Sudoku Squares and Chromatic Polynomials

Agnes M. Herzberg and M. Ram Murty

The Sudoku puzzle has become a very popular puzzle that many newspapers carry as a daily feature. The puzzle consists of a 9×9 grid in which some of the entries of the grid have a number from 1 to 9. One is then required to complete the grid in such a way that every row, every column, and every one of the nine 3×3 sub-grids contain the digits from 1 to 9 exactly once. The sub-grids are shown in Figure 1.
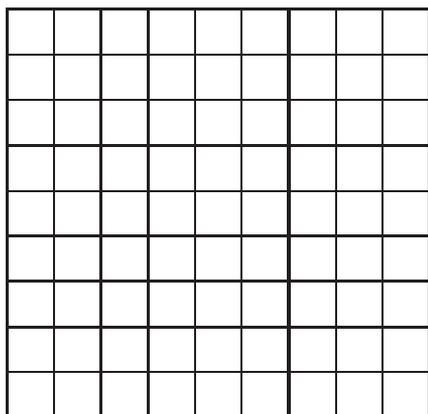
Figure 1. A Sudoku grid.

Agnes M. Herzberg is professor emeritus of mathematics at Queen's University, Canada. Her email address is herzberg@post.queensu.ca.

M. Ram Murty is professor of mathematics at Queen's University, Canada. His email address is murty@mast.queensu.ca.

Recall that a Latin square of rank n is an $n \times n$ array consisting of the numbers such that each row and column has all the numbers from 1 to n. In particular, every Sudoku square is a Latin square of rank 9, but not conversely because of the condition on the nine 3×3 sub-grids. Figure 2 (taken from [6]) shows one such puzzle with seventeen entries given.

Figure 2. A Sudoku puzzle with 17 entries.

For anyone trying to solve a Sudoku puzzle, several questions arise naturally. For a given puzzle, does a solution exist? If the solution exists, is it unique? If the solution is not unique, how many solutions are there? Moreover, is there a systematic way of determining all the solutions? How many puzzles are there with a unique solution? What is the minimum number of entries that can be specified in a single puzzle in order to ensure a unique solution? For instance, Figure 2 shows that the minimum is at most 17. (We leave it to

the reader that the puzzle in Figure 2 has a unique solution.) It is unknown at present if a puzzle with 16 specified entries exists that yields a unique solution. Gordon Royle [6] has collected 36,628 distinct Sudoku puzzles with 17 given entries.

We will reformulate many of these questions in a mathematical context and attempt to answer them. More precisely, we reinterpret the Sudoku puzzle as a vertex coloring problem in graph theory. This enables us to generalize the questions and view them from a broader framework. We will also discuss the relationship between Latin squares and Sudoku squares and show that the set of Sudoku squares is substantially smaller than the set of Latin squares.

Chromatic Polynomials

For the convenience of the reader, we recall the notion of proper coloring of a graph. A λ-coloring of a graph G is a map f from the vertex set of G to $\{1, 2, ..., \lambda\}$. Such a map is called a *proper coloring* if $f(x) \neq f(y)$ whenever x and y are adjacent in G. The minimal number of colors required to properly color the vertices of a graph G is called the *chromatic number* of G and denoted $\chi(G)$. It is then not difficult to see that the Sudoku puzzle is really a coloring problem. Indeed, we associate a graph with the 9×9 Sudoku grid as follows. The graph will have 81 vertices with each vertex corresponding to a cell in the grid. Two distinct vertices will be adjacent if and only if the corresponding cells in the grid are either in the same row, or same column, or the same sub-grid. Each completed Sudoku square then corresponds to a proper coloring of this graph. We put this in a slightly more general context. Consider an $n^2 \times n^2$ grid. To each cell in the grid, we associate a vertex labeled (i, j) with $1 \leq i, j \leq n^2$. We will say that (i, j) and (i', j') are *adjacent* if $i = i'$ or $j = j'$ or $\lceil i/n \rceil = \lceil i'/n \rceil$ *and* $\lceil j/n \rceil = \lceil j'/n \rceil$. (Here, the notation $\lceil \cdot \rceil$ means that we round to the nearest greater integer.) We will denote this graph by X_n and call it the *Sudoku graph* of rank n. A Sudoku square of rank n will be a proper coloring of this graph using n^2 colors. A Sudoku puzzle corresponds to a partial coloring and the question is whether this partial coloring can be completed to a total proper coloring of the graph.

We remark that, sometimes, it is more convenient to label the vertices of a Sudoku graph of rank n using (i, j) with $0 \leq i, j \leq n^2 - 1$. Then, (i, j) and (i', j') are adjacent if $i = i'$ or $j = j'$ or $[i/n] = [i'/n]$ and $[j/n] = [j'/n]$, where now $[\cdot]$ indicates the greatest integer function. That is, $[x]$ means the greatest integer which is less than or equal to x.

