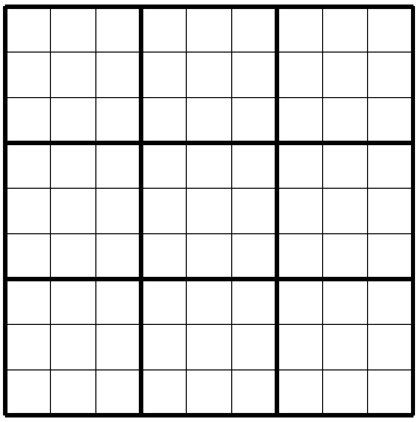
Mathematics of Sudoku I

Bertram Felgenhauer
Frazer Jarvis[*]

February 15, 2006

Introduction

Sudoku puzzles became extremely popular in Britain from late 2004. *Sudoku*, or *Su Doku*, is a Japanese word (or phrase) meaning something like *Number Place*. The idea of the puzzle is extremely simple; the solver is faced with a 9×9 grid, divided into nine 3×3 blocks:



In some of these boxes, the setter puts some of the digits 1–9; the aim of the solver is to complete the grid by filling in a digit in every box in such a way that each row, each column, and each 3×3 box contains each of the digits 1–9 exactly once.

It is a very natural question to ask how many Sudoku grids there can be. That is, in how many ways can we fill in the grid above so that each row, column and box contains the digits 1–9 exactly once.

In this note, we explain how we first computed this number. This was the first computation of this number; we should point out that it has been subsequently confirmed using other (faster!) methods.

First, let us notice that Sudoku grids are simply special cases of Latin squares; remember that a *Latin square of order* n is an $n \times n$ square containing each of the digits $1, \ldots, n$ in every row and column. The calculation of the number of Latin squares is itself a difficult problem, with no general formula known. The number of Latin squares of sizes up to 11×11 have been worked out (see references [1], [2] and [3] for the 9×9, 10×10 and 11×11 squares), and the methods are broadly brute force calculations, much like the approach we sketch for Sudoku grids below. (Brute force calculations are those where part of the calculation involves constructing all possible answers, and seeing which ones really do work.) It is known that the number of 9×9 Latin squares is $5524751496156892842531225600 \approx 5.525 \times 10^{27}$. Since this answer is enormous, we are going to have to be clever about how we do the brute force counting, in order to be able to get an answer in a sensible amount of computing time.

1 Initial observations

Our aim is to compute the number N_0 of valid Sudoku grids. In the discussion below, we will refer to the blocks as B1–B9, where these are labelled

[*]Contact address: Dr A.F.Jarvis, Department of Pure Mathematics, University of Sheffield, Sheffield S3 7RH, U.K., `a.f.jarvis@shef.ac.uk`

	B1			B2			B3	
	B4			B5			B6	
	B7			B8			B9	

Let's first observe that we can simplify the counting we need to do by *relabelling*. We could, for example, exchange all of the 1s and 2s in a valid Sudoku grid, and get another valid grid. We call this *relabelling* the 1s and 2s. If we like, we can relabel all of the digits; in particular, we can relabel any grid so that the top left block (B1) is given by

1	2	3
4	5	6
7	8	9

We will call a grid with this top left-hand block in *standard form.* This relabelling procedure reduces the number of grids by a factor of $9! = 362880$. The problem is reduced to counting the number $N_1 = \frac{N_0}{9!}$ of Sudoku grids in this standard form.

Our strategy is to consider all possible ways to fill in blocks B2, B3, given that B1 is in the standard form above. Then, for each of these possibilities for blocks B1–B3, we work out all possible complete grids with this given collection of blocks.

We will see that it happens that given blocks B2 and B3, there are other possibilities, B2′ and B3′ say, such that the number of ways of completing blocks B1, B2 and B3 to full grids is *the same* as the number of ways of completing blocks B1, B2′ and B3′. This means that we only need to count the number of ways to complete B1, B2 and B3 to a full grid, and we *do not need* to count the number of ways to complete B1, B2′ and B3′, as the answer will be the same. In fact, we will find that we will only need to work out the number of ways of completing a grid for remarkably few pairs B2 and B3.

2 Blocks B2 and B3

Here we try to catalogue efficiently the possibilities for blocks B2 and B3. Let's first work out how many possibilities there are for B2 and B3, and then explain how we can reduce the size of the list we need to count.

