



## 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](#) ... , [37](#), [38](#), [39](#) [Next](#)

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

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

Author	Message
<b>denis_berthier</b>	<div style="text-align: right;"> <a href="#">quote</a> <a href="#">edit</a> </div>

Joined: 19 Jun 2007  
 Posts: 937  
 Location: Paris, France

**gsf,**

I have a problem when I run `./sudz -f%g bands/*.sudz | ./suexg-cb ....`

It stops systematically after band 252, with a sequence of messages:

```

sudz: decode error
sudz: bands/253.sudz: corrupt
sudz: decode error
sudz: bands/254.sudz: corrupt
  
```

These bands are not corrupt: I can re-launch the process on these same bands, copied in another directory. These bands represent a very small part of the whole collection (and, of course, the results reported in my previous post include them).

I suspect some limit on the number of parameters passed to a Unix process: here  $1 (-f%g) + \text{the first } 252 \text{ bands} = 256 - 3(\text{internal parameters?})$

You explained that `.sudz` files can't be catenated.

Will the following ksh script (adapted from yours), generate correctly (at least, if not interrupted) all the bands in range 250-300 in a single `250-300.sudz` file:

**Code:**

```

typeset -Z3 band
for ((band=250; band<301; band++))
do
  if [[ ! -f $$band.sudz ]]
  then
    ./sudoku-darwin.i386 -gb$band -f%#ec >> 250-300.tmp
  fi
done
mv 250-300.tmp 250-300.sudz
print -u2 250-300.sudz
  
```

Last edited by denis\_berthier on Thu Oct 15, 2009 8:25 am; edited 1 time in total

[Back to top](#)

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

**Red Ed**

Posted: Thu Oct 15, 2009 6:48 am Post subject:

[quote](#)

Joined: 06 Jun 2005  
 Posts: 803

After 35 hours, my program's estimate of the number of proper minimal 30s per grid is  $1.8615e13$  with standard deviation 3.80% of that. The precision is equivalent to about 96 full scans of gsf's collection -- so actually, for the purpose of estimating the number of proper minimal 30s, I'm effectively operating at "only" **65 full scans a day**, not quite 70 as previously reported.

For comparison, Denis' estimate is  $1.8483e13$  with standard error 4.55% of that.

[Back to top](#)

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

**denis\_berthier**

Posted: Thu Oct 15, 2009 7:49 am Post subject:

quote edit

Joined: 19 Jun 2007  
Posts: 937  
Location: Paris, France**Red Ed,**

Is there any chance you can get (in reasonable time) sets of say 100 n-clue minimals for n=31, 32, .... 40?

It'd be interesting to see how the mean SER varies when n becomes so large.

[Back to top](#)

profile pm www

**eleven**

Posted: Thu Oct 15, 2009 8:36 am Post subject:

quote

Joined: 10 Feb 2008  
Posts: 534

Red Ed,

many thanks for the explanations and the code. All i could do now was to compile it and run it for 1000 24's to get the 30's and 31's. its as fast as i expected. I hope to find time soon to look at it closer.

Would i get an unbiased 31 set, if i take out one 31 from each 24 superset containing 31's ?

[Back to top](#)

profile pm

**Pat**

Posted: Thu Oct 15, 2009 10:29 am Post subject:

quote

Joined: 18 Jul 2005  
Posts: 1565**denis\_berthier wrote:****Red Ed,**

Is there any chance you can get (in reasonable time) sets of say 100 n-clue minimals for n=31, 32, .... 40?

hi **denis\_berthier**,

i hope you don't really need those with a high number of clues !!

minimal puzzles with 38 clues are available (but not random)the first 2 minimal **39s** were found by **Håvard** (2007.Aug.26)[Back to top](#)

profile pm

**denis\_berthier**

Posted: Thu Oct 15, 2009 10:52 am Post subject:

quote edit

Joined: 19 Jun 2007  
Posts: 937  
Location: Paris, France**Pat wrote:****denis\_berthier wrote:****Red Ed,**

Is there any chance you can get (in reasonable time) sets of say 100 n-clue minimals for n=31, 32, .... 40?

i hope you don't really need those with a high number of clues

Hi Pat,

No, I don't need them. The necessary upper bound of what I need is 30 (as I explained in a previous post) and I have all that I need for very precise SER or NRCZT statistics.

But it may be of independent interest to know the mean SER for n-clue minimals, n&gt;31.

