# Sudoku Players' Forums



# THE REAL DISTRIBUTION OF MINIMAL PUZZLES

**Goto page Previous 1, 2, 3, 4 Next**

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

**View previous topic :: View next topic**

| Author | Message |
|---|---|
| **ronk**<br><br>Joined: 02 Nov 2005<br>Posts: 2392<br>Location: Southeastern USA | Posted: Sat Jul 11, 2009 8:04 pm Post subject: Re: THE REAL DISTRIBUTION OF MINIMAL PUZZLES quote<br><br>Re: **THE REAL DISTRIBUTION OF MINIMAL PUZZLES**<br><br>I don't see how this topic belongs in the **Advanced solving techniques** forum. If possible, would a moderator please consider moving it to the **Puzzle/general** forum<br><br>**Back to top** profile pm |
| **Red Ed**<br><br>Joined: 06 Jun 2005<br>Posts: 575 | Posted: Sat Jul 11, 2009 8:07 pm Post subject: quote<br><br>Denis, would you like to reconsider your **m(n) = s(n) * cf(n) / sum(cf(n))** ?<br><br>It's quite obviously wrong, since both s(n) and cf(n)/sum(cf(n)) are less than one.<br><br>EDIT: @ronk - good call.<br><br>**Back to top** profile pm |
| **denis_berthier**<br><br>Joined: 19 Jun 2007<br>Posts: 701<br>Location: Paris, France | Posted: Sat Jul 11, 2009 8:08 pm Post subject: Re: THE REAL DISTRIBUTION OF MINIMAL PUZZLES quote edit<br><br>**ronk wrote:**<br>Re: **THE REAL DISTRIBUTION OF MINIMAL PUZZLES**<br><br>I don't see how this topic belongs in the **Advanced solving techniques** forum. If possible, would a moderator please consider moving it to the **Puzzle/general** forum<br><br>I opened it here because of its relation with the "rating rules" thread, but Ronk is right. Puzzle/general would be a better place. |

Back to top profile pm www

**denis_berthier**

Posted: Sat Jul 11, 2009 8:11 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 701
Location: Paris, France

> **Red Ed wrote:**
>
> Denis, would you like to reconsider your **m(n) = s(n) * cf(n) / sum(cf(n))** ?

Done

Back to top profile pm www

**Red Ed**

Posted: Sat Jul 11, 2009 8:19 pm Post subject: quote

<ignore post just deleted - I think you've now given me the answer I needed>

Joined: 06 Jun 2005
Posts: 575

Back to top profile pm

**Red Ed**

Posted: Sat Jul 11, 2009 8:28 pm Post subject: quote

Joined: 06 Jun 2005
Posts: 575

Since I defined s(n) and m(n), let me be absolutely clear: they are both just constants, not random variables arising from estimation procedures or suchlike. I take it that your **MP**, **cf(n)** and **sum(cf(n))** are also just constants. (Whether or not we know their values is immaterial.)

Given that, I really don't see how **m(n) = MP * s(n) * cf(n) / sum(cf(n))** "may seem to be the same thing as your formula, but it isn't". I'm going to treat them as one and the same unless you wish to debate it further.

The result of this is that we appear now to have a very specific point of contention: that the probability that a top-down search stops at level n is proportional to m(n) * choose(81,n). I'll let the air clear for a moment before giving evidence against the bit in brown (yuck ... running out of nice colours).

Back to top profile pm

**denis_berthier**

Posted: Sat Jul 11, 2009 8:34 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 701
Location: Paris, France

> **Red Ed wrote:**
>
> we appear now to have a very specific point of contention: that the probability that a top-down search stops at level n is proportional to m(n) * choose(81,n). I'll let the air clear for a moment before giving evidence against the bit in brown

That's what I've been expecting from the beginning. Tell me before I go to sleep!

Back to top profile pm www

**Red Ed**

Posted: Sat Jul 11, 2009 8:39 pm Post subject: quote

I'll do it tomorrow.

Joined: 06 Jun 2005
Posts: 575

I need to spend more time thinking how to estimate the number of 29- and 30-clue minimal puzzles for my method (which, btw, I note that you have not commented on, or at least not objected to, so far ...).

Back to top profile pm

**denis_berthier**

Posted: Sat Jul 11, 2009 8:46 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 701
Location: Paris, France

**Red Ed wrote:**

> I'll do it tomorrow.
> I need to spend more time thinking how to estimate the number of 29- and 30-clue minimal puzzles for my method (which, btw, I note that you have not commented on, or at least not objected to, so far ...).

Sorry, I don't know what your method is. I've never before been interested in puzzle generation. My involvement in this is recent, it is related to the discussion on "rating rules/puzzles" and probably short term.

BTW, the discussio here was not about your method and you haven't yet provided any counterargument to my proof.
If you only plan to say: you are wrong because I'm right, that'll be a little short.
See you tomorrow, then.

PS: I've once more modified my post at the bottom of the previous page, in order to eliminate any ambiguities.

Back to top profile pm www

**Red Ed**

Posted: Sun Jul 12, 2009 9:33 am Post subject: quote

Joined: 06 Jun 2005
Posts: 575

Good morning, Denis.

Let's keep going ...

**Quote:**

> And any non-indexed puzzle at floor n has probability Pn = 1/ {(N * 81! / n!) / (81 - n)!} = 1/N * 1/81! * n! * (81 - n)! of being reached.

