Comparison of classification results for collections of minimal Sudoku puzzles obtained by different random generators

The data and programs in this repository were the basis for the comparison results below.

The goal of this section is to elucidate how classification results for collections of minimal puzzles may vary according to the type of generator used to produce the collection.
It is also to show that some results remain valid independently of which random generator has been used.
The most striking result is undoubtedly that **all the puzzles (about 10,000,000) generated by all the types of random generators mentioned in this page (including controlled-bias - see section 2) can be solved by whips. This gives an idea of how powerful whips are.**
This is an old page from my website, with cosmetic updates for inclusion in this repository.

Thanks to Allan Barker, Mike Barker, Coloin, Eleven, gsf, Mike Metcalf, Red Ed, Ronk, StrmCkr, Tarek and ttt for their participation in discussions on the topics of this section in the Sudoku Player's Forum:
http://www.sudoku.com/boards/viewtopic.php?t=5995
Thanks to Guenter Stertenbrink for the (previously existing two versions of the) suexg generator I used here.

1) Definition of the random generators considered

Four generators have been used. Each of them produces uncorrelated samples of minimal puzzles.
But, as the distributions obtained are different, some of them must be biased. In the next section, we shall see that these four generators are biased.

- two "top-down" generators:
-- suexg-td (*the* classical top-down generator, present in the "PROGRAMS" folder)
--Allan Barker's new generator (private)
Allan's description of the two phases:
1) Generate complete grids
   1.1) Random placement of 81 clues
   1.2) Random Monte-Carlo pairswap convergence to valid solution
2) Generate minimal puzzles
   2.1) Random single clue removal to 55 givens
   2.2) Random single clue removal and test for single solution
   2.3) Random 1 pass removal of extra clues, to local minimum

- two "bottom-up" generators:
-- suexg-bu (present in the "PROGRAMS" folder)
More or less the same procedure as Mike's bottom-up below.
-- Mike Metcalf's bottom-up generator (private)
Mike's description:

1) To a blank grid add 18 random values at 18 random and distinct locations consistent with the basic constraints.
2) Find the number of solutions.
3) If zero solutions GOTO 1.
4) If one solution GOTO 9 (never happens, but you never know!).
5) If multiple solutions add a new clue and find number of solutions.
6) If zero solutions remove last clue added and GOTO 5.
7) If one solution GO TO 9.
8) If multiple solutions GOTO 5.
9) Visit each clue in turn randomly. If it is redundantremove it.

To produce minimal puzzles with suexg-td or suexg-bu, one can use the following (Unix) command line, supposing it has been compiled and named suexg.exe:
./suexg.exe seed number_of_puzzles [puzzles.txt]
The optional "puzzles.txt" part allows to write the puzzles in a file named "puzzles.txt" instead of seeing them in the terminal.
"seed" is any integer, used as the seed for the random numbers generator needed by the generator.
"number_of_puzzles" is the desired number of minimal puzzles. It outputs about 20-40 puzzles per second on a standard PC.
Example (to generate the sudogen0_1M collection): ./suexg.exe 0 1000000 > sudogen0_1M.txt

2) Definition of the collections of puzzles used for the statistics

The statistics below are based on the following collections of puzzles, generated respectively with the above 4 generators:
- sudogen0_1M (1,000,000 puzzles)
- rabrnd_1M (1,000,000 puzzles)
- suexg-bu-0_1M (1,000,000 puzzles)
- Mike#5 (64,000 puzzles)

Except for sudogen0_1M, for which all the computations were done on the full collection, smaller sub-collections have been used for some computations that would have taken too much time otherwise, as specified below:
- all the computations that involve only the number of clues and the SER are based on the whole collections,
- computations that involve the W+S rating are limited to 50,000 puzzles for rabrnd_1M, to 10,000 for suexg-bu-0_1M and to 50,000 for Mike#5.

These limitations are purely opportunistic and don't imply any a priori judgment on the 4 generators. They are largely enough in practice to validate the results stated below.

3) Results

In addition to the mean and standard deviation, two parameters of a distribution will be of interest: skewness and kurtosis.