In particular, we know that there is a trend more clues =&gt; higher mean SER in the whole 22-30 range.

- Is the trend still present above 30? Collections produced by Allan indicate a positive answer, but this is not completely conclusive as they are not random.

- Below 20, famous collections of 17s and 18s also indicate that the trend is there, but same problem: they are not random.

[Back to top](#)

profile pm www

**gsf**

Posted: Thu Oct 15, 2009 2:32 pm Post subject:

quote

Joined: 22 Sep 2005  
Posts: 3930  
Location: NJ USA

**denis\_berthier wrote:**

**gsf,**

I have a problem when I run `./sudz -f%g bands/*.sudz | ./suexg-cb ....`  
It stops systematically after band 252, with a sequence of messages:  
sudz: decode error  
sudz: bands/253.sudz: corrupt  
sudz: decode error  
sudz: bands/254.sudz: corrupt

These bands are not corrupt: I can re-launch the process on these same bands, copied in another directory. These bands represent a very small part of the whole collection (and, of course, the results reported in my previous post include them).

I suspect some limit on the number of parameters passed to a Unix process: here 1 (-f%g) + the first 252 bands = 256 - 3(internal parameters?)

You explained that .sudz files can't be catenated.

there is a memory leak somewhere between the bzip2 library and sudz  
until that's located you can work around (avoiding regeneration) by changing the sudz pipeline head to:

**Code:**

```
{
./sudz -f%g bands/0*.sudz
./sudz -f%g bands/1*.sudz
./sudz -f%g bands/2*.sudz
./sudz -f%g bands/3*.sudz
} | ./suexg-cb ....
```

this assumes the first 99 bands are named 001.sudz .. 099.sudz

[Back to top](#)

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

**eleven**

Posted: Thu Oct 15, 2009 4:27 pm Post subject:

[quote](#)

I had noticed increasing memory of gsf's sudoku, when uncompressing \*.sudz.

Since i also want to be able to break and restart it without losing too much time, i modified gsf's script to something like

**Code:**

```
typeset -Z3 band
for ((band=1; band<=416; band++))
do
    if [ ! -f done$1_$band.x ]; then
        ./sudoku-darwin.i386 -gb $band.sudz -f%#ec | ./suexg-cb $1 1 - >
tmp.dat
        cat tmp.dat >> puzzles_$1.dat
        touch done$1_$band.x
    fi
done
rm done$1_*.x
```

It can be started then with the seed as parameter.

[Back to top](#)

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

**Red Ed**

Posted: Thu Oct 15, 2009 5:53 pm Post subject:

[quote](#)

**denis\_berthier wrote:**

**Red Ed,**

Is there any chance you can get (in reasonable time) sets of say 100 n-clue minimals for n=31, 32, .... 40?

31s - sure, I can generate a few tens of them every hour; groups will be correlated, but that doesn't prevent you getting an unbiased estimate of the mean. 32+ ... mmm ... don't know yet; probably very little beyond 32 clues. I doubt that I'll find any 40s!

After 11 hours or so, my program estimates the number of proper minimal 31s to be 1.0960e12 with

Joined: 10 Feb 2008  
Posts: 534

standard error 9.68% of that. Denis had 1.0115e12 with much larger uncertainty (as you'd expect). Anyway, the two estimates are consistent, which is good.

**Quote:**

- Below 20, famous collections of 17s and 18s also indicate that the trend is there, but same problem: they are not random.

It might be reasonable to assume that the 17s found so far are ... let's guess ... 95%(?) of all of them out there, so SER stats should be reasonably representative of the whole population of 17s.

[Back to top](#)



**Red Ed**

Posted: Thu Oct 15, 2009 6:35 pm Post subject:



Joined: 06 Jun 2005  
Posts: 803

**eleven wrote:**

Would i get an unbiased 31 set, if i take out one 31 from each 24 superset containing 31's ?

No, you have to treat all puzzles output on an equal footing. You can get an unbiased set, but not an uncorrelated set. To estimate the mean of  $F(\text{puzzle})$  where  $F$  is any function you like, e.g. SER, just compute the mean over all puzzles output by the code. To estimate the standard error, I'd suggest bootstrap resampling.

If there's a need for it, I could incorporate code to produce the relevant estimates conditional upon the user providing a function **double F(int (\*puz)[9])** .

