

Frequency distribution of numbers in a sudoku

Because sudokus are fun.

Ilse van der Linden & Sander van Dorsten

Frequency distribution example (part 1)

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

Number distribution

1 : 3

2 : 2

3 : 3

4 : 2

5 : 3

6 : 5

7 : 3

8 : 5

9 : 4

givens = 30

Freq. distribution

1 : 0

2 : 2

3 : 4

4 : 1

5 : 2

Frequency distribution example (part 2)

freq. distribution

1 : 0

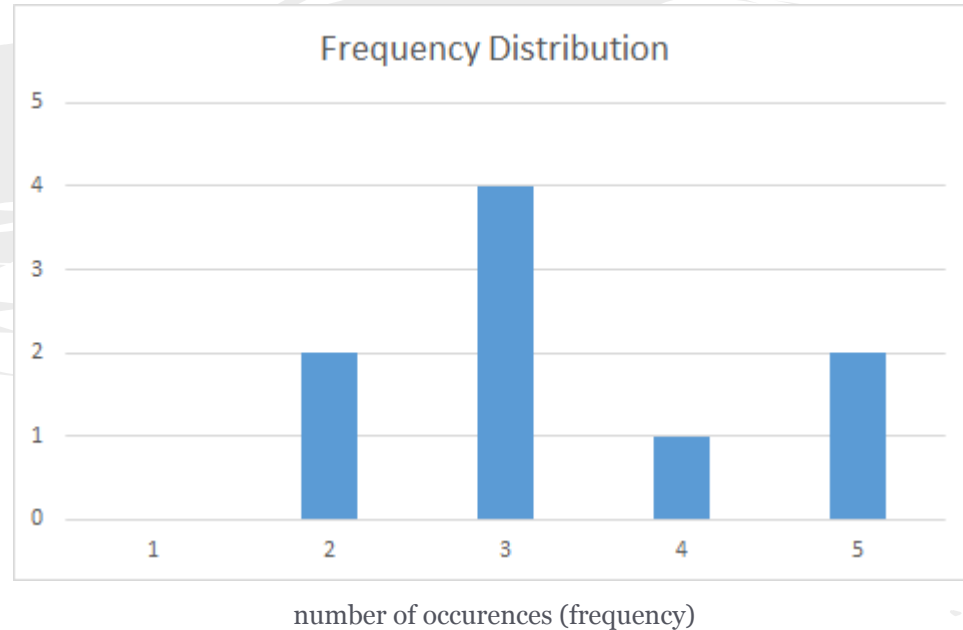
2 : 2

3 : 4

4 : 1

5 : 2

frequency
of
frequency





variance = $1.\overline{1}$



That's great... now what?



Our hypothesis

We predict that a sudoku with a **balanced frequency distribution** would be **easier** to solve by a SAT solver than a sudoku with an **unbalanced frequency distribution**, i.e. a **larger variance**.

Ingredients

- 3 sets of sudokus (17, 26, 33 givens)
- Each set contains 3500 sudokus
 - random freq. distribution → random variance
- Python implementation of picoSAT
- Magic

For every set of sudokus:

1. calculate freq. distribution \rightarrow variance
2. solve all sudokus
3. calculate runtime
4. repeat 2 and 3 and average over runtime

For 17 givens:

variance : occurences

0.32 : 396

0.54 : 1237

0.76 : 1121

0.98 : 694

1.20 : 52

variance : occurrences

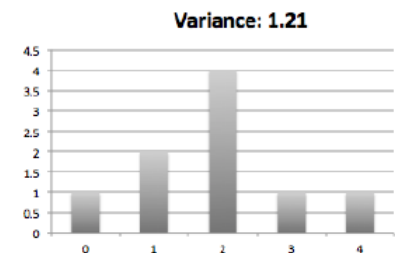
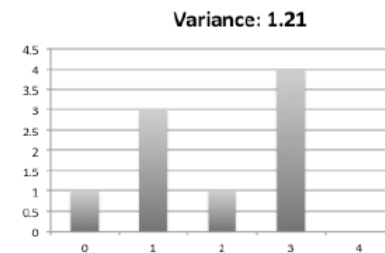
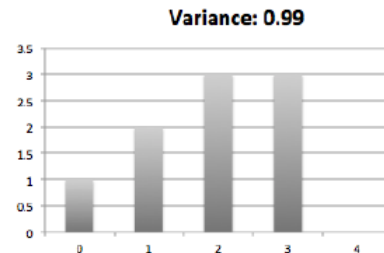
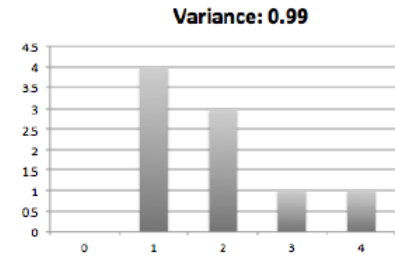
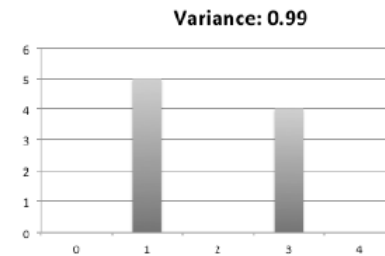
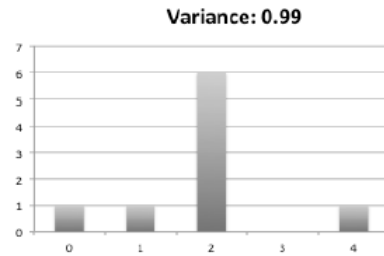
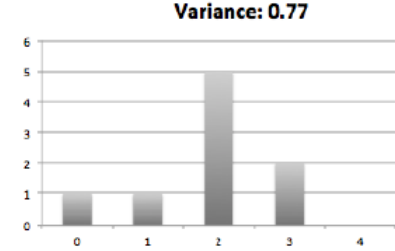
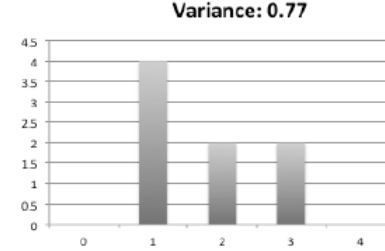
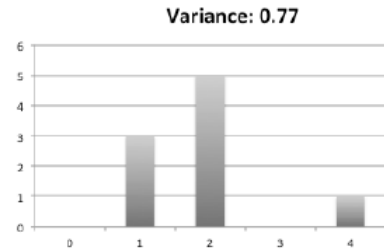
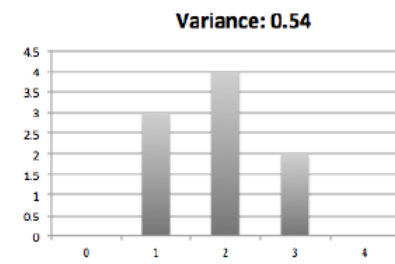
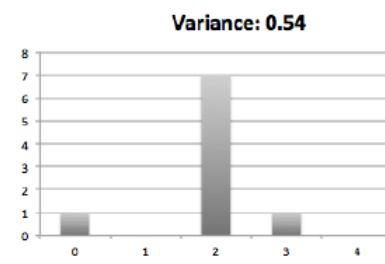
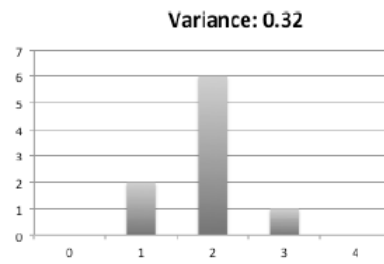
0.32 : 396

0.54 : 1237

0.76 : 1121

0.98 : 694

1.20 : 52



variance : occurrences : runtime in ms (average)

0.32 : 396 : 7.83

0.54 : 1237 : 8.34

0.76 : 1121 : 8.70

0.98 : 694 : 9.57

1.20 : 52 : 10.4



Does this work for given = 26? and 33?

(a) Results 17-givens

Variance	Runtime (ms)
0.32	7.83
0.54	8.34
0.77	8.70
0.99	9.57
1.21	10.4

(b) Results 26-givens

Variance	Runtime (ms)
0.10	5.04
0.32	5.28
0.54	5.30
0.77	5.31
0.99	5.37
1.21	5.40
1.43	5.41
1.65	5.48
1.88	5.60
2.10	5.64
2.32	5.62
2.54	5.57
2.77	5.68
2.99	5.88
3.21	5.90
3.43	5.40
3.65	6.40
3.88	6.00
4.10	5.80

Table 2: Results 33 givens

0.22	4.71
0.44	4.58
0.66	4.53
0.88	4.60
1.11	4.62
1.11	4.41
1.33	4.59
1.55	4.70
1.77	4.72
2.0	4.69
2.00	4.64
2.22	4.69
2.22	4.74
2.44	4.80
2.66	4.82
2.88	4.88
3.11	4.87
3.11	4.77
3.33	4.79
3.33	5.01
3.55	5.09
3.77	5.10
4.00	4.85
4.22	4.99
4.22	5.39
4.44	4.69
4.44	5.39
4.66	5.60
5.11	5.20



Conclusion

We predict that a sudoku with a **balanced frequency distribution** would be **easier** to solve by a SAT solver than a sudoku with an **unbalanced frequency distribution**, i.e. a **larger variance**.

- 17 givens: effect
 - 26 givens: some effect
 - 33 givens: minimal effect
-
- the more decisions, the more it matters



Thank you

Ilse van der Linden & Sander van Dorsten