Sudoku Players' Forums

? FAQ   **Q Search**   **Ⅲ Memberlist**   **⊞ Usergroups**
⚑ Profile   **⚑ You have no new messages**   **⚑ Log out [denis_berthier]**

Rating rules / Puzzles. Ordering the rules
Goto page **Previous** **1**, **2**, **3** ... , **10**, 11, **12** **Next**

[new topic] [post reply]   **Sudoku Players' Forums Forum Index** -> **Advanced solving techniques**

	View previous topic :: View next topic
Author	**Message**

---

m_b_metcalf

Joined: 15 May 2006
Posts: 2106
Location: Berlin

⬜ Posted: Wed Jun 24, 2009 2:11 pm   Post subject: [quote]

> **denis_berthier wrote:**
>
> > Is your generator available? Or is there any place where the basic principles it uses are defined? Or can you state them in a few words?
> >
> > In any case, I'm interested by your file of 110840 puzzles. It'd be a good way to check my results about the nrczt and SER ratings and to confirm that they don't depend on the generator.

I generate a grid by filling its first row with the nine digits in a random order, obtained using *random_number* (a Fortran 95 intrinsic function), and then adding subsequent rows of randomly arranged digits, sorting their order as necessary to avoid clashes with digits already in place on previous rows or already in the current box. If this cannot be done, a fresh set of random numbers is used. For the third and sixth rows, an additional check is required that the box constraint is also fulfilled. Also, a deadlock condition can arise on row 6 when the three digits missing from an almost completed box have already been assigned to a single column above the box. If this occurs, the program restarts at row 4.

I generate a puzzle by selecting, say, 60 clues from a grid at random and then removing further clues at random, checking the puzzle remains minimal at each step.

This is part of a larger program for which you would need a Fortran 95 compiler (of which there are two free ones available but I use the highly optimizing Intel v. 10).

If you send me a PM with your e-mail address I will send you the puzzle file.

Regards,

Mike Metcalf

Back to top   [👤 profile] [👥 pm]

---

denis_berthier

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

⬜ Posted: Wed Jun 24, 2009 2:22 pm   Post subject: [quote] [edit]

> **m_b_metcalf wrote:**
>
> > I generate a grid by filling its first row with the nine digits in a random order, obtained using *random_number* (a Fortran 95 intrinsic function), and then adding subsequent rows of randomly arranged digits, sorting their order as necessary to avoid clashes with digits already in place on previous rows or already in the current box. If this cannot be done, a fresh set of random numbers is used. For the third and sixth rows, an additional check is required that the box constraint is also fulfilled. Also, a deadlock condition can arise on row 6 when the three digits missing from an almost completed box have already been assigned to a single column above the box. If this occurs, the program restarts at row 4.
> > I generate a puzzle by selecting, say, 60 clues from a grid at random and then removing further clues at random, checking the puzzle remains minimal at each step.

I need some time to think about whether this is not biased.

> **m_b_metcalf wrote:**
>
> If you send me a PM with your e-mail address I will send you the puzzle file.

Thanks, this is the easiest way.

regards.

Back to top   profile   pm   www

coloin

Posted: Wed Jun 24, 2009 4:00 pm   Post subject: quote

Joined: 06 May 2005
Posts: 986
Location: Oxford

Well we investigated the generation of unbiased grids here, inc solution grids from puzzles made from suexg.

http://www.sudoku.com/boards/viewtopic.php?t=5128&start=0

Essentially if you have an average count of 11.58 U4s in your [solution] grids your grids are not biased. Grids made with suexg had average value 10.67

> **RedEd wrote:**
>
> Back to size-4 unavoidables: the average number is exactly 320784513273 / 27704267971, or roughly 11.58 per grid.

However I still feel that the method of removing clues to minimality predisposes slightly to smaller puzzles.

Take this random solution grid.

