers** are planted about 20 centimeters apart. Then about how many flowers are on a five meter by eight meter field? Well, roughly speaking, here each flower occupies a disjoint "region" of about 20 centimeters by 20 centimeters. And the entire field can be viewed as a scaled up version of one such region. So, the number of flowers on the whole field is the product of the two scaling factors, namely $(500/20) \times (800/20) = 25 \times 40 = 1,000$. So, there are about a thousand of them - lots of flowers!

Next, consider scaling in **three dimensions**. If we have a three-dimensional cube, and then expand it to be twice as large (in all directions), then it has three different scaling factors, so its volume is multiplied by $2 \times 2 \times 2 = 8$. That is, it is eight times as large!

And once again, this principle **applies to more than just cubes**, since any volume can be thought of as consisting of lots and lots of tiny disjoint cubes. For example, consider a **glass of beer** (a very relevant example for students!). If a second glass is twice as large in all directions, then it holds eight times as much beer – a fact which surprises many people. Or, suppose instead that a second glass is twice as tall, but only $2/3$ as wide and deep. Most people would think the second glass holds more beer. But in fact, it holds $2 \times (2/3) \times (2/3) = 8/9$ times as much. And $8/9 < 1$, so the second glass actually holds less.

Even more dramatically, consider a **cone-shaped glass** (like a fancy wine glass, or certain water-cooler cups). Suppose it is filled up to $2/3$ of its full height. Then what fraction of its volume is full? Well, since it is cone-shaped, its bottom two-thirds is identical to the entire cup, except scaled by $2/3$ in all directions. So, its bottom two-thirds holds $(2/3) \times (2/3) \times (2/3) = 8/27 \approx 30\%$ of its full volume. So, with a cone-shaped cup, if the bartender fills $2/3$ of its height, he is only giving you about 30% of a full glass of wine. Tell him to fill it up!

Another interesting application is to **mass**. A standard reference point is that a $10 \text{ cm} \times 10 \text{ cm} \times 10 \text{ cm}$ cube of water equals one litre, and weighs one kg. So what about a $1 \text{ m} \times 1 \text{ m} \times 1 \text{ m}$ cube of water? Well, 1 m is ten times as long as 10 cm. So, a $1 \text{ m} \times 1 \text{ m} \times 1 \text{ m}$ cube has volume $10 \times 10 \times 10 = 1,000$ litres, and weighs 1,000 kgs (about 2,205 lbs) – much too heavy to lift! Similarly, a $160 \text{ cm} \times 200 \text{ cm} \times 20 \text{ cm}$ waterbed has volume $16 \times 20 \times 2 = 640$ litres, and weighs a massive 640 kgs (about 1,411 lbs). This is why waterbeds can only be put into sturdy houses, and cannot be moved without first being drained.

Another perspective comes from picturing **crowds of people**. If you are in a one-dimensional lineup of people spaced about 0.5 meters apart, then the number of people standing within ten meters of you (in front or behind) is about $20/0.5 = 40$.

But if you are in a two-dimensional crowd of people (at a concert or party or dance), all spaced about 0.5 meters apart, then the number of people within ten meters of you is about $(20/0.5) \times (20/0.5) = 1,600$ – lots more! (Or, if you really want to consider just a circle of people around you, not a square, then the answer is more like $\pi(10/0.5)^2 \approx 1,257$.) Or, if a flock of birds, flying in three dimensions, are spaced about 0.5 meters apart, then the number of birds within ten meters of any one (central) bird is about $(20/0.5) \times (20/0.5) \times (20/0.5) = 64,000$, a massive number. That's scaling, in different dimensions.

Similarly, **stars** in our galaxy are approximately five light-years apart on average. So if stars are visible up to, say, one hundred light-years away in all directions, then the number of visible stars is roughly $(200/5) \times (200/5) \times (100/5) = 32,000$. (Or, if you want to count only a half-sphere of visibility, then it's about $\pi(100/5)^3/2 \approx 12,566$.) Starry starry night, indeed! So, scaling explains why on a clear night you can see so many stars, even though most stars are too far away to be visible.

Comparing two different **people** is also fun. Suppose that Person #2 is twice (say) as large as Person #1 in all directions, otherwise identical. Then how many times as high can Person #2 reach? Answer: 2 (since height is one-dimensional). How many times as much does Person #2 weigh? Answer: 8 (since weight is three-dimensional). How many times as many blades of grass does Person #2 trample if they each take one step on a lawn? Answer: 4 (since foot area is two-dimensional). How many times as much blood does Person #2's arteries contain? Answer: 8 (since volume is three-dimensional). How many times as much skin surface (e.g. for tattoos) does Person #2 have? Answer: 4 (since skin surface is two-dimensional). And so on. Again, these answers do not depend on the precise shape of the people or their feet or arteries, they just involve scaling. (And of course, similar answers apply to other scaling factors besides two.)

Next, consider **pressure**, i.e. the amount of force per unit area on e.g. the ground underneath our feet. Considering **snowshoes** is instructive. Snowshoes work by spreading our mass over a larger area, to reduce the pressure on each individual spot of snow.

Suppose I am wearing snowshoes which are twice as wide, and three times as long, as my regular boot. Then my same mass is spread over an area which is $2 \times 3 = 6$ times as large. So, the amount of pressure I exert on any one spot of snow is only $1/6$ as large. This is why I can (hopefully) stand on the top of the snow while wearing snowshoes, even if I would have sunk down deep in my normal boots.

Consider now a **giant lizard** (like Godzilla). Suppose the giant is one thousand times as large as a regular lizard, in all directions. Then it weighs $1,000 \times 1,000 \times 1,000 = 1,000,000,000$ (one billion) times as much(!). And it has one billion times as much blood, and so on. On the other hand, its foot area is $1,000 \times 1,000 = 1,000,000$ (one million) times as large. So, the amount of pressure that it exerts on the ground is $1,000,000,000$ divided by $1,000,000$, or $1,000$ (one thousand) times as much. That is why giants tend to crush whatever they step on. But because of scaling, the pressure is only multiplied by their scaling factor (e.g. a thousand), not by their full weight ratio (e.g. a billion).

Finally, consider **rainfall**. You might have noticed that rain amounts are normally reported as lengths, e.g. “downtown Toronto received 40.6 mm of rain today”. How could this be? Well, consider two different bins left out during a rainstorm. Suppose the second bin is twice as large as the first (in all directions). Then the area at the top of the bin is $2 \times 2 = 4$ times as large. So, the volume of rain collected by the second bin is 4 times as much. On the other hand, this rain is then spread over a base area which is also $2 \times 2 = 4$ times as large. This means that the height of the rain in the second bucket is $4/4 = 1$ times as high, i.e. the height of rain in the two buckets is identical! Similar considerations apply to any other buckets of any other sizes and shapes, as long as they have straight sides (so the area of each bucket is the same at the top and the bottom). So, when reporting an amount of rainfall as a length, that length is equal to the height of water that would be left in any straight-sided bucket of any size whatsoever. And that is why rainfall amounts can be reported as simple lengths, not as “volume per unit bucket” or some other complicated standard.

So, the next time someone asks you for an easy-to-understand example of how mathematical thinking applies to everyday life, tell them a tale or two about scaling!

REFERENCES

- [1] J.S. Rosenthal (2005), “Struck by Lightning: The Curious World of Probabilities” (book for the general public). HarperCollins Canada. <http://probability.ca/sbl/>
- [2] J.S. Rosenthal (2005), “The Magical Mathematics of Music”. Plus Magazine. <https://plus.maths.org/content/magical-mathematics-music>
- [3] J.S. Rosenthal (2008), “Monty Hall, Monty Fall, Monty Crawl”. Math Horizons (September 2008 issue), 5–7. <http://probability.ca/jeff/writing/montyfall.pdf>
- [4] J.S. Rosenthal (2011), “The Mathematics of Your Next Family Reunion”. <https://plus.maths.org/content/mathematics-your-next-family-reunion>
- [5] J.S. Rosenthal (2013), “The Rosenthal Fit: A Statistical Ranking of NCAA Men’s Basketball Teams”. TSN sports television channel. <http://www2.tsn.ca/ncaa/story/?id=418503>
- [6] J.S. Rosenthal (2014), “Statistics and the Ontario Lottery Retailer Scandal”. CHANCE Magazine 27(1). <http://probability.ca/lotteryscandal/> Reprinted in The Best Writing on Mathematics, Princeton University Press, 2015.
- [7] J.S. Rosenthal (2015), “Probability, Justice, and the Risk of Wrongful Conviction”. The Mathematics Enthusiast 12, 11–18. <http://probability.ca/jeff/ftpdir/probjustice.pdf>
- [8] J.S. Rosenthal (2015), “STA 201F: Why Numbers Matter”. <http://probability.ca/sta201>

ABOUT THE AUTHOR

Jeffrey Rosenthal received his PhD in Mathematics from Harvard University at the age of 24, and tenure at the University of Toronto at the age of 29. He received the 2006 CRM-SSC Prize, the 2007 COPSS Presidents’ Award, the 2013 SSC Gold Medal, and teaching awards at both Harvard and Toronto. He is a fellow of the Institute of Mathematical Statistics and of the Royal Society of Canada. Rosenthal’s book for the general public, Struck by Lightning: The Curious World of Probabilities, was published in sixteen editions and ten languages, and was a bestseller in Canada, leading to numerous media and public appearances, and to his work exposing the Ontario lottery retailer scandal.

ON ASSESSING REAL WORLD PREDICTION SKILL

BY DAVID ALDOUS

Professor in the Statistics Department at U.C. Berkeley

Introduction

Let me start with a puzzle. Is it possible to devise a quiz contest (on any topic, not necessarily mathematical) with the following properties?

Answers will be scored objectively – no subjective judgments (as would be needed for creative writing, for instance). Contestants who end with a better overall score will – beyond reasonable doubt – be better at the subject matter of the quiz.

The questions refer to substantive real-world matters, rather than fantasy (islands with liars and truth-tellers) or self-referential “how would most other contestants answer this question?” No person (or computer, etc) knows or will ever know the correct answer to any of the questions.