Recall that a graph is called *regular* if the degree of every vertex is the same. An easy computation shows that X_n is a regular graph with each vertex

having degree $3n^2 - 2n - 1 = (3n + 1)(n - 1)$. In the case $n = 3$, X_3 is 20-regular and in case $n = 2$, X_2 is 7-regular.

The number of ways of coloring a graph G with λ colors is well known to be a polynomial in λ of degree equal to the number of vertices of G. Our first theorem is that given a partial coloring C of G, the number of ways of completing the coloring to obtain a proper coloring using λ colors is also a polynomial in λ, provided that λ is greater than or equal to the number of colors used in C. More precisely, this is stated as Theorem 1.

Theorem 1. *Let G be a finite graph with v vertices. Let C be a partial proper coloring of t vertices of G using d_0 colors. Let $p_{G,C}(\lambda)$ be the number of ways of completing this coloring using λ colors to obtain a proper coloring of G. Then, $p_{G,C}(\lambda)$ is a monic polynomial (in λ) with integer coefficients of degree $v - t$ for $\lambda \geq d_0$.*

We will give two proofs of this theorem. The most direct proof uses the theory of partially ordered sets and Möbius functions, which we briefly review. A partially ordered set (or poset, for short) is a set P together with a partial ordering denoted by $\leq$ that satisfies the following conditions: (a) $x \leq x$ for all $x \in P$; (b) $x \leq y$ and $y \leq x$ implies $x = y$; (c) $x \leq y$ and $y \leq z$ implies $x \leq z$.

We will consider only finite posets. Familiar examples of posets include the collection of subgroups of a finite group partially ordered by set inclusion and the collection of positive divisors of a fixed natural number n partially ordered by divisibility. A less familiar example is given by the following construction.

Let G be a finite graph and e an edge of G. The graph obtained from G by identifying the two vertices joined by e (and removing any resulting multiple edges) is denoted G/e and is called the contraction of G by e. In general, we say that the graph G' is a contraction of G if G' is obtained from G by a series of contractions. The set of all contractions of a finite graph G can now be partially ordered by defining that $A \leq B$ if A is a contraction of B.

Given a finite poset P with partial ordering $\leq$, we define the *Möbius function* $\mu : P \times P \rightarrow \mathbb{Z}$ recursively by setting

$$\mu(x, x) = 1, \quad \sum_{x \leq y \leq z} \mu(x, y) = 0, \quad \text{if } x \neq z.$$

The main theorem in the theory of Möbius functions is the following. If $f : P \rightarrow \mathbb{C}$ is any complex valued function and we define

$$g(y) = \sum_{x \leq y} f(x),$$

then

$$f(y) = \sum_{x \leq y} \mu(x, y) g(x),$$

and conversely. We refer the reader to [5] for amplification of these ideas.

Proof of Theorem 1. We will use the theory of Möbius functions outlined to prove Theorem 1. Let (G, C) be given with G a graph and C a proper coloring of some of the vertices. We will say (G', C) is a subgraph of (G, C) if G' is obtained by contracting some edges of G with at most one end-point in C. This gives us a partially ordered set. The minimum contraction would be the vertices of C with the adjacencies amongst them preserved. We will also use the letter C to denote this minimum graph in our poset. For each subgraph (G', C) of this poset, let $p_{G',C}(\lambda)$ be the number of ways of properly coloring G' with λ colors with the specified colorings for C. Let $q_{G',C}(\lambda)$ be the number of ways of coloring (not necessarily proper) the vertices of G' using λ colors, with the specified colorings for C. Clearly, $q_{G',C}(\lambda) = \lambda^{v'-t}$, where v' is the number of vertices of G' and t is the number of vertices of C. If $\lambda \geq d_0$, then given any λ-coloring of (G, C), we may derive a proper coloring of a unique subgraph (G', C) simply by contracting the edges whenever two adjacent vertices of (G, C) have the same color assigned. In this way, we obtain the relation

$$q_{G,C}(\lambda) = \lambda^{v-t} = \sum_{C \leq G'} p_{G',C}(\lambda).$$

By Möbius inversion, we deduce that

$$p_{G,C}(\lambda) = \sum_{C \leq G'} \mu(C, G') \lambda^{v'-t},$$

and the right-hand side is a monic polynomial with integer coefficients, of degree $v - t$, as stated. □

We can also prove Theorem 1 without the use of Möbius functions. We apply induction on the number of edges of the graph (G, C). We consider three cases:

