Sudoku Players' Forums

THE REAL DISTRIBUTION OF MINIMAL PUZZLES

Goto page Previous 1, 2, 3 ... , 26, 27, 28 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: 874 Location: Paris, France	Posted: Thu Oct 01, 2009 12:21 pm Post subject: quote edit **eleven**, I tried your last modification and it works perfectly. ./sudoku -q- -f%v 002.sudz | ./suexg-cb-count-optim-stream.exe 0 1 - Here are the minimals produced, with the numbers of grids consumed: 12..5.....67....3.8.1.....1...8.5..6..948....5.6....5..8.26.....9352.....1...8 430573 .23...7.945.7..12...9.....5.1....6..36...1...9.8.....79....6...51.3.....63.2.. 47401 .2..5.7..456.....37.9....6...43.....6.5.7.....2..5.1...5.78.6.....29.....1.... 30316 .234...8.4..7....3..9.2...5....9..7....67.2..6.7.1...4...9..6.8..1.6..9...2..... 42183 12....78....789.....9.....2145.....3.....5...5.1.34....9..2.46.2....1.....7..3. 300683 We now have all the bricks to use gsf files as input to the controlled-bias generator. **gsf**, After the above, I tried to use the first 5 sudz files in the same run. I used cat 001.sudz 002.sudz 003.sudz 004.sudz 005.sudz > all.sudz (I know it can't be the smarter way of doing it, as it doubles the occupied disk space, but I don't know yet how to introduce a list of files instead of a single one in the command line) Replacing 002.sudz by all.sudz in the first command, I didn't get more minimals. Is there anything in the sudz file format that forbids to append them?
Back to top	profile pm www
eleven Joined: 10 Feb 2008 Posts: 506	Posted: Thu Oct 01, 2009 1:06 pm Post subject: quote Hm, either i have a serious bug or its true, that in the 133302 grids of band 299 you can find 2 times more minimals than in the 1007170 grids of band 1 (?!) [Edit:]The difference is not that big (i had an error in the command lines), but with about 10mio grids each the average number of grids needed in band 1 was 379342, and in band 299 93986. This has to be explained, before i can trust, that the program is correct.

Denis, i never had a MAC, but you can try to adopt gsf's script to run one band after the other.

Back to top profile pm

Pat

Posted: Thu Oct 01, 2009 2:45 pm Post subject: quote

Joined: 18 Jul 2005
Posts: 1550

> **denis_berthier wrote:**
>
> I don't know yet how to introduce a list of files instead of a single one in the command line

my understanding is that **gsf**'s software accepts any number of filenames on the command-line

i.e. this should work --

Code:

```
./sudoku -q- -f%v 001.sudz 002.sudz 003.sudz
004.sudz 005.sudz
```

Back to top profile pm

Allan Barker

Posted: Thu Oct 01, 2009 2:48 pm Post subject: quote

Joined: 21 Feb 2008
Posts: 346
Location: Bangkok

> **denis_berthier wrote:**
>
> As the controlled-bias generator consumes more than 200,000 complete grids per puzzle generated, 1,000,000 grids is far from enough. I fear we need the code, or at least an executable that can output a stream of complete grids.

Denis,

I have placed the MonteCarlo grid generator here
http://sudokuone.com/xsudo1/data.htm

The generator is in **montecar.c**.

It will compile with any generic gcc installation. There is a Windows batch file **mc.bat** with an example compile command line.

When you link to your other program just include the file **montecar.h** and comment out **main()** in montecar.c.

To use when linked in to your program:
1. Call run_montecarlo() to generate one grid.
2. Read the grid from SUD[9][9] as integers 1 through 9.

