The Maths of Sudoku and Latin squares

Professor Sarah Hart

21st November 2023

In today's lecture, we'll explore the mathematics relating to a certain kind of puzzle that millions of us tackle every day – the Sudoku. I'll show you its links to other kinds of number grids, like magic squares and so-called Latin squares, which have been studied for centuries, and we'll see some applications in areas as diverse as experiment design, coding theory, and even literature.

Solving Sudoku

Sudoku puzzles have been popular around the world for a long time. Here in the UK, they are a relatively recent phenomenon. The craze started in 2004, when The Times newspaper published its first Sudoku on November 12th. Within months more or less every British newspaper was printing daily puzzles. I'll briefly recap the rules for Sudoku, and an idea of the process for solving them. A Sudoku is a nine-by-nine square grid into each of whose cells is to be placed a number between 1 and 9. In a completed grid, every row, column, and each of the nine 3×3 square blocks, contains each number between 1 and 9 exactly once. In a Sudoku puzzle, some of the cells are filled in (by convention no more than thirty, and the filled cells have a 180 degree rotational symmetry about the centre cell) and you have to complete the grid.

		5	2		4	8		
4	1		9	8	6			5
		8	5			6	4	
	4		7	9	8	5	3	
8	3	6	a	4	5	7	9	2
	5		b	c	2		8	
	7	9	4		1	3		
5			6		9			
		4	8			9		

Stage 1

		5	2		4	8		
4	1		9	8	6			5
		8	5			6	4	
	4		7	9	8	5	3	d
8	3	6	1	4	5	7	9	2
	5		3	6	2		8	
	7	9	4		1	3		
5			6		9			
		4	8			9		

Stage 2

		5	2		4	8		3/9
4	1	e	9	8	6			5
		8	5			6	4	3/9
	4		7	9	8	5	3	
8	3	6	a	4	5	7	9	2
	5		b	c	2		8	
	7	9	4		1	3		
5			6		9			
		4	8			9		

Stage 3

For example, suppose at some point in your solution you have the grid in the configuration shown on the left – Stage 1. The row containing a must contain a 1, and all the other numbers are present. So $a = 1$. The column containing b now has every number except 3, so $b = 3$. Finally, the central block containing a, b, and c has every number except 6, so $c = 6$. There's a useful technique which goes by different names in different books, but I like to imagine a cake with three layers. Each layer or slice of three rows comprises also three blocks. In Stage 2, look at the middle slice, which is shaded. In the completed slice each number must appear exactly three times: once in each of the three rows and once in each of the three blocks. So, for instance, the number six appears twice so far in that block. Its third appearance cannot be in the middle or bottom row of the block because there are sixes already in those rows. And it cannot be in the left-hand or central block, for the same reason. So the final six must be in the top row of the right-hand block. There's only one vacant cell in that row and block, forcing $d = 6$. We can also look at vertical slices. In the right-hand vertical slice (shaded in Stage 3), we see that both 3 and 9 must be in the right-hand column of the uppermost block. But this time there are two possibilities. We conclude that those two cells contain 3 and 9, but we don't know which is which. This "pairing" is sometimes already enough, though, to allow us to deduce something about other cells. The shaded cell marked e in the second row, for example. In that row, currently we are missing a 3. We know that 3 is in either the top-right or bottom-right cell of the top-right block, so it cannot go anywhere else in the top-right block. So there's only one option: $e = 3$. There are other tricks and tips, and many books and websites dedicated to solving Sudoku, but of course my recommendation is *How to Solve Sudoku,* by our former Gresham Professor of Geometry, Robin Wilson.

There are lots of questions we can ask about Sudoku beyond solving them. How are they created? We must, for example, know there is not only a solution but a unique solution. How many grids are possible? What's the smallest number of initial cells that can be filled to enable us to solve the puzzle? But first, a little history. With the name Sudoku, it's tempting to imagine a centuries-old Japanese tradition, but in fact the origins of the puzzle seem to be in America in the late 1970s, when a magazine published some "number place" puzzles by Harold Garns. A Japanese puzzle company called Nikoli then started including puzzles like this in their magazine in 1984, where it became known as "Sudoku" which approximately means "number-single". Within a couple of years the set up was standardized and the puzzle took off. So, in historical terms, this is a very new puzzle. However, there are very ancient number grids and we're going to look at these next. They have the advantage of usually being considerably smaller than Sudoku, so they'll help us gear up for some of the questions around Sudoku that we've not get answered.