Code:

```
246518739791342658583769214875294361129635847634871925352187496918426573467953182
solution grid
```

Code:

```
lines:206083 average clues:24.342556 – similar to random grids
    20   9
    21   539
    22   7631
    23   36975
    24   71389
    25   60392
    26   23904
    27   4723
    28   490
    29   30
    30   1
```

because there are so many minimal puzzles in a single grid, let us look at a 40-clue subgrid...[25clue puzzle plus 15 more extra clues.]

Code:

```
.4......97....2.585...6..1...5....6....63...7.....192...21.....9.8.....3.679.....
25puzzle
24..18..97..3.2.585.3.6..1...5.94.6.1..63...7..4..1925..2187...9.8..6..3.679...82
25plus15puzz
```

If we generate 1000000 puzzles from this subgrid we get 372190 different puzzles [almost certainly there are more puzzles, especially with 27-29 clues]

Code:

```
puzz:372190 average clues:24.897902
```

```
            21   45
            22   1257
            23   21428
            24   105884
            25   148341
            26   77159
            27   16721
            28   1330
            29   25
```

compare with the clue stats from 500,000 puzz

Code:
```
puzz:257081 average clues:24.778035
21   45
22   1247
23   19460
24   81487
25   98436
26   46276
27   9416
28   699
29   15
```

If you generate "0nly" 16000 puzzles from the 40 clue subgrid you get this distribution

Code:
```
puzz:15266 average clues:24.392637
    21   27
    22   400
    23   2455
    24   5584
    25   4645
    26   1798
    27   340
    28   16
    29   1
```

there is a considerable skew towards smaller puzzles

C

Back to top   [profile] [pm]

denis_berthier   Posted: Wed Jun 24, 2009 4:57 pm   Post subject: [quote] [edit]

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

> **coloin wrote:**
> Well we investigated the generation of unbiased grids here, inc solution grids from puzzles
> made from suexg.
> http://www.sudoku.com/boards/viewtopic.php?t=5128&start=0
> Essentially if you have an average count of 11.58 U4s in your [solution] grids your grids are not
> biased. Grids made with suexg had average value 10.67

Thanks for the reference. I need more time to study this. The U4 argument is interesting for *solution grids*.
But the question I'm interested in here is puzzles, not solution grids.
I think it would be very difficult to test the unbiasedness of every step in the generation process of a
sequence of minimal puzzles. That's why I tried several direct tests, based on the autocorrelation of
numerical functions of the sequence.

As for your examples, I'm not sure what you want to prove, but no stats can be done based on a single
solution grid. Starting from a different one, you could obtain very different results.

Back to top   [profile] [pm] [www]

coloin   Posted: Wed Jun 24, 2009 10:50 pm   Post subject: [quote]

Joined: 06 May 2005
Posts: 986
Location: Oxford

Indeed, you are dealing with minimal puzzles, and that said, the degree of bias for the solution grids made by suexg is not that great.

However I was demonstrating that smaller puzzles are produced preferentially by the "clues removal" method of these programs.

I appreciate that taking puzzles from a single grid or subgrid might seem unwise.

I would hazard a guess that , w.r.t ratings, 1M "random" puzzles from a single grid would be as representative as 1M random puzzles from 1M random grids - but I could well be wrong. The clue distribution from this complete random grid is similar to your distribution.

I would suggest that the [lack of] correlation between the number of clues and the rating isnt just as clear cut as you as indicate.

here are the ratings from the initial 64000 of your 1M series, which you may chose to confirm.

Code:
```
173    28-puzzles Mean SER 5.12 Mean NRCZT-level 2.66
21972 24-puzzles Mean SER 3.68 Mean NRCZT-level 1.89

Total 1M puzzles Mean SER 3.78 Mean NRCZT-level 1.94
```

This doesnt look like a chance occurance - perhaps confirming that large puzzles tend have a higher than average rating, but I'm not sure why.