[Back to top](#)



**Red Ed**

Posted: Thu Oct 15, 2009 6:53 pm Post subject:



Joined: 06 Jun 2005  
Posts: 803

Re 32+ clues ... I'm just doing a run looking for 32s now. Three found after an hour and a bit, so getting 100 32s is doable. But 100 33s would probably be completely beyond the current algorithm. Ah well.

EDIT: restarted it with new feature to save off puzzles, e.g. these ...

**Code:**

```
320701008000900000000063400840096702000000000100074900080410200734620100200030800
320701008000900000000063400840096702000000000100074900080410200734608100200030800
---
60070000000094206090086040040560071020600000417902060030018004500000007000074030
60070000000094206090086040040560071020600000417902060030008004500000007000274030
600700000000942060900860400405600710206000004079020600300080205000000107500074030
600700000000942060900860400405600710206000004079020600300080205000000107000274030
---
0000014008059620717018000690006100900070380400500900000408001208900605050000000
```

[Back to top](#)



**denis\_berthier**

Posted: Fri Oct 16, 2009 4:44 am Post subject:



Joined: 19 Jun 2007  
Posts: 937  
Location: Paris, France

**gsf, eleven,**

I think there are two different problems.

As eleven, I had noticed increasing memory usage.

Strangely enough, this increasing memory appears (on my Mac) as "inactive" and it is not released when the process ends (nor even when I kill the shell in which it was running). (Real memory keeps at ~ 5.6 Mb and virtual at ~ 22 Mb.)

As a result, before I launch a new scan of the full collection, I reboot.

The problem I mentioned to gsf seems to be different: it happens even if memory is still available. I've launched new computations with the new gsf script, I have to wait until the end to see if everything is fine.

**eleven**, doesn't your script suppose all the 300-416 bands have been expanded in separate files?

[Back to top](#)



Red Ed

Posted: Fri Oct 16, 2009 6:26 am Post subject:



Joined: 06 Jun 2005  
Posts: 803

Some 32s to be getting on with:

(Many more need to be generated before there are *effectively* 100, i.e. the standard error of the number-of-32s-per-grid estimate drops to 10%)

```

320701008000900000000006340084009670200000000100074900080410200734620100200030800
320701008000900000000006340084009670200000000100074900080410200734608100200030800
---
600700000000942060900860400405600710206000004179020600300180045000000007000074030
600700000000942060900860400405600710206000004179020600300080045000000007000274030
600700000000942060900860400405600710206000004079020600300080205000000107500074030
600700000000942060900860400405600710206000004079020600300080205000000107000274030
---
00000140080596207170180006900061009000703804005009000000408001208900605050000000
00103764000000000604091307002006003140023506300015420205009704410072005070050260
00103764000000000604091307002006003140023506300015420205009704410072005070050260
00103764000000000604091307002006003140023506300015420205009704410072005070050260
00103764000000000604091307002006003140023506300015420205009704410072005070050260
00103764000000000604091307002006003140023506300015420205009704410072005070050260
00103764000000000604091307002006003140023506300015420205009704410072005070050260
---
040051308010860070000703020050106800000090000009500700074000000036480907801600040
040051308010860070000703020050106000000090000109500700074000080036480907800600040
040051308010860070000703020000106800000090000009500700074000000036480907801600240
040051308010860070000703020000106000000090000109500700074000080036480907800600240
040001308010860070000703020000106800000090000009520700074000000036480907801600240
040001308000860479000703020050106092020090000000500700074000000036480907800600040
040001308000860479000703020003106090020090000000500700074010000036480907800600040
040001308000860070000743021050106002000090000000500730074010000036480907800600040
040001308000060479000703020050176092020090000000500700074000000036480907800600040
040001308000060479000703020003176090020090000000500700074010000036480907800600040
04000130800006047900070302000176092020090000000500700074010000036480907800600040
040001308000060470000703020050176090000090000009500706074009000036480907800600040
040001308000060070000743020050176090000090000009500706074009000036480907800600040
040001008010860070000743020000106800000090000009500730074000000036480907801600240
---
078940000000702009000060001500000007607300094000406205050097000000614503016530000
---
000800509900052060000000400364000000720010040109000206000620034430109652000030090
---
000040060400962370000800000300020010020410000106078204708090100602084003000007006
---
090040056000700409005000720050107000270590000109230000907020008520000030061070200
---
40000130800074000950206047004300800002001009010000080005018903000006005000530986
---
60000100000700000050070040908000020770203804610640008000009700009610004801274090
---
730001008500002370000000060400506000000104050100327084040005810007000045005008036
730001008500002300000000060470506000000104050100327084040005810007000045005008036
730001008500002300000000060470506000000104050000327084940005810007000045005008036
730001008500002300000000060400506000000104050000327084940005810007000045205008036
73000100850000230000000006040050600000004050100327084040205810007010045005008036
73000100850000230000000006040050600000004050000327084940005810007010045205008036
730001000500802370000000060400506000000104050100327084040005810007000045005008036
730001000500802300000000060470506000000104050100327084040005810007000045005008036
730001000500802300000000060470506000000104050000327084940005810007000045005008036
73000100050080230000000006040050600000004050100327084040205810007010045005008036
030001008500802300000000060470506000000104050000327084940005810007000045005008036
0300010085008023000000000604005061000000104050000327084940005810007000045005008036
030001008500802300000000060400506000000104050000327084940005810007000045205008036
03000100850080230000000006040050600000004050100327084040205810007010045005008036
---
382050060605000000010060008000006000006419050009578046050097002700120085000000700
---
420001039010900070000050401004000000082094003009075802000000180058017300900008007
420001039010000070000050401004000000082094003009075802000009180058017300900008007
---
200960008700850409000043000400100002820094700109000604080000200902007050300000907
000961008700850409000043000400100002820094700109000604080000200902007050300000907