Magic Squares

The oldest square of numbers with special properties of some kind, at least the oldest I've been able to find, is the Lo Shu square dating back at least 2500 years. In Chinese legend, there was a great flood; when the people offered sacrifices to appease the river gods, a magic turtle emerged from the river with the square carved into its shell. This is an example of a *magic square*: the sum of the rows, columns and diagonals is always the same – the *magic number* –in this case it's 15. Over the next millennium magic squares appeared also in India and the Middle East, and subsequently spread to Africa and Europe. The Lo Shu square is "order 3" meaning it's a 3×3 square, and it's what's called a *classical* magic square, meaning it uses the numbers 1 to 9 (classical for order 4 means using the numbers 1 to 16, and for order n we'd use the numbers 1 to n^2). The oldest order 4 magic square (at least, the oldest that we can date accurately) appears in a 6th century CE book by the Indian writer Varahamihira. The square was given as an aid to making perfume. The sum of the rows, columns and diagonals is 18, and you are supposed to select four ingredients, and then follow any one of the rows, columns or diagonals to determine how much of each ingredient to use – the outcome in all cases being the same total quantity of perfume.

Over time, mathematicians started to find methods to generate and classify magic squares of different sizes or types. A Sudoku grid is an example of what's called a *semi-magic square*: its row and column sums are all the same (just $1+2+3+4+5+6+7+8+9=45$) but the diagonals don't necessarily have this sum. Perhaps the most famous magic square in Western art is shown in Dürer's engraving *Melencholia I*. There's a lot of symbolism in this image. Melancholy was traditionally associated with the planet Saturn, as too is geometry, and there's plenty of geometry in this picture! Dürer's mother died in 1514, the year the engraving was made, hence the melancholy. But why the square? It is a classical magic square, and the year the woodcut was made, 1514, is shown in the centre of the bottom row. In ancient tradition the nearest heavenly bodies were associated to a different order magic square. The planet associated to order 4 magic squares is Jupiter. And Jupiter, was a literally "jovial" counter to melancholy. This particular classical magic square was given for the reader to complete as an entertaining diversion by the mathematician Luca Pacioli in a book dating to about 1501 – he gives the first ten entries and leaves it to us to complete the rest. This may well have been where Dürer found it.

Let's think about classical magic squares a bit more. A classical magic square of order 4, like Dürer's, contains the numbers 1 to 16 in some order. So the sum of all the numbers in the square is $1+2+\cdots+15+16=136$. All four rows sum to the same magic number M so $4M=136$, and thus $M=34$. This is true of all order 4 classical magic squares. The same idea works for other orders. The magic number of an order 3 classical magic square is 15, for instance. In Dürer's square, all the rows, columns and diagonals sum to 34. But actually there's more. All the corners sum to 34, as do the 2×2 blocks at each corner, as well as the central 2×2 block. So it's a *very* magic square indeed.

The magic number of an order 2 classical magic square would be 5. But actually no such square can exist.

The row containing the number 1 must also contain the number 4, because the row sum is 5. But that's true of the column containing 1 as well, which would force 4 to appear twice. So, there are no order 2 classical magic squares. However, it can be proved that there are classical magic squares of all other orders. Several different methods exist but I'll show you one that works for odd orders that dates back at least 300 years, probably more. This method is sometimes known as the Siamese method, because it was popularised in the West by a French diplomat who had been taught it when he was ambassador to Siam (the old name for Thailand), and who believed that the method had originated in India. But it predates him by quite some time. Muhammad ibn Muhammad al Fulani al Kishnawi was a mathematician from the Fulani, or Fula, tribe from Katsina, in the north of what is now Nigeria, who spent much of his career in Egypt and died in Cairo in 1741. In a treatise on magic squares in 1732 he (among other things) listed all the order 3 classical magic squares, gave the method I'm about to explain, and illustrated it for magic squares of odd order up to 11. The method gives a classical magic square of order n for odd n. Place the numbers 1, 2, 3 in order as follows. Put 1 in the centre of the top row. At each stage move diagonally one square upwards and to the right. If you go past the top edge of the square, simply loop back to the bottom, like in one of those old-style arcade games where if you disappear off the top off the screen you reappear at the bottom; similarly, if you go past the right-hand edge, loop back to the left. If the position you should be moving to already has a number in it, then instead move down one square.