Back to top profile pm www

gsf

Posted: Thu Oct 01, 2009 2:55 pm Post subject: quote

denis_berthier wrote:

Joined: 22 Sep 2005
Posts: 3877
Location: NJ USA

> **eleven**,
> I tried your last modification and it works perfectly.
> ./sudoku -q- -f%v 002.sudz | ./suexg-cb-count-optim-stream.exe 0 1 -
>
> We now have all the bricks to use gsf files as input to the controlled-bias generator.
>
> **gsf**,
> After the above, I tried to use the first 5 sudz files in the same run.
> I used cat 001.sudz 002.sudz 003.sudz 004.sudz 005.sudz > all.sudz
> (I know it can't be the smarter way of doing it, as it doubles the occupied disk space, but I don't know yet how to introduce a list of files instead of a single one in the command line)
> Replacing 002.sudz by all.sudz in the first command, I didn't get more minimals. Is there anything in the sudz file format that forbids to append them?

good to see things moving along
the sudz files cannot be catenated
but you can specify more than one sudz file on the sudoku command line

Code:

```
./sudoku -q- -f%v *.sudz | ./suexg-cb-count-optim-stream.exe
```

Back to top profile pm www

gsf

Posted: Thu Oct 01, 2009 3:01 pm Post subject: quote

eleven wrote:

Joined: 22 Sep 2005
Posts: 3877
Location: NJ USA

> Hm, either i have a serious bug or its true, that in the 133302 grids of band 299 you can find 2 times more minimals than in the 1007170 grids of band 1 (?!) [Edit:]The difference is not that big (i had an error in the command lines), but with about 10mio grids each the average number of grids needed in band 1 was 379342, and in band 299 93986. This has to be explained, before i can trust, that the program is correct.

limiting observations to one band will most likely introduce noticeable bias
because all grids in the same band have the same exact top band
in addition, because of minlex ordering, all other bands in those grids have the property that

Code:

```
indexof(top-band) <= indexof(any-other-band)
```

this includes bands attained by swapping rows in a band and bands in the transpose

so, as the top band index increases, the heterogeneity of the other bands within a grid decreases

Back to top profile pm www

eleven

Posted: Thu Oct 01, 2009 3:26 pm Post subject: quote

Joined: 10 Feb 2008
Posts: 506

Thanks for the explanation, gsf.

So i will not look for a bug in the program now. Its still surprising for me that the grids from one set could have almost 4 times more minimals than those from the other.I did not expect a bigger difference between 2 (non automorphic) grids than say a factor 2.

Back to top profile pm

denis_berthier

Posted: Thu Oct 01, 2009 4:54 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 874
Location: Paris, France

Pat, gsf,
Thanks to your command lines, I have written an alias (for bash):
alias MIN="./sudoku -q- -f%v *.sudz | ./suexg-cb-count-optim46-stream.exe"
that will allow to launch the computation for the whole set of grids (when it is generated, in 2 weeks):
MIN seed 1 -
where "seed" is the seed for MWC
One could probably be a little smarter and write an alias that could be called simply as "MIN seed" (as 1 and - are fixed, there's no point in keeping them as arguments), but that's not blocking.

As there are now many modifications to suexg (all implemented by **eleven**), I keep track of them by using name extensions to suexg (version x.x, top-down):
cb: controlled-bias
count: output (puzzle, #grids consumed) pairs instead of only puzzles. (I'm not sure it is still useful to have this, but who knows!)
optim46: delete the first 46 clues without doing any checking
stream: the 3rd argument is sdtin instead of a filname

This morning, I've launched the generation of the first 300 bands; 12 hours later, I now have 9; the 10th is running.
I've also launched "MIN 0 -" on the 2nd processor. It has been running for one hour, allowing a rough estimate of how many puzzles one can hope per day. My first estimate is 1700-1800 (2.66 GHz CPU), which is almost 10 times better than the first version of suexg-cb.

Now, we just have to wait for the generation process to complete.

Last edited by denis_berthier on Thu Oct 01, 2009 5:10 pm; edited 1 time in total

Back to top profile pm www

denis_berthier

Posted: Thu Oct 01, 2009 5:07 pm Post subject: quote edit

Joined: 19 Jun 2007

> **Allan Barker wrote:**
> I have placed the MonteCarlo grid generator here

Posts: 874
Location: Paris, France

> http://sudokuone.com/xsudo1/data.htm
> ....
> 1. Call run_montecarlo() to generate one grid.
> 2. Read the grid from SUD[9][9] as integers 1 through 9.

