

Re: JSH: The will to lie

Source: <http://sci.tech-archive.net/Archive/sci.math/2006-09/msg00705.html>

- *From:* "Jesse F. Hughes" <jesse@xxxxxxxxxxxxxx>
 - *Date:* Mon, 04 Sep 2006 09:46:52 -0400
-

David Bernier <david250@xxxxxxxxxxxxxx> writes:

David Bernier wrote:

William Hughes wrote:

jstevh@xxxxxxx wrote:

<a long article in which he completely and conspicuously ignores the fact that a residue of 1 is more likely to be followed by a residue of 2>

How do you reconcile the observed fact that a residue of 1 is more likely to be followed by a residue of 2 with your claim that the sequence is random?

-William Hughes

I think it would be interesting to see by how much the sequence:
5-3*floor(5/3), ... 1000000007-3*floor(1000000007/3)
can be compressed by Winzip, Gzip, bzip2, etc.

In the primes mod 3 sequence, for primes > 3 and < 10⁷, we get a finite subsequence of 664,577 numbers in {1,2}.

I used a simplistic encoding 1 -> a byte '1', 2 -> a byte '2'.
A 664,577-byte file didn't even compress by a factor of 8 using the Winzip default. I think I got 84% reduction with the PPmD option; but that leaves 16%, or 0.16 > 1/8.

The obvious thing is to code in binary, and then compress. But I've put that on hold. Another possibility is to develop a custom compressor based on all of Tim Peters pretty considerable computations and comments. It's not clear to me how much pseudo-entropy is lost for subsequences of length $\sim \pi(10^9)$ using

Re: JSH: The will to lie

the correlations he found. I think a realistic probabilistic model for these correlations could be based on Markov chains, but I don't know how to compute entropies of Markov chains or processes.

Isn't the obvious compression algorithm for this sequence the one you give above?

The sequence is given by a simple algorithm. All you need to know is how many times you want to repeat that algorithm to generate the initial segment. So it seems to me that you can compress a sequence of length n to one of $\log_2(n) + C$ or so (where C accounts for the code that calculates the next prime and then mods by 3).

Or am I just mistaken on what you're doing. I assume you're looking at the Chaitin/Kolmogorov/Solovay complexity of the sequence to see how random it is in their sense. But since the sequence is defined by a simple algorithm, the answer is "not at all".

--

Jesse F. Hughes

"That's what's brutal about mathematics! When you're wrong, you can have spent years, and lots of effort, and come out at the end with nothing." -- James S. Harris on the path of self-discovery (?)

.