C

Back to top   [profile] [pm]

denis_berthier   Posted: Thu Jun 25, 2009 5:02 am   Post subject: [quote] [edit]

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

Coloin,

Here are the mean number of clues and standard deviation for each nrczt-level in the sudogen0_1M collection:

Code:
```
Level    #Puzzles      E(Clues)     s(Clues)
1_0      417624        24.29        1.089
1        120618        24.22        1.11
2        138371        24.35        1.12
3        168355        24.52        1.13
4        123153        24.60        1.14
5        24187         24.63        1.14
6        5511          24.68        1.13
7        1514          24.69        1.13
8        473           24.67        1.10
9        130           24.45        1.02
all      1M            24.38        1.11
```

Obviously, nothing can be concluded from this, the maximum deviation (0.31) of each subset from the global mean (24.38) being much smaller than the global standard deviation (1.11).

Conversely, here is, for each number of clues (except 20, 29 and 30, which don't have enough instances to be meaningful), the means and standard deviations for the nrczt-level and the SER:

Code:
```
#Clues  #Puzzles    E(Level)   s(Level)    E(SER) s(SER)
21      2428        1.58       1.04        3.03   2.0
22      34548       1.69       1.13        3.27   2.16
23      172512      1.78       1.20        3.44   2.28
```

```
24        342335      1.88      1.26      3.65      2.38
25        297838      2.01      1.31      3.90      2.46
26        122116      2.17      1.36      4.22      2.53
27        25315       2.38      1.40      4.63      2.56
28        2686        2.64      1.44      5.11      2.56
all       1M          1.94      1.29      3.77      2.42
```

Now, the question is, what conclusions can be drawn from this?

Firstly, it seems that there is indeed a tendency for a slightly higher mean complexity with increasing number of clues (whether we consider nrczt or SER).
But one should consider that the maximum deviation (0.7 for NRCZT, 1.44 for SER) of the mean of each subset from the global mean (NRCZT = 1.94 or SER = 2.42) is much smaller than the global standard deviation (1.29 for NRCZT, 2.42 for SER).
So, it seems that there is a trend, but it can't be used for predictions, due to the very low correlation coefficients (0.12) and the very large standard deviations.

Secondly, there is a different question: does this slight trend indicate any bias in the suexg generator?
There's no reason to suppose so. The slight trend (if confirmed) may be real, i.e. it may exist in the set of all the minimal puzzles.

Back to top   profile   pm   www

denis_berthier   Posted: Thu Jun 25, 2009 5:37 am   Post subject: quote   edit

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

Top sudogen0_1M

Here is now a short list of "interesting puzzles" from the sudogen0_1M collection.
They have been chosen because they satisfied at least one of the following "(relatively) extreme" conditions:
- SER > 9.1,
- NRCZT >= 9.