(1) Let us suppose that e is an edge connecting two vertices of G at most one of which is contained in C. We will use the following notation. $G - e$ will denote the graph obtained from G by deleting the edge e, but not its end-points. The graph obtained from G by identifying the two vertices joined by e (and removing any multiple edges) will be denoted G/e. With this notation, we have

$$p_{G,C}(\lambda) = p_{G-e,C}(\lambda) - p_{G/e,C}(\lambda)$$

because each proper coloring of G is also a proper coloring of $G - e$ and a proper coloring of $G - e$ gives a proper coloring of G if and only if it gives distinct colors to the end-points x, y of e. Thus, the number of proper colorings $p_{G,C}(\lambda)$ can be obtained from $p_{G-e,C}(\lambda)$ by subtracting those colorings that assign the same color

to both x and y, and this number corresponds to $p_{G/e,C}(\lambda)$. Each of $G - e$ and G/e have fewer edges than G. Thus, we may apply induction and complete the proof in this case.

(2) Now suppose that G has one vertex v_0 not contained in C. If this vertex is not adjacent to any vertex of C, then $G = C \cup v_0$, which is the disjoint union of C and the vertex v_0. Thus, we can color v_0 using any of the λ colors. Thus, $p_{G,C}(\lambda) = \lambda$ in this case. If this vertex is adjacent to d vertices of C, and these vertices use d_0 colors, then, $p_{G,C}(\lambda) = \max(\lambda - d_0, 0)$.

(3) Every vertex of G is contained in C. In this case, we already have a coloring of G and $p_{G,C}(\lambda) = 1$.

Thus, by induction on the number of edges of the graph, the theorem is proved. □

In a later section, we will examine the implications of this theorem for the Sudoku puzzle. For now, we remark that given a Sudoku puzzle, the number of ways of completing the graph is given by $p_{X_3,C}(9)$. A given Sudoku puzzle (X_3, C), has a unique solution if and only if this number $p_{X_3,C}(9) = 1$. It would be extremely interesting to determine under what conditions a partial coloring can be extended to a unique coloring. In this direction, we have the following general result.

Theorem 2. *Let G be a graph with chromatic number $\chi(G)$ and C be a partial coloring of G using only $\chi(G) - 2$ colors. If the partial coloring can be completed to a total proper coloring of G, then there are at least two ways of extending the coloring.*

Proof. Since two colors have not been used in the initial partial coloring, these two colors can be interchanged in the final proper coloring to get another proper extension. □

Theorem 2 implies that if C is a partial coloring of G that can be completed *uniquely* to a total coloring of G, then C must use at least $\chi(G) - 1$ colors. In particular, we have that in any 9×9 Sudoku puzzle, at least eight of the colors must be used in the "given" cells. In general, for the $n^2 \times n^2$ Sudoku puzzle, at least $n^2 - 1$ colors must be used in the "given" partial coloring in order that the puzzle has a unique solution.

Scheduling and Partial Colorings

Given a graph G with a partial coloring, we may ask if this can be extended to a full coloring of the graph. It is well known that coloring problems of graphs encode scheduling problems in real life. The extension from a partial coloring to a total coloring corresponds to a scheduling problem with additional constraints, where, for example, we may want to schedule meetings of

various committees in time slots, with some committees already pinned down to certain time slots. Of course, the corresponding adjacency relation is that two committees are joined by an edge if they have a member in common. This is a question that is of interest in its own right, and it seems difficult to determine criteria for when a partial proper coloring can be extended to a proper coloring of the whole graph.

A similar situation arises for frequency channel assignments. Suppose there are television transmitters in a given region and they need to be assigned a frequency channel for transmission. Two transmitters within 100 miles of each other are to be assigned different channels, for otherwise there will be interference in the signal. It is often the case that some transmitters have already been assigned their frequency channels and new transmitters are to be assigned new channels with these constraints. This is again a problem of completing a partial coloring of a graph to a proper coloring. Indeed, we associate a vertex to each transmitter and join two of them if they are within 100 miles of each other. A channel assignment corresponds to a "color" assigned to that vertex.

We can multiply our examples to many different "real life" contexts. In each case, the problem of completing a partial coloring of a graph to a proper coloring emerges as the archetypal theme.

Latin squares and Sudoku squares are then only special cases of this theme. However, they can also be studied independently of this context. The explicit construction of Latin squares is well known to have applications to the design of statistical experiments. In agricultural studies, for example, one would like to plant v varieties of plants in v rows and v columns so that the pecularities of the soil in which they are planted do not have bearing on the experiment. Agriculturists have always used a $v \times v$ Latin square to design such an experiment. This serves to balance the treatments of the experiment before randomization takes place.

