# How to Prove the Kolakoski Sequence is Equidistributed

Noah MacAulay

June 2026

## 1 Introduction

So last month I wrote a paper[1] about how to find the transformation symmetry groups of a series of automata associated with the Kolakoski sequence. But my *real* goal was to make progress towards a proof of the density conjecture, namely that the Kolakoski sequence has an asymptotically equal ratio of 1s and 2s. Since it seems unlikely I will return to this problem in the foreseeable future I thought I'd write down the high level strategy and how the group theory paper relates to it, in case anybody(human or AI model) wants to try to have a crack at it.

## 2 Measures of Randomness under RLD

Terminology: an $n$-differentiable sequence is the output of $n$-iterated run-length decoding(see my paper).

So here's the thought process that led to my current high-level strategy:

(a) Instead of proving that the Kolakoski sequence is equidistributed, we'll consider *all* $n$-differentiable sequences, and have the goal of proving that sufficiently long such sequences have some bound on closeness to equidistribution that goes to zero as $n$ increases. Then since the Kolakoski sequence is infinitely long and $n$-differentiable for all $n$, it must be equidistributed. This is how Chvatal derived his bounds. But for this to work for $n \to \infty$ we need to use induction.

(b) OK, so we'd like to use induction. But equidistribution(of 1s and 2s) does not imply equidistribution in the output string(e.g. 121212121212... $\to$ 122122122122122...). So we need some sort of higher order notion of randomness that we will prove increases under RLD. One obvious idea is "correlation between parity of position in string and 1/2", which translates to equidistribution after 1 iteration of RLD. But that itself isn't closed under RLD.

(c) Hmm, maybe we need to prove that 1/2 are equidistributed mod 3, 4, 5 also? Fourier randomness? But the problem with this is that $n$-differentiable things are *not* truly Fourier randomn. For instance there are no strings of "111", etc. The (likely) true correlation spectrum is studied in [2], it's pretty weird looking.

(d) So we need some sort of measure of "Kolakoski-randomness". Then we prove that it monotonically increases under RLD, approaching "maximal" as $n$ increases, and prove that this implies standard equidistribution(of 1s and 2s)

# 3 Group Theory is the Path to the Correct Measure of Randomness

Now once you start thinking about $n$-differentiable sequences it becomes clear that the "automaton" structure(desciribed in my paper) is the right way of thinking about it. Shen [2] used this to define an infinite family of notions of randomness: if you feed this string to a depth-$k$ automaton, how close are the states of the automaton to being equidistributed.

This is a good start. But it's still hard to prove anything here since the automata are pretty inscrutable objects, and it's not clear how to prove anything inductively. However! As outlined in my paper [1], the automata are permutation automata with a comprehensible symmetry group.

This suggests a new measure of randomness. Given a string $x$, we can consider its prefixes, and the group element associated to each one. (e.g. given the string 11221211 we consider a distribution over group elements given by the group element associated with 1, with 11, with 112, with 1122, with 11221....)

So we could formulate a conjecture that these distributions should approach uniformity as we iterated run-length decoding. There's a bunch of free parameters here, however. The full groups $\mathcal{K}_{1,2}^{\infty}$ and $\mathcal{J}_{1,2}^{\infty}$ are infinite so it's not clear what measure of uniformity to use.

One obvious idea is to consider the finite-level groups, consider how equidistributed your elements are on each finite group, and do some sort of double induction. Not totally clear how to proceed here though. I like the idea of using representation theory a bit better.

# 4 Representation Theory

I suspect the pathway to the right randomness measure goes through the finite-dimensional representations of $\mathcal{K}_{1,2}^{\infty}$ and $\mathcal{J}_{1,2}^{\infty}$. These can probably be determined inductively in a straightforward fashion.

Representation theory is often used to prove bounds on how quickly group elements approach "equilibrium" under a random walk, e.g. [4]. I believe this it due to its control over group convolutions, a la Fourier analysis.

Now, the next step after determining the irreps would be to analyze how they behave(as distributions) under run-length decoding. There are likely intricate relations between the distributions of different irreps.

Another conjecture I have(might be necessary to prove equidistribution of group elements) is that the distributions of group elements associated with all possible $2^n$ $n$-RLDs of 1, become uniformly distributed as $n$ approaches infinity. This is not the same as the previous conjecture – I'm saying, e.g. the set of all 2-RLDs of 1 is $\{11, 22, 1, 2\}$. We consider group elements associated with each *entire* string, not the prefixes. The set of *those* things becomes equidistributed(by whatever measure) as $n$ approaches infinity. Proving the whole thing may require some sort of intricate multi-way induction using this conjecture, the previous one, and the structure of the set of irreps.

Anyway I believe that these ideas have great promise as a path to a proof of Kolakoski equidistribution. I think skillful exploration could accomplish this in a few months to a year. So ambitious grad students, undergrads or amateurs may want to give it a try.

# References


[1] Noah MacAulay. "Group Theory of the Kolakoski Sequence". https://arxiv.org/abs/2605.14234

[2] Bobby Shen. "A uniformness conjecture of the Kolakoski sequence, graph connectivity, and correlations". https://arxiv.org/abs/1703.00180

[3] Vasek Chvatal. "Notes on the Kolakoski Sequence" https://users.encs.concordia.ca/ chvatal/93-84.pdf

[4] Persi Diaconis and Mehrdad Shahshahani. "Generating a Random Permutation with Random Transpositions"