Code:

```
#puzzle    Puzzle
           SER     NRCZT   #clues
707
4..3..6.8..5..1.3..16.4...7.7.....8......2....51..637...2.9.8.....1..........4..9
   9.2    12      25
2035
...7...637....4.9..6.95....8.....7...5.6.....9...2..16..3..8..224.....7.......9..
   9.2    10      24
88857
3.........41......69578.3..51.64.78.4....3.5...7......9....6..2...9..1...5.....6.
   9.2    12      26
117617
91.......7..2...8....3..16..7..6.2..5.1.........6....91.....5.......4.8.6.2.8.7..3
   9.2    12      24
124235
1......6..6..43...9....1..7..8..4.25.1...76......5.........8.2..4.9..5.8..3..7..
   9.3    12      24
143634
..5.3.4.2....54.9.1.........19.4..8.4.3..72....85....99..6.8...........7.27..5...
   9.2    10      26
175675
..85.....4......2..3..6.9.8....82...2.....3...7.9.1....9.8..6....52...7....7....1
   9.2    11      23
188892
...5...476...4..3..9....1....2...5..7.......8...6...1.9..46......6..1..28..9.24..
   9.2    13      24
200753
....6.3..4...2....1..1.4..8...86..9.3.....7.....4...816.7......8..9....7...34...2..
   9.2    10      24
298676
2..5...9..6...5........7.1.4.5.....6.8.3.6.4....8..7....1...43..2...89..9...1..5.
   9.2    13      25
346155
7...63....4....76...3.......9...45.23.4...7.47..8...1..72......1..492......3.51.
   9.2    10      27
```

```
420249
...37.4.....4..5.2.5.....872..9......6..24...78........4...17.5.1.......5.....8.9
   9.2    11    24
425702
..8..2....278...3.9....1....6..........3..95..5..1.4.61...592.4....8...5...1..7..
   9.2    12    25
431065
.3.....7.9....16...67.....1..9.467....61.5.4....3...1......8.......1...4..495716.
   9.2    12    27
618406
15.8....2...4.19......1..3.4....78...6..2.1......8..9....2.9..3....4.5...71....6
   9.3    12    25
631828
1.57....4..6..53.....8...5...7.5...94.......6.2.6.9.......7.....6...3.92.3....41.
   9.2    10    25
641898
..74.2...48.....9..1..5.....6.19.34.4........7..8.9...2.5...3.....1.6.94.3..7.
   9.2    10    26
717866
...51....5.4.......21....9..7..6.42...3..8....9.7....83....2.1.7..8....9.....7.5.
   9.2    11    24
739105
..7..6.8....2...46....3.2..41......27......5...2.19..4..16.5...38.12...95........
   9.2     9    26
792244
1.........43..5.....938...1...92.5..9..6.8.23......9....4..6.75...7.....6...1..9.
   9.2    11    25
810153
..64......7..5....9....6..8.....5.8..2.9..3....83..7..5.4.37.1.........7...12.56.
   9.2     9    25
845088
.8..1...6..3...5..2....7.....2.3..........8.947.8...2.6.....9.73..2.4.......5....
   9.2    12    22
942062
.4......96....98..7.......2..7.13....8.....2.1...5.9......2.5..5..4...78..1.7...4
   9.2    12    24
```

Back to top   [profile] [pm] [www]

---

m_b_metcalf   Posted: Thu Jun 25, 2009 6:42 am   Post subject: [quote]

Joined: 15 May 2006
Posts: 2106
Location: Berlin

For what it's worth, I generated an hour's worth of puzzles all from the same grid. The summary is:

Code:

```
Number:         54074
Average:    24.25319

      20        4
      21      194
      22     2443
      23    10709
      24    19050
      25    14842
      26     5588
      27     1117
      28      120
      29        7
```

Another game I played was to run a tiny sample of yesterday's puzzles through a program that pair-wise changes the values of the clues (but not their positions) and removes any clue that thereby becomes redundant. This I repeated several times. The result:

Code:

```
Iteration 1

   Number:          143
   Average:       20.45

      17        0
      18        3
      19       14
      20       64
      21       42
```

```
              22        17
              23         3
              24         0

   Iteration 2

      Number:           143
      Average:        20.09

              17         0
              18         3
              19        30
              20        71
              21        31
              22         6
              23         2
              24         0


   Iteration 3

      No change.
```

Regards,

Mike Metcalf

| **Back to top** | [profile] [pm] | |

| **denis_berthier** | Posted: Thu Jun 25, 2009 12:23 pm   Post subject: | [quote] [edit] |

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

> **m_b_metcalf wrote:**
> I generate a grid by filling its first row with the nine digits in a random order, obtained using *random_number* (a Fortran 95 intrinsic function), and then adding subsequent rows of randomly arranged digits, sorting their order as necessary to avoid clashes with digits already in place on previous rows or already in the current box. If this cannot be done, a fresh set of random numbers is used. For the third and sixth rows, an additional check is required that the box constraint is also fulfilled. Also, a deadlock condition can arise on row 6 when the three digits missing from an almost completed box have already been assigned to a single column above the box. If this occurs, the program restarts at row 4.