So this looks impossible at first sight – how can one grade objectively without knowing the answers? Now puzzles like this inevitably involve some kind of trick. But my trick is rather mild – an everyday quiz can be graded quickly, but for my quiz you have to wait a while to find your scores. If you can think of a less tricky such quiz, please let me know!

The Good Judgment Project

Here are 4 questions that people with an interest in world affairs might be pondering as I write (September 2017):

1. *Before 2018, will Russia officially announce that it is suspending its participation in or withdrawing from the Intermediate-Range Nuclear Forces Treaty?*
2. *Before 2018, will 5 or more countries experience 10 or more cases of poliovirus?*
3. *Before 2018, will there be a lethal attack on a US military vessel in the Red Sea, Gulf of Aden, Persian Gulf, or Gulf of Oman?*
4. *Before 2018, will China deploy a deep sea oil rig in another country's Exclusive Economic Zone without that country's permission?*

In the current *Good Judgment Project Classic Geopolitical Challenge* [1] participants are asked to assess the current probabilities of such future events. To reiterate, they are **not** asked to give a Yes/No prediction, but instead are asked to give a numerical probability, and to update as time passes and relevant news/analysis appears. Unlike school quizzes, you are free to use any sources you can – if you happen to be a personal friend of Vladimir Putin then you could ask him for a hint on the first question.

Do you think it is ridiculous to pose such questions to non-experts? If so, do you think that trial by jury is ridiculous? In both cases the point is to look at evidence and at expert opinion before giving an answer.

What makes this setting conceptually interesting is that no one will ever know the correct probabilities. Nevertheless one can judge participants' relative ability to assess such probabilities, after the outcomes are known. Explaining this paradox is the focus of this article.

Mean Squared Error

How can we assess someone's ability? We will use several very basic concepts from probability. A random variable X is, informally, a quantity with a range of possible numerical values, the actual value being determined by chance in some way. The *expectation* of X is a real number, written $E[X]$, analogous to the *average* of numerical data. If we seek to predict the value of a *random variable* X , our prediction has to be some constant x_0 . The (random) *squared error* of our prediction is the random variable $(X - x_0)^2$, and the expectation of that random variable, in symbols $E[(X - x_0)^2]$, is called the *mean squared error (MSE)* of the prediction. And the “best” predictor in the sense of minimizing the MSE is just the constant $x_0 = E[X]$.

As a specific example, for a single throw X of a fair die, the MSE from predicting x is

$$\frac{1}{6} \sum_{i=1}^6 (i - x)^2 = \frac{35}{12} + \left(x - \frac{7}{2}\right)^2$$

which is minimized at $x_0 = E[X] = 7/2$. The idea of using *squared* errors goes back to Gauss in the context of errors in astronomy observations, and is widely used in classical statistics because of its nice mathematical properties, which we will exploit in several ways.

An *event*, in the probability context, will either happen or not happen, and we can represent an event as a random variable, taking value 1 if the event happens and value 0 if not. This allows us to use “squared error” to score our predictions. If we predict 70% probability for an event, then our “squared error” is

$$\text{(if event happens)} (1.0 - 0.70)^2 = 0.09$$

$$\text{(if event doesn't happen)} (0.7 - 0)^2 = 0.49.$$

So suppose you participate in a *prediction tournament* like the Good Judgment Project. For simplicity let's suppose that participants just make a one-time *forecast*, a probability prediction, for each event. After the outcomes of all the events are known, your final score will be the average of these squared errors. As in golf, you are trying to get a **low** score.

In the next section I will argue that this is the right way to score. Just as in golf, your score really does indicate how good you are at the prediction game, give or take a small amount of luck.

A very little algebra

When you make a “probability p ” forecast for a certain event, your squared error score will be

$$\begin{aligned} \text{score} &= (1 - p)^2 \text{ if event occurs} \\ &= p^2 \text{ if not.} \end{aligned} \quad (1)$$

Suppose you actually believe the probability is q . What p should you announce as your forecast? Under your belief, your mean score (by the rules of elementary mathematical probability) equals $q(1 - p)^2 + (1 - q)p^2$ and a line of algebra shows this can be rewritten as

$$(p - q)^2 + q(1 - q). \quad (2)$$

Because you seek to minimize the score, and because all you are able to choose is p , you should announce $p = q$, your honest belief – with this scoring rule you cannot “game the system” by being dishonest, that is by announcing a value of p which is not your true belief for the probability.

Now write q for the true probability of the event occurring (recall we are dealing with future real-world events for which the true value q is unknown), and write p for your forecast probability. Then your (true) mean score, by exactly the same calculation, is also given by (2). The term $(p - q)^2$ is the “squared error” in your forecast probability.

Now consider two participants, A and B, making forecasts p_A and p_B for the same event which has (unknown) probability q . Then (2) implies that

$$E[\text{score (A)}] - E[\text{score (B)}] = (p_A - q)^2 - (p_B - q)^2. \quad (3)$$

In a prediction tournament there will be a large number n of events, with unknown probabilities ($q_i, 1 \leq i \leq n$) and with forecasts ($p_{A,i}, p_{B,i}, 1 \leq i \leq n$) chosen by the participants. We **would like to** measure how good a participant is by the average squared-error of their forecast probabilities

$$\text{MSE}(A) = \frac{1}{n} \sum_i (p_{A,i} - q_i)^2 \quad (4)$$

But this is impossible to know, because we don't know the q 's. However, (3) implies that for the final scores (the average of the scores on each event)

$$\begin{aligned} E[\text{final score (A)}] - E[\text{final score (B)}] \\ = \text{MSE}(A) - \text{MSE}(B). \end{aligned} \quad (4)$$

Now your actual final score is random, but by a “law of large numbers” argument, for a large number of events it will be close to its mean. Informally,

$$\begin{aligned} \text{final score (A)} &= E[\text{final score (A)}] \\ &\pm \text{small random effect.} \end{aligned} \quad (5)$$

Putting all this together,

$$\begin{aligned} \text{MSE}(A) - \text{MSE}(B) &= \text{final score (A)} - \text{final score (B)} \\ &\pm \text{small random effect.} \end{aligned}$$

Now we are done: the MSEs are our desired measure of skill, and from the observed final scores we can tell the relative skills of the different participants, up to a small amount of luck.

The mathematical bottom line

Rephrasing the argument above, an individual's score is conceptually the sum of three terms. Write q_i for the (unknown) true probability that the i 'th event happens.

- A term $\frac{1}{n} \sum_i q_i (1 - q_i)$ from irreducible randomness. This is the same for everyone, but we don't know the value.
- Your individual MSE (4), where "error" is (your forecast probability - true probability)
- Your individual luck, from randomness of outcomes.

The analogy with golf continues to be helpful. A golf course has a "par", the score that an expert should attain. Your score on a round of golf can also be regarded as the sum of three terms.

- The par score.
- The typical amount you score over par (your *handicap*, in golf language).
- Your luck on that round.

So a prediction tournament is like a golf tournament where no-one knows "par". That is, you can assess people's relative abilities, but we do not have any external standard to assess absolute abilities.

And the real world?

We've seen the mathematics, but what is the bigger picture? After all, one could just say it's obvious that some people will be better than others at geopolitical forecasts, just as some people are better than others at golf.

To me it is self-evident that one should make predictions about uncertain future events in terms of probabilities rather than Yes/No predictions. So it is curious that, outside of gambling-like contexts, this is rarely done. Indeed the only common context where one sees numerical probabilities expressed is the chance of rain tomorrow.

A major inspiration for current interest in this topic has been the work of Philip Tetlock. His 2006 book [2] looks at extensive data on how good geopolitical forecasts from political experts have been in the past (short answer: not very good). That book contains more mathematics along the "how to assess prediction skill" theme of this article.

What makes some people are better than others at forecasting, and can we learn from them? That is the topic of Tetlock's 2015 book [3], which reports in particular on an IARPA [5] sponsored study of a prediction tournament similar to the current one [1], though where participants were assigned to teams and encouraged to discuss with teammates. Their conclusions relate success to both cognitive style of individuals and to team dynamics.

Finally, readers of this magazine may be interested in a recent paper [4] claiming that Canadian strategic forecasters are better than their U.S. counterparts!

References

- [1] Good Judgment Open. www.gjopen.com
- [2] Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton University Press (2006).
- [3] Philip E. Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction*, Crown (2015).
- [4] David R. Mandela and Alan Barnes. Accuracy of forecasts in strategic intelligence, *Proceedings of the National Academy of Science* 111 10984 – 10989 (2014).
- [5] Intelligence Advanced Research Projects Activity. <https://www.iarpa.gov>

About the author

David Aldous has been Professor in the Statistics Dept at U.C. Berkeley, since 1979. A central theme of his research in mathematical probability is the study of large finite random structures, obtaining asymptotic behavior as the size tends to infinity via consideration of a suitable infinite random structure. He has recently become interested in articulating critically what mathematical probability says about the real world.



SNAKES & LADDERS

Playing Time

BY DODDY KASTANYA

A math enthusiast working as a Nuclear engineer

INTRODUCTION

The snake-and-ladder game is a traditional game which is well-known worldwide. It can be played by two or more participants. It is a board game which has a 10-by-10 grid. Each block in this grid is numbered, starting from 1 all the way to 100. Each player has a unique game piece to mark his/her position on the board. At the beginning of the game, all players are outside the board.

The number of steps taken will depend on the outcome of the roll of a die (or a pair of dice). If the game piece stops at the bottom of the ladder, the game piece will advance to the top of the ladder bringing the player closer to the goal (i.e., block number 100). If the game piece lands on the head of a snake, the game piece will be “eaten” by the snake bringing it down to its tail. As the game piece gets closer to the ultimate goal, the roll of the die needs to match the exact steps to reach the 100 block; otherwise, the game piece will not be moved and the player loses his/her turn. For example, when the game piece is at 97, the player needs to roll 3 (or less) in order to advance.