Skewness is a measure of the asymmetry of the distribution
- negative skew: the left tail is longer; the mass of the distribution is concentrated on the right of the mean; the distribution has relatively few low values.
- positive skew: the right tail is longer; the mass of the distribution is concentrated on the left of the

mean; the distribution has relatively few high values.
Details on Wikipedia (http://en.wikipedia.org/wiki/Skewness)

Kurtosis (also called "excess kurtosis" to avoid ambiguities) is a measure of how much the distribution shape differs from that of a Normal distribution with the same mean and standard deviation as X. More precisely:
- lower kurtosis means more of the variance is due to frequent modestly sized deviations;
- higher kurtosis means more of the variance is due to infrequent extreme deviations.
Kurtosis can vary between - 2 and + infinity.
Details on Wikipedia (http://en.wikipedia.org/wiki/Kurtosis)

3.1) Global results

generator	suexg-td (top-down)	Allan Barker's (top-down)	suexg-bu (bottom-up)	Mike Metcalf's (bottom-up)
collection name	sudogen0_1M	rabrnd_1M	suexg-bu-0_1M	Mike#5
size of sample used (unless other size stated)	1,000,000	1,000,000	1,000,000	64,000
mean(#clues)	24.38	24.38	23.87	23.89
standard deviation(#clues)	1.12	1.116	1.08	1.08
skewness(#clues)	0.08	0.11	0.11	
kurtosis(#clues)	0.007	0.014	0.026	
mean(W) (size of sub-sample used)	1.94	1.946 (50,000)	1.80 (10,000)	1.844 (50,000)
standard deviation(W) (size of sub-sample used)	1.29	1.29 (50,000)	1.24 (10,000)	1.24 (50,000)
skewness(W)	1.01	1.01 (50,000)	1.24 (10,000)	(50,000)
kurtosis(W)	0.27	0.34 (50,000)	0.92 (10,000)	(50,000)
mean(SER)	3.77	3.77	3.50	3.58
standard deviation(SER)	2.42	2.42	2.33	2.34
skewness(SER)	0.60	0.60	0.79	
kurtosis(SER)	-1.36	-1.35	-1.08	
correlation coeff (W, SER) (size of sub-sample used)	0.895	0.895 (50,000)	0.898 (10,000)	0.896 (50,000)
correlation coeff (W, log(#chains)) (size of sub-sample used)	0.946	0.946 (50,000)	0.946 (10,000)	0.96 (50,000)
correlation coeff (#clues,	0.115	0.117	0.096	0.112

W) (size of sub-sample used)		(50,000)	(10,000)	(50,000)
correlation coeff (#clues, SER)	0.120	0.121	0.11	0.094
max W (*)	13	13 (50,000)	11 (10,000)	11 (50,000)
max SER (*) (size of sub-sample used)	9.3	9.3	9.2	9.2

(*) comparison of maximum values for samples of different sizes is not meaningful

3.2) The W+S rating (mean, standard deviation, skewness and kurtosis) as a function of the number of clues

The W+S rating, a measure of complexity based on Subsets plus whips, has been defined in [PBCS].

generator	suexg-td (top-down)	Allan Barker's (top-down)	suexg-bu (bottom-up)	Mike Metcalf's (bottom-up)
collection name	sudogen0_1M	rabrnd_1M	suexg-bu-0_1M	Mike#5
size of sub-collection used	1,000,000	50,000	10,000	50,000
# clues	mean ||| standard deviation skewness ||| kurtosis # puzzles	mean ||| standard deviation skewness ||| kurtosis # puzzles ||| #puzzles x 20	mean ||| standard deviation skewness ||| kurtosis # puzzles	mean ||| standard deviation skewness ||| kurtosis # puzzles ||| #puzzles x 20
19	0	0	0	0
20	1.41 ||| 0.72 5e-5 ||| -0.06 44	0.97 ||| 0.05 (*) -4e-5 ||| -1.5 3 ||| 60	2.5 ||| 0.5 (*) 0.0 ||| -2.0 2	1.20 ||| 0.46 (*) 16 ||| 320
21	1.58 ||| 1.04 0.0037 ||| 1.52 2,428	1.73 ||| 1.23 (*) 0.003 ||| 0.43 117 ||| 2,340	1.65 ||| 1.00 0.011 ||| 0.72 88	1.64 ||| 1.05 393 ||| 7,860
22	1.69 ||| 1.13 0.048 ||| 1.43 34,548	1.70 ||| 1.17 0.05 ||| 1.76 1,733 ||| 34,660	1.63 ||| 1.15 0.13 ||| 1.83 824	1.65 ||| 1.13 4,032 ||| 80,640
23	1.77 ||| 1.20 0.22 ||| 0.83 172,512	1.78 ||| 1.19 0.22 ||| 1.07 8,611 ||| 172,220	1.70 ||| 1.20 0.42 ||| 2.84 2671	1.74 ||| 1.19 13,603 ||| 272,060
24	1.88 ||| 1.26	1.88 ||| 1.26	1.81 ||| 1.24	1.84 ||| 1.24