It seems that not all the rows are dealt with in the same way and this may introduce a bias, although which kind of bias it can be is not clear for me.

> **m_b_metcalf wrote:**
> I generate a puzzle by selecting, say, 60 clues from a grid at random and then removing further clues at random, checking the puzzle remains minimal at each step.

Am I wrong to say it is the same method as suexg?

It is very difficult to analyse how much the (possible) bias in the first step is smoothened by the second step.

Anyway, the results I obtain are significantly different on several points from those I obtained with suexg (see next post).

| **Back to top** | [profile] [pm] [www] | |

| **denis_berthier** | Posted: Thu Jun 25, 2009 12:25 pm   Post subject: | [quote] [edit] |

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

For a set of 50,000 random puzzles provided by m_b_metcalf with his generator, here is, for each number of clues, the means and standard deviations for the nrczt-level and the SER (these last variables don't have enough instances to be meaningful for #clues = 20, 29 and 30):
(I used only the 50,000 first puzzles in a series of 110,000).

Code:

```
#Clues #Puzzles  E(Level) s(Level)   E(SER) s(SER)
20      2        0.95     0.05        1.6   0.10
```

```
21      136      1.33      0.80      2.64      1.63
22      1801     1.34      0.84      2.57      1.58
23      8736     1.31      0.82      2.49      1.55
24      17145    1.26      0.78      2.41      1.50
25      14852    1.22      0.73      2.32      1.38
26      5952     1.19      0.69      2.26      1.31
27      1225     1.14      0.58      2.15      1.11
28      147      1.10      0.45      2.02      0.71
29      4        0.9       0.0       1.95      0.29
30      0
all     50k      1.25      0.76      2.41      1.48
```

E(clues) = 24.36
s(clues) = 1.12

E(NRCZT = 1.25
s(NRCZT) = 0.76

E(SER) = 1.95
s(SER) = 1.48

It can be noted that:
- contrary to the sudogen0_1M case, the mean level (NRCZT and SER) is (slightly) decreasing with the number of clues. But, here again, the maximum deviation of the mean value is much less than the global standard deviation. So, again, this may not be very meaningful;
- the mean number of clues is very close : 24.36 instead of 24.38 (and the standard deviations also (1.12 instead of 11.11);
- the mean NRCZT is much smaller :1.25 instead of 1.94
- the mean SER is much smaller :1.95 instead of 3.77

The classification according to NRCZT levels shows that:
- the max SER is 9.0
- the max NRCZT is 10 (but with only one instance; there's no puzzle with NRCZT = 9 and only one instance with NRCZT = 8)

Code:

```
Level  Number
1_0    32146
1      8055
2      5768
3      2464
4      1327
5      201
6      27
7      10
8      1
9      0
10     1
11     0
12     0
13     0
```

This shows that:
- there are many more puzzles at level 1_0 (i.e. solved by singles) than with suexg: 64.3% instead of 41.7%
- there are fewer puzzles at levels > 7 : 4/100.000 instead of 66.7/100.000.

Globally, m_b_metcalf's generator produces easier puzzles than suexg (wrt to the NRCZT or SER ratings). This may be related to a bias introduced by the first generation phase, but the nature of this (possible) bias is not clear for me.

Back to top   [profile] [pm] [www]

m_b_metcalf   Posted: Fri Jun 26, 2009 1:56 pm   Post subject: [quote]

Joined: 15 May 2006
Posts: 2106
Location: Berlin

denis_berthier wrote:

Average # of clues: 24.36532 vs 24,38

Code:

```
#Clues   #Puzzles        #Puzzles*1,000,000/51492 vs my results
    20      2                    39 vs 44
    21    142                 2,470 vs 2,428    (should be 2758!)
    22   1853                35,986 vs 34,548
    23   8996               174,797 vs 172,512
    24  17656               342,888 vs 342,335
    25  15306               297,250 vs 297,838
    26   6127               118,989 vs 122,116
    27   1257                24,412 vs 25,315
    28    149                 2,894 vs 2,686
    29      4                    78 vs 168
    30      0                     0  vs 10
```