Allan, thanks.
As my understanding of C is more than limited, could you add one or two lines such that it outputs the puzzle to stdout, in the standard ..5..7.... form?

Has the user any possibility of specifying the RNG seed?
If we want many puzzles, a single pass through all of gsf's list will not be enough, we need several passes, with different seeds (or the guarantee that different runs will use different sequences of randoms numbers).

Back to top | profile | pm | www

Allan Barker

Posted: Fri Oct 02, 2009 1:36 am Post subject: quote

Joined: 21 Feb 2008
Posts: 346
Location: Bangkok

> **denis_berthier wrote:**
>
> > **Allan Barker wrote:**
> >
> > > I have placed the MonteCarlo grid generator here
> > > http://sudokuone.com/xsudo1/data.htm
> > > ....
> > > 1. Call run_montecarlo() to generate one grid.
> > > 2. Read the grid from SUD[9][9] as integers 1 through 9.
> >
> > Allan, thanks.
> > As my understanding of C is more than limited, could you add one or two lines such that it outputs the puzzle to stdout, in the standard ..5..7.... form?
> >
> > Has the user any possibility of specifying the RNG seed?
> > If we want many puzzles, a single pass through all of gsf's list will not be enough, we need several passes, with different seeds (or the guarantee that different runs will use different sequences of randoms numbers).

OK, it now has a suex.. like command line where you specify a seed and a count, eg.

montecar randomseed N
montercar randomseed N [> myfile] or [| myapp]

There is also a binary montecar.exe on the website so no need to compile.
If you *do* link you can seed the RNG with a call to

iseed((unsigned long long)myseed).

The built in RNG uses a 64 bit key. Not as good as a Mersenne Twister I'm sure but it uses as many bits.

Back to top profile pm www

denis_berthier

Posted: Fri Oct 02, 2009 4:41 am Post subject: quote edit

Joined: 19 Jun 2007
Posts: 874
Location: Paris, France

In ~ 12 hrs, based on only the first 9 gsf files, the controlled-bias generator has produced 994 minimals (a little more than my first expectations, which now rise to ~ 2000 per day), with the following distribution:
mean= 25.66
standard-deviation= 1.135

nb-clues nb-instances
21 0
22 4
23 15
24 114
25 320
26 336
27 147
28 51
29 6
30 1
31 0

This gives an estimated value for the number-of-clues unbiased average: 26.64
This is consistent with the mean obtained from suexg-cb (~ 0.1 standard deviations above it) considering that:
- the sample is very small,
- as explained by gsf, the subset 0-9.sudz is biased

Back to top profile pm www

denis_berthier

Posted: Fri Oct 02, 2009 5:46 am Post subject: quote edit

Joined: 19 Jun 2007
Posts: 874
Location: Paris, France

eleven wrote:

> in the 133302 grids of band 299 you can find 2 times more minimals than in the 1007170 grids of band 1 (?!) [Edit:]The difference is not that big (i had an error in the command lines), but with about 10mio grids each the average number of grids needed in band 1 was 379342, and in band 299 93986.

With suexg-cb, I also noticed a very large variability in this number (X). Here are my stats (based on 350000 - 14000 puzzles; I've been using the version with this count, since you implemented it):
E(X) = 224687
sd(X) = 224133 (as large as the mean)
min(X) = 1
max(X) = 3544330 (no error: 3.5 millions - having more than 1 million is not so rare)

With Mike, we also noticed (on a small sample) that this number was significantly higher with his program - but the mean number of clues was consistent with

suexg-cb. One explanation may be that the suexg internal generator of complete grids produces grids with more minimals than normal (~20% more, as argued by Red Ed [add reference]) and suexg-cb has 20% more chances of finding one in the mean.
The question now is, does the mean number of minimals in the generation phase have an impact on the number-of-clues distribution?
Red Ed showed that, replacing the suexg generator of complete grids by his own, the impact was infinitesimal [Edit: added reference http://www.sudoku.com/boards/viewtopic.php?t=14615&postdays=0&postorder=asc&start=193.]