2.1 Top row of blocks

We want to list all the possible configurations for the top three blocks, given that the first block is of the standard form. Let's think about the top row of the second block. Either it consists (in some order) of the three numbers on row 2 of block B1, or on row 3 of block B1, or a mixture of the two. The possible top rows of blocks B2 and B3 are given by:

$\{4,5,6\}\|\{7,8,9\}$	$\{7,8,9\}\|\{4,5,6\}$
$\{4,5,7\}\|\{6,8,9\}$	$\{6,8,9\}\|\{4,5,7\}$
$\{4,5,8\}\|\{6,7,9\}$	$\{6,7,9\}\|\{4,5,8\}$
$\{4,5,9\}\|\{6,7,8\}$	$\{6,7,8\}\|\{4,5,9\}$
$\{4,6,7\}\|\{5,8,9\}$	$\{5,8,9\}\|\{4,6,7\}$
$\{4,6,8\}\|\{5,7,9\}$	$\{5,7,9\}\|\{4,6,8\}$
$\{4,6,9\}\|\{5,7,8\}$	$\{5,7,8\}\|\{4,6,9\}$
$\{5,6,7\}\|\{4,8,9\}$	$\{4,8,9\}\|\{5,6,7\}$
$\{5,6,8\}\|\{4,7,9\}$	$\{4,7,9\}\|\{5,6,8\}$
$\{5,6,9\}\|\{4,7,8\}$	$\{4,7,8\}\|\{5,6,9\}$

(where $\{a, b, c\}$ indicates the numbers a, b and c *in any order*).

The top row $\{4, 5, 6\}|\{7, 8, 9\}$ can be completed as follows:

1	2	3	{4,	5,	6}	{7,	8,	9}
4	5	6	{7,	8,	9}	{1,	2,	3}
7	8	9	{1,	2,	3}	{4,	5,	6}

giving $(3!)^6$ possible configurations (each set of three numbers can be written in $3! = 6$ different ways). The same is true for its reversal $\{7, 8, 9\}|\{4, 5, 6\}$. However, the other 18 possibilities behave differently; here, the top row of B2 consists of a mixture of some of the second row and some of the third row of B1. For example, the top row $\{4, 5, 7\}|\{6, 8, 9\}$ can be completed as:

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

where a, b and c stand for 1, 2 and 3, in some order, giving $3 \times (3!)^6$ possible configurations (b and c are interchangeable).

In total, we therefore have

$$2 \times (3!)^6 + 18 \times 3 \times (3!)^6 = 56 \times (3!)^6 = 2612736$$

possible completions to the top three rows.

Note that this means that the number of possibilities for the top three rows of a Sudoku grid is $9! \times 2612736 = 948109639680$.

2.2 Reduction

At this stage, we have a list of all possibilities for blocks B2 and B3. For each of these possibilities, we will try to fill in the remaining blocks to form valid Sudoku grids (the "brute force" part of the method). However, to run through all 2612736 possibilities for B2 and B3 would be extremely time-consuming. We need some way to reduce the number of possibilities which we need to consider. We will identify configurations of numbers in these blocks which give the same number of ways of completing to a full grid.

Luckily, there are a lot of things that we can do to the top three rows which preserve the number of completions to a full grid. We have already seen the relabelling operation. But there are others; for example, if we exchange B2 and B3, then every way of completing B2-B3 to a complete grid gives us a unique way to complete B3-B2 to a complete grid (just exchange B5 and B6, and B8 and B9). Indeed, we can permute B1, B2 and B3 in any way we choose. Although this changes B1, we can then relabel to put B1 back into standard form.

Furthermore, we can permute the columns within any block in any way we wish, performing the same operation to the columns in a completed grid. We can even permute the three rows of B1, B2 and B3, and again relabel to put B1 into standard form.

Lexicographical reduction

Take all of the 2612736 possibilities mentioned above. We catalogue them first as follows:

1. We begin by permuting the columns within B2 and B3 so that the top entries are in increasing order.
2. We then exchange B2 and B3 if necessary, so that the top left entry of B2 is smaller than that of B3.

The first step gives 6 ways to permute the columns in each block B2 and B3, so that, given any grid, there are $6^2 = 36$ grids derived from it with the same number of ways of completing; then the second essentially doubles this number. Overall, we are reducing the number of possibilities we need to consider by a factor of 72, giving 36288 possibilities for our catalogue. This is becoming more practicable, although more reductions are desirable.