In this context, if one were interested in also testing the role of various fertilizers on the growth of these plants, a Sudoku square might be used, where each sub-grid (or each band) has a different fertilizer applied to it, thus having each fertilizer on each treatment.

Explicit Coloring for X_n

In this section, we will show how one may properly color the Sudoku graph X_n. Recall that the *chromatic number* of a graph is the minimal number of colors needed to properly color its vertices. Thus, the complete graph K_n consisting of n vertices in which every vertex is adjacent to every other vertex has chromatic number n.

Theorem 3. *For every natural number n, there is a proper coloring of the Sudoku graph X_n using n^2 colors. The chromatic number of X_n is n^2.*

Proof. Let us first note that all the cells of the upper left corner $n \times n$ grid are adjacent to each other and this forms a complete graph isomorphic to K_{n^2}. The chromatic number of K_{n^2} is n^2 and thus, X_n would need at least n^2 colors for a proper coloring. Now, we will show that it can be colored using n^2 colors. As remarked earlier, it is convenient to label the vertices (i, j) with $0 \le i, j \le n^2 - 1$. Consider the residue classes mod n^2. For $0 \le i \le n^2 - 1$, we write $i = t_i n + d_i$ with $0 \le d_i \le n - 1$ and $0 \le t_i \le n - 1$ and similarly for $0 \le j \le n^2 - 1$, as well. We assign the "color" $c(i, j) = d_i n + t_i + n t_j + d_j$, reduced modulo n^2, to the (i, j)-th position in the $n^2 \times n^2$ grid. We claim that this is a proper coloring. To see this, we should show that any two adjacent coordinates (i, j) and (i', j') have distinct colors. Indeed, if $i = i'$, then we must show $c(i, j) \ne c(i, j')$ unless $j = j'$. If $c(i, j) = c(i, j')$, then $n t_j + d_j = n t_{j'} + d_{j'}$ which means $j = j'$. Similarly, if $j = j'$, then $c(i, j) \ne c(i', j)$ unless $i = i'$. If now, $[i/n] = [i'/n]$ and $[j/n] = [j'/n]$, then $d_i = d_{i'}$ and $d_j = d_{j'}$. If $c(i, j) = c(i', j')$, then

$$t_i + n t_j = t_{i'} + n t_{j'}.$$

Reducing this mod n gives $t_i = t_{i'}$. Hence, $t_j = t_{j'}$ so that $(i, j) = (i', j')$ in this case also. Therefore, this is a proper coloring. □

Counting Sudoku Solutions

We will address briefly the question of uniqueness of solution for a given Sudoku puzzle. It is not always clear at the outset if a given puzzle has a solution. In this section, we derive some necessary conditions for there to be a unique solution.

Figure 3 gives an example of a Sudoku puzzle which affords precisely two solutions.

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

Figure 3. A Sudoku puzzle with exactly two solutions.

We leave it to the reader to show that the puzzle in Figure 3 leads to the configuration in Figure 4.

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

Figure 4. The "solution" to the puzzle in Figure 3.

Clearly, one can insert either of the arrangements in Figure 5 to complete the grid. Thus, we have two solutions.

9	4
4	9

4	9
9	4

Figure 5. Two ways of completing the puzzle in Figure 4.

This observation leads to the following remark. If in the solution to a Sudoku puzzle, we have a configuration of the type indicated in Figure 6 in the same vertical stack, then at least one of these entries must be included as a "given" in the initial puzzle, for otherwise, we would have two possible solutions to the initial puzzle simply by interchanging a and b in the configuration.

a	b
b	a

Figure 6. A configuration leading to two solutions.

As remarked earlier, if the distinct number of "colors" used in a given Sudoku puzzle is at most seven, then there are at least two solutions to the puzzle. We noted that this was so because we could interchange the two unused colors and still get a valid solution. The multiplicity of solutions can also be seen from the chromatic polynomial. If d_0 is the number of distinct colors used, we have seen that $p_{X_3,C}(\lambda)$ is a polynomial in λ provided $\lambda \geq d_0$. Since the chromatic number of X_3 is 9,

we must have $p_{X_3,C}(\lambda) = 0$ for $\lambda = d_0, d_0 + 1, ..., 8$. As $p_{X_3,C}(\lambda)$ is a monic polynomial with integer coefficients, we can write

$$p_{X_3,C}(\lambda) = (\lambda - d_0)(\lambda - (d_0 + 1)) \cdots (\lambda - 8)q(\lambda),$$

for some polynomial $q(\lambda)$ with integer coefficients. Putting $\lambda = 9$ gives

$$p_{X_3,C}(9) = (9 - d_0)!q(9)$$

and the right hand side is greater than or equal to 2 if $d_0 \leq 7$. This gives us the stated necessary condition for there to be a unique solution, provided that there is a solution, which is a tacit assumption in every given Sudoku puzzle. In the last section, we give a Sudoku puzzle that uses only eight colors and has 17 given entries.