One of the great things about magic squares is that you once you have one, you can make more. Given a magic square, there are transformations we can apply to it that preserve its magic. For example, if we reflect it about a diagonal, then that interchanges rows with columns, and maps each of the diagonals onto itself. Therefore the row, column and diagonal sums of the new square will still all equal the magic number. Hence, we have created a new magic square. This transformation does not change the set of numbers that appear, and thus will also map classical magic squares to classical magic squares. Any magic-preserving transformation must obviously map the square to itself; in other words preserve the squareness of the shape. There are eight symmetries of a square, and each of these maps rows and columns to rows and columns, and diagonals to diagonals. This means that if we find one classical magic square of order n, we can find all its transformations and turn it into eight classical magic squares of order n.

Let's try and find all classical magic squares of order 3.

1		

Where can 1 go? Any row, column or diagonal containing 1 must have entries summing to 15. The only such sums are $1+5+9$ and $1+6+8$. If 1 is at a corner, it's involved in three sums; if at the centre then four sums. But only two are available. So 1 must be on an off-diagonal. We can rotate later to get back all the possibilities, but for the moment we can assume 1 is in the centre-left position.

1	5	9

The central row and central column must contain the pair 5,9 and the pair 6,8, in some order. The exact same reasoning as for 1 says 9 also is not on a diagonal. So we must have the central row being 1, 5, 9, and the left column being 6, 1, 8 or 8, 1, 6.

6		
1	5	9
8		

Reflecting in the horizontal if necessary we can assume 6,1,8. Then it's easy to complete the square. We can retrieve all possible order 3 squares by reflecting and rotating this seed square. So there are exactly eight order 3 classical magic squares.

How many classical magic squares of order n are there? The eight of order 3 are really just the same one reflected and rotated. So we tend to say that there's just one order 3 magic square "up to reflections and rotations". The calculations get increasingly more involved, but it's been calculated by a mixture of brains and brute force that, again viewing reflections and rotations of a square as being "the same", there are 880 classical magic squares of order 4 and 275,305,224 of order 5. The number of classical magic squares of order 6 is still unconfirmed, though the latest count, based on a very complicated brute-force search using multiple processors over several months, and as at the last update in September 2023, suggests there are 17,753,889,189,701,385,264, or 18 quintillion (to the nearest quintillion).

The other way we can make new magic squares (not classical ones though) is by adding. You can add squares of the same size by just adding corresponding entries. If square S has magic number M and T has magic number N, the square $S+T$ has magic number $M+N$. We've seen order 4 squares with magic numbers 18 (the perfume recipe) and 34 (Dürer's square). Filling the grid with 1's gives magic number 4. If we can place exactly one 1 in each row, column and diagonal, and zeroes elsewhere, we get a square with magic number 1. With these we can now create any magic number, often in several ways. For instance, for

23, we use $18 + 4 + 1 = 23$ to get the square with magic number 23 on the right below.

$$\begin{array}{|c|c|c|c|}\hline 2&3&5&8\\\hline 5&8&2&3\\\hline 4&1&7&6\\\hline 7&6&4&1\\\hline\end{array} + \begin{array}{|c|c|c|c|}\hline 1&1&1&1\\\hline 1&1&1&1\\\hline 1&1&1&1\\\hline 1&1&1&1\\\hline\end{array} + \begin{array}{|c|c|c|c|}\hline 0&1&0&0\\\hline 0&0&1&0\\\hline 1&0&0&0\\\hline 0&0&0&1\\\hline\end{array} = \begin{array}{|c|c|c|c|}\hline 3&5&6&9\\\hline 6&9&4&4\\\hline 6&2&8&7\\\hline 8&7&5&3\\\hline\end{array}$$