Denis,
Thanks for looking into this. As far as a bias/discrepancy is concerned, I have spent some time scratching my head. The grid generator was tested by **RedEd** some years ago and found satisfactory (I can no longer find the post). I think that a minor difference, possibly sufficient to cause the discrepancy, is in the way cells are removed. In an attempt to make my program more efficient, I make two passes over the clues, in the first removing those shown to be redundant using simple logic, and in the second removing those shown to be redundant using backtracking. In order to make a comparison, I have now generated a second file where removal is performed in a single pass. The summary for the same number of puzzles is:

Code:

```
Number:       51492
Average:   24.40608

    20      3    ->          58
    21    139               2699
    22   1663              32291
    23   8867             172201
    24  17309             336149
    25  15409             299250
    26   6395             124194
    27   1508              29032
    28    187               3632
    29     11                214
    30      1                 19
```

showing a small but clear difference. If you wish, I can send you this second file for you to analyse further.

Have we learnt something?

Regards,

Mike Metcalf

Back to top   [profile] [pm]

denis_berthier   Posted: Fri Jun 26, 2009 4:50 pm   Post subject: [quote] [edit]

m_b_metcalf wrote:

As far as a bias/discrepancy is concerned, I have spent some time scratching my head. The grid generator was tested by **RedEd** some years ago and found satisfactory (I can no longer find the post).

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

The only kind of test of unbiasedness for complete grids I heard of is the one referred to by Coloin in a previous post, related to U4s (http://www.sudoku.com/boards/viewtopic.php?t=5128&start=0).
But a statistical test is only a test. If negative, then there is a bias wrt to this test, but there may be no bias wrt to another test; if positive, then there's no bias wrt to this test, but there may be a bias wrt to another test.

It is like the Tower of Pisa. Depending on which side you're looking, either it leans or it doesn't. Except that it is much more complex: in our case, there can be hundreds of different angles of view.
Moreover, it is very unclear how passing or not the U4 test on complete grids can have any impact on a possible bias of minimal puzzles generated from them.

> **m_b_metcalf wrote:**
> I think that a minor difference, possibly sufficient to cause the discrepancy, is in the way cells are removed. In an attempt to make my program more efficient, I make two passes over the clues, in the first removing those shown to be redundant using simple logic, and in the second removing those shown to be redundant using backtracking. In order to make a comparison, I have now generated a second file where removal is performed in a single pass.

I guess there were lots of cases for which the first pass was enough (Could you check this ?). That'd explain why, in your first file, there were so many puzzles that could be solved with singles. Perhaps, it could explain also that there were so few hard puzzles (first phase "almost enough" most of the time?).
I think it is essential that the removal steps be completely random.

> **m_b_metcalf wrote:**
> If you wish, I can send you this second file for you to analyse further.

Yes, I'd like to do that. You can use the same mail.

> **m_b_metcalf wrote:**
> Have we learnt something?

Not very clear yet. But if the results with the new file are different, then we'll have learnt something.

Regards

Back to top   [profile] [pm] [www]

---

coloin   ▢ Posted: Fri Jun 26, 2009 8:13 pm   Post subject: [quote]

Very intriqing.......and I pretty much agree with all you comments.

Joined: 06 May 2005
Posts: 986
Location: Oxford

I was surprized as any with the stats on **m_b_m'**s puzzle generator.

As far as the solution grid is concerned - I feel that is not much important. The distribution of puzzles from a single grid behave similarly.
Perhaps this could be confirmed. I am continuing to generate *all* the puzzles from the 40-clue subgrid. By random generation of 2M puzzles I have over 500000 puzzles ......but there probably are many more.

