Sudoku Players' Forums

THE REAL DISTRIBUTION OF MINIMAL PUZZLES

Goto page Previous 1, 2, 3 ... , 32, 33, 34 Next

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

View previous topic :: View next topic

Author	Message
denis_berthier Joined: 19 Jun 2007 Posts: 905 Location: Paris, France	Posted: Thu Oct 08, 2009 2:59 am Post subject: quote **JPF wrote:** **denis_berthier wrote:** #clues #minimals 22 1.35529578042937e+33 23 5.5671277334354e+34 24 8.27561904095325e+35 25 4.79759013393657e+36 26 1.12755990408589e+37 27 1.14674603199864e+37 28 5.33292926612237e+36 29 1.23600037208473e+36 30 1.22733006549957e+35 31 4.1207263555285e+33 (*) * values based on a small sub-sample are not reliable Whoa, 15 significant digits ... What does it mean ? JPF Just that I was lazy and I took the output of my computer without proper formatting I have no idea of the precision. This also answers Red Ed's question. Back to top profile pm www
coloin Joined: 05 May 2005 Posts: 1080 Location: Devon UK	Posted: Thu Oct 08, 2009 6:45 am Post subject: quote Red Ed.....are you confining the 7 clues to within the solution grid ? if so.....a rather less subtle method than yours might be... http://magictour.free.fr/suexmu35.exe from dukuso can do a complete [minimal] +5 [within a solution grid] in minutes command suexmu35 <file.txt> 5 1

the file contains 2 81 strings
12121222111111111212222012111112221212
123456789solutiongrid.....

where 2 is a given and 1 a possible. 0 would be a clue already implied.

a + 7 would take too long - although if a few uncovered unavoidable sets were identified and coded i am sure it could be speeded up

please explain how the mean number of minimal 31s found this way helps...

Back to top profile pm

eleven

Posted: Thu Oct 08, 2009 8:13 am Post subject: quote

Joined: 10 Feb 2008
Posts: 528

I guess its the same principle like here.

[Edit:better to start with the 31 clue]
A minimal 31 clue has [corrected, thanks to Red Ed]choose(31,7) minimal (multisolution) 24 clues (in a fixed grid). The number of 24's=choose(81,24) per grid. If you get X minimal 31's from N random 24 clues, then the number of minimal 31's should be about
X * (choose(81,24)/N) / choose(31,7)
or something similar

Last edited by eleven on Thu Oct 08, 2009 10:16 am; edited 1 time in total

Back to top profile pm

Red Ed

Posted: Thu Oct 08, 2009 9:33 am Post subject: quote

Joined: 06 Jun 2005
Posts: 768

@coloin - ah-ha! - I hadn't spotted that program. Nice.

@eleven - if X/N is the mean number of proper minimal 31s per 24-clue subgrid then X/N * choose(81,24)/choose(31,24) is the mean number of proper minimal 31s per grid. And yes, it's the same principle. I'm hopeful that bottom-up will work better than top-down when searching for 31s. Going by Denis' figure for the number of 31s (accepting that it is a low-confidence estimate), there should be on average ~1/150 proper minimal 31s per 24.

@denis:

> **denis_berthier wrote:**
>
> I have no idea of the precision. This also answers Red Ed's question.

If those results are for N trials then you can regard each number of minimals found as a Bin(N,m[c]/N) random variable, where m[c] is the observed number of minimals. So the variance is m[c] x (1-m[c]/N), which is close to m[c]. So the standard deviation is a proportion ~sqrt(m[c]) of the observed number of minimals. Example: if you found 68 controlled-bias 30-clue minimals then the unbiased estimator for the number of minimals has standard deviation 8.2 or thereabouts. Maybe you'd like to update your web page: stats always look better with quantified confidence.

Back to top profile pm

eleven

Joined: 10 Feb 2008
Posts: 528

Posted: Thu Oct 08, 2009 10:15 am Post subject: quote

> **Red Ed wrote:**
>
> > choose(31,24)

Of course, thanks, i was thinking the wrong way around, should be (choose(31,7) in my post.