More Magic

How can we generalise the idea of a magic square? There are three main ways: go up in dimension, change the shape, or think about other kinds of "magic". In the first case, I'm talking about magic cubes, magic hypercubes, and so on. At least as far back as Fermat people were thinking about magic cubes – there's a 1640 letter from Fermat to Mersenne about it. We can consider other polygons, like magic hexagons, magic rectangles and so on. A square can be subdivided into smaller squares. The only other regular polygons for which a similar thing is true are regular hexagons and equilateral triangles, so our only real hope for magic regular polygons is those. However, magic triangles are not promising. The problem is that each of the corner triangles consists of a row, but is also contained in rows with more than one entry, as shown in the diagram. This forces all the corner entries, and thus all the row sums, to be zero. Already, then, we have shown that there are no classical magic triangles. Our last hope is hexagons. There are classical magic hexagons, there's a boring one-layer one. There's a three-layer one, which you can reflect and rotate just like with magic squares – this time there are twelve symmetries. But apart from that, there are no others of any size. There's a good Numberphile video explaining why, if you are interested. Apart from classical magic squares you can't get classical magic rectangles because if there are r rows and c columns, then r times the magic number must equal c times the magic number (both are equal to the sum of all the integers from 1 to rc). But if $rM = cM$ and r is different from c, then $M = 0$. And there's no way that is going to work! Finally we can consider other kinds of "magic", such as different constraints. For example, we could consider a kind of opposite of a magic square, where every row, column and diagonal must add to a different sum – these are called heterosquares, and antimagic squares in the special case that the row, column, diagonal totals form a sequence of consecutive integers. Or, there are "perimeter magic polygons": regular polygon with the consecutive positive integers from 1 to n placed along the perimeter in such a way that the sums of the integers on each side are constant. This at least works for other polygons apart from squares. But I want to stick with squares and think about the special case of order n semi-magic squares where each row and column contains each of the numbers 1 to n exactly once. Like Sudoku but fewer restrictions. These are called Latin squares.

Latin Squares

A Latin square is an $n \times n$ grid each of whose entries is an element from some set of size n (usually, for convenience, the numbers 1 to n), such that every element appears exactly once in each row and column. Sudoku grids are an example, but they are a special subset because they have the additional requirement of blocks. Latin squares are incredibly useful if you want to look at ways of trying different combinations. For example, imagine you want to try testing different crops. You need a reasonably large area of land to do this, and the conditions of the land are hard to make absolutely identical as soil conditions, moisture levels, aspect, can all vary to some extent. So, what you might want is, for instance, an order 5 Latin square on a hillside in Wales to test the effects of exposure on five different species of tree, as was done in 1929 at Beddgelert forest. There is one region for each tree in each row and column, so that this minimizes any small localized variations in conditions. There is at least one Latin square of every size, because you can make a kind of cyclic pattern. However this is not ideal for experiments because you want to lay out the crops (or whatever you are looking at) in as random a way as possible. To work out how many Latin squares there are, you have to first decide what counts as different. If we are arranging different categories, it doesn't really matter which one we label 1, which 2 and so on. Because of this, if we look at order 3 Latin squares as an example, we may as well assume the top row is 1, 2, 3, as this doesn't affect the underlying structure and we can reconstruct all possibilities later. With top row 1, 2, 3, the options for the middle-left entry are 3 or 2. The outcome is two possible squares:

1	2	3
2	3	1
3	1	2

or

1	2	3
3	1	2
2	3	1

Given that the numbers 1, 2, 3 can be assigned in six possible ways (123, 132, 213, 231, 312, 321), the total number of order 3 Latin squares is 12. The numbers grow fast, though not as insanely as classical magic squares: for order 4 there are 576, for order 5 it's 161,280, and there are 812,851,200 of order 6. By the time we get to order 9 there are 5,524,751,496,156,892,842,531,225,600, or about 5.525×10^{27}. (Not all of these give valid Sudoku grids, and we'll come back to that later.)

Why are these objects called Latin squares? The answer to that takes us back to puzzles. There was a card puzzle doing the rounds in France a few centuries back that has a bit of a Latin square ring to it. The challenge was to take the four highest cards (Jack, Queen, King, Ace) of each suit (hearts, diamonds, clubs, spades) in a deck of playing cards, and to arrange them in a four by four square array, such that each row and column contains exactly one Jack, one Queen, one King, and one Ace, as well as one heart, one diamond, one club and one spade. We can see that what's happening here is that the card values form a Latin square, but so too do the suits. Moreover, it's not just any pair of Latin squares. Each combination occurs exactly once (after all, unless you are a cheat, your pack of cards should not contain more than one Ace of spades, for instance). This puzzle has many solutions (1152, in fact, though it wasn't until the 20th century that they were all correctly classified, by British mathematician Kathleen Ollerenshaw). However, a similar but more difficult challenge started doing the rounds in the court of Catherine the Great of Russia, in the 18th century. This was the "thirty-six officers" problem. You had six ranks and six regiments, with one officer of each rank in each regiment, giving 36 in total. The problem was to arrange them in a six-by-six square such that, again, each row and each column has exactly one officer of each rank, and one officer from each regiment, with no combination of rank and regiment occurring twice – you aren't allowed two Colonels from Regiment 1, and so on. The interesting thing was that nobody could actually solve this puzzle. But fortunately, the great mathematician Leonhard Euler was in Petersburg at the time, and the story goes that Catherine asked him to have a look. And guess what? Euler couldn't do it either!