The rule of this game is straightforward. However, depending on how lucky (or unlucky) you are, you could be stuck playing this game for a long time; especially, if there are several players involved. So, what would be nice is to know what the typical playing time for this game is. This is exactly what we would like to do in this paper. So, without further ado, let's roll the dice and get on with it.

FINDING OUT TYPICAL PLAYING TIME

To estimate a meaningful average playing time, we would like to run a significant number of trials, millions of them if possible. For this purpose, we will recruit the help of computers to do the experiments for us.

There are a few things worth mentioning before introducing the algorithm to simulate a game. Firstly, the simulation is done only for a single player because it does not matter how many players are involved in the game, the game ends when a player reaches the goal (i.e., the 100 block). Secondly, though we can choose a pair of dice, a single die will be used in the evaluation. Thirdly, it is assumed that it takes five seconds to complete each turn; that is the time to roll the die and to move the game piece.

Algorithm for simulating a single game:

1. The game piece is placed outside of the board. This is equivalent of setting 0 as the current position for the game piece.
2. A random number between 1 and 6 is generated to simulate the roll of a die. The game piece is then moved according to the resulting number.
3. Four things can now happen. First, if the resulting position is a regular block, the player waits for his/her next turn. In this simulation, the player starts the next step by rolling the die again. Second, if the resulting position for the game piece is equal to one of the numbers marking the bottom of a ladder (the position of the bottom and top of each ladder has been determined prior to starting the game; they are determined as a part of setting up the board), then the position is updated to the location corresponding to the top of that particular ladder. Third, if the resulting position of the game piece is equal to one of the numbers marking the head of a snake, then the position is updated to the location corresponding to the tail of that particular snake. Finally, if the resulting position is a number higher than 100, the player loses his/her turn. In any case, the next step is executed once the final position is established. In the algorithm world, it is usually written in terms of a “go to” command. In this case: “go to Step 2”. The game is over when the resulting position is exactly 100.

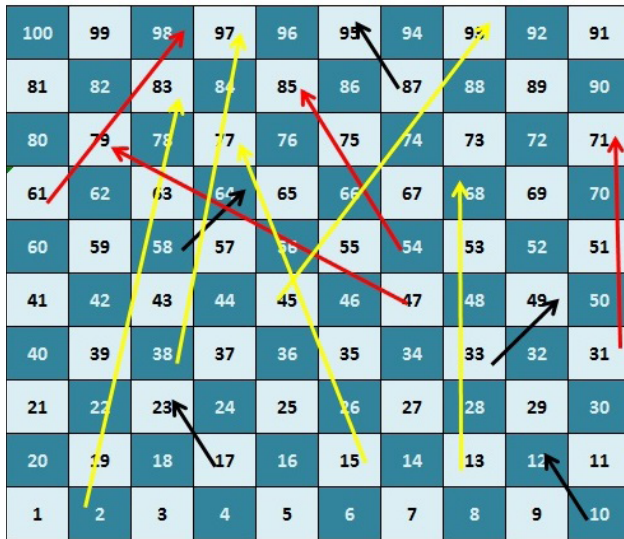


Fig. 1: Basic Ladder Configurations

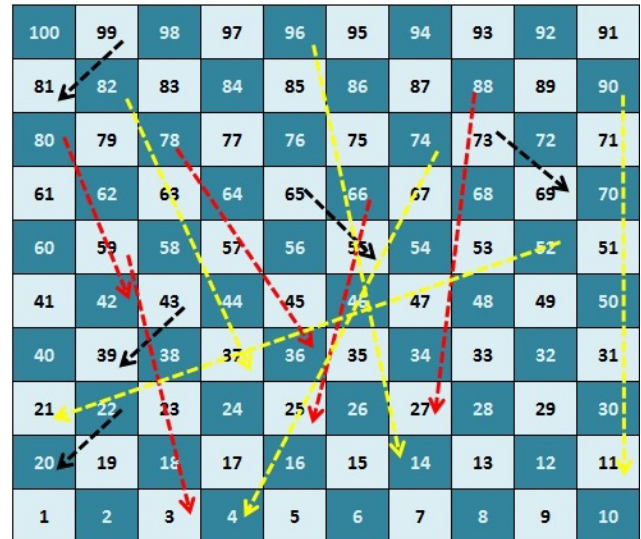


Fig. 2: Basic Snake Configurations

Now that the rules of the simulation have been established, we need to determine the number of simulations needed to get a meaningful average playing time and the number of sets needed to calculate confidence level for the estimated average playing time. For this particular game, 100 games per set are deemed sufficient and 1000 sets of 100 games will be used to build the confidence level.

The last piece of the puzzle is related to the design of the board itself. I am sure you have seen many variations of the board. Besides the artistic component of the design, the main differences among these boards lie on the number of snakes and ladders as well as their lengths. In the current evaluation, we would take a look at a board design which has exactly five snakes and five ladders. So, we need to introduce three basic configurations for the snakes and three basic configurations for the ladders.

These options and the possible combinations are nicely summarized in Table 1 (page 11). The name of the board reflects the size of the ladder followed by the size of the snake. For example, “S-L” means the board has five short ladders and five long snakes. Figure 1 shows a board with short ladders (solid-black lines), medium ladders (solid-red lines), and long ladders (solid-yellow lines). Similarly, Figure 2 depicts a board with short snakes (dotted-black lines), medium snakes (dotted-red lines), and long snakes (dotted-yellow lines). These basic configurations are then combined to get the nine configurations identified in Table 1.

RESULTS

Table 2 (page 11) summarizes the average playing time for various combinations of ladder and snake sizes. There are a couple of observations that could be made on the results presented in this table:

1. For a given size of snake (S, M, or L), the longer the size of the ladders, the shorter the average playing time will be. This result is not unexpected since a fix size of snakes will bring about the same chance of being eaten and pushed back toward the starting point. However, having a longer size ladder will help accelerate the game piece back toward the ultimate goal.
2. A similar observation cannot really be made for a given size of ladder. For the sets of short ladders, an observation could be made that the longer the snakes, the longer the average playing time will be. This is not unexpected since the longer the snakes, the chance of the game piece moves toward the beginning of the board. Once a game piece is “eaten”, the acceleration toward the goal by taking advantage of the existing ladders is practically unchanged since the layout of the ladder is fixed. However, this generalization does not extend to the medium and long ladders. For these cases, the combination with medium size snake requires the least amount of playing time. This result suggests that there is an optimum combination of the length of snakes and the length of ladders. Moreover, the placement of the ladders and snakes plays an important role in determining the playing time as well.

Size of Ladders			Size of Snakes			Case Name
Short	Medium	Long	Short	Medium	Long	
X			X			S-S
X				X		S-M
X					X	S-L
	X		X			M-S
	X			X		M-M
	X				X	M-L
		X	X			L-S
		X		X		L-M
		X			X	L-L

Table 1: Combination of Ladders and Snakes

Snake Sizes	Ladder Sizes		
	Short	Medium	Long
Short	256	214	178
Medium	626	210	151
Long	860	388	234

Table 2: Estimated Average Playing Time (in seconds)

Figure 3 (page 12) shows the variation of the estimated playing time for a specific combination of ladder and snake sizes, namely short ladders and medium snakes. The x-axis of this figure shows the estimated playing time (expressed in seconds) and the y-axis shows the frequency - or the number of occurrences - that falls under each category in the x- axis.

The average playing time for this particular combination is 626 seconds and it is located in the middle of the distribution. From this figure, one can see that while the game could take as long as around 800 seconds, there are still chances to finish the game in less than 500 seconds.

CAN WE MAKE IT QUICKER?

Unless you have ten or more friends who would like to play a single game of snakes-and-ladders together, or results indicate that this game is not too time-consuming.

An average game involving four people would likely last about an hour or so (provided that the lengths of the ladders and snakes are similar to the ones used here).

However, we still need to mention some ideas which could make this game end faster.

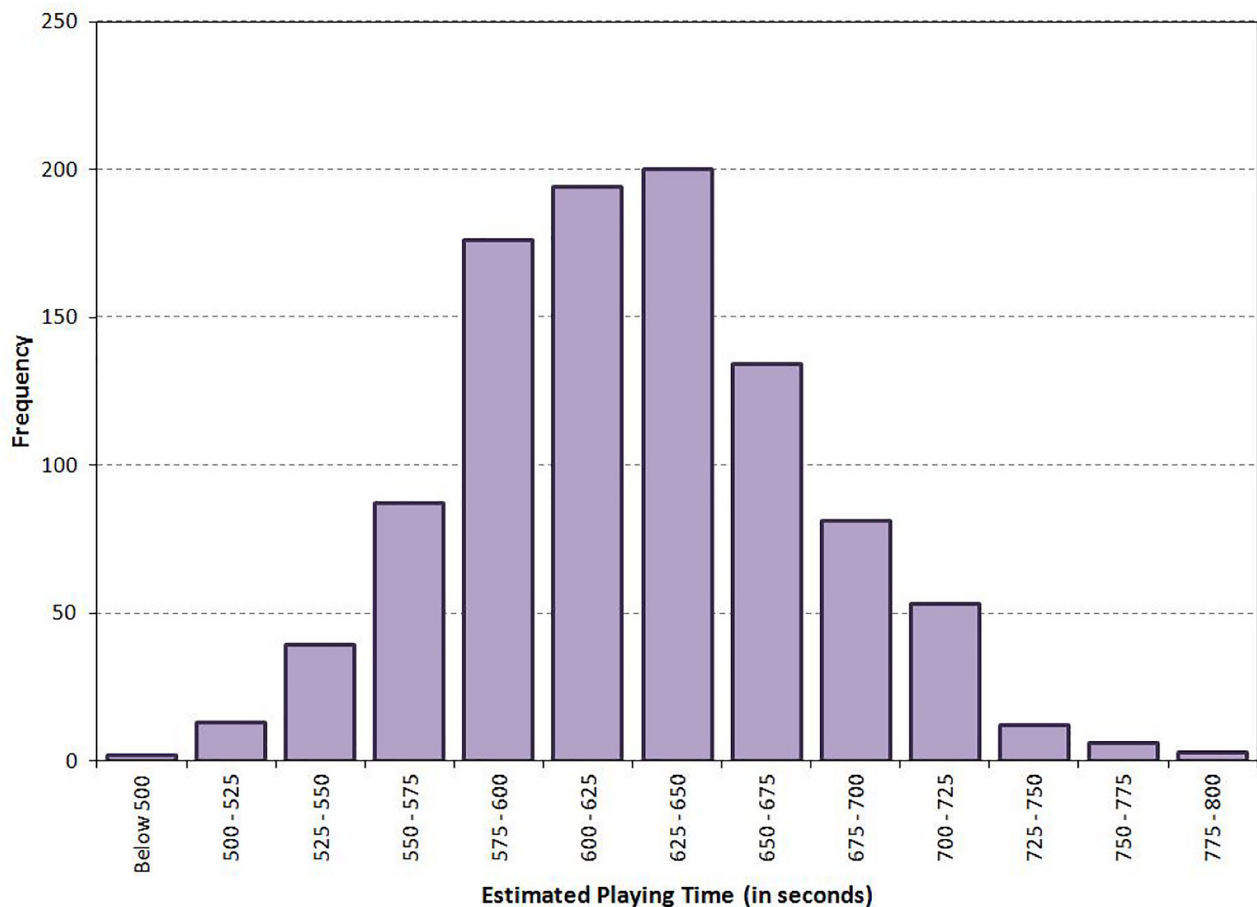


Fig. 3: Summary of Estimated Playing Times for a Combination of Short Ladders and Medium Snakes