	0.38 ||| 0.54 342,335	0.37 ||| 0.28 17,216 ||| 344,320	0.45 ||| 0.72 3679	 18,335 ||| 366,700
25	2.01 ||| 1.31 0.28 ||| 0.096 297,838	2.00 ||| 1.30 0.27 ||| 0.06 14,905 ||| 298,100	1.91 ||| 1.27 0.20 ||| -0.12 2106	1.98 ||| 1.29 10,439 ||| 208,720
26	2.17 ||| 1.36 0.008 ||| -0.31 122,116	2.21 ||| 1.39 0.09 ||| 0.36 6,054 ||| 121,080	2.06 ||| 1.35 0.04 ||| -0.58 550	2.11 ||| 1.35 2,761 ||| 55,220
27	2.38 ||| 1.40 0.01 ||| -0.52 25,315	2.37 ||| 1.39 0.01 ||| -0.5 1,221 ||| 24,420	2.63 ||| 1.33 -0.0004 ||| -1.30 71	2.37 ||| 1.41 406 ||| 8,120
28	2.64 ||| 1.44 0.0007 ||| 0.75 2,686	2.61 ||| 1.52 0.001 ||| -0.03 130 ||| 2,600	2.64 ||| 1.44 (*) 9e-5 ||| -1.36 9	2.11 ||| 1.33 (*) 14 ||| 280
29	2.86 ||| 1.45 1e-5 ||| -0.82 168	3.11 ||| 1.45 (*) 9 ||| 180	 0	3.00 ||| 0.00 (*) 1 ||| 20
30	3.19 ||| 1.10 (*) -4e-6 ||| -0.14 10	0.9 ||| 0.0 (*) 4e-5 ||| -0.21 1 ||| 20	 0	 0 ||| 0
all	**1.94 ||| 1.29** **1.01 ||| 0.27** 1,000,000	1.94 ||| 1.28 30,000	1.94 ||| 1.28 10,000	1.844 ||| 1.242 50,000 ||| 1,000,000

(*) values based on small samples are not meaningful

3.3) The SER rating (mean, standard deviation, skewness and kurtosis) as a function of the number of clues

The definition of the SER rating, based on the rules in Sudoku Explainer, has been recalled in the ratings page.

generator	suexg-td (top-down)	Allan Barker's (top-down)	suexg-bu (bottom-up)	Mike Metcalf's (bottom-up)
collection name	sudogen0_1M	rabrnd_1M	suexg-bu-0_1M	Mike#5
size of sample used	1,000,000	1,000,000	1,000,000	50,000
# clues	mean ||| standard deviation	mean ||| standard deviation	mean ||| standard deviation	mean ||| standard deviation