Counting Sudoku Squares of Rank 2

In a recent paper [3], Felgenhauer and Jarvis computed the number of Sudoku squares by a brute force calculation. There are

$$6,670,903,752,021,072,936,960$$

valid Sudoku squares. This is approximately 6.671×10^{21}, a very tiny proportion of the total number of 9×9 Latin squares which is (see [1])

$$5,524,751,496,156,892,842,531,225,600$$
$$\approx 5.525 \times 10^{27}.$$

This mammoth number of Sudoku squares can be cut down to size if we make a few simple observations. First, beginning with any Sudoku square, we can create $9! = 362880$ new Sudoku squares simply by relabelling. More precisely, if a_{ij} represents the (i, j)-th entry of a Sudoku square, and σ is a permutation of the set $\{1, 2, ..., 9\}$, then the square whose (i, j)-th entry is $\sigma(a_{ij})$ is also a valid Sudoku square. There are other symmetries one can take into account. For example, the transpose of a Sudoku square is also a Sudoku square. We can also permute any of the three bands, or the three stacks and the rows within a band, or the columns within a stack. When all these symmetries are taken into account, the number of essentially different Sudoku squares is the more manageable number

$$5,472,730,538,$$

approximately 5.47×10^9, as was shown in [4].

These calculations can be better understood if we consider the case of the number of distinct 4×4 Sudoku squares. Without loss of generality, we may as well label the entries in the first 2×2 block as in Figure 7.

It is not hard to see that there are at most 2^4 ways of completing this grid. However, we can be a bit more precise. One can show that taking into account the obvious symmetries already indicated, there are essentially only two 4×4 Sudoku squares! Indeed, since permuting the last two columns is an allowable symmetry, we may suppose without

Figure 7. A 4×4 Sudoku grid.

loss of generality that the last two entries in the first row are $(3, 4)$ in this order. Similarly, since we may interchange the last two rows, we may suppose that our square is as shown in Figure 8.

Figure 8. A 4×4 Sudoku puzzle.

Then, 1 and 4 are the only possible entries in the diagonal position $(3,3)$ and it is easily seen that the choice of 1 leads to a contradiction. This leads to the following arrangements.

Figure 9. Three 4×4 Sudoku puzzles.

The squares can be completed easily as shown in Figure 10.

However, the last configuration is equivalent to the second one upon taking the transpose and interchanging 2 and 3. Thus, there are really only two nonequivalent solutions for the 4×4 Sudoku square. From this reasoning, we also see that without taking into account the symmetries, we get a total of $4! \times 2 \times 2 \times 3 = 288$ Sudoku squares of rank 2.

In this context, it is interesting to determine the minimum number of cells filled in a 4×4 Sudoku puzzle that leads to a unique solution. Figure 11

gives one with four entries. Is there one with three? We will indicate a proof that there is not.

Figure 10. Three 4×4 Sudoku squares determined from Figure 9.

Figure 11. A 4×4 Sudoku puzzle with 4 cells filled in.

As we remarked earlier, the number of "colors" used in any partial coloring of the Sudoku graph of rank n must be at least $n^2 - 1$ in order that there be a unique solution. Thus, in the puzzle in Figure 11, we must use at least 3 colors. To prove that the minimum number for the rank 2 Sudoku graph is four, we must show if only three distinct "colors" are filled in, we do not have a unique solution. This is easily done by looking at the two inequivalent 4×4 Sudoku grids given in Figure 10 and checking the cases that arise one by one.

Permanents and Systems of Distinct Representatives

In this section, we will review several theorems on permanents and systems of distinct representatives that will be used in the next section in our enumeration of Sudoku squares. For further details, we refer the reader to [5].

If A is an $n \times n$ matrix, with the (i, j)-th entry given by a_{ij}, the *permanent* of A, denoted per A, is by definition

$$\sum_{\sigma \in \mathbb{s}_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

where $\mathbb{s}_n$ denotes the symmetric group on the n symbols $\{1, 2, ..., n\}$. The matrix A is said to be *doubly stochastic* if both the row sums and column sums are equal to 1.