So, maybe, band 299 has more minimals than band 1. I don't know if there is an easy way of checking this.

I haven't yet generated band 299; I currently have only the first 17 bands (as we need all of them, I dumbly generate them in ascending order). There's one interesting thing you could check about these two bands with so different behaviours: do they give you the same mean number of clues?

BTW, are you using a version of suexg-cb with the U4 test? If so, do you have an idea of how many puzzles you could get per day? Said otherwise, should I make it my standard version?

> **eleven wrote:**
> Denis, i never had a MAC...

It's just Unix (BSD) with a nice user interface in a nice aluminium box.

Last edited by denis_berthier on Fri Oct 02, 2009 6:41 am; edited 1 time in total

Back to top profile pm www

Red Ed

Joined: 06 Jun 2005
Posts: 748

Posted: Fri Oct 02, 2009 6:29 am Post subject: quote

> **denis_berthier wrote:**
> One explanation may be that the suexg internal generator of complete grids produces grids with more minimals than normal (~20% more, as argued by Red Ed [add reference]) and suexg-cb has 20% more chances of finding one in the mean.
> The question now is, does the mean number of minimals in the generation phase have an impact on the number-of-clues distribution?
> Red Ed showed that, replacing the suexg generator of complete grids by his own, the impact was infinitesimal.

The reference for both the 20% and "infinitesimal" (actually just very small) results is <here>.

Back to top profile pm

coloin

Posted: Fri Oct 02, 2009 9:37 am Post subject: quote

I believe bands 1-30 have a repeating minirow !

Joined: 06 May 2005
Posts: 1079
Location: Devon UK

This maybe explains a lot !

C

Back to top profile pm

eleven

Posted: Fri Oct 02, 2009 10:39 am Post subject: quote

Ah, interesting (but only 1.7 % are automorph).

Joined: 10 Feb 2008
Posts: 506

Maybe someone can check, if my UR4 function is correct. Runnning it at the band 299 puzzles, it only gave an average of 3.71818 UR4s/grid (9.26896 for the grids in band 1, 10.19 for 100k suexg grids, and 10.37 for suexg with prefilled diagonal).
You get the band 299 puzzles quickly by running gsf's program with
sudoku -gb299 -f '%#ec' -o 299.sudz
Then uncompress with
sudoku -q- -f%v 299.sudz > 299.dat

I had made another test comparing band 1 and 299, where i ran the sb program (as posted above with the MT RNG and U4 test) in parallel for the 2 grids files for 1 hour (3Gh dual core). The results were:

Code:

```
Band 1
23:       2
24:      14
25:      60
26:      66
27:      34
28:       3
179 puzzles, avg. clues 25.6983,
73.2 mio grids tested, avg tries for a puzzle 409165

Band 299
23:       5
24:      44
25:      87
26:      84
27:      35
28:       8
29:       3
266 puzzles, avg. clues 25.5113,
34.0 mio grids tested, avg tries for a puzzle 127193
```

This suggests, thet the 299 grids on average have more than 3 times more minimals with a lower clue average. You can see, that for the 299's the program scans less than half of the grids in the same time, because much fewer U4's were found.

Later (running 299 alone):

Code:

```
Band 1:
22:       0
```

```
23:       3
24:      21
25:      82
26:      83
27:      56
28:       7
29:       0
30:       1
100.5 mio grids, 253 puzzles, avg clues 25.7668, avg
tries 397450

Band 299:
22:       1
23:      20
24:     124
25:     272
26:     276
27:     122
28:      30
29:       4
30:       1
100 mio grids, 850 puzzles, avg clues 25.5459, avg
tries 117714
```

Rather different distributions. Note, that in both bands a 30 was found.

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

Goto page Previous 1, 2, 3 ... , 26, **27**, 28 Next

Page 27 of 28

Stop watching this topic

Jump to: General/puzzle 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