Permutation reduction

In fact, we haven't really made full use of all of the permutation and relabelling possibilities. As mentioned above, for each of the 36288 possibilities, we can consider all 6 permutations of the *three* blocks B1-B3, and all 6 permutations of columns within each block, making $6^4 = 1296$ possibilities in total. Having done this, our first block will not be in standard form, but we can relabel so that it is (relabelling B2 and B3 in the same way), and then use the lexicographical reduction on the result. Each of the 1296 permutations gives a new B2-B3 pair which has the same number of completions to a full grid. This provides a huge improvement again: computer calculations showed that this reduced the size of the list we need to test to just 2051 possible B2-B3 pairs. (The huge majority of these 2051 pairs arise from exactly $18 = 6^4/72$ of the 36288 possibilities. Some, however, arise from fewer, so it is necessary to store exactly how many of the 36288 possibilities give rise to the given blocks.)

But this is not all – we can do the same for the 6 permutations of the three *rows* of the configuration. That is, we can choose any permutation of these rows, and then relabel to put B1 back into standard form. It turned out that this gave a further reduction to testing just 416 possibilities for blocks B2 and B3.

Let's illustrate the procedure with an example. The top three rows might look something like the following:

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

Now let's permute the columns of B1, by exchanging the first two columns, say. (It's similarly possible to permute the 3 rows of all three boxes simultaneously, or to change the order of the boxes.) We get

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

Of course, the first block is no longer in standard form. But we can put it back in standard form by relabelling; that is, we relabel 1 as 2 and vice versa, 4 as 5 and vice versa, and 7 as 8 and vice versa. We get:

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

Now the first block is back in standard form, but blocks B2 and B3 aren't lexicographically reduced. We therefore sort the columns of B2, and the columns of B3, and then exchange B2 and B3 if necessary. We obtain:

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

This configuration has the same number of completions as the one we started with, because all operations we applied leave that number invariant. This means we can consider the two as equivalent for the purpose of the enumeration.

Column reduction

While this improvement is extremely useful, the calculation will be even faster if we can cut down our list further. Luckily, there are still more possibilities for improving our list. Here is a possible arrangement for the top three rows in a Sudoku grid:

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

Consider the positions marked for the numbers 8 and 9:

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

It is easy to see that any grid completing these three blocks also completes the following:

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

It follows that this has the same number of completions to a full grid as did the original arrangement. Consequently, the number of ways to extend the original three rows is the same as the number of ways of extending these three rows, and so we should only compute this once. Note that the same can be done for the pairs of numbers (1,2) in columns 4 and 7, (1,4) in columns 1 and 4, (5,8) in columns 2 and 9, and also (6,9) in columns 3 and 6. (It is necessary to relabel to get B1 back into standard form in all but the first of these examples.) We need that there is a 2×2 subrectangle in the grid with the same entries on the bottom row as on the top (but in the other order, of course).

An extension of this method allows us to identify any two configurations with a subrectangle of size $k \times 2$ (respectively, $2 \times k$) whose entries consist of two rows (respectively, columns) with the same numbers. Using this trick with just the 2×2 subrectangles reduced the list of 416 to 174. Using 2×3, 3×2 and 4×2 rectangles as well reduced this list to just 71.

In fact, all that we really need is that the *columns* of two configurations consist of the same numbers, in any order, to guarantee that two configurations can be completed in the same number of ways. The following arrangement:

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

has the same number of completions to a full grid as

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

(This needs to be relabelled to put it back into standard form, and then lexicographically reduced.)

In fact, we already had sufficiently few configurations that we didn't need to implement these further reductions. But, having made an exhaustive search of ways to complete our 71 representatives, we found that there were 44 distinct answers. Subsequent work by Ed Russell showed that these further reductions really do reduce the 71 classes to 44 classes.

Summary

At this stage, we have explained that every one of the original possibilities for blocks B1-B3 is equivalent (in the sense of having the same number of completions to a full grid) to one of a collection $\mathcal{C}$ of just 44. We just need to count the number of ways each of these 44 can be completed to a full grid. If C is one of the 44 possibilities, we will need to know the number of ways n_C that we can complete C to a full grid, and also the number of possibilities m_C for B1-B3 which are equivalent to C. Then N_0, the total number of Sudoku grids, will be the sum of all of the $m_C n_C$ for C in $\mathcal{C}$.