One would expect the game to finish faster when:

- The number of ladders is increased.
- The number of snakes is reduced.
- The sizes of the ladder are long.
- The sizes of the snake are short.
- We have significantly more ladders than snakes.

FINAL NOTES

The results presented in this paper are based on some idealized situations where certain uniformity was assumed. In real life, the expected average playing time would vary depending on various aspects such as the number of snakes, the number of ladders, non-uniform length of snakes and/or ladders, the distribution of the snakes and/or ladder, the number of player playing the game, and the use of a pair of dice.

Of course, in real life people (read: kids) do not really care about how long the game will last since they can stop it anytime they want to or when one of the players gets frustrated from being eaten by a snake one too many times... They just roll the die and have fun.

BIG DATA

BY CHRIS BUDD

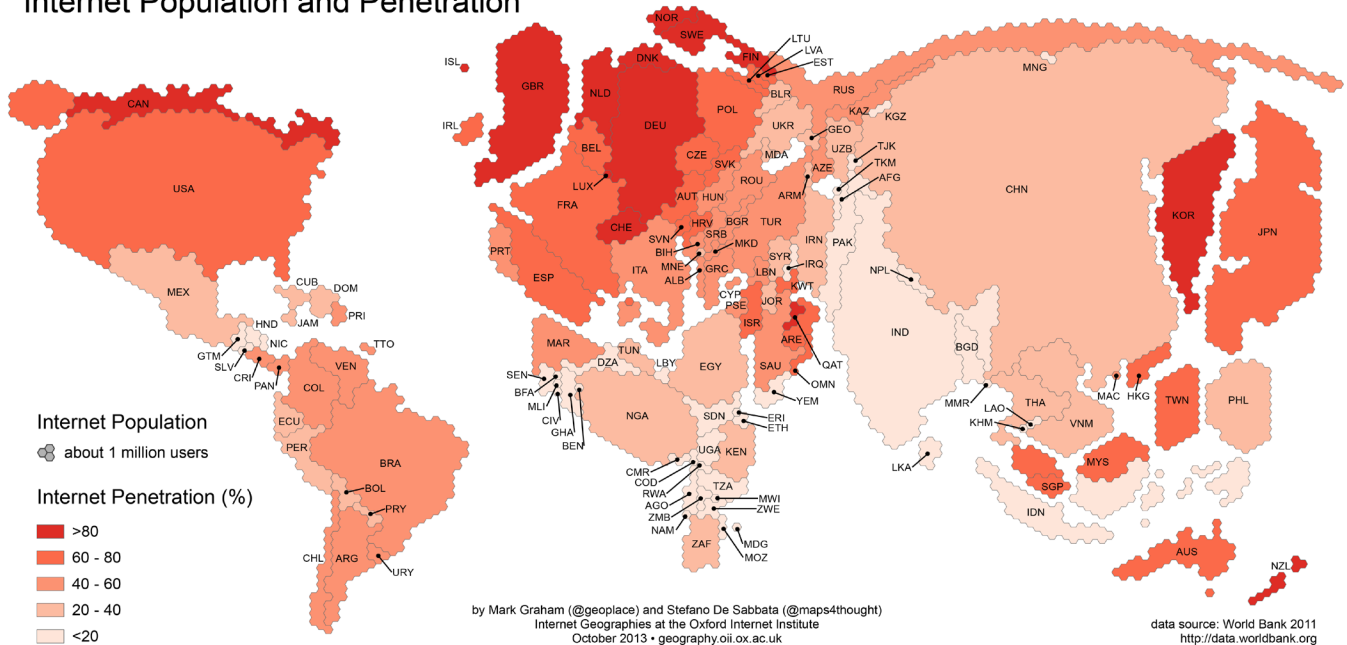
This article first appeared in Plus Magazine <http://plus.maths.org>.

We live in the information age. Most of what we do is hugely influenced by our access to massive amounts of data — whether this is through the Internet, on our computers, or on our mobile phones. The buzz word to describe this deluge of information is Big Data. In 2012 the UK government identified Big Data as one of the eight great technologies of the future. So what does the challenge of Big Data entail and how can we meet it?

WHERE DOES BIG DATA COME FROM

Perhaps the leading source of current Big Data is the Internet. According to a recent estimate, about 10^{21} bytes (a zettabyte) of information are added to the Internet every year, much of which is graphical in content. The internet penetration in the UK is over 80%, and in all but a few countries it is over 20%.

Internet Population and Penetration



This map illustrates the total number of Internet users in a country as well as the percentage of the population that has Internet access in 2011. Image: Stefano De Sabbata and Mark Graham, CC BY 3.0.

A major source of this data is the ever growing content on social media websites. For example, Facebook was launched in 2004. It now has 2 billion registered users (about a quarter of the world's population!) of which 1.5 billion are active. Around 2.5 billion pieces of content (around 500 terabytes of information) are added every day to Facebook, with most of this data stored as pictures. The search engine Google is estimated to seek information from around 15 exabytes (10^{15} bytes) of data (which it searches using a clever mathematical algorithm).



Smart watches can give you up-to-date health information wherever you are.

Another source of Big Data are mobile and smart phones. There are now more mobile phones than people in the world, with the potential for 25,000,000,000,000,000 possible simultaneous conversations. The forthcoming plans for a 5G network will offer data rates at 1 gigabyte per second simultaneously to tens of workers on the same office floor.

Another fast-approaching technology are sensors that can provide constant monitoring of, say, our state of health (with significant ethical implications). The 5G network will support several hundreds of thousands simultaneous connections for massive sensor deployments. Indeed the future is rapidly approaching: soon our devices will simply communicate with each other (for example the cooker talking to the dishwasher and also to the supermarket every time a meal is prepared) with little or no human interference — it's called the Internet of things.

Significant amounts of data, of significant interest to the social sciences, also come from the way that we use our devices and the information this gives about our lifestyles. Again there are significant ethical issues here. Every time we make a purchase on Amazon, use our bank on-line, switch on an electrical device, or simply use a mobile phone or write an email, we are creating data which contains information that can in principle be analysed. For example, our shopping habits can be determined, or our location tracked and recorded. Mathematics can be used at all stages of this, but we must never lose sight of the moral dimension in so doing.

THE NATURE OF BIG DATA

In one sense, Big Data has been the subject of mathematical investigation for at least 100 years. A classical example is meteorology, in which huge amounts of numbers need to be crunched to produce reliable weather forecasts. Similarly large data sets arise in climate models, geophysics and astronomy.

However, the data sets in these problems, while very large, are also well-structured and well-understood, with known levels of uncertainty. That's because they come from physical processes that, on the whole, scientists understand well.

The real challenges of understanding and dealing with Big Data come from the biological sciences, the social sciences and in particular from people-based activity. Such data is often garbled, incomplete, unreliable, complex, anecdotal and fast-arriving. It is often qualitative rather than quantitative, it isn't homogeneous, and it's about relations between things, rather than the things themselves, in a way that physical data isn't.

WHAT QUESTIONS DO WE WANT TO ASK OF BIG DATA?

How do we visualise Big Data, make speculations from it, model it, and understand it? How do we experiment on the systems that generate it, and ultimately how might we control those systems? The mathematical and scientific challenges behind these questions are as varied as they are important, and the very scale of Big Data makes automation necessary. This automation in turn relies on mathematical algorithms.

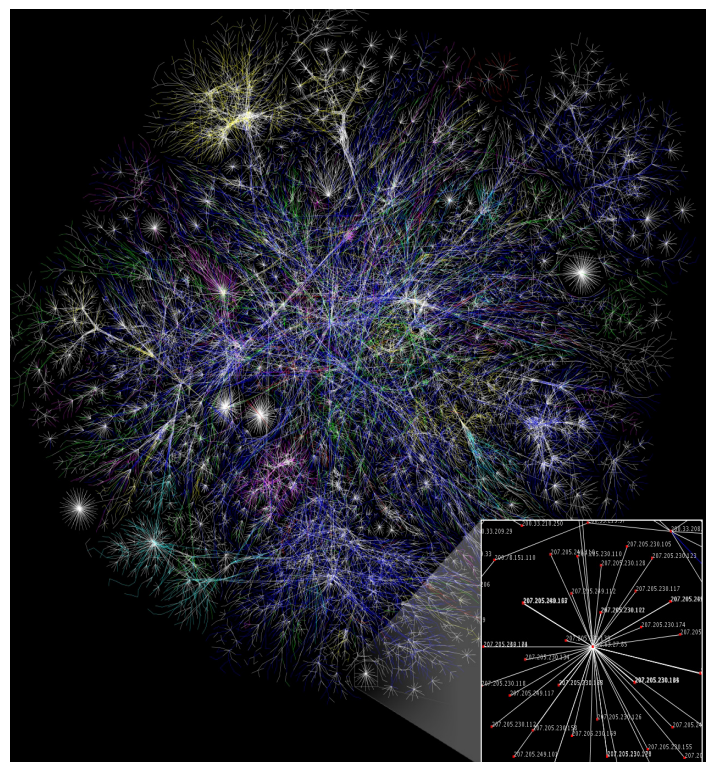
Questions we might ask from Big Data include:

- How do we rank information from vast networks in web browsers such as Google?
- How do we identify consumer preferences, loyalty or even sentiment and how do we make personalised recommendations?
- How do we model uncertainties in health trends for individual patients?
- How do we achieve and deal with real-time health monitoring (especially in the environment that 5G will lead to)?
- How to use smart data in energy supplies?