In 1926 B. L. van der Waerden posed the problem of determining the minimal permanent among all $n \times n$ doubly stochastic matrices. It was felt that the minimum is attained by the constant matrix all

of whose entries are equal to $1/n$. Over the years, this feeling changed into the conjecture that

$$(1) \qquad \operatorname{per} A \geq \frac{n!}{n^n},$$

for any doubly stochastic matrix A and was then referred to as the van der Waerden conjecture. By 1981 two different proofs of the conjecture appeared, one by D. I. Falikman and another by G. P. Egoritsjev. As for upper bounds, H. Minc conjectured in 1967 that if A is a $(0,1)$ matrix with row sums r_i, then

$$(2) \qquad \operatorname{per} A \leq \prod_{i=1}^{n} r_i!^{1/r_i}.$$

This conjecture was proved in 1973 by L. M. Brégman (see p. 82 of [5]). We will utilize both the upper and lower bounds for the permanents in our enumeration of Sudoku squares. The application will be via the theorem of Phillip Hall (sometimes called the *marriage theorem*), which we now describe.

Suppose that we have subsets $A_1, A_2, ..., A_n$ of the set $\{1, 2, ..., n\}$. We would like to select distinct elements $a_i \in A_i$. Such a selection is called a *system of distinct representatives*. The theorem gives necessary and sufficient conditions for when this can be done (see [5]). If for every subset S of $\{1, 2, ..., n\}$, we let

$$N(S) = \cup_{j \in S} A_j,$$

then a little reflection shows that it is necessary that $|N(S)| \geq |S|$ for there to exist a selection. (Here, $|S|$ indicates the cardinality of the set S.) Hall's theorem states that this is also sufficient. The number of ways this can be done is given by the permanent of the $n \times n$ matrix A defined as follows. It is a $(0,1)$ matrix whose (i, j)-th entry is 1 if and only if $i \in A_j$. We will refer to A as the *Hall matrix* associated with the sets $A_1, ..., A_n$.

These results can be used to obtain upper and lower bounds for the number of Latin squares of order n (as in [5]). Since we will need the lower bound in the next section, we give the details of this calculation.

For an $n \times n$ Latin square, the number of ways of filling in the first row is clearly $n!$. Suppose we have completed k rows of the Latin square. We now want to fill in the $(k+1)$-st row. For each cell i of the $(k+1)$-st row, we let A_i be the set of numbers not yet used in the i-th column. The size of A_i is therefore $n - k$. To fill in the $(k+1)$-st row of our Latin square is tantamount to finding a set of distinct representatives of the sets $A_1, ..., A_n$. The number of ways this can be done is the permanent of the corresponding Hall matrix A. Since $(n-k)^{-1}A$ is doubly stochastic, equation (1) shows that there are at least

$$\frac{(n-k)^n n!}{n^n}$$

ways of doing this. Taking the product over k ranging from 0 to $n - 1$ gives:

Lemma 4. *The number of Latin squares of order n is at least $n!^{2n}/n^{n^2}$.*

Corollary 5. *The number of Latin squares of order n^2 is at least*

$$n^{2n^4} e^{-2n^4 + O(n^2 \log n)}.$$

Proof. By Lemma 4, the number of Latin squares of order n^2 is at least

$$n^2!^{2n^2}/n^{2n^4}.$$

Using Stirling's formula,

$$\log n! = n \log n - n + \frac{1}{2} \log n + O(1),$$

we obtain the stated lower bound. □

Latin Squares and Sudoku Squares

We can now prove:

Theorem 6. *The number of Sudoku squares of rank n is bounded by*

$$n^{2n^4} e^{-2.5n^4 + O(n^3 \log n)}$$

for n sufficiently large.

Proof. The $n^2 \times n^2$ Sudoku square is composed of n^2 sub-grids of size $n \times n$. The entries in each sub-grid can be viewed as a permutation of $\{1, 2, ..., n^2\}$. Thus, a crude upper bound for the number of Sudoku squares is given by

$$[(n^2)!]^{n^2}.$$

We begin by observing that there are n "bands" in the Sudoku grid. ("Bands" are groups of the n successive rows.) The number of ways of completing the first band can be estimated as follows. We have $n^2!$ choices for the first row. The number of ways of filling the second row is calculated by evaluating the permanent of the following matrix. We have an $n^2 \times n^2$ matrix whose rows parametrize the cells of the second row. Let us note that for each cell, we have $n^2 - n$ possibilities since we already used n colors in the corresponding $n \times n$ sub-grid. The number of ways of filling in the second row is the number of ways of choosing a system of distinct representatives from this list of possibilities. More precisely, we consider the $n^2 \times n^2$ matrix whose rows parametrize the cells of the second row, and whose columns parametrize the numbers from 1 to n^2, and we put a 1 in the (i, j)-th entry if j is a permissible value for cell i and zero otherwise. This gives us a $(0,1)$ matrix whose permanent is the number of ways of choosing the set of distinct representatives (see [5]). By equation (2), this quantity is bounded by