Euler would have seen straight away that any solution to the 36 officers problem would need to have two Latin squares superimposed in such a way that each possible combination of pairs from the two sets occurs exactly once. He called such structures Graeco-Latin squares because he labelled the elements of one set with Greek letters, and the other set with Latin letters. (The name Latin square seems in fact to be a back-formation from this.) Nowadays we tend to call such pairs "orthogonal Latin squares". Euler tried to determine for which values of n these orthogonal Latin squares exist. He proved that there are no examples when $n = 2$, and that there are examples for odd n, and when n is a multiple of 4. He believed that $n = 6$ couldn't be done – that is, that the 36 officers problem is impossible, and he conjectured in 1782 that there are also no orthogonal Latin squares of order 10, 14, 18 and so on; numbers of the form $4k + 2$. Eventually, in 1906, Gaston Tarry proved, in a method aptly known as an exhaustion proof, that Euler was right about $n = 6$, so it was looking good for the conjecture. But amazingly in 1959 E. T. Parker, R. C. Bose, and S. S. Shrikhande showed that orthogonal Latin squares actually exist for all orders $n > 2$ except $n = 6$. This result was covered by Martin Gardner in the November 1959 edition of Scientific American, whose front cover features a 10×10 counterexample to Euler's conjecture. Thus, this mathematics was pretty recent when the French author George Perec used it as a crucial part of the structure of his book *La Vie Mode d'Emploi,* published in 1978 (translated by David Bellos as *Life, a User's Manual*). The narrative takes place in a building of ten floors with ten rooms on each floor. The stories of the 100 rooms of the building each feature unique combinations of different characteristics taken from lists of ten – for instance there was a list of ten fabrics. One of the major themes of the book is failure – there's a character who has made it his life's work, for example, to paint pictures of places around the world, have them made into jigsaw puzzles and then solve the puzzles, but he dies before he can complete the last one. So what better mathematical structure to base it on than one of the only times that Euler failed to be right. To make it even better, the choice of room for each chapter is given by following a knight's tour of a 10×10 chessboard. But the book misses out a room – it only has 99 chapters! Perec's cryptic explanation is delightful: "For this the little girl on pages 295 and 394 is solely responsible."

This literary use of Latin squares is of course my favourite application, but there are a couple of others I'll mention (and we already saw agricultural uses). Firstly, if you can find a pair of orthogonal Latin squares, perhaps you can find a whole collection of mutually orthogonal ones, where every pair is orthogonal. It can be shown that for order n squares, the theoretical maximum size of a collection of mutually orthogonal Latin squares is $n - 1$ (this is definitely not always attained, but it's been proved that it is attained when n is prime or a power of a prime). When you have more variables at play in experiments, it's often useful to be able to use multiple mutually orthogonal Latin squares like this. An early use was in the 1930s when trying to work out why a particular spindle in a cotton mill was defective. There were 5 spindles, each made of 4 components, and with the use of three mutually orthogonal Latin squares, one could test different combinations simultaneously in a very efficient way.