It is fair, I think, to say that many of the future advances in modern mathematics (together with theoretical computer science) will either be stimulated by the applications of Big Data or driven by the needs to understand Big Data. Many existing mathematical techniques (some of which until recently were considered as pure mathematics) are now finding significant applications in our understanding of Big Data. A key example of this is the mathematics of network theory.

NETWORKS EVERYWHERE

As the name suggests, network theory describes objects, called nodes, that are linked together by what are called edges. The nodes could be computers or websites, and the edges connections between the computers or links between the websites. The nodes can also be people and the connections their friends on Facebook or Twitter, or they could be mobile hand sets and the links conversations or simply a close proximity which might lead to interference. Network theory explains the nature of networks, allows us to search for connections between individual points in data sets, and can describe the movement of information around a network.



A partial map of the Internet based on 2005 data found on The Opte Project. CC BY 2.5.

Indeed, managing the mobile phone network (which is of course also used to download data) is a significant and growing application of the theory of graph colouring: finding ways of colouring the edges or nodes in a network according to specific constraints, such as adjacent nodes having to have different colours. For example, the colours might represent frequencies assigned to mobile phone transmitters, which have to be chosen to minimise interference and so need to be different for adjacent transmitters. Graph colouring was until recently regarded as belonging firmly in the domain of pure mathematics.

Other examples of networks which lead to big data include organisational networks (such as management networks, crime syndicates, even the voting behaviour in the Eurovision Song contest), technological networks (such as the power grid or electronic circuits), information networks (made up of genes, protein-protein interactions, word-of-mouth dissemination of information, myths and rumours), transport networks (such as airlines, food logistics, underground and overground rail systems), and ecological networks (food chains, diseases and infection mechanisms).

THE POWER OF NETWORK THEORY

Network theory can address many other questions related to Big Data. When you are dealing with very large networks it is not always easy to identify clusters — groups of nodes that are highly interlinked — or to segment the data into groups that share common features. Such information is vital in data mining and pattern recognition. It is especially relevant to the retail industry, who are interested in the behaviour and preferences of their customers, but can also be used to identify friendship groupings in social networks, to investigate the organisation of the brain, and even to finding voting patterns in the Eurovision Song contest. Network theory provides the algorithms both for identifying clusters and for segmenting data.



Information, gossip, infectious diseases: they all spread through social networks.

Such analysis can also help with another very significant problem encountered in many applications: linking data bases with different levels of granularity in space and time. An example is weather forecasting, where some of the data might be coming from satellites orbiting the Earth and transmitting MBytes of data a second. Other data might be from individuals in isolated ground stations who might only give a few measurements every day. Some of the data might even be historical such as records of sea captains 100 years ago. All three such data sets are useful and they have to be linked together in a seamless manner.

Equally important is the question of how connected a network is on the whole: are individual nodes connected to many others throughout the network, or are the connections sparse? What is the shortest path through the network? These questions are essential for efficient routing in the Internet, interpretation of logistic data, understanding the speed of word-of-mouth communications and even marketing. Network theory is also essential in searching for influential nodes in huge networks. Highly connected nodes — whether they represent people, websites, or airports — are hugely important for the robustness of a network, since removing them would significantly alter its overall connectedness. Such information can be used to break up terrorist organisations, stop epidemics from spreading, or keep air traffic rolling when a region is affected by bad weather.

WHAT ELSE CAN MATHS DO?

Network theory is just one of a variety of mathematical techniques used to study Big Data. Much of Big Data takes the form of images, so mathematical algorithms that classify, interpret, analyse and compress images are extremely important. Statistical methods have long been used to analyse and interpret images, but there has recently been a significant growth in novel mathematical algorithms, drawing on ideas from pure mathematics people previously thought had no direct applications in the real world. Some of these algorithms are based on the analysis of complicated equations, leading to some powerful and unexpected applications of highly technical tools from the relevant theory of equations. Algebraic topology, an area of maths that investigates properties of shapes using algebra, plays a very useful role in classifying images. And techniques from category theory, an area that investigates mathematical structures and concepts on a highly abstract level, can be used to “parse” an image to see how the various components fit together. In the context of machine learning this allows for machines to “perceive” what the objects in an image are and to make “reasoned” decisions about it.

This is only a short list. There are many other areas of mathematics and computer science that have also found applications in the study of Big Data. So watch this space! I am confident that we will see great advances in pure, applied and computational maths arising from these challenges.

ABOUT THE AUTHOR

Chris Budd OBE is Professor of Applied Mathematics at the University of Bath, Vice President of the Institute of Mathematics and its Applications, Chair of Mathematics for the Royal Institution and an honorary fellow of the British Science Association. He is particularly interested in applying mathematics to the real world and promoting the public understanding of mathematics.

He has co-written the popular mathematics book *Mathematics Galore!*, published by Oxford University Press, with C. Sangwin, and features in the book *50 Visions of Mathematics* ed. Sam Parc.



WORLDS IN YOUR POCKET: 2D Screens to 3D Touch

BY HARRISON CUI

Undergraduate student in Computer Science at Cornell

A simple touchscreen can drop you right in the middle of a football game or a psychological thriller. Growth in the mobile gaming space and advances in the processing power of handheld devices have led to the development of complex 3D games for cell phones and tablets. Popular games like Madden, Five Nights at Freddy's, and Minecraft Pocket Edition enable players to interact with complex 3D worlds, all through a small 2D screen. How do they do it?

The touchscreen is your window into the game's world. The 3D world is projected onto the 2D screen in a specific viewing direction, from a certain location, which can be changed to show a different perspective of the 3D world. This is similar to using a camera to take a picture of an object: moving or turning the camera shows a different picture of the object. When you touch the screen, the game engine must figure out which object you want to interact with. This task is handled by a method known as ray casting, in which a ray is projected from the touch point on a 2D display screen into the imaginary 3D world generated beyond the screen. The ray continues until it encounters an object, and the game engine then knows with which object you wish to interact.



Fig. 1

This method enables gamers to select blocks with a simple touch of the screen as shown in Figure 1. Mathematically, a ray can be described by

$$\langle\langle x_0, y_0, z_0 \rangle\rangle + t \langle\langle v_x, v_y, v_z \rangle\rangle$$

where (x_0, y_0, z_0) is the point the ray starts from and $\langle\langle v_x, v_y, v_z \rangle\rangle$ is the direction the ray points in. The variable t here is a “parameter”—corresponding points along the ray are obtained by substituting values of t from zero up to the smallest value of t where the ray hits something.

The difficult part is thinking up an efficient method for ray casting. Traditional ray casting has a major drawback: it is computationally expensive and can very easily take up 95% of rendering time. A more practical method for finding the closest object in the direction of the casted ray is known as the Voxel Traversal Algorithm [1].

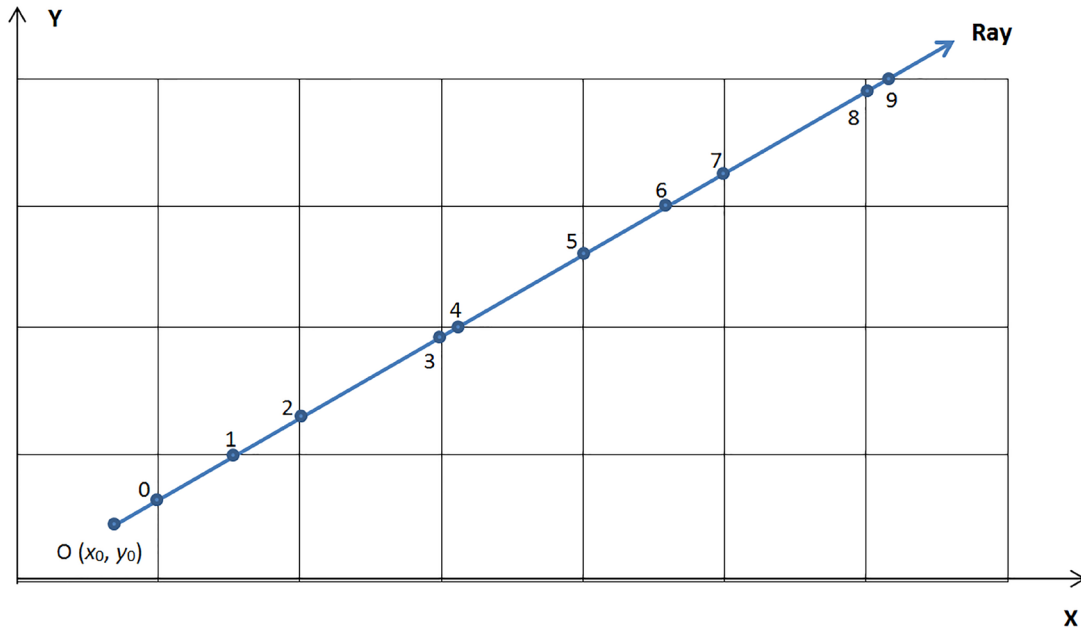


Fig. 2

We divide the 3D world into small cubes with a unit length of one. These cubes are generally referred to in the game community as voxels. To make the algorithm more understandable, we can present it in 2D as shown in Figure 2 where the voxels are represented as squares rather than cubes.