	skewness ||| kurtosis # puzzles	skewness ||| kurtosis # puzzles	skewness ||| kurtosis # puzzles	skewness ||| kurtosis # puzzles ||| #puzzles x 20
19	0	3.0 ||| 0.0 (*) 1	0	0
20	2.75 ||| 1.70 (*) 7e-5 ||| 1.08 44	2.98 ||| 1.94 (*) 8e-5 ||| 0.76 56	3.01 ||| 1.99 0.0004 ||| 0.25 281	2.60 ||| 1.61 (*) 16 ||| 320
21	3.03 ||| 2.00 0.003 ||| 0.28 2,428	3.19 ||| 2.11 0.003 ||| -0.20 2,392	3.02 ||| 2.00 0.012 ||| 0.28 8,559	3.26 ||| 2.16 393 ||| 7,860
22	3.27 ||| 2.16 0.037 ||| -0.50 34,548	3.27 ||| 2.17 0.036 ||| -0.48 33,980	3.17 ||| 2.12 0.095 ||| -0.28 82,423	3.19 ||| 2.13 4,032 ||| 80,640
23	3.44 ||| 2.28 0.15 ||| -0.88 172,512	3.43 ||| 2.27 0.15 ||| -0.86 171,607	3.32 ||| 2.24 0.27 ||| -0.69 276,719	3.35 ||| 2.46 13,603 ||| 272,060
24	3.65 ||| 2.38 0.24 ||| -1.21 342,335	3.63 ||| 2.37 0.24 ||| -1.20 342,747	3.52 ||| 2.34 0.29 ||| -1.08 363,806	3.57 ||| 2.37 18,335 ||| 366,700
25	3.90 ||| 2.46 0.15 ||| -1.48 297,838	3.90 ||| 2.46 0.15 ||| -1.48 298,548	3.79 ||| 2.44 0.12 ||| -1.41 205,938	3.83 ||| 2.45 10,439 ||| 208,720
26	4.22 ||| 2.53 0.03 ||| -1.68 122,116	4.23 ||| 2.53 0.03 ||| -1.68 122,583	4.16 ||| 2.52 0.016 ||| -1.66 54,493	4.12 ||| 2.52 2,761 ||| 55,220
27	4.63 ||| 2.56 -0.0006 ||| -1.74 25,315	4.62 ||| 2.56 -0.0003 ||| -1.73 25,124	4.58 ||| 2.55 -6e-6 ||| -1.74 7,213	4.58 ||| 2.58 406 ||| 8,120
28	5.11 ||| 2.56 -0.0009 ||| -1.59 2,686	5.16 ||| 2.51 -0.001 ||| -1.55 2,798	5.14 ||| 2.47 -0.0002 ||| -1.54 538	4.44 ||| 2.77 (*) 14 ||| 280
29	5.43 ||| 2.46 -0.0001 ||| -1.28 168	5.40 ||| 2.54 -8e-5 ||| -1.40 156	6.43 ||| 1.84 (*) -4e-5 ||| 1.57 24	6.00 ||| 0.00 (*) 1 ||| 20
30	6.46 ||| 1.74 (*) -2e-5 ||| 3.6 10	5,69 ||| 2.02 (*) -9e-6 ||| 0.22 8	4.75 ||| 2.45 (*) 0.0 ||| -2.0 2	0 ||| 0
all	**3.77 ||| 2.42 0.60 ||| -1.36** 1,000,000	**3.77 ||| 2.42 0.60 ||| -1.35** 1,000,000	**3.53 ||| 2.34 0.79 ||| -1.08** 1,000,000	3.567 ||| 2.35 50,000 ||| 1,000,000

(*) values based on small samples are not meaningful

4) General conclusions

The most striking results are:

- ***the possibility of solving by whips all the puzzles generated by these generators;***

- ***the close similarity between the 2 top-down generators***:
  - same mean number of clues,
  - same distribution of the number of clues,
  - close mean SER and mean W, globally and for each number of clues;

- ***the close similarity between the 2 bottom-up generators:***
  - same mean number of clues,
  - same distribution of the number of clues,
  - close mean SER and mean W, globally and for each number of clues;

- ***the clear difference between the top-down and bottom-up generators:***
  - notably larger mean number of clues for the top-down generators,
  - notably different distributions of the number of clues,
  - different mean SER and mean W for each number of clues, with a slightly higher SER and W for the top-down generators,
  - globally larger mean SER and mean W for the top-down generators;

- ***for the 4 generators, a number of clues between 20 and 30 for all the puzzles*** (exceptionally, one puzzle with 19 clues with Allan's

generator);

- ***for the 4 generators, a small upward trend for the mean SER and mean W when the number of clues increases from 21 to 29*** (not enough data to allow any conclusion below 21 or above 29 - except for sudogen0_1M, for which the trend appears in the whole 20-30 span)*;

- ***nevertheless, for the 4 generators, a very small (~ 0.1) correlation coefficient between the number of clues and the SER or W, which implies that the number of clues can't be used for predictions of the (SER or W) complexity of an individual puzzle***.

(*) In addition to the above tables, the upward complexity trend seems to be confirmed for (not completely random) collections of puzzles: for the 17-clue puzzles (based on Gordon's Royle collection of 36,628 17-clue puzzles) and for the 30- to 35- clue puzzles, based on collections of ~ 1,000 puzzles each generated by Allan Barker's Monte Carlo procedure;
- 17-clue mean SER: 2.55
- 30-clue mean SER: 6.32
- 31-clue mean SER: 6.67
- 32-clue mean SER: 7.89
- 33-clue mean SER: 7.14
- 34-clue mean SER: 7.25
- 35-clue mean SER: 7.54

But how these collections can be related to those obtained with the above generators is unclear.

The following general results are true for all the generators considered above:

- ***all the minimal puzzles built with random generators can be solved by whips***;
- ***at least 99% of the minimal puzzles can be solved by whips of length 5 or less***;
- ***at least 99,9% of the minimal puzzles can be solved by whips of length 7 or less***;
- ***the SER rating provides (after conversion) a statistically good approximation of the W+S rating (which is harder to compute, but purely logic, intrinsic and invariant under symmetry)***.