Code:
```
c:\Suxx>clusta file1.txt
lines:517883 average clues:25.028385
   21   45     [k]
   22   1258 [?k]
   23   22473
   24   127112
   25   211041
   26   123824
   27   29565
   28   2502
   29   63              [k=constant, all puzzles prob found]
```

targetting some of the larger puzzles......

Code:
```
c:\Suxx>clusta file2.txt
lines:52969 average clues:26.128849
   24   1144
   25   10662
   26   24009
   27   14641
   28   2405
   29   108
```

Most of the small puzzles have been found, but the larger ones are more difficult. An example of this is that a 30-clue puzzle was not generated randomly. ....but one has come out using gsfs program - which generates to completeness [sudoku -m -qFN -f%v file.txt] and this is still running, currently the stats are.

Code:

```
c:\Suxx>clusta file3.txt
lines:430886 average clues:25.167914
    21   10
    22   680
    23   14118
    24   89594
    25   173371
    26   117494
    27   32196
    28   3317
    29   105
    30   1
```

If this 40 clue subgrid is behaving itself....this may well be better reflection of the average puzzle size. But of course this is an assumption.

gsf also has a puzzle generator which could be analysed. Im not sure of the function which tries to generate randomly............

C

Back to top   [profile] [pm]

| **denis_berthier** | Posted: Sat Jun 27, 2009 5:50 am   Post subject: | [quote] [edit] |

Thinking again of all this: it should be easy to check if the U4-test is meaningful wrt to the complexity of puzzles (measured as their NRCZT or SER levels). We just have to compute the correlation coefficient between U4 and NRCZT (or SER). 10,000 puzzles should be enough.

Joined: 19 Jun 2007
Posts: 601
Location: Paris, France

The only thing I'm missing for this is a (preferably fast) program that computes the U4 number of a (list of) complete grid(s). Do you known if any such program is available?

Back to top   [profile] [pm] [www]

| **m_b_metcalf** | Posted: Sat Jun 27, 2009 10:21 am   Post subject: | [quote] |

> **denis_berthier wrote:**
> I think it is essential that the removal steps be completely random.

Joined: 15 May 2006
Posts: 2106
Location: Berlin

This is clealy true, but begs the question as to what happens for an obviously biased sample. To this end, I have produced a third file in which the clue removal is again carried out in one pass, but systematically starting in r1c1 and working through to r9c9. This depletes preferentially row 1, column 1 and box 1. A 'typical' puzzle looks like this:

Code:

```
. . . . . . . . .
. . . . . . . . 4
. . . . . 8 . 7 5
. . 1 . 7 . . . .
. 2 . . . 1 . 4 .
. . 4 . . 5 2 9 .
. 3 . . . 4 . 1 .
. 4 . . 8 . . 6 7
. 9 5 . 1 7 . 2 8   SE=7.2
```

The result summary, for the same number, 51492, of puzzles is:

Code:

```
Number:        51492
```

```
   Average:    25.00264

       20       0
       21      37
       22     682
       23    4318
       24   12466
       25   17015
       26   11648
       27    4309
       28     905
       29     109
       30       3
```

where we see immediately the large increase in the average number of clues.

> **denis_berthier wrote:**
>
>> **m_b_metcalf wrote:**
>>
>> If you wish, I can send you this second file for you to analyse further.
>
> Yes, I'd like to do that. You can use the same mail.

Done.

Regards,

Mike Metcalf

Back to top   profile   pm

Display posts from previous: All Posts ⬍   Oldest First ⬍   Go

newtopic   postreply   **Sudoku Players' Forums Forum Index ->**   All times are GMT
                            Advanced solving techniques   **Goto page** Previous 1, 2, 3 ... , 10, 11, 12 Next
Page 11 of 12

Stop watching this topic

                                         Jump to: Advanced solving techniques   ⬍   Go

                                              You **can** post new topics in this forum
                                              You **can** reply to topics in this forum
                                              You **can** edit your posts in this forum
                                              You **can** delete your posts in this forum
                                              You **can** vote in polls in this forum