Imagine the ray as a particle starting out at the initial touch point—the origin of the ray, (x_0, y_0) —and moving in the positive x -direction with a velocity v_x and in the positive y -direction with a velocity v_y . Voxels that do not contain objects represent air blocks the particle can freely pass through, whereas voxels that do contain objects represent impassible blocks the particle cannot move through. The particle's goal is to detect the first voxel it cannot pass. The particle can go no further than the borders of an impassible voxel; thus, the algorithm only requires the particle to check the edges of the voxels to determine whether an object exists at that voxel.

However, this presents an interesting problem: how can we effectively calculate the distance the particle must travel to reach the next voxel in its linear path? Because we know the square voxels have sides with a unit length of one, the particle must either have an x or y coordinate of integer length when it encounters the border of a voxel. Then, to determine the first voxel the particle encounters, we must find the first integer x or y coordinate that the particle reaches. We are able to calculate the time, t_{x0} , for the particle to reach the first integer x -coordinate, with the equations:

$$t_{x0} = \begin{cases} \frac{\lceil x_0 + 1 \rceil - x_0}{v_x} & \text{if } v_x > 0; \\ \frac{\lfloor x_0 - 1 \rfloor - x_0}{v_x} & \text{if } v_x < 0. \end{cases}$$

Similarly, we are able to calculate the time, t_{y0} , for the particle to reach the first integer y -coordinate, with the equation:

$$t_{y0} = \begin{cases} \frac{\lfloor y_0 + 1 \rfloor - y_0}{v_y} & \text{if } v_y > 0; \\ \frac{\lfloor y_0 - 1 \rfloor - y_0}{v_y} & \text{if } v_y < 0. \end{cases}$$

The smaller time value represents the time it takes for the particle to encounter its first new voxel. If either v_x or v_y is zero, its respective times go to infinity. The time for the other one will always be less than infinity, so the ray will continue traveling only in one direction. Suppose t_{x0} is less than t_{y0} . Then the first voxel the particle encounters will be in time t_{x0} . When it reaches this voxel, we know the next coordinate with an integer x -coordinate it approaches will be a distance of 1 unit away — the length of a side of the voxel. The additional time, t_x , required to reach this point can be calculated with the equation:

$$t_x = \frac{1}{v_x}$$

Thus, the total time it takes to reach this integer x -coordinate is simply:

$$t_{x0} + t_x$$

And the total time it takes to reach an n th integer x -coordinate can be expressed as:

$$t_{x0} + nt_x$$

Similarly, with the same logic, we can represent these equations in terms of y as well:

$$t_{y0} = \frac{1}{v_y}$$

$$t_{y0} + t_y$$

$$t_{y0} + nt_y$$

Thus, we have a set of numbers that represents the times it takes for the particle to reach each consecutive integer x -coordinate along the ray:

$$\{ t_{x0}, t_{x0}+t_x, t_{x0}+2t_x, t_{x0}+3t_x, \dots, t_{x0}+nt_x \}$$

Similarly, we have a set of numbers that represents the times it takes for the particle to reach each consecutive integer y -coordinate along the ray:

$$\{ t_{y0}, t_{y0}+t_y, t_{y0}+2t_y, t_{y0}+3t_y, \dots, t_{y0}+nt_y \}$$

To move the particle to a new voxel along the ray, we first calculate the times it takes to reach the next integer x -coordinate and y -coordinate. For example these times are $t_{x0} + t_x$ and t_{y0} at the next step. The smaller of the two is the time it takes for the particle to move to the neighboring voxel. Because we know the times it takes for the particle to reach each new voxel, calculating the distance the particle must travel to reach them is simple.

$$x = v_y t$$

$$y = v_x t$$

Now that we have described the algorithm in a 2D case, we can easily extend this to a 3D case by simply adding a z-axis.

Thus, the algorithm presents an efficient method of determining the point in space at which an object exists. Using a minimal number of computations, it determines the distance to the next voxel and checks whether that voxel is an air block or an obstacle. If it's an air block, the algorithm continues to the next voxel; however, if there's an obstacle, the algorithm terminates.

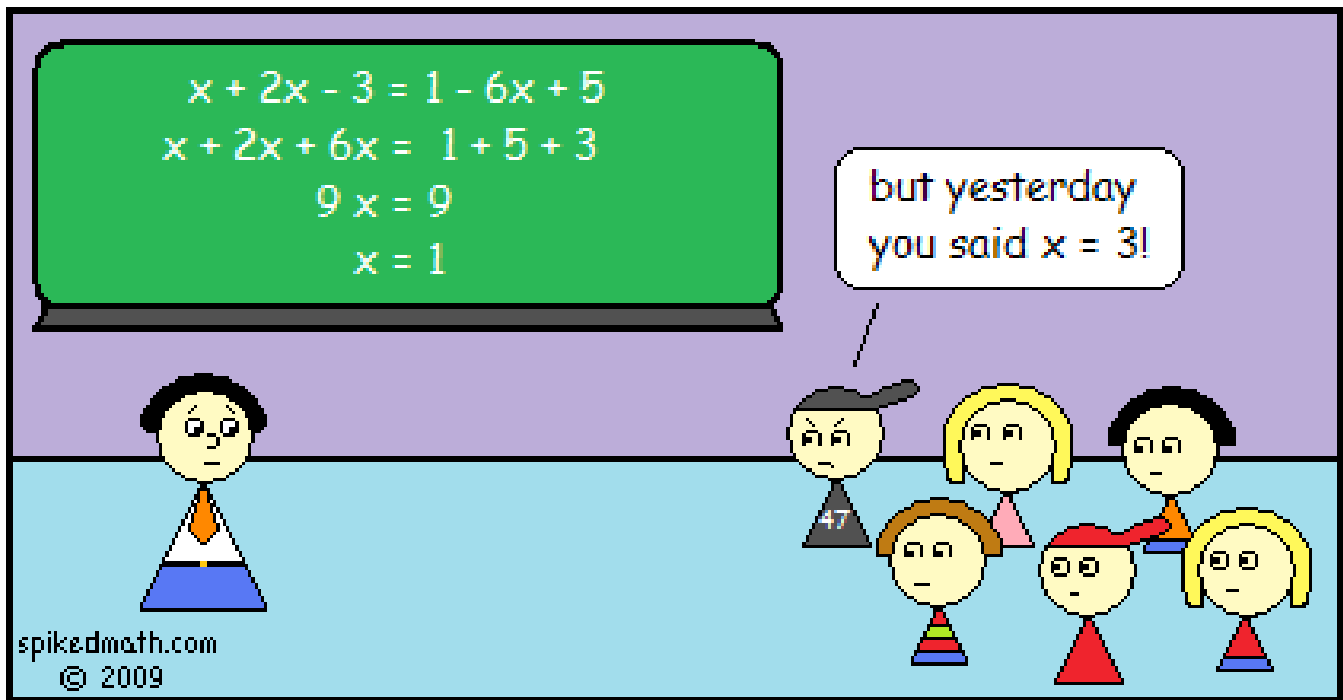
Since voxels in video games are generally tiny, further optimizations such as clumping voxels together into super-voxels (maybe a 100x100x100 cube of voxels) can make this algorithm even faster. These super-voxels can then be divided into two groups: the empty super-voxels containing all air blocks and the non-empty super-voxels containing some obstacles. This enables the same checking system on a larger scale, but we can now check less often.

We can apply similar optimization to various areas other than math and computer science—like determining the point in space in which breakfast exists after waking up late for school. Next time you do anything, take a lesson from voxels and try discovering a more efficient method. You might free up time for some math!

REFERENCES

[1] John Amanatides, Andrew Woo. A Fast Voxel Traversal Algorithm for Ray Tracing. <http://www.cse.yorku.ca/~amana/research/grid.pdf>

[2] Cast ray to select block in voxel game <http://gamedev.stackexchange.com/questions/47362/cast-ray-to-select-block-in-voxel-game>



SIMPLE SET GAME PROOF STUNS MATHEMATICIANS

BY ERICA KLARREICH FOR QUANTAMAGAZINE.ORG

A NEW SERIES OF PAPERS HAS SETTLED A LONG-STANDING QUESTION related to the popular game in which players seek patterned sets of three cards.



Image Credit: Olena Shmahalo, Quanta Magazine

In a series of papers posted online in recent weeks, mathematicians have solved a problem about the pattern-matching card game Set that predates the game itself. The solution, whose simplicity has stunned mathematicians, is already leading to advances in other combinatorics problems.

Invented in 1974, Set has a simple goal: to find special triples called “sets” within a deck of 81 cards. Each card displays a different design with four attributes — color (which can be red, purple or green), shape (oval, diamond or squiggle), shading (solid, striped or outlined) and number (one, two or three copies of the shape). In typical play, 12 cards are placed face-up and the players search for a set: three cards whose designs, for each attribute, are either all the same or all different.

Occasionally, there’s no set to be found among the 12 cards, so the players add three more cards. Even less frequently, there’s still no set to be found among the 15 cards. How big, one might wonder, is the largest collection of cards that contains no set?

The answer is 20 — proved in 1971 by the Italian mathematician Giuseppe Pellegrino. But for mathematicians, this answer was just the beginning. After all, there's nothing special about having designs with only four attributes — that choice simply creates a manageable deck size. It's easy to imagine cards with more attributes (for instance, they could have additional images, or even play different sounds or have scratch-and-sniff smells). For every whole number n , there's a version of Set with n attributes and 3^n different cards.

For each such version, we can consider collections of cards that contain no set — what mathematicians confusingly call “cap sets” — and ask how large they can be. Mathematicians have calculated the maximal size of cap sets for games with up to six attributes, but we'll probably never know the exact size of the largest cap set for a game with 100 or 200 attributes, said Jordan Ellenberg, a mathematician at the University of Wisconsin, Madison — there are so many different collections of cards to consider that the computations are too mammoth ever to be carried out.

Yet mathematicians can still try to figure out an upper bound on how big a cap set can be — a number of cards guaranteed to hold at least one set. This question is one of the simplest problems in the mathematical field called Ramsey theory, which studies how large a collection of objects can grow before patterns emerge.

