presenter: James Propp
title: The Ehrenfeuct-Mycielski sequence

problem:

Given a (possibly empty) finite binary sequence $(a)_1^{n-1} = (a_1, a_2, \ldots, a_{n-1})$ ($n \geq 1$), define its "maximal repeated suffix" as the longest terminal sub-word that occurs earlier in the sequence (possibly overlapping the terminal sub-word), i.e., as the longest (possibly empty) terminal sub-word of the form $(a)_{n-j}^{n-1}$ (with $j \geq 0$) such that $(a)_{n-j}^{n-1} = (a)_{n-j-i}^{n-1-i}$ for some $i > 0$; define the "predicted value" of a_n as a_{n-i}, where i is as small as possible. That is, one predicts that a_n will be the bit the occurred immediately after the most recent occurrence of the maximal repeated suffix in $(a)_1^{n-1}$

The Ehrenfeucht-Mycielski sequence is the sequence (grown from the seed 0) whose nth bit is **never** equal to the predicted value:

$$0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, \ldots$$

This appears in the OEIS as sequence A007061 (though in Sloane's version every term is increased by 1).

Klaus Sutner's article "The Ehrenfeucht-Mycielski sequence" (available on the web at `www.cs.cmu.edu/∼sutner/papers/em-sequence.ps.gz`) shows abundant evidence of order in this sequence (see Figure 4 for instance), but it remains unproved that the density of 1's in the sequence is equal to 1/2 (or even exists at all). Sutner does show that the upper density of 1's is at least 0.11.

Empirically, it seems not only that 1's occur with density 1/2, but every bit-string of length k occurs with density $1/2^k$.

One can create other sequences like the Ehrenfeucht-Mycielski sequence by starting with an arbitrary finite sequence as the initial seed and decreeing that all subsequent terms should differ from their predicted values. I looked at the four sequences grown from the seeds (0,0,0), (0,0,1), (0,1,0), and (0,1,1); the sequences appeared to be pairwise uncorrelated, in the sense that for any pair of them, the number of positions in the first n terms in which one sees a in the first sequence and b in the second (with a, b in $\{0, 1\}$) is close to $n/4$. (I did not use the other four seeds of length three because sequences grown from complementary seeds remain complementary and hence cannot be uncorrelated.)