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 768

Posted: Thu Oct 08, 2009 11:59 am Post subject: quote

> **eleven wrote:**
>
> > **Red Ed wrote:**
> >
> > > First, do some preprocessing to check that the subgrid is minimal (i.e. no clue is implied by the others) ...
> >
> > I would not know how to do this quickly. suexk needed 53 sec to calculate the numbers of solutions for 240 multisolution 24's on my PC, so one test would need about 5 secs.

I can check something like ten 24-clue subgrids/second for minimality using template-based processing.

Back to top profile pm

eleven

Joined: 10 Feb 2008
Posts: 528

Posted: Thu Oct 08, 2009 1:48 pm Post subject: quote

Interesting, do you have a link for that "template-based processing" ?

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 768

Posted: Thu Oct 08, 2009 2:21 pm Post subject: quote

I'll post code in a few days. Bed time now!

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 905
Location: Paris, France

Posted: Thu Oct 08, 2009 9:37 pm Post subject: quote

> **Red Ed wrote:**
>
> > **denis_berthier wrote:**
> >
> > > I have no idea of the precision. This also answers Red Ed's question.
> >
> > If those results are for N trials then you can regard each number of minimals found as a Bin(N,m[c]/N) random variable, where m[c] is the observed number of minimals. So the variance is m[c] x (1-m[c]/N), which is close to m[c]. So the standard deviation is a proportion ~sqrt(m[c]) of

> the observed number of minimals. Example: if you found 68 controlled-bias 30-clue minimals then the unbiased estimator for the number of minimals has standard deviation 8.2 or thereabouts. Maybe you'd like to update your web page: stats always look better with quantified confidence.

I was speaking of the absolute number of n-clue minimals, which isn't really a topic of interest for me and isn't even mentioned on my web page. Recently, you said you had your own estimates for these numbers. How do they compare with mine?

As for my web pages, your remark would be valid (and I'd agree with the above results) if they mentioned an estimate of the suexg-cb number-of-clue distribution, but they only mention the numbers of n-clue instances in the suexg-cb sample, which are raw experimental data.
If I wanted to give a confidence interval for the unbiased distribution, the binomial model wouldn't apply and a more complex computation would be required to get the standard deviations from those of the cb-distribution. It isn't impossible to get an estimate of the *precision* (using ydx-xdy/y^2) - not the standard deviation - but that's a matter of priority and I'm currently much more interested in checking the insensitivity to the source of complete grids.
Monitoring the various runs of programs for obtaining sufficiently many minimals with different sources of grids is very time consuming, not very intellectually stimulating, but I can see no other way of getting better estimates and of "proving" this insensitivity.

Back to top profile pm www

Red Ed

Joined: 06 Jun 2005
Posts: 768

Posted: Thu Oct 08, 2009 10:48 pm Post subject: quote

Re comparison: <here>

Re Binomial: that *is* the right model, although my use of it above is a tad informal. The formal treatment is unpleasant: <here>

My analysis is applied to the controlled-bias counts; then for the absolute number of minimals you just multiple up by the usual scale factor.

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 905
Location: Paris, France

Posted: Thu Oct 08, 2009 11:20 pm Post subject: quote

To be clear:
Binomial is (obviously) the right model for the n-clue puzzles (n fixed) produced by the controlled-bias generator.
It can't be the right model for the unbiased minimals for the very simple reason that, as there isn't even a random process defined to produce unbiased random samples of minimals, no model of this inexistent process can be defined.

> **Red Ed wrote:**
> The analysis is applied to the controlled-bias counts;

So, OK on this. But

> **Red Ed wrote:**

> then for the absolute number of minimals you just multiple up by the usual scale factor.

I don't know what you call "usual scale factors".
Nothing was usual before I defined the controlled-bias generator and introduced the "correction coefficients".
If what you call "usual scale factors" are these correction coefficients, just multiplying by them (after normalisation) could only give a crude estimate.

Back to top profile pm www

eleven

Posted: Fri Oct 09, 2009 2:25 am Post subject: quote

Joined: 10 Feb 2008
Posts: 528

> **denis_berthier wrote:**
>
> Then everything seems alright: as it is only there for optimisation, it isn't really important if it misses a few U4, is it?