3 The brute force counting

The brute force part of the calculation comes next; for each of the possible 44 B2-B3 configurations, we have to find how many ways we can complete B1–B3 to a full grid. Essentially, we just try all possible ways to fill the remaining boxes and see which are Sudoku grids, but we can again make our lives a little easier by insisting that the first column of blocks B4 and B7 is lexicographically ordered (that is, by permuting the middle three rows we can assume that the left-hand entries of B4 are in numerical order, and similarly for B7; we can also exchange B4 and B7). This speeds up our calculation by a factor of 72 once again.

We could have reduced the search a little further by doing more of the reduction steps given above. However, the predicted running time at this point was sufficiently low that the possible speed-up gained by going through a catalogue of possibilities for all of blocks B4 and B7, reduced by some of the methods listed above, was hardly worth implementing.

Recall that at this stage, the top three blocks B1–B3, and also the first column of blocks B4 and B7 are filled. Now we just try each of the possible ways to complete this information to a full Sudoku grid, and count how many of them work. This was programmed very efficiently by the first author using a *backtracking algorithm*. This proved to be a very efficient method, exhausting the possibilities for a given configuration of blocks B2 and B3 to just under 2 minutes on a single PC.

3.1 The result

At this stage, we had a list of 44 configurations for the top three rows. We know that *every* possibility for the top three rows can be completed to a full grid in the same number of ways as one of these 44. For example, we had computed that 178848 of the 2612736 possibilities for the top three rows in standard form can be completed to a full grid in the same number of ways as the particular choice:

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

We also computed that this particular choice of top three rows can be completed to a full grid in $72 \times 97961464 = 7053225408$ ways. The same calculations were made for all of the 44 possibilities on our list. This then works out, all at once, how many ways there are to complete 178848 of the 2612736 possibilities for the top rows. We do the same for the other 43, to find exactly how many ways we can complete all of the top three rows to a full grid. This computes the number of grids with B1 in standard form; we need to multiply by $9! = 362880$ to get the total number.

In total, there are $N_0 = 6670903752021072936960 \approx 6.671 \times 10^{21}$ valid Sudoku grids. Taking out the factors of 9! and 72^2 coming from relabelling and the lexicographical reduction of the top row of blocks B2 and B3, and of the left column of blocks B4 and B7, this leaves $3546146300288 = 2^7 \times 27704267971$ arrangements, the last factor being prime. Subsequently, Ed Russell verified the result; it has now been verified by several other people as well. Although our original method required a couple of hours of computer time, Guenter Stertenbrink and Kjell Fredrik Pettersen have subsequently developed a method which completes the entire calculation in less than a second!

Appendix: A heuristic argument

Here is a simple heuristic argument which happens to give almost exactly the right answer. It is due to Kevin Kilfoil.

First, it is clear that there are $N = (9!)^9$ ways to fill in each of the blocks B1–B9 in such a way that each block has the digits 1–9.

We also know the number of ways to fill in the three blocks B1–B3 so that each block has the digits 1–9 and also each row has the digits 1–9; we calculated this above as 948109639680. The same will be true for blocks B4–B6 and B7–B9. So the number of ways to fill in each of the blocks B1–B9 in such a way that each block has digits 1–9, and each row has digits 1–9, is 948109639680^3.

It follows that the proportion of all of the N possibilities which also satisfy the row property is

$$k = \frac{948109639680^3}{(9!)^9}.$$

This will also be the proportion of the N grids satisfying the column property. A Sudoku grid is just one of the N grids that has both the row and column property. *Assuming that these are independent*, this would give the total number of Sudoku grids as

$$Nk^2 = \frac{948109639680^6}{(9!)^9} \approx 6.6571 \times 10^{21}.$$

In fact, this answer cannot be correct (it is not even an integer), and the problem is that the row and column probabilities are *not* quite independent. However, this prediction is really very close to the actual answer we found above: the difference from our exact value is just 0.2%.

The programs and data are stored at `http://www.afjarvis.staff.shef.ac.uk/sudoku/`.

References

[1] S.F.Bammel, J.Rothstein, *The number of* 9×9 *Latin squares*, Discrete Mathematics 11 (1975) 93–95

[2] B.D.McKay, E.Rogoyski, *Latin squares of order 10*, Electronic J. Combin. 2 (1995), Note 3, approx 4pp. (electronic)

[3] B.D.McKay, I.M.Wanless, *The number of Latin squares of order eleven*, Ann. Combin. 9 (2005) 335–344