A much more modern use of mutually orthogonal Latin squares relates to error-detecting and errorcorrecting codes. These are ways of transmitting information where you want to be able to detect and maybe even correct errors. Suppose you want to send a Yes/No command to your remote space probe or whatever. You could say 1 means go and 0 means stop. But signals can get corrupted in their travel across millions of miles. So it's safer to duplicate: send 11 or 00. Now if the receiver receives 10 or 01 it knows there's been an error. But it wastes time having to send back "huh?" all that distance. So we'd like to be able in fact to correct errors. If we send each message three times: 111 or 000, then if there's an error and we receive say 110, we can spot there's an error and (assuming there's been at most one error), we can deduce that the original message was 111. Experiments with transmissions can give us a good idea of what it's safe to assume. For longer messages, sending the whole thing three times is very wasteful. So what we'd like is to have a code, with the maximum number of codewords, that can still detect and correct an error. The way we correct errors is to find the "closest" true codeword to the message we receive, where "closest" here means the codeword that differs in the fewest places from what we receive (the number of places two words differ is called their Hamming distance after the mathematician Richard Hamming). To make sure there's a unique closest word, we need each pair of codewords to differ in at least three places. That way, with one error, there's no way we can differ from more than one codeword in only one place. As an example, let's say you have an alphabet of q letters, and words with four letters. What's the most codewords you can have, and still ensure the "distance 3" requirement? Suppose you have two codewords $abcd$ and $efgh$. If it happened that $ab = ef$, then these words would differ in at most two places. So that can't happen. This means no two codewords can have the same two first letters. The number of possible different first two letters is q^2. So that's the maximum number of codewords in this scenario. What's totally brilliant is that it can be shown that there is a code like this (an alphabet of q letters, words of length 4, with q^2 codewords) if and only if there is a pair of orthogonal Latin squares of order q. As a toy example, remember those two different Latin squares of order 3? With them, we can construct a code with nine codewords, each four digits long, which will allow us to spot up to two errors and correct up to one. All you do is take the two squares, and the codewords are the row and column of each cell followed by the entries in the squares. Notice here that any two digits determine the whole codeword. If you know the row and column, you know the two entries in that cell. If you know the row and one entry, then because it's a Latin square that tells you the column and then you can find the other entry. Similarly if you know the column and one entry. And if you know just the two entries, then as these squares are mutually orthogonal, that allows you to determine the row and column. Because any two digits determine the codeword uniquely, it means each pair of codewords must have at least three of their digits different, and this is what allows us to correct an error. And actually this can be extended: there is a code over an alphabet of q letters, with q^2 words of length L, where each pair of codewords differs in at least $L - 1$ places, if and only if there are $L - 2$ mutually orthogonal Latin squares of order q. If you want to correct E errors, then you need to set $L = 2E + 2$. The result says there is a code that corrects E errors over an alphabet of q letters with q^2 words of length $2E + 2$ (each differing in at least $2E + 1$ places), if and only if there are $2E$ mutually orthogonal Latin squares of order q. So to correct 2 errors, you'd need 4 mutually orthogonal Latin squares. The smallest q this can be done for is $q = 5$, and the four mutually orthogonal Latin squares of order 5 give you a code over an alphabet of five letters, with 25 words of length 6 each differing in 5 places.

Back to Sudoku

How many Sudoku grids are there? The number of Latin squares of order 9 is about 5.525×10^{27}. But not all of these give valid Sudoku grids. How can we begin to count? The numbers for order 9 get very big so to give you an idea of what's going on, we'll compare with the order 4 Sudoku. Here, each of the numbers 1 to 4 must appear exactly once in each row, column, and 2×2 block. For Sudoku we have a block rule, a row rule, and a column rule. Let's call a block grid a grid that satisfies the block rule.

The number of possible ways to fill a block (in isolation) is $4 \times 3 \times 2 \times 1 = 4! = 24$ (the exclamation mark is pronounced "factorial" and $n!$ is just the product of all the whole numbers from 1 to n).

There are four blocks so that gives $4!^4 = 331{,}776$ block grids. For the standard order 9 case, it's $9!^9$. But of course most of these won't obey the row and column rules. Let's work out what proportion of the block grids also have exactly one of each number in each row. Once we've chosen the left hand two blocks (in $4!^2 = 576$ ways) then there are two ways to complete each of the rows, so $2 \times 2 \times 2 \times 2 = 16$ ways to complete the whole grid, meaning the number of block-grids that satisfy the row rule is $576 \times 16 = 9216$. This means the proportion of the block grids that satisfy the row rule is $\frac{16 \times 4!^2}{4!^4} = \frac{16}{576} = \frac{1}{36}$. By the same reasoning, $\frac{1}{36}$th of the block-grids satisfy the column rule. The somewhat more complicated calculation for the order 9 case gives a fraction of $\frac{1}{n}$ where n is a very large number (exactly 128,024,064,000,000 in fact). Now there's a nice estimate by Kevin Kilfoil, which is really close to the true answer. We make the false-but-not-far-out assumption that satisfying the row rule is *independent* from satisfying the column rule. If two events are independent that means we can multiply the probabilities. The probability of tossing a coin and rolling a dice and getting a head and a six is $\frac{1}{2}\times\frac{1}{6}=\frac{1}{12}$, because they are independent. In this case if we have independence it would mean the probability of a block grid satisfying both the row rule and the column rule is $\frac{1}{n}\times\frac{1}{n}$. This would give an estimate for the number of standard Sudoku grids of $\frac{1}{n}\times\frac{1}{n}\times 9!^9$, which is $6{,}657{,}084{,}616{,}885{,}512{,}582{,}463.488 \approx 6.657\times 10^{21}$. Obviously, this can't be exactly correct as it's not a whole number. The true calculation is more involved as the row and column rules are in fact not entirely independent of each other and the calculation of exactly which of the grids that satisfy the block and row rules also satisfy the column rules is longer and requires brute force computer tests. However, the true answer is only 0.2% away from the estimate at $6{,}670{,}903{,}752{,}021{,}072{,}936{,}960 \approx 6.671\ \times\ 10^{21}$.