$$(n^2 - n)!^{\frac{n}{n-1}}.$$

Proceeding similarly gives the number of possiblities for the third row as

$$(n^2 - 2n)!^{\frac{n}{n-2}}.$$

In this way, we obtain a final estimate of

$$\prod_{k=0}^{n-1} (n^2 - kn)!^{\frac{n}{n-k}}$$

for the number of ways of filling in the first band consisting of n rows of a Sudoku square of rank n. Suppose now that $(i - 1)$ of the n bands have been completed. We calculate the number of ways of completing the i-th band. Here, we will give an upper bound. The number of possible entries for the first cell of the i-th band is

$$n^2 - (i - 1)n.$$

Thus, as before, the number of ways of completing the first row of the i-th band is bounded by

$$(n^2 - (i - 1)n)!^{\frac{n^2}{n^2-(i-1)n}},$$

since for each cell, we must exclude the entries already entered in the column to which the cell belongs. Similarly, the number of ways of completing the second row of the i-th band is at most

$$(n^2 - ((i - 1)n + 1))!^{\frac{n^2}{n^2-((i-1)n+1)}}.$$

We proceed in this way until the i-th row of the i-th band to get an estimate of

$$(n^2 - ((i - 1)n + i))!^{\frac{n^2}{n^2-((i-1)n+i)}}.$$

For the $(i + 1)$-th row, we change our strategy and exclude the numbers already entered in the sub-grid in which the cell belongs. This means we have entered in entries already in the subgrid and we must exclude these to get an estimate of

$$(n^2 - in)!^{\frac{n^2}{n^2-in}},$$

for the number of ways of filling in the i-th row of the i-th band. Proceeding in this manner, we get that the number S_n of Sudoku squares of rank n is bounded by

$$\prod_{i=1}^{n} (n^2 - (i - 1)n)!^{\frac{n^2}{n^2-(i-1)n}}$$

$$(n^2 - [(i - 1)n + 1])!^{\frac{n^2}{n^2-[(i-1)n+1]}} \cdots$$

$$\cdots (n^2 - [(i - 1)n + i])!^{\frac{n^2}{n^2-[(i-1)n+i]}}$$

$$(n^2 - in)!^{\frac{n^2}{n^2-in}} \cdots (n^2 - (n - 1)n)!^{\frac{n^2}{n^2-(n-1)n}}.$$

Thus,

$$\frac{\log S_n}{n^2} \leq \sum_{i=1}^{n} \left(\sum_{j=0}^{i} \frac{\log(n^2 - [(i - 1)n + j])!}{n^2 - (i - 1)n + j} \right.$$

$$\left. + \sum_{k=i}^{n-1} \frac{\log(n^2 - kn)}{(n^2 - kn)} \right).$$

We estimate the sum using Stirling's formula. It is

$$\leq \sum_{i=1}^{n} \left(\sum_{j=0}^{i} \log(n^2 - [(i - 1)n + j]) + \sum_{k=i}^{n-1} \log(n^2 - kn) \right)$$

$$- n^2 + O(\log^2 n).$$

This is easily seen to be

$$\leq \sum_{i=1}^{n} \left(i \log(n^2 - (i - 1)n) + \sum_{k=i}^{n-1} \log(n^2 - kn) \right)$$

$$- n^2 + O(\log^2 n).$$

Thus, $(\log S_n)/n^2 + n^2 + O(n \log n)$ is

$$\leq \frac{1}{2} n^2 \log n +$$

$$\sum_{i=1}^{n} \left(i \log(n - i + 1) + (n - i) \log n + \sum_{k=i}^{n-1} \log(n - k) \right).$$

The summation over k is

$$\log(n - i)! = (n - i) \log(n - i) - (n - i) + O(\log n),$$

by Stirling's formula. Combining all of this gives

$$\frac{\log S_n}{n^2} \leq 2n^2 \log n - 2.5n^2 + O(n \log n),$$

from which the theorem is easily derived. □

We have the following corollary.

Corollary 7. *Let p_n be the probability that a randomly chosen Latin square of order n^2 is also a Sudoku square. Then*

$$p_n \leq e^{-0.5n^4 + O(n^3 \log n)}.$$

In particular, $p_n \to 0$ as n tends to infinity.

Proof. By Theorem 6, the number of Sudoku squares of rank n is at most

$$n^{2n^4} e^{-2.5n^4 + O(n^3 \log n)}.$$

The number of Latin squares of rank n^2 is by Corollary 5, at least

$$n^{2n^4} e^{-2n^4 + O(n^2 \log n)}.$$

Thus, the probability that a random Latin square of order n^2 is also a Sudoku square is bounded by

$$e^{-0.5n^4 + O(n^3 \log n)},$$

which goes to zero as n tends to infinity. □

In fact, the theorem shows that the number of Sudoku squares is substantially smaller than the number of Latin squares.

