



Sudoku Players' Forums

[FAQ](#) [Search](#) [Memberlist](#) [Usergroups](#)

[Profile](#) [You have no new messages](#) [Log out \[denis_berthier \]](#)

THE REAL DISTRIBUTION OF MINIMAL PUZZLES

Goto page [Previous](#) [1](#), [2](#), [3](#) ... , [33](#), [34](#), [35](#) [Next](#)

[newtopic](#)

[postreply](#)

[Sudoku Players' Forums Forum Index -> General/puzzle](#)

[View previous topic :: View next topic](#)

Author

Message

eleven

Posted: Fri Oct 09, 2009 5:04 pm Post subject:

[quote](#)

Joined: 10 Feb 2008
Posts: 529

denis_berthier wrote:

eleven,
Thanks for these corrections.
Did you test their impact on speed?

No, but that the U4 test is slightly slower should well be compensated by finding more U4's, so i guess it becomes a bit faster.

[Back to top](#)

[profile](#) [pm](#)

Red Ed

Posted: Sat Oct 10, 2009 7:49 am Post subject:

[quote](#)

Joined: 06 Jun 2005
Posts: 776

Initial indications are that [the supersets method](#) might be able to do one 24, and (mostly implicitly) all its 31-clue supersets, every five seconds or so. If debugging and testing bear that out, it would be good news.

[Back to top](#)

[profile](#) [pm](#)

denis_berthier

Posted: Sat Oct 10, 2009 11:32 am Post subject:

[quote](#) [edit](#)

Joined: 19 Jun 2007
Posts: 911
Location: Paris, France

EXTENDED RESULTS WITH THE CONTROLLED-BIAS GENERATOR, IN WHICH GSF'S LIST IS TAKEN AS THE SOURCE OF COMPLETE GRIDS

The controlled-bias generator has now accomplished 31 full scans of the exhaustive gsf's collection of (isomorphism classes of) complete grids, resulting in a total of 658,850 controlled-bias minimal puzzles.

Here are the results

Code:

```
#clues #instances % unbiased %
```

```

          in cb sample  in cb sample  (estimated)
19         0             0.0           0.0
20         0             0.0           0.0
21        22            0.0033         3.80e-05
22       698            0.106          0.0033
23     12319            1.87           0.149
24     78337            11.89          2.289
25    201464            30.58          13.42
26    222854            33.82          31.98
27    112136            17.02          32.77
28    27320             4.15           15.40
29    3472              0.53            3.58
30     224              0.034            0.40
31      4              0.00061          0.012

controlled-bias mean= 25.666
controlled-bias standard-deviation= 1.116

unbiased mean= 26.575
unbiased standard-deviation= 1.114

```

For ease of comparison, I recall below the results when the generator part is the suexg internal generator instead of gsf's collection (sample of 500,000 puzzles):

Code:

```

#clues  #instances      %          unbiased %
         in cb sample  in cb sample  (estimated)
19         0             0.0           0.0
20         2             0.0004         1.58e-06
21         4             0.0008         9.20e-06
22        615            0.123          0.0039
23       9848            1.9696          0.159
24      60576            12.12           2.356
25     154024            30.80           13.66
26     168070            33.61           32.10
27     83911             16.78           32.65
28     20234             4.05            15.18
29     2566              0.51            3.52
30      147              0.024            0.35
31       3              0.0006          0.012

controlled-bias mean= 25.65
controlled-bias standard-deviation= 1.117

unbiased mean= 26.56
unbiased standard-deviation= 1.113

```

In both cases, we have the same stats for the SER and NRCZT:

SER:

controlled-bias mean = 4.15 unbiased mean= 4.48

controlled-bias standard-deviation = 2.49 unbiased standard-deviation = 2.53

NRCZT:

controlled-bias mean= 2.14 unbiased mean = 2.31

controlled-bias standard-deviation = 1.34 unbiased standard-deviation = 1.38

Red Ed: notice that, in both cases, there are more 27s than 26s in the estimated unbiased distribution - and the difference seems to be clearly beyond sampling errors.

[Back to top](#)

[profile](#) [pm](#) [www](#)

Red Ed

Posted: Sat Oct 10, 2009 11:40 am Post subject:

[quote](#)

Joined: 06 Jun 2005
Posts: 776

Curious result re 26 vs. 27. I'll do a significance test some time (I assume that you won't).

What source of randomness are you using to select which clue to remove next in the deletion phase?

[Back to top](#)

[profile](#) [pm](#)

denis_berthier

Posted: Sat Oct 10, 2009 12:00 pm Post subject:

[quote](#) [edit](#)

Joined: 19 Jun 2007
Posts: 911
Location: Paris, France

Red Ed wrote:

Curious result re 26 vs. 27.