“The cap set problem we think of as a model problem for all these other questions in Ramsey theory,” said Terence Tao, a mathematician at the University of California, Los Angeles, and a winner of the Fields Medal, one of mathematics' highest honors. “It was always believed that progress would come there first, and then once we'd sorted that out we would be able to make progress elsewhere.”

Yet until now, this progress has been slow. Mathematicians established in papers published in 1995 and 2012 that cap sets must be smaller than about $1/n$ the size of the full deck. Many mathematicians wondered, however, whether the true bound on cap set size might be much smaller than that.

They were right to wonder. The new papers posted online this month showed that relative to the size of the deck, cap set size shrinks exponentially as n gets larger. In a game with 200 attributes, for instance, the previous best result limited cap set size to at most about 0.5 percent of the deck; the new bound shows that cap sets are smaller than 0.000043 percent of the deck.

Previous results were “already considered to be quite a big breakthrough, but this completely smashes the bounds that they achieved,” said Timothy Gowers, Fields medalist and mathematician at the University of Cambridge.

There's still room to improve the bound on cap sets, but in the near term, at least, any further progress is likely to be incremental, Gowers said. “In a certain sense this completely finishes the problem.”

GAME, SET, MATCH

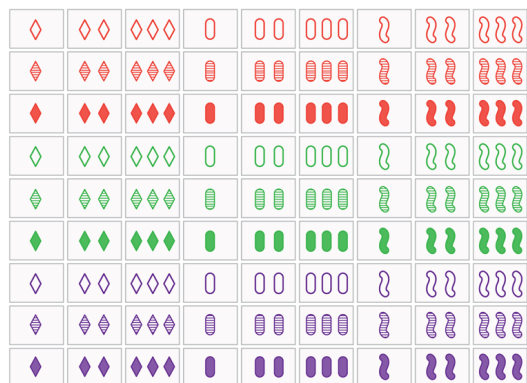
To find an upper bound on the size of cap sets, mathematicians translate the game into geometry. For the traditional Set game, each card can be encoded as a point with four coordinates, where each coordinate can take one of three values (traditionally written as 0, 1 and 2). For instance, the card with two striped red ovals might correspond to the point (0, 2, 1, 0), where the 0 in the first spot tells us that the design is red, the 2 in the second spot tells us that the shape is oval, and so on. There are similar encodings for versions of Set with n attributes, in which the points have n coordinates instead of four.

The rules of the Set game translate neatly into the geometry of the resulting n -dimensional space: Every line in the space contains exactly three points, and three points form a set precisely when they lie on the same line. A cap set, therefore, is a collection of points that contains no complete lines.

SET POINT

The goal of Set is to find special triples called “sets” within a deck of 81 cards. Each card displays a different design with four attributes — color (red, purple or green), shape (oval, diamond or squiggle), shading (solid, striped or outlined) and number (one, two or three copies of the shape). In typical play, 12 cards are placed face-up and the players search for sets of three cards whose designs, for each attribute, are either all the same or all different. Occasionally, there's no set among the 12 cards, so more cards are added. A collection of cards with no set is called a cap set.

THE FULL DECK



SET OR NO SET

Some examples below:

Are the attributes all the same or all different?				
Color	×	×	all different	all the same
Shape	all different	all different	all different	all the same
Shading	all different	all different	all different	all the same
Number	×	all different	all different	all different
	Not a set	Not a set	A set	A set

CAP SET

A simple way to build a fairly large cap set is to include only cards that show two of the three choices for each attribute. This cap set will be $(2/3)^n$ as big as the whole deck, where n is the number of attributes.



Image Credit Lucy Reading-Ikkanda for Quanta Magazine

Previous approaches to getting an upper bound on cap set size used a technique called Fourier analysis, which views the collection of points in a cap set as a combination of waves and looks for the directions in which the collection oscillates. “The conventional wisdom was that this was the way to go,” Tao said.

Now, however, mathematicians have solved the cap set problem using an entirely different method — and in only a few pages of fairly elementary mathematics. “One of the delightful aspects of the whole story to me is that I could just sit down, and in half an hour I had understood the proof,” Gowers said.

The proof uses the “polynomial method,” an innovation that, despite its simplicity, only rose to prominence on the mathematical scene about a decade ago. The approach produces “beautiful short proofs,” Tao said. It’s “sort of magical.”

A polynomial is a mathematical expression built out of numbers and variables raised to powers — for instance, $x^2 + y^2$ or $3xyz^3 + 2$. Given any collection of numbers, it’s possible to create a polynomial that evaluates to zero at all those numbers — for example, if you pick the numbers 2 and 3, you can build the expression $(x - 2)(x - 3)$; this multiplies out to the polynomial $x^2 - 5x + 6$, which equals zero if $x = 2$ or $x = 3$. Something similar can be done to create polynomials that evaluate to zero at a collection of points — for instance, the points corresponding to Set cards.

At first glance, this doesn’t seem like a very deep fact. Yet somehow, these polynomials often seem to contain information that isn’t readily visible from the set of points. Mathematicians don’t fully understand, Ellenberg said, just why this approach works so well, and which types of problems it can be useful for. Until a few weeks ago, he added, he considered cap set “an example of a problem where the polynomial method really has no purchase.”

That changed on May 5, when three mathematicians — Ernie Croot of the Georgia Institute of Technology, Vsevolod Lev of the University of Haifa, Oranim, in Israel, and Péter Pál Pach of the Budapest University of Technology and Economics in Hungary — posted a paper online showing how to use the polynomial method to solve a closely related problem, in which each Set attribute can have four different options instead of three. For technical reasons, this problem is more tractable than the original Set problem.

In this game variant, for any collection of cards with no set, Croot, Lev and Pach considered which additional cards could be laid down on the table to complete a set. They then built a polynomial that evaluates to zero on these completion cards, and figured out an ingeniously simple way to split the polynomial into pieces with smaller exponents, which led to a bound on the size of collections with no sets. It was a “very inventive move,” Ellenberg said. “It’s always incredibly cool when there’s something truly new and it’s easy.”

The paper soon set off a cascade of what Ellenberg called “math at Internet speed.” Within 10 days, Ellenberg and Dion Gijswijt, a mathematician at Delft University of Technology in the Netherlands, had each independently posted papers showing how to modify the argument to polish off the original cap set problem in just three pages. Yesterday, they posted a joint paper combining their results. The trick, Ellenberg said, is to realize that there are many different polynomials that evaluate to zero on a given set of points, and that choosing just the right one gets “a little bit more juice out of the method.” A cap set, the new proofs establish, can be at most $(2.756/3)^n$ as large as the whole deck.

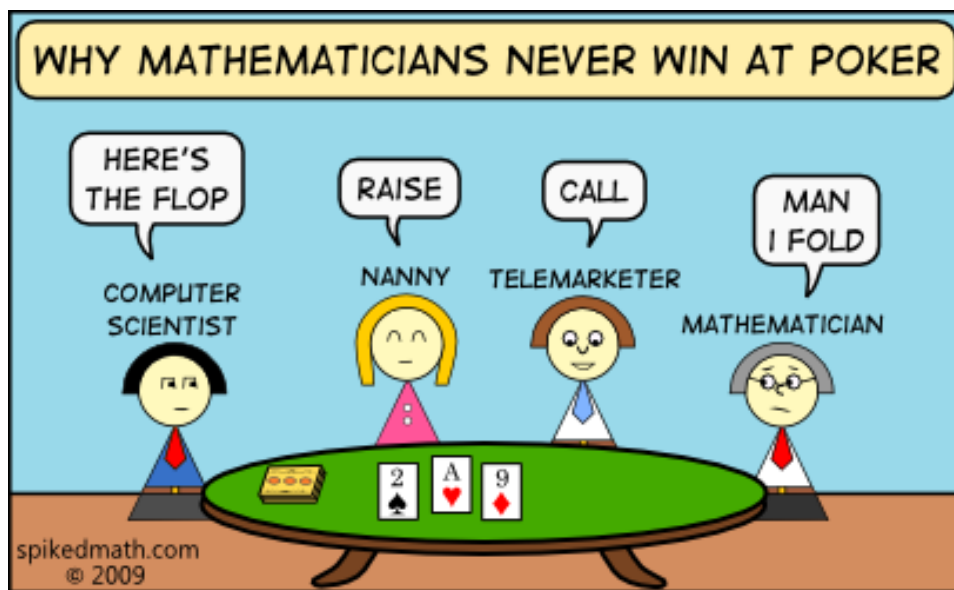
Mathematicians are now scrambling to figure out the implications of the new proof. Already, a paper has been posted online showing that the proof rules out one of the approaches mathematicians were using to try to create more efficient matrix multiplication algorithms. And on May 17, Gil Kalai, of the Hebrew University of Jerusalem, wrote an “emergency” blog post pointing out that the cap set result can be used to prove the “Erdős-Szemerédi sunflower conjecture,” which concerns sets that overlap in a sunflower pattern.

“I think a lot of people will be thinking, ‘What can I do with this?’” Gowers said. Croot, Lev and Pach’s approach, he wrote in a blog post, is “a major new technique to add to the toolbox.”

The fact that the cap set problem finally yielded to such a simple technique is humbling, Ellenberg said. “It makes you wonder what else is actually easy.”

SPECIAL THANKS TO QUANTA MAGAZINE

Reprinted with permission from QuantaMagazine.org, an editorially independent publication of the Simons Foundation whose mission is to enhance public understanding of science by covering research developments and trends in mathematics and the physical and life sciences.



THE THREE RADII PROBLEM

BY TAMAR BARABI

Teacher: Shawn Gavriel Morris, Democrati Lev HaSharon Israel

INTRODUCTION

My name is Tamar, I am a tenth-grade student at the Democratic school in Lev HaSharon, and am in the 4-5 credit mathematics course.