Concluding Remarks

It is interesting to note that the Sudoku puzzle is extremely popular for a variety of reasons. First, it is sufficiently difficult to pose a serious mental challenge for anyone attempting to do the puzzle. Secondly, simply by scanning rows and columns, it is easy to enter the "missing colors", and this gives the solver some encouragement to persist. The novice is usually stumped after some time. However, the puzzle can be systematically solved by keeping track of the unused colors in each row, in each column, and in each sub-grid. A simple process of elimination often leads one to complete the puzzle. Some of the puzzles classified under the "fiendish" category involve a slightly more refined version of this elimination process, but the general strategy is the same. One could argue that the Sudoku puzzle develops logical skills necessary for mathematical thought.

What is noteworthy is that this simple puzzle has given rise to several problems of a mathematical nature that have yet to be resolved. We have already mentioned the "minimum Sudoku puzzle" problem, where we ask if there is a Sudoku puzzle with 16 or fewer entries that admits a unique solution.

We have already commented that if only 7 or fewer colors are used, the puzzle does not admit a unique solution. One may ask if there is a puzzle using only 8 colors with a minimum number of entries. In Figure 12, we give such a puzzle that again has only 17 given entries (taken from [6]).

Figure 12. A Sudoku puzzle using only 8 colors.

These questions suggest the more general question of determining the "minimum Sudoku" for the general puzzle of rank n. It would be interesting to determine the asymptotic growth of this minimum as a function of n. Our discussion shows that this minimum is at least $n^2 - 1$. Is it true that the minimum is $o(n^4)$?

Another interesting problem is to determine the number of Sudoku squares of rank n. More precisely, if $p_{X_n}(\lambda)$ is the chromatic polynomial of the Sudoku graph of rank n, what is the asymptotic behavior of $p_{X_n}(n^2)$? If S_n is the number of Sudoku squares of rank n, it seems reasonable to conjecture that

$$\lim_{n \to \infty} \frac{\log S_n}{2n^4 \log n} = 1.$$

It is clear that this limit, if it exists, is less than or equal to 1. We have already noted the various symmetries of the Sudoku square. For example, applying a permutation to the elements $\{1, 2, ..., n^2\}$, gives us a new Sudoku square. Thus, starting from one such square, we may produce $n^2!$ new Sudoku squares. There are also the band permutations, and these are $n!$ in number, as well as the stack permutations, which are also $n!$ in number. We may permute the rows within a band as well as columns within a stack and each of these account for $n!^n$ symmetries. In addition, we can take the transpose of the square. In total, this generates a group of symmetries, which can be viewed as a subgroup of $\mathfrak{S}_{n^4}$. It would be interesting to determine the size and structure of this subgroup. If we agree that two Sudoku squares are *equivalent* if one can be transformed into the other by performing a subset of these symmetries, then an interesting question is the asymptotic behavior of the number S_n^*, of the number of inequivalent Sudoku squares of rank n.

The question of determining the asymptotic behavior of S_n and S_n^* seems to be as difficult as the enumeration of the number of Latin squares of order n. In this context, there are some partial results. A $k \times n$ Latin rectangle is a $k \times n$ matrix with entries from $\{1, 2, ..., n\}$ such that no entry occurs more than once in any row or column. Godsil and McKay [2] have determined an asymptotic formula for the number of $k \times n$ Latin rectangles for $k = o(n^{6/7})$. This suggests that we consider the notion of a *Sudoku rectangle* of rank (k, n) with n^2 columns and kn rows with entries from $\{1, 2, ..., n^2\}$ such that no entry occurs more than once in any row or column or each $n \times n$ sub-grid. It may be possible to extend the methods of [2] to study the asymptotic behavior of the Sudoku rectangles of rank (k, n) for k in certain ranges.

References

[1] S. BAMMEL and J. ROTHSTEIN, The number of 9×9 Latin squares, *Discrete Math.***11** (1975), 93–95.

[2] C. D. GODSIL and B. D. MCKAY, Asymptotic enumeration of Latin rectangles, *J. Comb. Theory Ser. B* **48** (1990), no. 1, 19–44.

[3] B. FELGENHAUER and A. F. JARVIS, Mathematics of Sudoku I, *Mathematical Spectrum* **39** (2006), 15–22.

[4] E. RUSSELL and A. F. JARVIS, Mathematics of Sudoku II, *Mathematical Spectrum* **39** (2006), 54–58.

[5] J. H. VAN LINT and R. M. WILSON, *A Course in Combinatorics,* Cambridge University Press, 1992.

[6] See http://www.csse.uwa.edu.au/~gordon/sudokumin.php.