Why curious?

As the distribution is very close to normal with mean 26.575, it is on the contrary very consistent that 27, which is a little closer to the mean than 26, has a probability a little larger.

Red Ed wrote:

What source of randomness are you using to select which clue to remove next in the deletion phase?

How do you imagine this could have any impact on the ratio 27s/26s?

[Back to top](#)

[profile](#) [pm](#) [www](#)

Red Ed

Posted: Sat Oct 10, 2009 12:01 pm Post subject:

[quote](#)

Joined: 06 Jun 2005
Posts: 776

You first, Denis. Please answer the question.

[Back to top](#)

[profile](#) [pm](#)

Red Ed

Posted: Sat Oct 10, 2009 12:15 pm Post subject:

[quote](#)

Joined: 06 Jun 2005
Posts: 776

Here's an analysis of the difference between our results for 26s and 27s.

26s first:

- 222854 minimals => std err \sim 1/472
- Scale up by choose(81,26) / (31 x 5472730538) to give number of proper minimal 26s per grid
- => you estimate 1.4872e15 w. std err 0.0031e15
- I had estimated 1.4847e15 w. std err 0.0088e15
- So **we are in agreement on the total number of 26s**

Now 27s:

- 112136 minimals => std err \sim 1/335
- Scale up by choose(81,27) / (31 x 5472730538) to give number of proper minimal 27s per grid
- => you estimate 1.5244e15 w. std err 0.0046e15
- I had estimated 1.4364e15 w. std err 0.0433e15
- So **our agreement or not on the number of 27s is inconclusive**

The issue appears to be that I expended much less energy on the 27s than you did, so my estimate of the number of 27s was quite uncertain. We're 2 standard deviations apart, which is pretty serious but perhaps not conclusive. So I'm re-running my 27s estimator now to collect enough data to bring down the std err and make comparison easier. Oh, for a quad-core machine ...

~~~~~

### Now for some more analysis - this time comparing the proportions distributions ...

Let's apply my [top-down sensitivity analysis](#) to Denis' results posted above. We can work directly with the "#instances in cb sample" values, since the extension to number-of-minimals-in-a-grid involves only constant scale factors.

If source 'A' is gsf's collection of grids and source 'B' is *suexg* then:

#### Code:

|       | Source 'A' |        | Source 'B' |        |
|-------|------------|--------|------------|--------|
| Clues | Mean       | StdDev | Mean       | StdDev |
| 24    | 78337      | 279.89 | 60576      | 246.12 |
| 26    | 222854     | 472.07 | 168070     | 409.96 |

Then using 'R' ...

#### Code:

```
n=1e6
a24=rnorm(n, 78337,279.89)
a26=rnorm(n,222854,472.07)
b24=rnorm(n, 60576,246.12)
b26=rnorm(n,168070,409.96)
sum(a24/a26 < b24/b26)*100/n
```

... the answer is consistently about 99.996%

In other words: the data collected in Denis' experiments show, with  $\sim$ 99.996% confidence, that *suexg*'s estimate of the relative proportion of 24-clue vs. 26-

clue minimals places more weight on 24-clue minimals than is the case for gsf's complete collection of solution grid classes.

This demonstrates again that the solution grid source affects the proportions distribution.

It's been noted previously that the difference is small. To quantify that, let's just observe that the ratios of 26s to 24s in Denis' data were **13.97** (A) vs. **13.62** (B). [EDIT: and, as Denis notes, this is a relative difference of only 2.5% or so.]

Last edited by Red Ed on Sat Oct 10, 2009 8:29 pm; edited 1 time in total

[Back to top](#)

 [profile](#)  [pm](#)

**denis\_berthier**

☐ Posted: Sat Oct 10, 2009 2:24 pm Post subject:

 [quote](#)  [edit](#)

Joined: 19 Jun 2007  
Posts: 911  
Location: Paris, France

**Red Ed wrote:**

To quantify that, let's just observe that the ratios of 26s to 24s in Denis' data were 13.97 (A) vs. 13.62 (B).

To really quantify that you'd better observe that **13.62 is only 2.5 % smaller than 13.97**.

So, the conclusion one should draw is rather that, in the lower part of the distribution, there is a very small (2.5%) difference between the 2 samples.

Last edited by denis\_berthier on Sat Oct 10, 2009 7:59 pm; edited 1 time in total

[Back to top](#)

 [profile](#)  [pm](#)  [www](#)

**Red Ed**

☐ Posted: Sat Oct 10, 2009 3:51 pm Post subject:

 [quote](#)

[Deleted]

Joined: 06 Jun 2005  
Posts: 776

Last edited by Red Ed on Sat Oct 10, 2009 8:09 pm; edited 1 time in total

[Back to top](#)

 [profile](#)  [pm](#)

**denis\_berthier**

☐ Posted: Sat Oct 10, 2009 3:58 pm Post subject:

 [quote](#)  [edit](#)

Joined: 19 Jun 2007  
Posts: 911  
Location: Paris, France

**Red Ed wrote:**

Now 27s:

- 112136 minimals => std err  $\sim$  1/335
- Scale up by choose(81,27) / (31 x 5472730538) to give number of proper minimal 27s per grid
- => you estimate 1.5244e15 w. std err 0.0046e15
- I had estimated 1.4364e15 w. std err 0.0433e15
- So **our agreement or not on the number of 27s is inconclusive**

...  
We're 2 standard deviations apart, which is pretty serious but perhaps not conclusive.

In the previous post, you drew conclusions from small differences. But here, you refuse to draw conclusions from very large ones.

We are not 2 standard deviations apart. The value you get from your small sample is exactly 19.13 standard deviations apart from what I get from my very large one (taking your computations as such).

If that's not conclusive, especially for a value (27) very close to the mean, what will be conclusive?

And even if there were only 2 standard deviations, as you compute it, the probability would be of magnitude  $\sim 1\%$ .

Last edited by denis\_berthier on Sat Oct 10, 2009 8:02 pm; edited 3 times in total

[Back to top](#)

[profile](#) [pm](#) [www](#)

**Red Ed**

Posted: Sat Oct 10, 2009 5:02 pm Post subject:

[quote](#)

Joined: 06 Jun 2005  
Posts: 776

The significance is lessened due to my experiments being small and my results having large uncertainty, so it really is closer to -2 sigma than -19 sigma (calculation:  $\sqrt{sd\_me^2 + sd\_you^2}$ ). The p-value, i.e. likelihood of a difference this big happening at random, is  $\sim 2.2\%$ . That's worryingly small when we should expect our two methods to give the same answer -- hence me re-running my experiment for longer to get a better estimate.

Last edited by Red Ed on Sat Oct 10, 2009 8:15 pm; edited 1 time in total

[Back to top](#)

[profile](#) [pm](#)

**denis\_berthier**

Posted: Sat Oct 10, 2009 5:33 pm Post subject:

[quote](#)

[edit](#)

[Deleted]

Joined: 19 Jun 2007  
Posts: 911  
Location: Paris, France

Last edited by denis\_berthier on Sat Oct 10, 2009 8:03 pm; edited 1 time in total

[Back to top](#)

[profile](#) [pm](#) [www](#)

**Red Ed**

Posted: Sat Oct 10, 2009 5:43 pm Post subject:

[quote](#)

[Deleted]

Joined: 06 Jun 2005  
Posts: 776

Last edited by Red Ed on Sat Oct 10, 2009 8:08 pm; edited 1 time in total

[Back to top](#)

[profile](#) [pm](#)

**denis\_berthier**

Posted: Sat Oct 10, 2009 6:31 pm Post subject:

[quote](#)

[edit](#)

[Deleted]

Joined: 19 Jun 2007  
 Posts: 911  
 Location: Paris, France

Last edited by denis\_berthier on Sat Oct 10, 2009 8:04 pm; edited 1 time in total

[Back to top](#)

 [profile](#)  [pm](#)  [www](#)

**eleven**

📅 Posted: Sat Oct 10, 2009 7:24 pm    Post subject:

 [quote](#)**Red Ed wrote:**

Initial indications are that [the supersets method](#) might be able to do one 24, and (mostly implicitly) all its 31-clue supersets, every five seconds or so. If debugging and testing bear that out, it would be good news.

Joined: 10 Feb 2008  
 Posts: 529

Using a singles solver only i could check about 33% very quickly: In almost 10% of random 24's one of the clues is forced by the others through singles. And in 23% more than 6 singles are forced by the givens.

Also the about 29% puzzles, where 4-6 singles are forced, could be checked brute force rather quickly.

But the rest i would have to leave to Red Ed 😊

[Back to top](#)

 [profile](#)  [pm](#)

Display posts from previous:

 [newtopic](#) [postreply](#)

**Sudoku Players'**  
**Forums Forum**  
 Index ->  
**General/puzzle**

All times are GMT  
 Goto page [Previous](#) [1](#), [2](#), [3](#) ... , [33](#), [34](#), [35](#) [Next](#)

**Page 34 of 35**[Stop watching this topic](#)

Jump to:

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

Powered by phpBB © 2001, 2005 phpBB Group