Some time ago, I solved an exercise in geometry, and I noticed that it could be solved in a simpler way, with a much simpler theorem, which I later realized had not yet been formulated. I checked with Shawn, my teacher, I asked relatives outside of Israel who work in mathematics, and I got advice from my parents. We sent the article to Professor Ron Livne at Hebrew University in Jerusalem, and to other experts, and I came to understand that the theorem truly had not been previously formulated anywhere, despite the fact that it is very basic and logical. With some help from Shawn and my father, who is also a teacher, I wrote the following proof with some additional conclusions.



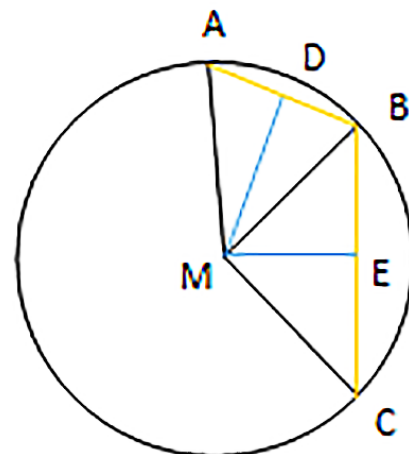
THE THEOREM:

If three or more line segments of equal length leave a single point and reach the boundary of a circle, the point is the center of the circle and the lines are its radii.

PROOF FOR THE THEOREM:

Proof: We will prove the theorem for the case of three lines

Suppose we are given a point M and three points, A , B and C lying on a circle such that $AM=BM=CM$.



We must show that **M** is the center of the **circle**, and that **MA**, **MB**, **MC** are **radii**

1. We will draw the chords **AB** and **BC**.
2. The triangles **AMB** and **BMC** are isosceles. Derived from the given information.
3. We will draw segments to the base of the isosceles triangles, **MD** and **ME**
4. **MD** and **ME** are perpendicular bisectors of the segments **AB** and **BC**, respectively. This derives from points **2** and **3**, and from the fact that “in an isosceles triangle, the perpendicular bisector to the base passes through the third vertex/
5. **ME** and **MD** pass through the center of the circle. This derives from point **4** and the theorem the perpendicular bisector to any chord in a circle passes through the center of the circle.
6. **M** is the center of the circle. This derives from point **5**, because **M** is the only common point of the two line segments.
7. **MA**, **MB**, and **MC** are radii of the circle. This derives from point **6** and from the fact that each of them moves from **M** to the edge of the circle.

Q.E.D (6 and 7)

APPLICATIONS OF THE THEOREM

CONCLUSION A:

No more than one circle may pass through three points.

EXPLANATION:

Through the negation—If two circles passed through the same three points, we would reach the situation in which three congruent line segments or more could pass between two points—the two centers of the circles—something which contradicts the theorem.

CONCLUSION B:

A special case of the theorem—from a point outside the circle, it is not possible to pass more than two equal line segments to the circle’s circumference.

EXPLANATION:

When more than two lines that pass from a point to the edge of the circle are equal, according to the theorem we proved, the point must be the center of the circle.

CONCLUSION C:

If two equal chords bisect each other in a circle, they are diameters.

EXPLANATION:

If equal line segments intersect, we receive four equal line segments leaving a point to the edge of the circle; according to the theorem, the point will be the edge of the circle, and a segment that passes through the center of the circle is a diameter.

2017 MATH QUICKIES

Solutions will be published in the next issue of PI IN THE SKY.

1 Find the number of the nonempty subsets of the set of the first nine positive integers which contain the same number of even and odd integers.

2 Find the tens digit of $2^{2016} + 2^{2017}$.

3 If $a_n = a_{n-1} + 3a_{n-2}$, $n \geq 2$ and $a_1 = a_2 = 1$, find the remainder when a_{2016} is divided by 5.

4 Let x, y, z be any real numbers such that $3x + y + 2z \geq 3$ and $-x + 2y + 4z \geq 5$ Find the minimum value of $7x + 5y + 10z$.

5 Find all real numbers x which are solutions of the equation:

$$\sqrt[3]{x^2 + x} + \sqrt[3]{x^2 + x + 2} = 2$$

6 Find the minimum value of $\frac{c}{a+b} + \frac{a}{b+c}$, where $a, b, c \in [2,3]$ and $a + b + c = 8$.

7 Find the maximum area of a convex quadrilateral having sides of length 2, 5, 10 and 11.

8 Find the maximum number of non-congruent integer-side rectangles, which can be obtained when an 8 x 8 square is cut into pieces with all cuts parallel to its sides.

PRIZE!

PIMS is sponsoring a prize of \$100 to the first high school student (from within the PIMS operating region: Alberta; British Columbia; Manitoba; Saskatchewan; Oregon; Washington) who submits the largest number of correct answers before December 1, 2017. Submit your answers to: pims@uvic.ca

2015 MATH QUICKIES

1. Find all positive integers n for which $\sqrt{5n+2}$ is an integer.

SOLUTION:

If a positive integer k is congruent to $0, \pm 1, \pm 2 \pmod{5}$ then $k^2 \equiv 0$ or $k^2 \equiv \pm 1 \pmod{5}$ and consequently there is no positive integer such that $5n+2 = k^2$.

2. Find all the integers m such that $m^2 + 3m - 50$ is divisible by 19^2 .

SOLUTION:

We first have

$$m^2 + 3m - 50 = (m + 11)(m - 8) + 38$$

If m is an integer such that $m^2 + 3m - 50$ is divisible by 19^2 then we must have

$$19|(m + 11)(m - 8) \Leftrightarrow 19|(m + 11) \text{ or } 19|(m - 8)$$

and since $(m + 11) - (m - 8) = 19$ we conclude that $19|(m + 11)$ and $19|(m - 8)$ and hence $19^2|(m + 11)(m - 8)$. Therefore $19^2|38$, which is a contradiction. Consequently there is no integer m with the requested property.

3. Find the number of positive integers ≤ 1000 which are not divisible by any of 5, 7, and 11.

SOLUTION:

The number of positive integers which are ≤ 1000 and divisible by 5, 7, 11, 35, 55, 77, 385 = $5 \times 7 \times 11$ is respectively 200, 142, 90, 28, 18, 12, 2. The requested number is $1000 - (200+142+90)+(28+18+12)-2 = 624$.

4. Let x, y be any real numbers. Find the smallest possible value of $|x - 3| + |x - y + 2| + |100 - y|$.

SOLUTION:

For any real numbers a, b we have the triangular inequality $|a| + |b| \geq |a + b|$. Hence

$$|x - 3| + |x - y + 2| + |100 - y| \geq |x - 3 + y - x - 2 + 100 - y| = 95$$

for any real numbers x, y . The equality is attained if $5 \leq x + 2 \leq y \leq 100$.

5. Let $P(x)$ be a polynomial of degree four and let $a \geq 1, b \geq 1$ be distinct numbers such that $P(a) = P(1 - a), P(b) = P(1 - b)$. Show that $P(x) = P(1 - x)$, for any real number x .

SOLUTION:

The numbers $a, b, 1 - a, 1 - b$ are distinct and they are roots of the polynomial $Q(x) = P(x) - P(1 - x)$. Since Q is a degree three polynomial with four distinct roots, we conclude that it should be the zero polynomial.

6. Let a_1, a_2, \dots, a_n be real nonnegative numbers such that $a_1 + a_2 + \dots + a_n = k$. Find the maximum value of $a_1a_2 + a_2a_3 + \dots + a_{n-1}a_n$.

SOLUTION:

We have

$$\begin{aligned} a_1a_2 + a_2a_3 + \dots + a_{n-1}a_n &\leq (a_1 + a_3 + \dots)(a_2 + a_4 + \dots) \\ &\leq \frac{[(a_1 + a_3 + \dots) + (a_2 + a_4 + \dots)]^2}{4} \\ &= \frac{(a_1 + a_3 + \dots + a_n)^2}{4} = \frac{k^2}{4} \end{aligned}$$

The maximum value is attained if for example, $a_1 = a_2 = \frac{k}{4}$

7. Let $M = \{1, 2, 3, \dots, 2016\}$ and k a positive integer. Find the minimum value of k for which any subset of M with k elements contains at least two distinct numbers such that one of them is a multiple of the other.

SOLUTION:

If $k \leq 1008$ then the set $\{1009, 1010, \dots, 2016\}$ contains 1008 elements and does not contain at least two distinct numbers one of them a multiple of the other. Let us show that any subset of M with 1009 elements contains at least two distinct numbers such that one of them is a multiple of the other. If $A = \{a_1, \dots, a_{1009}\}$ is any subset of M with 1009 elements then for any $a_i \in A$ there is a nonnegative integer b_i such that

$$1009 \leq a_i 2^{b_i} \leq 2016$$

Since there are 1008 integers between 1009 and 2016, by the pigeonhole principle we must have two distinct numbers $a_i < a_j$ in A such that

$$a_i 2^{b_i} = a_j 2^{b_j}$$

Hence $b_i > b_j$ and $a_j = a_i 2^{b_i - b_j}$, that is a_j is a multiple of a_i . Therefore, the requested minimum value of k is 1009.

8. The point P is inside a convex quadrilateral $ABCD$ of area 168 such that $PA = 9$, $PB = PD = 1$ and $PC = 5$. Find the perimeter of the quadrilateral.

SOLUTION:

Let $\alpha, \beta, \gamma, \delta$ denote the angles $\angle APB, \angle BPC, \angle CPD$ and respectively, $\angle DPA$. The area of the quadrilateral can be written as

$$2 \cdot 128 = 9 \cdot 12 \sin \alpha + 12 \cdot 5 \sin \beta + 5 \cdot 12 \sin \gamma + 12 \cdot 9 \sin \delta$$

Since $9 \cdot 12 + 12 \cdot 5 + 5 \cdot 12 + 12 \cdot 9 = 2 \cdot 168 = 336$ the above equality holds if and only if $\alpha = \beta = \gamma = \delta = 90^\circ$. Therefore, by Pythagorean theorem we obtain $AB = 5$, $BC = 13$, $CD = 13$ and $DA = 15$. The perimeter of the quadrilateral is 46.



Pacific Institute *for the*
Mathematical Sciences



MIX
Paper from
responsible sources
FSC® C014956