The other mathematical question is about how to tell a partially completed grid is solvable, and if there's a unique solution. If we have seven entries, for example, then in any valid solution the roles of the remaining two numbers could be switched, so that's obviously not enough. What's the minimum number of entries we need in order to have a chance at a unique solution? This was open for a long time but the answer is 17. It was found by showing that any grid with 16 entries that had at least one valid completion had more than one valid completion, and then finding a grid with 17 entries that had a unique valid solution. However, the number of "minimal grids" is still an open question. These are grids where removal of any number makes the Sudoku have more than one solution. At the other extreme, it's very surprising but the maximum number of entries you have to fill in before guaranteeing a unique solution is 77. The reason is that if you have a configuration with two numbers, say 1 and 2, at opposite corners of a rectangle such that the two 1, 2 pairs also lie in the same blocks, then removing these leaves a Sudoku with two solutions. This is best possible because if only three entries are removed, then at least one row or column will have only one missing number, and thus the grid can be completed.

There are many other variants. Bigger Sudoku, Venn-doku, Cube-doku, magic Sudoku (where they are also magic squares, like the one I showed in the lecture that additionally has the Lo-Shu square as the central block), mutually orthogonal Sudoku (yes, these exist) and so on. The only limit is your imagination!

Next time…

This has been the second lecture in my series *Games, Puzzles, Paradox, and Proof*. In my next lecture we're going to look at mathematical riddles, puzzles and paradoxes, and how the ideas behind them have led to some fascinating and deep mathematics.

References and Further Reading

- Robin Wilson's *How to Solve Sudoku*, published by Brilliant Books, ISBN 9781904902621.
- You can read about antimagic squares at https://mathworld.wolfram.com/AntimagicSquare.html
- There's a very interesting paper about the ancient order 4 magic square we mentioned *Varāhamihira's pandiagonal magic square of the order four,* by Takao Hayashi https://www.sciencedirect.com/science/article/pii/031508608790019X
- See Fermat's magic cube: https://demonstrations.wolfram.com/FermatsMagicCube/
- The Numberphile video on magic hexagons is at https://youtu.be/ZkVSRwFWjy0
- You can read the gory details about the minimum number of entries in a valid Sudoku puzzle in the paper *There is no 16-Clue Sudoku: Solving the Sudoku Minimum Number of Clues Problem via Hitting Set Enumeration, by* Gary McGuire, Bastian Tugemann, Gilles Civario at https://arxiv.org/pdf/1201.0749.pdf
- Latest update on magic squares of order 6 is at https://magicsquare6.net/doku.php?id=magicsquare6
- I made my magic Sudoku with the help of https://www.sudoku-solutions.com/, which checks any partial grid you enter to see if it is valid and has a unique solution. So you can (try to) make your own, subject to whatever crazy constraints you like.
- If you want to get a lot further into more of the mathematics relating to Sudoku, try *Taking Sudoku Seriously: The math behind the world's most popular pencil puzzle*, by Jason Rosenhouse and Laura Taalman (Oxford University Press, 2011).

A note on images used in the lecture

To the best of my knowledge all the images used are either in the public domain, or may be used for educational purposes under fair use rules, or were created by me. The only image I wasn't able to source definitively was the photograph of the Latin square laid out in 1929 at Beddgelert Forest in Wales; I believe it's from a Forestry Commission report of circa 1945.