I found the time now to test it better and indeed there is a bug, which let it miss about 4-5% (i was too greedy to save some time).
The function find4unavoid() should be:

Code:

```
void find4unavoid(int i, int j)
{
  int k,l;
  k=1+9*((i-1)%9)+(i-1)/9;l=1+9*((j-1)%9)+(j-1)/9;
  if (k == 22 && l == 58)
    printf("22/58, %d, %d, %d, %d\n", A[k+1], A[l+2], A[l+1],
A[k+2]);
  if (A[i]==A[j+9]&&A[j]==A[i+9])add4unavoid(i,j,i+9,j+9);
  if (A[i]==A[j+18]&&A[j]==A[i+18])add4unavoid(i,j,i+18,j+18);
  if
(A[i+9]==A[j+18]&&A[j+9]==A[i+18])add4unavoid(i+9,j+9,i+18,j+18);
  if (A[k]==A[l+1]&&A[l]==A[k+1])add4unavoid(k,l,k+1,l+1);
  if (A[k]==A[l+2]&&A[l]==A[k+2])add4unavoid(k,l,k+2,l+2);
  if
(A[k+1]==A[l+2]&&A[l+1]==A[k+2])add4unavoid(k+1,l+1,k+2,l+2);
}
```

I also corrected it in the original post with the code here

Back to top profile pm

denis_berthier

Posted: Fri Oct 09, 2009 4:40 am Post subject: quote

Joined: 19 Jun 2007
Posts: 905
Location: Paris, France

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

Back to top profile pm www

denis_berthier

Posted: Fri Oct 09, 2009 5:02 am Post subject: quote

Joined: 19 Jun 2007
Posts: 905
Location: Paris, France

FIRST RESULTS WITH THE CONTROLLED-BIAS GENERATOR ON THE FULL GSF COLLECTION OF COMPLETE GRIDS

The controlled bias generator (version suexg-cb-optim48-U4) has now completed 5

full scans of the gsf collection of equivalence classes of solution grids.
Here are the results.

Code:

```
#clues  #instances  raw %      unbiased %
        in sample
19      0           0.0        0.0
20      0           0.0        0.0
21      1           0.00094    1.07e-05 (*)
22      108         0.101      0.0031
23      1977        1.855      0.148
24      12710       11.93      2.30
25      32690       30.67      13.48
26      35840       33.63      31.8418652278533
27      18309       17.18      33.14
28      4351        4.08       15.19
29      555         0.52       3.54026692334375
30      31          0.029      0.34 (*)
31      1           0.00094    0.018 (*)

raw mean= 25.665
raw standard-deviation= 1.114

unbiased mean= 26.57
unbiased standard-deviation= 1.11
```

SER:
unbiased-average = 4.48
unbiased-standard-deviation = 2.527

NRCZT:
unbiased-average = 2.31
unbiased-standard-deviation = 1.38

As can be checked, this is very close to the results obtained when suexg itself is used as the grid generator instead of gsf.
Additional scans are running and more precise results expected for next week.

Back to top

profile pm www

Red Ed

Joined: 06 Jun 2005
Posts: 768

Posted: Fri Oct 09, 2009 8:59 am Post subject: quote

denis_berthier wrote:

> **Red Ed wrote:**
>
> > then for the absolute number of minimals you just multiple up by the usual scale factor.
>
> I don't know what you call "usual scale factors".
> ...(elided)...
> If what you call "usual scale factors" are these correction coefficients, just multiplying by them (after normalisation) could only give a crude estimate.

If m[c] out of N trials yield c-clue minimals then the estimated average number of minimals per grid is m[c] x choose(81,c)/N. If we know the uncertainty in m[c] then by multiplying by choose(81,c)/N we know the uncertainty in the estimated average

number of minimals. That's all.

Back to top profile pm

Display posts from previous: All Posts Oldest First Go

newtopic postreply Sudoku Players' Forums Forum Index -> General/puzzle

All times are GMT - 8 Hours

Goto page Previous 1, 2, 3 ... , 32, 33, 34 Next

Page 33 of 34

Jump to: General/puzzle Go