```

```

---
520901000001002509000000000040610700070405080150200304780190205204506000000008000
---
00000070070940360901000000003078090080190043100400080508020000090307850710500004
---
054001300030740500906000001003079100000410083000005902090100000308097205000038000
054001300030700560906000001003009056000016703000005902090020000308097205000038000
054001300030000560906000001003079056000016703000005902090020000308097205000038000
---
006000070050740061004063509000004195900000040045000700097310000000520007021090800
006000070050740061004063509000004195900000040045000700007310000000520907021090800
---
030000500006000070000050468300020705002030040059400802600004007200005980000280654
---
70090000008100200090284007140610508080000405010928004050062081020000030000000005
---
30700102901070008000800370143000009280603000000004003000406200600017905701005000
---
500000029009702461002900000048100096000078040107409052000506010006000005900007000
500000020009702461000900500048100096000070040107409852000596010006300000900007000
074000029009702061000900500048105096000000040107409852000506010006010000900007000
070000029009702461002900508048100096000000040167409052000506010006000000900007000
0700000290097024610029005000481000960000080001074090520005060100060000009000207000
0700000290097024610029005000481000960000080001074090520005060100060000009000207030
070000029009702461002900500048100096000008000107409052000506010006000000900207030
070000029009702461002900500048100096000000040107409852000506010006000000900207000
070000029009702461002900500048100096000000040107409052000506010006000000900207030
070000020009702461600900078048100096000000040107409052000506010006000000951007000
000000029009702461002900578048100096000000040167409052000506010006000000900007000
000000029009702461002900508048100096000000040167409052000506010706000000900007000
000000029009702461002900508048100096000000040107409052000506010706010000900007000
000000029009702461002900500048100096000000040107409852000506010706010000900007000
000000029009702061600900508048100096000000040107409052080506010706010000900007000
000000020009702461002900508048105096000000040107409052000506017006000000901007000
000000020009702461002900500048105096000000040107409852000596017006000000900007000
000000020009702461002900500048105096000000040107409852000506017006000000901007000
000000020009702461002900500048100096000000040107409852000596017706000000900007000
000000020009702461002900500048100096000000040107409052000596017706300000900007000
---
03200100950980200100096000003000042000004100100200906408100000306008215050006004
---
008001009600002000027850400450100007206470105070000006005204700000530010002600050
008001009600002000027850000450100007206470105070000006065204703000530010002000050
---
000000400090800001410960508004008005028490013100230800050020000286007000000509600
000000400090800001410960508004008005028490013100230800050020000280307000000509600
000000400090800001410960508004008005028490013100230800000020000286007000040509600
000000400090800001410960508004008005028490013100230800000020000280307000040509600
000000400090800000410960508004008005028490013100035800050020000280307000000509602
000000400090800000410960508004008005028490013100035800000020000280307000040509602
---
35706002000870000100205000050307001020630507400042003000000007600000090701580240
35706002000870000100205000050300001020630507400042003000000007600007090701580240
---
06080170900000208082000000004100000090004130000029004780000043032497810001008200
---
00600042000486007071090050080720000006018004004500000000020000600419850051070290
006000420004860070710900000807200010009180740045000000000000000600419850051070290
00000042050486007071090050080720000006018004004500000000020000600419850051070290
000000420504860070710900500807200000060080040145000000000520000600419850001070290
00000042050486007071090000080720001006018004004500000000020000600419850051070290
00000042000486007071090000080720001000918074004500000000020000600419850051070290
---
004000570890700000010003409540087190008100000170600840300006914450010000000000050
004000570890700000010003409540080190008100700170600840300006914450010000000000050
---
000000020608740350000003706004009008006305940009600230040000813800000500503090002
000000020608740300000003706004009008006305940009600230040000813800000590503090002
---
000851309380000570009000000804210007090407000160005200400529013600104000000000002
---
803001000400802000010043000304000015082500043109004286908000050000000000071090032
803001000400802000010043000004000015082500043109004286938000050000000000071090032
---