This is false. "Puzzles" (subgrids) at floor n will only be reached if the top-down searcher gets to floor n at all. Therefore, your **Pn** seems to have no relevance to the top-down search process that we are attempting to model. Since your whole premise is based around the claimed **Pn+1/Pn** ratios, I think we need to be clear about what you intend by the quoted statement above.

I would've thought the correct formula was **Pn** = 1/N * s(n)/m(n). Write m(n) = eps(n) * choose(81,n); then **Pn** = 1/N * 1/81! * n! * (81 - n)! * s(n)/eps(n). You appear to be assuming that s(n)/eps(n) is a constant.

Back to top profile pm

**denis_berthier**

Posted: Sun Jul 12, 2009 10:48 am Post subject: quote edit

Joined: 19 Jun 2007
Posts: 701
Location: Paris, France

**Red Ed wrote:**

> Good morning, Denis.

Bonjour, Red Ed. Here, it's already afternoon.

**Red Ed wrote:**

> **Quote:**
>
> > And any non-indexed puzzle at floor n has probability Pn = 1/{(N * 81! / n!) / (81 - n)!} = 1/N * 1/81! * n! * (81 - n)! of being reached.
>
> This is false. "Puzzles" (subgrids) at floor n will only be reached if the top-down searcher gets to floor n at all. Therefore, your **Pn** seems to have no relevance to the top-down search process that we are attempting to model.

You already asked the same thing yesterday.
What happens below B (i.e. after a minimal puzzle is reached) is irrelevant. All that we need are the probabilities on B and above.
But you can imagine that the top-down generator continues to go downwards after outputting the minimal puzzles. The formula is thus valid everywhere.
Mahematically, there is a random process going on and the generator preys on it for minimal puzzles. Wether you consider this process is stopped at B or not is irrelevant.
See my web pages for an updated version(http://www.carva.org/denis.berthier/HLS/Classification/index.html)

**Red Ed wrote:**

> I would've thought the correct formula was **Pn** = 1/N * s(n)/m(n). Write m(n) = eps(n) * choose(81,n); then **Pn** = 1/N * 1/81! * n! * (81 - n)! * s(n)/eps(n). You appear to be assuming that s(n)/eps(n) is a constant.

What is eps(n) ?

Back to top profile pm www

**coloin**

Posted: Sun Jul 12, 2009 11:58 am Post subject: quote

Joined: 06 May 2005
Posts: 1027
Location: Devon UK

Well, what a lot has been written in a short period. Im relieved that weve moved on to an extent and addressed the topic in ernest.

I cant fully follow the preceeding discussions [yet] though.

We have seen that puzzles can be generated from a full complete valid grid - removing clues to reach minimality.

Analysis of the clue distributions from 40-clue subgrids reveals how biased this sampling process is.

In terms of clues size and ? hardness.

You would think that statistically the chances of a [minimal] 24-puzzle being produced would be the same as any other [minimal] 24-puzzle - but I believe this is not the case. As the clue numbers reduce the pruning tree will be different [there will be an increasing number of "essential" "non-removable" clues as you advance] this will be specific to each of the many subpuzzles in the search tree.

From a 40-clue subgrid, investigating the inital puzzles produced, investigating the number of duplicates [ more in low-clue puzzles] might reveal that easier puzzles come out sooner and more often.

I will report back the results of these hypothyses. Although perhaps it is academic because I cant see us ever getting random puzzles this way.

C

Back to top profile pm

**denis_berthier**

Posted: Sun Jul 12, 2009 12:48 pm Post subject: quote edit

**Coloin**,
If I remember well, you once proposed your own estimation of the mean number of clues of a minimal puzzle.
But I can't find where and I don't remember your arguments.
Could you please recall both briefly?

Joined: 19 Jun 2007
Posts: 701
Location: Paris, France

Back to top profile pm www

**Red Ed**

Posted: Sun Jul 12, 2009 1:16 pm Post subject: quote

**denis_berthier wrote:**

> What happens below B (i.e. after a minimal puzzle is reached) is irrelevant. All that we need are the probabilities on B and above.

It is highly relevant. Your **P24** formula, for example, assumes that all choose(81,24) subgrids at level 24 are "reachable" ... but they're *not* - e.g. because some have multiple solutions. But you can't say that level 24 is "below B" and thus "irrelevant" because there are minimal puzzles at that level that whose probabilities we need to know.

Joined: 06 Jun 2005
Posts: 575

**denis_berthier wrote:**

> **Red Ed wrote:**
>
> > I would've thought the correct formula was **Pn** = 1/N * s(n)/m(n). Write m(n) = eps(n) * choose(81,n); then **Pn** = 1/N * 1/81! * n! * (81 - n)! * s(n)/eps(n). You appear to be assuming that s(n)/eps(n) is a constant.
>
> What is eps(n) ?

It's defined as m(n)/choose(81,n) in the second of my sentences quoted above. In words, it's the probability that a random n-clue subgrid is a minimal puzzle.

I still maintain that your **Pn** formula is out by a factor of s(n)/eps(n).

Back to top profile pm

**Red Ed**

Posted: Sun Jul 12, 2009 1:29 pm Post subject: quote

Joined: 06 Jun 2005
Posts: 575

**coloin wrote:**

> You would think that statistically the chances of a [minimal] 24-puzzle being produced would be the same as any other [minimal] 24-puzzle - but I believe this is not the case.

"Not the case" is correct. I didn't want to raise that myself in discussion with Denis because he talks about variables X(n) obtained by averaging over all n-clue puzzles in a sample; his focus appears not to be on the variance among individual puzzles.

Back to top profile pm

Display posts from previous: All Posts Oldest First 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