```

```

30060009000900300090040000400100700050000036170036054004078210720510600000620007
00060009000900300090040000400100700050000036170036054004078210720510600500620007
---
03095046000080207009000350005321070082000400010900000000000010082100640010426030
---
043801009005042300600903020704000000850006000009400800096000130000010905001009240
043801009005042300000903020704000000850006000069400800096000130000010905001009240
040801009005042300600903020704000000850006000009400803096000130000010905001009240
---
52074036030680047000000300880000010260210005007000083040530000026740008000000600
---
0200000088040526005010000000521098000253910000080000056100290000600010000325860
02000000880405260050100000005210980002539000100080000056100290000600010000325860
000000008804052600001060420005210980002509000100080000056100290000600010000325860
000000008804052600001060020005210980002539000100080000056100290000600010000325860
---
520701300000000000001960520084007005072010080150090200000089700000000014010074950
---
003901420040000301800040060704100650086000010050000000007300100090000036018690240
---
080000400001040360000803070008090000090510837100000004060005902802009005015020780
---
00506102801005200070000000134000089027009005300903804000000005500619204000080600
00500102801005200070004000134000089027009005300903804000000005500619204000080600
0050010280100520007000400013400009027809005300903804000000005500619204000080600
---
030800409070900360090063508000000000000610840060084050006090200380200000210030680
030800409070900360090063008000500000000010840160084050706090200380200000210000680
030800409000900360091063528000000000000610800060084950056090204080200000010000680
---
000040300003002009080900021830079052500000000067405000008214030072500900340007000
000040300000062009080903420834070052500000000067005803000204030072000900340097000
000040300000002009080903021830079052500000000067405000008214030072500900340007000
000040300000002009080903021830079052500000000067005090000204030072030900340097080
0000400000003062009080903420834070052500000000067005803000204030072000900340097000
0000400000003062009080903420834070052500000000067005800000204030072500900340097000
0000400000003002009080900021830079052500000000067405003008214030072500900340007000

```

[Back to top](#)**denis\_berthier**

Posted: Fri Oct 16, 2009 6:51 am Post subject:



Joined: 19 Jun 2007  
 Posts: 937  
 Location: Paris, France

**Red Ed wrote:**

Some 32s to be getting on with:  
 (Many more need to be generated before there are *effectively* 100, i.e. the standard error of the number-of-32s-per-grid estimate drops to 10%)

Do you mean that there may still be some (relatively strong) bias in this collection?

If not, this is very unexpected: the mean SER is smaller than for the 30s (but the sd is higher)

30s: mean = 6.76, sd = 1.7 (based on 481 unbiased 30s)

32s: mean = 6.18, sd = 2.3

[Back to top](#)**Red Ed**

Posted: Fri Oct 16, 2009 7:14 am Post subject:



Joined: 06 Jun 2005  
 Posts: 803

Yes, the collection is *effectively* fairly small and so may happen to be clustered in an unrepresentative part of "32-space". The process by which the 32s are constructed is unbiased, but as we both know you need a lot of outputs before drawing conclusions with any certainty.

The estimate of the mean SER produced from any number of outputs will be unbiased.

However, quantifying confidence in that estimate is a bit fiddly: bootstrap resampling is one method.

[Back to top](#)

Display posts from previous:



General/puzzle

Goto page [Previous](#) [1](#), [2](#), [3](#) ... , [37](#), [38](#), [39](#) [Next](#)

Page 38 of 39

[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